Project Review

Consortium Project on
Seismic Inverse Methods for Complex Structures
Breckenridge, Colorado
May 7-10, 1996

prepared by

CWP Students and Research Leaders:

AbdulFattah Al-Dajani  Vladimir Grechka  Björn Rommel
Tariq Alkhalifah  Herman Jaramillo  Andreas Rüger
Gabriel Alvarez  Edward Jenner  Trino Salinas
Norm Bleistein  Ken Larner  John A. Scales
Tong Chen  Zhaobo Meng  John Stockwell
Jack K. Cohen  Fernando Moraes  Ilya Tsvankin
Maarten V. de Hoop  Alejandro Murillo  Alberto Villarreal
H. Lydia Deng  Konstantin Osypov  Lan Wang
Tagir Galikeev  Gabriel Perez  Meng Xu

CWP Staff:

Jo Ann Fink  Ellen Yarnell  Barbara McLennon
Project Assistant  Administrative Assistant  Publications Specialist

Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401
Acknowledgments

This project review is prepared for the sponsors of the Consortium Project at the Center for Wave Phenomena. The Consortium Project provides substantial funding for the overall research and educational program at the Center. The Center also receives funds from the following government agencies:

- Office of Naval Research
  - Ocean Acoustics Division
  - Applied Analysis Division
- Army Research Office
- National Science Foundation
- Department of Energy,
  Advanced Computational Technology Initiative joint research with:
  - Los Alamos National Laboratory
  - Lawrence Livermore National Laboratory
  - Oak Ridge National Laboratory

We are extremely grateful for the support of these agencies and for that of the Consortium sponsors who are listed here and at the back of this book.

Amerada Hess Corporation             Landmark Advance Products Group
Amoco Production Company            Marathon Oil Company
Aramco                                Maxus Energy
BP Exploration Inc.                   Norsk Hydro A.S.
Bureau of Geophysical Prospecting    Oryx Energy Company
Compagnie Générale de Géophysique    Pennzoil Exploration & Production
Chevron Petroleum Technology         PGS Tensor Geophysical Company
Conoco Inc.                          Shell E & P Technology Company
Elf Aquitaine                         Silicon Graphics Inc.
Exxon Production Research Company    Statoil
Geco-Prakla Geophysical Co.           Talisman Energy Inc.
IBM Corporation                      Texaco USA
Instituto Mexicano del Petróleo      Union Pacific Resources
Intevep S.A.                          Unocal
Japex                                 Western Geophysical
Introduction

This is the twelfth edition of the Project Review report on the Consortium Project at the Center for Wave Phenomena. By almost any measure, this has been the most successful year in our history. One only has to browse the Table of Contents to realize that there are about 50% more papers in this Project Review book than there were in last year's. Because of the number of papers, we have organized them by topic. A scan of the titles reveals both the breadth and the depth of our program. At the Project Review meeting, there will be additional papers available that were completed too late to make it into this volume.

For the first time, we will use a poster session on the first afternoon of the meeting to supplement the two days of talks; we simply have too much that we want to report to our sponsorship to squeeze into our standard two days of presentations.

Center support

The support base for the Center has dramatically increased. In particular, five new companies have joined the Consortium project since we reported our status at last year's meeting: Instituto Mexicano del Petróleo, Maxus Energy, Pennzoil, Silicon Graphics, and Talisman Energy. Furthermore, Landmark/Advance has rejoined the project. A full list of sponsors appears on the acknowledgement page at the beginning of this volume and a list of technical contacts appears at the end of this volume. Our government support has also increased significantly. A list of those sponsors appears in the acknowledgement page at the beginning of this volume.

Our industrial support and our government support for both research and education complement one another; each gains from—and strengthens—the other.

We have received funding from the National Science Foundation for computer equipment, as part of a joint proposal between John Scales and four people in the CSM Department of Mathematical and Computer Sciences. For CWP, the biggest impact has been the purchase of an SGI Power Challenge. This is a 4-processor machine with a total of 384 MBytes of memory. In addition, we have a dual-channel 32 GByte disk vault. SGI has lent us an Indigo-2 XZ workstation to support this system.

We have been honored with funding from the SEG Foundation to support Samizdat Press, managed at CWP by John Scales and from a distance by Martin Smith. The SEG Foundation has also provided funding to support the management of Seismic Unix (SU) by Jack K. Cohen and John Stockwell. We received partial support from Blackhawk Geophysical for a student. We have received funding from Bee Bednar, personally, along with a matching grant from Amerada Hess in support of Samizdat Press.

CWP on the Web

Our CWP World Wide Web site is accessed by 7000 to 9000 hits per week (are there that many geophysicists out there?), while Samizdat Press —our online publishing arm—also receives around 8000 contacts per week. Some of this increase is due to our mirroring the GOCAD page.

---

1This increase has dictated a two-column format and smaller font to produce a volume that is still manageable. Feedback on this experimental format would be appreciated.
On the other hand, these numbers exclude access from within CWP. Also, they do not include hits on a mirror site of ours running at Edinburgh.

Computing environment

Last year, we indicated that we were starting the process of upgrading our local area computer network. Since then, we have purchased fifteen pentium PC’s (90mhz - 120mhz) and also have obtained the SGI equipment mentioned above. We expect all of these machines to be upgraded to 64 MBytes by the time of this meeting. We still have 28 functioning NeXTs, but it is a race to stay ahead, as this number dwindles. We are grateful for John Stockwell’s cannibalizing instincts that keep these machines running.

Samizdat Press

The number of available free online publications in Samizdat Press is increasing. New material includes Ilya Tsivankin’s course notes, Darrell Whitley’s GA tutorial, a new version of Brian Kennett’s continuum notes with figures, a new version of Scales and Smith’s inverse theory book (and software). Furthermore, in mid-May there will be a new book by John Wahr on geodesy and geodynamics, and a set of lecture notes by George Backus is in preparation. Check it out at samizdat@dix.mines.edu or at www.cwp.mines.edu/bookshelf.html.

SU

SU now boasts 535 sites in 35 countries. Also, the testimonials we receive confirm that these are not all casual acquisitions. People are using SU and providing feedback to Jack Cohen and John Stockwell that indicates wide acceptance and approval, despite SU’s limitations as compared to full-fledged commercial packages.

SU has a mirror site in South Korea maintained by Sang-yong Suh (sysuh@kigam.re.kr). Access is through ftp.kigam.re.kr.

Freeware

SU and the CWP Objected Oriented Optimatization library (COOOL), as well as other software, are available on line. Browse the offerings at timna.mines.edu/cwpccodes/.

Visitors

John Anderson, of Mobil, who was a visitor here for two years, successfully defended his thesis here at Mines in April. Thus, what was an accepted “common law” reality for years has now been confirmed with an official title.

Fabien Bosquet recently completed his eighteen-month visit with us and returned to Elf-Accuitaine. He left behind a culture imbued with the virtues of GOCAD as well as some wonderful software. We are taking full advantage of it.

Currently, Martijn de Hoop is in residence as a visiting faculty member in the Department of Mathematical and Computer Sciences. He has taught a course this semester on his approach to inverse problems. We anticipate a set of lecture notes for this course to be published as a report sometime in the fall.
Three post-docs have also joined us: Björn Rommel, Vladimir Grechka and Konstantin Osypov.

Students

CWP supports eleven students with funds from the Consortium project and government research contracts. Four additional students working with CWP faculty have support from companies.

Publications

Since the last Project Review meeting, seven research papers and one set of lecture notes have been published. In addition to the papers in last year's and this year's Project Review books, ten research reports have been written and are in various stages of submission, review, and/or acceptance in journals or proceedings. The most recent reports, papers and software are listed at the end of this introduction. The list includes five Master's and two PhD theses.

We anticipate an exciting, information-packed Project Review meeting, with distinguished visitors and guest speakers. In particular, some of our ACTI colleagues from the national laboratories will attend, so that they will be exposed to the full scope of our research program and meet our industrial sponsors.

Norman Bleistein, Director
Center for Wave Phenomena
April 1996

Papers and Reports Distributed


Fei, T., and Larner, K., 1995, Elimination of numerical dispersion in finite-difference modeling and migration by flux-corrected transport: Geophysics, 60, no. 6, 1830-1842.


Uzcategui, O., 1995, 2-D depth migration in transversely isotropic media using explicit operators: Geophysics, vol. 60, no. 6, 1819-1829.


Cohen, J.K., 1995, Analytic study of the effective parameters for determination of the NMO velocity function in transversely isotropic media: submitted to Geophysics, 12/9/95. CWP-191P.


Alkhalifah, T., 1996, Analytic insights into the anisotropy parameter $\eta$.

Velocity analysis using nonhyperbolic moveout in transversely isotropic media

Tariq Alkhalifah  
Center for Wave Phenomena  
Colorado School of Mines

ABSTRACT

P-wave reflections from horizontal interfaces in transversely isotropic (TI) media have nonhyperbolic moveout. Alkhalifah and Tsvankin (1993) showed that such moveout as well as all time-related processing in TI media with a vertical symmetry axis (VTI media) depend on only two parameters, called $V_{rmo}$ and $\eta$. They also showed that these two parameters can be estimated from the P-wave dip-moveout behavior of surface seismic data.

Alternatively, one could use the nonhyperbolic moveout for parameter estimation. The quality of resulting estimates depends largely on the departure of the moveout from hyperbolic and its sensitivity to the estimated parameters. The size of the nonhyperbolic moveout in TI media is primarily dependent on the anisotropy parameter $\eta$. An “effective” version of this parameter provides a useful measure of the nonhyperbolic moveout even in isotropic $v(z)$ media. Moreover, effective $\eta$, $\eta_{eff}$, is used to show that nonhyperbolic moveout associated with typical TI media (e.g., shales, with $\eta \approx 0.1$) is larger than that associated with typical $v(z)$ isotropic media. The departure of moveout from hyperbolic is increased when typical anisotropy is combined with vertical inhomogeneity.

The stability of the nonhyperbolic-moveout inversion depends largely on the range of offsets used in the inversion procedure. Larger offset-to-depth ratios ($X/D$) provide more nonhyperbolic information, and therefore, increased stability and resolution in the inversion. The $X/D$ values (e.g., $X/D > 1.5$) needed for obtaining stability and resolution are within conventional acquisition limits, especially for early zero-offset times.

Although estimation of $\eta$ using nonhyperbolic moveouts is not as stable as that by the dip-moveout method of Alkhalifah and Tsvankin (1993), particularly in the absence of large offsets, it does offer some flexibility: it can be applied in the absence of dipping reflectors and may also be used to estimate lateral $\eta$ variations. Application of the nonhyperbolic inversion to data from offshore Africa demonstrates its usefulness, especially in estimating lateral and vertical variations in $\eta$.

Key words: transverse isotropy, nonhyperbolic moveout, velocity analysis

Introduction

Often, the earth’s subsurface is dominated by interfaces that are horizontal or sub-horizontal. Therefore, reflections from such interfaces are the primary source of information in much seismic data. Moveout from sub-horizontal reflectors, for example, provides useful velocity information. Fortunately, reflection moveouts from horizontal interfaces are generally well represented by truncated Taylor’s series-type characterizations of moveout in transversely isotropic (TI) media with vertical symmetry axis (Hake et al., 1984; Tsvankin and Thom-
Alkhalifah and Tsvankin (1995) demonstrated that, for TI media with vertical symmetry axis (VTI media), just two parameters are sufficient for performing all time-related processing, such as normal moveout (NMO) correction (including non-hyperbolic moveout correction, if necessary), dip-moveout (DMO) correction, and prestack and poststack time migration. Taking $V_P$ to be the $P$-wave velocity in the horizontal direction, the two anisotropy parameters are

$$\eta \equiv 0.5 \left( \frac{V_P^2}{V_{nmo}^2} - 1 \right) = \frac{\epsilon - \delta}{1 - 2\delta},$$

(1)

and the short-spread NMO velocity for a horizontal reflector

$$V_{nmo} = V_P \sqrt{1 - 2\eta}.$$  

(2)

Here, $V_P$ is the $P$-wave vertical velocity, and $\epsilon$ and $\delta$ are two of Thomsen's (1986) dimensionless anisotropy parameters.

Moreover, Alkhalifah and Tsvankin (1995) show that these two parameters, $\eta$ and $V_{nmo}$, are obtainable solely from surface seismic $P$-wave data, specifically from estimates of stacking velocity for reflections from interfaces having two distinct dips (the DMO method). The two-parameter representation and inversion also holds in $\psi(z)$ media. Alkhalifah (1996b) used the DMO inversion method to invert for vertical variations in $\eta$. However, the DMO inversion of Alkhalifah and Tsvankin (1995) and Alkhalifah (1996b) works only when reflectors with at least two distinct dips (e.g., a fault and a gently dipping reflector) are present, as long as one of the dips is not close to 90 degrees.

Hake et al. (1984) derived the three-term Taylor's series expansion of the reflection moveout from horizontal reflectors in VTI media. The presence of the third term in their expansion implies nonhyperbolic moveout. Tsvankin and Thomsen (1994) recast the three-term expansion as a function of Thomsen's (1986) parameters, which provided a more compact equation. Moreover, using an asymptotic fit, Tsvankin and Thomsen suggested a correction factor that approximates the deleted higher-order terms of the Taylor's series expansion, thus stabilizing the moveout at long offsets. Tsvankin and Thomsen (1995) studied the problem of inverting for Thomsen's anisotropy parameters ($V_P$, $\epsilon$, and $\delta$) using the nonhyperbolic moveout of reflections from horizontal interfaces. They found that such an inversion using only $P$-wave data would be highly ill-conditioned because of the tradeoff between the vertical velocity and anisotropic coefficients, which cannot be overcome by using even twice the spread-to-depth length, which is in agreement with the two-parameter dependency of time-related processing (Alkhalifah and Tsvankin, 1995). They also point to the ambiguity in resolving the second- ($A_2$) and fourth order ($A_4$) coefficients of the Taylor's series expansion using traveltime moveout for ray angles up to 45 degrees. Although such ambiguity exists in the general sense of using traveltimes to invert for $A_2$ and $A_4$, it can be somewhat overcome by extracting $A_2$, which is simply the reciprocal of the NMO velocity squared, from conventional velocity semblance analysis, and in turn using it in the inversion for $A_4$, or in my case $\eta$. In fact, as we will see later, a 2-D semblance scan over both parameters proves to be a reliable method. Neidell and Taner (1971) have stated the clear benefits of semblance analysis for parameter extraction.

Byun et al. (1989) applied a two-parameter velocity analysis on synthetic vertical-seismic-profile (VSP) data using a "skewed" hyperbolic moveout formula for horizontal reflectors. Although their velocity-analysis approach showed promise, their nonhyperbolic (or skewed hyperbolic) formula was a coarse approximation to the actual moveout in TI media (Tsvankin and Thomsen, 1994). For example, their formula required knowledge of the vertical $P$-and $S$-wave velocities, whereas the true moveout is very much independent of these two parameters (Alkhalifah and Tsvankin, 1995).

Use of the deviation of moveout from hyperbolic for parameter estimation in general depends on the size of the deviations, as well as, on the sensitivity of the nonhyperbolic moveout to the estimated parameters and the absence of complicating factors such as lateral velocity variation. Here, I compare the size of nonhyperbolic moveout reflections from horizontal interfaces in VTI media with those associated with typical vertically inhomogeneous isotropic media. Then, I invert for estimates of $\eta$ using the nonhyperbolic moveout, and discuss the sensitivity of the inversion to errors in the measured parameters, namely $V_{nmo}$ and traveltime. I also apply semblance analyses over nonhyperbolic trajectories to estimate both $V_{nmo}$ and $\eta$. The study includes field data applications that exemplify the usefulness of this method.

### Nonhyperbolic Moveout in Layered Media

Hake et al. (1984) derived a three-term Taylor's series expansion for the moveout of reflections from horizontal interfaces in homogeneous, VTI media. If we ignore the contribution of the vertical shear-wave velocity, $V_S$, which is negligible (Tsvankin and Thomsen, 1994; Alkhalifah and Lerner, 1994; Tsvankin, 1995; Alkhalifah, 1996c), their equation can be simplified when expressed in terms of $\eta$ and $V_{nmo}$, as follows:
\[ t^2(X) = t_0^2 + \frac{X^2}{V_{nmo}^2} - \frac{2\eta X^4}{t_0^2 V_{nmo}^2 (1 + 2\eta X^2)}. \]  \tag{3}

Here, \( t \) is the total travel time, \( t_0 \) is the two-way zero-offset traveltime, and \( X \) is the offset. The first two terms on the right correspond to the hyperbolic portion of the moveout, whereas the third term approximates the non-hyperbolic contribution. Note that the third term (fourth-order in \( X \)) is proportional to the anisotropy parameter \( \eta \), which therefore controls nonhyperbolic moveout directly.

Tsankin and Thomsen (1995) derived a correction factor to the nonhyperbolic term of Hake et al. (1984) equation that increases the accuracy and stabilizes traveltime moveout at large offsets in VTI media. The more accurate moveout equation, when expressed in terms of \( \eta \) and \( V_{nmo} \), is given by

\[ t^2(X) = t_0^2 + \frac{X^2}{V_{nmo}^2} + \frac{2\eta X^4}{t_0^2 V_{nmo}^2 (1 + AX^2)}, \]  \tag{4}

where

\[ A = \frac{2\eta}{t_0^2 V_{nmo}^2 \left( \frac{V_H^2}{V_{nmo}^2} - \frac{V_T^2}{V_{nmo}^2} \right)}, \]

and the horizontal velocity, \( V_h \), from equation (1) for homogeneous media, is given by

\[ V_h = V_{nmo} \sqrt{1 + 2\eta}. \]

Through simple manipulation, equation (4) reduces to

\[ t^2(X) = t_0^2 + \frac{X^2}{V_{nmo}^2} - \frac{2\eta X^4}{t_0^2 V_{nmo}^2 \left[ t_0^2 V_{nmo}^2 (1 + 2\eta X^2) \right]^2}, \]  \tag{5}

(Alkhalifah and Tsankin, 1995). The difference between equations (4) and (5) will become evident in the \( v(x) \) case, where \( V_h \) will be defined in two different ways. Equation (5) is accurate for large offset-to-depth ratios. Note that, by setting \( X = 0 \) in the denominator of the fourth-order term, equation (5) reduces to equation (3). The additional \( X \) factor in the denominator produces an expansion that approximates the influence of the terms (beyond the fourth-order) that were omitted in the Taylor's series expansion, and therefore increases the moveout accuracy at large offset. The higher-order approximation is based on the fact that as \( X \) becomes very large (goes to infinity), while \( t_0 \) is finite, equation (3) reduces to

\[ t^2(X) = \frac{X^2}{V_h^2}. \]  \tag{6}

Therefore, equation (6) is asymptotically exact, because the ray path in this case is horizontal.

Equations (3) and (4) can be used as well in layered media, with a small-offset approximation of the type made by Dix (1955). Hake et al. (1984) and Tsankin and Thomsen (1994) provided key equations for moveout in layered VTI media, but in terms of conventional elastic coefficients and Thomsen's (1986) parameters, respectively. Here, I recast their expressions in terms of the practical anisotropy parameters \( \eta \) and \( V_{nmo} \).

First, as usual, the normal-moveout velocity involves a root-mean-squared (rms) average of velocities in the previous layers. Specifically

\[ v_{nmo}^2(t_0) = \frac{1}{t_0} \int_0^{t_0} v_{nmo}^2(\tau) d\tau, \]  \tag{7}

where all lower-case variables \( V \), including \( v_{nmo} \), correspond to interval-velocity values, and the integration is over time for the vertical raypath. Therefore, \( v_{nmo} \) is the interval NMO velocity given by

\[ v_{nmo}(\tau) = v(\tau) \sqrt{1 + 2\delta(\tau)}, \]

and \( v(\tau) \) is interval vertical velocity. On the other hand, upper-case variables \( V \), including \( V_{nmo} \), correspond to quantities that are averages for the entire vertical column from the surface to the reflector of interest. (Recall, here \( V_{nmo} \) refers to the NMO velocity for horizontal or near horizontal reflectors.)

Next, starting with Tsankin and Thomsen's expansion for the coefficient of the fourth-order term in Hake's equation for moveout in layered VTI media, in Appendix A I find that equations (3) and (4) continue to hold, with \( \eta \) replaced by

\[ \eta_{eff}(t_0) = \frac{1}{2} \left( \frac{\eta}{1 + 2\eta} \right) \int_0^{t_0} v_{nmo}^4(\tau) \left[ 1 + 8\eta(\tau) \right] d\tau - 1. \]  \tag{8}

Here, \( \eta(\tau) \) is the instantaneous value of the anisotropy parameter \( \eta \) as a function of the vertical reflection time.

In homogeneous isotropic media \( \eta(\tau) = 0 \), expressions (7) and (8) inserted into equation (3) reduce to the familiar three-term expansion given by Taner and Koehler (1969), as shown in Appendix B. Now, in equation (4) Tsankin and Thomsen (1994) suggested that \( V_h \) be computed using the following rms relation

\[ V_h^2(t_0) = \frac{1}{t_0} \int_0^{t_0} v_h^2(\tau) d\tau, \]  \tag{9}

where

\[ v_h(\tau) = v_{nmo}(\tau) \sqrt{1 + 2\eta(\tau)}. \]

Note that, in this case, equation (4) is described by three effective parameters \( (V_{nmo}, V_h, \text{and} \eta_{eff}) \). This would complicate velocity analysis and inversion applications requiring, among other things, three-parameter searches. On the other hand, a slightly different definition of \( V_h \), given by

\[ V_h(t_0) = V_{nmo}(t_0) \sqrt{1 + 2\eta_{eff}(t_0)}, \]  \tag{10}
will reduce the number of effective parameters to two \((V_{\text{mne}}\text{ and } \eta_{\text{eff}})\), simplifying the equation for later uses. In this case, equation (5) holds for layered media with \(\eta\) replaced by \(\eta_{\text{eff}}\), which is computed using equation (8).

The right side of Figure 1 shows the percent error in the computed moveout corresponding to reflections from (a) the bottom of the second layer, and (b) the bottom of the third layer in the model shown on the left side of the figure. Clearly, for \(X/D < 2\), the moveout corresponding to \(\eta_{\text{eff}}\), calculated using equation (3) (gray curve) has smaller error (better approximates the exact moveout) than does the moveout corresponding to equation (4), using \(V_6\) calculated from equation (9) (black curve). Both approximations give better results than does moveout described by equation (3) (dashed curve), modified from Hake et al. (1984). \(V_{\text{mne}}\) for all three approximate curves is the same, and is calculated using equation (7). Therefore, not only did the modified \(V_6\) expression (10) simplify the problem by reducing the number of required effective parameters, it apparently provided a better approximation to the exact moveout. Although only one example is shown here, this conclusion holds for many other \(v(\tau)\) VTI models tested.

A definition of \(V_6\) that is more in line with the asymptotic approximation used to produce equation (4) is to take \(V_6\) as the maximum horizontal velocity among the overlaying layers. Such a definition of \(V_6\), however, is not practical for typical seismic spreads because, although the definition is accurate asymptotically, it can considerably overestimate \(V_6\) at practical offsets (e.g., for models that include a thin layer with a high \(v_6\)).

Isotropic media are simply a subset of VTI media, where \(\epsilon = \delta = 0\). Therefore, equations (3) and (5) can be used to approximate moveout in isotropic layered media. Thus, although the anisotropy parameter \(\eta(\tau)\) equals zero throughout, because the medium is inhomogeneous, \(\eta_{\text{eff}}(t_0)\), as given by equation (8), is nonzero. In fact, equation (3) reduces to the familiar three-term expression given by Taner and Koehler (1969) for isotropic media (see Appendix B). Therefore, the value of \(\eta_{\text{eff}}\) can as well be used to describe the departure from hyperbolic moveout caused by the inhomogeneity in layered media, above a certain reflector. For isotropic \(v(\tau)\) media, equation (4) yields hyperbolic moveout using Tsvankin and Thomsen's definition of \(V_6\) [equation (9)], and therefore ignoring nonhyperbolic moveouts associated with vertical inhomogeneity. Better estimates of the moveout are achieved by using the new definition of \(V_6\) [equation (10)]. However, for isotropic media, the nonhyperbolic moveout given by equation (3) becomes slightly less accurate than that in equation (3) (see Appendix B). The reason for the reduced accuracy is that the correction factor introduced by Tsvankin and Thomsen is based on the anisotropy assumption only. Therefore, although it produces highly accurate moveout description for homogeneous VTI media, equation (5) has increased errors when inhomogeneity is introduced to the model (Appendix B). Fortunately, the errors arising from using equation (5) for all models shown (using examples with strong vertical inhomogeneity) are nevertheless less than 0.5% for \(X/D < 2\), rather independent of the strength of anisotropy.

Properties of the Nonhyperbolic Moveout

From equations (3) and (5), the value of \(\eta_{\text{eff}}\) for a given \(V_{\text{mne}}\) and \(t_0\) directly describes the degree of nonhyperbolic moveout in both anisotropic and isotropic layered media. For \(\eta_{\text{eff}} = 0\), the fourth-order term in equation (3) vanishes, and the moveout is hyperbolic. This is the case in homogeneous isotropic or elliptically isotropic media.

Referring to Figure 1, one cannot distinguish between the amount of nonhyperbolic moveout due to anisotropy and that due to inhomogeneity. If this medium, with its large vertical inhomogeneity, were strictly isotropic \((\epsilon = 0\) and \(\delta = 0)\) in each layer), then \(\eta_{\text{eff}}\), calculated using equation (8), would equal 0.06. In contrast, the presence of anisotropy resulted in \(\eta_{\text{eff}} = 0.19\). The difference between the two \(\eta_{\text{eff}}\) values, however, is not directly attributable to anisotropy because the relation between these factors is nonlinear.

The value of 0.06 for \(\eta_{\text{eff}}\) in this three-layer example results from a strong vertical inhomogeneity. I find \(\eta_{\text{eff}}\) values associated with more typical \(v(\tau)\) (average gradient of 0.6 s^{-1}) isotropic media to be much smaller than 0.1, a common value for typical TI media; thus, nonhyperbolic moveout is less severe for typical isotropic \(v(\tau)\) media than for common homogeneous VTI media. For example, let us approximate the velocity increase with depth in an isotropic medium \(\eta(\tau) = 0\), by a constant velocity gradient, \(a\), with \(v_0\) as the velocity at the surface; that is,

\[ v(\tau) = v_0 + a\tau. \]

For such a medium, velocity can be expressed in terms of two-way vertical traveltimes, \(t_0\) as

\[ v(t_0) = v_0 e^{0.5a t_0}. \]

Through straightforward derivation using equations (7) and (9), we find that for such a constant-gradient medium,

\[ \eta_{\text{eff}}(t_0) = \frac{1}{\tau} \left( \frac{0.5a t_0}{\tanh(0.5a t_0)} - 1 \right). \]

Here, \(t_\tau\) is the hyperbolic tangent function. Note that
\( \eta_{\text{eff}} \) is independent of \( v_0 \); therefore any linear velocity function with the same velocity gradient will lead to the same degree of nonhyperbolic moveout, independent of \( v_0 \).

Figure 2 shows \( \eta_{\text{eff}} \) values as a function of vertical time for three values of velocity gradient, \( a \). All three curves show modest values of \( \eta_{\text{eff}} \) when compared with \( \eta \) for typical homogeneous TI media [e.g., Taylor sandstone, where \( \eta = 0.156 \) (Thomsen, 1986)]. This result supports the contention that anisotropy typically introduces larger departure from hyperbolic moveout than does velocity layering. From Figure 2, it seems that the nonhyperbolic moveout at later times is large in \( v(z) \) media, but as \( t_0 \) increases (\( t_0 > 2 \) s), since the maximum offset usually remains constant, \( X/D \) reduces; therefore, the significance of nonhyperbolic moveout becomes smaller (Al-Chalabi, 1974). In other words, although \( \eta_{\text{eff}} \) increases with \( t_0 \) in Figure 2, the decrease in \( X/D \) reduces its influence, and therefore nonhyperbolic moveout due to smooth vertical inhomogeneity is small at all times.

In contrast, for homogeneous TI media, \( \eta_{\text{eff}} \) is constant, and the nonhyperbolic moveout is largest at early times where \( X/D \) is large.

Following steps similar to those used in deriving equation (11), we can derive an analytical expression of
for factorized TI media \( \eta = \delta, \) and \( \epsilon \) are independent of position (Cerveny, 1989; Alkhalilifah and Lander, 1994) with linear velocity variation. In factorized TI (FTI) media, we have

\[
v_{nmo}(\tau) = v(\tau)\sqrt{1 + 2\eta}, \tag{12}
\]

and

\[
v_h(\tau) = v_{nmo}(\tau)\sqrt{1 + 2\eta},
\]

where \( \eta \) is a constant \( \eta \) value for the factorized TI medium. For such a medium, with constant vertical gradient in the vertical \( P \)-wave velocity, \( \eta \) is given by

\[
\eta = \frac{1}{8}[1 + 8\eta]\frac{0.5a_t t_0}{\tanh(0.5a_t t_0)} - 1. \tag{13}
\]

The difference between the \( \eta \) values for FTI media [equation (13)] and for isotropic media [equation (12)] is given by

\[
\Delta \eta = \eta - \eta_{IS} = \frac{0.5a_t t_0}{\tanh(0.5a_t t_0)}. \tag{14}
\]

Based on equation (3), \( \Delta \eta \) describes the difference in the moveout curves for the two media. Moreover, since the moveout curve described by equation (3) is a reasonable approximation to the zero-offset diffraction curve, even in layered media, the difference between the moveout curves associated with VTI media and those of isotropic media can provide some insight into errors that can result from using an isotropic migration in VTI media.

**Estimating Anisotropy Using Nonhyperbolic Moveout**

If the maximum offset is large enough (offsets in marine surveys nowadays often exceed 4 km) relative to reflector depth and the resolution of the data is high, it is possible to estimate the degree of nonhyperbolic moveout due to anisotropy.

For \( X/D < 1 \), in layered VTI media the moveout is approximately hyperbolic and is given by

\[
t^2_h = t^2_0 + \frac{X^2}{V_{nmo}^2}.
\]

Subtracting equation (5) from this equation, and replacing \( \eta \) with \( \eta \), results in

\[
\Delta t^2 = t^2_h - t^2_0 = t^2_0 V_{nmo}^2 [2\eta t^2_0 V_{nmo} + (1 + 2\eta)X^2]. \tag{15}
\]

the amount of time-squared deviation attributable to the nonhyperbolic moveout. A straightforward manipulation of equation (15) results in

\[
\eta = \frac{\Delta t^2 V_{nmo}^2 (t^2_0 V_{nmo} + X^2)}{2X^2 (X^2 - \Delta t^2 V_{nmo}^2)}, \tag{16}
\]

Figure 3. Calculated \( \eta \) values as a function of error in the NMO velocity for offset-to-depth ratios, \( X/D = 1.5 \) (dashed black curve), \( X/D = 2 \) (solid gray curve), and \( X/D = 2.5 \) (solid black curve). Here, \( t_0 = 2.0 \) s, and \( \eta \) for the model is 0.1.

To accuracy governed by that of equation (5). Note that this expression is singular for \( X = 0 \). Clearly, no \( \eta \) information can be extracted from small offsets; the stability in estimating \( \eta \) is expected to increase with offset.

To estimate \( \eta \) using equation (16), one must first obtain \( V_{nmo} \), the short-spread NMO velocity corresponding to a horizontal reflector. This velocity can be obtained using conventional velocity analysis based on a moveout spread that satisfies \( X/D < 1 \). Assuming that an accurate \( V_{nmo} \) is obtained, then \( \Delta t^2 \) can be measured from the reflection moveout in the seismic data. One way to measure \( \Delta t^2 \) for use in equation (16) is to apply an NMO correction using \( V_{nmo} \) and compute

\[
\Delta t^2 = t^2_0 - t^2_{cor}, \tag{17}
\]

where \( t_{cor} \) corresponds to the moveout traveltime after NMO correction. Clearly, if the true moveout is hyperbolic, then \( t_{cor} \) equals \( t_0 \) and therefore \( \Delta t^2 = 0 \).

If we assume horizontal layering, the accuracy of the derived \( \eta \) depends primarily on the accuracy of the measured \( V_{nmo} \) and \( \Delta t^2 \), which in turn depends on the accuracy of \( V_{nmo} \). Therefore, the sensitivity measure must combine the effects of errors in \( V_{nmo} \) on both \( \Delta t^2 \) and \( \eta \). Figure 3 shows the sensitivity of \( \eta \) to errors in the measured \( V_{nmo} \) (i.e., from velocity analysis), calculated using equation (16). This example corresponds to a homogeneous medium with \( \eta = 0.1 \), \( V_{nmo} = 2.0 \) km/s, and \( t_0 = 2.0 \) s.

As expected, errors are smaller when longer offsets are used in estimating \( \eta \), again as long as velocity does not vary laterally. Therefore, any inversion technique (i.e., least-squares) based on the nonhyperbolic method to obtain \( \eta \) from measurements at different offsets should benefit from weighting factors that favor the
far offsets. Further, for fixed offset, the errors clearly increase with increase in either \( t_0 \) or \( V_{nmo} \) since an increase in either implies a reduced ratio \( X/D \).

Note in Figure 3 that even at zero error in \( V_{nmo} \) the inverted \( \eta_{cm} \) is not exactly 0.1. The error can be attributed to the difference between the forward time calculation, which involves exact ray tracing, and the inversion process based on equation (16), which is an approximation. This difference in this homogeneous model is largest for about \( X/D=2 \).

In any case, the errors caused by equation (16), at the correct \( V_{nmo} \), are small for typical anisotropies. Therefore, I rely on this analytical representation to accomplish most of the inversions in this paper.

The above approach of estimating \( V_{nmo} \) and \( \eta \) is introduced to develop insights in the nonhyperbolic inversion problem. A more practical approach, based on the 2-D semblance analysis method, is discussed later.

**Semblance Analysis Based on Hyperbolic Moveout**

Semblance analysis is less sensitive to traveltime errors than is traveltime inversion, and generally produces more stable results. The semblance coefficient is defined as the ratio of the output energy over a window of a stack of traces to the input energy in the unstacked traces. In mathematical terms, \( S_k \), the semblance coefficient, is

\[
S_k = \frac{\sum_{i=k-N/2}^{k+N/2} \sum_{j=1}^{M} f_{ij}(i,j)}{\sum_{i=k-N/2}^{k+N/2} \sum_{j=1}^{M} f_{ij}(i,j)}
\]

where \( f_{ij} \) is the recorded data in trace \( i \) at the time sample \( j \), and \( j \) is a function of the zero-offset time sample \( l \) and the trace (offset) \( i \). The window size \( N+1 \), usually set at about half the dominant period of the wavelet, is used to smooth the semblance-spectrum estimates. The semblance coefficient has a maximum value of unity (when all traces are identical) and a minimum of zero. Semblance summation in this form is biased against randomness and sudden variations in amplitude and polarity. Also, unlike simple summation as in conventional stacking, it is insensitive to the overall trace amplitude. Specifically, events with identical moveout, but differing in amplitude by a scaling factor, produce the same semblance response.

Estimating \( V_{nmo} \) through semblance velocity analysis is based on summing data over hyperbolic trajectories controlled by the trial moveout velocity, which defines \( j(i,l) \). Therefore, the velocity panel that shows the highest amplitude (stack power), for a specific time, through summation or some semblance measure, is the stacking velocity. In homogeneous isotropic media, where the moveout is hyperbolic, the stacking velocity is identical to the NMO velocity of the medium. In anisotropic, as well as inhomogeneous media (Al-Chalabi, 1974), the moveout is no longer hyperbolic, and the nonhyperbolic portion of the moveout can distort estimates of stacking velocity so that they differ from the NMO velocity, with difference proportional to the size of the nonhyperbolic moveout. As demonstrated earlier, nonhyperbolic moveouts are larger for typical anisotropy than for typical vertical inhomogeneity. Therefore, the difference between the stacking and the NMO velocity is expected to be larger in anisotropic media. For \( \eta > 0 \), which is the typical situation, moveouts at far offsets deviate from the nearly hyperbolic trajectory at lower times, resulting in higher stacking-velocity estimates from velocity analysis.

The size of the deviation of stacking-velocity estimates depends primarily on the range of offsets used in the velocity analysis process.

Figure 4 shows velocity analysis panels for various offset-to-depth ratios (\( X/D \)) used in the analysis. Here, the reflection is at \( t_0=2.0 \) s, \( \eta \) for the model is 0.1, and \( V_{nmo}=2.0 \) km/s. Random noise with an rms signal-to-noise ratio of 3 was added to all synthetic examples used in this paper. Estimated stacking velocities increase with increasing \( X/D \); for \( X/D=2 \), even the reflection time is distorted from the actual zero-offset time of 2 s. While this spread-length bias increases with increasing \( X/D \), the ability to resolve the velocity also increases with increasing offset used in the analysis process. For \( X/D=1 \), which is typically used in conventional velocity analysis, the stacking velocity from Figure 4 is estimated to be 2.03 km/s, 1.5 percent higher than \( V_{nmo} \). If a smaller \( X/D \) is used (i.e., \( X/D=0.5 \)), the error in estimating \( V_{nmo} \) becomes less than one percent, although the resolution is poorer. Theoretically, as offset approaches zero the stacking velocity should approach \( V_{nmo} \). Practically, however, as the range of offsets used decreases, velocity analysis suffers from reduced resolution. The trade off between resolution and accuracy in stacking velocity depends mainly on the peak frequency of the wavelet. Once a choice is made regarding this trade off, we can use Figure 3 to relate possible \( V_{nmo} \) errors to the accuracy expected in inverting for \( \eta \) from traveltimes picks.

Figure 5 shows semblance results from summing again over hyperbolic trajectories, controlled by the stacking velocity, for a single reflection event with zero-offset time of 2 s. The general model is the same as in Figure 4 with (a) \( \eta = 0 \) (isotropic model), and (b) \( \eta = 0.1 \). In both cases \( V_{nmo}=2.0 \) km/s. The vertical axis in Figure 5 corresponds to the maximum offset used in the semblance-analysis process. For smaller maximum off-
Figure 4. Velocity analysis panels for various offset-to-depth ratios ($X/D$) used in the analysis. Here, the reflection is at $t_0=2.0$ s (at depth, $D=2$ km), $\eta$ for the model is 0.1, and $V_{nmo}=2.0$ km/s. The peak frequency of the Ricker wavelet used in the analysis here, and throughout the paper, is equal to 40 Hz.

Figure 5. Velocity-analysis panels for various maximum offsets used in the analysis. Here, the reflection is at $t_0=2.0$ s, $V_{nmo}=2.0$ km/s, and (a) $\eta=0$, and (b) $\eta=0.1$. The peak frequency of the Ricker wavelet used the analysis is equal to 40 Hz.

sets, the resolution is poor and the velocity is unresolvable. As the maximum offset increases, so does the resolution. Nevertheless, a clear shift of the best-fit stacking velocity occurs in the TI model, a direct influence of nonhyperbolic moveout. The shift is dramatic as we approach $X/D=2$ (at offset 4.0 km). Also, as the maximum offset used increases, the semblance power decreases, because the best-fit hyperbolic moveout fails to simulate the true nonhyperbolic moveout. As the amplitude decreases, the contribution of noise to the analysis process would increase. In the next section, I demonstrate, through a semblance analysis over nonhyperbolic moveout, how to reduce the errors in estimating $V_{nmo}$ that arise in long-offset data.

Semblance Analysis Based on Nonhyperbolic Moveout

To use the semblance coefficient with a nonhyperbolic moveout trajectory I simply describe $\gamma(l,l)$ using the nonhyperbolic moveout equation (5) instead of the hyperbolic one. However, in this case, the moveout depends on two parameters rather than one, thus expanding the dimensionality of the search. The nonhyperbolic scan below is applied over $V_h$ and $V_{nmo}$, rather than $\eta$ and $V_{nmo}$, so that both axes have the same units to simplify comparison of resolution and accuracy.

Figure 6 shows the semblance coefficient as a function of $V_{nmo}$ and $V_h$ for a model with a horizontal reflector at depth 2.0 km beneath a homogeneous TI medium with $V_{nmo}=2.0$ km/s and $\eta=0.2$. The zero-offset reflection time of the Ricker wavelet was at 2 s, and the scan was done by setting $t_0 = 2.0$ s. A 3-D scan would require a search over zero-offset time, as well. Figure 6a corresponds to a maximum $X/D$ of 1.5 used in the semblance summation, Figure 6b corresponds to maximum $X/D=2.0$, and Figure 6c corresponds to maximum $X/D=2.5$. As expected, resolving power (reciprocally related to the overall size of the elongated nearly elipsoidal-shaped darkened region) increases with larger maximum
offsets. In fact, because this elongated region tilts further from the vertical as \(X/D\) increases, the ability to resolve \(V_h\) increases considerably when larger offsets are included. (The maximum semblance response for any of the three maximum \(X/D\) could be picked at \(V_{nmo}=2.0\) km/s and \(V_h=2.18\) km/s. (The confidence in this pick increases with increasing offsets used in the analysis.) These values of \(V_{nmo}\) and \(V_h\) result in \(\eta = 0.095\), which is close to the actual value of 0.1. The slightly low estimate for \(\eta\) arises from using the nonhyperbolic equation (5) which is an approximation (although a good one) of the actual moveout.

A practical approach, which can reduce the cost of a 3-D scan over \(V_{nmo}\), \(V_h\) and \(t_0\) is an iterative 2-D technique, where we scan once over \(V_{nmo}\), fix the interpreted values of \(V_{nmo}\), and then do another scan, this time over \(V_h\). The results of the \(V_h\) analysis would then be used to scan again over \(V_{nmo}\) and so on until a convergence criterion is met. As has been suggested by May and Stratley (1979), convergence is guaranteed through using an orthonormal basis (i.e. the Legendre polynomials) to represent the moveout polynomial [equation (3)].

The NMO velocity obtained from semblance analysis is usually more accurate than that extracted by fitting a hyperbolic curve, in a least square sense, to the moveout of a reflection. We should expect a similar result when nonhyperbolic moveouts are used in the semblance. In the analysis based on least-squares fitting of traveltimes, Tsvankin and Thomsen (1995) concluded that the second and fourth order coefficients of the Taylor’s series expansion of the TI moveout are not resolvable from traveltime moveout curves in VTI media.

The reason that the semblance approach reduces this ambiguity in resolving the anisotropy parameters discussed by Tsvankin and Thomsen is basically described by the concept of objective function (a function of the unknown parameters formed so that maximum or minimum value of the function corresponds to the solution of the problem).

The objective function for the semblance analysis approach, given by the semblance responses in Figure 6 for 40-Hz peak frequency, has a more stable maximum than does the objective function calculated based on a least-square traveltime fitting of the moveout over the same range of offsets (Figure 7). In fact the least-square method is more sensitive to the shortcomings of the moveout approximation than is the semblance approach. This is obvious by observing the amount of shift of the maximum (or minimum) from the true values for the model. The velocity-analysis objective function is also more stable and less sensitive to noise and traveltime errors than the least-squares traveltime fitting approach. Figure 8 shows (a) a semblance analysis, and (b) a least-squares traveltime fitting objective function after subjecting the synthetic data of Figure 6 to random traveltime shifts between 0 and 0.5\%(\(=10\) ms), as might happen after a poor static correction. These traveltime shifts have a mean of 0.25\%(\(=5\) ms). Clearly, the objective function of the least-square fitting approach is much more influenced by the errors (shifted from the true value) than is the semblance approach. The resolution of the objective function for the semblance approach also depends largely on the peak frequency of the wavelet. The dominant frequency of 40 Hz, used here, is quite representative of
Anisotropy and Inhomogeneity

In typical TI media, η is positive, and the fourth-order term in equation (5) is negative. Similarly, in typical isotropic media in which velocity varies with depth, the fourth-order term is again negative, producing a similar moveout behavior. Specifically, both $t^2 - X^2$ curves are convex upwards (Hake et al., 1984). Although $\eta_{\text{eff}}$ for isotropic media is usually smaller than that in TI media, such a result will raise problems in deciding how much of the inverted $\eta_{\text{eff}}$ to attribute to anisotropy and how much to inhomogeneity. Nevertheless, we expect that, often, the dominant portion of the nonhyperbolic moveout can be attributed to anisotropy, and if inhomogeneity is resolvable through other techniques (i.e., conventional velocity analysis), then a better assessment of the contribution from anisotropy can be obtained.

The relative sensitivity of dip-moveout processing to anisotropy and vertical inhomogeneity differs from that of nonhyperbolic moveout. For example, primarily for moderate to low dips, a dip-moveout impulse response for typical anisotropy ($\eta > 0$) can be approximated as a stretched version of the one for homogeneous isotropic media, while the $v(\tau)$ isotropic impulse response is a squeezed version of the homogeneous isotropic one (Alkhalifah, 1996a). Therefore, the presence of both anisotropy and inhomogeneity in a medium leads to DMO actions that are opposite to one another, while the actions of both anisotropy and inhomogeneity increase the nonhyperbolic moveout of reflections from horizontal interfaces.

Suppose only one reflection is strong enough to show measurable nonhyperbolic moveout in a $v(\tau)$ VTI medium. A reasonable approach might be to consider the medium PTI, and therefore obtain a constant $\eta_{\text{PTI}}$. Using equation (8) and setting $\eta(\tau)$ to be constant ($=\eta_{\text{PTI}}$) results in

$$\eta_{\text{PTI}} = \frac{1}{8} \left[1 + 8\eta_{\text{eff}}\int_{t_0}^{\infty} \frac{V_{\text{nmo}}(t_0)}{V_{\text{nmo}}(\tau)} d\tau - 1\right],$$

which enables one to deduce an average $\eta$ corresponding solely to anisotropy. Equation (18) is important because often, as we will see later in the field example especially
at early recording times \( t_0 < 2 \) s, only one reflection will
be strong enough to show measurable nonhyperbolic moveouit. (Lateral velocity variation, of course, would com-
plicate this interpretation.)

Advantages of the Nonhyperbolic Inversion Method

Alkhalifah and Tsvankin (1995) have developed a pro-
cedure for estimating \( \eta \) and \( V_{nmo} \) in layered TI media
using the short-spread moveout behavior for dipping re-
flectors (DMO method). Even vertical variations of \( \eta \) can
be estimated using such a DMO method (Alkhalifah,
1996b). That DMO-based procedure is probably more
stable in inverting for the anisotropy parameters than is
the method described above, especially in the absence
of large offsets and at later times, where \( X/D \) is small.

Unlike the DMO method, however, this nonhyper-
bolic moveout method does not require dipping reflectors,
and therefore, it is more flexible; also, it can be applied
to a broader range of field data. Moreover, it is easier to
use to obtain lateral variations of \( \eta \). For example, statistical
estimation of the lateral variation in \( \eta \) can be made
from data at many CMP locations.

Given the tradeoff between \( V_{nmo} \) and \( \eta \) in equa-
tion (1), the errors in estimating \( V_h \), using the nonhyper-
bolic moveout, are generally small. The horizontal veloc-
ity is the necessary quantity for migration of a vertical
reflector to its true position, and using the nonhyper-
bolic inversion it is usually estimated at higher accuracy
than is \( V_{nmo} \). Therefore, with the nonhyperbolic move-
out method one can better construct the time-migration
impulse response than with isotropic methods.

One area in which \( \eta \) measurements from nonhyper-
bolic moveout can play a major role is in the presence of
very steep (near vertical) reflectors, such as flanks of salt
domes in the Gulf of Mexico, where, in addition, reflect-
ions from interfaces with intermediate dips may not be
available. Alkhalifah and Tsvankin (1993) showed that
the DMO method fails to yield accurate values of \( \eta \) for
such steep dips, primarily because the moveout for such
reflections in TI media is not distinguishable from that in
isotropic media or in any other anisotropic model. There-
fore, the moveout for such dips becomes somewhat inde-
dependent of the anisotropy parameter, \( \eta \). The nonhyper-
bolic moveout for events from sub-horizontal reflectors
potentially can provide \( \eta \) information for improved mi-
gration of data from steep reflectors.

This nonhyperbolic method, however, is based on
the assumption of lateral homogeneity, with some toler-
ance, as is usually the case with \( v(z) \) algorithms, to mild
lateral inhomogeneity (i.e., smooth lateral variations).
Therefore, strong lateral inhomogeneities will cause prob-
lems for the method, requiring a much more advanced
treatment, which is beyond the scope of this paper. In
media with strong lateral inhomogeneity, \( \eta \) is still pos-
sibly measurable, and can be used to aid in making non-
hyperbolic moveout correction, but it has no simple in-
terpretation in terms the medium properties.

Field-Data Example

Figure 9 shows a seismic line from offshore Africa
provided by Chevron Overseas Petroleum, Inc. The line
was processed using conventional NMO and DMO algo-
rithms without taking anisotropy into account. Veloci-
ity analysis shows a general vertical velocity increase
with depth that can be simplistically modeled with a
constant gradient of about \( 0.7 \) s\(^{-1}\). As Alkhalifah and
Tsvankin (1995) demonstrate, these data are influenced
by the presence of anisotropy. Moreover, using the DMO
method for estimating \( \eta \) and \( V_{nmo} \), Alkhalifah (1996a)
shows that the anisotropy is strongest above \( t_0=2 \) s, in
a massive shale formation.

Figure 10 shows CMP gathers after applying iso-
tropic homogeneous DMO and NMO correction. The
NMO correction is based on the velocities obtained from
conventional velocity analysis, using a spread given by
\( X/D < 1 \). DMO was applied to reduce even the small distor-
tion of the stacking velocity caused by the mild dip (\( \pm 6 \)
degrees) of the reflector at about \( t_0=1.8 \) s. The two sub-
parallel events prior to \( t_0=2.0 \) s show significant depart-
ture from hyperbolic moveout. If the deviations in Fig-
ure 10 were caused by NMO velocity overcorrection (us-
ing lower than true velocities), then these moveout curves
would have departures from \( t_0 \) proportional to \( X^2 \). The
fact that these curves are practically straight for \( X/D < 1 \),
implies that they are controlled by higher-order terms of
the Taylor’s expansion (e.g., \( X^4 \)).

A detailed portion of Figure 10 (Figure 11) helps in
picking reflection times, and therefore, measuring \( \Delta t \).
Some of the moveouts (e.g., the reflection at \( t_0=1.86 \) s
at CMP 700) have a slight initial plunge prior to the
larger offsets where the nonhyperbolic behavior domi-
nates the moveout. This initial plunge results from using a
\( V_{nmo} \) value in the NMO correction that is higher than
the true value. As suggested in Figure 4, the higher ve-
locity is probably due to spreadlength bias, which arises
from the attempt to fit nonhyperbolic moveout with hy-
perbolic curves. Analysis over nonhyperbolic moveout
should overcome such a problem as well as provide an
estimate of the nonhyperbolic portion of the spread.

Figure 12 shows the semblance response using non-
hyperbolic moveouts as a function of \( V_{nmo} \) and \( V_h \) (sim-
Figure 9. CMP-stacked seismic line (offshore Africa) after application of NMO correction along with homogeneous and isotropic DMO; the distance between CMP's is 12.5 m.

Similar to that of Figure 6), for the same reflection events shown in Figure 11, at CMP locations 700, 800, and 900. Note that, among the three locations, $V_{\text{true}}$ decreases monotonically from 2.15 km/s at CMP location 700 to 2.07 km/s at CMP 900. This decrease corresponds mainly to the decrease in the zero-offset times of these reflections, which reflects the general velocity increase with depth obtained by Alkhalifah (1996a). On the other hand, $V_{\text{a}}$ or $\eta$ have their highest value at the middle location, CMP location 800. At this CMP location, $\eta_{\text{eff}}=0.28$, whereas at CMP locations 700 and 900 $\eta_{\text{eff}}$ equals 0.2 and 0.16, respectively. These values of $\eta$ include the combined influence of anisotropy and inhomogeneity. Using equation (18) in an attempt to remove the influence of vertical inhomogeneity, I estimate $\eta_{\text{eff}}=0.15$, 0.22, and 0.13 for CMP locations 700, 800, and 900, respectively, which I attribute primarily to the anisotropy. These values are on average higher than those obtained by Alkhalifah and Tsvankin (1993) and Alkhalifah (1996a with the DMO-based approach. However, their measurements correspond primarily to the region near CMP location 900, which has an $\eta$ value that is more in agreement with their calculations. Finally, considering the overall small lateral velocity variation, the resulting estimate of $\eta$ suggests the presence of anisotropy. Further analysis of the relative importance of anisotropy and lateral velocity variation, however, is necessary.

Figure 13 shows a schematic plot of the raypath from the source down to the reflection point and back up to the receiver, for the maximum offset used in the semblance analysis at CMP 800 in Figure 12. The ray bending is caused by the vertical increase in velocity with depth. Unlike NMO velocity analysis, the $\eta$ estimates rely on information from large offsets; therefore, the subsurface influence on the $\eta$ estimate is not laterally local (i.e., near the CMP location of measurement). Although use of large
Figure 10. CMP gathers at locations 700, 800, and 900 after NMO correction and isotropic homogeneous DMO. The NMO correction is based on velocities obtained from conventional velocity analysis with $X/D < 1$.

Offsets can help improve the resolution of the inversion, but can hamper the lateral resolution of estimating $\eta$.

To get a better understanding of the lateral variations in $\eta$ in the above field example, we should compare CMP locations that are at least 3 km apart (the distance corresponding to the maximum offset used in the analysis). Moreover, the data would have better served us if $\eta$ estimates were made at many more CMP locations and were then averaged (smoothed) over 3 km intervals, with for example a Gaussian window.

Figure 14 shows another portion of the data set from offshore Africa that is dominated by horizontal (or sub-horizontal) events. The large number of strong horizontal reflectors should provide us with an excellent setting for applying the layer stripping approach discussed earlier. This data set also includes offsets up to 4.3 km, which will help in boosting the resolution of the semblance analysis at later times. Nevertheless, the measurements at the later times still suffer from lower resolution, due to smaller $X/D$, as well as increased layer-stripping errors that propagate from the top to the bottom of the section. Figure 15 shows four sample nonhyperbolic semblance responses calculated at times 1.24, 1.86, 2.28, and 2.99 s, at CMP location 300. Picking the $V_{nmo}$ and $V_h$ values corresponding to the maximum semblance responses for these times, as well as other ones, and inserting them into equations (7) and (8), yields the velocity and $\eta$ curves shown in Figure 16 (black curves). Because these are marine data, $V_{nmo}$ and $\eta$ are set to equal $1.3 \text{ km/s}$ and zero, respectively, at the surface. The gray curves, on the other hand, describe the upper and lower limits of possible parameter values corresponding to the uncertainties in picking $V_{nmo}$ and $V_h$ (e.g., picking within the dark region in Figure 15). As above, I ignore the influence of lateral velocity variation on the results. However, the lateral velocity variation in this region is mild (< 2%).

Note in Figure 15 that the semblance resolution, especially for $V_h$, decreases with increasing zero-offset time due to the reduction in $X/D$. This will degrade the accuracy of picking, resulting in errors in the interval values that increase with vertical time. Also, as with any other layer-stripping application, the interval values at the later times have errors accumulated from measurements at the earlier times. So, in Figure 16, $\eta$ values beyond $t_0=2.0$
Figure 11. Detail of Figure 10. The black curve is the approximate location of the zero-crossing of the reflection wavelet. It indicates the general shape of the reflection moveout.

Figure 12. Nonhyperbolic velocity analysis for CMP locations 700, 800, and 900. Here, \( t_0 \), which varies from one CMP to another, has been extracted from Figure 11. The gray curves again correspond to contour lines describing \( \eta \).
s are not that reliable. On the other hand, the increase of \( \eta \) up to \( t_0 = 1.8 \) s is believable, because this increase is maintained even when the measured values at \( t_0 = 1.86 \) s in Figure 15 are perturbed within the range of acceptable picks (the black region). For example, although \( V_4 \) was evaluated at the maximum semblance to equal 2.05 km/s, one can assume a margin of error, corresponding to the black region, of about \( \pm 0.04 \) km/s, which corresponds to about a 2 percent error. The limits of this margin are given by the gray curves in Figure 16. Within this margin of error, the \( \eta \) curve in Figure 16 always increases up to the maximum value at 1.8 s, but the particular form of the increase depends on the \( V_4 \) value. Therefore, the increase in \( \eta \) until 1.8 s is more-less an accurate representation of the anisotropy variation under CMP 300. The reliability of the results at these times was further enhanced by observing similar measurements from neighboring CMP locations (i.e., CMP locations 200 and 250).

Conclusions

The nonhyperbolic moveout behavior of reflections from horizontal interfaces is an important source of velocity information for processing, especially in anisotropic media. In most anisotropic media, the nonhyperbolic moveout is relatively large, larger, in fact, than that in typical vertically inhomogeneous isotropic media. Therefore, it is usually observable and measurable, and thus, can be used to invert for medium parameters. Although estimates of \( \eta \) derived from nonhyperbolic moveout method alone might not be reliable to use directly in lithology interpretation, such results can play a major role in processing and supporting the estimates of \( \eta \) and \( V_{nmo} \) obtained from the DMO method (Alkhalifah and Tsvankin, 1995), as well as in providing \( \eta \) values in areas where the dipping features required by the DMO method are absent.

The effective \( \eta \) values obtained from the NMO method can be used as an indicator of the relative contribution to nonhyperbolic moveout due to anisotropy as opposed to that due to vertical heterogeneity. This indicator demonstrates that nonhyperbolic moveout associated with typical TI media (\( \eta \approx 0.1 \)) is greater than that associated with typical \( v(z) \) isotropic media; therefore, the importance of applying the nonhyperbolic moveout correction in such TI media prior to stacking is larger than that in \( v(z) \) media. The importance of anisotropy relative to that of lateral velocity variation remains an important area for future study.

The process of extracting \( \eta \) values from the nonhyperbolic moveout behavior of reflections is sensitive to errors in the measured \( V_{nmo} \), with the sensitivity reducing as offsets used in the inversion increase. It also reduces at smaller vertical time corresponding to smaller depths, and thus larger offset-to-depth ratio. In TI media, with positive \( \eta \), nonhyperbolic moveout tends to give an overestimate of the value of \( V_{nmo} \) obtained us-
ing velocity analysis, depending mainly on the offset-to-depth ratio, $X/D$, used in the velocity analysis. For typical $X/D < 1.0$, such increases do not exceed 2 percent. Depending on the ratio $X/D$ used in the inversion for $\eta$, this overestimate of $V_{hmo}$ results in an estimate of $\eta$ that is low, by no more than 0.04 for a model with $\eta = 0.1$. However, the estimates $\eta$ and $V_{hmo}$ can be used iteratively to improve further estimates of one another. At increased cost, semblance analysis over nonhyperbolic trajectories can reduce such errors and thereby providing better estimates of $V_{hmo}$ and $\eta$.

The nature of seismic data, which are dominated by horizontal and sub-horizontal reflections, makes this method more widely applicable than is the DMO method of Alkhalifah and Tsvankin (1995), which relies on the presence of reflections with different dips. The nonhyperbolic moveout method, in homogeneous media, needs only one reflector, preferably a horizontal one, for the inversion to work. Therefore, it can be applied almost anywhere, which helps in estimating lateral variations in $\eta$. Use of larger offsets in the nonhyperbolic inversion, however, which are necessary to stabilize the inversion, can degrade the lateral resolution of $\eta$.

One area in which $\eta$ measurements from nonhyperbolic moveout can play a major role is in the presence of very steep (near vertical) reflectors, such as flanks of salt domes in the Gulf of Mexico, where in addition, reflectors from interfaces with intermediate dip may not be available. In this situation, the DMO method of Alkhalifah and Tsvankin (1995) fails for such steep dips, primarily because the moveout for such a reflection in TI media is not distinguishable from that in isotropic media. Therefore, nonhyperbolic moveout from horizontal events may be the only information in surface seismic data that can provide us with $\eta$ estimates necessary to better migrate such dips.

Using the nonhyperbolic moveout method on data from offshore Africa helped to estimate $\eta$ both vertically and laterally. However, the accuracy of the inversion reduces dramatically with depth, due to the reduced $X/D$, and due to layer-stripping errors.
Acknowledgment

I am grateful to Ken Larner, Ilya Tssvinkin, John Toldi and Vladimir Grechka for helpful discussions. I thank John Toldi and Chris Dale of Chevron Overseas Petroleum, Inc. for providing the field data. Special thanks are also due to the Center for Wave Phenomena, Colorado School of Mines, for its technical support, and also to KACST, Saudi Arabia, for its financial support. Financial support for this work also was provided in part by the United States Department of Energy, (this support does not constitute an endorsement by DOE of the views expressed in this paper).

REFERENCES


Alkhalifah, T., 1996a, Transformation to zero offset in transversely isotropic media: Geophysics, expected to be published in the July-August issue.


Alkhalifah, T., 1996c, Analytic insights into the anisotropy parameter $\eta$: Center for Wave Phenomena, Colorado School of Mines (CWP-202).


Nédélec, N. S., and Taner, M. T., 1971, Semblance and other coherency measures for multichannel data: Geophysics, 36, 482-497.


APPENDIX A: Effective \( \eta \) in Layered Media

For multi-layered TI media, the exact quartic coefficient \( A_4 \) of the Taylor's series expansion is given by (Hake et al., 1984; Tsvankin and Thomsen, 1994)

\[
A_4 = \frac{\left( \sum \left[ v_{nmo}^{(i)} \right]^2 \Delta t^{(i)} \right)^2 - t_0 \sum \left[ v_{nmo}^{(i)} \right]^4 \Delta t^{(i)}}{4 \left( \sum \left[ v_{nmo}^{(i)} \right]^2 \Delta t^{(i)} \right)^4} + \frac{t_0 \sum A_{4i} \left[ v_{nmo}^{(i)} \right]^8 \Delta t^{(i)} \Delta t^{(i)2}}{\left( \sum \left[ v_{nmo}^{(i)} \right]^2 \Delta t^{(i)} \right)^2},
\]

(A1)

which includes (in the first term) ray-bending due to the layered structure. Here, \( \Delta t^{(i)} \) and \( v_{nmo}^{(i)} \) are the two-way zero-offset time and the NMO velocity for a given layer \( i \), respectively. \( A_{4i} \) is the quartic coefficient \( A_4 \) in layer \( i \). When the contribution of the vertical shear-wave velocity is ignored, it is given by

\[
A_{4i} = -\frac{2\eta^{(i)}}{[\Delta t^{(i)}]^{5/2} [v_{nmo}^{(i)}]^{4}},
\]

(A2)

where \( \eta^{(i)} \) is the \( \eta \) value for a given layer \( i \). On the other hand, again ignoring the contribution of the vertical shear-wave velocity, \( A_4 \) is given by

\[
A_4 = -\frac{2\eta_{\eta \eta \eta}}{t_0^2 V_{nmo}^2}.
\]

(A3)

Plugging \( A_{4i} \) into equation (A1) and using integration instead of summation gives

\[
A_4(t_0) = \frac{t_0^2 V_{nmo}^2(t_0) - t_0 \int_0^{t_0} v_{nmo}^c(\tau) d\tau}{4t_0^2 V_{nmo}^2} - \frac{8t_0 \int_0^{t_0} \eta(\tau) v_{nmo}^c(\eta(\tau)) d\tau}{4t_0^2 V_{nmo}^2}.
\]

(A4)

Substituting equation (A3) into equation (A4), with straightforward manipulation, results in

\[
\bar{\eta}_{\eta \eta}(t_0) = \frac{1}{8} \frac{1}{t_0^2 V_{nmo}(t_0)} \int_0^{t_0} v_{nmo}(\tau) [v_{nmo}^c(\tau)] \eta^2(\tau) d\tau
\]

(A5)

APPENDIX B: Nonhyperbolic Moveout in Isotropic Layered Media

Typically, reflections in isotropic \( v(z) \) media are approximated by hyperbolic moveout characterized by root-mean-square (rms) velocity. If the subsurface velocity has large variation in depth, the hyperbolic moveout assumption is less accurate, and an additional term in the Taylor series expansion is needed to better simulate the moveout. The three-term expansion (Tanner and Koehler, 1969; Shah and Levin, 1973; Al-Chalabi, 1974) for isotropic media is given by

\[
t = t_0 + \frac{X^2}{V_{nmo}^2} + \frac{(1 - \frac{V_{nmo}^2}{V_{nmo}^2})X^4}{4t_0^2 V_{nmo}^2},
\]

(B1)

where \( V_{nmo} \) is calculated using equations (7), and \( V_i \) is calculated using the following

\[
V_i(t_0) = \frac{1}{t_0} \int_0^{t_0} v_i(\tau) d\tau.
\]

(B2)

The interval quantities \( v_{nmo}(\tau), \) and \( v_i(\tau), \) in this case, are equal. The first two terms in equation (B1) correspond to hyperbolic moveout, and the additional term is the nonhyperbolic term, which provides better accuracy at the larger offsets. Actually, equation (3), with \( \eta_{\eta \eta \eta} \) given by equation (8), reduces to equation (B1) for isotropic media.

The left portion of Figure A1 shows an isotropic \( v(z) \) model consisting of three layers. The right portion shows the percent error in the computed moveout corresponding to reflections from (a) the bottom of the second layer, and (b) the bottom of the third layer. The time errors resulting from using equation (B1) (or equation (3)), given by the black curve, are overall less than those resulting from using equation (5), given by the gray curve. Both, however, are better than using a hyperbolic moveout, given by the dashed curve. Nevertheless, the difference between the two nonhyperbolic approximations is small, even for such a strong vertical inhomogeneity. This difference reduces dramatically for smoother inhomogeneity. The additional \( X \) factor in the denominator of the fourth-order term in equation (5), although extremely important for VTI models, is generally ineffective for isotropic models.
Figure A1. Left: Model with three isotropic layers. Right: Percent time error in moveout corresponding to reflections from (a) the bottom of the second layer, and (b) the bottom of the third layer. The grey curve corresponds to equation (5); the black curve corresponds to equation (3); the dashed curve corresponds to the hyperbolic moveout. $V_{nmo}$ for all three curves is calculated using equation (7).
Micro-local, non-linear, resolution analysis of generalised Radon transform inversions in anisotropic media

Maarten V. de Hoop and Norman Bleistein
Center for Wave Phenomena,
Colorado School of Mines,
Golden CO 80401-1887, USA.

April 10, 1996

Abstract

The resolution analysis of generalised Radon transform (GRT) inversions of seismic data is carried out in general anisotropic media. The GRT inversion formula is derived from the ray-Born approximation of the wave field for volume scatterers. However, by considering in the resolution analysis scattering surfaces instead, the inversion provides reflection/transmission coefficients as functions of scattering angles and azimuths. Those coefficients can be subjected to any type of A(mplitude) V(ersus scattering) A(ngles) analysis, in a non-linear sense to obtain information about the medium discontinuities across the scattering surface.

Keywords: migration/inversion, Kirchhoff approximation, generalised Radon transform.
1 Introduction

In the process of enhancing resolution, both spatially as well as in elastic parameters, current acquisition techniques take reflection-seismic measurements at increasingly larger scattering angles. To achieve a better resolution, however, one must account for the interplay between heterogeneity and anisotropy, at different length scales. Also, to accommodate inversion of wide angle data, one has to consider a non-linear approach, at least near the relevant reflectors in the configuration. We use the generalised Radon transform or GRT formulation to achieve this goal.

First, the ray-Born approximation of the direct scattering problem is cast into an integral of surface integral representations. Then the inversion of the linearized scattering problem is carried out. The resolution analysis of the linearized GRT inversion serves as the basis to arrive at a micro-local non-linear formulation: its range of validity is extended by applying a stationary phase analysis with respect to migration dip, defined as the normalized gradient of total travel time along the scattering characteristic. In this process, carrying out the GRT inversion on Kirchhoff-type data, an explicit adjustment of the inversion formula is found.

By ordering the data, for each image point, into common scattering-angle/azimuth gathers, the GRT inversion formula can be modified to obtain reflection/transmission coefficients at specular ray geometries. Any amplitude versus scattering angles, AVA, analysis can then be applied to those coefficients to derive information about the medium perturbation. Several parametrizations of this perturbation have been developed to reveal to which quantities the coefficients are sensitive.

The idea of estimating angle-dependent reflectivity has been developed by various authors, see for example Geoltrain and Chovet [1] and Lumley [2]. A more rigorous discussion of such an approach can be found in the book of Bleistein et al. [3]. In our paper, we derive a GRT-based inversion procedure that accomplishes the goal of constructing full reflection, and possibly transmission, coefficients as functions of scattering angle and azimuth, in closed form.

GRT-type inversion formulas for the linearized inverse problem have been developed by Norton and Linzer [4], by Beylkin [5, 6, 7, 8], by Miller et al. [9, 10], by Beylkin et al. [11], and by Rakesh [12] for the acoustic case. The extension to the elastic case was discussed by Beylkin and Burridge [13], and anisotropy was considered by De Hoop et al. [14] and Spencer and De Hoop [15]. Inversion formulas aiming to estimate reflectivity rather than the medium perturbation were developed by Cohen and Bleistein [16], by Bleistein and Cohen [17], and by Bleistein [18]. This paper brings the two approaches together. How to implement the GRT inversion procedures numerically can be found in De Hoop and Spencer [19].
2 The basic equations

Notation

First, we introduce some basic notation. Choose coordinates in the configuration according to

\[ x = (x_1, x_2, x_3) = \text{Cartesian position vector,} \]
\[ s = (s_1, s_2, s_3) = \text{source point,} \]
\[ r = (r_1, r_2, r_3) = \text{receiver point,} \]
\[ t = \text{time.} \]

The medium is described by

\[ \rho(x) = \text{density,} \]
\[ c_{ijkl}(x) = \text{elastic stiffness tensor.} \]

while the wave field is described by

\[ u(x, t) = (u_1(x, t), u_2(x, t), u_3(x, t)) = \text{displacement vector,} \]

and generated by a source distribution given by

\[ f(x, t) = (f_1(x, t), f_2(x, t), f_3(x, t)) = \text{body-force source density.} \]

In the remainder of the paper, we will employ the summation convention.

The displacement in a heterogeneous anisotropic medium satisfies the elastodynamic wave equation

\[ \rho \partial^2_t u_i - \partial_j (c_{ijkl} \partial_k u_k) = f_i, \quad (2.1) \]

with summation over repeated lower case indices, here and below. Let

\[ G(x, x', t) = (G_{ip}(x, x', t)) \quad (2.2) \]

be the causal Green's tensor, which satisfies (cf. Eq. (2.1))

\[ \rho \partial^2_t G_{ip} - \partial_j (c_{ijkl} \partial_l G_{kp}) = \delta_{ip} \delta(x - x') \delta(t), \quad G_{ip} = 0 \text{ for } t < 0. \quad (2.3) \]
Asymptotic ray theory

Here, we summarize the formulation of anisotropic ray theory for the evaluation of the Green’s tensor. Let

\[ G_{tp}(x, x', t) = \sum_{N} A^{(N)}(x, x') \xi^{(N)}(x) \xi^{(N)}(x') \delta(t - \tau^{(N)}(x, x')) \]

\[ + \text{ terms smoother in } t . \] (2.4)

In this equation, the arrival time \( \tau^{(N)} \) and the associated polarization vector \( \xi^{(N)} \) satisfy

\[ (\rho \delta_{ik} - c_{ijkt}(\partial_t \tau^{(N)})(\partial_j \tau^{(N)})) \xi_k^{(N)} = 0 \quad \text{(at all } x) , \] (2.5)

which implies the eikonal equation

\[ \det(\rho \delta_{ik} - c_{ijkt}(\partial_t \tau)(\partial_j \tau)) = 0 \quad \text{(at all } x) . \] (2.6)

The polarization vectors are assumed to be normalized so that \( \xi_i^{(N)} \xi_i^{(N)} = 1 \). Define the slowness vector \( \gamma^{(N)} \) by

\[ \gamma^{(N)}(x) = \nabla_x \tau^{(N)}(x, x') . \] (2.7)

Then, Eq.(2.6) constrains \( \gamma \) to lie on the sextic surface \( \mathcal{A}(x) \) given by

\[ \det(\rho \delta_{ik} - c_{ijkt}\gamma_t \gamma_j) = 0 . \] (2.8)

\( \mathcal{A}(x) \) consists of three sheets \( \mathcal{A}^{(N)}(x), N = 1, 2, 3, \) each of which is a closed surface surrounding the origin. An individual sheet is described by (cf. Eq.(2.8))

\[ 2 \mathcal{H} = \rho - \xi_i c_{ijkt}\gamma_t \gamma_j \xi_k = 0 . \] (2.9)

The scalar amplitudes \( A \) must satisfy the transport equation

\[ \partial_j(c_{ijkt}\xi_i^{(N)} \xi_k^{(N)} \mathcal{A}^{(N)})^2 \partial_t \tau^{(N)} = 0 , \] (2.10)

where \( N, \) again, indicates the mode of propagation, that is, the sheet of the slowness surface on which the corresponding slowness vector lies.

The characteristic or group velocities \( v^{(N)} \) are normal to \( \mathcal{A}^{(N)}(x) \) at \( \gamma^{(N)} \) and satisfy

\[ v^{(N)} \cdot \gamma^{(N)} = 1 ; \quad v^{(N)} = \frac{\nabla_x \gamma^{(N)} \mathcal{H}}{\gamma \cdot \nabla_x \gamma^{(N)} \mathcal{H} \bigg|_{\mathcal{H}=0}} . \] (2.11)

The normal or phase speeds are given by

\[ V^{(N)} = \frac{1}{|\gamma^{(N)}|} . \] (2.12)
The unit phase direction follows as

$$\alpha^{(N)} = V^{(N)} \gamma^{(N)}.$$  

From Eq.(2.11) it follows that

$$V^{(N)} = |v^{(N)}| \cos \chi,$$  

(2.13)

where $\chi$ is the angle between $v^{(N)}$ and $\gamma^{(N)}$.

The amplitudes can be expressed in terms of certain Jacobians,

$$A = \frac{1}{4\pi \rho(x) \rho(x') M^{1/2}} \quad \text{with} \quad M = \frac{|v(x') V(x) | \frac{\partial x}{\partial q_1} \wedge \frac{\partial x}{\partial q_2}}{\frac{\partial \gamma}{\partial q_1} \wedge \frac{\partial \gamma}{\partial q_2}},$$  

(2.14)

in which $A$ and $M$ carry the superscript $(N)$. Here, $(q_1, q_2)$ parameterize the rays originating from the source. One can verify that the dimension of $A$ is $[\text{time}]^2 \times [\text{mass}]^{-1}$, which upon multiplication by force, with dimensions $[\text{mass}] \times [\text{length}] \times [\text{time}]^{-2}$, gives the dimension of displacement, i.e., $[\text{length}]$.

**Source and receiver Green's functions**

In the integral representation for the scattered field, we need the Green's functions originating both at the source and the receiver points. Further, the gradient of total travel times from the source to a scattering point to the receiver are required in preparation of the GRT inversion. We introduce these functions here.

Set

$$\tilde{G}(x, t) = G(x, s, t), \quad \tilde{G}(x, t) = G(r, x, t).$$  

(2.15)

Employing asymptotic ray theory in both Green's functions, we introduce the notation

$$\tilde{A}^{(N)}(x) = A^{(N)}(x, s), \quad \tilde{A}^{(M)}(x) = A^{(M)}(r, x)$$  

(2.16)

in the case of scattering from incident mode $N$ to outgoing mode $M$.

According to Eq.(2.7), the slowness vectors at $x$ are given by

$$\tilde{\gamma}^{(N)}(x) = \nabla x_\tau^{(N)}(x, s) \quad \tilde{\gamma}^{(M)}(x) = \nabla x_\tau^{(M)}(r, x);$$  

(2.17)

the associated phase directions are given by

$$\tilde{\alpha}^{(N)} = \frac{\tilde{\gamma}^{(N)}}{|\tilde{\gamma}^{(N)}|}, \quad \tilde{\alpha}^{(M)} = \frac{\tilde{\gamma}^{(M)}}{|\tilde{\gamma}^{(M)}|}$$  

(2.18)
and the phase speeds (cf. Eq.(2.12)) are given by

\[ \hat{V}^{(N)} = \frac{1}{|\hat{\gamma}^{(N)}|}, \quad \hat{V}^{(M)} = \frac{1}{|\hat{\gamma}^{(M)}|}. \] (2.19)

We also define the two-way travel time \( T^{(N,M)} \) and its gradient,

\[ T^{(N,M)}(r, y, s) \equiv \tau^{(N)}(y, s) + \tau^{(M)}(r, y), \quad \Gamma^{(N,M)}(r, x, s) \equiv \nabla_x T^{(N,M)}(r, x, s). \] (2.20)

From Eq.(2.17) we see that

\[ \Gamma^{(N,M)}(r, x, s) = \hat{\gamma}^{(N)}(x) + \hat{\gamma}^{(M)}(x). \] (2.21)

The direction of \( \Gamma^{(N,M)} \),

\[ \nu = \frac{\Gamma^{(N,M)}}{||\Gamma^{(N,M)}||}, \]

will be the migration dip, which we referred to in the introduction. The different quantities are illustrated in Figure 2.1.
3 Medium description: micro-local perturbation

To analyze a typical geological setting, we consider the following, micro-local, representation,

$$c^{(1)}(x) \rightarrow c^{(1)}(x, \phi(x)),$$  \hspace{1cm} (3.1)

of the medium’s perturbation. Here, \(\phi\) is a smooth function of \(x\), some (curved) level surfaces of which describe a family of interfaces. The gradient of the perturbation is assumed to vary rapidly normal to the level surfaces of \(\phi\) and smoothly along them, implying that

$$\nabla_x c^{(1)} = (c^{(1)})' (\nabla_x \phi) + \text{smoother terms in } x, \quad (c^{(1)})' = \partial_y c^{(1)}.$$  \hspace{1cm} (3.2)

This representation extends to a small ball around any point, in particular the image point \(y\), say, under investigation. The derivative \((c^{(1)})'\) is understood in the distributional sense. Typically, it will have a Dirac-distribution type behavior across any geological interface. Loosely, the derivative can be interpreted as the difference in medium properties across a level surface.

We will omit the first argument of \(c^{(1)}\) in the remainder of this paper, and take only the most rapidly varying term of the perturbation’s derivatives into account. Imaging reflectors or interfaces amounts to mapping this leading-order behavior, \((c^{(1)})'\). We will confirm below that our inversion procedure does this, and also provides estimates of angularly dependent reflection coefficients.

4 The single scattering equation

In this section, we introduce the Born approximation representing the singly scattered wavefield. We then show how to recast this volume integral representation into a surface integral for the response to the most singular element of the scattering process, namely, reflecting the discontinuities of the medium parameters.

We begin the analysis with the volume-scattering representation of the ray-Born approximation for the scattered displacement field. Let \((r, s) \in \partial R \times \partial S\); ideally, the boundaries \(\partial R, \partial S \sim S^2\) are closed surfaces surrounding the heterogeneous domain \(D\) in which the medium is unknown. Then (De Hoop et al. [14])

$$u^{(1)}_{pq}(r, s, t) = - \int_{D_{pq}} \tilde{\epsilon}^{(\hat{N})}(r) \tilde{\epsilon}^{(\hat{N})}(s) A^{(\hat{N})}(x) \times$$
$$\left( w^{(\hat{N})}(x) \right)^T c^{(1)}(x) \delta''(t - T^{(\hat{N})}(r, x, s)) \, dx,$$  \hspace{1cm} (4.1)

where \(N, M \in \{1, 2, 3\}, \)

$$A^{(\hat{N})}(x) = \rho(x) \tilde{A}^{(\hat{N})}(x) \tilde{A}^{(\hat{N})}(x),$$  \hspace{1cm} (4.2)
contains the amplitudes,
\[ \mathbf{w}^{(N,M)} = \left\{ \xi_{N}^{(N)} \xi_{M}^{(M)}, \frac{1}{2} \left[ \tilde{a}_{ij}^{(N)} \tilde{a}_{kl}^{(N)} + \tilde{a}_{jk}^{(N)} \tilde{a}_{li}^{(N)} \right] \right\}, \]
\[ \tilde{a}_{ij}^{(M)} = \frac{1}{2} \nu_{ij}^{(M)} \left( \xi_{i}^{(N)} \xi_{j}^{(N)} + \xi_{j}^{(N)} \xi_{i}^{(N)} \right), \quad \tilde{a}_{kl}^{(N)} = \frac{1}{2} \nu_{kl}^{(N)} \left( \xi_{k}^{(N)} \xi_{l}^{(N)} + \xi_{l}^{(N)} \xi_{k}^{(N)} \right), \]
describes the contrast-source radiation patterns, and
\[ \mathbf{c}^{(1)} = \left\{ \rho^{(1)}, \frac{c_{ijkl}^{(1)}}{\rho \nu_{ijkl}^{(N)}} \right\}, \]
represents the relative medium perturbation. Here, \( \nu_{ijkl}^{(L)} \) denotes the (local) normal speed of mode \( L \) in the background medium for \( \gamma \) averaged over all phase directions. The notation \( \nu \) is meant to emphasize that the quantity is angle independent, which is important for retaining the actual medium perturbation from \( \mathbf{c}^{(1)} \) and the GRT inversion to be applicable.

In the micro-local setting, substituting Eq.(3.1) into Eq.(4.1), we have
\[ u_{pq}^{(1)}(r, s, t) = -\int_{D} \xi_{pq}^{(N)}(r) \xi_{qs}^{(N)}(s) A^{(N,M)}(x) \times \]
\[ \left( \mathbf{w}^{(N,M)}(x, \alpha^{(N)}(x), \alpha^{(M)}(x)) \right)^{T} \mathbf{c}^{(1)}(\phi(x)) \delta''(t - T(x)) \, dx, \quad (4.3) \]
where, for convenience, we employ the shorthand notation
\[ T(x) = T^{(N,M)}(r, x, s), \quad \Gamma(x) = (\nabla x T^{(N,M)})(r, x, s), \quad (4.4) \]
see Eq.(2.20). To make use of the properties of the gradient of the medium's perturbation (cf. Eq.(3.2)), we will partially integrate expression (4.3). Since
\[ (\nabla x T)(x) \delta''(t - T(x)) = -\nabla x \delta'(t - T(x)), \quad (4.5) \]
we have,
\[ \delta''(t - T(x)) = -\frac{1}{\nu \cdot \nabla x T(x)} \nu \cdot \nabla x \delta'(t - T(x)), \quad (4.6) \]
for arbitrary \( \nu \in S^{2} \) as long as \( \nu \cdot \nabla x T \neq 0 \). Hence,
\[ u_{pq}^{(1)}(r, s, t) \approx -\int_{D} \xi_{pq}^{(N)}(r) \xi_{qs}^{(N)}(s) A^{(N,M)}(x) \times \]
\[ \left( \mathbf{w}^{(N,M)}(x, \alpha^{(N)}(x), \alpha^{(M)}(x)) \right)^{T} \mathbf{c}^{(1)}(\phi(x)) \left( \frac{\nu \cdot \nabla x \phi}{(\nu \cdot \Gamma)} \right) \delta'(t - T(x)) \, dx. \]
Here, the approximation arises from neglecting lower order terms as in Eq.(3.2). These will produce asymptotically lower order contributions to the wavefield.

It is assumed that \( \tilde{\nu} = \nu(x) \) is slowly varying in space, and may be chosen equal to the local geological dip,

\[
\nu_{\phi} \equiv \frac{\nabla x \phi}{|\nabla x \phi|}.
\] (4.8)

On the other hand, we can choose \( \tilde{\nu} = \nu \), which we always know. In the inverse scattering problem, the geological dip is unknown and has to be determined. Below, we show that at stationarity the geological and migration dips must be parallel.

**Surface integral representation**

Now, we show how to recast the volume integral representation (4.7) into an integral over surface integrals over the level surfaces of \( \phi \). To this end, we choose curvi-linear coordinates \( \sigma = \sigma(x) \), \( \sigma = (\sigma_1, \sigma_2, \sigma_3) \), \( \mu = 1, 2 \), such that the \( \sigma_\mu \) are coordinates in the level surfaces of \( \phi \) and \( \sigma_3 \) is the local coordinate in the \( \nu_{\phi} \)-direction. If \( \sigma_3 \) represents the actual value of the level of \( \phi \), we set \( \sigma_3 = L \). The volume form is given by

\[
dx = \frac{1}{|\nabla x \phi|} \, dL \, d\Sigma(x), \quad d\Sigma(x) = |\partial_{\sigma_1} x \wedge \partial_{\sigma_2} x| \, d\sigma_1 d\sigma_2 ;
\]

the transformation from Cartesian to curvi-linear coordinates yields the Jacobian

\[
\frac{\partial(x)}{\partial(\sigma)} = |\partial_{\sigma_3} x \cdot (\partial_{\sigma_1} x \wedge \partial_{\sigma_2} x)| = \frac{|\partial_{\sigma_1} x \wedge \partial_{\sigma_2} x|}{|\nabla x \phi|}.
\] (4.9)

Now write

\[
(c^{(1)})'(\phi(x)) = \int_{\mathbb{R}} (c^{(1)})'(L) \delta(\phi(x) - L) \, dL .
\] (4.10)

Substituting Eq.(4.10) into Eq.(4.7), interchanging the order of integration, and using the property (cf. Eq.(4.9))

\[
\int_{\mathcal{D}} \ldots \delta(\phi(x) - L) \, dx = \int_{s=L} \ldots \frac{1}{|\nabla x \phi|} \, d\Sigma(x) ,
\]

we obtain

\[
u_{p,q}^{(1)}(r, s, t) 
\approx - \int_{\mathbb{R}} dL \left[ \int_{\mathcal{D}} \tilde{\xi}_p^{(N)}(s) A^{(N)}(x, \tilde{\nu}(\tilde{x})) \nu \tilde{\tau}(\tilde{x}) (z, \tilde{\nu}(\tilde{x})) \right] \left( c^{(1)} \right)'(L) \times 
\]

\[ \frac{(\vec{v} \cdot \nabla_x \phi)}{(\vec{v} \cdot \Gamma)} \bigg|_x \delta(\phi(x) - L) \delta'(t - T(x)) \, dx \]

\[ = - \int_{\mathbb{R}} dL \left[ \int_{\phi = L} \xi_{z_0}^{(M)}(r) \xi_{z_0}^{(N)}(s) A^{(N-M)}(x) \times (w^{(N-M)}(x, \tilde{\alpha}(N)(x), \tilde{\alpha}(M)(x)) \right] \right) \right) \right) \end{align}}

\[ \frac{(\vec{v} \cdot \nabla_x \phi)}{(\vec{v} \cdot \Gamma)} \bigg|_x \delta'(t - T(x)) \, d\Sigma(x) \right] \]

\[ - \int_{\mathbb{R}} dL \left[ \int_{\phi = L} \xi_{z_0}^{(M)}(r) \xi_{z_0}^{(N)}(s) A^{(N-M)}(x) \times (w^{(N-M)}(x, \tilde{\alpha}(N)(x), \tilde{\alpha}(M)(x)) \right] \right) \right) \end{align}}

\[ \frac{(\vec{v} \cdot \nabla_x \phi)}{(\vec{v} \cdot \Gamma)} \bigg|_x \delta'(t - T(x)) \, d\Sigma(x) \right] \]

(4.11)

As a consequence of Eq.(4.11), we also have

\[ \partial_t \nu^{(1)}_{p} (r, s, t) = - \int_{\mathbb{R}} dL \left[ \int_{\phi = L} \xi_{z_0}^{(M)}(r) \xi_{z_0}^{(N)}(s) A^{(N-M)}(x) \times (w^{(N-M)}(x, \tilde{\alpha}(N)(x), \tilde{\alpha}(M)(x)) \right] \right) \right) \end{align}}

\[ \frac{(\vec{v} \cdot \nabla_x \phi)}{(\vec{v} \cdot \Gamma)} \bigg|_x \delta''(t - T(x)) \, d\Sigma(x) \right] \]

(4.12)

The singular support of the medium perturbation, \( \phi(x) = L \), and the isochrone surfaces, \( T(x) = t \), are illustrated in Figure 4.1.

5 Reflection and transmission coefficients

To identify reflection and transmission coefficients in Eq.(4.12), we first extract phase velocities at the scattering point from the amplitudes,

\[ A_u(x) = A^{(N-M)}(x) \left[ \tilde{\nabla}^{(N)}(x) \tilde{\nabla}^{(M)}(x) \right] \right) \right) \end{align}}

\[ d \Sigma(x) \right] \]

then, Eq.(4.12) can be written in the form

\[ \partial_t \nu^{(1)}_{p} (r, s, t) = - \int_{\mathbb{R}} dL \left[ \int_{\phi = L} \xi_{z_0}^{(M)}(r) \xi_{z_0}^{(N)}(s) A_u(x) \times (w^{(N-M)}(x, \tilde{\alpha}(N)(x), \tilde{\alpha}(M)(x)) \right] \right) \right) \end{align}}

\[ \frac{(\vec{v} \cdot \nabla_x \phi)}{(\vec{v} \cdot \Gamma)} \bigg|_x \delta''(t - T(x)) \, d\Sigma(x) \right] \]

\[ R_{L}^{(N-M)}(x, \tilde{\alpha}(N)(x), \tilde{\alpha}(M)(x)) \right) \right) \right) \end{align}}

\[ \frac{(\vec{v} \cdot \nabla_x \phi)}{(\vec{v} \cdot \Gamma)} \bigg|_x \delta''(t - T(x)) \, d\Sigma(x) \right] \]

(5.2)
in which,

$$R_L^{(N,M)}(\cdot, \tilde{\alpha}^{(N)}(\cdot), \tilde{\alpha}^{(M)}(\cdot)) = \frac{(w^{(N,M)})^T (c^{(1)})' \nabla \phi}{\left[ \tilde{\nu}^{(N)}(\tilde{\nu}^{(M)})^3 \right]^{1/2} \Gamma^2 (\vec{v} \cdot \nu)^2} ,$$

(5.3)

represents the scattering coefficient for the $N$, $M$ conversion at $x$ with $\phi(x) = L$, i.e., $R_L^{(N,M)}$ really depends on $\sigma_\mu(x)$, $\mu = 1,2$. Observe that the scattering coefficient contains a function that may be singular on the level surfaces of $\phi$. In fact, the integration over $L$ picks up the singular support of $(c^{(1)})'$.

**Specular reflection and transmission**

Now, let $c^{(1)}$ contain a step function in $L$ (across a curved interface at $L = L_0$). Then $(c^{(1)})'(L) = (\Delta c^{(1)}) \delta(L - L_0)$. At the specular point for given $\nu_\phi$, the pair $\tilde{\alpha}^{(N)}(\cdot), \tilde{\alpha}^{(M)}(\cdot)$ satisfies Snell's law,

$$\tilde{\alpha}^{(M)}(\cdot) \cdot (I - \nu_\phi \nu_\phi) = -\tilde{\alpha}^{(N)}(\cdot) \cdot (I - \nu_\phi \nu_\phi) .$$

(5.4)

(In an isotropic medium with $M = N$ the solution is simply given by

$$\tilde{\alpha}_s^{(N)} = -\tilde{\alpha}^{(N)}(\cdot) \cdot (I - 2\nu_\phi \nu_\phi) .$$
representing ordinary reflection). At the specular point, the migration dip coincides with the geological dip, \( \nu = \nu_\phi \).

Substituting the solution, \( \hat{\alpha}(N) \), into the scattering matrix Eq.(5.3), yields the linearized reflection/transmission coefficients \( R_{L_0}^{(N,M)}(\nu, \hat{\alpha}(N)(\nu)) \delta(L - L_0) = R_L^{(N,M)}(\nu, \hat{\alpha}(N)(\nu), \hat{\alpha}(N)(\nu)) \) .

In fact, with \( \bar{\nu} = \nu_\phi \), Eq.(5.2) is equivalent to the Kirchhoff-Born approximation, which satisfies the principle of reciprocity. For notational convenience, we will absorb the Dirac distribution in \( R_L^{(N,M)} \).

The Kirchhoff approximation

Here, we show how to go from the linearized Born representation Eq.(4.12) to the non-linear Kirchhoff approximation. For general reference, we introduce the relevant scattering and specular angles: in addition to \( \nu \), we set

\[
\cos \theta = \hat{\alpha}(\bar{N}) \cdot \hat{\alpha}(\bar{M}), \quad \psi = \text{third Euler angle}.
\]

Then, at any point in \( D \), we have a mapping

\[
(\hat{\alpha}(\bar{N})(\nu), \hat{\alpha}(\bar{M})(\nu)) \rightarrow (\nu, \theta, \psi).
\]

If \( \nu = \nu_\phi \), the ray geometry is specular; let the associated specular scattering angles be defined as

\[
\cos \bar{\theta} = \hat{\alpha}(\bar{N}) \cdot \nu_\phi, \quad \cos \bar{\theta} = \hat{\alpha}(\bar{M}) \cdot \nu_\phi, \quad \theta_\phi = \bar{\theta} - \bar{\phi}.
\]

In the non-reciprocal, ‘non-linear’ Kirchhoff approximation, the scattering matrix is simply replaced by the full reflection/transmission coefficients at specular, cf. Eq.(5.2) with \( \bar{\nu} = \nu_\phi \),

\[
\partial_t u_{pq}^{(1)}(r, s, t) = -\int_{L_0} \left[ \int_{\phi_L} \hat{\varepsilon}_p^{(M)}(r) \hat{\varepsilon}_q^{(N)}(s) A_u(x) \times \right.
\]

\[
R_L^{(N,M)}(x, \hat{\alpha}(N)(x)) (\nu_\phi \cdot \Gamma) \delta''(t - T(x)) \left\{ \frac{d\Sigma(x)}{\nabla x_\phi} \right\} dL.
\]

This representation is more adequate than Eq.(5.2) for wide-angle scattering. The time derivative is taken to pave the way for the Radon transform inversion. (In three dimensions, one needs the second derivative of the Dirac distribution.) In our further analysis, we actually employ the reciprocal representation Eq.(5.2) in which \( R_L^{(N,M)} \) is replaced by the full reflection/transmission coefficient at specular. Note that due to the singular function contained in \( R_L^{(N,M)} \), cf. Eq.(5.5), the integration over \( L \) in Eq.(5.9) reduces to a sum over scattering surfaces or interfaces.
6 Stationary phase analysis of the direct scattering problem

By applying stationary phase arguments, the integral in Eq.(5.2) can be evaluated. The analysis confirms the consistency with asymptotic ray theory in configurations with a family of surface scatterers.

We choose rotated Cartesian coordinates $(x_\mu, z)$, $\mu = 1, 2$ in the neighborhood of a yet to be determined specular point $y(L) \in \{ \phi = L \}$, such that

$$z \parallel \nu_\phi, \quad \{x_\mu\} \perp \nu_\phi$$

(6.1)

and

$$\nu = \nu_\phi$$

see Figure 4.1. The function $T(y(L))$ has a derivative given by

$$\frac{dT}{dL} = \frac{|\nabla x T|}{|
abla x \phi|} y(L),$$

(6.2)

while, by the implicit function theorem, the function $L y(t)$ satisfying

$$T \left( y(L y(t)) \right) = t$$

exists.

Taylor expansions of the level and isochrone surfaces including the curvature terms yield

$$0 = \phi(x) - \phi(y) = |\nabla x \phi(y) \cdot z + \frac{1}{2} x_\mu \phi_{\mu\nu}(y) x_\nu; \quad \phi_{\mu\nu}(y) = \frac{\partial^2 \phi}{\partial x_\mu \partial x_\nu} y$$

$$T(x) = T(y) + |\nabla x T(y) \cdot z + \frac{1}{2} x_\mu T_{\mu\nu}(y) x_\nu; \quad T_{\mu\nu}(y) = \frac{\partial^2 T}{\partial x_\mu \partial x_\nu} y.$$  

(6.3)

(Summations are carried out over $\mu, \nu$.) The first equality in (6.3) amounts to the representation of the level surface $\{ \phi = L \}$,

$$z = -\frac{x_\mu \phi_{\mu\nu} x_\nu}{2 |\nabla x \phi|},$$

which upon substitution in the second equality yields

$$T(x) = T(y) + \frac{i}{2} |\nabla x T(y) \cdot x_\mu \gamma_{\mu\nu}(y) x_\nu$$

(6.4)

with

$$\gamma_{\mu\nu} = \frac{T_{\mu\nu}}{|\nabla x T|} - \frac{\phi_{\mu\nu}}{|\nabla x \phi|}$$
and \( x, y \) in the same level surface. Note that the matrix \( \Upsilon \) may vary with the level \( L \), and can be negative or positive definite, or indefinite. The case of vanishing \( \Upsilon \), leading to a caustic analysis, will be postponed to a future paper. Otherwise, for \( t \) near \( T(y) \), we have the intermediate result

\[
\partial^2_t \int_{\phi=L} H(t - T(x)) \, d\Sigma(x)
= \partial^2_t \int_{\mathbb{R}^2} H(t - T(y) - \frac{1}{2} \nabla_x T(y) \cdot x, \Upsilon_{\mu \nu}(y)x_\nu) \, dx_1 \, dx_2
= \frac{2\pi \delta^*(t - T(y))}{|\Gamma(y)| \sqrt{|\det(\Upsilon(y))|}} ,
\]

where

\[
\delta^* = \begin{cases} 
\delta & \text{if } \Upsilon \text{ positive} \\
\mathcal{H}\delta & \text{if } \Upsilon \text{ indefinite} \\
-\delta & \text{if } \Upsilon \text{ negative}
\end{cases}
\]

In the above \( \mathcal{H} \) denotes the Hilbert transform; the notation \( {}^* \) indicates an action on the time dependence.

We will use Eq.(6.5) to evaluate the integral in Eq.(4.11) eventually. First, note that Eq.(6.5) implies

\[
\int_{\mathbb{R}} dL \left[ \partial^2_t \int_{\phi=L} (c^{(1)})'(L) \left( \frac{\bar{\nu} \cdot \nu_\phi}{(\bar{\nu} \cdot \Gamma)} \right) H(t - T(x)) \, d\Sigma(x) \right]
= \int_{\mathbb{R}} dL \left( c^{(1)}(L) \frac{2\pi}{|\Gamma| \sqrt{|\det(\Upsilon)|}} \left( \frac{\bar{\nu} \cdot \nu_\phi}{(\bar{\nu} \cdot \Gamma)} \right) y(L)^{\delta^*[t - T(y(L))]} \right)
= \frac{2\pi}{|\Gamma| \sqrt{|\det(\Upsilon)|}} \left( \frac{\bar{\nu} \cdot \nu_\phi}{(\bar{\nu} \cdot \Gamma)} \right) y(Ly(t)) \left( \frac{dL}{dL} \right)^{-1} (c^{(1)})^* ,
\]

where, for any \( f \),

\[
\left( \frac{dL}{dL} \right)^{-1} f \bigg|_{Ly(t)}^* = \int_{\mathbb{R}} f(L) \delta^*[t - T(y(L))] \, dL .
\]

Substituting Eq.(6.2) into Eq.(6.7), and using the result in Eq.(4.11) implies

\[
u^{(N,M)}_{(s,t)}(r, \Phi, s) = -\frac{2\pi |\nabla x_\phi|}{|\det(\Upsilon)| |\Gamma|^3} \left( \xi^{(N,M)}(r, \Phi, s) A^{(N,M)}(s) \times (w^{(N,M)}(\cdot, \Phi(\cdot), \Phi_s(\cdot)))^T y(Ly(t)) (c^{(1)})^* \bigg|_{Ly(t)}^* ,
\]

\[
\equiv u^{(N,M)}_{ps}(r, \Phi, s)
\]

\[
\equiv u^{(N,M)}_{ps}(r, \Phi, s) = -\frac{2\pi |\nabla x_\phi|}{|\det(\Upsilon)| |\Gamma|^3} \left( \xi^{(N,M)}(r, \Phi, s) A^{(N,M)}(s) \times (w^{(N,M)}(\cdot, \Phi(\cdot), \Phi_s(\cdot)))^T y(Ly(t)) (c^{(1)})^* \bigg|_{Ly(t)}^* ,
\]

\[
\equiv u^{(N,M)}_{ps}(r, \Phi, s)
\]

\[
\equiv u^{(N,M)}_{ps}(r, \Phi, s) = -\frac{2\pi |\nabla x_\phi|}{|\det(\Upsilon)| |\Gamma|^3} \left( \xi^{(N,M)}(r, \Phi, s) A^{(N,M)}(s) \times (w^{(N,M)}(\cdot, \Phi(\cdot), \Phi_s(\cdot)))^T y(Ly(t)) (c^{(1)})^* \bigg|_{Ly(t)}^* ,
\]

\[
\equiv u^{(N,M)}_{ps}(r, \Phi, s)
\]
since at the specular point we have $\nu = \nu_\phi$ and $\nu_\phi \cdot \Gamma = |\Gamma|$. This formula is an extension of the convolutional model approximation in one-dimensional space to three dimensions.

In terms of the linearized reflection/transmission coefficients, we have

$$u_{pq}^{(\tilde{N} \tilde{M})}(r, s, t) = -\frac{2\pi}{\sqrt{\det(Y)|\Gamma|}} \tilde{\xi}_p^{(\tilde{N} \tilde{M})}(r) \tilde{\xi}_q^{(\tilde{N} \tilde{M})}(s) A \xi_{\xi}(\cdot) R_{L}^{(\tilde{N} \tilde{M})} |y = y(L), L = Ly(t) \text{ } (6.9)$$

Note that the $^*$ relates to the KMAH index.

## 7 Inversion based on the GRT

In preparation of the inverse transformation, we introduce the scalar quantity

$$\partial_t u^{(\tilde{N} \tilde{M})}(r, s, y) = \frac{\tilde{\xi}_p^{(\tilde{N} \tilde{M})}(r) \partial_t u_{pq}^{(\tilde{N} \tilde{M})}(r, s, t) T^{(\tilde{N} \tilde{M})}(r, y, s) \tilde{\xi}_q^{(\tilde{N} \tilde{M})}(s)}{A_n(y)}. \text{ } (7.1)$$

Then, using Eq.(5.2), we get

$$\partial_t u^{(\tilde{N} \tilde{M})}(r, s, y) \simeq -\int_\mathbb{R} \int_{x=L} R_{L}^{(\tilde{N} \tilde{M})}(x, \tilde{\alpha}^{(\tilde{N} \tilde{M})}(x), \tilde{\chi}^{(\tilde{N} \tilde{M})}(x)) (\tilde{\nu} \cdot \nu_\phi)(\nu \cdot \Gamma) |\nabla^2 \phi| |x \times$$

$$\delta''(T(y) - T(x)) \frac{\text{d}\Sigma(x)}{|\nabla^2 \phi|} \text{d}L; \text{ } (7.2)$$

we expand

$$T(y) - T(x) = \Gamma(y) \cdot (y - x) + \ldots \simeq \Gamma(y) |y - x| \cdot \nu, \text{ } (7.3)$$

which describes the isochrone tangent plane at $y$. We have

$$\delta''(|\Gamma(y)||y - x| \cdot \nu) = |\Gamma(y)|^{-3} \delta''((y - x) \cdot \nu).$$

**Linearized inversion**

The basis of the GRT inversion is Gel'fand's plane wave expansion, which we write in the form

$$\int_{S^2} \int_{\phi=L} \delta''((y - x) \cdot \nu) \frac{\text{d}\Sigma(x)}{|\nabla^2 \phi|} \text{d}\nu = \int_{S^2} \int_D \delta(\phi(x) - L) \delta''((y - x) \cdot \nu) \text{d}x \text{d}\nu$$

$$= -8\pi^2 \delta(\phi(y) - L). \text{ } (7.4)$$

We will use this expansion to invert Eq.(4.12). The inversion is accomplished by setting up a system of 22 equations, using the simplification according to Eq.(7.1), and integrating the result.
both over $\hat{\alpha}$ and $\tilde{\alpha}$, i.e.,

$$w^{(\tilde{N}M)}(y, \hat{\alpha}^{(\tilde{N})}(y), \hat{\alpha}^{(M)}(y)) \partial_t u^{(\tilde{N}M)}(r, s, y) \left| \frac{\Gamma^4}{(\nu \cdot \nu_\phi)} \right| y \frac{\partial(\hat{\alpha}, \tilde{\alpha})}{\partial(s, r)} \right| y$$

$$= -\int_{\mathbb{R}^3} dL \left[ \int_{\mathbb{S}^2} w^{(\tilde{N}M)}(y, \hat{\alpha}^{(\tilde{N})}(y), \hat{\alpha}^{(M)}(y)) w^{(\tilde{N}M)}(x, \hat{\alpha}^{(\tilde{N})}(x), \hat{\alpha}^{(M)}(x))^T (c^{(1)})'(L) \times \right]$$

$$\left| \begin{array}{c} (\nu \cdot \Gamma) \\ \nu \cdot \nu_\phi \\ y \end{array} \right| x \left| \begin{array}{c} \Gamma(y)^3 \\ \partial(\hat{\alpha}, \tilde{\alpha}) \\ \partial(\nu, \theta, \psi) \end{array} \right| \delta''((y - x) \cdot \nu) \, d\Sigma(x) \, d\nu \, d\theta \, d\psi. \quad (7.5)$$

Define the matrix

$$\Lambda_y(y) \equiv \int_{\mathbb{S}^2} w^{(\tilde{N}M)}(., \hat{\alpha}^{(\tilde{N})}(.), \hat{\alpha}^{(M)}(.)^T \partial_{(\nu, \theta, \psi)} \left| \begin{array}{c} \partial(\hat{\alpha}, \tilde{\alpha}) \\ \partial(\nu, \theta, \psi) \end{array} \right| d\nu d\theta d\psi. \quad (7.6)$$

at any image point $y$. Then, employing Eq.(7.4) in Eq.(7.5) and setting $\tilde{\nu} = \nu$ (recall that $\tilde{\nu}$ has been arbitrary until now), yields

$$(c^{(1)})'(\phi(y)) | \nabla_x \phi | y = \int_{\mathbb{R}^3} (c^{(1)})'(L) \delta(\phi(y) - L) | \nabla_x \phi | y \, dL \approx \quad (7.7)$$

$$\frac{1}{8\pi^2} \int_{\partial S \times \partial R} \Lambda_{y}^{-1} \left| \begin{array}{c} (c^{(1)})'(L) \\ (\nu \cdot \nu_\phi) \end{array} \right| y \left| \begin{array}{c} \Gamma^4 \\ \nu \cdot \nu_\phi \end{array} \right| y \frac{\partial(\hat{\alpha}, \tilde{\alpha})}{\partial(s, r)} \right| y \right| d\nu d\theta d\psi.$$
Resolution analysis

Backsubstituting the Kirchhoff-Born scattering formula Eq.(4.12) with \( \vec{\nu} = \nu \), into the inversion formula Eq.(7.8) and substituting the one-sided Fourier representation of the Dirac distribution, yields the resolution operator,

\[
\langle (c^{(1)})' (\phi(y)) | \nabla x \phi | y \rangle = \int_{\mathbb{R}^3} \int_{\Omega = L} \mathcal{R}(y, x) (c^{(1)})' (\phi(x)) | \nabla x \phi | x \frac{d\Sigma(x)}{\nabla x \phi | x} \, dL, \tag{7.9}
\]

with matrix kernel \( \mathcal{R} \),

\[
\mathcal{R}(y, x) = \text{Re} \sum_j \frac{1}{8\pi^3} \int \exp[i\Phi(y, x, \Theta)] a(y, x, \Theta) \chi_j(\Theta) d\Theta. \tag{7.10}
\]

Here, \( \Theta = (\omega, \mathbf{s}, \mathbf{r}) \) and

\[
\int_{\mathbb{R}^7 \times \partial S \times \partial R} \cdots d\Theta = \int_{\partial S \times \partial R} \int_{\mathbb{R}^7} \cdots |\Gamma^{(N)}(r, y, s)|^3 \omega^2 \, d\omega \, ds \, dr. \tag{7.11}
\]

The resolution operator expresses how well the reconstruction can be accomplished within the framework of the linear theory.

The phase function \( \Phi \) of the Fourier integral operator with kernel Eq.(7.10) is simply given by

\[
\Phi(y, x, \Theta) \equiv \omega \left( \Gamma^{(N)}(r, y, s) - \Gamma^{(N)}(r, x, s) \right), \tag{7.12}
\]

while the amplitude function arises as the matrix

\[
a(y, x, \Theta) \equiv \frac{A_y(x)}{A_y(y)} \xi_p^{(N)}(r, y) \xi_q^{(N)}(r, x) \xi_q^{(N)}(s, y) \xi_q^{(N)}(s, x) \times
\]

\[
\left[ (\nu^{(N)}(r, y, s))^{-1} w^{(N)}(r, y, \hat{\alpha}^{(N)}(y), \hat{\alpha}^{(N)}(y)) \right] \times
\]

\[
(w^{(N)}(r, \hat{\alpha}^{(N)}(x), \hat{\alpha}^{(N)}(x))). \tag{7.13}
\]

In anticipation of introducing the scattering coefficients (cf. Eq.(5.3)), we introduce the one-dimensional array of functions \( a_R \), satisfying

\[
a_R(y, x, \Theta) R_L^{(N)}(x, \hat{\alpha}^{(N)}(x), \hat{\alpha}^{(N)}(x)) = a(y, x, \Theta) (c^{(1)})' (\phi(x)) | \nabla x \phi | x. \tag{7.14}
\]
We will interpret $\Theta$ in the spatial Fourier domain. To this end, we carry out two coordinate transformations. First, we employ the ray induced mapping

$$s = s(\tilde{N}, \tilde{M}, \nu, \theta, \psi), \quad r = r(\tilde{N}, \tilde{M}, \nu, \theta, \psi) \quad \text{for fixed } y,$$

(7.15)

with

$$\int_{\partial S^2 \times \partial R} \frac{\partial (\tilde{\alpha}, \tilde{\alpha})}{\partial (s, r)} \bigg|_y ds dr = \int_{S^2 \times S^2} \frac{\partial (\tilde{\alpha}, \tilde{\alpha})}{\partial (\nu, \theta, \psi)} \bigg|_y d\nu d\theta d\psi.$$

Thus, at $y$, $\Theta$ is mapped on $(\omega, \nu, \theta, \psi)$, and we set

$$a(y, x, \omega, \nu, \theta, \psi) \frac{\partial (\tilde{\alpha}, \tilde{\alpha})}{\partial (s, r)} \bigg|_y = a(y, x, \omega, s, r) \frac{\partial (\tilde{\alpha}, \tilde{\alpha})}{\partial (\nu, \theta, \psi)} \bigg|_y.$$

We identify

$$T^{(N\tilde{M})}(r, x, s) \quad \text{with} \quad T^{(N\tilde{M})}(x, \nu, \theta, \psi),$$

which function determines also the dual generalised Radon transform.

Second, the frequency $\omega$ is transformed to the wavenumber $k_z$ according to

$$k_z \equiv \omega |\Gamma^{(N\tilde{M})}(r, y, s)|;$$

(7.16)

then

$$\int_{\mathbb{R}^+} \cdots |\Gamma^{(N\tilde{M})}(r, y, s)|^3 \omega^2 \, d\omega = \int_{\mathbb{R}^+} \cdots k_z^2 \, dk_z.$$

(7.17)

We now identify the wave vector

$$\Theta' \equiv k_z \nu \in \mathbb{R}^3 \quad \text{with} \quad d\Theta' = k_z^2 \, dk_z \, d\nu,$$

(7.18)

and we will consider $(\theta, \psi)$ as parameters, i.e., $\Theta \rightarrow (\omega, \nu, \theta, \psi) \rightarrow (\Theta', \theta, \psi)$. The Jacobian of the latter transformation is written as (cf. Eq.(7.17))

$$\frac{\partial (\Theta')}{\partial (\omega, \nu)} \bigg|_y = h(y, \nu) \omega^2, \quad h(y, \nu) = |\Gamma^{(N\tilde{M})}(r, y, s)|^3;$$

formally,

$$h = |\det \left( \begin{array}{ccc} \Gamma^{(N\tilde{M})} & \partial_{\nu_1} \Gamma^{(N\tilde{M})} & \partial_{\nu_2} \Gamma^{(N\tilde{M})} \end{array} \right)|.$$

(7.19)

The inverse transformation to frequency, the so-called Stolt mapping, is given by

$$\omega(\Theta') = \frac{\Theta' \cdot \Gamma^{(N\tilde{M})}(r, y, s)}{|\Gamma^{(N\tilde{M})}(r, y, s)|^2},$$

(7.20)
since
\[ \Theta' = \omega \Gamma^{(K,M)}(r, y, s). \]

We identify
\[ \Phi(y, x, \Theta) \text{ with } \Phi(y, x, \Theta', \theta, \psi). \]

In the phase space with coordinates \((x, \Theta')\), Eq.(7.10) gives rise to the resolution equation
\[
\langle (c^{(1)})'(\phi(y)) | \nabla x \phi | y \rangle = \Re \sum_j \int_{S^2} \frac{1}{8\pi^3} \int_{\mathbb{R}^+} \int_{\Theta = L} \int_{S^2} 
\exp[i\Phi(y, x, \Theta', \theta, \psi)] a(y, x, \Theta', \theta, \psi) \chi_j(\Theta', \theta, \psi) d\Theta' \times 
\frac{d\Sigma(x)}{|\nabla x \phi| x} dL \, d\theta \, d\psi. \tag{7.21}
\]

In this expression, for \(x\) near \(y\), the dependencies of \(\Phi\) on \(\theta, \psi\) disappear to leading order:
\[ \Phi = \Theta' \cdot (y - x) + \cdots. \]
The integration over \(\Theta'\) represents the spatial resolution, whereas the integration over \(\theta, \psi\) represents the parameter resolution per migration dip. Below, we will constrain \(\chi_j\) to be a function of \(k_r = |\Theta'|, \theta, \psi\) alone.

**Stationary phase analysis of the resolution operator**

Inside the integral over \(\omega\) in Eqs.(7.10)-(7.11), we consider \(\omega\) to be large. Then, we apply a four-dimensional stationary phase analysis with respect to the integrations over \((\sigma_1, \sigma_2, \nu) \in S(L) \times S^2\), cf. Eq.(7.21).

We choose polar coordinates on the \(\nu\)-sphere,
\[ \nu = (\sin \theta' \cos \psi', \sin \theta' \sin \psi', \cos \theta'). \]

We extract \(L = \sigma_3\) and \(k_r = |\Theta'|\) from the set of phase space coordinates,
\[ (x, \Theta') \rightarrow (\sigma_1, \sigma_2, L, k_r, \theta', \psi'), \quad \eta \equiv (\sigma_1, \sigma_2, \theta', \psi'). \]

resubstitute \(k_r = \omega |\Gamma(y)|\), and set
\[ \Phi(y, x, \Theta', \theta, \psi) = \omega \Phi'(y, L, \eta, k_r, \theta, \psi). \]

Writing the coordinates explicitly, the resolution equation (7.21) takes the form
\[
\langle (c^{(1)})'(\phi(y)) | \nabla x \phi | y \rangle = \Re \sum_j \int_{S^2} \frac{1}{8\pi^3} \int_{\mathbb{R}^+} \int_{S(L) \times S^2} 
\exp[i\omega \Phi'(y, L, \eta, k_r, \theta, \psi)] a(y, L, \eta, k_r, \theta, \psi) \chi_j(k_r, \theta, \psi) h(y, \nu) \times 
\frac{d\eta}{|\nabla x \phi| x} \frac{d\sigma_3}{dL} \omega^2 \, d\omega \, dL \, d\theta \, d\psi, \tag{7.22}
\]
where
\[ d\eta = \Sigma(x) \sin \theta' d\theta' d\psi' \]
\[ \tag{7.23} \]
For given \((\theta, \psi)\), the phase \(\Phi'\) is stationary with respect to the variables of integration \(\eta\) if
\[ \partial_{\sigma_{\mu}} \Phi' = 0 \quad \text{and} \quad \partial_{(\theta', \psi')} \Phi' = 0 \]
\[ \tag{7.24} \]
here,
\[ \partial_{\sigma_{\mu}} \Phi' = -\Gamma \cdot \partial_{\sigma_{\mu}} x \]
\[ \tag{7.25} \]
while
\[ \partial_{\theta_{\nu}} \Phi' = \left[ \gamma^{(M)}(y) - \gamma^{(M)}(x) \right] \cdot \partial_{\theta_{\nu}} r + \left[ \gamma^{(N)}(y) - \gamma^{(N)}(x) \right] \cdot \partial_{\theta_{\nu}} s \]
\[ \tag{7.26} \]
\[ \partial_{\psi_{\nu}} \Phi' = \left[ \gamma^{(M)}(y) - \gamma^{(M)}(x) \right] \cdot \partial_{\psi_{\nu}} r + \left[ \gamma^{(N)}(y) - \gamma^{(N)}(x) \right] \cdot \partial_{\psi_{\nu}} s \]
\[ \tag{7.27} \]
The stationary points are denoted as \(\sigma^{0}_{\mu} = \sigma^{0}_{\mu}(y)\) and induce the mapping \(x(\sigma^{0}_{\mu}, L)\) for any \(L\). The solution of the first equation in (7.24) with (7.25) implies that the stationary migration dip, \(\nu^{0}\), must be parallel to the geological dip, \(\nu_{\phi}\), i.e.,
\[ \nu^{0} = \pm \nu_{\phi} \]
\[ \tag{7.28} \]
one solution, \(\sigma^{0}_{\mu}\), of the second equation is easy to identify:
\[ x(\sigma^{0}_{\mu}, L) = y \]
at this value one expects the peak contribution. At \(x(\sigma^{0}_{\mu}, L)\), given \(\nu^{0}\), \((\theta, \psi)\) will determine the stationary values of \((s, r)\). We denote the stationary points by \(\eta^{0} = (\sigma^{0}_{\mu}, \nu^{0}(x(\sigma^{0}_{\mu}, L)))\), and the stationary point set by \(H^{0}\), which contains at least two elements (cf. Eq.(7.28)).

Applying the four-dimensional stationary phase approximation to Eq.(7.22) amounts to
\[ \left( (c^{(1)})'(\phi(y)) |v x \phi| y \right) = \sum_{\eta_{j} \in H^{0}} \text{Re} \sum_{j} \int_{S^{2}} \frac{1}{8\pi^{3}} \int_{\mathbb{R}} \int_{\mathbb{R}^{+}} \left( \frac{2\pi}{\omega} \right)^{2} \]
\[ \exp \left[ i \omega \Phi'(y, L, \eta^{0}, k_{r}, \theta, \psi) + i \frac{\pi}{4} \text{sign}(\nabla_{\eta} \nabla_{\eta} \Phi') \right] \frac{\alpha(y, L, \eta^{0}, k_{r}, \theta, \psi)}{\sqrt{\text{det}(\nabla_{\eta} \nabla_{\eta} \Phi')^{0}} \}
\chi_{J}(k_{r}, \theta, \psi) h(y, \nu^{0}) \times
\]
\[ \left( c^{(1)} x(\sigma^{0}_{\mu}, L)) |v x \phi| x(\sigma^{0}_{\mu}, L) \right) \frac{\partial_{\sigma} x \wedge \partial_{\sigma} x}{|\nabla x \phi|} \left| x(\sigma^{0}_{\mu}, L) \right| \omega^{0} d\omega dL d\theta d\psi \].
\[ \tag{7.29} \]
Here \( \text{sig} \) denotes the number of positive eigenvalues minus the number of negative eigenvalues of a matrix. In Eq. (7.29), \( k_r \) is the stretch of frequency with \( |\Gamma(y)|^0 \), the norm of the gradient of travel time in case the migration dip is stationary. In terms of scattering coefficients (cf. Eq. (7.14)), for \( x(\sigma^0_\mu, L) \) near \( y \in \{ \phi = L \} \), expression (7.29) reduces to

\[
\langle (c^{(i)})'(\phi(y)) \mid \nabla x \phi \mid y \rangle \simeq \sum_{\nu^0 = \pm \nu_\phi} \frac{i}{2} \sum_J \int_{S^2} \int_{\mathbb{R}} \text{Re} \frac{1}{\pi} \int_{\mathbb{R}^+} \exp \left[ i\omega \Phi'(y, L, \eta^0, k_r, \theta, \psi) + i\frac{\tau}{4} \text{sig} \left( \nabla_\eta \nabla_\eta \Phi' \right)^0 \right] \chi_J(k_r, \theta, \psi) dk_r \times \\
\frac{a_R(y, L, \eta^0, k_r, \theta, \psi) R^{(NM)}_L(\eta^0, k_r, \theta, \psi) h(y, \nu^0)}{|\Gamma(y)|^0 \sqrt{\text{det} \left( \nabla_\eta \nabla_\eta \Phi' \right)^0}} \times \\
\left| \frac{\partial_\sigma \mathbf{x} \wedge \partial_{\sigma^2} \mathbf{x}}{|\nabla x \phi|} \right| x(\sigma^0_\mu, L) \quad dL \quad d\theta \quad d\psi. 
\]

(7.30)

Recognizing the bandlimited Dirac distribution,

\[
I_{\Delta, J}(\nu^0 \cdot (y - x)) \equiv \frac{1}{\pi} \times \\
\int_{\mathbb{R}^+} \exp \left[ i\omega \Phi'(y, L, \eta^0, k_r, \theta, \psi) + i\frac{\tau}{4} \text{sig} \left( \nabla_\eta \nabla_\eta \Phi' \right)^0 \right] \chi_J(k_r, \theta, \psi) dk_r 
\]

we find that the reconstruction amounts to a weighted integration of scattering coefficients.

\[
\langle (c^{(i)})'(\phi(y)) \mid \nabla x \phi \mid y \rangle \simeq \sum_{\nu^0 = \pm \nu_\phi} \frac{i}{2} \sum_J \int_{E_\theta \times E_\psi} \int_{\mathbb{R}} \text{Re} I_{\Delta, J}(\nu^0 \cdot (y - x)) \frac{1}{|\nabla x \phi|} \left| x(\sigma^0_\mu, L) \right| \times \\
\frac{a_R(y, L, \eta^0, k_r, \theta, \psi) R^{(NM)}_L(\eta^0, k_r, \theta, \psi) h(y, \nu^0)}{|\Gamma(y)|^0 \sqrt{\text{det} \left( \nabla_\eta \nabla_\eta \Phi' \right)^0}} \times \\
\left| \frac{\partial_\sigma \mathbf{x} \wedge \partial_{\sigma^2} \mathbf{x}}{|\nabla x \phi|} \right| x(\sigma^0_\mu, L) \quad dL \quad d\theta \quad d\psi, 
\]

(7.32)

where we have accounted for the fact that the range of integration over \( \theta, \psi \) will be limited by the acquisition geometry. The ranges are given by \( E_\theta = E_\theta(\nu^0), E_\psi = E_\psi(\nu^0, \theta) \).
In Section 9, we will evaluate the Hessian $\det(\nabla_{\eta} \nabla_{\eta} \Phi')^0$ and show that $\text{sgn}(\nabla_{\eta} \nabla_{\eta} \Phi')^0 = 0$ at $\nu^0 = \pm \nu_\phi$. Hence, in case the partition is complete, we have

$$\sum_j \text{Re} \ I_{\Delta,j}(\nu^0 \cdot (y - x)) \frac{1}{|\nabla_{\eta} \phi_j|} = \frac{1}{|\nabla_{\eta} \phi_j|} \delta(\nu^0 \cdot (y - x)) \approx \delta(\phi(y) - \phi(x)). \quad (7.33)$$

Due to the point symmetry of the slowness surface at the image point $y$, we can replace the summation $\sum \nu^0 = \pm \nu_\phi \frac{1}{2}$ by the substitution $\nu^0 = \nu_\phi$.

8 GRT inversion of Kirchhoff data

Now, we will analyze the resolution operator from a different perspective. To accommodate for the non-linear reflection/transmission coefficients, we substitute our Kirchhoff-like approximation, a mixture of Eqs.(5.2) and (5.9), into the inversion formula Eq.(7.7). The result is an equation very similar to Eq.(7.32). We obtain

$$\langle (c^{(1)}) (\phi(y)) | \nabla_x \phi | y \rangle \approx \sum_{\nu^0 = \pm \nu_\phi} \frac{1}{2} \int_{E_R \times E_R} \int_R r_L^{(\tilde{\Sigma}; L)} \, dL \, d\theta \, d\psi, \quad (8.1)$$

where

$$r_L^{(\tilde{\Sigma}; L)} = \sum_j \text{Re} \ I_{\Delta,j}(\nu^0 \cdot (y - x)) \frac{1}{|\nabla_{\eta} \phi_j|} \bigg| x(\sigma^0_{\mu}, L) \bigg| \times$$

$$\frac{a_R(y, L, \eta^0, k, \theta, \psi) R_L^{(\tilde{\Sigma}; L)}(\eta^0, k, \theta, \psi) h(y, \nu^0)}{|\Gamma(y)|^0 \sqrt{\text{det}(\nabla_{\eta} \nabla_{\eta} \Phi')^0}} \times$$

$$|\partial_{\sigma_2} x \wedge \partial_{\sigma_2} x| \bigg| x(\sigma^0_{\mu}, L) \bigg| . \quad (8.2)$$

Componentwise, the reflection/transmission coefficients can be identified, viz., by undoing the multiplications by $a_R$.

9 The Hessian

In this section we will evaluate the Hessian of $\Phi'$ at a stationary point. First, note that $\partial_{\nu_\sigma} \partial_{\nu_\sigma} \Phi' = 0$ at the stationary point $x = y$ (then $\Phi \equiv 0$). Hence,

$$\det \left( \nabla_{\eta} \nabla_{\eta} \Phi' \right)^0 = \left[ \det \left( \nabla_{\nu_\sigma} \nabla_{\sigma} \Phi' \right)^0 \right]^2 . \quad (9.1)$$

On the other hand, note that

$$\left( \partial_{\nu_\sigma} \partial_{\sigma_2} \Phi' \right) = \left( \partial_{\nu_\sigma} \Gamma \right) \cdot \left( \partial_{\sigma_2} x \right) . \quad (9.2)$$
This matrix can be written in the form
\[
\begin{pmatrix}
- \partial_\nu \Gamma & - \\
- \partial_\nu \Gamma & - \\
\end{pmatrix}
\begin{pmatrix}
\partial_\sigma x \\
\partial_\sigma x \\
\end{pmatrix},
\]
hence, also,
\[
|\Gamma| \det [(\partial \nu \Gamma) \cdot (\partial \sigma x)] = \det \begin{pmatrix}
- \partial_\nu \Gamma & - \\
- \partial_\nu \Gamma & - \\
\end{pmatrix}
\begin{pmatrix}
\partial_\sigma x \\
\partial_\sigma x \\
\nu_\phi \\
\end{pmatrix}.
\] (9.3)

In this expression, the first matrix on the right-hand side can be identified as \(h\) (see Eq.(7.19)) and the second one with the Jacobian of the coordinate transformation (Eq.(4.9)). Hence, using Eq.(9.1), we arrive at the identity
\[
|\Gamma| \sqrt{\det (\nabla_\eta \nabla_\eta \Phi^0)} = h(\nu^0, \nu^0) |\partial_\sigma x \wedge \partial_\sigma x| x(\sigma^0, L).
\] (9.4)

In view of Eq.(9.1), the positive and negative eigenvalues must come in pairs, so that \(\text{sig} (\nabla_\eta \nabla_\eta \Phi^0)^0\) must be equal to 0 or 4. On the other hand, we have intrinsically assumed that the gradient of two-way travel does not vanish, so that \(h \neq 0\). In accordance with Eq.(9.4), hence, the determinant cannot vanish. This implies that under continuous deformations of the interface and ray geometries, the signature of the Hessian cannot change from 0 to 4 or vice versa (this would require an eigenvalue to become zero). In the case of flat level surfaces, it can be shown that the signature equals zero, which now implies that
\[
\text{sig} (\nabla_\eta \nabla_\eta \Phi^0)^0 = 0
\]
for any \(\phi\).

The modified GRT inversion

Substituting Eq.(9.4) into Eq.(8.2) yields
\[
\tau_L^{(N,M)} = \sum_J \Re I_{\Delta,J}(\nu^0 \cdot (y - x)) \frac{1}{|\nabla x|} x(\sigma^0, L) \times
\]
\[
a_R(y, L, \eta^0, k_\tau, \theta, \psi) R_L^{(N,M)}(\eta^0, k_\tau, \theta, \psi).
\] (9.5)

To arrive at this expression, at the stationary dip \(\nu^0 = \nu_\phi\), we could have set \((\nu \cdot \nu_\phi) = 1\) in Eq.(7.7) to begin with. Then, a priori knowledge about the geological dip is not required.
In fact, the dip is estimated from the image resulting from the partial reconstruction discussed in Section 7. The natural way to rewrite inversion formula (7.8) with Eq.(8.1) is

\[
\int_{\mathbb{R}^2} r_L^{(N, N')} dL \simeq \frac{1}{8\pi^2} \int_{S^2} \left( \Lambda y \cdot w^{(N, N')}(y) \hat{\alpha}^{(N)}(y), \hat{\alpha}^{(N')}(y) \right) \partial_\nu u^{(N, N')}(r, s, y) |\Gamma|^4 \frac{\partial (\hat{\alpha}, \hat{\alpha})}{\partial (\nu, \theta, \psi)} \bigg|_y d\nu.
\]

Here, we employ the mappings defined in Eq.(7.15). Note that this inversion formula comprises a system of equations; from each equation, in principle, the reflection/transmission coefficient at specular can be determined (we estimate the geological dip from the image and we have re-ordered the inversion by scattering angle and azimuth, which implies that we know \( a_R \) at stationary). The redundancy can be employed to verify or improve the estimate of the stationary dip. How the stationary dip appears in the spectrum of the medium's perturbation is discussed in Appendix A.

10 Discussion

We have shown, by carrying out a precise resolution analysis, that it is feasible to extract information about the angular dependent reflection/transmission coefficients from a GRT-based migration/inversion. We did not have to linearize nor expand the coefficients. In fact, the outcome of the resolution analysis is a multiple set of images for the reflection/transmission coefficients for the available range of specular scattering angles. Any type of AVA analysis can then be applied to interpret those images. In the derivation we have made use of the fact that the surface integral representations are linear in the scattering coefficients; these coefficients reduce, at specular, to the reflection/transmission coefficients.

The GRT approach employs a somewhat unusual ordering of data, viz., in common \((\theta, \psi)\) gathers. The inversion formula reduces to a two-dimensional integration over migration dip. The \((\theta, \psi)\) sorting, however, varies with the image point. It bears resemblance with the sorting in common offset, though. The use of such a sorting, however, necessitates the calculation of an additional Jacobian.

11 Acknowledgement

The authors would like to thank Herman Jaramillo for making Figure 2.1.
Appendix A. The spectrum of the medium perturbation

In principle, the geological dip can be directly estimated from an image. However, the geological dip is also hidden in the spectrum of the medium perturbation Eq.(4.10).

Setting (cf. Eq.(6.1))

\[ z \parallel \nu_\phi, \quad \{x_\mu\} \perp \nu_\phi \]

at \( y \in D \), the medium perturbation spectrum is of the form (cf. Eq.(4.10))

\[
\partial_\phi \bar{c}^{(1)}(k) = \int_{\mathbb{R}^2} dL \left( c^{(1)} \right)'(L) \int_D \delta(\phi(x) - L) \exp \left[-i \left( k_\mu x_\mu + k_z (y(L) + z) \right) \right] dx , \tag{A.1}
\]

where in fact \( k = \Theta' = k_\tau \nu \). Near the level surface \( \phi = L \), we employ the expansion

\[ \phi(x) = L + |\nabla x|_{\phi} z + \frac{1}{2} x_\mu \phi_{\mu \nu} x_\nu ; \tag{A.2} \]

we implicitly assume that the integral over \( L \) is windowed around \( y \) such that in the window or neighborhood \( \phi_{\mu \nu} \) may depend on \( L \) but \( \nu_\phi \) does not. At the level surface on which \( y \) lies, we have \( y = 0 \). Then,

\[
\partial_\phi \bar{c}^{(1)}(k) \simeq \int_{\mathbb{R}^2} dL \left( c^{(1)} \right)'(L) \exp[-ik_z y(L)] \times \tag{A.3}
\]

\[
\int_D \delta(|\nabla x|_{\phi} z + \frac{1}{2} x_\mu \phi_{\mu \nu} x_\nu) \exp\left[-i \left( k_\mu x_\mu + k_z (x_\mu \phi_{\mu \nu} x_\nu) \right) \right] dx x_1 dx_2 dz
\]

\[
= \int_{\mathbb{R}^2} dL \left( c^{(1)} \right)'(L) \exp[-ik_z y(L)] \int_{\mathbb{R}^2} \exp\left[-i \left( k_\mu x_\mu - k_z (x_\mu \phi_{\mu \nu} x_\nu) \right) \right] dx x_1 dx_2 \frac{1}{|\nabla x|_{\phi}}
\]

\[
= \int_{\mathbb{R}^2} 2\pi \frac{\exp[\pi i \csc(k_z \phi_{\mu \nu})/4]}{|k_z| \sqrt{\det(\phi_{\mu \nu})}} \exp\left[i k_\mu \frac{\nabla x_{\phi}}{k_z \phi_{\mu \nu}^{-1} k_\nu} \right] \times \frac{1}{k_z^2} \delta(\nu - (\nu \cdot \nu') (\nu - (\nu \cdot \nu')) \left( c^{(1)} \right)'(k_\tau (\nu - (\nu \cdot \nu'))),
\]

where \( \csc \), as in the main text, represents the sum of signs \( (\pm 1) \) of eigenvalues of \( k_z \phi_{\mu \nu} \). Note that \( k_\mu = 0 \) corresponds with the geological dip direction.

If the level surfaces of \( \phi \) were flat and \( \nu_\phi = \nu \) fixed, we would get

\[
\partial_\phi \bar{c}^{(1)}(k) = \delta(k - (k \cdot \nu') \nu') \left( c^{(1)} \right)'(k \cdot \nu') = \frac{1}{k_z^2} \delta(\nu - (\nu \cdot \nu') (\nu - (\nu \cdot \nu')) \left( c^{(1)} \right)'(k_\tau (\nu - (\nu \cdot \nu'))),
\]

which, in the Radon domain, implies

\[
\int_D \partial_\phi \bar{c}^{(1)}(\phi(x)) \delta''(y \cdot \nu - x \cdot \nu) \ dx
\]
\[
\frac{1}{\pi} \delta(\nu - (\nu \cdot \nu^1)\nu^1) \operatorname{Re} \int_{R^+} (c^{(1)})'(k_\tau(k_\tau - \nu \cdot \nu^1)) \exp(ik_\tau \varphi) \, dk_\tau \bigg|_{\varphi = y \cdot \nu} = y \cdot \nu
\]

\[
= -\frac{1}{\pi} \delta(\nu - (\nu \cdot \nu^1)\nu^1) \operatorname{Re} \int_{R^+} (c^{(1)})'(k_\tau) \exp(ik_\tau \varphi) \, dk_\tau \bigg|_{\varphi = y \cdot \nu^1} \quad \text{(A.4)}
\]

This formula shows that the GRT algorithm can reveal the geological dip explicitly. We assumed a proper coordinate system on \(S^2\), such that

\[
\int_{S^2} \delta(\nu - (\nu \cdot \nu^1)\nu^1) \, d\nu = 1.
\]
References


Characterization of the interwell space by Bayesian methods

Fernando S. de Moraes and John A. Scales
Center for Wave Phenomena
Colorado School of Mines

ABSTRACT
When facing the task of mapping rock attributes in a reservoir, geophysicists rely on information from the subsurface, such as well logs, and from surface geophysical methods. The most common approach to this problem is to interpret each data set separately and integrate the results manually. Bayesian methods provide a formal way of integration by using probability density functions to represent the information contained in each data set. However, difficulties associated with the construction and manipulation of high-dimensional probabilities have prevented wide application of these methods.

Using first principles of probability theory, one can derive the general Bayesian formulation for the geophysical inverse problem. This formulation can then be specialized to explore the much smaller length scale involved in the subsurface information when compared with the geophysical information. This difference makes it possible to use the subsurface information to independently construct probability densities for specific parameters that account for neighboring information such as well logs and other geological data. The principle of maximum entropy is used to construct these probability densities, thus giving a unified theoretical treatment that generalizes geostatistical methods.

When the specialized Bayesian formulation is used, n-dimensional distributions are replaced by approximations to their marginals, leading to a series of one-dimensional problems involving one parameter at a time. This gives a method that is able to systematically integrate diverse geological and geophysical information while maintaining practical feasibility.

Key words: Bayesian inference, Maximum entropy, Geostatistics.

CONTENTS
1 Introduction
2 Bayesian methodology
   2.1 Basic formulation
3 Prior Probabilities
   3.1 Maximum entropy
   3.2 Geostatistical approach
4 Approaches to the solution
   4.1 Nuisance parameters
   4.2 Fixed parameters
5 Example
6 Conclusions
7 Acknowledgments

1 INTRODUCTION
With the depletion of new exploratory frontiers, there is a growing demand for technology capable of investigating more subtle plays or of giving a more detailed description of reservoirs. Examples of these advances are new discoveries made in the Gulf of Mexico, using 3-D prestack migration methods to image under the salt, and the increasing use of geostatistics to interpolate well data (e.g., porosity), also using surface data (e.g., seismic data) to improve reservoir descriptions. The main goal of this work is to investigate how these demands can be incorporated into geophysical studies aimed at making inferences about the geology of a particular area through the solution of an inverse problem.
Traditionally, inversion methodologies have adopted the perspective of the geophysical data set. That is, we seek to learn about the resolution and the level of noise in the data in order to select a suitable parameterization of the subsurface (Backus and Gilbert, 1968; Parker, 1977). Equivalent results also can be accomplished by regularization, which employs damping and smoothing operators (Constable et al., 1987). This approach faces the fact that the geophysical data can resolve only certain averages of the true earth, thus we must lower demand for resolution in a given model to the appropriate level either by re-parameterization or by smoothing. The final results are usually very coarse models that may not serve the inference objectives. For instance, the inversion may generate interpretative models with questionable applicability in describing the geology, which are not useful given the current demands discussed above.

To meet these demands, inversion methodologies cannot rely on only one data set. Instead, additional information from well logs, petrophysics and geological interpretation are necessary to produce satisfactory results, once the parameterization is fixed by the demands for knowledge of a particular area. The order of the day is to explain aspects of the geology, not the geophysical data. To reach an appropriate description of the geology, we usually need parameterizations on a much finer scale than those determined by analysis of the resolution of the geophysical data. Then, according to this framework, the geophysical information becomes relatively less important when a large amount of additional information is available. In such cases, the geophysical data play a role well described as aiding the interpolation of geological attributes. That is, the final picture will be mainly determined by the subsurface information, but will also satisfy the averages that are well resolved by the geophysical data.

Thus, the main challenge is to develop inversion methodologies that can rigorously integrate any additional data sets that carry subsurface information such as well logs, geological sections and core measurements. This task is complicated by the diversity of these data sets. For example, geological information is usually qualitative or semi-quantitative, while all inverse calculations need quantitative data. Furthermore, different types of geophysical data are in different units, they respond to a change in different geological attributes and they have essentially different mathematical formulations.

In addition, to properly solve an inverse problem, it is necessary to address the uncertainties present in the calculations. These include the noise in the geophysical data, the uncertainties in the prior information and in the mathematical model (forward problem). The ability to include all uncertainties in an inverse calculation gives the means of judging proposed models. That is, it provides answer to questions such as: how reliable are the estimates? Or do we need more information (data) to resolve a given objective?

In principle, both issues of integration and uncertainty analysis can be handled by the Bayesian approach. But the burden becomes how to construct the probabilistic models and extract information from a combination of them, which is never a trivial task for high-dimensional problems. This paper attempts to address these difficulties of the Bayesian methodology by replacing multidimensional distributions with approximations to their marginals. Thus, instead of solving the multidimensional problem directly, we can replace it by a series of one-dimensional problems for one parameter at a time. This is possible when we take advantage of the much smaller length scales in the subsurface (local) information in comparison with the geophysical (global) information, thus allowing for inferences on one parameter independently of the others based on subsurface information alone.

In the next section, we derive the general Bayesian formulation of the geophysical inverse problem from the first principles of probability theory. Then, in Section 3, we discuss two methods for deriving distributions for the parameters that account for all the local information: the principle of maximum entropy and kriging. After that, in Section 4, we show how the general Bayesian formulation can be modified to incorporate the probability densities derived from the subsurface information. Finally, we solve a simple computational problem that illustrates the methodology.

### 2 Bayesian Methodology

The goal of scientific inference is to combine information from physical theories, data and other background knowledge to draw conclusions about a given physical system. Because the total information is always incomplete, that is, insufficient for precise conclusions, the mechanism of inference must involve plausible reasoning rather than deductive reasoning. Of course, it is always possible to modify the original problem so that the information available may provide an exact solution, but then, only a mathematical problem is being solved, not an inference one.

The application of the Bayes' theorem to inference problems addresses the same goal: draw conclusions about a physical system upon incomplete information. But it is not a stand-alone tool for inference, it is just a natural result of using probability theory to conduct plausible (probable) reasoning, which is what we need to
solve problems. In this context, probabilities are used to represent degrees of belief or plausibility about the truth of a proposition.

These ideas were rigorously substantiated by Cox (1946) and Cox (1961). In these works, he started from a desiderata of consistency to show that any method of inference that uses real numbers to represent degrees of plausibility must be based on the basic rules of probability theory (see below); otherwise, it will be inconsistent.

Despite this relatively recent result, the foundations of probability theory as logic, as we know it today, began over two centuries ago. To better understand the context of Bayesian inference, it is useful to make a brief excursion through the sequence of developments in the field. According to the historical accounts given by Jaynes (1978), the Bayes' theorem was derived to solve a problem that it did not solve completely. The problem is that before we can apply the basic rules of probability we need some initial probability assignments. There is nothing within the basic rules that tells how to assign those initial probabilities; thus an extension to these rules is needed. The first formal attempt to do that was the principle of insufficient reason presented by Bernoulli in 1713.

Bernoulli's principle states that if the available evidence gives no reason to consider one proposition neither more nor less favorable than another, the only honest possible probability assignment lends them equally likely. This principle can be extended to consider multiple sets of propositions in which case the probability assignment is the ratio between the number of favorable possibilities \( m \) and the total number of possibilities \( N \), that is, \( p = \frac{m}{N} \). But Bernoulli himself recognized the enormous limitations with his principle, namely, that it requires the ability to break a given problem into a set of exhaustive equally likely possibilities, which is an impossible task for most problems of inference. He then reasoned as Jaynes tells us: "if a probability \( p \) cannot be calculated in the manner \( p = \frac{m}{N} \) by direct application of the principle of insufficient reason, then in some cases we may still reason backwards and estimate the ratio \( \frac{m}{N} \) approximately by observing frequencies in many trials". Thus Bernoulli started seeking a formal relation between theoretical probabilities and observable frequencies, which yielded the weak law of large numbers for binomially-distributed variables. Later developments by Laplace and de Moivre (Gaussian approximation to the binomial distribution) ended up forming the basis for the central limit theorem. But all these results still assume independence, and they also relate to the direct problem. That is, they all describe the situation where, from known population numbers \( m \) and \( N \), we obtain the theoretical probability that particular sample numbers \( m' \) and \( N' \) can be observed. This indicates that the difficulties observed with the principle of insufficient reason are still present.

The real challenge was to find the solution for the inverse problem: from observable facts to obtain a theoretical probability of a given proposition. Pursuing this problem, Bayes found the beta distribution as a solution for the inversion of the binomial distribution (a particular case of the Bayes' theorem), and Laplace found the Bayes' theorem in the form we know it today. The Bayes theorem represents the complete solution for the inversion problem, which was aimed to avoid use of the principle of insufficient reason for direct assignment of probabilities. Ironically, however, the only useful results obtained by Laplace employed a particular case of the Bayes' theorem that relied on priors of the type \( \frac{1}{N} \) given by that same principle. According to Jaynes, this was not because Laplace failed to obtain the general form of the Bayes formula, which he wrote down, but rather because he did not have any principle for finding priors in cases where the prior information fails to render the possibilities equally likely. Important breakthroughs were achieved in the works of Shannon (1948) and Jaynes (1957) with the generalization of the principle of maximum entropy to assign prior probabilities.

This clearly illustrates the point that Bayes' theorem is only a consequence of using probability theory as logic, which within itself does not contain rules for assigning probabilities so the calculations can get started. This is why additional principle such as the principle of insufficient reason and maximum entropy are needed.

To introduce the fundamental equations of probability theory, consider the propositions \( A \) and \( B \) plus some background information \( C \). The sum rule tells that for an exhaustive, mutually exclusive set of probabilities, the sum of their individual probability is the unity. This can be expressed simply in terms of proposition \( A \) and its complement \( \overline{A} \) by

\[
P(A | C) + P(\overline{A} | C) = 1. \tag{1}
\]

The product rule gives the probability for \( AB \), which is commutative, is given by

\[
P(AB | C) = P(A | BC)P(B | C) = P(B | AC)P(A | C). \tag{2}
\]

From (1) and (2), we can find the sum rule for non-exclusive propositions \( A, B \) as

\[
P(A + B | C) = P(A | C) + P(B | C) - P(AB | C). \tag{3}
\]

Bayes' theorem comes directly from the commutativity of the product rule, which is essentially a logical statement. Thus, if the problem is to draw conclusions about the truth of proposition \( A \), we can equate the two terms in Equation (2) to get
or

\[ P(A | B, C) P(B | C) = P(B | AC) P(A | C) \] (4)

\[ P(A | B, C) = \frac{P(B | A, C) P(A | C)}{\sum_{j=1}^{M} P(B | A_j, C) P(A_j | C)}. \] (5)

If proposition \( A \) can be broken down into alternative propositions denoted by \( A = \{ A_1, A_2, \ldots, A_M \} \) and the problem is to draw conclusions about a particular \( A_i \), we get

\[ P(A_i | B, C) = \frac{P(B | A_i, C) P(A_i | C)}{\sum_{j=1}^{M} P(B | A_j, C) P(A_j | C)}. \] (6)

where the denominator is still just \( P(B | C) \). These different forms of the Bayes' theorem are only two out of great number of possibilities, which depend on the specific question being asked. They further illustrate that a single application of these elementary forms of Bayes' theorem gives only a probability, not a probability distribution.

This is an important issue when dealing with continuous probabilities, where the given quantities appearing in the conditional probabilities do not take on simple values. They take on a distribution of values, which makes necessary a marginalization process before a variable can be set to a particular number. To substantiate the understanding of such issues, it is useful to investigate the relationship between marginal and conditional probability distributions. Thus, consider the joint distribution of \( x \) and \( y \) as given by \( p(x, y) \). The marginal distribution for \( x \) can be written as

\[ \int p(x, y) \, dy = \int p(x | y) \, p(y) \, dy = p(x), \] (7)

while the conditional probability for \( x \) given \( y \) is

\[ p(x | y = y^*) = \frac{p(x, y^*)}{\int p(x, y) \, dx \, _{y=y^*}} = \frac{p(x, y^*)}{p(y^*)}, \] (8)

where the superscript * will always denote a particular value taken by a variable. However, if we let \( p(y) = \delta(y - y^*) \), then Equation (7) yields

\[ \int p(x | y) \delta(y - y^*) \, dy = p(x | y = y^*), \] (9)

which is the same as the conditional probability. A generalization of this for any \( p(y) \) is given by the mean-value theorem (see e.g., Gradshteyn and Ryzhik, 1980, p. 211), which is valid if \( p(x | y) \) is a bounded, continuous function and \( p(y) \) is positive and integrable in the interval \([y_1, y_2]\). In this case, we are guaranteed the existence of a value \( \xi \in [y_1, y_2] \), such that

\[ \int_{y_1}^{y_2} p(x | y) \, p(y) \, dy = p(x | y = \xi), \] (10)

using the fact that \( p(y) \) integrates to one. From this, it becomes clear that a marginal is always equivalent to a certain conditional distribution, corresponding to an unknown value of the variable being given or integrated out (Figure 1). Only for the particular case of the delta function can the value \( y^* \) be immediately substituted for \( y \).

2.1 Basic formulation

Using these ideas, it is possible to derive a Bayesian formulation for geophysical inverse problems. The first step is to define the information that will enter the calculation. This leads to a joint state of knowledge that can be represented by a joint probability distribution, which can be decomposed using the product rule given in Equation (2).

The geophysical inverse problem combines prior information (\( I \)), information from theory and from data measurements. Thus, consider the parameterized earth model represented by the vector \( \mathbf{m} \in \mathbb{R}^M \), and data predicted from theory and obtained by measurements as given by the vectors \( \mathbf{d}_t \), \( \mathbf{d}_o \in \mathbb{R}^N \), respectively.

It also important to establish all sources of uncertainty involved in \( \mathbf{d}_t \) and \( \mathbf{d}_o \). Those can be introduced as noise terms in the expressions for the data vectors, which we define as given by

\[ \mathbf{d}_t^* = \mathbf{d}_t + \mathbf{n}_m + \mathbf{n}_t, \] (11)

\[ \mathbf{d}_o^* = \mathbf{d}_o + \mathbf{n}_o, \] (12)

where \( \mathbf{n}_m \) are the errors associated with the model parameterization, \( \mathbf{n}_t \) the errors generated by approximations in the forward modeling theory, \( \mathbf{n}_o \) are the observational errors, \( \mathbf{d}_t^* \) are predicted data values and \( \mathbf{d}_o^* \) the observed data values. The predicted data values are the result of computations \( \mathbf{d}_t = \mathbf{f}(\mathbf{g}(\mathbf{m})) \), where \( \mathbf{g} \) is the forward modeling operator. The motivation to use the theory to predict data is the underlying assumption that perfect theoretical data \( (\mathbf{d}_t) \) match the observed data exactly in the absence of observational errors; that is, \( \mathbf{d}_o = \mathbf{d}_t \). From this we might as well write that

\[ \mathbf{g}(\mathbf{m}) = \mathbf{d}_t + \mathbf{n}_m + \mathbf{n}_t, \] (13)

\[ \mathbf{d}_o^* = \mathbf{d}_t + \mathbf{n}_o. \] (14)

With the above definitions, the representation of the joint state of knowledge is given by the joint probability \( p(\mathbf{m}, \mathbf{d}_t, \mathbf{d}_o | \mathcal{I}) \), which can be decomposed by the product rule to yield

\[ f(\mathbf{m} | \mathbf{d}_t, \mathbf{d}_o, \mathcal{I}) = \] (15)
\[
\begin{align*}
\frac{s(m | \mathcal{I}) g(d_o | m, \mathcal{I}) r(d_t | m, d_o, \mathcal{I})}{h(d_t, d_o | \mathcal{I})} .
\end{align*}
\] (16)

The interpretation of each term in Equation (16) is:

- \( r(d_t | m, d_o, \mathcal{I}) \) is the probability of the data predicted by the theory. It accounts for all modeling errors, including geological noise and approximations in the forward modeling operator. It also corresponds to the so-called \textit{likelihood} function, which name is used to emphasize the inference role of this distribution when the parameters \( m \) are considered as variables instead of given quantities. We discuss this distribution in more detail in the next section.

- \( s(m | \mathcal{I}) \) is the prior distribution for the model parameters. It accounts for any information at hand before the acquisition of the geophysical data.

- \( h(d_t, d_o | \mathcal{I}) \) is a term originating from the left side of Equation (15). Thus, for consistency reasons, it cannot be assigned. It must be determined by the terms in the right-hand side of the equality, which is why it is only a normalization factor.

- \( g(d_o | m, \mathcal{I}) \) is the data regarded as a probability distribution. It represents the ranking of data vectors according to their plausibility, given our knowledge of the model parameters and prior information. It does not involve the theoretical data, and therefore avoids using the forward modeling operator. If viewed as a likelihood function, it can represent only the qualitative interpretation that can be done using a display of a geophysical data such as seismic sections or gravity maps.

By using this Bayesian formulation and based on the arguments given in last section, it becomes clear that observed data cannot be just substituted into a Bayes' formula, since they are not exactly known numbers, but rather a distribution of values given by \( g \). Observed data must be introduced by a marginalization process, which can be described using Equation (13) by

\[
\int_R f(m | d_t, d_o, \mathcal{I}) h(d_t, d_o | \mathcal{I}) \, dd_o =
\]

\[
\int_R g(d_o | m, \mathcal{I}) r(d_t | m, d_o, \mathcal{I}) \, dd_o .
\] (17)

Equation (17) is similar to that given by Tarantola (1987), but it has the advantage of having been derived from the product rule of probability theory.

To examine Equation (17) further, it is useful to consider two important particular cases. The first case is when the probability \( g(d_o | m, \mathcal{I}) \) is strongly concentrated; i.e., can be approximated by \( \delta(d_o - d^*_o) \). Then, the integrals in Equation (17) can be written as

\[
\int_R f(m | d_t, d_o, \mathcal{I}) h(d_t, d_o | \mathcal{I}) \, dd_o =
\]

\[
s(m | \mathcal{I}) \int_R r(d_t | m, d_o, \mathcal{I}) \delta(d_o - d^*_o) \, dd_o ,
\] (18)

which can be evaluated to yield

\[
f(m | d_t, \mathcal{I}) = \frac{s(m | \mathcal{I}) r(d^*_o | m, \mathcal{I})}{h(d_t | \mathcal{I})} ,
\] (19)

using the assumption that \( d_t = d_o \).

The other important case is when the predicted data distribution is strongly concentrated: that is, modeling errors are negligible. In this case the probability for \( d_o \) is given by \( \delta(d_o - d^*_o) \) or equivalently by \( \delta(d_o - d^*_t) \). Then Equation (17) becomes

\[
\int_R f(m | d_t, d_o, \mathcal{I}) h(d_t, d_o | \mathcal{I}) \, dd_o =
\]

\[
s(m | \mathcal{I}) \int_R \delta(d_o - d^*_t) g(d_o | m, \mathcal{I}) \, dd_o ,
\] (20)

which can be evaluated as

\[
f(m | d_t, \mathcal{I}) = \frac{s(m | \mathcal{I}) g(d^*_t | m, \mathcal{I})}{h(d_t | \mathcal{I})} .
\] (21)
The similarities between Equations (19) and (21) are striking. We discuss these two extreme cases in more detail in the next two sections.

This formulation also reveals that the usual formulation of a Bayesian inverse problem given in the literature

$$ f(m | d_0, I) = \frac{s(m | I) r(d_0 | m, I)}{h(d_0 | I)} \quad (22) $$

has some logical difficulties in the way that the observed data values $d_0^s$ are substituted for $d_0$. To see that, recall that we can substitute numbers for variables in conditional probabilities, and fully account for the uncertainties (i.e., achieve the correspondence to a marginalization procedure), only when the associated distribution is strongly concentrated. Otherwise, conditional probabilities give only partial account for the uncertainties, since they correspond to looking at the intersection of the joint distribution with a particular hyperplane plane $d_0 = d_0^s$ (see Figure 1). Thus, if the data are sufficiently precise to justify immediate substitution, what is the probability assignment for $r$? If the data are not precise and we do make the substitution, how can we guarantee that the particular conditional probability really gives a realistic picture of the uncertainties for $m$? Questions such as these do not appear when the theoretical data $d_i$ are introduced.

### 2.1.1 The likelihood function

To better understand the meaning of the likelihood function, consider the expression for the theoretical data given by Equation (11). It is a sum of the forward modeling equation and some error terms. The forward model $g(m)$ can be taken as precise numbers, since computer errors can be made arbitrarily small, at least in principle. Therefore, we have that

$$ r(d_i | m, d_0, I) = r(n_i | m, d_0, I) $$

$$ = r(d_i - g(m) | m, d_0, I), $$

where $n_i = -(m_n + n_g)$ (see Equation (11)). The above expression may seem strange at first glance, but all it is saying is that the probability for $d_i$ remains invariant upon translations, which are indistinguishable by the prior information. However, $d_i$ can only be assessed through the observed data through the expression given in Equation (14). Then, by substitution, we get

$$ r(d_i - g(m) | m, d_0, I) $$

$$ = r(d_i^s - n_i - g(m) | m, d_0, I), $$

$$ = r(d_i^s - g(m) | m, d_0, I), \quad (23) $$

which gives the same as when the assumption $d_i = d_i^s$ is directly applied to the first equation.

To investigate further, consider again the case where observational errors are negligible, as in Equation (18). The integration over $d_0$ for the likelihood determined by Equation (24) gives

$$ f(m | d_0, I) = \frac{s(m | I) r(d_0^s - g(m) | m, I)}{h(d_0 | I)}. \quad (25) $$

This likelihood function has the same form as usually given in the literature, although it involves different arguments.

If a normal distribution with zero mean and covariance matrix $C_i$ is assigned to the data misfit, we can write

$$ r(d_i | m, I) = (2\pi)^{-\frac{N}{2}} |C_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (d_i^s - g(m))^T C_i^{-1} (d_i^s - g(m)) \right\}. \quad (26) $$

For the linear problem, the computed data are given by

$$ d_i = Gm + n_i, \quad (27) $$

where $G$ is a $N \times M$ matrix. Then, Equation (26) becomes

$$ r(d_i | m, I) = (2\pi)^{-\frac{N}{2}} |C_i|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (d_i^s - Gm)^T C_i^{-1} (d_i^s - Gm) \right\}. \quad (28) $$

For the linear case, we can rewrite Equation (28) to emphasize its inference role (i.e., $m$ unknown), using that

$$ (d_i^s - Gm)^T C_i^{-1} (d_i^s - Gm) $$

$$ = (d_i^s - Grh)^T C_i^{-1} (d_i^s - Grh) $$

$$ + (m - rh)^T G^T C_i^{-1} G (m - rh), \quad (29) $$

where $rh = (G^T C_i^{-1} G)^{-1} G^T C_i^{-1} d_i^s$ is an estimated parameter vector. This assumes that $G^T C_i^{-1} G$ is invertible. If we substitute this relation into Equation (28), we get

$$ r(d_i | m, I) = (2\pi)^{-\frac{N}{2}} |C_i|^{-\frac{1}{2}} \exp \left\{ S(rh) \right\} $$

$$ \exp \left\{ -\frac{1}{2} (m - rh)^T G^T C_i^{-1} G (m - rh) \right\}, \quad (30) $$

where $S(rh) = (d_i^s - Grh)^T C_i^{-1} (d_i^s - Grh)$ is the estimated misfit value. This is an important form of likelihood function due to the fact that the location parameters $rh$ are completely determined by the data. Because of that it is called in Bayesian literature the data translated likelihood.

### 2.1.2 The observed data distribution

The arguments for the construction of the data distribution ($g$ in Equation 17) are similar to those used for the likelihood function. The exception here is that we are not allowed to introduce the forward modeling equation.
Since $d_i$ does not enter the definition of $g$, however, we show below that for the extreme case where the modeling errors are negligible, the forward modeling equation is introduced into $g$ by marginalization.

Using the definition given in Equation (12), we can write
\[
g(d_0 | m, I) = g(n_0 | m, I) = g(d_0^* - d_0 | m, I),
\]
which can be justified in the same way as before: the uncertainty about $d_0$ does not change upon translations, which are not distinguished by the prior information.

To investigate further, we return to the case where modeling errors are negligible, given by Equation (20). For this case, the likelihood function is a delta function $\delta(d_i - d^*_i)$, or $\delta(d_0 - d^*_0)$ when using the assumption that $d_1 = d^*_1$. The computed data values $d^*_i$ are given by the forward modeling equation $g(m)$. When this and Equation (31) are introduced into Equation (17), the integration over $d_0$ gives
\[
f(m | d_0, I) = \frac{s(m | I) g(d_0^* - g(m) | m, I)}{h(d_0^* | I)}. \tag{32}
\]
Next, consider again the example of a normal distribution with zero mean and covariance matrix $C_0$ being assigned for $g$. Then, we can write
\[
g(d_0 | m, I) = (2\pi)^{-\frac{N}{2}} | C_0 |^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [d_0^* - g(m)]^T C_0^{-1} [d_0^* - g(m)] \right\}. \tag{33}
\]
All other equations of the previous section can be derived from this, in the same way. The only notable difference is that, for this case, we are dealing with the covariance matrix for the observational errors ($C_0$) as opposed to the covariance matrix of the modeling errors ($C_i$) of the previous section.

### 3 Prior Probabilities

As discussed in the previous sections, inverse problems face the task of producing a detailed picture of the subsurface based on the geophysical data and other prior information. While Bayesian methods seem to be the appropriate framework to combine all the available information, practical difficulties associated with high-dimensionality of the probability distributions limit its application. Thus, further developments are needed to overcome the limitations.

One possible direction to approach the problem is by looking into the characteristics of the geophysical and the subsurface information. When we do this, it is possible to recognize that they are defined over fundamentally different length scales. That is, geophysical information has a global nature, which means that all parameters must be considered simultaneously in order to solve an inverse problem. In contrast, it is possible to explore the subsurface information to draw conclusions about a particular parameter of the model in an isolated way. In fact, this local aspect of the subsurface information has been routinely exploited in other fields such as geostatistics and it is fundamentally an interpolation problem. In this section, two different methods for deriving local probability distributions, the maximum entropy and the indicator kriging, are discussed and their relationships investigated. Then, in the next section, we show how to incorporate local distributions into the full geophysical inverse problem.

A schematic representation of local probabilities computed from a well log is given in Figure 2. At a well location, if the measurement errors are negligible when compared with the inversion errors, the local probability is approximately a delta function. As we go away from the borehole, the uncertainty increases, as reflected in broader and broader local distributions. The intensity of the attenuation of the probabilities, as we go away from the borehole, is given by the degree of spatial variability of the medium and by the interpolation algorithm.

The ability to incorporate local distributions into the geophysical inverse corresponds also to the ability to incorporate an unlimited amount of prior information. This becomes clear after going over the methodologies for computing these distributions. Also, when necessary, the same methodologies used to derive the local distributions can be used to construct the full multidimensional distribution. To see this, consider the case where a set of direct measurements of the desired parameter has been made. If we include these measurements in the prior information $I$ and denote the model parameters by $m_j$, $j = 1, \ldots, M$, we find that the local probabilities are actually an approximation for the conditional probability given by $f(m_j | I)$. Because these distributions are estimated one at a time, the current distribution being estimated can also be conditioned to the mean value of others previously estimated distributions. Then the multivariate distribution can be obtained by taking the product of all the individual distributions, which is just an application of the product rule:
\[
f(m | I) = f(m_1 | I) f(m_2 | m_1, I) \cdots f(m_{M-1} | m_1, m_2, \ldots, m_{M-2}, I) f(m_M | m_1, m_2, \ldots, m_{M-1}, I). \tag{34}
\]
But the main goal is to work directly with the local distributions. Thus, we begin the discussion on how to come up with the local probabilities by examining the maximum entropy method. Then, we look into geostatistical...
methods that have close association to maximum entropy methods

3.1 Maximum entropy

The information entropy was first defined by Shannon (1948) to measure the amount of information (or conversely the uncertainty) in a given distribution. Later, Jaynes (1957) found that the information entropy can be used as formal rule for assigning prior probabilities, as we discuss below. Because entropy is a measure of the uncertainty of a probability distribution, the maximum entropy principle provides the most conservative distribution that agrees with all the given constraints. According to Shannon (1948), the entropy of a discrete probability density function is given by

\[ H(p) = - \sum_{i=1}^{n} p_i \log p_i. \]

Many scientific problems involve continuous variables. Thus, it is important to extend the definition of entropy to the continuous case. This extension, which can be found in Jaynes (1963), leads to the relative entropy. However, careful examination of the issues involved suggests that, for inference purposes, relative entropy should be adopted even in the discrete case when the correspondence to continuous variables is important (Gouveia et al., 1996). As the name suggests, relative entropy is a measure of uncertainty in probability \( p \) relative to another probability \( q \), which is known a priori. Thus, if we consider these probabilities associated with the variable \( x \), the relative entropy is given by

\[ H(p; q) = - \sum_{i=1}^{n} p_i \log \frac{p(x_i)}{q(x_i)}, \]

in the case \( x \) is a discrete variable, or

\[ H(p; q) = - \int_{\mathbb{R}} p(x) \log \frac{p(x)}{q(x)} \, dx, \]

when \( x \) is a continuous variable.

An equivalent measure is given by the cross-entropy, which is the negative of the relative entropy as defined above. The cross-entropy was first defined by Kullback (1959) under the name of directed divergence. Shore and Johnson (1981) provide an extensive collection of properties of the cross-entropy with proofs.

Maximization of the continuous entropy functional (36) is the variational problem over \( p \), given by max \( H(p; q) \), subject to the normalization

\[ \int_{\mathbb{R}} p(x) \, dx = 1, \]

and to other constraints given in the form of expectations \( \langle f_k(x) \rangle \)

\[ \int_{\mathbb{R}} f_k(x) p(x) \, dx = \mu_k, \quad k = 1, \ldots, K, \]

where \( \mu_k \) is a numerical value that can be computed from the available data. Usually \( f_k(x) = x^n \), where \( n = 1, 2, \ldots \), which makes \( \langle f_k(x) \rangle \) the moments of the distribution. The solution for this problem (Jaynes, 1957, Gouveia et al., 1996) is given by

\[ p(x) = q(x) \exp \left[ -\lambda_0 - \sum_{k=1}^{K} \lambda_k f_k(x) \right], \]

or
\[ p(z) = Z^{-1} q(z) \exp \left[ -\sum_{k=1}^{K} \lambda_k f_k(z) \right], \quad (40) \]

with
\[ Z \equiv \exp(\lambda_0) = \int_{\mathbb{R}} q(z) \exp \left[ -\sum_{k=1}^{K} \lambda_k f_k(z) \right] dz. \quad (41) \]

The complete solution requires the determination of the Lagrange multipliers \( \lambda_k \) and the reference prior \( q(z) \). A well-known result of the maximum entropy problem is that when only the first two moments of the distribution are given as constraints, the maximum entropy distribution relative to a uniform distribution approaches the normal distribution, as the limits of the uniform distribution go to infinity (Gouveia et al., 1996).

### 3.1.1 Conditional probabilities

To show a simple example of how the principle of maximum entropy can be used to derive local probability densities, consider a set of measurements denoted by the vector \( w = (w_1, \ldots, w_N) \in \mathbb{R}^N \), which provide some information about a single parameter \( m \). The measurements need not to correspond to the same attribute. Thus, we can define the vector
\[ x = \begin{bmatrix} m \\ w \end{bmatrix}, \]

and its corresponding covariance matrix
\[ C = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix}, \quad (42) \]

where
\[ C_{11} = C_{m,m} ; \quad C_{22} = C_{w,w} ; \quad C_{12} = C_{m,w} = C_{w,m}, \]

which are \( 1 \times 1 \), \( N \times N \) and \( N \times 1 \) matrices, respectively. Also, the corresponding inverse covariance matrix is defined as \( V = C^{-1} \), which can be partitioned in the same form as Equality (42) and each part is given by

\[ V_{11} = (C_{11} - C_{12} C_{22}^{-1} C_{21})^{-1}, \]
\[ V_{12} = -C_{12}^T C_{11} C_{21} (C_{22} - C_{21} C_{11}^{-1} C_{12})^{-1}, \]
\[ V_{21} = -C_{22}^T C_{21} (C_{11} - C_{12} C_{22}^{-1} C_{21})^{-1}, \]
\[ V_{22} = (C_{22} - C_{21} C_{11}^{-1} C_{12})^{-1}, \quad (43) \]

assuming that all matrices are invertible. As mentioned in the last section, the conditional probability \( f(m \mid w) \), when only the first two moments are being considered, can be written as the ratio of two Gaussians
\[ f(m \mid w) = \frac{G(x)}{G(w)}, \quad (44) \]

where
\[ G(x) = (2\pi)^{-\frac{N}{2}} |V|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (x - \bar{x})^T V (x - \bar{x}) \right], \quad (45) \]

\[ \bar{x} = \begin{bmatrix} m_0 \\ \bar{w} \end{bmatrix} \]

are the mean values for \( x \), and
\[ G(w) = (2\pi)^{-\frac{N}{2}} |C_{22}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (w - \bar{w})^T C_{22}^{-1} (w - \bar{w}) \right]. \quad (46) \]

Equation (44) can be expanded as
\[ f(m \mid w) = \alpha \exp \left\{ -\frac{1}{2} \left[ V_{11} (m - m_0)^2 + 2 (m - m_0) V_{12} (w - \bar{w}) + (w - \bar{w})^T V_{22} (w - \bar{w}) \right] \right\}. \]

The mean of a normal distribution also coincides with the mode. Thus we may find the conditional expectation \( m \mid w_1, \ldots, w_N \) by maximizing the above conditional distribution, which is equivalent to minimizing the argument of the exponential function. This process represented by
\[ V_{11} (m - m_0) + V_{12} (w - \bar{w}) = 0. \]

Note that \( V_{11} = (C_{11} - C_{12} C_{22}^{-1} C_{21})^{-1} \) is just a scalar, and
\[ V_{12} = -C_{12}^T C_{11} C_{21} (C_{22} - C_{21} C_{11}^{-1} C_{12})^{-1}, \]

a vector in \( \mathbb{R}^N \). From this, we can find an estimator for the conditional expectation as given by
\[ \hat{m} = m_0 - V_{12}^T V_{12} (w - \bar{w}). \quad (47) \]

To find the conditional variance \( \sigma^2_{\hat{m}} \mid w_1, \ldots, w_N \), we can just rewrite the argument of the exponential function in terms of \( \hat{m} \), beginning with the expansion
\[ f(m \mid w) = \alpha \exp \left\{ -\frac{1}{2} \left[ V_{11} (m^2 - 2m_0 m + m_0^2) + 2 V_{12} (w - \bar{w}) m - 2 m_0 V_{12} (w - \bar{w}) \right. \right\} \]
\[ + (w - \bar{w})^T \left( V_{22} - C_{22}^{\hat{m}} \right) (w - \bar{w}) \right\}, \quad (48) \]

or
\[ f(m \mid w) = \alpha \exp \left\{ -\frac{1}{2} \left[ V_{11} (m^2 - m_0^2) - 2 V_{12} (m - m_0) V_{12} (w - \bar{w}) + 2 m_0 V_{12} (w - \bar{w}) \right. \right. \]
\[ + (w - \bar{w})^T \left( V_{22} - C_{22}^{\hat{m}} \right) (w - \bar{w}) \right\}. \quad (49) \]
The second term of the argument can be recognized as the estimator given by Equation (47). By making the substitution and completing the squares we get

\[
f(m \mid w) = \alpha \exp \left\{ -\frac{1}{2} \left[ V_{\overline{1}}(m - \bar{m}) \right]^2 - (w - \bar{w})^T V_{\overline{12}}^{-1} V_{\overline{12}}(w - \bar{w}) \right\}
\]

or

\[
f(m \mid w) = \alpha \exp \left\{ -\frac{1}{2} \left[ V_{\overline{1}}(m - \bar{m}) \right]^2 - (w - \bar{w})^T (V_{\overline{22}} - C_{\overline{22}})^{-1} \left( V_{\overline{12}}(w - \bar{w}) \right)
\right\}.
\]

Since the second term is constant, it can be incorporated into the multiplicative constant \( \alpha \) to yield

\[
f(m \mid w) = \alpha \exp \left\{ -\frac{1}{2} \left[ V_{\overline{1}}(m - \bar{m}) \right]^2 \right\},
\]

where

\[
\left( \sigma_{m, i}^2 \mid w_1, \ldots, w_N \right) = V_{\overline{11}}^{-1},
\]

The estimator for the conditional mean, given by Equation (47), can be recognized as the simple kriging equation (see, e.g., Equation (8) in Lesson II of Journel, 1989). Thus, we see that when we have the ability to specify the mean and covariance between variables we can easily obtain the local conditional moments, which lead to the specification of the local normal marginal distribution.

However, in many situations it is useful to specify other higher-order moments, such as

\[
\left\{ (m \mid w_1, \ldots, w_N) \right\},
\]

\[
\left\{ (m^2 \mid w_1, \ldots, w_N) \right\},
\]

\[
\left\{ (m^3 \mid w_1, \ldots, w_N) \right\} \text{ and}
\]

\[
\left\{ (m^4 \mid w_1, \ldots, w_N) \right\}.
\]

If we did that, we could make use of the maximum entropy principle to find the corresponding probability density function (pdf). In this way, information such as skewness and kurtosis would be included in the calculations. McCullagh (1987) discusses methods for calculation of the conditional higher-order moments.

A simple alternative approach is provided by the indicator kriging, which we discuss next.

### 3.2 Geostatistical approach

Geostatistical methods for estimating local probability distributions are based on the so-called indicator variables, which is defined below. The strengths of the method are twofold. First, is the ability to account for diverse information, which is one of the main difficulties in inversion. Second is that this procedure is specially designed to give approximations to the local distributions we seek. To understand how this is done, let's see how indicators are used to represent probabilities.

#### 3.2.1 Indicator representation of probabilities

We begin this section by recalling that for one particular value \( y^* \) taken by the variable \( y \), we have in general that the probability of the event \( y \leq y^* \) is given by

\[
P(y \leq y^*) = F(y) \text{ and } f(y) = \frac{dF(y)}{dy},
\]

where capital letters \( (F) \) are used to represent cumulative distribution functions (cdf) and lower case letters \( (f) \) are used for probability density functions.

In our problem, we can look at a single model parameter \( m \) (e.g., rock density) and its value \( m^* \) at a particular point \( r \). The task is to estimate the probability that \( m \leq m^* \) given some information in the neighborhood. Thus, if we again consider a data vector \( w = (w_1, \ldots, w_N) \), corresponding to subsurface measurements of the same attribute as \( m \) (e.g., density log), we can apply the indicator transformation, which can be defined by

\[
i_j(\mu) = \begin{cases} 
0, & m = w_j > \mu, \\
1, & m = w_j \leq \mu,
\end{cases}
\]

where \( \mu \) is a selected threshold value for \( m \). Since \( i \) is a binary variable, its probabilities are given by the Bernoulli distribution

\[
f(i) = P(m \leq \mu) | [1 - P(m \leq \mu)].
\]

Therefore, the expected value of \( i \) is

\[
\langle i(\mu) \rangle = 1 \cdot P(m \leq \mu) + 0 \cdot P(m > \mu),
\]

\[
= P(m \leq \mu) = F(m),
\]

where \( F(m) \) is the cdf for the model parameter. This means that we can approximate the probability \( F(m) \) by averages of indicators. One simple average is just

\[
P(m \leq \mu) = \frac{1}{N} \sum_{j=1}^{N} i_j(\mu),
\]

which gives the same importance to any value \( i \) regardless of its location or the threshold value \( \mu \). If we want to infer the uncertainty about \( m \) given indicators in the neighborhood, the spatial variability should be accounted for (i.e., the closer the samples, the more likely they are to have similar values). So a weighted version of (56) can be defined as
Figure 3. Schematic figure showing: (a) the indicator transformation of rock density measurements (full dot) at three threshold values, (b) indicator interpolation to a location represented by the empty dot for each threshold map, (c) the threshold values corresponding to the transformation in (a) and (d) a sketch of a plot of the cumulative distribution for density obtained by the indicator regression in (b).

\[ P(m \leq \mu) = \sum_{j=1}^{N} a_j(\mu) i_j(\mu). \]  

This version, which corresponds to an interpolation problem for \( i(\mu) \), provides an estimate for \( P(m \leq \mu) \) at \( r \) for a given threshold value \( \mu \). If this procedure is repeated for a set of appropriately chosen threshold values \( \mu_k, k = 1, 2, ..., K \), we can build up the discrete approximation for the cdf for the model parameter, namely \( F(\mu) \approx F(m) \). For that, the requirements are: the application of the indicator transform to all the \( N \) measurements \( w \), for each of the \( K \) threshold values, and the determination of the \( N \times K \) weights \( a_j(\mu) \) (Figure 3). It is important to notice that in order for equation (57) to provide the building block for a legitimate cdf, the following order relations should be observed:

\[ F(\mu_k) \in [0, 1] \quad \text{for all } k, \]
\[ F(\mu_k) \leq F(\mu_l) \quad \text{for all } \mu_k \leq \mu_l. \]  

The approximation of \( F(m) \) through the indicator formalism can be better understood pictorially, by looking at the indicator transform \( i(\mu) \) for different thresholds, at a particular location, as a stack values. This corresponds to examining Figure 3 (a) along a vertical line at a given rock density observation (e.g., 2.04 g/cm\(^3\)). In fact, at a location where we have made an observation of the model parameter (e.g., core density
measurement), the corresponding cdf is a step function at the observed value \( w_j \) (Figure 4). This corresponds to a probability density distribution given by a delta function at \( w_j \), implying an error-free measurement. The indicator variable gives the exact position of the step if there is a threshold value \( \mu_k = w_j \). At another location where a sample value is not available, the indicator variable is determined by an interpolation procedure (Figure 3 (b)) such as Equation (57). Thus, we have that

\[
P(m \leq \mu) = \langle i(\mu) \rangle \approx i^*(\mu),
\]

where \( i^* \) is the estimated value for \( i \).

3.2.2 Indicator coding for different types of information

Indicator variables give a great flexibility in representing information. To see that, consider three levels of knowledge in any earth science study: the knowledge prior to any experiment on the area under study, that from secondary information and that from direct observation of the parameter under study.

The knowledge prior to any experiment comes from information of physics or previous experiments in different areas under similar settings. The knowledge from secondary information is gained by performing experiments (geophysical or geological) on parameters that are somehow related to the ones which we are specifically interested in. For example, in rock density estimation, indirect information on density can be given by the knowledge of seismic velocity. Also, qualitative geological information in the form of lithologic sections derived from drillholes and well logs fall in the same category. Finally, the direct observations are made by in situ measurements, which can be either density logs or core measurements. In summary, if we consider the rock-density estimation problem, useful information can be provided by

- density measurements from well logs or core analysis,
- velocity values derived from seismic data interpretation,
- physical interval constraints,
- geological information on rock types and expertise.

The indicator coding of all these distinct types of information is just a generalization of the indicator transform (53), given by

\[
i(\tau, \mu) = \begin{cases} 
0, & m(\tau) > m_l, \\
\text{undefined}, & m(\tau) \in [m_l, m_u], \\
1, & m(\tau) \leq m_u,
\end{cases}
\]

where \( m_l \) and \( m_u \) are the lower and upper limits of an interval defining the precision of the corresponding inform-
Figure 5. Cumulative distribution and the corresponding probability density function for: (a) interval constraints (type 2) and (b) prior distribution (type 3).

<table>
<thead>
<tr>
<th>$\mu_0$</th>
<th>1</th>
<th>1</th>
<th>1</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_1$</td>
<td>1</td>
<td>1</td>
<td>0.9</td>
<td>0.8</td>
</tr>
<tr>
<td>$\mu_2$</td>
<td>1</td>
<td>?</td>
<td>0.6</td>
<td>0.8</td>
</tr>
<tr>
<td>$\mu_3$</td>
<td>0</td>
<td>?</td>
<td>0.4</td>
<td>0.7</td>
</tr>
<tr>
<td>$\mu_4$</td>
<td>0</td>
<td>0</td>
<td>0.2</td>
<td>0.4</td>
</tr>
<tr>
<td>$\mu_5$</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>$\mu_6$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>threshold</th>
<th>type 1</th>
<th>type 2</th>
<th>type 3</th>
<th>posterior cdf</th>
</tr>
</thead>
</table>

Table 1. Schematic representation for the prior cdfs of the three types described in the text and the corresponding updated posterior cdf.

ation. If the associated errors can be neglected, $m_i = m_u$ and the interval has zero amplitude. For this case Equation (60) reduces to (58), corresponding to a step cdf (Figure 4), which will be referred as a cdf of “type 1.” When $m_i < m_u$, the “undefined interval” can be left empty (Figure 5a) to represent an interval constraint limiting the possible occurrence of $m$, or it can be filled with values between 0 and 1 denoting a prior cdf for $m$ (Figure 5b). The former cdf will be referred as a cdf of “type 2,” and the latter will be referred as a cdf of “type 3.”

Making reference to the cdfs just defined, it is possible to associate the information, given above, to the proper type. Direct rock density measurements can be coded as a cdf of type 1, when the associated errors are negligible in comparison to those of an inversion procedure. Quantitative secondary information can either be coded as cdfs of type 1 or type 3, while qualitative information, such as geological information, should be represented as type 2 or 3. For quantitative secondary data, such as seismic velocity values, the cdf type depends on how precisely the secondary information can determine the value for $m$. It usually yields cdfs of type 3. Finally, expertise can be coded either as type 2 or type 3.

To summarize the three types for coding the subsurface information, it is possible to represent each cdf type in the form of indicator columns (i.e., cdfs at various locations), each line corresponding to a threshold value $\mu_k$. An indicator interpolation scheme is the updating procedure leading to the approximation of the cdf for $m$ at a given location $r$. The interpolation step can be interpreted as a maximum entropy calculation in the same
way as presented in Section 3.1.1, since it usually relies on first and second order-statistics of indicators varying continuously on the interval $[0,1]$. Table 1 is a schematic representation of this process, considering the case of nine threshold values. The right-hand column is a representation of the updated cdf, which synthesizes all the information available.

Deutsch and Journel (1992) provide an extensive suite of Fortran routines that implement indicator kriging and cokriging, which can be used to estimate the local conditional cdfs.

4 APPROACHES TO THE SOLUTION

As we have seen, the Bayesian approach to inverse problems involves probability distributions in high dimensions both in the data and model spaces. The formal solution is the product of these multivariate probability distributions, and the desired information about a particular parameter can be obtained by marginalization. This requires a sophisticated set of tools, such as Monte Carlo integration and importance sampling, to carry out multidimensional integrations (e.g., Scales and Tarantola, 1994; Mosegaard and Tarantola, 1995).

We seek alternative approaches to the estimation of the parameter vector $m$, without solving the full multivariate problem just discussed. In addition, the solution must incorporate all the prior information ($I$), processed into local conditional distributions using the methods discussed in the previous section. To proceed further we have to face a problem arising from the fact that the geostatistical regression is of local nature. Univariate conditional probabilities are estimated for individual parameters, independently of the other, considering only the information in the neighborhood. In contrast, the geophysical inverse problem cannot be done for a single parameter, independently, due to forward modeling relation given by Equation (11). This contrast is basically the origin of the alternative formulations discussed below, which hinge on the idea of reducing the dimensionality of the geophysical problem, using marginalization theory.

The idea is to seek the solution one parameter at a time in a way similar to the geostatistical problem. To carry out this approach, the parameter vector $m$ must be divided into two parts. One part is just the specific parameter to be estimated in the current iteration, denoted by $m_1$, and the other part is composed of the remaining components of $m$, denoted by $m_2$. That is,

$$m = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix},$$

where $m_1 \in R$ and $m_2 \in R^{M-1}$.

Two distinct treatments for $m_2$ can be considered. To treat $m_2$ as nuisance parameters*, or as fixed quantities known a priori. Now the statement for our problem is: we want a measure of the uncertainty of a particular subset of parameters of the geophysical model $m$; given the synthetic data $d_s$, the observations $d_o$ and the prior information $I$. Then, the general Bayesian formulation given by Equation (17) can be expanded further to give, for the first case

$$\int_R f(m_1, m_2 | d_s, d_o, I) h(d_i, d_o | I) dd_o = t(m_1 | I) v(m_2 | m_1, I)$$

and for the second case

$$\int_R f(m_1, m_2 | d_s, d_o, I) h(d_i, d_o | I) dd_o = t(m_1 | I) v(m_2 | m_1, I)$$

For simplicity, let us investigate more closely the particular case when the modeling errors are negligible given by Equation (21). When we apply the division of the parameter vector and treat $m_2$ as nuisance parameters, this equation becomes

$$p(m_1, m_2 | d_s, I) = t(m_1 | I) \frac{u(d_i, m_2 | m_1, I)}{h(d_i | I)},$$

or when $m_2$ are fixed parameters, it becomes

$$p(m_1 | m_2, d_s, I) = t(m_1 | I) \frac{u(d_i, m_2 | m_1, I)}{v(m_2 | I) g(d_o | m_1, m_2, I)},$$

where

$$u(d_i', m_2 | m_1, I) = v(m_2 | m_1, I) g(d_o | m_1, m_2, I).$$

Both formulations lead to a product of two terms. One of them is the prior distribution $t(m_1 | I)$ that corresponds to the local estimation problem described in Section 3. Thus, the objective here is to study the other term, involving the distributions $u$ and $g$ ($h$ is the normalizing distribution), to determine a computing scheme for the posterior distribution for the parameter $m_1$.

To put the developments so far into perspective, we have discussed three alternative Bayesian formulations for the geophysical inverse problem that are given by Equations (17), (64) and (65). The first alternative is...

\* Term usually employed in Bayesian inference to denote parameters one is obligated to infer, but has no immediate interest in.
the full calculation of the posterior distribution; it does not involve the division of the parameter vector \( \mathbf{m}_2 \). To use this approach and the methods for deriving priors (one-dimensional distributions of the type \( t(\mathbf{m}_1; \mathcal{I}) \)) of the previous section, it is necessary to employ a scheme, such as Equation (34), to build the multivariate prior distribution for the parameters from each univariate distribution. But then, the difficulties associated with the high dimensionality, that we have been trying to avoid, need to be addressed. The second approach that treats \( \mathbf{m}_2 \) as nuisance parameters is the most complete because it reduces the dimensionality of the problem and also takes into account all uncertainties associated with the parameters. Because of that all practical developments follow the nuisance parameter approach, which is discussed in detail in the next section. Finally, the third approach that considers \( \mathbf{m}_2 \) as numbers fixed a priori is limited, since it does not fully account for all the uncertainties (see Figure 1 and associated discussions). This approach looks only at a particular conditional probability for the parameter \( \mathbf{m}_1 \) given the values for \( \mathbf{m}_2 \). Because of that, this approach is only briefly discussed below.

### 4.1 Nuisance parameters

For this case, the \( \mathbf{m}_2 \) parameters must be eliminated from the problem to produce a final result representing a marginal distribution for \( \mathbf{m}_1 \). This can be accomplished by integration since

\[
\pi(\mathbf{m}_1 \mid \mathcal{I}) = \int f(\mathbf{m}_1, \mathbf{m}_2 \mid \mathcal{I}) \, d\mathbf{m}_2,
\]

where \( R \) represents the proper domain of integration of \( \mathbf{m}_2 \). If we apply this idea to Equation (64), we get

\[
\pi(\mathbf{m}_1 \mid \mathcal{I}) = \int \pi(\mathbf{m}_1 \mid \mathcal{I}) g(d_1 \mid \mathbf{m}_1, \mathbf{m}_2, \mathcal{I}) \, d\mathbf{m}_2,
\]

where \( \kappa \) is the normalizing constant that incorporates the distribution \( h \). In the integrand of this equation, we have a pdf on the data space, which is the likelihood function \( g \), and a multivariate prior conditional distribution for \( \mathbf{m}_2 \) \( \mathcal{I} \). The latter is a \((M-1)\)-dimensional distribution, which means that we still need to handle integration in high dimensional space. But the difference here is that this is now an iterative procedure, where at each step a different parameter is considered as \( \mathbf{m}_1 \) and the others \( \mathbf{m}_2 \) will be eliminated from the problem. Thus, it is intuitive to expect that we may discard some information about the parameters \( \mathbf{m}_2 \) as long as sufficient information about \( \mathbf{m}_1 \) is introduced through \( t \). Following this idea, the prior information can be divided into two parts: a part that defines only a normal distribution (e.g., mean and covariance information) and another that complements this information (e.g., higher-order moment information) in such way that

\[
\mathcal{I}_T = \mathcal{I}_N + \mathcal{I}_C.
\]

That is, the total information \( \mathcal{I} \) equals the logical sum of normal information \( \mathcal{I}_N \) and its complement \( \mathcal{I}_C \). Furthermore, we may assume that for parameters \( \mathbf{m}_2 \) only the information \( \mathcal{I}_N \) is used in each step, then we can write

\[
\pi(\mathbf{m}_1 \mid \mathcal{I}_T) = \pi(\mathbf{m}_1 \mid \mathcal{I}_N)
\]

and

\[
\pi(\mathbf{m}_1 \mid \mathcal{I}_N) = \pi(\mathbf{m}_2, \mathbf{m}_1 \mid \mathcal{I}_N).
\]

Equation (69) can be substituted into Equation (67) to yield

\[
\int_R g(d_1 \mid \mathbf{m}_1, \mathbf{m}_2, \mathcal{I}) \, d\mathbf{m}_2.
\]

The integrand plays the role of the likelihood function for the one-dimensional posterior \( \pi \). This can be emphasized by writing Equation (70) after the integration, as given by

\[
\pi(\mathbf{m}_1 \mid d, \mathcal{I}_T) = \pi(\mathbf{m}_1 \mid \mathcal{I}_N) \, s(d_1 \mid \mathbf{m}_1, \mathcal{I}_N).
\]

Because of that, for easy reference, the function \( s \) will be referred by the name of extended likelihood.

To illustrate this approach, consider the simple case where the distributions \( p \) and \( g \) are normal and the forward model is linear (Equation (27)). Thus, the pdf \( p \) in Equation (70) is normal with mean \( \mathbf{m} \) and covariance matrix \( \Sigma(\mathbf{N}(\mathbf{m}, \Sigma_{\mathbf{m}})) \), which can be written as

\[
f(\mathbf{m} \mid \mathcal{I}) = (2\pi)^{-(M-1)/2} |C_{\mathbf{m}}|^{-\frac{1}{2}} \exp \left[ -\frac{1}{2} (\mathbf{m} - \mathbf{m}_1)^T C_{\mathbf{m}}^{-1} (\mathbf{m} - \mathbf{m}_1) \right].
\]

If the covariance matrix of the above equation is partitioned as

\[
\Sigma(\mathbf{m}) = \begin{bmatrix} C_{\mathbf{m}_{11}} & C_{\mathbf{m}_{12}} \\ C_{\mathbf{m}_{21}} & C_{\mathbf{m}_{22}} \end{bmatrix},
\]

where \( C_{\mathbf{m}_{11}} \), \( C_{\mathbf{m}_{12}} \), \( C_{\mathbf{m}_{21}} \), and \( C_{\mathbf{m}_{22}} \) are \( (M-1) \times 1 \), \( (M-1) \times (M-1) \), \( (M-1) \times 1 \), and \( (M-1) \times (M-1) \) matrices, respectively, then Theorem 10.6.1 of Graybill (1983), which gives the expression for the marginal of a normal distribution, allows us to write

\[
\pi(\mathbf{m}_1 \mid \mathcal{I}_N) = \frac{1}{\sqrt{2\pi C_{\mathbf{m}_{11}}}} \exp \left[ -\frac{1}{2} (\mathbf{m}_1 - \mathbf{m})^T C_{\mathbf{m}_{11}}^{-1} (\mathbf{m}_1 - \mathbf{m}) \right].
\]
By adopting these models, it is possible to perform the integral in Equation (70) analytically, what we discuss next.

4.1.1 Normal extended likelihood

In addition to the specification of the model for $p$, consider the likelihood function given by Equation (33) for $g$. Using these models, the integral in Equation (70), denoted by $I$, can be written as

$$I = (2\pi)^{-\frac{N+1}{2}} |C_0|^{-\frac{1}{2}} |C_m|^{-\frac{1}{2}} \int_R \exp \left\{ -\frac{1}{2} \left[(d_0^* - Gm)^T C_0^{-1}(d_0^* - Gm) + (m - \bar{m})^T C_m^{-1}(m - \bar{m}) \right] \right\} \, dm_2. \tag{74}$$

If we now apply the same manipulation to get the data-translated likelihood in Equation (30) we get

$$I = (2\pi)^{-\frac{N+1}{2}} |C_0|^{-\frac{1}{2}} |C_m|^{-\frac{1}{2}} \exp [S(\bar{m})] \int_R \exp \left\{ -\frac{1}{2} \left[(m - \bar{m})^T C_p^{-1}(m - \bar{m}) \right] \right\} \, dm_2, \tag{75}$$

where

$$C_p = (G^T C_0^{-1} G + C_m^{-1})^{-1}. \tag{76}$$

the estimated parameter vector is given by

$$\bar{m} = \bar{m} + C_p M^T C_0^{-1}(d_0^* - G\bar{m}) \tag{77}$$

and the estimated misfit by

$$S(\bar{m}) = (d_0^* - G\bar{m})^T C_0^{-1}(d_0^* - G\bar{m}) + (m - \bar{m})^T C_m^{-1}(m - \bar{m}). \tag{78}$$

Our main problem now is to evaluate the integral for the parameters $m_2$. This task is facilitated with the above manipulation, since if we suppress all the constant terms in Equation (75) we are left with

$$I_2 = \int_R \exp \left\{ -\frac{1}{2} \left[(m - \bar{m})^T C_p^{-1}(m - \bar{m}) \right] \right\} \, dm_2. \tag{79}$$

We now let the inverse of the estimated covariance be given by

$$R = C_p^{-1} = G^T C_0^{-1} G + C_m^{-1}. \tag{80}$$

Using the same matrix partition as before, we can write

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{bmatrix}, \quad C_p = \begin{bmatrix} C_{p11} & C_{p12} \\ C_{p21} & C_{p22} \end{bmatrix}. \quad \text{and} \quad C_p = \begin{bmatrix} C_{p11} & C_{p12} \\ C_{p21} & C_{p22} \end{bmatrix}. \tag{81}$$

Now, again according to the theorem of Graybill (1983) we have

$$I_2 = (2\pi)^{-\frac{N+1}{2}} |R_{22}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} [m - \bar{m}_1]^T C_{p11}^{-1}(m - \bar{m}_1) \right\}. \tag{82}$$

Substituting this result back into Equation (73) we finally get

$$I = (2\pi)^{-\frac{N+1}{2}} |C_0|^{-\frac{1}{2}} |C_m|^{-\frac{1}{2}} |R_{22}|^{-\frac{1}{2}} \exp \left\{ \frac{1}{2} S(\bar{m}) \right\} \exp \left\{ -\frac{1}{2} (m - \bar{m}_1)^T C_{p11}^{-1}(m - \bar{m}_1) \right\}. \tag{83}$$

4.1.2 Nature of the approximation

We can better understand the nature of the approximation made in Equation (69), considering the definitions for $I_T$, $I_N$ and $I_C$ in Equation (68), by fully expanding the conditional probability $v$ according to

$$P(m_2 | m_1, I_T) = P(m_2 | m_1, I_N + I_C) = \frac{S_1}{S_2}. \tag{84}$$

where in general

$$S_1 = P(m_2, m_1 | I_N) + P(m_2, m_1 | I_C) - P(m_2, m_1 | I_N + I_C), \tag{85}$$

$$S_2 = P(m_1 | I_N, I_C) P(I_C). \tag{86}$$

However, according to the definition of $I_N$ and $I_C$, they are independent, in which case

$$S_1 = P(m_2, m_1 | I_N) + P(m_2, m_1 | I_C) - P(m_2, m_1 | I_N + I_C). \tag{87}$$

and

$$S_2 = P(m_1 | I_N) P(I_N) + P(m_1 | I_C) P(I_C). \tag{88}$$

In either case

$$P(I_N) + P(I_C) = 1, \tag{89}$$

or

$$P(I_N) + P(I_C) = 1. \tag{90}$$

Thus the statement made by Equation (69) becomes clear. It says that the weight of the information $I_N$ for $m_2$ is such that $P(I_N) \ll P(I_C) - P(I_N, I_C)$ or $P(I_N) \ll P(I_C)$, depending on the case. This amounts to having practically no information about cross-correlations of order higher than two for the parameters. But we still can make use of the higher-order information on individual parameter through the probability density $t$ in Equation (70).

Analytically, this can be illustrated using the maximum entropy distribution, Equation (40) of Section 3. Substituting that equation for $t$ in Equation (70), using
Equation (73) for \( q \) gives
\[
\frac{t(m_1; \mathcal{I}_X)}{q(m_1; \mathcal{I}_X)} \propto \exp \left[ -\left( \lambda_1 \frac{m_1}{C_{m_1}} \right) m_1 + \left( \frac{1}{2C_{m_1}} \right) m_1^2 - \sum_{n=3}^{\infty} \lambda_n m_n^3 \right].
\] (88)

A very important particular case to consider is when all we have is actually \( \mathcal{I}_X \) (i.e., \( \mathcal{I}_C \) vanishes). This corresponds to a maximum entropy problem constrained by the first two moments of the unknown distribution, which leads to a normal distribution. More precisely, the Lagrange multipliers \( \lambda_n = 0 \), for \( n = 3, 4, \ldots \) and
\[
\lambda_1 = -\frac{\mu}{\sigma^2}
\]
and
\[
\lambda_2 = \frac{1}{2\sigma^2},
\]
where \( \mu \) and \( \sigma^2 \) are the mean and the variance supplied to the maximum entropy problem. Of course to be consistent we need \( \mu = \bar{m}_1 \) and \( \sigma^2 = C_{m_1} \), which yields a ratio of unity for \( \lambda_2 \). Thus, all that is left is the extended likelihood function, which is simply a multinormal Bayesian inversion procedure. To summarize, when all we know is really only the first and second-order moments, Equation (70) reduces to the more traditional Gaussian Bayesian formula (see, e.g., Tarantola, 1987).

4.2 Fixed parameters

To analyze this case, we can use the same simplifications of the previous section to rewrite (65) as
\[
p(m_1; \mathcal{I}_X) = \frac{t(m_1; \mathcal{I}_X)}{q(m_1; \mathcal{I}_X)} \frac{v(m_1, m_2)}{v(m_2)} g(d_1; m_1, m_2),
\]
or if we rewrite it in terms of the extended likelihood function
\[
s(m_1, m_2; \mathcal{I}_X) = \frac{t(m_1, m_2; \mathcal{I}_X)}{q(m_1, m_2; \mathcal{I}_X)} \frac{v(m_1, m_2)}{v(m_2)} g(d_1; m_1, m_2).
\] (89)

To analyze this expression, the ratio in the second factor can be represented by
\[
\frac{v(m_1, m_2; \mathcal{I}_X)}{v(m_2)} g(d_1; m_1, m_2) = \frac{v(H_0)}{v(H_1)} g(H_0; m_1, m_2) = \frac{v(H_0)}{v(H_1)},
\] (90)
which is similar to the ratios used for simple hypothesis testing with \( H_0 \) being the simple hypothesis\(^\dagger\) and \( H_1 \); the simple alternative\(^\ddagger\). This term provides a measure of how significantly the parameters \( m \) influence the fit to the data. This may suggest a connection between this ratio and resolution. For instance, if the parameter \( m \) has no significance in explaining the observed data it can be removed, causing the ratio to become just unity. Conversely, if the ratio is close to one, it is an indication that parameter \( m \) plays no significant role in explaining the data. To proceed further in the analysis, let us consider again the same normal distributions of the previous section. Then, the second term of Equation (89), denoted by \( f_1 \), can be written as
\[
f_1 = \frac{\exp \left[ -\frac{1}{2} \left( \delta d^T C_\mathcal{I}^{-1} \delta d + \delta m^T C_m^{-1} \delta m \right) \right]}{\exp \left[ -\frac{1}{2} \left( \delta d^T C_\mathcal{I}^{-1} \delta d + \delta m^T C_m^{-1} \delta m \right) \right]},
\] (91)
with
\[
\delta d = d'_0 - G m,
\]
\[
\delta d' = d'_0 - G' m_2,
\]
and
\[
\delta m = m - \bar{m},
\]
\[
\delta m_2 = m_2 - \bar{m}_2.
\]
\( G' \) is the forward operator \( G \) without the columns corresponding to the parameter vector \( m_2 \). If we rewrite the previous expression in terms of the data translated likelihoods we have
\[
f_1 \propto \exp \left[ -\frac{1}{2} \left( (m - \bar{m})^T C_p^{-1} (m - \bar{m}) \right) \right]
\]
with
\[
C_p = (G^T C_\mathcal{I}^{-1} G + C_m^{-1})^{-1},
\]
and
\[
C_p^{-1} = (G^T C_\mathcal{I}^{-1} G + C_m^{-1})^{-1}.
\]

At this point the usefulness of this approach is not completely clear. For inference purposes, it is not very attractive in comparison with the approach of the previous section, because it only makes partial use of the uncertainty. This is because we are considering only a particular cross-section of the joint distribution corresponding to values for the parameters \( m_2 \) fixed \( a \ priori \) (see Figure 1).

5 EXAMPLE

To develop a better understanding of the proposed methodology, the results of some calculations are presented, beginning from a simple problem. Consider the problem of density inversion from gravity data, where the sources are six rectangular cells of constant density (Figure 6).
The true density contrast is derived by imposing an exponential correlation function on a sequence of uncorrelated Gaussian pseudo-random numbers. The formula for the gravity of prismatic bodies, which is available from many different sources (see, e.g., Telford et al., 1976, p. 74), is used to compute the synthetic gravity field.

In addition to the earth model and observed field, it is also necessary to generate the prior information. Besides assuming that the exponential correlation function is known, we simulate the uncertainties for the density in each cell based on selected probability density functions (solid line plots in Figure 7). This is done by setting the mean of the selected pdfs to the true density contrasts, what makes it possible to generate samples whose averages approximate these true values and whose higher-order moments give a synthetic degree of uncertainty (e.g., variance, skewness and kurtosis). The range of possible values taken by the random samples is limited to a certain interval, what introduces an error between two means computed from each pdf using its natural domain and the truncated domain, respectively. This error can be made small by choosing the truncation interval large enough. The numbers for this example were computed from one thousand samples generated from each distribution truncated at $[-7, 7]$. If we consider the physical property density, this range of density contrasts makes sense only if we think in terms of chemical elements and not in terms of rocks. Table 2 shows the true value for the density contrast and the values obtained considering the particular choice of truncation interval.

The averages computed from the samples generated using the selected pdfs can be used as prior information
Figure 8. First five hundred samples generated from the selected theoretical distributions (Figure 6) to represent the uncertainty about the density contrast in each cell.

Table 2. Comparison between the true value set for the density contrast in each cell and values derived from the truncated distributions and from the samples drawn from these distributions.

<table>
<thead>
<tr>
<th></th>
<th>True values</th>
<th>Truncated mean</th>
<th>Sample mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.50634</td>
<td>1.50237</td>
<td>1.51462</td>
</tr>
<tr>
<td>2</td>
<td>0.26004</td>
<td>0.26004</td>
<td>0.25899</td>
</tr>
<tr>
<td>3</td>
<td>2.03991</td>
<td>2.04381</td>
<td>2.03918</td>
</tr>
<tr>
<td>4</td>
<td>1.79565</td>
<td>1.79561</td>
<td>1.82480</td>
</tr>
<tr>
<td>5</td>
<td>1.28441</td>
<td>1.24882</td>
<td>1.16753</td>
</tr>
<tr>
<td>6</td>
<td>0.17437</td>
<td>0.14920</td>
<td>0.21767</td>
</tr>
</tbody>
</table>

for a maximum entropy distribution calculation. The results from this calculation provide the local uncertainty for each parameter. Thus within the proposed methodology,
the next task, according to Equation (70), is to derive the
global part of the problem using the normal approximation,
where marginals can be computed analytically.

The global part of the problem is just the full Gauss-
ian Bayesian inversion, which is composed of the like-
lihood function and a prior, both taken to be normal
distributions. The likelihood function is constructed in
the usual way, according to Equation (33), with the data
error covariance matrix assumed to be known. The prior
distribution is built by combining the information from
the exponential correlation function between the par-
eters and the first two sample moments. This informa-
tion is used to derive the covariance matrix, which is just
correlations \( \rho \) scaled by the sample variances \( \sigma \), i.e.,

\[
C_m = \begin{bmatrix}
\sigma_1^2 \rho_{11} & \cdots & \sigma_1 \sigma_{\rho_{12}} \\
\vdots & \ddots & \vdots \\
\sigma_{\rho_{12}} & \cdots & \sigma_{\rho_{22}}
\end{bmatrix}
\]

and the vector of initial guesses, which are given by the
first sample moments. These two functions are combined
in analytical expression arising from the integration of
the parameters \( m_0 \), which is given in Equation (82).
Next, to get the final answer (i.e., the posterior distribu-
tion) for a specific parameter, we need to take the product
of this equation and the maximum entropy distribution,
normalized by the marginal of the Gaussian prior \( q \). To
get all the parameters, it is necessary to iterate as many
times as the number of parameters.

The maximum entropy distribution calculation fol-
low the work of Mead and Papanicolaou (1984) modified
to include a uniform reference prior. The particular im-
plementation for this example uses the Newton method
with line search. Sample moments up to the fourth order
are used as input to the optimization routine. The iter-
ations of the algorithm are stopped when the moments of
the computed maximum entropy distribution agree with
the input sample moments to the order \( 10^{-6} \) or better.
For this run, it usually took 6 to 7 iterations for con-
vergence. The final approximations of the theoretical dis-
bributions are shown by dashed lines in Figure 7. The overall
agreement between the estimated and true distributions
is good, except for the Cauchy distribution (6). However,
this is not very important since the estimated distribu-
tion carries the right moment information, according to
the convergence criterion. It only tells that moments up
to fourth order fall short of being able to correctly rep-
resent the Cauchy distribution, but this is not part of
the goals. The only goal is to make inferences about the
density contrast. The decision as to what are the rele-
vant moments for this goal have already been made a
\textit{priori}. In fact, later results show that the gain for going
up to fourth order is marginal for this problem; the abil-
ity to do so is important nevertheless, since it permits
the introduction of diverse prior information.

To understand the behavior of the proposed estim-
ation methodology, three levels of uncorrelated Gaussian
random numbers are used to corrupt the synthetic gravity
data. The standard deviations for each noise level are
respectively 1, 10 and 100% of the maximum syn-
thetic gravity value. Then, several measures of central
tendency are computed as estimates for the density con-
trast in each cell. They are the mean and the mode of
the posterior distribution and the mean of the Gaussian
Bayesian problem alone, which are given in Table 3. At
very low noise level (1%), there is a general agreement
between the three estimates. As we increase the noise
level to 10% of the maximum gravity value, the mean
and the mode of the posterior distribution start to pull
apart, but the posterior means and the Gaussian means
are still essentially equivalent. At extreme noise condi-
tion, the 100% level, the mean of the posterior distribu-
tions is overall slightly closer to the true values than are
the Gaussian means, and the modes become extremely
biased. This reflects the degradation of the gravity in-
formation, which allows for more features contained in
the prior information to show up. Overall the posterior
means and the Gaussian means are equivalent, with only
a marginal advantage for the posterior means for the case
of extreme noise. When the posterior means have pulled
away from the true parameter in comparison with the
Gaussian means, the numbers are written in bold face in
Table 3. In some of those occasions the effect seems to be
systematic.

It is important to point out that the fact that the
posterior contains more features of the distributions used
to generate the samples as noise level increases is just
a natural consequence of the lack of resolution data. This is a feature desired only in the sense that
it is supplying information not contained in the gravity
data, since our goal is not to reproduce those theoretical
distributions of Figure 7. What we want is to draw in-
ferrances about the density contrast of the cells. Thus, to
see how the methodology is performing, we must look for
the errors associated with the density estimates. Table 4
shows a comparison between the prior, the Gaussian and
the posterior variances for the density contrast in each
cell. Overall, there is a marginal reduction in the pos-
terior variances when compared with the Gaussian var-
iances. However, for few occasions shown as bold num-
bers in Table 4, the posterior variances increase. These
increases tend to be associated with the asymmetrical
distributions.

As observed above, at low noise level the prior in-
formation is made irrelevant by the gravity data and we
Parameter estimates

<table>
<thead>
<tr>
<th></th>
<th>1% noise</th>
<th></th>
<th>10% noise</th>
<th></th>
<th>100% noise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>true</td>
<td>mean</td>
<td>mode</td>
<td>true</td>
<td>mean</td>
<td>mode</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>true</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1.506</td>
<td>1.491</td>
<td>1.491</td>
<td>1.410</td>
<td>1.408</td>
<td>1.097</td>
</tr>
<tr>
<td>2</td>
<td>0.280</td>
<td>0.291</td>
<td>0.291</td>
<td>0.197</td>
<td>0.189</td>
<td>0.210</td>
</tr>
<tr>
<td>3</td>
<td>2.050</td>
<td>2.163</td>
<td>2.163</td>
<td>2.316</td>
<td>2.310</td>
<td>1.990</td>
</tr>
<tr>
<td>4</td>
<td>1.796</td>
<td>1.659</td>
<td>1.669</td>
<td>1.912</td>
<td>1.906</td>
<td>1.999</td>
</tr>
<tr>
<td>5</td>
<td>1.284</td>
<td>1.135</td>
<td>1.135</td>
<td>0.784</td>
<td>0.549</td>
<td>0.861</td>
</tr>
<tr>
<td>6</td>
<td>0.174</td>
<td>0.303</td>
<td>0.303</td>
<td>0.160</td>
<td>0.160</td>
<td>0.396</td>
</tr>
</tbody>
</table>

Table 3. Values of central tendency taken from the posterior distribution (means and mode) and from the Gaussian Bayesian solution ($\tilde{m}$). Bold numbers indicate that the posterior mean moved away from the true values in comparison with $\tilde{m}$.

Variances

<table>
<thead>
<tr>
<th></th>
<th>1% noise</th>
<th></th>
<th>10% noise</th>
<th></th>
<th>100% noise</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Prior</td>
<td>Gaussian</td>
<td>Posterior</td>
<td>Gaussian</td>
<td>Posterior</td>
<td>Gaussian</td>
</tr>
<tr>
<td>1</td>
<td>1.19684</td>
<td>0.00168</td>
<td>0.00168</td>
<td>0.04814</td>
<td>0.04727</td>
<td>0.33775</td>
</tr>
<tr>
<td>2</td>
<td>0.02227</td>
<td>0.00061</td>
<td>0.00062</td>
<td>0.00629</td>
<td>0.00567</td>
<td>0.01450</td>
</tr>
<tr>
<td>3</td>
<td>2.04926</td>
<td>0.03321</td>
<td>0.03302</td>
<td>0.33629</td>
<td>0.33938</td>
<td>0.73396</td>
</tr>
<tr>
<td>4</td>
<td>0.81267</td>
<td>0.00947</td>
<td>0.00947</td>
<td>0.11950</td>
<td>0.12284</td>
<td>0.29358</td>
</tr>
<tr>
<td>5</td>
<td>1.22102</td>
<td>0.03151</td>
<td>0.03214</td>
<td>0.34828</td>
<td>0.25422</td>
<td>0.61830</td>
</tr>
<tr>
<td>6</td>
<td>3.21465</td>
<td>0.01227</td>
<td>0.01124</td>
<td>0.20426</td>
<td>0.19330</td>
<td>1.14011</td>
</tr>
</tbody>
</table>

Table 4. Comparison between the prior, the Gaussian and the posterior variances for the parameters at different noise levels. Bold numbers indicate posterior variances that are greater than their Gaussian counterpart.

are basically free to choose any one of the estimates. At high noise level, however, this is not the case, the choice for estimates can make a difference. If we use only the Gaussian approach, the decision is still simple. Take the mean and the variance for the estimates and the error bars, respectively. However, for the posterior distribution, the decision is not straightforward. What should we pick for the estimate and the associated error for the density contrast? Here that we have the true values available, it is clear that the mean is the best approximation. We can understand this intuitively by thinking that for asymmetrical distributions the mode can be extremely biased towards high or low values, depending on the nature of the asymmetry. This happens, for example, for a monotonic distribution such as the exponential, where the mode will always be zero regardless of its mean (see Figure 7). For the error bars, the issues are basically the same. The asymmetry of the posterior distribution makes it inappropriate to have centered error bar of the type mean $\pm n\sigma$, where $n$ is an integer and $\sigma$ is the standard deviation. Instead, as illustrated in Figures 10 – 12, we can take 95% interquartile intervals, which can be computed independently of any estimates for the density contrasts. This is done by finding the parameter values corresponding to probability 1 – 0.025 and 0.025, which can be computed by inversion of the posterior cumulative distribution (Figure 13).

However, density estimates are necessary to compute the synthetic gravity field used in the fitting procedure, which is shown in Figure 9 with the synthetic, the noisy and the estimated gravity fields overlapped. We use the mean of each posterior distribution to compute the synthetic gravity field.

Another useful type of analysis is to perform several runs of the inversion scheme for different noise values with the same standard deviation. A possible choice from the previous section is the 10% noise level. The results are shown in Tables 5 and 6. The overall behavior of
Figure 9. The synthetic (dot-dashed line), the noisy (dashed line) and the estimated (solid line) gravity fields for 1(a), 10(b) and 100% (c) of the maximum synthetic gravity values. The estimated field uses the posterior mean for the parameters.

<table>
<thead>
<tr>
<th>Parameter estimates</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

Table 5. Comparison between the Gaussian means and the posterior means in several inversion runs considering different noise realizations with the same variance. Bold numbers indicate that the posterior mean (second column) moved away from the true values in comparison with in (first column).

the solution is basically the same as discussed above. It is interesting to note that the differences between the Gaussian and posterior means tend to be systematic.

6 CONCLUSIONS

Bayesian methods provide an adequate framework to integrate information of diverse origin and to carry uncertainty analysis associated with an inverse problem. The general Bayesian formulation for the geophysical inverse problem can be obtained by the application of the product rule of probability theory which carries logical mechanisms of commutativity. Furthermore, when we introduce a new variable for predicted data, all particular cases, such as noiseless observed data or perfect forward modeling, comes out in a coherent way. The greatest difficulty with Bayesian methods comes from the high dimensionality of the parameter and data vectors, which leads to probability distributions with the same dimensions. When we explore the possibility of using local prior information to derive priors for each parameter independently of the others, we can minimize the dimensionality problem. This is done by using part of the local information to construct a high-dimensional normal distribution that can be mar-
Figure 10. Inversion results depicted by the posterior marginal for each parameter. The 95% interquantile regions are represented by the shaded areas and the true value for the parameter are given by the solid circles. This example uses the 1% of the maximum gravity value as standard deviation for the noise.

<table>
<thead>
<tr>
<th>Variances</th>
<th>First run</th>
<th>Second run</th>
<th>Third run</th>
<th>Fourth run</th>
<th>Fifth run</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.06899</td>
<td>0.06725</td>
<td>0.08705</td>
<td>0.08433</td>
<td>0.05096</td>
</tr>
<tr>
<td></td>
<td>0.04999</td>
<td>0.03201</td>
<td>0.03162</td>
<td>0.07586</td>
<td>0.07378</td>
</tr>
<tr>
<td>2</td>
<td>0.01051</td>
<td>0.00866</td>
<td>0.01196</td>
<td>0.00922</td>
<td>0.01006</td>
</tr>
<tr>
<td></td>
<td>0.00943</td>
<td>0.00764</td>
<td>0.00980</td>
<td>0.01150</td>
<td>0.01022</td>
</tr>
<tr>
<td>3</td>
<td>0.39925</td>
<td>0.36971</td>
<td>0.43777</td>
<td>0.40204</td>
<td>0.31180</td>
</tr>
<tr>
<td></td>
<td>0.29838</td>
<td>0.37332</td>
<td>0.35329</td>
<td>0.51726</td>
<td>0.47890</td>
</tr>
<tr>
<td>4</td>
<td>0.14969</td>
<td>0.149331</td>
<td>0.20168</td>
<td>0.18589</td>
<td>0.16488</td>
</tr>
<tr>
<td></td>
<td>0.16360</td>
<td>0.11519</td>
<td>0.11688</td>
<td>0.19007</td>
<td>0.18522</td>
</tr>
<tr>
<td>5</td>
<td>0.32999</td>
<td><strong>0.41494</strong></td>
<td>0.27640</td>
<td><strong>0.28377</strong></td>
<td>0.29010</td>
</tr>
<tr>
<td></td>
<td>0.21888</td>
<td>0.22422</td>
<td>0.22884</td>
<td>0.28123</td>
<td><strong>0.33119</strong></td>
</tr>
<tr>
<td>6</td>
<td>0.23107</td>
<td>0.21816</td>
<td>0.30497</td>
<td>0.28345</td>
<td>0.16861</td>
</tr>
<tr>
<td></td>
<td>0.16166</td>
<td>0.17777</td>
<td>0.16985</td>
<td>0.34031</td>
<td>0.47809</td>
</tr>
</tbody>
</table>

Table 6. Comparison between the Gaussian variances and the posterior variances for several inversion runs using different noise values with the same variance. Bold numbers indicate the cases where the posterior variance is greater than the Gaussian variance.
Figure 11. Inversion results depicted by the posterior marginal for each parameter. The 95% interquantile regions are represented by the shaded areas and the true value for the parameter are given by the solid circles. This example uses the 10% of the maximum gravity value as standard deviation for the noise.

ginalized analytically. After marginalization we are left with only a one-dimensional distribution for a specific parameter that carries the geophysical information. This distribution can be combined with the rest of the local information by simple multiplication. In this way, we replace a high-dimensional inverse problem by series of estimation steps, involving one parameter at a time.

7 ACKNOWLEDGMENTS

, the many discussions with Richard O. Hansen and the advises from Kadri Dagdelen on geostatistical methodologies.

The authors are grateful for the discussions with Richard O. Hansen, Kadri Dagdelen, Wences Gouveia and Ken Larner. This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, the sponsors of the Gravity and Magnetics Project, both at the Colorado School of Mines, the Shell Foundation and the Army Research Office. In addition, the first author acknowledges the support of the Conselho Nacional de Desenvolvimento Cientifico e Tecnológico (CNPq, Brazil).

References


for generating smooth models from electromagnetic sounding data. Geophysics, 52, 289–300.


Figure 13. Points of the cumulative distribution computed from the posterior distribution for each parameter, using an adaptive method.

five lessons. In Short Course in Geology, Volume 8, Washington, D. C. American Geophysical Union.
Entropy-based complexity of optimization problems

H. Lydia Deng  
Center for Wave Phenomena and Department of Mathematical & Computer Sciences  
Colorado School of Mines  
Golden, Colorado 80401, USA

John A. Scales  
Center for Wave Phenomena and Department of Geophysics  
Colorado School of Mines  
Golden, Colorado 80401, USA

ABSTRACT

Optimization problems arise in every scientific field, whether as a statement of a fundamental principle, such as the principle of maximum entropy, or as an encapsulation of a given problem, such as maximizing the coherence of the output of some image-processing algorithm. It is often the case in practice that the function to be optimized cannot be specified in closed form in terms of elementary functions, but must be evaluated pointwise via a computer program. It goes without saying that it is desirable to have some way of selecting appropriate optimization tools for a given problem. In order to do so we must have some means of characterizing the structure or complexity of generic optimization problems. Further complicating matters is the high dimensionality of many of the functions encountered. Unable to visualize these high-dimensional functions directly, we propose a measure of complexity based on statistical analysis of samples of models found by some searching algorithms; we believe this measure, to some extent, quantifies how hard the resulting optimization problem is likely to be. Then, we use this measure to analyze some analytic functions. Finally, we analyze the multi-resolution analysis (MRA) of a highly multi-modal function arising in exploration seismology from the standpoint of this new entropy-based complexity measure.

Key words: Optimization, Objective Function, Global Search, Complexity, Topography, Entropy

What Makes an Optimization Hard?

We consider the problem of optimizing a function $F$ (the objective or cost function) mapping $\mathcal{M} \subset \mathbb{R}^N$ into $\mathcal{Y} \subset \mathbb{R}$. We refer to $\mathcal{M}$ as the model space. In most applications, the function $F$ cannot be expressed in closed form in terms of elementary functions, but must be evaluated pointwise. Depending on the application, the problem may be to find the global extremum of $F$, a single local extremum, or a collection of local extrema. If the structure of $F$ is unknown, then optimization is fundamentally a matter of search in the model space. In order to be able to treat such a broad variety of situations, we begin with an abstract statement of the search algorithm.

Algorithm 1. General Search (GS)

\[ \{m_t\} = GS(F, P, T, S) \]

Let $F : \mathcal{M} \subset \mathbb{R}^N \to \mathcal{Y} \subset \mathbb{R}$, $P = \{m^k_t\}_{t=1}^{\ldots,K}$ be an initial population with size $K \geq 1$, $T$ be a transition operator, and $S$ be a stopping criterion.

(i) Iteratively apply the transition operator to generate a new population of models at each iteration, so that $m^k_t = T m^k_{t-1}$;
(ii) Repeat (i) until $S$ is satisfied. The final set of models $\{m_r\}$ are the output of the search.

This is a general statement of searching, which can be considered as an evolution of a population of models on an $N$-dimensional lattice, where each node $m$ has a function value $F(m)$ attached to it. The transition operator $T$ is the rule that determines to which nodes each model walks at each step. For different choices of the initial population $P$ and transition operator $T$, we have different optimization algorithms.

There is no generally agreed upon characterization of what makes an optimization problem hard. Hardness has to do partly with our goals — do we need a global extremum or will a good local extremum do; partly with the structure of the function — does it have lots of local extrema, how broad is the basin of attraction of the extrema we seek; and partly with the dimensionality of the problem — exhaustive search will be infeasible except for low-dimensional problems, etc. Hardness comprises all of these things and more. Since the functions we are interested in can usually be evaluated only point-wise, some degree of global sampling is essential in order to achieve the characterization we seek.

Unfortunately, there are very few systematic comparisons of global optimization/sampling algorithms, either with other global algorithms, or with repeated application of more elementary methods. For instance, if a simulated annealing algorithm achieves a given result in a iterations, what can we conclude if a straightforward combination of random sampling and hill-climbing can achieve as good a result in fewer function evaluations? Has the sophisticated algorithm really done what we thought it was doing? The sheer size of the problems faced defeats simple answers to such questions. For a combinatorial problem with $N$ unknowns and each of which can take $m$ possible values, the number of possible models is $m^N$.

Imagine the surface of an objective function being a high-dimensional landscape with hills and basins of different depths and widths scattered on the surface. From our experience, the performance of searching algorithms depends to a large extent on certain topographical features of this landscape, such as the number of basins (local extrema), the widths and depths of these basin of attractions, etc. When the objective function is high-dimensional and can only be sampled point-wise, how best to acquire and analyze this topographical information is not clear.

Some hard combinatorial optimization problems

Hard combinatorial global optimization problems arise in various scientific and engineering fields. Some of the most widely studied include the spin-glass problem in statistical physics, the traveling-salesman problem (TSP) in computer science, and the residual statics problem in exploration seismology. Here we briefly describe these three problems. In this paper, we focus on the last problem; however, we believe that our results can be used to classify other hard optimization problems as well.

Spin-glasses

Spin-glasses (Edwards & Anderson, 1975; Fischer & Hertz, 1991) are disordered magnetic materials in which the orientation of nearby magnetic dipoles may be either parallel or anti-parallel. Models of spin-glasses typically consist of lattices of spins with each spin pointing either up or down. Suppose the entire system has $N$ spins, each up or down; there are a total $2^N$ possible configurations. Each configuration has a total energy given by a Hamiltonian (Hertz et al., 1991),

$$H = -\sum_{i,j} J_{ij} (s_i \times s_j), \quad s_i, s_j = \pm 1, \quad (1)$$

where $s_i$ and $s_j$ are the orientations (up or down) of the two spins, $J_{ij}$ is the energy weighting factor. The goal is to find a configuration, $\{s_i\}_{i=1, \ldots, N}$ that minimizes the Hamiltonian.

Traveling salesman problem (TSP)

The TSP is a well-known NP-complete optimization problem*. Given $N$ cities with distance (or cost) $C_{ij}$ between them. The task is to find the minimum-length (or cost) closed tour that visits each city exactly once and returns to its starting point. The objective function can be formulated as

$$F = \sum_{i=1}^{N} \sum_{j=i}^{N} C_{ij} x_{ij}, \quad x_{ij} = 0, 1, \quad (2)$$

where the unknowns $x_{ij}$ is 1 if the path connecting $i$ and $j$ is traveled and 0 if otherwise. For this problem, both the number of possible models and computation time grows exponentially with the number of cities $N$. For an unsymmetric TSP with $N = 100$ cities, there are $100!$ possible tours, which is approximately $10^{157}$.

* NP-complete is a class of optimization problems for which there are no known polynomial-time algorithms. Empirically, the computation time for solving an NP-complete problem grows exponentially with the dimensionality $N$. 


Residual statics

In exploration seismology, statics are the time shifts in seismic reflection data caused by heterogeneous material properties in the near surface. Figure 1 shows synthetic seismic traces (i.e., seismogram) that differ only by random shifts simulating the contamination of such heterogeneity. Essentially, the goal of residual statics estimation is to look for the time shift of each trace by maximizing the alignment of the traces. Consider an example where we need to align three identical traces. Fixing the first trace, we look for time-shifts for the second and third traces, $t_1$, $t_2$, so that the sum of squares of the stacked traces (stacking-power) is maximized. Figure 2 shows an example of such a two-dimensional objective function, which has multiple hills and basins of attractions scattered on the landscape. In practice, however, the stacking-power objective function is high-dimensional and highly multi-modal. Sophisticated global searches have generally been thought to be necessary for such problems (Rothman, 1985; Rothman, 1986).

Global search strategies

Among the searching methods defined via Algorithm 1, there are two extreme strategies, hill-climbing (HC) and uniform Monte Carlo (UMC). HC search is a local search applied to a single model (population size $K = 1$). An initial population $P = m_0$ is selected (possibly at random) and the transition operators $T$ are deterministic operators, such as conjugate gradient or downhill simplex, which follow a path downhill (or uphill for maximizing) as far as possible. For objective functions containing more than one local extrema (multi-modal), the final optimal model $m$ using HC strongly depends on the choice of the initial model $m_0$. UMC, other the other hand, selects points with uniform probability in the model space. The transition operation $T$ is simply the selection of new points at random and therefore makes no use of information from previous generations. Thus, if there are $M$ parameters and each of them can take $n$ possible values, the probability of finding a particular model is proportional to $n^{-M}$ for each function evaluation.

Many global search strategies have been developed that yield a compromise between two extremes; almost all of these incorporates stochastic elements, especially in the construction of transition operators. It is important for the success of global searches that the transition operators can make the best use of information provided by the current samples while avoiding being trapped in local extrema. Among all these strategies, the most widely used are Simulated Annealing (SA), (Kirkpatrick et al., 1983) and Genetic Algorithms (GA) (Holland, 1975) and random hill-climbing (RHC), to be defined shortly.

SA and GA searching strategies use stochastic transition operators $T$ that are biased towards good samples from the previous generations. Such schemes can therefore be considered as biased random walks on objective-function surfaces. Many variations of SA and GA can be found in the literature (Aarts & Korst, 1989; van Laarhoven & Aarts, 1987; Goldberg, 1989). Although the asymptotic convergence results are known for both SA and GA, these results are hardly useful in practice.

RHC searches, on the other hand, apply determin-
istic transition operations $T$ to a randomly chosen population of initial models $P$. RHC explores locally in multiple areas of objective functions, and the resulting samples are a set of local/global extrema. This search algorithm can be described as

**Algorithm 2. Random Hill Climbing**

\[
(M^t)_{k=0,\ldots,M} = RHC(F, K, \epsilon, cmaz)
\]

Let the initial population size be $K$. Let the stopping criterion $S$ be that either gradients of all samples are reduced to $\epsilon$ or the number of iterations reaches $cmaz$. Let $T_{local}$ be the local hill-climbing transition operator.

1. Choose initial models $P = \{m^0_k\}_{k=1,\ldots,K} \in \mathcal{M}$ uniformly at random, where $K \gg 1$;
2. Apply Algorithm 1, $\{m^t\} = GS(F, P, T_{local}, S)$.

The final population contains $M$ distinct models, $\{m^t\}_{k=0,\ldots,M}$.

In this paper, all RHC numerical results use the nonlinear Conjugate Gradient as transition operators (Deng et al., 1995).

**Why complexity?**

Highly multi-modal problems such as residual statics are usually treated by Monte Carlo methods. However, it has been observed by Whitby et al. (1995b) that for this problem, a RHC scheme can be more effective than a sophisticated GA. But it is not clear whether this is a general feature of statics or of the particular problem studied.

Clearly, the effectiveness of SA and GAs is somehow related to the surface landscape of the objective functions to be optimized. This situation is summarized by Kaufmann (1993):

Annealing works well only in landscapes in which deep energy wells also drain wide basins. It does not work well on either a random landscape or a “golf course” potential, which is flat everywhere save for a unique “hole”. In the latter case, the landscape offers no clue to guide search.

Recombination (in GAs) is useless on uncorrelated landscapes but useful under two conditions (1) when the high peaks are near one another and hence carry mutual information about their joint locations in genotype space and (2) when parts of the evolving system are quasi-independent of one another and hence can be interchanged with modest chances that the recombined system had the advantage of both parents.

In addition, as RHC searches use hill-climbing transition operators, it is easy to see that the performance of RHC is also largely affected by topographical features of objective functions described above, such as “wide basins”, “deep wells”, and “random landscape”, etc.

In order to study the performance of global optimization algorithms, it is important to quantify these qualitative descriptions of high-dimensional function surfaces. Characterizing the topographical features of high-dimensional functions is the main goal of this research. We believe that a quantitative measure of these features will be useful in designing optimal transition operators.

In this paper, we propose a measure of complexity of high-dimensional functions in the context of information. We first review some related work and discuss the criteria used to define our measure. We then show examples of applying this criterion to some analytical functions. Finally, we use this measure to analyze the behavior of a multi-resolution analysis (MRA) of the seismic statics problem.

**Measures of Complexity**

**Previous work on complexity**

Chavents (1991) developed sufficient conditions for an objective function to be locally convex. These conditions are based on the distance $\times$ curvature induced by the objective function on trajectories. This local convexity criterion could be generalized to global samples of an objective function, to provide a global measure of complexity.

A measure of the complexity can also be defined by analyzing the distribution of function values for a population of samples found by a given search algorithm (Wolpert & Magready, 1995; Magready & Wolpert, 1995). Such a measure therefore evaluates the performance of a searching algorithm for a given objective function. Based on this, the authors draw two conclusions on complexity of general combinatorial optimization problems:

1. No algorithm has a better (or worse) average performance than do other algorithms for all possible objective functions. This result is referred as the no free lunch (NFL) theorem (Wolpert & Magready, 1993).
2. No optimization problem is intrinsically harder than other optimization problems when averaged over all possible search algorithms. However, there do exist the optimal search algorithms for a specific optimization problem (Magready & Wolpert, 1995).

Although the Wolpert-Magready measure and conclusions are informative, they do not characterize the topography of the objective function.

Another proposal is that functions can be characterized by their spatial correlation properties (Weinberger, 1990; Stadler, 1992a). Several typical combinatorial optimization problems were investigated by studying the correlation in landscapes: the TSP (Stadler, 1992c), graph-bipartitioning problem (Stadler, 1992b), and the NK model problems, a spin-glass like problem in biology.
(Kauffman & Weinberger, 1989). Using correlation features of the objective function's landscape as a criterion, they study the effectiveness of some global algorithms for certain types of landscapes.

In addition, analyzing the topography of high-dimensional energy functions is important in physics. Berry and Breitengraser-Kunz (1995) studied topography and dynamics of multidimensional inter-atomic potential surfaces by analyzing a population of local minima, each of which has two saddle points connected to them. By connecting these samples in a certain order, the high-dimensional function surface is represented by a series of one-dimensional lines. By looking at these one-dimensional plots, the topography information is represented by the width and depth of the primary, secondary or tertiary basins of attractions (Berry & Breitengraser-Kunz, 1995).

Furthermore, the complexity of high-dimensional Hamiltonians can also be studied by means of entropy (Falciioni et al., 1995). For an N-dimensional Hamiltonian, some local extrema are first found by some local search algorithm. Contributions of these local-minima to the complexity is represented by a probability distribution \( P^{(k)}(N) \), where

\[ P^{(k)}(N) \propto \Delta^{(k)}(N). \]

Here, \( \Delta^{(k)}(N) \) is the estimated width of the kth basin of attraction. The complexity of the N-dimensional surface can be characterized by the following entropy,

\[ S(N) = - \sum_k P^{(k)}(N) \ln P^{(k)}(N) = \left( \ln \left( \frac{1}{\Delta(N)} \right) \right). \tag{3} \]

**Information measure of the complexity**

From our experience and previous studies, the hardness of optimization problems largely depends on topography of the objective functions. Therefore, the complexity of an optimization procedure can be considered equivalent as the topographical complexity of the objective function. The topographic features of a landscape are mostly attributed to the distribution of local extrema, such as,

(i) number of basin of attractions (local extrema);
(ii) width of each basin of attractions;
(iii) relative depths of the basins (function values of the extrema).

As shown in Algorithm 2, RHC explores various regions of the model space and takes initial samples downhill to the bottom of basins on the surface of functions. Therefore, the results of systematic RHC searches can be used to characterize the complexity of objective functions. However, the entropy-based complexity, as defined in equation (3), represents global features of the objective function landscape only by the number of basins of attractions and the widths of these basins. It does not take into account the relative depths of each basins. Figure 3 shows two functions with the same number of local minima and widths of basins of attractions. Using the criterion in equation (3), they would have the same complexity. However, the difficulty of minimizing these functions is different: the left function has identical basins of attractions, while the one on the right has a dominant global minimum and decreasingly important local minima away from the center.

We define our entropy-based complexity measure as follows:

**Definition 1. Entropy-Based Complexity**

Let \( F: \mathcal{M} \subseteq \mathbb{R}^N \rightarrow \mathcal{Y} \subseteq \mathbb{R} \). Let \( \{m_i\}_{i=1, \ldots, M} = \text{RHC}(F, K, \epsilon, \text{emax}) \) be the distinct converged models of an RHC search. \( \{h_i\}_{i=1, \ldots, M} \) be the histogram of the final population, and \( \{y_i = F(m_i)\}_{i=1, \ldots, M} \) be their corresponding function values. Let \( y_m = \min(y_i) \), and \( \sigma = (\sum y_i^2)^{1/2} \). Define the entropy-based complexity \( C_e \) as

\[ C_e = - \sum_{i=1}^M P(m_i) \ln(P(m_i)) \]

\[ = \left( \ln \left( \frac{1}{P(m_i)} \right) \right), \tag{4} \]

where \( P(m_i) \propto h_i \cdot v(y_i) \), in which

\[ v(y_i) = \begin{cases} 1, & \text{if } \sigma = 0; \\
\exp(-\frac{y_i - y_m}{\sigma^2}), & \text{otherwise};
\end{cases} \]

and \( P(m_i) \) is normalized to \( \sum_{i=0}^{M-1} P(m_i) = 1 \).

\( C_e \) defined in Definition 1 measures the amount of information gained about the global features of the objective function after an RHC. \( P(m_i) \) is a probability representing contributions to the global information by the ith converged model found by the RHC. This probability is proportional to the width of basin of attractions \( (h_i) \), weighted by the relative depths of the basins \( (v(y_i)) \). This
The complexity measure is similar to the entropy measure used in characterizing Hamiltonians (Falcioni et al., 1995), except that we take into account the distribution of function values at each local extrema.

Let us consider some examples of the application of this measure to a family of two-dimensional functions.

\[ F(m) = m^T Am + b^T w(f, m), \]  

(5)

where \( A \) is a square matrix, \( b \) is a constant vector and \( w(f, m) \) is an oscillatory vector function

\[ w_i = -\cos(f_i m_i), \quad i = 0, 1 \]

where \( f_i \) is the frequency in each direction. The function \( F(m) \) in equation (5) is multi-modal, and the local minima are caused by the oscillating term \( w(m) \). Hill-climbing searches of this function may converge to local extrema for arbitrary initial models. The difficulties of finding the global extremum, however, vary with the coefficients. For simplicity, we only consider equation (5) when

\[ A = a I_2, \quad b = b, \quad I; \quad f = f I_1, \]

where \( I_2 \) is a 2 \( \times \) 2 identity matrix and \( I_1 = (1, 1)^T \). In this case, \( m = (m_0, m_1) \) and equation (5) becomes

\[ F(m_0, m_1) = a (m_0^2 + m_1^2) - b \cos(f_0 m_0) - b \cos(f_1 m_1). \]

We perform RHC on equation (6) with a population of 500 models. The searches are stopped when the residuals are reduced to \( \epsilon = 10^{-5} \) or when the number of non-linear Conjugate Gradient iterations reaches \( c_{\text{max}} = 200 \).

Figure 4. Random hill-climbing with population of \( K = 500 \). (a) and (b) show the function surface defined in equation (6) when \( a = 0, b = 1 \) and \( a = 1, 0.001 \) respectively. (c) and (d) show the convergence of 500 initial models in the 2-D model space on functions (a) and (b). In the first case \( C_e = 0.01 \), in the second \( C_e = 4.11 \).

Figure 5. Random hill-climbing with population of \( K = 500 \). (a) and (b) show the function surface defined in equation (6) when \( a = 0.1, b = 1 \) and \( f = 1, 10 \) respectively. (c) and (d) show the convergence of 500 initial samples in the 2-D model space on functions (a) and (b). In the first case \( C_e = 2.14 \) and second case \( C_e = 3.95 \).

When the frequency of the oscillation is zero, \( f = 0 \), and \( F(m_0, m_1) \) becomes quadratic with a unique global minimum. Figures 4(a) and (b) show the function surfaces when \( a = 0, b = 1 \) and \( a = 1, 0.001 \) respectively. Both functions are quadratic. Noticing the scale of the two plots, however, Figure 4(b) is much flatter than (a). This difference in the curvature of surfaces influences the result of optimization significantly. Figures 4(c) and (d) show the converged models for the RHC search when the stopping criterion are met. In Figure 4(c), we see that all 500 models converge to the global minimum when the curvature of the quadratic function is large enough. As a result, \( C_e = 0 \). However, when the curvature is small, as in Figure 4(b), the maximum number of iterations is exceeded and the final models are scattered about the domain. In this case \( C_e = 4.1 \). This demonstrates that even for unimodal objective functions, the hardness of global search could also depend on the curvature of the landscape.

When the frequency \( f \) is not zero, function \( F \) is multi-modal. Figures 5(a) and (b) show the function surface when \( a = 0.1, b = 1 \) and \( f = 1, 10 \) respectively; both functions have the same quadratic term, but have different number of local minima. Figures 5(c) and (d) show the converged models using the RHC search when
stopping criterion are met. The converged models are more clustered at bottoms of basins on the low-frequency function surface than on the high-frequency one, and their entropy-based complexities $C_e$ are 2.14 and 3.93, respectively. Figure 6 shows the complexity $C_e$ as a function of the spatial frequency $f$ when the curvature $a$ is fixed. We see that $C_e$ increases as the number of local extrema increases.

Fixing the spatial frequency, the complexity of a multi-modal function is also influenced by the spatial curvature. Figure 7 shows the function surfaces of equation (6) when $f = 10$, $b = 1$ and $a = 1$, 10 in (a) and (b) respectively; both functions have the same number of local minima as does in Figure 5(b). Figures 7(c) and (d) show the converged samples when the stopping criterion are met. Compared with Figure 5(d), the converged models are more clustered with the increase of spatial curvature despite of the fact that all three functions have the same number of local minima in the model space. The entropy-based complexity $C_e$ of these two functions are 3.17 and 1.02, respectively. Figure 8 shows the complexity $C_e$ as a function of the spatial curvature $a$ for the same function when $f = 10$ and $b = 1$.

Applications to High-dimensional Test Functions
In this section, we use the complexity $C_e$ to study two commonly used optimization test functions, the Rosenbrock function and the Griewank function.
\[ R(x) = \sum_{i=1}^{N-1} \left[ 100(x_i - x_{i-1}^2)^2 + (1-x_{i-1})^2 \right], \]  
(7)

where \( x = (x_0, \ldots, x_N) \). Although unimodal, the long and narrow basin is a challenge for searching algorithms. Figure 9 shows the function surface and its contour when \( N = 2 \). When \( N \geq 2 \), the function is still unimodal, but it is not easy to see how the increase of dimensionality alters the difficulty of optimization.

One way of studying the spatial curvature of functions is by looking at the ratio of largest and smallest eigenvalues (condition number) of the Hessian. The Hessian for equation (7) is a tri-diagonal matrix,

\[
\begin{pmatrix}
  a_0 & c_0 & 0 & \cdots & 0 \\
  b_1 & a_1 & c_1 & 0 & \cdots \\
  0 & b_2 & a_2 & c_2 & 0 \\
  \vdots & \vdots & \vdots & \vdots & \vdots \\
  0 & 0 & \cdots & b_{N-1} & a_{N-1}
\end{pmatrix}
\]  
(8)

where

\[
a_0 = 2 - 1200 x_0^2 - 400 x_1, \\
a_i = 202 + 1200 x_{i-1}^2 - 400 x_i, \quad 0 < i < N - 1 \\
a_{N-1} = 200 \\
b_i = -400 x_{i-1} \\
c_i = -400 x_i, \\
\]

At the global minimum \((1, 1, \cdots, 1)\), the tri-diagonal matrix equation (8) becomes Toeplitz except for \( a_0 \) and \( a_{N-1} \):  

\[
\begin{align*}
  a_i &= 1002, \quad 0 < i < N - 1 \\
  b_i &= -400, \quad 0 < i < N - 1 \\
  c_i &= -400, \quad 0 \leq i < N - 1 \\
  a_0 &= 802 \\
  a_{N-1} &= 200.
\end{align*}
\]

The condition number of the Hessian at the global minimum reaches an asymptote with increasing dimension, as shown in Figure 10. Figure 11 shows the complexity measure \( C_r \) as a function of the number of dimensions; it shows the same asymptotic trend as does the condition number. Thus the increasing complexity for low dimensions is the result of increasing ill-conditioning of the Hessian and has nothing to do with local minima.

**High-dimensional Griewank functions**

The Griewangk function is another test function being used for testing optimization algorithms (Whitley et al., 1995a):

\[
g(x) = 1 + \sum_{i=1}^{N} \frac{x_i^2}{4000} - \prod_{i=1}^{N} \cos(x_i) / \sqrt{i}
\]  
(9)
The cosine term makes equation (9) multimodal. Figure 12 shows a one-dimensional slice of the Griewangk function along the diagonal of the hypercube for dimensions 1, 3, 5, 9. Whitley and Mathias (1995a) observed such slices and concluded that “as the dimensionality increases the local optima induced by the cosine decrease in number and complexity”.

However, such pictures can be misleading since they tell us only about low-dimensional projects of the function. Figure 13 shows slices of the same functions when all but one variables are fixed to be 0. The increasing dimensionality does not change the shape of the slices. On the other hand, Figure 14 shows slices of this function when all but one variables are fixed to be 2. Oscillation of the function has been reduced with the increasing dimensionality, but the curvature has not been increased as it was in Figure 12. Therefore, studying the overall performance of high dimensional functions could be tricky. We compute $C_e$ for the Griewangk function with a population 500 models in the domain of $-5 \leq z_i \leq 5$, $i=0, \ldots, N-1$. Figure 15 shows the resulting complexity $C_e$ for dimensions up to 50. This result gives us more confidence that we really understand the dimensional-dependence of complexity than by simply looking at hyper-planes.

**Complexity of Residual Statics Problems**

**A synthetic residual statics problem**

Let us return to the residual statics problem previously described. Figure 2 shows a statics objective function with two unknowns. In practice, however, the time-shifts of the traces are not independent. The statics of each trace are caused by the combined time distortion of near-source and near-receiver heterogeneities (source-statics
and receiver-statics). Figure 16 illustrates the similarity of travel paths near each source and each receiver.

The recorded reflection seismic signals are usually sorted into midpoints $y$ (of the source and receiver locations) and offsets $h$ (half distance between the source and receivers). Letting $\mathbf{s}$ and $\mathbf{r}$ be unknown vectors of source- and receiver-statics, this optimization problem can be formulated as

$$
\max_{\mathbf{s}, \mathbf{r}} F(\mathbf{s}, \mathbf{r}) = \sum_y \sum_{h_1 \neq h_2} \Phi_{y, h_1, h_2}(\tau(\mathbf{s}, \mathbf{r}));
$$

(10)

where $\Phi_{y, h_1, h_2}(\tau)$ is the cross-correlation between traces (after nominal correction for normal moveout) of offsets $h_1$ and $h_2$ at midpoint $y$ evaluated at

$$
\tau = s_i(y, h_1) + t_j(y, h_1) - s_i(y, h_2) - t_j(y, h_2);
$$

and $i(y, h)$ and $j(y, h)$ are the source and receiver indices for midpoint $y$ and offset $h$, respectively. Function $F(\mathbf{s}, \mathbf{r})$ in equation (10) is usually called stacking-power function.

Figure 17 shows the recording geometry of one example synthetic data set. This data set has 20 sources, 35 distinct receivers and 320 traces. All, traces are identical except for random source and receiver statics. These are generated by repeatedly shifting a single trace of field data. Thus, the objective function of equation (10) has 55 unknowns. When there are no statics in the data, the global maximum of the function is at the origin $(\mathbf{s}_i = \mathbf{0}, \mathbf{r}_i = \mathbf{0})$.

Figure 18 shows 2-D hyper-planes of the stacking-power function from slices in which all parameters are fixed at their correct values except the 10th and 11th source statics (left) and the 10th source and 20th receiver statics (right). For both slices, the function appears to have many local maxima, regularly scattered throughout the slices. On the other hand, Figure 19 shows the stacking-power function as functions of the 10th and 11th source statics (left) and the 10th source statics and the 20 receiver statics, where all other unknowns statics are chosen at random. The topography of the function is much less regular in this case.
Behavior of the multi-resolution analysis

Rather than use a Monte Carlo global optimization method to solve the statics problem, Deng (1995) has proposed simplifying the optimization via a multi-resolution analysis (MRA). The idea is to use a wavelet decomposition to generate successively simpler representations of the seismic data, thereby eliminating progressively more local extrema from the objective function. We now apply the entropy-based measure of complexity to this multi-resolution analysis and see if we can get a deeper understanding of just what is being accomplished by the MRA.

First let us define a multi-scale RHC algorithm:

Algorithm 3. Multi-scale Random Hill-Climbing

\((\{m\} = MRHC(F, L, \varepsilon, cmaz))\)

Let \(\{S_i\}_{i=L, \ldots, 0}\) be a sequence of increasingly smooth operators to be defined below, and \(S_0\) be an identity.

(i) Let \(f_L = S_L F\); choose an initial population \(\{m^s\}_{s=1, \ldots, K}\) with size \(K\) at random; apply Algorithm 2, so \(\{m^s\}_{s=1, \ldots, M} = RHC(f_L, K, \varepsilon, cmaz)\), and \(i = L-1\).

(ii) Let \(f_i = S_i F\), \((L > i \geq 0)\) and \(\{m^s\} =

(iii) Decrease \(i\) by 1, repeat (ii) until \(i = 0\). The final set of models \(\{m^s\}\) is the solution.

The smoothing operators \(\{S_i\}_{i=L, \ldots, 0}\) could be a sequence of low-pass filters with increasingly wider pass-band (Chen, 1994; Bunks et al., 1993), or a sequence of increasingly fine wavelet operators (Deng, 1995). The sequence of smoothing operators should be such that the resulting functions, \(\{f_i\}_{i=L, \ldots, 0}\), have the same global feature as the objective function \(F\) with increasing number of local optima, and \(f_0 = F\). Deng (1995) showed that this could be achieved using the shift-invariant basis of Saito and Beylkin (1993).

To achieve a more comprehensive picture of the performance of the multi-scale algorithm, we applied the Multi-scale Random Hill-Climbing algorithm to the 55 parameter statics problem using a population of 1000 models with 0 - 5 levels of wavelet decomposition. Figure 20 shows histograms of converged models achieved with a non-linear conjugate gradient approach. It can be seen that several levels of decomposition result in fewer distinct converged models. However, too much decomposition (level 5) may result in a function that is too flat for optimization. Figure 21 shows \(C_i\) calculated from one RHC for decomposition levels varying from 0 to 5, where level 0 represents using the original data. Repeated experiments on different initial populations and stopping criteria display the same shape of the \(C_i\) curve, though each \(C_i\) may vary according to population size and stopping criterion. These results indicate that complexity of
Figure 21. Complexity $C_e$ as a function of the level of wavelet decomposition.

the optimization is reduced via MRA, up to a point (level 4, in this example).

Conclusions & Future Work

We have developed an entropy-based measure of the complexity of function surfaces and used it to study the difficulty associated with generic optimization problems. We have shown the application of this measure to various analytic test functions as well as a practical problem in exploration seismology. In addition, we have used this measure to gain insights into the behavior of multi-resolution analysis.

Our criterion is based on a statistical analysis of the results of random hill-climbing, and measures the amount of information the hill-climbing can achieve about the topography of the function. The front-end for this calculation is the specification of an initial population of points in the domain of the function. At present we choose these with uniform probability on the interior of this domain. It is likely that there are more efficient ways of specifying these points in high dimensional spaces, for example, by quasi-Monte Carlo methods, or by subdividing the domain into small sub-domains and extrapolating the results.

The factors contributing to the information encapsulated by this measure are: the number of basins of attractions on the function landscape and the widths and depths of these basins. The number of distinct converged models from the random hill-climbing can estimate number of local optima, the histogram of the converged models is used to approximate widths of the basin of attractions, and the distribution of function values of each basin of attractions is the weighting factor of each probability. When $C_e = 0$, the function is unimodal with only one extremum. On the other hand, when the function is a constant, the complexity $C_e$ is maximum: $C_e = \ln K$, where $K$ is the population size of the random hill-climbing.

Ideally, this complexity, as a characterization of the topography of surfaces, should be independent of the searching algorithm and computing time. In practice, however, we can estimate statistically the topography of functions from only limited samples. Further, as the dimensionality increases, volumes become increasingly compressed near their boundaries. So it may be difficult to choose the initial population of models in such a way as to ensure efficient sampling of a function's domain.

At first glance it seems disturbing that a measure of complexity would be influenced by numerical issues such as ill-conditioning. But as a practical matter it is often difficult to untangle the results of ill-conditioning (flat landscape) from those of multi-modality, and so while it remains to investigate the sensitivity of our measure to the stopping criteria and search algorithms used, we believe that these are important aspects of what makes an optimization problem hard.

Acknowledgment

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, the Shell Foundation and Oak Ridge National Laboratory.

References


Estimating residual statics by optimizing a complexity-reduced stacking-power function

H. Lydia Deng  
*Center for Wave Phenomena and Department of Mathematical & Computer Sciences, Colorado School of Mines, Golden, Colorado 80401, USA*

Bin Wang & Keh Pann  
*Mobil Exploration & Producing Technical Center, Dallas, Texas 75247, USA*

**ABSTRACT**

Near-surface weathering layers introduce time anomalies in seismic data, which cause poor quality in stacked seismic sections. The ultimate goal of statics estimation in seismic data processing is to obtain high-quality stacked seismic sections. One measure of the stacking quality is stacking-power function. Therefore, statics estimation is a mathematical optimization problem where we look for the time-shifts among traces such that the stacking power is maximized.

We begin by reviewing both linear and non-linear approaches of conventional residual static estimation under the "umbrella" of this optimization. The time-picking approach is a over-simplified linearization of the stacking-power function where only one travel time in a correlation is used, while the stacking-power approach uses all trace information in correlations for complicated non-linear optimizations.

We then propose using partial information from the correlations, *envelope functions*, to formulate a new objective function. We conjecture that this new objective function approximates the global structure of its corresponding stacking-power function, and that the number of local maxima is significantly reduced. Statics are estimated by optimizing this simplified objective function using local-search algorithms.

This new algorithm is tested on two field data sets where routine processing failed to give good answers. It shows that this approach is particularly appropriate for data where large statics are involved.

**Key words:** Residual Statics, Objective Function, Stacking Power, Optimization, Envelope

**Introduction**

Land seismic data are generally contaminated by lateral variation of near-surface irregularities, which cause misalignment of reflection events prior stacking. This misalignment of events can usually be corrected by estimating surface-consistent statics-shifts for each source and receiver location. The ultimate goal of statics correction is to obtain high-quality stacked seismic sections. One measure of the stacking quality is the sum of squares (power) of the stacked sections. Therefore, residual statics estimation can be formulated as a mathematical optimization problem where we look for the time-shifts among traces such that the stacking-power function is maximized.

Conventional residual statics estimation by solving a linear system (Taner et al., 1974; Wiggins et al., 1976;
Larner et al. (1979) has been widely used by industry due to its computational simplicity. However, the only information used from data in this approach is one time-shift in each correlation. When data have poor signal-to-noise ratio, this time-picking may be difficult due to the presence of ambiguities. Choosing wrong peaks of correlations causes the alignment of wrong events in the seismic section, which is known as "cycle skips". This problem is especially serious where large statics (say, more than half a dominate period) exist in data. We recognize that the conventional linear-system can be considered as a result of linearization from the general algorithm of maximizing the stacking power. The condition of this approximation is that all seismic traces are identical except for small statics shifts.

Ronen and Claerbout (1985) proposed using stacking-power function directly as a measure for estimations. Using the stacking power as an objective function, seismic trace information is used rather than only one picked time from each correlation. A commonly acknowledged problem of the stacking-power criterion is that the objective function to be optimized is a high-dimensional, complex surface with many local maxima. Global search algorithms, such as simulated annealing (SA) (Rothman, 1985; Rothman, 1986) and genetic algorithms (GA) (Stork & Kusuma, 1992; Smith et al., 1992; Whitley et al., 1995b), are needed in the optimization of stacking-power functions. However, these global search algorithms tend to be computationally expensive; there is no guarantee that these global searches would converge to global maxima within a finite computation time. Unsuccessful global search may find local maxima, which causes "cycle skips" on the static corrected section.

Efforts have been made to simplify objective functions in order to avoid local optima in global searches for such difficult optimization problems. A multi-grid scheme was suggested to sub-sampled, low-frequency versions of data in seismic waveform inversion (Saleck et al., 1993; Chen, 1994; Bunks et al., 1995). Multi-resolution analysis (MRA) (Deng, 1995) using a shift-invariant wavelet base (Saio & Beylkin, 1993) has been also applied to seismic data for the purpose of simplifying objective functions. Although some successes have been reported, it has not been carefully studied that whether the objective functions were so severely distorted that they lose their qualification as a measure of stacking quality. Furthermore, Eolson (1992) pointed out that band limiting data cause values of local optima in stacking-power function approaching that of global optimum, therefore, ambiguity in optimization could be increased when searching on such a high-dimensional surface. Refer to Deng & Scales (1996) for a study of such issues.

In this paper, we propose estimating residual statics by optimizing a simplified objective function, the reduced stacking-power. This new objective function approximates global feature of its corresponding stacking-power function by using a smooth approximation for each correlation functions.

This residual statics estimation technique is applied on two field data sets, which have serious statics problems. After the estimated statics are corrected, results of stacking show better lateral coherence than the output from ProMax.

**Complexity of Residual Statics Estimation**

**Stacking power as an objective function**

Generally speaking, the ultimate goal for statics corrections is to improve stacked images; it is reasonable to use a measure of stack quality as an objective function in residual statics estimations. One measure of stacking quality is the sum of squares of the stacked traces — stacking power (Nadeil & Taner, 1971; Ronen & Claerbout, 1985; Rothman, 1985). A stacking power of \( N \) seismic traces may be written as

\[
F(\mathbf{y}) = \sum_{i=0}^{N-1} \left( \sum_{\tau=0}^{N-1} d_i(t + \tau) \right)^2,
\]

(1)

where \( d_i(t) \) is the \( i \)-th trace and each component of vector \( \mathbf{y} \) is the time-shift for a trace. It can be easily shown that maximizing the function in equation (1) is equivalent to maximizing a summation of cross-correlation functions,

\[
F(\mathbf{y}) = \sum_{i=0}^{N-1} \sum_{\tau=0}^{N-1} |\Phi_{ij}(\tau_j - \tau_i)|,
\]

(2)

where \( \Phi_{ij}(\tau) \) is the cross-correlation of the \( i \)-th and \( j \)-th traces evaluated at \( \tau = \tau_j - \tau_i \). Equation (2) is mostly used in implementations because cross-correlations \( \Phi_{ij}(\tau) \) may be pre-computed and stored.

In most practical residual statics problems, each component of \( \mathbf{y} \) is not completely independent to each other. For example, under the surface-consistent assumption, each trace shift \( \tau_i \) is a linear combination of unknown static-shifts of the corresponding source and receiver. Furthermore, when the residual normal moveout and subsurface structure are considered, \( \tau_i \) should be the linear function of all these contributing factors (Taner et al., 1974; Wiggins et al., 1976). In essence, a general residual statics problem is a mathematical optimization problem in which we look for parameter vectors \( \mathbf{y} \) so that

\[
\max_{\mathbf{y}} F(\mathbf{y}) = \sum_{i} \sum_{j \neq i} \Phi_{ij}(\tau_{ij}(\mathbf{y})),
\]

(3)
where $\tau_{ij}(\bar{\tau})$ is a linear function of the unknown parameters. The function $F(\bar{\tau})$ in equation (3) is referred to as the *stacking-power function*.

**Behavior of stacking-power functions**

In practice, stacking-power functions are generally highly multi-modal, high-dimensional surfaces with many local maxima. Smith *et al.* (1992) shows a two-dimensional projection of one such objective function, which presents an irregular shape with many local maxima.

By observing equation (3), however, we see that stacking-power functions have some special properties:

1. Each term $\Phi_{ij}$ in the summation of $F$ is a one-dimensional function, whose values are pre-computed. The function $F$ has maxima at points where all correlation functions exhibit minima (see Appendix A).
2. $F$ is *separable*, i.e., there are no non-linear interactions between variables. Therefore, a change of variable can be found so that the optimal value for each parameter can be determined independently of all other parameters (Whitely *et al.*, 1995a).
3. Not all maxima of $F$ are produced by maxima of one-dimensional function $\Phi_{ij}$ because the arguments of $\Phi_{ij}(\tau)$ are not completely independent (see Appendix A). However, for the purpose of statics estimation, we are not interested in those local-maxima that are not caused by the alignment of traces.

To demonstrate these observations, we start with a simple, schematic statics problem: consider three identical seismic traces which can be shifted independently of each other, and look for the independent time-shifts ($\tau_i$, $i = 0, 1, 2$) that can best align the traces. Without loss of generality, we use the first trace as a reference ($\tau_0 \equiv 0$) and look for time-shifts of other two traces ($\tau_1$ and $\tau_2$). The stacking power is a two-dimensional function,

$$F(\tau_1, \tau_2) = \Phi_{01}(\tau_1) + \Phi_{02}(\tau_2) + \Phi_{12}(\tau_2 - \tau_1).$$

If the three traces are identical, all correlation functions are auto-correlations. The left of Figure 1 shows one example of such correlations. The two-dimensional stacking-power function in this case is shown on the right of Figure 1, which demonstrates that some local maxima exist even in such a simple case. The converged models of local-search algorithms will strongly depend on the initial models, and the estimated statics may easily be at a local maximum if a local search is used. On the other hand, Figure 1 shows that the observable maxima are scattered regularly on the function surface. The global maximum is located at the point where maximum positive peaks of all three correlation functions coincide (the correct time shift). And the observable local maxima are regularly distributed at locations where positive peaks of all three correlation functions coincide. Among those local maxima, those that coincide with the maximum positive peak of one correlation function have significant values. In Figure 1, the dominate stripes (vertical, horizontal, and diagonal) are located at areas where either of the three correlation functions is at the highest maxima.

Observations from Figure 1 suggest that contributions to significant maxima of stacking-power function are mostly due to positive peaks of the correlations from which the stacking-power function is made of. Small and negative correlation values are of little interest to us.

**Linearization of stacking-power functions**

Equation (3) is generally a high-dimensional function in residual statics estimation of seismic data, it can be simplified by an approximation. Let $T_{ij}$ be the maximum-correlation time between the $i$th and $j$th traces. When $\tau_{ij}(\bar{\tau}) - T_{ij}$ is small, equation (3) can be approximated with the Taylor expansion,

$$F(\bar{\tau}) \approx \sum_{i} \sum_{j \neq i} \left[ \Phi_{ij}(T_{ij}) + \frac{\Phi'_{ij}(T_{ij})}{2}(\tau_{ij}(\bar{\tau}) - T_{ij})^2 \right],$$

where $\tau = T_{ij}$ corresponds the maximum of correlation function $\Phi_{ij}(\tau)$ with $\Phi'_{ij}(T_{ij}) = 0$ and $\Phi''_{ij}(T_{ij}) < 0$. With the above linearization, maximization of function $F(\bar{\tau})$ is approximated by a least-squared problem,

$$\min_{\tau} \sum_{i=0}^{N-1} \sum_{j=i+1}^{N-1} | \Phi'_{ij}(T_{ij}) | (\tau_{ij}(\bar{\tau}) - T_{ij})^2 .$$

If all traces are identical except for static-shifts, correlation function $\Phi_{ij}(\tau)$ is a shifted version of an auto-correlation $\Phi(\tau)$, with the highest maximum at $T_{ij}$, i.e., $\Phi_{ij}(\tau) = \Phi(\tau - T_{ij})$. Therefore, the above least-squared problem becomes,
\[
\min_{\bar{z}} \sum_{i} \sum_{j \neq i} (\tau_{ij}(\bar{z}) - T_{ij})^2.
\] (5)

Let \( \mathbf{A} \) be the linear operator representing the linear combination \( \tau_{ij}(\bar{z}) \), equation (6) is equivalent to solving the following linear system,

\[
\mathbf{A} \bar{z} = \bar{T},
\] (6)

where \( \bar{T} \) is a vector whose components are the maximum correlation time \( T_{ij} \). Equation (6) is a general formulation of the traditional linear residual statics estimation. When the surface consistent are assumed as well as the residual normal moveout and the subsurface structural variations, equation (6) is the tradition linear system (Taner et al., 1974; Wiggins et al., 1976).

According to the observations from the previous section, the stacking-power function is separable. Optimization of this high-dimensional function can be reduced to a series of one-dimensional optimization. The traveltime picking in the conventional statics estimation is, in effect, the procedure of these one-dimensional optimization. For the three-trace example, this approach is equivalent to isolating the three dominant stripes (horizontal, vertical and diagonal) in Figure 1, and looking for the intersection of them, which is at the origin for this example.

This approach is the most commonly used in industry for residual statics estimation due to its computational efficiency. However, this is an over-simplified approach. The only information used from the seismic data is the time corresponding to the maximum peak of correlations. Results of statics estimation rely entirely on the time-pickings \( T_{ij} \). When seismic data are contaminated by severe noise and large statics, time-picking becomes difficult and ambiguous. If wrong peaks are picked from the correlation function, cycle skips would occur and seismic events would be misaligned.

For problems with severe noise and large statics, Ronen and Claerbout (1985) proposed using equation (3) as the objective function for the optimization. However, global searches, such as simulated annealing (SA) and genetic algorithms (GAs), are generally needed for searching the global maximum of stacking-power functions (Rothman, 1985; Rothman, 1986; Smith et al., 1992). These global searches require intensive computation, and the behavior is unpredictable.

Envelope approach of the stack-power function

Looking for a compromise between using “too little” (linearized approach) and “too much” (global optimization) information from correlations, we seek ways of using “partial” information given by correlation function. Since the complexity of stacking powers is mostly due to the complexity of correlation functions, suppressing high-frequency components of the oscillatory correlations will result in a complexity-reduced stacking-power function. For obtaining smooth approximation of a signal, we can have the following three options:

- applying low-pass filtering to the signal, including using multi-resolution analysis (Deng, 1998). This approximation is to cut off high-frequency components, and preserving low-frequency components;
- replacing the signal with a slowly varying one by using Hilbert transform; this approximation is, in effect, to shift the spectrum to be centered at 0 frequency by removing the carrier frequency of the signal;
- interpolating positive peaks of the signal, formulating a smooth signal by preserving large positive information.

Using high-cut filter could be dangerous when the correlation is narrow-band or frequency band of noise is lower than that of the signal. On the other hand, computing the envelope by Hilbert transform is heavily influenced by large negative values in correlations. As we see from the previous section, these values are of little interest to us in residual statics problems. We choose the last option for achieving a simplified objective function.

Finding all positive peaks of a one-dimensional correlation function, we can obtain a slowly varying function by interpolating these peak points. This function represents global feature of large correlation values; we refer this smooth approximation of a one-dimensional signal as the envelope. Using the envelope functions, we define a similar function as the stacking power,

\[
\hat{F}(\bar{\tau}) = \sum_{i=0}^{N-1} \sum_{j=m+1}^{N-1} [\hat{\Phi}_{ij}(\tau_j - \tau_i)],
\] (7)

where function \( \hat{\Phi}_{ij}(\tau) \) is the envelope of correlation \( \Phi_{ij}(\tau) \). The function in equation (7) uses “partial” information from each correlation functions; that is this function ignores all information given by correlations except for all positive peaks. Figure 2 shows a correlation function of two identical traces and its envelope, which is obtained by using the cubic-convolution interpolation scheme (Keys, 1981; Keys & Pann, 1993). Envelope functions contain fewer peaks than does correlation function themselves. Therefore, one expect that the summation of these envelope functions, equation (7), contains fewer local maxima than does the stacking-power function, i.e. the complexity of this function is reduced from that of the original. We call this function reduced stacking-power function.

In order for equation (7) to be a useful objective function, it needs to be a qualified measure for the stack-
Figure 2. A correlation function between two identical seismic traces and its envelope function.

Figure 3. The correlation function, \( \exp\left(-\frac{(x/10)^2}{100}\right) \cos(x) \) and its envelope are shown on the left. The two-dimensional stacking-power function \( F(\tau_1, \tau_2) \) formed by this correlation is shown on the right.

Figure 4. The contour plot of the interpolated function from positive maxima of the original stacking-power function \( F(\tau_1, \tau_2) \) is shown on the left. And the contour plot of the reduced stacking-power function is shown on the right.

Optimizing the reduced stacking power

Conjecture 1. Reduced stacking-power functions have their complexity reduced from the stacking-power functions; they contain fewer local maxima than do stacking-power functions.

Conjecture 2. Reduced stacking-power functions maintains the major global/local maxima from those of the corresponding original stacking-power functions.

Conjecture 3. Reduced stacking-power maintains the global feature of its corresponding stacking-power function.

Figure 3 shows a narrow-band correlation function,

\[
\Phi(\tau) = \exp\left(-\frac{x^2}{100}\right) \cos(x),
\]

and its envelope. For simplicity in this test, the envelope function is obtained by linearly interpolating local peaks of the correlation function, both of them are shown on the left of Figure 3. Using equation (8) as the correlation functions for the three-trace alignment problem, the stacking power function \( F(\tau_1, \tau_2) \) is shown on the right of Figure 3. Figure 4 shows the corresponding reduced stacking-power function the right. The left of Figure 4 shows the interpolated contours from all positive maxima of the stacking-power function in Figure 3. It can be seen that these two plots have the same global feature.

Surface consistent reduced stacking-power function

For most residual statics problems, it is often assumed that for waves traveling through the near-surface, only those with approximately vertical raypaths with respect the interface can be transmitted to deep subsurface structure, therefore can be recorded in our reflection seismic data. Although not being always true, it is a good enough assumption for many practical problems. Therefore, residual statics are surface consistent, that is each source or receiver contributes the same statics shift to all traces it involves, regardless of the ray paths connected with it. Suppose the ith shot is fired at time \( t = 0 \) and let the data recorded at the jth receiver, after offset-dependent moveout is removed, be denoted by \( d_{ij} \). Let \( f_{ij}(t) \) be the same data but without noise \( n_{ij} \) and without static shifts. Ignoring the lateral variation of the sub-surface structure and residual moveout, the recorded trace can be written as

\[
d_{ij}(t) = f_{ij}(t - s_i - r_j) + n_{ij}(t),
\]
where \( s_i, \tau_j \) are unknown static-shifts need to be estimated from seismic traces. In equation (9), source and receiver statics generally can not be recovered by using the conventional linearized approximation unless both statics and noise contamination to the data are not severe.

In solving residual statics problems, seismic data are often sorted into a common-midpoint (CMP) domain; equation (3) is in a form of

\[
F(\vec{s}, \vec{\tau}) = \sum_y \sum_{h_1 \neq h_2} \Phi_{h_1, h_2}(\tau(\vec{s}, \vec{\tau})),
\]

where \( \vec{s}, \vec{\tau} \) are shot and receiver estimated statics vectors, \( y \) denotes the midpoint index, \( t \) is the time index over a specified window, and \( h \) is the offset index, and \( \Phi_{h_1, h_2}(\tau) \) is the cross-correlation between traces \( d_{y,h_1}(t) \) and \( d_{y,h_2}(t) \) and \( \tau(\vec{s}, \vec{\tau}) \) is a linear function of source- and receiver-statics. When ignoring the residual normal-moveout and time-shift caused by subsurface structure, function \( \tau(\vec{s}, \vec{\tau}) \) is determined by the recording geometry,

\[
\tau = s_i(y, h_1) + \tau_j(y, h_1) - s_i(y, h_2) - \tau_j(y, h_2),
\]

and \( i(y, h) \) and \( j(y, h) \) are the source and receiver indices for midpoint \( y \) and offset \( h \), respectively. Therefore, solving a residual statics problem is a nonlinear optimization where we look for \( \vec{s}, \vec{\tau} \) that

\[
\text{max}_{\vec{s}, \vec{\tau}} F(\vec{s}, \vec{\tau}).
\]

Following the same analysis as the previous section, the reduced stacking-power function with the surface-consistent constraints can be defined as,

\[
\hat{F}(\vec{s}, \vec{\tau}) = \sum_y \sum_{h_1 \neq h_2} \hat{\Phi}_{h_1, h_2}(\tau(\vec{s}, \vec{\tau})),
\]

where \( \hat{\Phi}_{h_1, h_2}(\tau) \) is the envelope of correlation functions of \( \Phi_{h_1, h_2}(\tau) \). To keep the smoothness and prevent the overshoot of the interpolated envelope functions, the cubic convolution interpolation scheme originally developed by Keys (1981; 1993) are used throughout the project.

A new residual-statics estimation algorithm

Following the analysis and conjectures of the previous section, we design an algorithm for estimating residual statics for NMO corrected, CMP data contaminated by surface-consistent statics as follows,

Algorithm 1. (Residual Statics using Reduced Stacking Power)

(i) (Pre-compute and store:) For all cross-correlation \( \Phi(\tau) \), find positive-peaks on all cross-correlation between traces for all CMP. Use cubic convolution (Keys & Pann, 1993) to form envelope of the sequences \( \Phi(\tau) \).

(ii) (Gross estimation of large magnitude statics:) local optimization on the reduced stacking-power function \( \hat{F}(\vec{s}, \vec{\tau}) \) for the initial model \( \vec{s}_0, \vec{\tau}_0 \); correct data for the estimated statics.

(iii) (Refinement:) After large magnitude statics are corrected in step (ii), results can be refined by optimizing the stacking-power function using local optimization with 0 initial guess, or the conventional linearized statics estimation. Depending upon results from step (ii), this step may not be necessary.

In case of loss of resolution of the reduced stacking-power function, the last step of refinement is important for further improvement of the stacking quality.
Figure 7. Contour plots of stack powers as a function of static time shift at the 10th source and the 20th receiver points, with all other statics held to be correct. The left figure shows that of the original stacking-power function, while the right figure shows that of the reduced stacking-power function (using envelopes).

Figure 9. Comparison of the added statics and estimated statics. Horizontal axis is the index of unknowns: 0–19 are sources and 20–54 are receivers. The solid-thin line shows the added statics, while the thick-dashed line shows the estimated statics.

Figure 6 and 7 show behavior of the original stacking power function and its reduced form from slices. Similar to Figure 4, the original stacking power appears to have many local maxima, which are scattered regularly on the 2-D slices. On the other hand, complexity of the reduced stacking power function is largely reduced observing from these 2-D slices.

When the above dataset are contaminated by severe noise and large source and receiver statics, the traces are no longer identical. The left of Figure 8 shows the stacked section of this data set contaminated by surface-consistent statics and 50% additive random noise (signal-to-noise ratio is 2). The noise is band-limited Gaussian noise which has the same bandwidth as the original signal. The added statics are generated by a uniformly distributed random number generator between $[-100, 100]$ ms. The added and estimated statics are shown in Figure 9. The estimated statics in general agree well with the added statics; the observed large error at beginning and end of receiver statics are caused by the edge effect — low folds. The stacked section after estimated residual statics are corrected is shown on the right of Figure 8, where horizontal events show up except those at the beginning and end of CMPs.

**Synthetic Examples**

Figure 5 shows the stacking chart of the synthetic data set used for testing the residual statics estimation algorithm. This data set has 20 shots, 35 distinctive receivers and 320 traces; all traces are generated by copying and shifting of a field trace. Correlations of these traces are, therefore, shifted auto-correlation of the field trace. This auto-correlation function and its envelope were shown in Figure 2. For this problem, the stacking-power function and its reduced form both have dimension of 55. When no statics are added to the data, the global maximum of the objective function is at the origin of the model space (all statics are zero).

Not being able to visualize the 55-dimensional objective function in this problem, we can observe slices where we hold all unknowns being correct except two.

**Alberta foothill data set**

We now test Algorithm 1 on a set of field data from Alberta foothill. This data set is contaminated by large statics which cause the cycle skips of strong events in stacked sections. Figure 10 shows stacked sections, where residual statics are corrected using ProMax. A large cycle skip can be observed at about CMP 420.

This data set has an average of 60 folds. We choose 50-CMP gathers from this dataset, which was obtained by 88 sources and 88 receivers. Therefore, this statics
Figure 10. Alberta foothill data: output of statics-corrected stacked sections from ProMax. An obvious cycle skip shows at the circled area.

Figure 11. Left figure shows 50 stacked CMPs from the Alberta foothill data set before residual statics are corrected. The right figure shows the same stacked section after residual statics are corrected using the proposed algorithm.
Figure 12. Paradox basin data: output of statics-corrected stacked sections from ProMax. An obvious cycle skip shows at the circled area.

Figure 13. A stacked section from Paradox data before residual statics are corrected is shown on the left. And the stacked Paradox data after statics are estimated and corrected using the proposed algorithm is shown on the right.
estimation problem is an optimization of 176 dimensions. The stacked section for CMP 401–450 are shown on the left of Figure 11. Discontinuity of events are severe in this section. Especially, events between time 1.5–2.0 s show an obvious cycle skip. The statics-corrected stacked section using the proposed residual statics correction algorithm is shown on the right of Figure 11.

The difference of event-shapes between Figures 10 and 11 is caused by an elevation correction while doing stacking in ProMax. For solving such a 176-dimensional global optimization problem, running time is around 48 minutes on an SGI Indy.

Paradox basin data set

Another data set was tested on the Algorithm 1. Figure 12 shows a ProMax output of statics-corrected stacked section. This output shows a strong cycle skip at around CMP 2110, especially for the event of 0.8 s.

This data set has an average fold of 15, we processed 100 CMPs with 35 sources and 146 receivers. A stacked section containing 100 CMPs before the residual statics are corrected is shown on the left of Figure 13. The residual statics estimation is an optimization problem of 181 dimensions, for which we used a non-linear Conjugate Gradient algorithm. The right figure in 13 shows the stacked section after residual statics are estimated and corrected using the proposed algorithm. The strong event at around 0.8 s are almost continuous across the section and some stratigraphic details can be seen.

However, continuity of the 0.8 s is not satisfactory at around CMP 2110. This indicates the loss of resolution in reduced stacking-power function. The second iteration of optimizing the stacking-power function is needed for refinement. Figure 14 shows the result of this second iteration using non-linear conjugate-gradient search with 0 initial guess. The lateral correlation of the strong event at around 0.8 s has notably improved.

Each iteration of the statics estimation for this data set is about 25 minutes on an SGI Indy.

Conclusion & Discussion

Residual statics estimation is an important stage in seismic data processing. As this procedure is designed for improving high quality stacks, the problem of residual statics estimation can be formulated as an optimization problem which looks for time-shifts among traces (or sources and receivers) that gives the maximum value of the sum of squares (power) of the stacked section. We have studied both linear and non-linear conventional approaches for residual statics estimation. We notice that the conventional linear approaches have over-simplified the problem by using one traveltime datum for each correlation of two traces, and conventional non-linear approach had complicated the problem by using all information given by such correlations.

By studying the nature of stacking-power function, we recognize that this function can be simplified by using partial information given by correlations. Because of the special objective in this non-linear optimization problem, only positive peaks are interesting to us. In this paper, we propose using the envelope of correlations to form a reduced stacking-power function as the objective function for optimization. We also observe that the proposed objective function has fewer local maxima than its corresponding original stacking power, and this complexity-reduced function approximates the global structure of the original.

Using the surface-consistent constraint, we have applied the algorithm to a set of synthetic data which is generated by one field recorded trace, but contamin-
Optimizing the reduced stacking power

References


Acknowledgment

This project was initially supported by the summer program for university students at Mobil Exploration & Producing Technical Center during the summer of 1995. We appreciate Dave Hinkley and Rachel Ho for their generous help throughout the project. Our thanks also go to Bob Evans and Bob Peterson for their support and permission for publishing this work. We are also grateful to owners of the two dataset, Thistle Inc. and Mobil Canada, for their generous offer of the data and their permission for release.

The first author also appreciates her adviser, Dr. John Scales, who has pushed the direction of studying and simplifying objective functions in optimization problems. She was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines and Oak Ridge National Laboratory.

ated by large statics (up to ±100 ms) and 50% band-limited Gaussian noise. The frequency band of the noise is the same as that of the signal. We demonstrate that Algorithm 1 deals well with data contaminated by such noise and large statics. We have also applied the new algorithm to two sets of field data, where ProMax residual-correction output show cycle-skips on stacked sections. The new algorithm was able to resolve the residual statics correctly.

The effectiveness of residual statics algorithms usually depend on the redundancy of data; CMPs with low-fold result in poor estimated statics. Long-wavelength components of statics are generally either in the null space or are poorly constrained by residual-statics algorithms. We do not address above issues.

The basic assumption of the stacking-power criterion is that offset-dependent moveout has been correctly adjusted before the optimization. However, this assumption may not be valid if the sub-surface structure is complex or velocity models are poor. In addition, the assumption of surface consistency may not be appropriate in many cases. The robustness of the proposed algorithm needs to be studied when these conditions are not ideal.

There are several alternatives in the implementation of the Algorithm 1. For saving memory used to hold all correlation values, we could only store the picked maxima for all correlations, and they can be interpolated on the fly while evaluating the reduced stacking-power function or its derivative. Another approach is to use coarser grids of the optimization due to smoothness of envelope functions.
APPENDIX A: Maxima of stacking-power functions

For a generic stacking-power function

$$F(\vec{x}) = \sum_i \sum_{j \neq i} \Phi_{ij}(\tau_{ij}(\vec{x})), \quad (A1)$$

each component of its gradient vector can be written as

$$\frac{\partial F(\vec{x})}{\partial x_k} = \sum_i \sum_{j \neq i} \Phi_{ij}'(\tau_{ij}(\vec{x})) \frac{\partial \tau_{ij}}{\partial x_k}, \quad \text{for } \forall k. \quad (A2)$$

Let us study the relationship between the maxima of the stacking-power function and those of the correlation functions. We can prove the following theorem.

**Theorem 1.** If there exists $\vec{x}_0$ so that

$$\nabla \Phi_{ij}(\vec{x}) |_{\vec{x}=\vec{x}_0} = 0, \quad \forall i, j, \quad (A3)$$

then,

$$\nabla F(\vec{x}) |_{\vec{x}=\vec{x}_0} = 0. \quad (A4)$$

Theorem 1 states that if a point in the domain is the maximum of all correlation functions simultaneously, then it must be a maximum of the stacking-power function. This theorem can be proved as follows.

Since $\nabla \Phi_{ij}(\vec{x}) |_{\vec{x}=\vec{x}_0} = 0$, for $\forall i, j$, we can also write that for each $\Phi_{ij}$,

$$\frac{\partial \Phi_{ij}(\vec{x})}{\partial x_k} |_{\vec{x}=\vec{x}_0} = \Phi_{ij}'(\vec{x}) \frac{\partial \tau_{ij}}{\partial x_k} = 0, \quad \text{for } \forall k. \quad (A5)$$

From equation (A2), we have that $\nabla F(\vec{x}) |_{\vec{x}=\vec{x}_0} = 0$. Therefore, $\vec{x}_0$ is a maximum point of the stacking-power function equation (A1). □

However, the converse of Theorem 1 is not true. That is, not all maxima of the stacking-power function are the maxima of correlation functions. It is easy to see that if equation (A4) is true, equation (A3) is not necessarily true.

Therefore, while oscillations in seismic data cause global/local extrema of stacking-power functions, they are not the only contributing factor. There exist local maxima on stacking-power functions that are not caused by stationary points of each correlations. These spurious maxima are those we are not interested in residual statics estimation.

To demonstrate this point, Figure A1 shows a 2-D projection of a 55-dimensional stacking-power function in which all parameters are fixed except two components. The left figure shows the slice when the components are fixed at correct values. In this case, the apparent maxima are regularly distributed on the function surface. On the other hand, the right figure shows the projection when the fixed components are chosen at random. In this case, the structure is evidently more irregular.
Seismic waveform nonlinear inversion in stratified elastic media

Wences Gouveia
Department of Geophysics
Center for Wave Phenomena
Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
In this work I estimate impedance contrasts of a layered elastic medium using a nonlinear inversion procedure. The inversion procedure adopts the Bayesian methodology in which the “solution” of the inverse problem is characterized by an a posteriori probability density function on the space of models m. This function consists of the product of two probability density functions. The first, the likelihood function L(m), defines what it means for a model to fit the data. The second, the a priori probability density function ρ(m), incorporates any available prior information on the underlying model.

I carry on the inversion assuming that all the uncertainties of the problem are Gaussian. As a consequence, the problem reduces to a nonlinear weighted least-squares with a penalty function. The weighting and the influence of the penalty function on the inverse solution are determined by the covariances matrices of those probability functions. I estimate the (data) covariance matrix of the likelihood function directly from the noise in the data, and the (model) covariance matrix of the a priori probability function from a set of “synthetic” well log measurements. Results are illustrated for the inversion of a synthetic data generated for a 50-layer elastic model, in the presence of band-limited noise. The uncertainties of this inverse calculation are formally assessed by analyzing the covariance matrix of the a posteriori distribution.

Key words: Inversion, Bayesian Inference, Reflectivity, Model Covariances

Introduction
The nonlinear inversion of elastic seismic waveforms is a procedure that relies on the minimization of a misfit function that quantifies the difference between observed data and synthetic data, generated by a suitable modeling algorithm. Essentially the model parameters are systematically perturbed by a nonlinear optimization algorithm until a pre-specified misfit threshold is reached. Several studies have been done in the literature addressing this problem. The reference list that follows includes some of those works. Tarantola (1986) pioneered this approach, in the field of exploration seismology. Mora (1987) addressed some of the practical complications that arise with this method due to lack of proper coverage of the subsurface. He also showed how the solution of the inverse problem can be improved when transmitted data (Vertical Seismic Profiles) were used in conjunction with surface seismic data. Scales et al. (1988) discussed the statistical meaning of L2 norms and studied robust inverse methods in which the data misfit function was built with other norms than the L2 norm. Crase et al. (1989) performed a similar study where he illustrated with numerical examples the lack of robustness of the L2 norm when compared to the L1, secant and Cauchy norms when inverting a noisy data set. Pan et al. (1994) applied this technique for lithology estimation of a layered target zone from compressional reflection data.

The possibility of being able to systematically incorporate diverse a priori information and make quantitative statements of resolution is the main driving force of this technique. However the non-uniqueness of the solution, the high computational cost and sometimes the lack of
robustness with respect to noise are all significant to its application on a more routine basis. Here, noise is defined as anything not taken into account by the forward modeling procedure. Examples are random events in the data (observational errors) or features of the subsurface not considered by the modeling algorithm (modeling errors).

With respect to reducing the space of possible solutions for a given inverse problem, prior information has to be incorporated into the procedure. This can be done in different ways. One can make assumptions about the underlying model, for instance about smoothness. This is a one of the possible approaches in the Tikhonov regularization technique (Tikhonov and Arsenin, 1977). The degree of smoothness in the model is controlled by the so-called regularization parameter. In this case, the "solution" to the inverse problem is chosen to be the model associated with the largest regularization parameter that still fits the observed data up to a pre-specified threshold. This model represents the smoothest one that still fits the data.

The Bayesian approach for seismic inversion represents another way to incorporate prior information. In this technique (Jaynes, 1968; Tarantola; 1987 and others), the solution of the inverse problem is formulated as a probability distribution over the space of models, namely the a posteriori probability function. This function is the product of two probability density functions. The likelihood function that measures the misfit between observed and synthetic data, and the a priori probability function, that implements a suitable statistical model for the model parameters. For example, if it is reasonable to assume that the model parameters can be represented by a Gaussian random process, the a priori probability function adopts the form of a multidimensional Gaussian probability density function. This function is a formal mechanism for incorporation of information about the model parameters into the inversion procedure.

In this work I study the inverse solution of the nonlinear elastic amplitude inverse problem in the presence of noise. I show preliminary results in which a synthetic seismic data set, a shot gather generated for a 50-layer elastic model, is inverted in the presence of noise. A smoothed version of the velocity (for P- and S-waves) and density profiles are used as a initial guess. The uncertainty of the inverse results are assessed by inspecting the covariance matrix of the a posteriori probability distribution. This procedure provides error bars on the estimates of impedances and densities.

This work is organized as follows. First, I introduce the modeling procedure utilized in the inversion procedure. Next, the inverse problem is formulated in the framework of Bayesian estimation, in which different model parameterizations are presented and discussed. Moreover, the data misfit function is studied, by displaying its hyperplanes. The objective of this analysis is to provide some insight on the extent of different models that can be considered "solutions" to the inverse problem. Following that, I show numerical examples of the inversion of a 50-layer elastic model and evaluate the uncertainties of the solution. Since the data was corrupted with Gaussian random noise, the likelihood function would be a Gaussian distribution if the forward model were linear, not the case in this work. To compute the uncertainties of the inverse solution I make a Gaussian approximation of this function about the model that maximizes the a posteriori probability density.

The forward modeling procedure

In this work I will be addressing the nonlinear inversion of seismic waveform data under the assumption that the underlying model is elastic, isotropic and laterally homogeneous. Among the available modeling techniques, the reflectivity method (Fuchs and Müller, 1971 and Kennet, 1983) seems to provide the most adequate algorithm for this purpose. The main reasons for this are the following.

• The reflectivity method provides an analytic solution for the displacement field. This allows the analytic computation of the Frechet derivatives, which components are the derivatives of the forward modeling operator with respect to the model parameters.
• The reflectivity method is an offset-frequency-domain approach, and as a consequence allows a high degree of parallelism which can be exploited in the inversion algorithm.
• In recent work, Widmaier et al. (1995) proposed a statistical approach for the computation of the transmissivity response of a stack of elastic layers for a slant-incident plane wave. Since, in the reflectivity method the seismic source is decomposed into plane waves, one might consider the possibility of replacing the overburden above a given target reflector by such a transmissivity function for each one of the plane waves. Although this is not attempted in the present work, it might provide an interesting correction for the overburden.

In this section I will introduce the reflectivity method in a schematic way, keeping the usage of mathematical expressions to a minimum. A detailed description of this procedure can be found in the excellent tutorial given in Müller (1985). The geometry of the problem in cylindrical coordinates is shown in Figure 1. The elastic isotropic layered medium is completely characterized by
the P-wave (α) and S-wave (β) velocities, the density (ρ) and the thickness (dz) for each layer. Consider for a moment a homogeneous medium above and below the source depth z_s. In this case, the upgoing (z < z_s) and downgoing (z > z_s) displacement potentials due to a point force located at depth z_s with orientation [F_x, F_y, F_z] in Cartesian coordinates are given by (Müller, 1985)

For z > z_s :
\[ \Phi^e_s = \frac{1}{4\pi\rho_m} \int_{0}^{\infty} (\epsilon_1 A_{s1} J_0(\omega r) + \epsilon_2 A_{s2} J_1(\omega r)) \exp[-i\omega z_s(z - z_s)] du \]
\[ \Psi^e_s = \frac{1}{4\pi\rho_m} \int_{0}^{\infty} \frac{i}{j\omega} (\epsilon_1 C_{s1} J_0(\omega r) + \epsilon_2 C_{s2} J_1(\omega r)) \exp[-i\omega z_s(z - z_s)] du \]
\[ \chi^e_s = \frac{1}{4\pi\rho_m} \int_{0}^{\infty} \eta F_s C_s J_1(\omega r) \exp[-i\omega z_s(z - z_s)] du \]
\[ A_{s1} = u, \quad A_{s2} = \frac{u}{i\omega z_s}, \quad C_{s1} = \frac{u}{b_m}, \quad C_{s2} = \frac{i}{u} \]
\[ F_s = \frac{i}{\beta b_m}, \]
(1)

For z < z_s :
\[ \Phi^e_s = \frac{1}{4\pi\rho_m} \int_{0}^{\infty} (\epsilon_1 B_{s1} J_0(\omega r) + \epsilon_2 B_{s2} J_1(\omega r)) \exp[-i\omega z_s(z - z_s)] du \]
\[ \Psi^e_s = \frac{1}{4\pi\rho_m} \int_{0}^{\infty} \frac{1}{j\omega} (\epsilon_1 D_{s1} J_0(\omega r) + \epsilon_2 D_{s2} J_1(\omega r)) \exp[-i\omega z_s(z - z_s)] du \]
\[ \chi^e_s = \frac{1}{4\pi\rho_m} \int_{0}^{\infty} \eta F_s J_1(\omega r) \exp[-i\omega z_s(z - z_s)] du \]
\[ B_{s1} = -u, \quad B_{s2} = \frac{u}{i\omega z_s}, \quad D_{s1} = \frac{u}{b_m}, \quad D_{s2} = \frac{i}{u} \]
\[ F_s = \frac{i}{\beta b_m}. \]
(2)

Here, \( \Phi \), \( \Psi \), and \( \chi \) are the scalar potentials in the frequency (\( \omega \)) domain for P-waves, SV-waves and SH-waves, respectively. \( u \) is the horizontal slowness, \( a_m \) and \( b_m \) are the vertical slownesses of layer \( m \), \( \epsilon_1 = F_3, \epsilon_2 = F_1 \cos \phi + F_2 \sin \phi \) and \( \eta = -F_1 \sin \phi + F_2 \cos \phi \), where \( \phi \) is the azimuthal angle. \( J_0(\cdot) \) and \( J_1(\cdot) \) are the Bessel functions of the first kind of zero-th and first-order, respectively. For simplicity, consider only the P-SV propagation modes. It is convenient to define the following source amplitude vectors
\[ S_{l,2}^e = \begin{bmatrix} A_{l,2} \\ C_{l,2} \end{bmatrix}, \quad S_{l,2}^\ast = \begin{bmatrix} B_{l,2} \\ D_{l,2} \end{bmatrix}. \]
(3)

The synthesis of the wavefield at the surface requires a two-step procedure. First the upgoing wave \( V_{1,2} \) at \( z = z_m \) should be computed. It is not difficult to show (Müller, 1985) that \( V_{1,2} \) are given by
\[ V_{1,2} = (I - R^- R^+)^{-1} (S^e_{1,2} + R^- S^d_{1,2}). \]
(4)

Here, \( R^- \) and \( R^+ \) are the reflectivity matrices for the lower and upper stack of layers, respectively. Such matrices are derived recursively from considerations on only the continuity of the displacement and the traction fields across the interfaces. By the same approach it is possible to compute the transmissivity \( T^+ \) of the upper stack of layers and compute the potential at the surface, which is given by
\[ V_{1,2} = T^+ V_{1,2}. \]
(5)

The far-field displacement field at the surface is directly derived from the potentials (Aki and Richards, 1980),
\[ \begin{bmatrix} u_r \\ u_v \end{bmatrix}(r, \omega) = \omega \sum_{i=1}^{n} \int_{0}^{\infty} J_i^i UV_i^0 du, \]
(6)

Here, \( u_r \) and \( u_v \) represent the radial and vertical components of the displacement field, and the following definitions have been used:
\[ U = \begin{bmatrix} u \\ b_0 \\ a_0 \\ -u \end{bmatrix}, \quad J_i = \begin{bmatrix} -J_i(\omega r) & 0 \\ 0 & iJ_i(\omega r) \end{bmatrix}, \]
(7)

Equation (6) is used to compute the P-SV displacement field in the frequency domain for the inverse problems addressed in this paper. In the algorithm implemented here however, all data misfit computations are done in the time domain.

**The Bayesian Inversion Scheme**

In the Bayesian approach for parameter estimation, the solution of the inverse problem is represented by the \textit{a posteriori} probability density function \( \sigma(m) \) defined on
the space of models. As mentioned before, this probability function consists of the product of two probability functions: the likelihood $L(m)$ and the a priori $\rho(m)$ probability density functions. Consider the particular case where such distributions assume the following forms in a $n$-dimensional model space:

$$L(m) = \kappa_1 \exp \left[ -\frac{1}{2} (g(m) - d_{obs})^T C_D^{-1} (g(m) - d_{obs}) \right],$$

$$\rho(m) = \kappa_2 \exp \left[ -\frac{1}{2} (m - m_0)^T C_M^{-1} (m - m_0) \right].$$ \hspace{1cm} (8)

Those expressions arise in the situation when all the uncertainties of the problem are normally distributed. The likelihood density function would be Gaussian if the forward modeling operator $g(m)$ were linear, which is not the case in this study. In the above expression $\kappa_1$ and $\kappa_2$ are normalization constants. $C_D$ is the data covariance matrix, incorporating the variances and covariances of the noise in the observed data $d_{obs}$. The a priori density function is Gaussian, defined by its mean model $m_0$ and its model covariance matrix $C_M$. This matrix incorporates the 2nd-order statistics of the parameters.

The main objective of the inversion procedure is to find the model (or models) associated with significant maxima of $\sigma(m) \propto L(m) \rho(m)$. Optimization techniques are often used to determine those maximum a posteriori (MAP) estimators. Note that this procedure is equivalent to minimizing of the following objective function:

$$\min_m \Theta(m) = \frac{1}{2} [(g(m) - d_{obs})^T C_D^{-1} (g(m) - d_{obs})$$

$$+ (m - m_0)^T C_M^{-1} (m - m_0)].$$ \hspace{1cm} (9)

At this point the similarity of the minimization problem defined by Equation (9) and the regularized nonlinear least-squares problem defined as

$$\min_m \| C_M^{-1/2} (g(m) - d_{obs}) \|^2 + \lambda \| R (m - m_0) \|^2.$$ \hspace{1cm} (10)

should be clear. In the latter, $R$ is a regularization matrix and $\lambda$ the regularization parameter. Often regularization is used to incorporate smoothness constraints into the inverse problem for stabilization purposes when the data are corrupted by noise. Common choices for $R$ are the first- and second-difference operators (Scales et al., 1990). The regularization parameter $\lambda$ controls the influence of the regularization matrix $R$ in the solution of the inverse problem. This parameter characterizes the trade-off between data misfit and degree of smoothness. As shown in Gouveia (1995) in the Bayesian scheme the use of the regularization technique implies a structure of the model covariance matrix given by

$$C_M = [R^T R]^{-1}.$$ \hspace{1cm} (11)

One of the advantages of the Bayesian approach is the formal assessment of the uncertainties of the inverse solution. It is possible to show that, for a linear forward modeling operator, the a posteriori covariance matrix has the form (Tarantola, 1987)

$$C_M = [G^T C_D^{-1} G + C_M^{-1}]^{-1},$$ \hspace{1cm} (12)

where $G$ is the linear (or linearized) forward modeling operator and $C_M$ its hermitian. Linearizing $g(m)$ about the MAP model, when meaningful, can be used to estimate the uncertainties of the inverse solution in the nonlinear case. This is the approach used in this work to compute the uncertainties of the inverse problem solution.

### Optimization Technique

The objective function in Equation (9) will be minimized by a nonlinear version of the conjugate gradient method (Fletcher, 1987). This method is a local, iterative, optimization procedure that works by updating an iterate $m_n$ in the following way

$$m_{n+1} = m_n + \alpha \phi_n.$$ \hspace{1cm} (13)

Here, $\phi_n$ is a search direction, which can be expressed as a function of the gradient of the objective function $\Theta(m)$ at iterations $n$ and $n-1$, and $\alpha$, the step length. This parameter is determined by a line search procedure (Dennis and Schnabel, 1987). The gradient of $\Theta(m)$ is given by:

$$\nabla g(m) = \Gamma(m) C_D^{-1} (g(m) - d_{obs}) + C_M^{-1} (m - m_0).$$ \hspace{1cm} (14)

Here, $\Gamma(m)$ is the Frechét derivative matrix which components are defined as $\frac{\partial g(m)}{\partial m_i}$. A common procedure to obtain $\Gamma(m)$ is by standard perturbation analysis of the model parameters and subsequent use of the Born approximation. This procedure casts the computation of the Frechét derivatives as a forward modeling problem with special boundary and initial conditions (Tarantola, 1987). This approach was not taken in this work. Here, the Frechét derivatives are computed by analytical differentiation of the displacement field of Equation (6). This yields a recursive algorithm similar to, but more expensive than, the forward modeling procedure. In theory, the approach used here is more accurate since no approximations are made in the computation of the derivatives.

### Model Parameterization

Several studies have been done to understand the information available in seismic waveform data with respect to subsurface elastic parameters (Tarantola, 1987; Jannane et al. 1989; and others). The following model parameterizations were used in those studies:
• Lamé parameters $\lambda$ and $\mu$, and mass density;
• $P$-wave and $S$-wave velocities, and mass density;
• $P$-wave and $S$-wave impedances, and mass density, and
• $P$-wave impedance, Poisson’s ratio, and mass density.

The basic conclusion is that seismic data contain information on the long wavelength of velocities and on short wavelength of impedances. The fact that long wavelength of velocities influences the travel times of the propagating waves has as a direct consequence a non-quadratic objective function for the parameters describing the background velocities. Gradient-based methods provide little hope for solving such problems, and global optimization techniques might be required. However the objective function for the parameters describing the short wavelengths of impedances (and also velocities) is approximately quadratic and, in principle, could be tackled by gradient-based optimization algorithms.

Another important issue is the coupling among different parameters. Ideally, in an inverse problem, the model parameters should be independent of each other. In a recent work, Debinski and Tarantola (1995) studied the level of coupling among the model parameters in the waveform seismic inverse problem, for different choices of parameterization. Their main conclusion is that it is possible to interpret seismic amplitudes equally well with small values of velocity and large values of density, or with large values of velocity and small values of density. In other words these parameters trade off. The level of coupling is considerably lower when $P$-wave and $S$-wave impedances and density, or $P$-wave impedance, Poisson’s ratio and density are used as model parameters.

The results obtained in Debinski and Tarantola (1995) were confirmed by numerical simulations done in this work. Consider the model illustrated in Figure 2. The unknowns are the elastic parameters of the eight layers delimited by the arrows shown in the figure. The vertical component of the displacement due to a vertical point force generated for this model is illustrated in Figure 3.

This data set was inverted considering just the likelihood function $L(\mathbf{m})$ (i.e. for an uniform $\rho(\mathbf{m})$), using two distinct parameterizations. Figure 4 shows the MAP solution when the $P$-wave impedance, $S$-wave impedance and density were used as model parameters, and Figure 5 shows the MAP model obtained when the parameterization was done in terms of $P$-wave velocity, $S$-wave velocity and density. These results were obtained with the same initial guess, an smoothed version of the true model parameters. Although the norm of the residual (shown in Figure 6) of the final solution is smaller when velocities and densities are used as model parameters, the MAP model for this case disagrees more with the true model than when impedences and densities are used to parameterize the model. Apparently, the trade-off (coupling) between $P$- and $S$-wave velocities is larger than the coupling between $P$- and $S$-wave impedances. Clearly, this is a manifestation of the null-space of the problem. In both examples the density was poorly resolved, with slightly better results when impedances were used as model parameters. When the MAP solution shown in Figure 5 is used to generate $P$- and $S$-wave impedance profiles, I obtained the result illustrated in Figure 7. Such profiles show a better agreement with the true model than do the velocity profiles illustrated in Figure 5. However, the agreement is not as good as the one of the MAP model.
obtained when impedances are used to parameterize the model space. This suggests that the (quasi) "uncoupling" between $P$- and $S$-wave impedances can just be partially recovered if the MAP model obtained from an inversion procedure carried on in a velocity-density space is used to generate the impedance profiles.

These results led to the choice of the parameterization of the model space in terms of impedances and density. Note that the computational cost of the algorithm is about the same as the one when velocities and density are used as model parameters. As a final remark, since a gradient-based optimization method will be used for the minimization of the objective function, no attempt will be made to estimate the long-wavelength component of the model space. This information, assumed to be available from velocity analysis or from travel time tomo-

Figure 4. MAP model when $P$-wave impedance, $S$-wave impedance and density are used as model parameters.

Figure 5. MAP model when $P$-wave velocity, $S$-wave velocity and density are used as model parameters.

graphy, will be incorporated into the inverse problem via the initial model $\mathbf{m}_0$.

The Objective Function

In this section, the convexity of the objective function is studied in order to provide some insight on its resolving power. For the model illustrated in Figure 8 (the same model was used in Pan et al. (1994)), hyperplanes of the objective function are plotted in Figure 9. Here, just the two layers shown by the arrows in Figure 8 are considered unknown in the inverse problem. Each hyperplane in Figure 9 was built by changing a pair of parameters (one from each layer) systematically and keeping the other constants at their correct values. Due to small $S$-wave velocity contrast in this model, the hyperplane as-
Waveform inversion

Figure 6. Data misfit as a function of the conjugate-gradient iteration number.

Figure 7. Impedance profiles derived from the velocity and density profiles shown in Figure 5.

associated with the S-wave velocities is basically flat. The other two hyperplanes, associated with P-wave velocities and density, exhibit long valleys characterizing the large range of models that are related to similar values of the objective function. Such models form the null-space of the likelihood function. For larger size problems it can be expected that the size of the null space will increase. Therefore, it is necessary to use a priori information in order to differentiate among models that fit the data up to a pre-specified threshold. In this work this a priori information is provided by the probability density $p(m)$. This is the subject of the following section.

Figure 8. Elastic layered model associated with the hyperplanes displayed in Figure 9.

Amplitude Bayesian Inversion with a priori Information

Methodology

As mentioned before, the a priori information on the underlying model is incorporated into the Bayesian framework via the probability density function $p(m)$. In this work it will be assumed that $p(m)$ can be modeled by a Gaussian distribution. Such assumption implies that, up to the 2nd-order moments (mean and covariances) of the model parameters will be incorporated as information to the inverse problem. The mean, or most likely a priori model, will also correspond to the initial guess of the inverse problem, since it should represent the best model available to the problem at hand. With respect to the model covariance matrix $C_M$, very often assumptions about its structure are usually introduced to make the solution technique more tractable. Here, I assume that a realization of the model parameters is available in the form of a set of well-log measurements. Considering that the well log is representative of such a realization, and also assuming it is an ergodic, stationary random process (at least within some depth window), the model covariance matrix can be estimated by the following expression (Priestley, 1989):

$$C_M[i,j] = \frac{1}{n} \sum_{k=1}^{n} m_k m_{k-\tau}. \quad (15)$$

Here, $\tau = |i - j|$, and $m_i$ represents the value of the P-wave impedance, S-wave impedance or density at the depth $i \, dz$, where $dz$ is the depth sampling. Notice that the covariance matrix is a function of just the lag of the depth series, a property of a stationary process. Ergodi-
Figure 9. Hyperplanes of the objective function. The innermost contour level represents a data misfit of 0.1. All plots are normalised to amplitude 1.

Figure 10. 50-layer elastic model. The medium for depths less than 1 km is homogeneous with P-wave velocity, S-wave velocity and density equal to 3 km/s, 2 km/s and 1 gm/cm³, respectively.

Figure 11. Model covariance matrices for P-wave impedance (a), S-wave impedance (b) and density (c). The impedances were obtained from the velocity and density profiles displayed in Figure 10.

city is required when just one well log is available; as a consequence the covariance matrix $C_{ii}$ has to be estimated as a depth average of this single well log. Ergodicity means that ensemble averaging (i.e., the averaging of many well logs) for the estimation of the model covariance can be replaced by spatial averaging of a single
Figure 12. Vertical component of the displacements generated for the elastic model of Figure 10.

Figure 13. Vertical component of the displacements generated for the elastic model of Figure 10 corrupted by band-limited noise.

Figure 14. Noise seismogram (a) and the data covariance matrix (b). For visualization purposes the data covariance matrix is just plotted for the first half second of data.

realization. One possible future direction of this work is the estimation of the model covariance matrix when several well logs are available. In this case the ergodicity assumption can be relaxed.

As an example, consider the impedance and density profiles shown in Figure 10. Those are to be considered synthetic well logs, built by sampling a multi-dimensional Gaussian distribution. This figure displays the target zone for the inversion. The model covariance matrices computed for these profiles are shown in Figure 11. Using the reflectivity method, I generated a synthetic data set for the elastic model in Figure 10. These data are shown in Figure 12. After addition of Gaussian noise (band-limited to the same frequency of the signal) these data are shown in Figure 13. One seismic trace \( s(t) \) containing only noise was used to construct the data covariance matrix \( C_D \) of the likelihood density function. This matrix can be estimated in the same way as the model covariance matrix. Again, under the assumptions of stationarity and ergodicity, this estimator is given by:

\[
C_D[i,j] = \frac{1}{n} \sum_{k=1}^{n} d_k d_{k+i-j}.
\]

Here, \( \tau = |i-j| \), and \( n \) is the number of data points. \( d_k \) is the \( k \)-th data sample recorded at time \( k \Delta t \), where \( \Delta t \) is the time sampling interval. Figure 14 illustrates the noise seismogram and its associated data covariance matrix computed using the above equation.

Results

I inverted the data set in Figure 13 to estimate \( P \)- and \( S \)-wave impedances and mass density for the 50 layers shown in Figure 10. This procedure was based on the minimization of the objective function \( \Theta(m) \) defined in Equation (9). The data and model covariance matrices used in the objective function are shown in Figures 11 and 14, respectively. The initial guess used here was the true model smoothed using a 100 m running average. The MAP model that fit the data up to one standard deviation of the noise (the \( \chi^2 \) criterion), along with the initial guess and the true model, are shown in Figure 15. The error bars, computed from the covariance matrix of the a posteriori probability density \( \sigma(m) \) (Equation (12)) is also show. The actual procedure used to compute such error bars will be discussed in detail in the next section. The data misfit as a function of the conjugate-gradient iteration number is displayed in Figure 16, showing that the MAP model fits the observed data up to one standard deviation. The final residual (the difference between the observed data set and the data set modeled for the MAP model) is illustrated in Figure 17. No coherent events are
Figure 15. The MAP model estimated from the data set illustrated in Figure 13. The true model and the initial guess used in the inverse problem are also displayed.
Waveform inversion

![Graph](image)

**Figure 16.** The normalized data misfit as a function of the conjugate gradient iteration number. The data misfit threshold (related to the standard deviation of the noise) is also indicated in the figure.

![Image](image)

**Figure 17.** The difference between the observed data (Figure 13) and the synthetic data generated for the MAP model (Figure 15). The same plotting scale is used for both data sets.

seen in the figure, which consists mainly of the noise initially added to the data set generated for the true model. Figure 15 shows a good agreement between the MAP model and the true underlying model.

**Uncertainty Analysis**

The impedance and density models shown in Figure 15, not considering the error bars, constitute an incomplete solution to the inverse problem. In the situation, which is inevitably always the case, of noise-corrupted data, and a non-empty null space of the forward modeling operator (i.e., different models being mapped onto the same, or very similar, data sets), there is not an unique solution to an inverse problem. In the Bayesian formulation the solution to the inverse problem is not a model, but the *a posteriori* probability distribution $\sigma(m)$. Therefore, considering valid the approximations made in the construction of this distribution, it provides a direct mechanism to assess the spectrum of models that are likely to form the solution of the inverse problem. The *a posteriori* probability distribution takes into account the noise in the observed data, the null space of the forward modeling operator and the *a priori* information available to the inverse calculation. However, in complicate problems as the non-linear reflection seismic waveform data inversion, it is convenient (and probably mandatory due to computational limitations) to make further approximations to allow the manipulation of this probability function. Here, I resorted to linearizing the forward modeling operator (the reflectivity integral) about the MAP solution. Once this is done, the *a posteriori* probability function defined in this work is truly Gaussian, with model covariance defined in Equation (12), repeated here for convenience:

$$C_M' = \left[G^H C_D^{-1}G + C_M^{-1}\right]^{-1}. \tag{17}$$

Recall that the optimization of the *a posteriori* probability function (or equivalently, the optimization of the objective function $\Theta(m)$, defined in Equation (9)) is carried on without any linearization. Such a linearization is just performed after the MAP model is encountered. Figure (18) displays the *a posteriori* model covariance for this problem. As Equation (17) shows, this matrix is computed by inverting the quantity $\left[G^H C_D^{-1}G + C_M^{-1}\right]$, which can present eigenvalues small enough to make the inverse numerically unstable. When that happens, which is the case in this work, a possible way to get around this problem, is by a numerical stabilization procedure, like computing the generalized inverse. This approach was taken to compute the matrix displayed in Figure (18).

The minimum singular value allowed in the calculation was 0.001. The square-root of the main diagonal of this matrix provides the standard deviations on each one of the model parameter estimations. According to the analysis done here, all models within those error bars can be considered probable solution to the inverse problem. As show in Figure (15), the true solution lies within the error bars for most of the depth range in which the inversion was performed. The error bars on density values are the largest ones, emphasizing the limited information on those parameter in seismic amplitudes. Actually in this case, the initial guess to the inverse problem also lies with the error bars, indicating the the inverse did not improve the initial density model. It only provided ranges within each the true density might lie.
computing the *a posteriori* covariance matrix. From which I derived error bars on the solution of the inverse problem. The final result, illustrated in Figure 15 for the noise corrupted data shown in Figure 13 fits the data to one standard deviation and reproduces the true model to a good extent. In general the true model lies within the error bars computed for the MAP model. The error bars on density are the largest ones, reflecting the poor ability of seismic amplitudes to resolve density variations in the subsurface.

The inversion algorithm presented in this work is computationally intensive. The most demanding procedure is the calculation of the Frechet derivatives. However, due to the frequency domain approach, the level of parallelism is very high, making a parallel implementation of the algorithm suitable to distributed- or shared-memory machines. The current version of the inversion code is sequential, but its parallelization should be straightforward.

**Acknowledgments**

I thank Prof. John Scales for the constant encouragement, and Doug Hart for the line search code used in this work.

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, the Army Research Office and the National Science Foundation under grant DMS-9506603.

**References**


Fuchs, K. and Müller, G., 1971, Computation of synthetic seismograms with the reflectivity method and
Resolution of seismic waveform inversion: Bayes versus Occam

Wences Gouveia and John A. Scales
Department of Geophysics and Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80411 USA

ABSTRACT
In Bayesian inference, information about models is posited a priori. This information, which may very well include features in the null space of the forward problem, affects both the computed models and the resulting resolution estimates. In Occam’s inversion, on the other hand, the goal is to construct the smoothest model consistent with the data. This is not to say that one believes a priori that models are really smooth, but rather that the length scale associated with the smoothing is an indirect measure of resolution, since small-scale features not required to fit the data tend to be eliminated. In some cases the mathematical machinery of Bayesian inference resembles that of Occam’s inversion, but the goals and interpretations of the two methods are rather different. To understand better the similarities and differences of these two approaches, we show an application of both methods to the problem of inferring the Earth’s near surface elastic properties from reflection seismic data. On the one hand we derive a priori information about the Earth’s layering from fine-scale borehole measurements; coupled with information about the noise in the data and the elastic forward modeling operator, we are able to compute the Bayesian a posteriori probability distribution. Pseudo-random realizations of this posterior probability may have features that are implied by the a priori information as well as the data, even if the former are not well resolved by the data. Then we solve the Occam’s inversion problem by determining the maximum (second difference) regularization that allows for the data to be fit. In this case we estimate the resolution in terms of the degree of model smoothness implied by the data. In the Occam calculation we do not attempt to incorporate a priori information about the models, and the resulting averaged Earth model is very similar to the MAP model computed via Bayesian methods. But the error estimates associated with these models are rather different.

Key words: Seismic Waveform Inversion, Occam’s Inversion, Bayesian Inference

1 INTRODUCTION
Solving an inverse problem means making inferences about physical systems from real data. This requires taking into account three different kinds of information.

- What is known about the parameters independently of the data? In other words, what does it mean for a model to be reasonable or unreasonable?
- How accurately are the data known? That is, what does it mean to “fit the data”?
- How accurately is the physical system modeled? Does the model include all the physical effects that contribute significantly to the data?

We will describe two fundamentally different strategies for solving inverse problems in the context of reflection seismic waveform inversion. In the Bayesian approach, we encapsulate prior information about layered Earth models in the form of probability distributions; distributions which are independent of the
data and may describe variations in material properties at length scales well below the resolution of the data. When such distributions are combined with probabilistic information about data uncertainties (both random and theoretical), it is possible to derive a final (posterior) probability distribution assimilating both types of information. All questions of resolution are answered via this posterior probability.

On the other hand, the construction of such prior probabilities is a controversial matter. So an alternative, and we think quite reasonable, approach is to find the most featureless models that are consistent with the data and whatever other deterministic information that is available. This featurelessness is achieved via Tikhonov regularization with a discrete second difference matrix. Resolution is implicitly determined by the length scale over which smoothing is possible while still fitting the data. This approach is called Occam’s inversion by Constable et al. (1987).

### 1.0.1 Bayesian Inversion

In the Bayesian approach the three categories of information mentioned above are encapsulated as probability distributions. The net result of this is an equation assimilating all available information into a posterior probability distribution on the space of models \( m \) (Tarantola, 1987)

\[
\sigma(m) = \rho_M(m) \int_D \frac{p_D(d) \Theta(d|m)}{\mu_D(d)} \, dd
\]

where \( \sigma \) is the posterior probability, \( \rho_M \) is a prior distribution on the space of models, \( p_D \) is the distribution of data uncertainties, \( \Theta \) is a conditional distribution which incorporates errors in the theory (i.e., in the forward model), and \( \mu_D \) is a normalizing distribution sometimes taken to be the Jaynes null state of information. Everything on the right side of Equation 1, except for \( \rho_M \), can be thought of as a likelihood function, measuring the degree of data fit.

Equation (1) is rather general. For purposes of this paper we will simplify this by assuming that all the uncertainties in the problem can be described by stationary Gaussian distributions. In the Gaussian case the likelihood function reduces to

\[
\sqrt{\frac{(2\pi)^{-n}}{\det C_D}} \exp \left[ -\frac{1}{2} (g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \right]
\]

where \( d_{\text{obs}} \) is the vector of observed data which dimension is \( n \), \( C_D \) is the data covariance matrix and \( g(m) \) is the forward operator. Similarly, the Gaussian prior distribution reduces to

\[
\sqrt{\frac{(2\pi)^{-m}}{\det C_M}} \exp \left[ -\frac{1}{2} (m - m_{\text{prior}})^T C_M^{-1} (m - m_{\text{prior}}) \right]
\]

where \( m \) is the number of model parameters and \( C_M \) is the covariance matrix describing the distribution model parameters about the a priori model \( m_{\text{prior}} \). In the event that all uncertainty is Gaussian, the full posterior probability is the normalized product of Equations 3 and 2. If the forward operator is linear, then this distribution is itself a Gaussian.

In the Bayesian approach the information implied by \( \rho_D(m) \) could very well be unresolved by the data. But if we really believe that we can encapsulate valid information about the Earth in \( \rho_D(m) \), then it should be incorporated into the calculation. Conceptually this is simply a matter of looking at all the models that are reasonable according to \( \rho_D(m) \) and seeing which ones fit the data.

### 1.0.2 Occam’s Inversion

Let us accept the exponent of Equation (2) as being a useful measure of data fit irrespective of any Bayesian interpretation of the models. Then the set of models \( m \) such that

\[
(g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \leq \epsilon
\]

subject to

\[
(g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \leq \epsilon
\]

where \( R \) is a discrete second-difference operator. We can implement this practically as a weighted least squares problem with a Lagrange multiplier to control the trade-off between model smoothness and data fit. By increasing the model smoothness until we can no longer fit the data to the prescribed tolerance, we will have found our model.

### 1.0.3 Occam versus Bayes

A similar weighted least squares problem arises when one attempts to find the model that maximizes the Bayesian posterior when using Gaussian distributions. Indeed the discrete difference operator \( R \) does define some sort correlation among the model parameters: the smoother the
model, the more correlated its parameters. If we were to solve the Occam problem, and thereby determine the maximum degree of smoothness consistent with the data (as defined by Equation (4)), and then incorporate this smoothness into the Bayesian problem as an a priori covariance matrix $C_M^{12} = \lambda R^2 R$, where $\lambda$ is a constant, then the two problems would be almost equivalent. However, this is a rather extreme example because the whole reason for trying to solve the problem from a Bayesian point of view is to have the ability to incorporate data-independent model information. And if such information is available should we not be able to construct a more informative prior covariance matrix $C_M$ than would be implied by the data alone, as in Occam?

How to incorporate realistic deterministic prior information into an Occam style inversion is an interesting issue, but not one we will address in this paper. Here we consider the problem of using elastic reflection seismograms (synthetically generated with known Gaussian errors) to infer $P$-wave impedance, $S$-wave impedance and density of the upper crust. We will use 4 seconds of data to image the Earth down to 1.5 km, and typical $P$-wave wavelengths are on the order of 300 m.

On the one hand, we will use Occam to find the smoothest model that fits the data (achieves a normalized $\chi^2$ of one). For the Bayesian problem, on the other hand, we will construct an a priori covariance matrix for layered Earth models using well logs, a form of geologic information commonly available to exploration geoscientists. Since well logs are made using much higher frequencies than the reflection seismic data, they have resolution below 1 m. (However, to keep our life simple, we will use a 10 m discretization.) As a result, the a priori covariance matrix will contain information about layered Earth models well below the resolution of the surface seismic data. It is our goal to see to what extent this prior information influences the computed models and their resolution, and to make a fair comparison of the results of realistic Bayesian and Occam calculations. Finally, we obtain analytic expressions for the bias of the estimates obtained in the Occam's and Bayesian approaches.

2 FORMULATION OF THE TWO PROBLEMS

For both the Bayesian and Occam calculations we make the following assumptions:

- The observed data $d_{obs}$ are contaminated by Gaussian errors with known covariance $C_D$.
- The models $m$ are elements of $R^n$ where $m$ is fixed throughout. In particular $m = 150$, which is three times the number of layers, there being one unknown $P$-wave impedance, $S$-wave impedance and density for each of the 50 layers.
- The forward operator $g(m)$ is known and exact. In our case this is an elastic reflectivity-based synthetic seismogram algorithm (Fuchs and Müller, 1971).
- The Bayesian a priori model $m_{prior}$ is taken to be a long wavelength approximation of the true model and is used as the initial model in the Occam calculation.

Now let us define the specifics of the two calculations. The goal of the Bayesian calculation is to compute the posterior probability distribution:

$$
\sigma(m) \propto \exp\left[-\frac{1}{2} \left( (g(m) - d_{obs})^T C_D^{-1}(g(m) - d_{obs}) 
+ (m - m_{prior})^T C_M^{-1}(m - m_{prior}) \right) \right],
$$

(7)

Due to the nonlinearity of $g$, resolution analysis of this probability distribution is computationally expensive. To simplify the calculations we do a nonlinear optimization to find the model ($m_{map}$) that maximizes $\sigma(m)$. A Gaussian approximation of this density is then made about this model, yielding (Tarantola, 1987):

$$
\sigma(m) \propto \exp\left[ -\frac{1}{2} (m - m_{map})^T C_{M'}^{-1}(m - m_{map}) \right],
$$

(8)

where

$$
C_{M'} = [C_M^{-1} + C_D^{-1}]^{-1},
$$

(9)

is the a posteriori covariance matrix, a resolution measure that quantifies the uncertainties about $m_{map}$. $G$ is the Fréchet derivative of $g$, evaluated at this model. Therefore the Bayesian problem, which we refer to as Problem 1, can be stated as:

**Problem 1. Bayesian Inversion**

- Compute $\max_m \sigma(m)$ defined in in Equation (7) by solving a nonlinear optimization problem.
- Linearize the forward operator $g$ about the resulting MAP model $m_{map}$.
- Compute the a posteriori covariance matrix.

We consider now the Occam's approach to inversion. Let $R$ be a a Tikhonov second difference regularization matrix. Possible forms for this operator are

$$
R = \begin{bmatrix}
-2 & 1 & 0 & 0 & \cdots & 0 & 0 \\
1 & -2 & 1 & 0 & \cdots & 0 & 0 \\
0 & 1 & -2 & -1 & \cdots & 0 & 0 \\
& & & \ddots & \ddots & \ddots & \\
0 & 0 & 0 & 0 & \cdots & 1 & -2
\end{bmatrix},
$$

(10)
or
\[
R = \begin{bmatrix}
0 & 0 & 0 & 0 & \ldots & 0 & 0 \\
1 & -2 & 1 & 0 & \ldots & 0 & 0 \\
0 & 1 & -2 & 1 & \ldots & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \ldots & 1 & -2 & 1 \\
0 & 0 & 0 & \ldots & 0 & 0 & 0
\end{bmatrix}.
\] (11)

However, there exists an important difference between these two regularization operators. The latter maps constant and linearly increasing vectors to a zero vector, and therefore has the same null space as the continuous operator. Operator (10) is not singular. It will certainly regularize the inverse problem, no matter what the null space of \(G\) is. Here we use (10). Next we state the Occam’s problem.

**Problem 2.** Occam’s Inversion

For fixed \(\lambda\) solve the weighted, regularized least squares problem

\[
\min_m \Gamma(m) = \frac{1}{2} \left[ (g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) \right] + \lambda (m - m_{\text{prior}})^T R^T R (m - m_{\text{prior}}). 
\]

Increase \(\lambda\) until

\[
(g(m) - d_{\text{obs}})^T C_D^{-1} (g(m) - d_{\text{obs}}) = \epsilon
\]

can no longer be achieved.

This regularized nonlinear least squares problem can be solved using the same optimization algorithm as used in the Bayesian MAP calculation. Then it’s simply a matter of repeating the calculation until we find the optimal regularization parameter \(\lambda\). The model \(m_{\text{occam}}\) associated with the largest value of \(\lambda\) that still fits the data is the solution to this problem.

### 3 Resolution

In the Bayesian calculation, all questions of resolution are addressed via the posterior probability density \(\sigma(m)\). Even assuming Gaussian errors and prior information, for a sufficiently nonlinear forward operator \(g\), we may have to resort to Monte Carlo integration or sampling methods to extract confidence sets or other measures of resolution from \(\sigma(m)\). In this paper we are going to sidestep this issue by assuming that we can make a Gaussian approximation to \(\sigma(m)\) in a neighborhood of the \(m_{\text{map}}\). This is equivalent to linearizing the forward operator about this model. That being the case, then the complete Bayesian picture of resolution is provided by a posteriori covariance matrix

\[
C_M = \left[ \sigma^2 C_D^{-1} G + C_M^{-1} \right]^{-1}.
\] (12)

In the numerical examples we will extract information from this equation in two different ways. First, by looking at the (square roots) of the diagonal elements, we can put 1-\(\sigma\) error bars on the MAP model. But these error bars can be misleading because they ignore correlations among the parameters. So we also generate a tour of pseudo-randomly simulated models sampled according Equation (12) and \(m_{\text{map}}\). By looking at the distribution of models sampled in this way, we get a clear picture of which features in the model are well resolved and which are not.

Now, since the prior covariance matrix \(C_M\) appears in Equation (12), it is obvious that the information we extract from the fine scale features in the well log influence the resolution; to a greater or lesser extent depending on the relative weights associated with \(G^T C_D^{-1} G\) and \(C_M^{-1}\). This is in contrast to the Occam’s solution, which seeks a level of model smoothness based solely on data fit. It seems reasonable then that the model produced by Occam would, in general, be smoother than the models sampled from the Bayesian posterior. Which is not to say that the MAP model will be less smooth than the Occam model. The MAP model, being the center of the Gaussian distribution, and thus an average, may very well be significantly smoother than a typical realization of this distribution. This is another argument for sampling models from the Bayesian posterior. If we are Bayesians, then we believe that the Earth models we generate should reflect the fine-scale information contained in the log, even if it is unresolvable by the data. Whereas if we are Occamists, then we try to find the broadest average of the Earth that is consistent with the data. That way we can be confident that the features that are in the model are required to fit the data. So in the Occam calculation such a degree of smoothness is one measure of resolution.

An alternative measure in the Occam’s inversion is done by propagating the data uncertainties, represented by \(C_D\), into the model space. That provides a covariance matrix that quantifies the uncertainties on the estimate of the subsurface parameters obtained by solving Problem 2.

As in the Bayesian case the assumption of linear forward modeling is necessary for such an uncertainty analysis. This can be accomplished by linearizing \(g\) about \(m_{\text{occam}}\), hence reducing Problem 2 to a weighted, regularized linear system of equations that relates model and data perturbations by

\[
\delta \tilde{m} = \delta d,
\]

where

\[
\tilde{A} = \begin{bmatrix} C_D^{-1} G \\ \sqrt{\lambda} R \end{bmatrix},
\]
\[ \delta d = \left[ \begin{array}{c} C_D^{-\frac{1}{2}} \delta d \\ 0 \end{array} \right] \\
= \left[ C_D^{-\frac{1}{2}} (d_{obs} - g(m_{prior})) \right]. \]

The uncertainties of the estimate \( m_{\text{occam}} \) are given by:

\[ E \left[ \delta m \delta m^T \right] = E \left[ \bar{A}^T \delta d (\bar{A}^T \delta d)^T \right] \]

\[ = \bar{A}^T E \left[ \delta d \delta d^T \right] \bar{A}^T \]

\[ = \bar{A}^T C_D^{-\frac{1}{2}} E \left[ \delta d \delta d^T \right] C_D^{-\frac{1}{2}} \bar{A}^T \]

\[ \approx \bar{A}^T \bar{A}^T. \]  

Here, \( \bar{A}^T \) is the pseudo-inverse of \( \bar{A} \), and \( \bar{A}^T \) is \( \bar{A}^T \) with the last \( N \) columns set to zero, where \( N \) is the dimension of the model space. \( \delta m \) is the solution obtained from Equation (13). Notice that since \( G \) is obtained by linearizing \( g(m) \) about \( m_{\text{occam}} \), \( E \left[ \delta d \delta d^T \right] \approx C_D \), neglecting linearization errors. Thus, the last equation in (15).

**NUMERICAL CALCULATIONS**

Here, we consider the problem of estimating \( P \)- and \( S \)-wave impedance, and density profiles from a synthetic surface seismic data set, generated for the elastic velocity model illustrated in Figure (2). These data, shown in Figure (1), was corrupted by correlated Gaussian noise. Figure (2) plots only the target zone for the inversion. 50 layers of 0.01 km thickness will be considered in this problem. In the following we compare the solutions to problems 1 and 2 when inverting this synthetic data set. We approach these problems in two ways. In the first we don't resort to any kind of linearization, and tackle the problem via nonlinear optimization procedures. In Appendix A we linearize the problem about model \( m_{\text{prior}} \), hence obtaining a linear system of equations that is solved iteratively. It is worth mentioning that the criterion to accept a model as a solution is based on data fitness, which is measured according to Equation (4), using the nonlinear forward operator in the first case, and its linearization in the second case.

The nonlinear case

**Problem 1**

Gouveia (1996), discusses in detail the results obtained when the inversion was carried on under the Bayesian framework, in a nonlinear fashion. In his implementation, no assumptions were made about the structures of the covariance matrices \( C_D \) and \( C_N \). Those were built directly from the correlation of the noise in the observed data (assumed known in this experiment) and from the correlation length of the true impedance and density profiles (Figure (2)). Figure (3) illustrates the MAP model, along with the true subsurface parameters and the model \( m_{\text{prior}} \). The error bars in the figure are the square-roots of the main diagonal of the \( a \) posteriori covariance matrix (9), as discussed before. The following conclusions can be drawn from this experiment.
The MAP model estimated from the data set illustrated in Figure 1. The true model and the initial guess used for the nonlinear optimization calculation is also shown.

- The MAP solution reconstructs the subsurface parameters with a good accuracy. Such a model fits the observed data to one standard deviation:
  - the error bars include the true model, for most of the depth range within which the inversion was attempted;
  - the P-wave impedance error bars are the smallest, reflecting the fact that they are the best resolved parameters in this experiment, and
  - the density error bars are the largest, indicative of the poor resolution of this parameter available in seismic amplitudes. Moreover, since the initial guess for the density profile is also within the error bars, it is as good a model as the MAP model. Therefore, the only information on density provided by the inversion was the error bars, representing the range within which the true density might lie.

Figure 4. Regularization parameter $\lambda$ plotted as a function of the standard deviation of the data misfit associated with the optimum model obtained from the regularized nonlinear inversion. The dotted line corresponds to one standard deviation of the noise. Therefore the intersection of the solid and dotted curves corresponds to having fit the data to one standard deviation on average. This graph was smoothed over a $\lambda$ interval of length five.

Problem 2

In order to find the model associated with the largest regularization parameter $\lambda$, a search procedure should be implemented. The one used here, was simply to increase this parameter to a value at which the data could no longer fit. Figure (4) shows the standard deviation of the data misfit associated with the optimum model, as a function of the regularization parameter. The model $m_{\text{occam}}$ associated with the optimal regularization parameter (indicated in the figure by the arrow) is shown in Figure (4). The figure also displays the true model and the initial guess. This result should be compared with the model displayed in Figure (3). Figure (6) plots $m_{\text{map}}$ and $m_{\text{occam}}$ in the same graph to make the comparison easier. Although both models are similar, it is not difficult to see the higher accuracy of $m_{\text{map}}$ compared to $m_{\text{occam}}$. (Compare P-wave impedances at the depths of 1.25 and 1.30 km, and S-wave impedances at depth 1.34 km, for instance). This illustrates, at least for this specific example, that the use of a priori information consistent with the true subsurface parameters (in this case, the correlation length of the medium) can emphasize true features in the inverse problem solution. Such features might be smoothed out by the Occam's procedure since they might not be required to fit the data.

Although $m_{\text{occam}}$ is smoother than $m_{\text{map}}$, by con-
Figure 5. The Occam's model estimated from the data in Figure 1 using a non-linear, regularized optimization procedure. The true model and the initial guess are also displayed.

Figure 6. Comparison of the models obtained from Bayesian and Occam inversions.

On the Resolution of the Estimates

In this section we discuss the resolution provided by the Bayesian and Occam techniques by examining the covariance matrices about $\mathbf{m}_{\text{map}}$ and $\mathbf{m}_{\text{occam}}$ for the nonlinear problem. In particular, with respect to the Bayesian technique, we sample the a posteriori probability distribution $\sigma(\mathbf{m})$, thus generating pseudo-realizations of subsurface models that are consistent with the a priori information and observed data.

The resolution of the the forward modeling operator $g(\mathbf{m})$ is discussed in Appendix B. There we discuss which features of the model space (long- or short-wavelength components of impedance and density) can be resolved.
by the inversion procedure just considering the observed data. We do so via the Singular Value Decomposition (SVD) of this operator linearized about a given model. Although any model would do in this analysis, we chose the model \( m_{\text{map}} \) displayed in Figure (3).

\textbf{Resolution Analysis in Bayesian Inference}

The \textit{a posteriori} covariance matrix is displayed in Figure (7). As explained in (Gouveia, 1996) the direct inversion of the quantity \( G^H C_D^{-1} G + C_M^{-1} \) can be numerically unstable due the presence of small eigenvalues. What Figure (7) actually shows is the pseudo-inverse of \( G^H C_D^{-1} G + C_M^{-1} \), in which the smallest eigenvalue allowed to enter in the calculation was 0.001. Once \( m_{\text{map}} \) and \( C_M \) are computed it is possible to sample the \textit{a posteriori} distribution and generate pseudo-random realizations of the subsurface that are consistent with the observed data and the \textit{a priori} information. This is done by taking the inner product of the lower triangular part of the \( LU \) decomposition of \( C_M \), with an uncorrelated Gaussian pseudo-random sequence. Typical realizations are displayed in Figure (8). As expected the \( P \)-wave impedance show the smaller variance. The density realizations show the largest variance, emphasizing that small information on density is available from seismic amplitudes. The \( S \)-wave impedance realizations are somewhere in between. Another way to display the pseudo-random models is shown in Figure (9). In this figure the realizations are shown side by side. Thus, larger lateral continuity implies a higher degree of confidence in the estimates of the subsurface parameter at a particular depth. Figure (9) illustrates the main purpose of the Bayesian inversion approach. With this procedure, we aim at assessing the non uniqueness of the inverse problem, due to noise in the observed data and to the null-space of the forward modeling operator. By generating realizations of the subsurface that are consistent with the available information, we are able to see a spectrum of plausible models. These simulations bear some similarity with stochastic simulations done in geostatistics (Isaaks and Srivastava, 1989). However, to our knowledge, the forward modeling operator is not directly considered in geostatistical simulations.

\textbf{Resolution Analysis in Occam’s Inversion}

The regularization parameter \( \lambda \) is directly related to resolution. The larger the regularization parameter associated with a given model, the larger its correlation length, and hence more limited its resolution. The covariance matrix \( E' \delta\hat{m} \delta\hat{m}^T \) is of course function of this parameter, as well as the regularization operator \( R \) and the modeling operator. The Occam’s covariance matrix of the model perturbations is shown in Figure (10). This quantity provides a measure of resolution that is spatial and parameter-dependent. The wider diagonals of \( E' \delta\hat{m} \delta\hat{m}^T \), as compared to the model covariance obtained in the Bayesian case (Figure (7)), imply a larger correlation length (i.e., smoothness) for the model perturbations. As a final remark, note that the diagonal of this covariance matrix becomes progressive wider as we consider \( P \)-wave impedance (upper right-hand quadrant), \( S \)-wave impedance (central quadrant) and density (lower left-hand quadrant). This reflects the difference in resolution for each one of the parameters.

\textbf{Bias of the Estimates}

Let \( \Omega \) be an estimator of some parameter \( \theta \). The estimator is unbiased if

\[ E[\theta - \Omega] = 0. \] (16)

Next we compute the bias for the Occam and Bayesian inverse problems. To keep the discussion as simple as possible we will assume in this section that the inverse problem is linear, that the forward operator is known exactly and that the discretization errors are negligible.

In statistics one of the main motivations for the computation of bias is to obtain an unbiased estimator from a biased one. However this will not be possible here. The reason is that the amount of bias depends on the true
Figure 8. Pseudo-random realizations of the \textit{a posteriori} probability distribution. Shown are $P$-wave impedance (top), $S$-wave impedance (middle) and density (bottom).
Figure 9. Random realizations of the a posteriori probability distribution. Lateral continuity implies on a high confidence of the estimates.
Bayes versus Occam

Figure 10. Covariance matrix obtained for the Occam's inversion.

subsurface. Nonetheless, the bias calculations that follow, provide further understanding on the inverse procedures studied in this work.

Biased of an Exactly Linear Inverse Problem

For a linear inverse problem with exact forward operator, the true model and observed data are related by:

$$G \text{m}^{\text{true}} = d + n,$$

where $n$ is a noise term. The least-squares estimator $\hat{m}$ of the true model is

$$\hat{m} = G^\dagger d,$$

where $G^\dagger$ is the pseudo-inverse of $G$. Therefore the bias is:

$$E[\text{m}^{\text{true}} - \hat{m}] = E[\text{m}^{\text{true}} - G^\dagger d] = E[\text{m}^{\text{true}} - G^\dagger (G \text{m}^{\text{true}} - n)] = (I - G^\dagger G) \text{m}^{\text{true}} + G^\dagger E[n],$$

assuming zero mean noise. Since the matrix $I - G^\dagger G$ is a projector onto the null space of $G$, in this case the bias of the generalized inverse model is the component of the true model in the null space of the forward problem. Thus the bias will zero if:

- The operator $G$ is invertible and so $G^\dagger = G^{-1}$, or
- $\text{m}^{\text{true}}$ is orthogonal to the null space of $G$, so that the true model has no component in the null space of $G$.

Now, if the true earth model were constant, then the second condition would imply that the row sums of the resolution matrix $G^\dagger G$ must be zero. But, of course, in general this is not the case.

Bias of the Bayesian estimate

Now in the Bayesian calculation we do not really compute estimators of the true model; we compute probabilities associated with the true model. So while we may certainly compute the expected difference between the unknown true model and any particular model, it is not clear whether this case can be interpreted as a bias. Nevertheless, it is an interesting exercise to compute this difference for the MAP model and for now we will think of this as the analogue of bias in the Bayesian calculation.

For a linear problem, the MAP model is the least squares solution of

$$\langle (G^T C_D^{-1} G + C_M^{-1}) \rangle \text{m} = G^T C_D^{-1} d.$$  \hspace{1cm} (20)

Thus

$$\text{m}_{\text{MAP}} = A^\dagger G^T C_D^{-1} d,$$  \hspace{1cm} (21)

where $A^\dagger$ is the pseudo-inverse of $G^T C_D^{-1} G + C_M^{-1}$. The expected difference between this model and the true earth model is:

$$E[\text{m}^{\text{true}} - \text{m}_{\text{MAP}}] =$$

$$E[\text{m}^{\text{true}} - A^\dagger G^T C_D^{-1} d] =$$

$$E[\text{m}^{\text{true}} - A^\dagger G^T C_D^{-1} (G \text{m}^{\text{true}} - n)] =$$

$$[I - (A^\dagger G^T C_D^{-1} G)] \text{m}^{\text{true}} + A^\dagger G^T C_D^{-1} E[n] =$$

$$[I - (A^\dagger G^T C_D^{-1} G)] \text{m}^{\text{true}}.$$  \hspace{1cm} (22)

We will now consider two cases. First assume that the matrix $G^T C_D^{-1} G + C_M^{-1}$ is invertible. That being the case, we have:

$$A^\dagger = (G^T C_D^{-1} G + C_M^{-1})^{-1},$$

and thus,

$$G^T C_D^{-1} G = A^{-1} - C_M^{-1}.$$  \hspace{1cm} (23)

Using this last expression in (22) it is easy to show that:

$$E[\text{m}^{\text{true}} - \text{m}] = C_M^{-1} C_M^\dagger \text{m}^{\text{true}},$$  \hspace{1cm} (24)

where $C_M^\dagger$ is the a posteriori model covariance matrix defined in Equation (9). Equation (24) has an interesting interpretation. Note that the bias is proportional to the product of the a posteriori and a priori model covariance matrices. In the situation that $C_M^\dagger = C_M$, i.e., no new information was provided by the observed data, the bias is equal to the true model $\text{m}^{\text{true}}$, regardless of the a priori model covariance $C_M$. As the inversion becomes
more informative, the elements of the matrix \( C_{M'} \) get smaller and so does the bias.

In the situation where the matrix \( G^T C_{D'}^{-1} G + C_{M}^{-1} \) is not invertible, the interpretation is less clear. In this case \( A' \neq (G^T C_{D'}^{-1} G + C_{M}^{-1})^{-1} \), and the expected difference between the true and the MAP model is given by Equation (22), which can be rewritten as:

\[
E[\hat{m}^{\text{true}} - \hat{m}] = m^{\text{true}} - (A'G^T C_{D'}^{-1} G) m^{\text{true}}. \tag{25}
\]

Note that if \( m^{\text{true}} \) lies in the null space of \( G \), then \( G m^{\text{true}} = 0 \) and the bias is equal to \( m^{\text{true}} \). If that is not the case, the bias will be a complicated function of the forward modeling operator and covariance matrices.

**Bias of the Occam Estimator**

For a linear Occam inversion, model and observed data are related by:

\[
(G^T C_{D'}^{-1} G + \lambda R^T R) m = G^T C_{D'}^{-1} d. \tag{26}
\]

In this case we seek an estimate of an average of the true model \( \hat{m}^{\text{true}} \), as opposed to an estimate of the true model itself. Hence, a more appropriate definition of bias is \( E[\hat{m}^{\text{true}} - m_{\text{occam}}] \). A possible description (certainly not the only one) of the true subsurface average \( m^{\text{true}} \) is:

\[
m^{\text{true}} = m^{\text{true}} - R m^{\text{true}} = (I - R) m^{\text{true}}, \tag{27}
\]

where \( R \) is a roughness operator. This operator can be directly related to the regularization operator \( R \) used in the Occam’s procedure, by \( R = \lambda R^T R \). Under this definition, one can show that the bias is given by:

\[
E[\hat{m}^{\text{true}} - m_{\text{occam}}] = \frac{1}{(G^T C_{D'}^{-1} G - \lambda R^T R)^{-1} - I} \lambda R^T R m^{\text{true}}, \tag{28}
\]

assuming that the matrix \( G^T C_{D'}^{-1} G + \lambda R^T R \) is invertible. Equation (28) indicates that the Occam’s estimate has little bias since \( m^{\text{true}} \) is an average, thus smooth by construction, and \( R^T R \) is a differentiation operator.

**Conclusions**

If we use a parameterization of model space which is sufficiently fine to avoid discretization errors, then there will inevitably be features described by these models that are not well constrained by data. In Occam’s inversion we compute the smoothest model that is consistent with the data and interpret the degree of smoothness as a direct measure of resolution. In Bayesian inversion we compute a probability distribution of models that is consistent with the data and use error bars derived from this distribution to assess the resolution provided by the inverse procedure. Bayesian inversion, however, allows one to incorporate probabilistic statements of data-independent information, if such information is available. In reflection seismology we believe such information is often available from in situ measurements of rock properties, geologic models, or other data. Since some of this information is clearly in the null space of the reflection seismic data, its incorporation into the calculation affords a clear illustration of the differences between the Occamist and Bayesian points of view.

We have shown a synthetic example of a reflection seismic inverse problem carried out from both a Bayesian and an Occamist point of view. On the one hand, we have computed a Gaussian a priori distribution of layered Earth models by estimating a covariance matrix from hypothetical well log measurements. These measurements are performed at a much finer resolution than the seismic wavelength. With this information we computed the maximum a posteriori model via a nonlinear optimization procedure. Making a Gaussian approximation to the posterior distribution about this MAP model allows us to compute the a posteriori covariance, from which we derive both 1-σ error bars and a tour of typical models pseudo-randomly sampled from this posterior. For the Occam calculation we optimized the nonlinear least squares data misfit function, regularized with a discrete second order operator, for a sequence of increasingly large values of the smoothness. This gives the smoothest model that fits the data. And while the length scale of the smoothing is a measure of resolution, we derived a generalized covariance matrix from the augmented linear system. The bandwidth of this matrix, being essentially a depth-dependent correlation length, offers a more detailed measure of the resolution provided by the data alone.

The result is that the MAP model is similar, although slightly superior, to the Occam model. Moreover, models pseudo-randomly simulated from the Bayesian posterior distribution have features that are not required to fit the data since they are influenced by the data independent prior information. Whereas, by construction, the Occam inversion does not generate an Earth model per se, but rather an estimate of the coarsest average of the Earth that is consistent with the data. We also compute the bias of the solutions in the two approaches. Strictly speaking the meaning of bias is not clear in the Bayesian case, but it is easy to compute the expected difference between the true model and any particular model; we do this for the MAP model and show that the result depends only on the ratio of the posterior to prior model covari-
ance. For the Occam's estimate we compute the bias with respect to the average of the true subsurface, instead of the true subsurface itself.

Acknowledgments
This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, the Army Research Office, the Shell Foundation, and the National Science Foundation under grant DMS-8506603.

References

APPENDIX A: The Linearized inversion
In this appendix we solve Problems 1 and 2 by linearizing the forward modeling operator about \( m_{\text{prior}} \).

A1 Problem 1
The linearized version of Problem 1 is

\[
(G^T C_D^{-1} G + C_Z^{-1}) \delta m = G^T C_D^{-1} (g(m_{\text{prior}}) - d_{\text{obs}}),
\]

\[
m_{\text{map}} = m_{\text{prior}} + \delta m. \tag{A1}
\]

\( G \) is the forward modeling operator \( g(m) \) linearized about \( m_{\text{prior}} \). This weighted, regularized linear system of equations was solved by a standard conjugate gradient procedure. The result, initial guess and true model are illustrated in Figure (A1). Comparison with the model obtained in the nonlinear approach (Figure (3)), shows that the linearized result is considerably poorer in accuracy. A direct comparison is obviously not fair due to the large differences in computational cost of both procedures. In this particular problem the nonlinear implementation costs approximately ten times more than the linearized version. The main objective here is to illustrate the higher accuracy in the results provided by the nonlinear approach to the problem of extracting impedance and density information from seismic amplitudes.
Figure A2. Regularization parameter $\lambda$ plotted as a function of the standard deviation of the misfit of the linear system (A2) associated with the optimum model obtained from the regularized, linearized inversion. The dotted line corresponds to one standard deviation of the noise. Therefore the intersection of the solid and dotted curves corresponds to having fit the data to one standard deviation, on average. This graph was smoothed over a $\lambda$ interval of length five.

A2 Problem 2

As expected the linear system obtained here has a similar form to the one obtained in the Bayesian case. It can be written as:

$$\left(G^T C_D^{-1} G + \lambda R^T R \right) \delta m = G^T C_D^{-1} \left(g(m_{\text{prior}}) - d_{\text{obs}}\right).$$

(A2)

$m_{\text{occam}} = m_{\text{prior}} + \delta m.$

As in the nonlinear case, Figure (A2) shows the regularization parameter as a function of the data misfit. The model $m_{\text{occam}}$ associated with the arrow in the figure is shown in Figure (A3). If this result is compared with the one obtained in the nonlinear version of the Occam's procedure (Figure (5)), we see that the linearization errors are not negligible. As in the Bayesian case, the model obtained with the nonlinear optimization (Figure (5)) is far superior than the one obtained in the linearized inversion. Figure (A4) shows $m_{\text{map}}$ and $m_{\text{occam}}$ in the same graph. A few comments can be made:

- $m_{\text{occam}}$ is considerably smoother than $m_{\text{map}}$. This model is obtained by solving a regularized least-squares linear system just once, as opposed to the nonlinear case, where a sequence of such systems are solved. This implies a more extensive averaging that will make $m_{\text{map}}$ obtained in the nonlinear case smoother.
- Both solutions are relatively poor in accuracy, which can be verified by inspection of Figures (A1) and (A3).
- $m_{\text{map}}$ preserves the general trend of the underlying model to a larger extent than the regularized inversion does. That can be attributed to the use of a priori information that is consistent to the true subsurface parameters.

Figure A3. The Occam's model estimated from the data in Figure 1, using a linearized, regularized inversion procedure. The true model and the initial guess are also displayed.

APPENDIX B: Resolution of the forward modeling operator

In this appendix we present an SVD analysis of the forward modeling operator $g(m)$. Let $G = Q\Sigma V^T$ be the SVD decomposition of the linearized operator $G$. Considering that we have a linear system $Gm = d$, the solution model $m$ can be represented as:

$$m = \sum_{i=1}^{\infty} \frac{q_i d}{\sigma_i} v_i,$$

(B1)
Bayes versus Occam

long-wavelength (smoother) components can be better resolved than short-wavelength (rougner) components from seismic amplitudes. This analysis should, however, be coupled with the coefficients displayed in the bottom part of Figure (B1). Notice that the longest-wavelength components of both impedance and density make a very small contribution (due to the small value of the coefficient) to the expansion (B1). Therefore, this portion of the parameter spectrum is not well resolved by the data, though model eigenvectors are associated with large singular values. In essence, what can be learned from this SVD analysis is that long- to middle-wavelengths of $P$-wave impedance and middle- to small-wavelength of both $S$-wave impedance and density could be resolved by the forward modeling operator $G$.

Figure A4. Comparison of the models obtained from Bayesian and regularized linearized inversions.

where $q_i$ and $v_i$ are the $i$-th columns of the matrices $Q$ and $V$, respectively, $\sigma_i$ is the $i$-th singular value of $G$. Model eigenvectors $v_i$ associated with large values of $\sigma_i$ are well resolved. Figure (B1) displays the model eigenvectors of $G$. In the upper part of the figure we show the normalized singular values. In the lower part we show the coefficients $\frac{q_i}{\sigma_i}$ of the expansion (B1). In order to form the solution model $m$ in Equation (B1) we add model eigenvectors with their respective weight until the superposition fits the data to one standard deviation. The cutoff is indicated by the arrow in Figure (B1).

Studying the spectrum of model eigenvectors we conclude that for both impedance and density profiles,
Figure B1: Singular Value Decomposition of the elastic waveform inversion. The arrow indicates the last model eigenvectors interpreted in the solution such that the data is 1.5 standard deviation of the noise. For each of the parameters the model eigenvectors are normalized to one.
Transverse isotropy versus lateral heterogeneity in inversion of $P$-wave reflection traveltimes

Vladimir Grechka
Center for Wave Phenomena
Colorado School of Mines

ABSTRACT

Transverse isotropy causes pronounced nonhyperbolic moveout of long-spread $P$-wave reflection data. Lateral heterogeneity may affect the moveouts in much the same way as transverse isotropy. For this reason, a given $P$-wave reflection moveout may be interpreted equally well in terms of parameters of homogeneous transversely-isotropic (TI) or laterally heterogeneous (LH) isotropic models.

In weakly laterally heterogeneous transversely isotropic media, the common-midpoint (CMP) moveout of a $P$-wave reflected from a horizontal interface has the form similar to that in homogeneous TI media. However, both the normal-moveout (NMO) velocity and the quartic moveout coefficient are affected by lateral heterogeneity which exhibits itself in the derivatives of the zero-offset traveltimes $t_0$ and the NMO velocity $V_{nmo}$ with respect to the lateral coordinate.

Despite the presence of heterogeneity, nonhyperbolic velocity analysis can be performed in the same way as in homogeneous TI models. However, to find the anisotropic parameter $\eta$ from the quartic moveout term, the second derivative of $V_{nmo}$ and the fourth derivative of $t_0$ are needed. Since these derivatives are calculated from the data by numerical differentiation, they further reduce stability in the estimation of $\eta$ as compared to that in homogeneous TI media. Consequently, $P$-wave reflection traveltimes cannot be simultaneously inverted for transverse isotropy and lateral heterogeneity, even in the simplest model of a single plane layer.

Key words: transversely isotropic laterally heterogeneous media, nonhyperbolic $P$-wave reflection moveout

Introduction

The presence of seismic anisotropy in different rocks and on various scales makes it necessary to generalize conventional isotropic velocity analysis to account for anisotropy. This has recently been done and a number of approaches to anisotropic traveltime inversion and velocity analysis were developed (Byun et al., 1988; Sena, 1991; Tsvankin and Thomsen, 1993; Alkhalifah and Tsvankin, 1995; Alkhalifah, 1996). Alkhalifah and Tsvankin (1996) showed that the normal-moveout velocity $V_{nmo}$ and the anisotropic parameter $\eta$, which can be obtained using surface $P$-wave reflections from interfaces of different dips, allow one to perform all time-processing procedures in TI media, including NMO, DMO, and time migration. Although this DMO-inversion approach provides a stable way of estimating the parameter $\eta$, it requires the presence of at least two dipping reflectors and cannot be used for dips exceeding 70-80°. The same parameters, $V_{nmo}$ and $\eta$, can be extracted from nonhyperbolic $P$-wave moveout reflected from horizontal interfaces (Alkhalifah, 1996). Although the accuracy of estimation of $\eta$ in the latter case is not sufficient for lithology discrimination (Grechka and Tsvankin, 1996), the results of nonhyperbolic velocity analysis can be used for imaging of steep reflectors such as faults and flanks of salt domes.

All the above-mentioned approaches are based on the assumption of lateral homogeneity of the TI models.
on the scale of a common midpoint gather. It could be expected that the presence of lateral heterogeneity would influence the P-wave reflection moveout in a very similar manner as transverse isotropy does. It is shown in this paper that a given P-wave reflection moveout in a weakly TI medium can be exactly reproduced in isotropic laterally heterogeneous model with weak lateral heterogeneity. This result indicates that there is no way to distinguish between TI and LH media based on a single P-wave moveout.

If several moveouts for adjacent common midpoints in a TI LH model are available, it is theoretically possible to separate the effects of anisotropy and heterogeneity and estimate the parameters of the model. However, this procedure requires numerical differentiation of the zero-offset traveltimes and the NMO velocity with respect to the horizontal coordinate. In this case, the estimation of anisotropic parameter $\eta$ is expected to be even less stable than that for homogeneous TI models. As was shown by Grechka and Tsvankin (1996), even in the simplest model of a single homogeneous TI layer, the values of the anisotropic parameter $\eta$ found from nonhyperbolic velocity analysis are subject to relative errors of about 50 - 80%. The presence of heterogeneity magnifies these errors further and can lead to completely unrealistic values of $\eta$.

### Approximations of P-wave reflection moveout in homogeneous weakly TI media

An approximate equation for the common midpoint (CMP) P-wave moveout $t(x)$ reflected from the horizontal interface in homogeneous TI layer with vertical symmetry axis was obtained by Tsvankin and Thomsen (1994), rewritten by Alkalifah and Tsvankin (1995)

$$
t^2(x) = t_0^2 + \frac{x^2}{V_{nmo}^2} - \frac{2\eta x^4}{V_{nmo}^2(t_0^4V_{nmo}^2 + (1 + 2\eta)x^2)}.
$$

(1)

In this equation, $x$ is the offset, $t_0$ is the zero-offset traveltine, $\eta$ is the anisotropic parameter which relates to Thomsen's (1986) parameters $\epsilon$ and $\delta$ as

$$
\eta = \frac{\epsilon - \delta}{1 + 2\delta}.
$$

(2)

and $V_{nmo}$ is the normal moveout velocity that can be expressed via vertical velocity $V_0$

$$
V_{nmo}^2 = V_0^2(1 + 2\delta).
$$

(3)

Under the assumption of weak anisotropy, i.e. when the parameters

$$
\{\epsilon, \delta, \eta\} \ll 1,
$$

(4)

equation (1) can be linearized (Tsvankin and Thomsen, 1994)

$$
t^2(x) = t_0^2 + \frac{x^2}{V_{nmo}^2} - \frac{2\eta x^4}{V_{nmo}^2(t_0^4V_{nmo}^2 + x^2)},
$$

(5)

where $\eta$ from equation (2) reduces to the difference

$$
\eta = \epsilon - \delta.
$$

(6)

Although the linearized moveout equation (5) is less accurate than Alkalifah and Tsvankin's (1995) equation (1), it can be used to gain insight into the influence of anisotropy on P-wave reflection traveltimes. Figure 1 compares traveltine $t(x)$ calculated using equation (5) and traveltine $t_{nmo}(x)$ computed numerically without weak anisotropy assumption and shows the errors of approximation (5).

It is instructive to derive equation (5) one more time applying the method based on the Frechêt derivatives of the traveltimes (Grechka and McMechan, 1996b) with respect to velocity variations. The same approach will be used in the following sections to obtain traveltimes in laterally heterogeneous media.

I begin with the equation for reflection moveout $t_{iso}(x)$ in a homogeneous isotropic medium

$$
t_{iso}(x) = \frac{\sqrt{x^2 + \delta^2}}{V_0},
$$

(7)
where \( V_0 \) is the isotropic P-wave velocity, \( a = 2h \) is the doubled reflector depth, and \( z \) is the the offset. Consider the velocity perturbation
\[
V = V_0 \left( 1 + \frac{\Delta V}{V_0} \right),
\]
where
\[
|\Delta V/V_0| \ll 1.
\]
(9)
The traveltimes
\[
t(x) = t_{iso}(x) + \Delta t(x)
\]
for the medium with the perturbed velocity can be approximated by the first two terms of the Taylor series in \( \Delta V/V_0 \)
\[
t(x) = \frac{\sqrt{x^2 + a^2}}{V_0} \left[ 1 - \frac{\Delta V}{V_0} \right].
\]
(11)
It follows from Fermat’s principle [see also Backus and Gilbert (1969)] that, if the quadratic and higher order terms with respect to \( \Delta V \) are neglected in equation (11), the perturbations of the ray trajectory can be neglected as well. That is why the factors \( \sqrt{x^2 + a^2}/V_0 \) are the same in equations (7) and (11).

Note that equation (11) does not assume any specific velocity perturbation; only inequality (9) has to be satisfied. I will apply equation (11) to general velocity perturbations caused by anisotropy and lateral heterogeneity. In this section, only anisotropic velocity perturbations will be considered.

It is important to mention that equation (11) is linear with respect to \( \Delta V \). This means that the final result of the derivation will also be linear with respect to anisotropic parameters which are responsible for the velocity perturbation \( \Delta V \) and assumed to be small.

Here, I basically repeat the derivation of equation (5) given in Appendix A of Tsvankin and Thomsen (1994). Consider a homogeneous TI layer with a vertical symmetry axis. The phase velocity \( v(\phi) \) as the function of the phase angle \( \phi \) is expressed, in the weak anisotropy limit \( (|\epsilon| \ll 1, |\delta| \ll 1) \), as (Thomsen, 1986)
\[
v(\phi) = V_0 (1 + \delta \sin^2 \phi \cos^2 \phi - \epsilon \sin^4 \phi),
\]
where \( V_0 \) is the vertical velocity. Although the phase angle \( \phi \) differs from the group angle \( \theta \) in a weakly anisotropic TI medium, the group velocity \( V(\theta) \) as the function of the group angle \( \theta \) has the same form as in equation (12)
\[
V(\theta) = V_0 (1 + \delta \sin^2 \theta \cos^2 \theta + \epsilon \sin^4 \theta).
\]
(13)
Since trigonometric functions of \( \theta \) in equation (13) are directly expressed via the offset \( x \) and the doubled reflector depth \( a \)
\[
\sin^2 \theta = x^2/(z^2 + a^2)
\]
and
\[
\cos^2 \theta = a^2/(x^2 + a^2),
\]
equation (13) is written in the form
\[
V = V_0 \left[ 1 + \frac{x^2(\epsilon + a^2 \delta)}{(z^2 + a^2)^2} \right].
\]
(16)
where the second term in the brackets can be treated as small velocity perturbation \( \Delta V \) in equation (8). This means that we can apply general equation (11) to the velocity function (16) and obtain P-wave reflection traveltimes in the form
\[
\frac{\sqrt{x^2 + a^2}}{V_0} \left( 1 - \frac{x^2(\epsilon + a^2 \delta)}{(z^2 + a^2)^2} \right).
\]
(17)
The next step is to show that equation (17) is equivalent to equation (5). Raising equation (17) to the fourth power and keeping only terms linear in \( \epsilon \) and \( \delta \), I obtain
\[
t(x) = \frac{1}{V_0^2} \left( a^4 + 2a^2x^2(1 - 2\delta) + x^4(1 - 4\epsilon) \right).
\]
(18)
Denoting \( t_s^0 = a^2/V_0^2 \) and \( V_{nmo}^2 = V_0^2 (1 + 2\delta) \), I find from equation (18)
\[
t(x) = (t_s^0 + \frac{x^2}{V_{nmo}^2})^2 - 4z^4 (\epsilon - \delta)/V_{nmo}^2.
\]
(19)
Finally, a linear approximation of the square root of equation (19) gives exactly equation (5).

The derivation above justifies the approach of obtaining traveltimes based on the velocity perturbations. Below, I will apply the same method to account for lateral heterogeneity.

**Interpretation of TI moveout in terms of lateral heterogeneity**

Let us examine equation (5) from a somewhat different standpoint. Assume that we do not know anything about the origin of this equation and try to interpret it in terms of isotropic velocity heterogeneity. First, a specific kind of heterogeneity should be chosen because a general velocity function \( V(z, z) \) (where \( x \) is the lateral coordinate and \( z \) is the depth) cannot be determined from a single moveout. Comparing equations (13) and (16), we see that velocity dependence on the propagation angle due to anisotropy can be alternatively treated as the offset dependence of the isotropic velocity field. Therefore, equation (16) suggests that laterally varying velocity \( V(z, z) \) should be chosen to simulate the influence of anisotropy. Moreover, equation (16) indicates that lateral heterogeneity, which affects traveltimes similarly to weak anisotropy, must be weak as well. Hence, the appropriate model is a laterally heterogeneous (LH) medium with velocity...
\[ V(x) = V_0 \left[ 1 + c(x) \right] \]  
(20)

where \( |c(x)| \ll 1 \) is a dimensionless function.

Since the velocity (20) has the generic perturbation form (8), the approximation of traveltime can be expressed in the form similar to equation (11). I generalize this equation to account for lateral velocity heterogeneity. First, it is necessary to specify the coordinate of the source because now the traveltimes depend not only on the offset but also on the source position. For the source located at the zero coordinate and the offset \( x \), equation (7) can be generalized as

\[ t(x) = \frac{\sqrt{x^2 + z^2}}{V_0} \int_0^{x} \frac{d\xi}{V(\xi)}. \]  
(21)

Equation (21) assumes that, in the case of weak heterogeneity, integration can be performed along straight rays.

Substitution of equation (20), in the linear approximation with respect to \( c \), gives

\[ t(x) = \frac{\sqrt{x^2 + z^2}}{V_0} \left[ 1 - \frac{1}{x} \int_0^x c(\xi) d\xi \right]. \]  
(22)

where division by zero for \( x \to 0 \) does not occur because

\[ \lim_{x \to 0} \frac{1}{x} \int_0^x c(\xi) d\xi = c(0). \]  
(23)

Equation (22) presents the common-shot-point (CSP) reflection moveout in a weakly LH medium, which now is generally different from the CMP moveout. The latter can be obtained by shifting the integration limits in equation (22) and, for the midpoint located at zero coordinate, the CMP moveout is

\[ t(x) = \frac{\sqrt{x^2 + z^2}}{V_0} \left[ 1 - \frac{1}{V_0} \int_{-z/2}^{z/2} c(\xi) d\xi \right]. \]  
(24)

Equations (22) and (24) have the form that is quite similar to that of equation (11). The second terms in the brackets in equations (22) and (24) express the impact of the lateral heterogeneity on the reflection travel times.

Let us come back to the initial problem of this section and find the isotropic lateral velocity distribution \( V(x) \) that reproduces moveout (3) in a homogeneous TI medium. Comparing equations (17) and (22) for CSP moveout, we conclude that the lateral perturbation \( c(x) \) in velocity is expressed as

\[ c(x) = \frac{d}{dx} \left[ \epsilon^2 (x^2 + a^2 \delta) \right] \frac{(x^2 + a^2)}{(x^2 + a^2)^2} \]  
(25)

or, in the linear approximation,

\[ c(x) = \left[ \delta \xi V_{\text{nmo}}^2 + \epsilon^2 \xi V_{\text{nmo}}^2 - z^2 \right] \left[ (\xi V_{\text{nmo}}^2 + x^2) \right]^{-1}. \]  
(26)

where \( \delta \) and \( \epsilon \) should be treated now as the parameters of the isotropic lateral heterogeneous velocity field.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{image.png}
\caption{Functions \( c \) describing lateral velocity heterogeneity found for reflection traveltimes (5) for TI models: solid – Taylor sandstone; dashed – Dog Creek shale.}
\end{figure}

Equation (26) indicates that the velocity heterogeneity \( c(x) \), reproducing moveout (5) in a homogeneous TI medium, is the symmetric function of the offset \( x \). It is not surprising because the velocity function (16), corresponding to transverse isotropy, is symmetric as well. Figure 2 demonstrates the functions \( c(x) \) that give the same reflection traveltimes as those in the homogeneous TI models of Taylor sandstone and Dog Creek shale. It indicates that, indeed, lateral velocity heterogeneity is weak.

Figure 3 shows that traveltimes in TI media and corresponding LH media are remarkable close. This justifies the validity of the approximations made. On the other hand, this result indicates that the moveouts in homogeneous TI and isotropic LH media are practically indistinguishable and a single TI moveout curve can be interpreted in terms of parameters of lateral heterogeneity.

\begin{equation}
V(x, \xi) = V_0 \left[ 1 + c(\xi) + \frac{x^2 (\xi V_{\text{nmo}}^2 + a^2 \delta (\xi))}{(x^2 + a^2)^2} \right].
\end{equation}
(27)

where \( x \) is the offset, \( \xi \) is the horizontal coordinate of a point on the ray, \( a = 2h \), and \( V_0 \equiv V_0(0, \xi) \) is the vertical velocity at the offset \( x = 0 \). Equation (27) can be obtained by combining equations (16) and (20). It assumes again that the ray trajectories are straight lines and the
Figure 3. The relative difference \( e(z/h) = t_{LH}/t_{TI} - 1 \) (in %) between the travel times \( t_{LH} \) computed in TI media using approximation (3) and travel times \( t_{LH} \) computed in LH media with velocity distribution given by equations (20) and (26): solid – Taylor sandstone; dashed – Dog Creek shale. The travel times in LH models were computed using the algorithm described by Grechka and McMechan (1996a).

A linear approximation with respect to small parameters \( c(\xi) \), \( \epsilon(\xi) \), and \( \delta(\xi) \) is applicable.

To obtain the expression for the CMP P-wave reflection moveout in weakly LH TI medium, I use the Fréchet derivatives of travel times (Grechka and McMechan, 1996b) with respect to \( c(\xi) \), \( \epsilon(\xi) \), and \( \delta(\xi) \) to perturb travel times in homogeneous isotropic medium [equation (7)].

The moveout takes similar to equation (24) form

\[
t(x) = \sqrt{\frac{z^2 + a^2}{V_0^2}} \left[ 1 - \frac{1}{z} \int_{-z/2}^{z/2} c(\xi) \, d\xi \right] - \left( \frac{x}{x^2 + a^2} \right)^{1/2} \left[ \frac{a^2}{z^2} \left( \epsilon(\xi) + a^2 \delta(\xi) \right) \right] (28)
\]

and contains one additional term due to anisotropy. Using the expansions of the functions \( c(\xi) \), \( \epsilon(\xi) \), and \( \delta(\xi) \) in Taylor series in the vicinity of \( \xi = 0 \), equation (28) can be written as (Appendix A)

\[
t^2(x) = t_0^2 + \frac{x^2}{V_{nmo}^2} - \frac{q x^4}{V_{nmo}^2 (t_0^2 V_{nmo}^2 + x^2)}. \quad (29)
\]

In equation (29), \( t_0 \) is the two-way zero-offset traveltime,

\[
V_{nmo}^2 = V_0^2 (1 + 2 \delta_0 + \frac{t_0^2 V_0^2 V_{nmo}^2}{12}) \quad (30)
\]

is the normal-moveout velocity, the coefficient of the quartic term

\[
q = 2 \eta_0 + \frac{1}{12} t_0^2 V_{nmo}^2 \left( \delta_0'' + \frac{V_0''}{V_{nmo}} \right)
\]

the parameter

\[
\eta_0 = \epsilon_0 - \delta_0, \quad (32)
\]

and the superscripts \( (\cdot)'' \) and \( (\cdot)''' \) denote the second and the fourth derivatives of corresponding functions with respect to the CMP-coordinate \( x \) at the midpoint.

It is worthwhile to compare equations (5) and (29) to examine the influence of the lateral heterogeneity on traveltimes. Clearly, equations (5) and (29) have the same form. The difference between them is that the terms containing lateral derivatives of \( V_0 \) and \( \delta_0 \) appear in expression (30) for NMO velocity and in the coefficient \( q \) of the quartic moveout term [equation (31)]. Note that only the even-order derivatives are present. The odd-order derivatives were canceled out because of symmetry of the integration limits in equation (28). This physically means that, under the assumption of weak heterogeneity, only symmetrical (with respect to the midpoint) variations of velocity and anisotropic parameters \( \epsilon \) and \( \delta \) affect the CMP moveout.

Another important feature of the moveout equation (29) is that the values of \( \eta_0 \) and \( \delta_0 \) in the quartic term are combined in the single parameter \( \eta_0 \) [equation (32)]. The physical meaning of the parameter \( \eta_0 \) was explained by Alkhalifah and Tsvankin (1993): it expresses the deviation of the P-wave slowness curve from an ellipse in homogeneous TI media and, namely because of this deviation, the reflection moveout in TI media is nonhyperbolic. In the problem under examination, velocity varies along the reflected rays for non-zero offsets because of lateral heterogeneity. These velocity variations can produce the same effect on the traveltimes, causing nonhyperbolicity of the moveout, even in the absence of anisotropy. This is explicitly exhibited by the quartic term in equation (29). Therefore, the nonhyperbolicity of the reflection moveout in LH TI media can be caused by a number of reasons and can not be described in terms of just single parameter \( \eta_0 \).

A possible approach to nonhyperbolic velocity analysis in LH TI media

The moveout equation (29) in LH TI media has similar form to equation (5) which is applicable in the homogeneous TI models. Therefore the nonhyperbolic velocity analysis applied by Alkhalifah (1996) in the case of homogeneous TI media may serve as a starting point for the forthcoming analysis. Alkhalifah (1996) used the fact that the moveout (1) is the function of three parameters \( t_0, V_{nmo}, \) and \( \eta \) (which is equivalent to \( \eta_0 \) in the notation...
for LH TI media) and suggested to determine $V_{nmo}$ and $\eta$ by a two-dimensional semblance scan for a given value of $t_0$. Although the obtained values of $\eta$ are not as stable as those extracted from the NMO velocities measured at two different dips (Alkhalifah and Tsvankin, 1993), they still provide some restraints on $\eta$ and can be used in time migration.

In the presence of lateral heterogeneity, the reflection moveout (29) no longer depends only on $t_0$, $V_{nmo}$, and $\eta$; the derivatives of vertical velocity and anisotropic parameter $\delta$ with respect to lateral coordinate should be accounted for as well. Evidently, there are more unknowns than equations. However, all the parameters now are functions of the coordinate $y$ of the midpoint, therefore we can try to utilize the information about lateral variations of $t_0(y)$ and $V_{nmo}(y)$ to reduce the number of unknowns. This can be done by taking into account the fact that, under the assumption of weak heterogeneity, the two-way zero-offset traveltime

$$t_0(y) = \frac{a}{V_0(y)},$$

where $a = 2h$ is the doubled reflector depth and $V_0(y)$ is the vertical velocity as the function of lateral coordinate. Equation (33) explicitly exhibits the assumption that the reflector is horizontal, i.e., its depth $h$ is independent of $y$. The linear approximation of the derivatives of $t_0$ gives

$$\frac{1}{t_0(y)} \frac{d^2 t_0(y)}{dy^2} = -\frac{1}{V_0(y)} \frac{d^2 V_0(y)}{dy^2}.$$

Therefore, all the ratios $V_0^{(k)}(y)/V_0(y)$ can be considered as the known quantities as long the derivatives $t_0^{(k)}(y)$ are calculated by consecutive differentiation of $t_0(y)$.

Equations (34) and equation (B4) for the second derivative of $\delta_0$ from Appendix B allow one to rewrite equations (30) and (31) in the form

$$V_{nmo}^2 = V_0^2 (1 + 2\delta_0 - \frac{1}{12} t_0 V_0^{(IV)}),$$

$$q = 2\eta_0 + \frac{\delta_0 V_{nmo}}{12} \left[ V_{nmo}'' + \frac{7}{240} t_0 V_{nmo}^{(IV)} \right].$$

Traveltime $t(x)$ in LH TI media is described now in terms of the same parameters $t_0$, $V_{nmo}$, and $\eta$ as traveltime (5) in homogeneous TI media and in terms of the derivatives of $t_0$ and $V_{nmo}$ that appear due to lateral heterogeneity.

Equation (29) can be, in principle, used for nonhyperbolic velocity analysis in LH TI media. However, it requires the knowledge of the fourth derivative of the zero-offset traveltime $t_0^{(IV)}(y)$ and the second derivative of the normal-moveout velocity $V_{nmo}''(y)$ with respect to lateral coordinate $y$. Since $t_0(y)$ and $V_{nmo}(y)$ are essentially discrete functions which are known only at the CMP coordinates, the derivatives $t_0^{(IV)}$ and $V_{nmo}''$ have to be found by numerical differentiation. Clearly, any errors in $t_0^{(IV)}$ and $V_{nmo}''$ will propagate into estimated values of $\eta$.

As discussed by Grechka and Tsvankin (1996), the parameter $\eta$ found from nonhyperbolic velocity analysis in homogeneous TI media is extremely sensitive to small long-period traveltimes errors because of the trade-off between $\eta$ and $V_{nmo}$. Evidently, the presence of higher-order derivatives of $t_0$ and $V_{nmo}$ in the quartic moveout term (36) further reduces the stability of $\eta$ estimation. This indicates that P-wave reflection traveltime data in LH TI media do not provide enough constraints to estimate both anisotropy and lateral heterogeneity and separate their effects on the traveltimes.

**Discussion and conclusions**

In this paper, I examined the simultaneous influence of transverse isotropy and lateral heterogeneity on P-wave reflection moveout from a horizontal interface. I showed that the dependence of group velocity on the ray direction in homogeneous TI media can be alternatively treated as the dependence of isotropic velocity on the offset for a laterally heterogeneous model of some specific kind. Therefore, it is not surprising that transverse isotropy and lateral heterogeneity can mimic each other's influence and produce very similar effects on traveltimes.

It is important to point out that the nonhyperbolic reflection moveout in weakly TI media can be reproduced in isotropic LH models and lateral heterogeneity turns out to be weak as well. In contrast, vertical velocity heterogeneity must be rather strong to cause noticeable nonhyperbolicity of the moveout (e.g., Alkhalifah (1966)) because the nonhyperbolicity appears in vertically heterogeneous media only due to ray bending.

If the TI medium is laterally heterogeneous, the reflection moveout contains additional terms responsible for the lateral variations in the vertical velocity $V_0$ in the vicinity of the midpoint. These terms may cause pronounced nonhyperbolicity of the moveout even in isotropic medium.

Nonhyperbolic velocity analysis in LH TI media can be performed based on equation (29) in exactly the same fashion as it was done by Alkhalifah (1996) for homogeneous or vertical inhomogeneous TI media. This analysis provides estimates of the normal-moveout velocity $V_{nmo}$ and the coefficient $q$ of the quartic moveout term. However, in the presence of lateral heterogeneity, the coefficient $q$ contains not only the anisotropic parameter $\eta$, but also the terms expressing the impact of heterogeneity. These terms contain derivatives of the zero-offset traveltime $t_0^{(IV)}(y)$ and the NMO velocity $V_{nmo}(y)$ with
respect to lateral coordinate $y$. Calculation of the derivatives may seem straightforward, however, they are to be computed by numerical differentiation of $t_0$ and $V_{nmo}$ determined from the data, which may lead to substantial errors. It is obvious from equation (36) that any errors coming from numerical evaluation of the derivatives $t_0^{(V)}$ and $V_{nmo}$ will distort the values of $\eta_0$. Furthermore, as was shown by Grechka and Tsvankin (1996), the estimation of the quartic term $q$ itself ($q$ equals to $\eta$ in the case of homogeneous TI media) suffers from the trade-off between $q$ and $V_{nmo}$ which leads to relative errors in $q$ of about 50 – 80% even for the simplest single-layer model. Combination of these two kinds of errors may give rise to entirely unrealistic estimations of $\eta$.

It is known (Alkhaliifah and Tsvankin, 1995) that the values of $\eta$ obtained using NMO velocities from dipping reflectors in homogeneous TI media are more stable than those found from nonhyperbolic velocity analysis of horizontal events. Similar approach is expected to give better results in the presence of lateral heterogeneity as well.

Acknowledgments

I am grateful to members of A(nisotropy)-team of the CWP for helpful discussion and to Ilya Tsvankin, Andreas Rüeger, and Tariq Alkhaliifah for reviewing the manuscript.

References


APPENDIX A: The moveout equation in LH TI media

To calculate the integrals in equation (28), I have to select some specific expressions for functions $c(\xi)$, $e(\xi)$, and $\delta(\xi)$. Bearing in mind that the final result should look similar to the polynomial-like equation (3), I have chosen to use Taylor series in the vicinity of $\xi = 0$:

\begin{equation}
\frac{c(\xi)}{c(0)} = \sum_{k=0}^{\infty} \frac{c^{(k)}(0)}{k!} \xi^k,
\end{equation}

\begin{equation}
\frac{e(\xi)}{e(0)} = \sum_{k=0}^{\infty} \frac{c^{(k)}(0)}{k!} \xi^k,
\end{equation}

and

\begin{equation}
\frac{\delta(\xi)}{\delta(0)} = \sum_{k=0}^{\infty} \frac{\delta^{(k)}(0)}{k!} \xi^k.
\end{equation}

The quantities $c^{(k)}(0)$, $c^{(k)}(0)$, and $\delta^{(k)}(0)$ are the derivatives of the order $k$ of vertical velocity $V_0(\xi)$ and anisotropic parameters $e(\xi)$, $\delta(\xi)$ with respect to $\xi$ at $\xi = 0$:

\begin{equation}
V_0^{(k)}(\xi) \equiv \frac{d^k V_0(\xi)}{d\xi^k} \bigg|_{\xi=0},
\end{equation}

\begin{equation}
e^{(k)}(\xi) \equiv \frac{d^k e(\xi)}{d\xi^k} \bigg|_{\xi=0},
\end{equation}

and

\begin{equation}
\delta^{(k)}(\xi) \equiv \frac{d^k \delta(\xi)}{d\xi^k} \bigg|_{\xi=0}.
\end{equation}

Notice that the summation begins with 1 in equation (A1) and with 0 in equations (A2) and (A3). This is because the zero-order term, which is equal to 1 in equation (A1), has already explicitly taken into account in equation (28).
After substitution of series (A1) - (A3) into the moveout equation (28), calculation of the integrals, and performing exactly the same steps as those that have been done for equation (17), I obtain nonhyperbolic moveout equation in weakly LH TI media in the form

$$t^2(x) = t_0^2 + \frac{x^2}{V_{nmo}^2} - \frac{q x^4}{V_{nmo}^2 (t_0^2 V_{nmo}^2 + x^2)},$$  \hspace{1cm} (A7)

where \( t_0 \) is the two-way zero-offset traveltime,

$$V_{nmo}^2 = V_0^2 (1 + 2\delta_0 + \frac{t_0^2 V_0^2}{12} \delta_0^{'''}),$$ \hspace{1cm} (A8)

is the normal-moveout velocity,

$$q = 2\eta_0 + \frac{1}{12} t_0^2 V_{nmo}^2 (\delta_0^{'''} + \frac{V_0^{'''} V_{nmo}'}{V_{nmo}})$$
$$+ \frac{1}{960} t_0^4 V_{nmo}^4 V_0^{(IV)},$$ \hspace{1cm} (A9)

is the coefficient of the quartic term,

$$\eta_0 = \epsilon_0 = \delta_0,$$ \hspace{1cm} (A10)

and the superscripts \((\cdot)'''\) and \((\cdot)^{(IV)}\) denote the second and the fourth derivatives of corresponding functions with respect to the CMP-coordinate \( x \) at the midpoint. These derivatives are identical to derivatives (A4) - (A6) with respect to \( \xi \) because they are taken at the same point \( x = \xi = 0 \).

APPENDIX B: Calculation of the second derivative \( \delta''_0 \)

Equation (30) for the normal-moveout velocity can be linearized as

$$V_{nmo} = V_0 (1 + \delta_0 + \frac{t_0^2 V_0}{24} \delta_0^{'''}).$$ \hspace{1cm} (B1)

Substitution of equation (34) for the second derivative \( V_0^{'''} \) gives

$$V_{nmo} = V_0 (1 + \delta_0) - \frac{t_0^2 V_0^{'''} \delta_0^{'''}}{24}.$$ \hspace{1cm} (B2)

Taking the second derivative of equation (B2) and replacing \( V_0 \) with \( V_{nmo} \), in the linear approximation, I obtain

$$\delta_0^{'''} = -\frac{V_0^{'''}}{V_0} \frac{V_{nmo}^{'''}}{V_{nmo}} = \frac{1}{24} t_0^2 V_{nmo}^{(IV)}.$$ \hspace{1cm} (B3)

Applying equation (34) again, I finally find

$$\delta''_0 = \frac{t_0^2}{t_0} + \frac{V_{nmo}^{'''}}{V_{nmo}} = \frac{1}{24} t_0^2 V_{nmo}^{(IV)}. \hspace{1cm} (B4)$$
Feasibility of nonhyperbolic moveout inversion in transversely isotropic media

Vladimir Grechka
Center for Wave Phenomena
Colorado School of Mines

Ilya Tsvankin
Center for Wave Phenomena, Department of Geophysics
Colorado School of Mines

ABSTRACT

Inversion of reflection travel times in anisotropic media can provide estimates of anisotropic coefficients required for seismic processing and lithology discrimination. Nonhyperbolic P-wave moveout for vertical transverse isotropy (VTI media) is controlled by the parameter $\eta$ that is also responsible for the influence of anisotropy on all time-processing steps including dip-moveout (DMO) correction and time migration. Here, we have examined the feasibility of $\eta$ estimation using semblance analysis along analytically defined nonhyperbolic moveout curves on long CMP spreads.

Despite the high accuracy of the moveout approximation itself, the inverted value of $\eta$ is highly sensitive to small long-period errors in reflection travel times caused, for instance, by horizontal inhomogeneities or errors in the statics correction. While the normal-moveout velocity $V_{nmo}$ is well-determined by reflection moveout, small traveltime distortions on the order of 3-4 ms lead to errors in $\eta$ of about $\pm 0.1$ (80% or more), even in the simplest model of a single VTI layer. This ambiguity in the inversion for $\eta$ is caused by the insufficient magnitude of the nonhyperbolic moveout term for the P-wave that causes an interplay between the parameters $V_{nmo}$ and $\eta$. It should be mentioned that due to the character of this trade-off, the horizontal velocity is better constrained by long-spread moveout than $\eta$ (although not as tightly as the NMO velocity).

In vertically inhomogeneous media, interval values of $\eta$ can be obtained from a Dix-type differentiation scheme that involves the nonhyperbolic term of the moveout curve. This procedure, however, leads to a dramatic amplification of the already significant errors in $\eta$, especially for layers with a relatively small thickness. Therefore, although the inverted value of $\eta$ can be regarded as a crude measure of anisotropy above the reflector, it can hardly be used for interval parameter estimation or detailed lithology inversion.

Since all kinematically equivalent models obtained from nonhyperbolic moveout inversion provide an adequate description of long-spread reflection moveout, they can be used in poststack time migration. However, in most cases parameter estimation for imaging purposes can be accomplished by a much more stable dip-moveout (DMO) inversion method developed by Alkhalifah and Tsvankin (1995). Only in the case of near-vertical reflectors, which cannot be used in the DMO inversion, nonhyperbolic moveout provides a better way of building a starting model for anisotropic migration.

Key words: vertical transverse isotropy, nonhyperbolic reflection moveout, inversion
Introduction

Conventional hyperbolic moveout analysis, routinely used for building isotropic velocity models, does not provide enough information for velocity estimation in anisotropic media. If the medium is transversely isotropic with a vertical symmetry axis (VTI), the normal-moveout (NMO) velocities from horizontal reflectors of all three pure modes \((P, SV, SH)\) are different from the corresponding vertical velocities (Thomsen, 1986):

\[
V_{nmo}[P\text{-wave}] = V_{p0} \sqrt{1 + 2\delta},
\]

(1)

\[
V_{nmo}[SV\text{-wave}] = V_{s0} \sqrt{1 + 2\sigma},
\]

(2)

\[
V_{nmo}[SH\text{-wave}] = V_{s0} \sqrt{1 - 2\gamma},
\]

(3)

where \(V_{p0}\) and \(V_{s0}\) are the vertical \(P\)- and \(S\)-wave velocities, respectively, and \(\epsilon, \delta\) and \(\gamma\) are Thomsen (1986) anisotropy parameters. The coefficient \(\sigma \equiv \left(\frac{V_{s0}}{V_{p0}}\right)^2 (\epsilon - \delta)\) is the combination of the anisotropic parameters largely responsible for \(SV\)-wave signatures (Tsvankin and Thomsen, 1994).

>From equations (1)-(3) it is clear that NMO velocities and short-offset reflection moveout as a whole are not sufficient to distinguish between the true vertical velocities and anisotropic coefficients, even both \(P\) and shear data have been recorded. This explains the importance of using nonhyperbolic (long-spread) reflection moveout in the reconstruction of the anisotropic velocity field. In the presence of transverse isotropy, moveout becomes nonhyperbolic even in a single VTI layer, unless the anisotropy is elliptical. Tsvankin and Thomsen (1994) have developed a nonhyperbolic moveout equation that reduces to the exact quartic Taylor series for the \(r^2 - \sigma^2\) curve at moderate source-receiver offsets and converges at infinitely large offsets as well. This equation provides an accurate approximation for \(P\)-wave long-spread reflection moveout in horizontally stratified VTI media with no restrictions on the magnitude of the velocity variations. Although, as shown by Tsvankin and Thomsen (1994), kinematic signatures of \(P\)-waves in VTI media are governed by three parameters \((V_{p0}, \epsilon,\) and \(\delta\)), \(P\)-wave reflection moveout depends on just two combinations of them: the NMO velocity from a horizontal reflector \((V_{nmo})\) and the anisotropic coefficient \(\eta\) responsible for the nonhyperbolic portion of the moveout curve (Alkhalifah and Tsvankin, 1995):

\[
V_{nmo} = V_0 \sqrt{1 + 2\delta},
\]

(4)

\[
\eta \equiv \frac{1}{2} \left( \frac{V_{hor}^2}{V_{nmo}^2} - 1 \right) = \frac{\epsilon - \delta}{1 + 2\delta},
\]

(5)

* For brevity, the qualifiers in "quasi-\(P\)-wave" and "quasi-\(SV\)-wave" will be omitted.

Therefore, \(P\)-wave moveout data are not sufficient to resolve the true vertical velocity, no matter how large the maximum offset is. Tsvankin and Thomsen (1995) demonstrated that the only way to resolve the vertical velocity and anisotropic coefficients from reflection travel-times is to include long-spread moveout of the \(SV\)-wave in the inversion procedure.

However, Alkhalifah and Tsvankin (1995) proved that it is not necessary to know the individual values of the anisotropic parameters and the vertical velocity for \(P\)-wave time processing. All time-processing steps, including NMO, DMO, and time migration, are fully determined by the two parameters \((V_{nmo} and \eta)\) responsible for reflection moveout. Alkhalifah and Tsvankin (1995) suggested to obtain \(V_{nmo}\) and \(\eta\) from moveout velocities measured at two different dips. In the most common case, \(V_{nmo}\) can be found directly by conventional semblance analysis of horizontal events, which allows to determine \(\eta\) from the NMO velocity of a single dipping event. Although this methodology provides a stable way of estimating the parameter \(\eta\), it requires the presence of dipping reflectors (such as fault planes) in the whole depth range of interest and cannot be used for dips exceeding 75-80°.

Therefore, it is important to investigate the possibility of obtaining \(\eta\) using long-spread \(P\)-wave moveout from horizontal reflectors. Alkhalifah (1996) suggested to determine \(V_{nmo}\) and \(\eta\) (or \(V_{nmo}\) and \(V_{hor}\)) by a two-dimensional semblance scan based on the nonhyperbolic moveout equation described above. He showed that this inversion procedure gives reasonably accurate results for data acquired over a horizontally homogeneous VTI layer. However, he also found that the extracted values of \(\eta\) are sensitive to small errors in \(V_{nmo}\), even if the spreadlength is twice as large as the reflector depth. This trade-off stems from the ambiguity in the determination of the quartic \(P\)-wave moveout term that was first noticed by Tsvankin and Thomsen (1995).

Here, we investigate the feasibility of \(\eta\) estimation using long-spread \(P\)-wave moveout data acquired in the presence of random and correlated noise. We start with an analytic study of the trade-off between \(V_{nmo}\) and \(\eta\) and proceed with numerical examples illustrating the influence of realistic errors in travel-times on the inverted values of \(\eta\). Our analysis shows that \(\eta\) is extremely sensitive to low-frequency variations in the travel-times that correspond to small deviations of the medium from the homogeneous VTI model assumed in the inversion procedure.
Linear analysis of the problem

Both in the analytic and numerical parts of this work we have extensively used the nonhyperbolic equation for reflection moveout \( t^2(x^2) \) in VTI media suggested by Tsvankinim and Thomsen (1994). Some of the numerical results below, however, are obtained by anisotropic ray tracing. Long-spread P-wave moveout in a single horizontal VTI layer can be represented in the following way (Alkhalifah and Tsvankinim, 1995):

\[
t^2(x) = t_0^2 + \frac{x^2}{V_{nmo}^2} - \frac{2\eta x^4}{V_{nmo}^2 t_0^2 V_{nmo}^2 + (1 + 2\eta) x^2},
\]

where \( t_0 \) is the two-way vertical traveltme and \( x \) is the source-receiver offset. The only difference between equation (6) and the original expression developed by Tsvankinim and Thomsen (1994) is that the contribution of \( V_{so} \) to the nonhyperbolic term is ignored (also, the generic coefficients \( \epsilon \) and \( \delta \) were replaced with \( \eta \)). To study the dependence of the traveltime \( t(x) \) on the parameters \( t_0, V_{nmo}, \) and \( \eta \), we calculate the normalized Frechét derivatives from equation (6):

\[
D_t \equiv \frac{t_0}{t(x)} \frac{\partial t(x)}{\partial t_0} = \frac{1}{t^2(x)} \left[ 1 + \frac{2\eta x^4}{(t_0^2 V_{nmo}^2 + (1 + 2\eta) x^2)^2} \right],
\]

\[
D_V \equiv \frac{V_{nmo}}{t(x)} \frac{\partial t(x)}{\partial V_{nmo}} = \frac{-x^2}{t^2(x) V_{nmo}^2} \left[ 1 - \frac{2\eta x^2 (2t_0^2 V_{nmo}^2 + x^2)}{(t_0^2 V_{nmo}^2 + (1 + 2\eta) x^2)^2} \right],
\]

and

\[
D_\eta \equiv \eta \frac{\partial t(x)}{t(x)} \frac{\partial t(x)}{\partial \eta} = \frac{-\eta x^4 (t_0^2 V_{nmo}^2 + x^2)}{t^2(x) V_{nmo}^2 (t_0^2 V_{nmo}^2 + (1 + 2\eta) x^2)^2}.
\]

The derivatives (7)-(9) represent the relative (percentage) variations in the traveltime \( t(x) \) caused by normalized perturbations in the parameters \( t_0, V_{nmo}, \) and \( \eta \), respectively. Figure 1 displays these derivatives for horizontal offsets up to \( x = 2h \) (\( h \) is the reflector depth). Clearly, the vertical traveltime \( t_0 \) is the best-defined of the three parameters, while \( \eta \) has the least influence on the traveltimes, even at large offsets. Figure 1 also shows that the derivatives \( D_V \) and \( D_\eta \) have a similar character and, therefore, \( V_{nmo} \) and \( \eta \) may, to some extent, compensate each other's influence on the traveltimes. This possible trade-off between \( V_{nmo} \) and \( \eta \) can also be seen from the moveout equation (6) itself. The absolute values of the derivative \( D_V \) in Figure 1, however, are much greater than the values of the derivative \( D_\eta \). This suggests that variations in \( t(x) \) due to perturbations in \( \eta \) may be compensated by much smaller perturbations in \( V_{nmo} \) of the opposite sign. It is obvious from Figure 1 that \( \eta \) is the least-constrained parameter of the moveout curve, even on relatively long spreads feasible for reflection surveys.

Although the analysis based on the Frechét derivatives provides some insight into the possibility of resolving \( V_{nmo} \) and \( \eta \) from the reflection traveltimes, it has two shortcomings. First, because of the linear nature of the Frechét derivatives, this analysis is limited to small perturbations of \( V_{nmo} \) and \( \eta \) from chosen initial values. Another disadvantage is that the Frechét derivatives vary with the initial values of \( V_{nmo} \) and \( \eta \). For instance, if we perturb a purely isotropic (or elliptically anisotropic) model, this approach breaks down because \( D_\eta = 0 \) [equation (9)], and it is necessary to take the second derivative of the traveltime into account.
Figure 2. The influence of $V_{nmo}$ and $\eta$ on the $P$-wave reflection traveltimes in a horizontal VTI layer. The contours display the maximum difference $\Delta t$ (in ms) between the traveltimes for a reference model with $t_0 = 1.0$ s, $V_{nmo} = 2.0$ km/s, and $\eta = 0.1$ ($V_o = 2.0$ km/s, $\epsilon = 0.1$, $\delta = 0$, and the reflector depth $h = 1$ km) and models with the same $t_0$ but different $V_{nmo}$ and $\eta$ shown on the axes; the spreadlength $s_{max} = 2$ km ($s_{max}/h = 2$). The moveout for the reference model is calculated exactly using ray tracing, while the traveltimes for all other models are computed from equation (6).

Reflection moveout as a function of $V_{nmo}$ and $\eta$: numerical study

To obtain a quantitative picture of the influence of $V_{nmo}$ and $\eta$ on long-spread moveout, we continue with a numerical analysis of the nonhyperbolic moveout equation (6).

Figure 2 shows the dependence of the reflection moveout on variations in $V_{nmo}$ and $\eta$ for a model that has a typical moderate positive value of $\eta$. Similar plots illustrating the sensitivity of $P$-wave moveout to the model parameters were given in Tsukanov and Thomsen (1994). However, Tsukanov and Thomsen had been mostly interested in the moveout inversion for the vertical velocity and had not recast the moveout in terms of the coefficient $\eta$ that was introduced later by Alkalifah and Tsukanov (1995).

Two features of the contours of the maximum traveltime differences $\Delta t$ in Figure 2 should be noted. First, the contours form an elongated valley in the coordinates $(V_{nmo}, \eta)$, along which the reflection moveout remains almost independent of the two parameters. While this valley corresponds to a relatively narrow range of NMO velocities, it spans a wide range of $\eta$ values. Apparently, a small change in $V_{nmo}$ can be compensated by a much more pronounced variation in $\eta$ of the opposite sign without changing the traveltimes in a measurable way.

Second, the center of the contours $\Delta t_{max}$ (corresponding to the minimum of $\Delta t_{max}$ equal to 0.45 ms in Figure 2) is slightly shifted from the correct position $(V_{nmo} = 2.0$ km/s, $\eta = 0.1)$ due to small deviations of the moveout approximation (6) from the exact raytraced traveltime curve for the reference model. Equation (6) was designed by Tsukanov and Thomsen (1994) to ensure the correct behavior of the moveout for infinitely large offsets ($x/h \rightarrow \infty$). At intermediate offsets ($x/h \approx 2$), however, this moveout approximation can be somewhat improved by empirically changing the coefficient in the denominator of the nonhyperbolic term. We found that the coefficient $2$ in front of $\eta$ in the $x^2$ term should be replaced by $3.5$ to minimize deviations from the exact moveout up to the offset $x = 2h$. This modification is equivalent to changing the horizontal velocity in the model, while keeping the correct values of the quadratic and quartic moveout terms. Thus, the version of equation (6) that we will use in the rest of the paper is as follows:

$$t^2(x) = t_0^2 + \frac{x^2}{V_{nmo}^2} - \frac{2\eta x^2}{V_{nmo}^2 t_0^2 V_{nmo} + (1 + 3.5\eta) x^4}. \quad (10)$$

Figure 3 shows the contours similar to the ones in Figure 2 but now calculated using the modified approximation (10) for three different spreadlength-to-depth ratios. Evidently, the correction in the traveltime equation has moved the center of the contours $\Delta t_{max}$ in Figure 3 to the correct position corresponding to $V_{nmo} = 2.0$ km/s and $\eta = 0.1$. The smallest traveltime residuals ($\Delta t_{max}$) at the center of the contours are equal to 0.11 ms, 0.26 ms, and 1.27 ms in Figures 3a, 3b, and 3c, respectively.

As expected, Figure 3 indicates that $\eta$ becomes better constrained with the increasing maximum offset. As $x_{max}/h$ grows, the contours corresponding to a given maximum time residual become tighter, keeping $\eta$ (as well as $V_{nmo}$) closer to the correct value. Also, for long spreads the contours tilt toward the $V_{nmo}$-axis making the variation in $\eta$ for a fixed $\Delta t_{max}$ less pronounced. The meaning of this result becomes clear if we recall that $\eta$ is responsible for the nonhyperbolic moveout term that makes a significant contribution to the traveltime only at relatively large offsets exceeding the reflector depth. Evidently, the inversion of nonhyperbolic moveout for $\eta$ becomes more stable with increasing spreadlength (e.g., Alkalifah, 1990). However, we will restrict the maximum offset in the subsequent examples by $x_{max} = 2h$
since it is seldom possible to use longer spreads in practice.

In addition to analyzing the maximum traveltime differences, it is instructive to examine the moveout curves on the whole CMP spread. Figure 4 shows the ray-traced traveltime curves for the models labeled by the letters A, B, and C in Figure 3b. Despite a significant variation in \( \eta \) (from 0.04 to 0.17) among the three models, the maximum values of the traveltime differences \( (\Delta t_{BA} = t_B - t_A \) and \( \Delta t_{BC} = t_B - t_C \) picked from Figure 3b are less than 4 ms. Figure 4, generated using an exact ray-tracing algorithm, confirms this result: the moveouts \( t_A, t_B, \) and \( t_C \) are very close to each other at all offsets with the maximum deviations \( \max(\Delta t_{BA}) = 2.9 \) ms and \( \max(\Delta t_{BC}) = 3.1 \) ms. Note that the maximum separation between the curves occurs at the intermediate offset equal to the reflector depth rather than at the maximum offset \( z_{\text{max}} \).

The conclusions drawn from Figure 4 have important consequences for evaluating the feasibility of nonhyperbolic moveout inversion. Since the moveout curves corresponding to a wide range of \( \eta \) values are almost indistinguishable from each other at all offsets, the inversion for \( \eta \) can hardly be expected to be stable. Even relatively small traveltime errors may result in significant deviations of \( \eta \) from the correct value. As pointed out by Tsvankin and Thomsen (1995), the magnitude of the nonhyperbolic moveout term for P-waves may not be sufficient (even on these relatively long spreads) to obtain the quartic moveout coefficient (and hence \( \eta \)) with acceptable accuracy. Below, we study the influence of different kinds of traveltime errors on the actual inversion of long-spread moveout for \( V_{\text{rms}} \) and \( \eta \).

**Inversion of nonhyperbolic moveout**

If reflection traveltimes on a certain spreadlength are known, the parameters \( t_0, V_{\text{rms}}, \) and \( \eta \) can be obtained, for instance, by least-squares fitting of equation (10) to the moveout curve (Tsvankin and Thomsen, 1994). However, due to the presence of noise and the need to automate data processing, picking of reflection traveltimes on CMP data is seldom done in practice. Instead, the best-fit moveout (stacking) velocity and the corresponding hyperbolic moveout curve are usually found from semblance analysis on CMP gathers with the spreadlength close to the reflector depth (Tanner and Koehler, 1969; Neidell and Tanner, 1971). Similarly, the inversion for \( \eta \) can be accomplished by semblance analysis along nonhyperbolic moveout curves described by equation (6) or (10) (Alkhalifah, 1996). For a given value of the vertical traveltime \( t_0 \), we need to carry out a two-
Figure 4. (a) The exact reflection moveouts ($t_A$, $t_B$, and $t_C$) for the models marked by A, B, and C in Figure 3b. The model parameters are: $V_{nmo} = 1.95 \text{ km/s}$, $\eta = 0.17$ (Model A, gray curve), $V_{nmo} = 2.0 \text{ km/s}$, $\eta = 0.1$ (Model B, solid black), and $V_{nmo} = 2.05 \text{ km/s}$, $\eta = 0.04$ (Model C, dashed); (b) The traveltimes differences $\Delta t_{BA} = t_B - t_A$ (gray dots) and $\Delta t_{BC} = t_B - t_C$ (black dots).

Figure 5. (a) A synthetic seismogram of the $P$-wave reflected from the bottom of a VTI layer with the parameters $V_0 = 2 \text{ km/s}$, $\epsilon = 0.1$, $\delta = 0$, and $h = 1 \text{ km}$. The source pulse is a Ricker wavelet with the central frequency 40 Hz. (b) The semblance contours are calculated using equation (10) for $t_0 = 1 \text{ s}$. 
dimensional semblance scan over \( V_{nm0} \) and \( \eta \) to find a model that yields the maximum value of semblance. We have used the following equation to calculate the semblance coefficient:

\[
S(t_0, V_{nm0}, \eta) = \frac{\sum_{t_0}^{T + t_0} \sum_{t_0}^{T - t_0} [\sum_{x = x_{\text{min}}}^{x_{\text{max}}} A(x, t)]^2}{M \sum_{t_0}^{T + t_0} \sum_{t_0}^{T - t_0} \sum_{x = x_{\text{min}}}^{x_{\text{max}}} A^2(x, t)}, \quad (11)
\]

The summation of the amplitudes \( A(x, t) \) and their squares \( A^2(x, t) \) in equation (11) is performed along the moveout curves \( t(t_0, V_{nm0}, \eta, x) \) [described by equation (10)] originating at the vertical traveltimes \( t_0 \) within a smoothing window \( T \) centered at time \( t_0 \). In all examples below, we have used 50 equally spaced traces \( (M = 50) \) with the maximum offset \( x_{\text{max}} = 2h = 2.0 \text{ km} \). The smoothing window was \( T = 32 \text{ ms} \), which corresponds to 8 time samples. The amplitudes \( A(x, t) \) between time samples were obtained by linear interpolation.

The results of this nonhyperbolic semblance search for a typical single-layer VTI model are displayed in Figure 5. The maximum semblance in this example is equal to 0.87 and corresponds to the correct values of \( V_{nm0} \) and \( \eta \). The shapes of the semblance contours in Figure 5b and the contours of the maximum traveltime difference in Figure 3b look quite similar. This can be expected since models with poorly resolved traveltimes curves should yield close values of semblance. Clearly, \( V_{nm0} \) and \( \eta \) can simultaneously vary along the diagonal “valley” in Figure 5b without producing significant changes in the value of semblance.

To study the stability of the nonhyperbolic moveout inversion, we have added an error \( \tau(x) \) to the reflection traveltimes \( t(x) \) from Figure 5 (without changing the amplitudes) and performed the two-dimensional semblance analysis of the modified seismogram. The function \( \tau(x) \) was chosen in the form

\[
\tau(x) = \tau_0 \sin \frac{2\pi k}{x_{\text{max}}} x, \quad (12)
\]

where \( x_{\text{max}}/k \) is the period of the error function, and \( \tau_0 = 3 \text{ ms} \). Although the magnitude of the traveltime error \( \tau_0 \) is extremely small, it can be expected to cause errors in \( \eta \) of about 50\% (Figure 4).

Figure 6 shows the results of the nonhyperbolic semblance analysis for a range of \( k \) values, as well as for a completely random traveltime error.

The first example (Figure 6a) was computed for a traveltimes error \( \tau(x) \) with the period four times greater than the maximum offset \( x_{\text{max}} \) (\( k = 1/4 \)). In this case, the error \( \tau(x) \) grows monotonically with \( x \), and the distorted traveltime \( t(x) + \tau(x) \) is slightly larger than \( t(x) \) at all offsets. Naturally, such an error leads to a higher value of \( t_0 \) and a lower \( V_{nm0} \) of the “best-fit” model corresponding to the maximum of the semblance function. In this particular example, the best-fit values are \( t_0 = 1.012 \text{ s} \) (instead of the correct value \( t_0 = 1.0 \text{ s} \)), \( V_{nm0} = 1.97 \text{ km/s} \) (instead of \( V_{nm0} = 2.0 \text{ km/s} \), and \( \eta = 0.12 \) (instead of \( \eta = 0.1 \)). Note, that the semblance maximum moves along the diagonal valley in Figure 5b, with the increase in \( \eta \) compensated by the smaller value of \( V_{nm0} \).

The period of the traveltime error \( \tau(x) \) in Figure 6b is twice as large as \( x_{\text{max}} \), making the traveltime \( t(x) + \tau(x) \) close to the reflection moveout for Model A from Figure 3 (see the gray curve in Figure 4a). Although the magnitude of the traveltime error remains extremely small (3 ms), the semblance maximum moves substantially along the \( \eta \) axis yielding the best-fit value of \( \eta = 0.18 \), which is far different from the correct \( \eta = 0.1 \). The influence of this pronounced increase in \( \eta \) on reflection moveout has been practically canceled out by a much smaller (about 3\%) decrease in \( V_{nm0} \) with respect to the actual value of 2 km/s. The value of \( t_0 \) corresponding to the maximum of semblance in this example is almost equal to the correct \( t_0 = 1.0 \text{ s} \).

As expected from the proximity of the traveltime curves, the pair \( (V_{nm0}, \eta) \) corresponding to the semblance maximum in Figure 6b practically coincides with the parameters of Model A from Figure 3 (\( \eta = 0.17 \)). Evidently, long-period traveltime errors of the magnitude that can be considered as insignificant in the practice of data processing have a drastic influence on the inverted value of \( \eta \). Time distortions of this character can be easily caused by a small horizontal heterogeneity in the velocity field, long-period errors in the statics correction etc.

Figure 6c shows the result of the semblance analysis with the error function \( \tau(x) \) reversed compared to that in the previous plot (e.g., \( \tau(x) \) has the same period and magnitude as in Figure 6b but the opposite sign). In this case, the traveltime \( t(x) - \tau(x) \) is close to that for Model C from Figure 3 (see the dashed curve in Figure 4a), and the coordinates of the semblance maximum \( (t_0 = 0.998 \text{ s}, V_{nm0} = 2.04 \text{ km/s}, \eta = 0.05) \) are close to those of point \( C \) in Figure 3b \( (V_{nm0} = 2.05 \text{ km/s}, \eta = 0.04) \). Again, it should be emphasized that low-frequency noise with a magnitude of just 3 ms reduces the inverted value of \( \eta \) in half.

Further increase in the frequency of the noise trace \( \tau(x) \) (Figure 6d) leads to the deviation of the semblance maximum in the opposite direction compared to what we observed in Figure 6b, where we had a lower-frequency error with the same sign. For the example in Figure 6d, the period of the traveltime error is equal to the maximum offset \( x_{\text{max}} \), and the noisy trace has larger traveltimes at small offsets and smaller — at offsets exceeding \( h = x_{\text{max}}/2 \). The explanation of the inversion result in
Figure 6. The results of semblance analysis for the model from Figure 3 after the addition of a traveltime error [equation (12)] with $\tau_0 = 3$ ms and different normalized periods. (a) $k = 1/4$; (b) $k = 1/2$; (c) $k = 1/2$ but the sign of the error is opposite ($\tau_0 = -3$ ms); (d) $k = 1$; (e) $k = 2$. (f) random traveltime error with the zero mean and a maximum magnitude of 4 ms.
this case is more complicated than for those in Figures 6a – 6c since it involves a trade-off between all three move-
out parameters: $t_0$, $V_{nmo}$, and $\eta$. The semblance algorithm tries to provide the best possible fit to the distorted move-
out at small offsets by increasing the vertical time (the best-fit $t_0 = 1.014$ s instead of the correct $t_0 = 1.0$ s).
Since $t_0$ became greater, $V_{nmo}$ must also increase (from 2 km/s to 2.04 km/s) to yield smaller times at inter-
mediate and large offsets. (Figure 1 explains this character-
of the trade-off between $t_0$ and $V_{nmo}$: the derivatives of
$t(x)$ with respect to these parameters are of the op-
posite sign.) A higher value of $V_{nmo}$, as before, leads to a
decrease in $\eta$ from the correct value $\eta = 0.1$ down to
0.06.

For the value $k = 2$ in Figure 6e, the spreadlength
comprises two periods of the error function $\tau(x)$. Since
the nonhyperbolic moveout described by equation (10) is
roughly equivalent (in terms of the spatial frequency) to
the fourth-order polynomial in the offset $x$, it has a lower
spatial frequency than the traveltime function $t(x) = \tau(x)$.
Therefore, no changes in $V_{nmo}$ and $\eta$ can provide an ac-
ccurate fit to the moveout curve, and equation (10) per-
forms some kind of averaging of the traveltimes. This
leads to a decrease in the value of the maximum sem-
b lance compared to the three previous examples, but the
best-fit model in Figure 6e ($t_0 = 1.0$ s, $V_{nmo} = 1.98$ km/s
and $\eta = 0.12$) is relatively close to the exact solution.

Finally, semblance results for a purely random
traveltime error are shown in Figure 6f. In this case, due
to the insensitivity of the shape of the semblance objec-
tive function to random errors, the position of the sem-
b lance maximum provides an accurate estimate of both
parameters, although the value of semblance decreases
to 0.62.

Our error estimates made for a model with $\eta = 0.1$
remain valid for the whole range of plausible $\eta$ values.
For instance, Figure 7 shows the semblance contours for
a model with a higher value of $\eta = 0.2$. In the absence
of traveltime errors (Figure 7a), the semblance analysis
yields the correct model parameters $V_{nmo} = 2.0$ km/s,
$\eta = 0.2$. However, a small low-frequency error [equa-
tion (12)] with a magnitude of 3 ms shifts the location of
the semblance maximum to the point $V_{nmo} = 1.94$ km/s,
$\eta = 0.3$ (the best-fit $t_0 = 0.998$ s in both cases).

The results of semblance analysis presented in Fig-
ures 6 and 7 confirm the conclusion drawn from the study
of reflection traveltimes (Figures 3 and 4) that time vari-
ations on the order of several first milliseconds may cause
errors in $\eta$ of about 50% and more. Figures 6 and 7 indi-
cate one more interesting feature of nonhyperbolic mo-
veout inversion. If we fix $\eta$ at the correct value and ex-
amine the semblance variation along the horizontal lines

![Figure 7](image-url). The semblance contours for a model with the same $t_0 = 1$ s and $V_{nmo} = 2$ km/s as in Figure 6, but for $\eta = 0.2$.
(a) without traveltime errors (b) for the traveltime error (12) with the magnitude $t_0 = 3$ ms and normalized period $k = 1/2$. 
\( \eta = 0.1 \) in Figure 6 and \( \eta = 0.2 \) in Figure 7, we notice that the maximum semblance lies very close (within 1\%) of the correct value of \( V_{nmo} \). This means that the errors discussed above are entirely due to the trade-off between \( V_{nmo} \) and \( \eta \) and are unlikely to appear in hyperbolic semblance analysis on conventional-length spreads.

The character of the trade-off between \( V_{nmo} \) and \( \eta \) suggests that the horizontal velocity may be better constrained by long-spread moveout than \( \eta \). Indeed, \( V_{hor} \) can be expressed through \( V_{nmo} \) and \( \eta \) as [equation (3)]

\[
V_{hor} = V_{nmo} \sqrt{1 + 2\eta}.
\] (13)

Since within the family of kinematically equivalent models higher values of \( V_{nmo} \) are compensated by lower values of \( \eta \), estimates of the horizontal velocity can be expected to be more stable than those of \( \eta \).

Figure 8 shows the contours from Figure 7, but this time plotted as functions of the horizontal and NMO velocity. Again, \( V_{nmo} \) and \( V_{hor} \) can simultaneously vary along an elongated ”valley” in the contours without substantially changing the value of semblance. In the absence of the traveltime errors (Figure 8a) the maximum semblance is located at the correct point \( V_{nmo} = 2.0 \) km/s and \( V_{hor} = 2.36 \) km/s. If the error of magnitude 3 ms with the normalized period \( k = 1/2 \) is added to the traveltimes, the semblance maximum shifts to \( V_{nmo} = 1.94 \) km/s and \( V_{hor} = 2.45 \) km/s, which corresponds to errors of 3.0\% in \( V_{nmo} \) and 3.8\% in \( V_{hor} \). Therefore, the horizontal velocity is better constrained by long-spread moveout than \( \eta \) but less tightly than the NMO velocity.

Errors of nonhyperbolic velocity analysis in layered media

The discussion above was limited to the simplest single-layer model. For more realistic horizontally layered models, Tsvankin and Thomsen (1994) have introduced a Dix-type differentiation procedure designed to estimate interval values of the quartic moveout term. Their approach, expressed in terms of the parameter \( \eta \) (Appendix A), can be summarized as follows:

1) Using long-spread moveout from the the \( i \)-th (\( i = 1, \ldots, N \)) interface, find \( t_0(i) \) and the effective values of \( V_{nmo}(i) \) and \( \eta(i) \) from traveltime fitting or semblance analysis based on equation (10). (Clearly, for the most shallow boundary (\( i = 1 \)), the effective \( V_{nmo}(1) \) and \( \eta(1) \) are equal to their interval values \( V_{nmo,1} \) and \( \eta_1 \).)

2) Determine the interval NMO velocities in any layer \( i = 2, \ldots, N \) from the Dix (1955) equation

\[
V_{nmo,i}^2 = \frac{V_{nmo}^2(i) t_0(i) - V_{nmo}^2(i - 1) t_0(i - 1)}{t_0(i) - t_0(i - 1)}.
\] (14)

3) Calculate from equation (A9) the quantities

\[ V_{nmo}, V_{hor} \].

Figure 8. The same contours as in Figure 7 but plotted in coordinates \((V_{nmo}, V_{hor})\).
\[ f(i) = V_{nmo}^2(i)[1 + 8\eta(i)]. \]  

(15)

4) Compute the interval values of \( \eta \) from equation (A10):

\[ \eta_i = \frac{1}{8V_{nmo,i}^2} \left[ \frac{f(i)t_0(i) - f(i-1)t_0(i-1)}{t_0(i) - t_0(i-1)} - V_{nmo}^2(i) \right]. \]  

(16)

To illustrate errors in the estimation of the interval values of \( \eta \), we consider a model composed of two identical layers with \( V_{nmo}(1) = V_{nmo}(2) = 2.0 \text{ km/s} \) and \( \eta(1) = \eta(2) = 0.1 \) (the reflection from the first interface may be assumed to be caused by a jump in the density). The vertical traveltimes for the two reflections are \( t_0(1) = 0.8 \text{ s} \) and \( t_0(2) = 1 \text{ s} \). If there are no traveltime errors and we start with the correct values of the parameters for both interfaces, equations (14) – (16) yield the correct interval values \( V_{nmo,2} = 2.0 \text{ km/s} \) and \( \eta_2 = 0.1 \). If, because of the slightly erroneous traveltimes (see Figure 6b), we obtain for the reflection from the deeper boundary \( t_0(2) = 0.998 \text{ s} \), \( V_{nmo}(2) = 1.94 \text{ km/s} \), \( \eta(2) = 0.18 \), equations (14) – (16) give highly inaccurate values of \( V_{nmo,2} = 1.67 \text{ km/s} \) and, especially, \( \eta_2 = 0.79 \).

Therefore, Dix-type differentiation of nonhyperbolic moveout leads to a dramatic amplification of errors in \( \eta \) that become more severe with decreasing thickness of the layer of interest.

### Discussion and conclusions

Long-spread P-wave moveout from horizontal reflectors in VTI media depends on the vertical traveltime \( t_0 \), normal-moveout velocity \( V_{nmo} \), and the “anellipticity” parameter \( \eta \). While \( V_{nmo} \) controls hyperbolic moveout on conventional-length spreads, the coefficient \( \eta \) is responsible for the nonhyperbolic portion of the moveout curve at large offsets that exceed the reflector depth. Inversion of long-spread P-wave moveout for \( \eta \) represents the only way of obtaining information about anisotropy from P-wave reflection traveltimes in horizontally-homogeneous VTI media. The importance of recovering \( \eta \) from surface data cannot be overestimated since this parameter (in combination with \( V_{nmo} \)) is sufficient to perform all time-processing steps for P-waves in VTI media, including NMO correction, dip-moveout (DMO) removal, and time migration (Alkhalifah and Tsvankin, 1995). Also, \( \eta \) represents a potentially powerful lithology indicator that can be used, for instance, to discriminate between shales and sands.

Here, we have studied the possibility of recovering \( \eta \) using nonhyperbolic semblance analysis of P-wave reflection data based on the analytic approximation for long-spread moveout suggested by Tsvankin and Thomsen (1994). Although this approximation is close to the exact reflection traveltimes on spreadlengths feasible for reflection surveys, we have improved its accuracy even further by slightly modifying the denominator of the nonhyperbolic term. The high quality of the moveout equation, however, turned out to be insufficient for reliable determination of \( \eta \) from P-wave moveout data, even in the simplest model of a single VTI layer. In agreement with the results of Tsvankin and Thomsen (1995), we have found a family of kinematically equivalent VTI models with relatively close values of the NMO velocity and significantly different values of \( \eta \). In one of our model examples, we have increased the value of \( \eta \) from 0.1 to 0.17 and reduced the NMO velocity by 3% without changing the traveltimes by more than 3 ms at any offset (the spreadlength was twice as large as the reflector depth).

To confirm this observation, we have carried out actual inversion for \( V_{nmo} \) and \( \eta \) by nonhyperbolic semblance analysis of reflections from the bottom of a single horizontal VTI layer. For noise-free data, the maximum semblance corresponds to the actual values of both parameters, although the difference in semblance within the family of equivalent models is very small. Addition of random traveltime errors reduces the value of semblance but does not move the semblance maximum from the correct values of \( V_{nmo} \) and \( \eta \). This result is quite understandable since suppression of random noise is a well-known feature of semblance velocity analysis. However, the nonhyperbolic semblance inversion breaks down in the presence of long-period traveltime errors, even if the magnitude of these errors is extremely small. For instance, the error function with the spatial half-period equal to the maximum offset and the magnitude of 3 ms causes an 80% (0.08) distortion in the value of \( \eta \). Long-period errors of such a small magnitude can be caused, for instance, by insignificant horizontal heterogeneities or errors in the statics correction. It should be emphasized that since the nonhyperbolic moveout inversion is carried out on long spreads (twice as large as the reflector depth), it is much more influenced by lateral heterogeneity than conventional hyperbolic velocity analysis.

We conclude that the magnitude of P-wave nonhyperbolic moveout on spreads feasible for reflection seismic studies is not large enough to overcome the trade-off between \( V_{nmo} \) and \( \eta \). It should be emphasized that \( V_{nmo} \) remains well-constrained by P-wave traveltime data, even in the presence of significant anisotropy, with errors not exceeding 3% in all our examples. However, the inversion for \( \eta \) in this simplest single-layer model is rather unstable, with error bars easily reaching \( \pm 0.1 \). Alkhalifah (1996) emphasizes that semblance velocity analysis is less sensitive to errors in reflection traveltimes than purely kinematic (e.g., least-squares) inversion al-
Author/Editor: V. Grechka & I. Tsvankin

Page: 156

The text continues as follows:

...moveout inversion in this case may be used to build a starting model for anisotropic migration.

Another potential application of nonhyperbolic moveout inversion is determination of Thomson's parameter $\varepsilon$ from reflection data in the case when the vertical velocity is known (e.g., from check-shots or well-logs). Large errors in $\eta$ stem from the interplay between the horizontal and NMO velocity in the moveout inversion, despite the fact that both velocity values are relatively well-constrained by nonhyperbolic moveout. Since the vertical velocity is not influenced by this trade-off, it can be combined with the horizontal velocity recovered from long-spread moveout to provide an estimate of $\varepsilon$. Of course, interval values of $\varepsilon$ may be substantially distorted due to the amplification of errors in the Dix-type differentiation procedure.

Acknowledgments

We are grateful to the members of the A(nisotropy)-team of CWP and especially to Tariq Alkhalifah for many helpful discussions. The support of this work was provided by the members of the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP, Colorado School of Mines, and by the United States Department of Energy (project "Velocity Analysis, Parameter Estimation, and Constraints on Lithology for Transversely Isotropic Sediments") within the framework of the Advanced Computational Technology Initiative.

References

Alkhalifah, T. 1996, Velocity analysis using nonhyperbolic moveout in transversely isotropic media. CWP-203, this volume.


APPENDIX A: Estimation of interval η in layered media

Tsankin and Thomsen (1994) showed that P-wave reflection traveltimes in vertically inhomogeneous VTI media are well-approximated by the single-layer expression (6), if all moveout coefficients are replaced by the effective quantities that reflect the influence of layering:

\[
\begin{align*}
\frac{t^2(x, i)}{t_0^2(i)} &= A_2(i) + A_4(i)\frac{x^2}{1 + A^*(i)z^2}, \\
A_2(i) &= \frac{1}{V_{nmo}^2(i)}, \\
A^*(i) &= \frac{A_4(i)}{V_{nmo}^2(i) - A_2(i)}, \\
V_{hor}(i) &= V_{nmo}(i)\sqrt{1 + 2\eta(i)},
\end{align*}
\]

where the vertical traveltimes and the parameters \(A_2(i), A_4(i),\) and \(A^*(i)\) are calculated for the stack of layers above the \(i\)-th interface.

The quadratic coefficient \(A_2(i)\) is reciprocal to the squared NMO velocity

\[
A_2(i) = \frac{1}{V_{nmo}^2(i)},
\]

and \(A^*(i)\) is related to the effective quartic moveout coefficient \(A_4(i)\) by

\[
A^*(i) = \frac{A_4(i)}{V_{nmo}^2(i) - A_2(i)};
\]

\(V_{hor}(i)\) is the horizontal velocity. Equations (A1)–(A3) are identical to the corresponding expressions for a homogeneous VTI medium.

The NMO velocity \(V_{nmo}(i)\) in a layered VTI medium is given by the conventional Dix (1955) equation, while \(A_4(i)\) can be obtained from an exact averaging formula that takes into account both anisotropy and vertical inhomogeneity (Hake et al., 1984). It is convenient to express \(A_4(i)\) through \(\eta\) and NMO velocity in the same way as in a single VTI layer (Alkhalifah, 1996):

\[
A_4(i) = -\frac{2\eta(i)}{t_0^2(i)V_{nmo}^2(i)},
\]

where \(\eta(i)\) now represents an effective parameter that absorbs the influence of anisotropy and layering. An explicit expression for \(\eta(i)\) can be obtained from the equation for \(A_4(i)\) given in Hake et al. (1984) and Tsankin and Thomsen (1994).

Since the meaning of the horizontal velocity in stratified media is not quite clear [see the discussion in Tsankin and Thomsen (1994) and Alkhalifah (1996)], we assume that we can simply use equation (13) valid for homogeneous media:

\[
V_{hor}(i) = V_{nmo}(i)\sqrt{1 + 2\eta(i)},
\]

where \(V_{nmo}(i)\) and \(\eta(i)\) are the effective quantities defined above. Alkhalifah (1996) shows this approximation for the horizontal velocity does not reduce the accuracy of the moveout equation.

Since \(A_4(i)\) and \(V_{hor}(i)\) are now related to \(V_{nmo}(i)\) and \(\eta(i)\) by the same equations as in a single VTI layer, we can rewrite equation (A1) in the form of equation (6):

\[
\begin{align*}
\frac{t^2(x, i)}{t_0^2(i)} &= A_2(i) + \frac{x^2}{V_{nmo}^2(i)} \left[ \eta(i) \frac{x^2}{1 + A^*(i)z^2} \right],
\end{align*}
\]

To make the moveout equation even more accurate, we replace the coefficient 2 in the denominator of the non-hyperbolic term by 3.5, as in equation (10), and apply equation (A6) to determine the effective quantities \(V_{nmo}(i)\) and \(\eta(i)\) through nonhyperbolic semblance analysis.

The interval NMO velocity is then recovered from the effective \(V_{nmo}(i)\) using the conventional Dix formula. Next, we obtain the interval values of \(\eta\) by applying the Dix-type equations presented by Tsankin and Thomsen (1994). For the interval \(A_{4,i}\) we have

\[
A_{4,i} = -\frac{H_i}{V_{nmo}^3(i)}
\]

where \(H_i\) is defined through the following Dix-type equation:

\[
H_i = \frac{f(i) t_0(i) - f(i - 1) t_0(i - 1)}{t_0(i) - t_0(i - 1)} - V_{nmo}^4(i),
\]

\[
(f(i) = V_{nmo}^4(i)[1 - 4 A_4(i) t_0^2(i) V_{nmo}^2(i)].
\]

Substitution of equation (A4) into equation (A9) allows one to express \(f(i)\) through the effective \(\eta(i)\):

\[
f(i) = V_{nmo}^4(i) [1 - 8\eta(i)].
\]

Finally, we find the interval value \(\eta\) from equations (A4), (A7), and (A8):

\[
\eta = \frac{H_i}{8 V_{nmo}^2(i)} \left[ f(i) t_0(i) - f(i - 1) t_0(i - 1) \right] t_0(i) - t_0(i - 1) - V_{nmo}^4(i)]
\]
Uncertainties in seismic inverse calculations

John A. Scales
Department of Geophysics and Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado, USA

ABSTRACT

Solving an inverse problem means making inferences about physical systems from real data. This requires taking into account three different kinds of information or uncertainty:

- What is known about the parameters independently of the data? In other words, what does it mean for a model to be reasonable or unreasonable?
- How accurately are the data known? That is, what does it mean to "fit the data"?
- How accurately is the physical system modeled? Does the model include all the physical effects that contribute significantly to the data?

Exploration seismologists have a large amount of information available that could be used to refine inferences about the subsurface. Some of this information could readily be encapsulated as hard constraints, while in other cases, probabilities, whether derived directly from data or personal prejudices, seem more appropriate. In this paper I will discuss how such information can be used to quantify the uncertainties in inverse calculations.

Key words: Inference, inverse theory, uncertainty, seismic inversion

A Simple Example

To motivate the discussion, let us consider first a problem of linear tomographic inversion.

The Forward Problem

Acoustic sources and receivers are lowered into two vertical boreholes as shown in Figure 1. Sources are excited in one borehole, seismograms are recorded in the other, and the goal is to use the time-of-flight of waves to infer the elastic properties of the medium between the boreholes. To some approximation the wave propagation can be treated kinematically, so the travel time of a wave (e.g., the direct arrival) is related to the unknown slowness between the boreholes by

\[
t(s(x, y, z)) = \int_{\text{ray}(s(x, y, z))} s(x, y, z) \, dl.
\]

Figure 1. A simple model of cross-hole or X-ray tomography
The details of the calculation are not important here, they are described in various places (Bording et al. (1987), for example). The point is that within certain approximations (small perturbations, high frequency, etc.) the forward operator mapping model parameters to data can, when discretized, be represented as a linear system of equations:

\[ \delta t = J \cdot \delta s \]  

(2)

where \( \delta t \) is the vector containing the differences between the observed travel times and those computed using an approximate slowness model \( s_0 \), \( \delta s \) is a vector of perturbations to \( s_0 \) (what we’re seeking), and \( J \) is the Jacobian derivative matrix of the forward operator.

The Inverse Problem

Now that we have a tractable, linear forward problem, mapping model parameters (the components of \( \delta s \)) into predicted data, we can consider how to use the travel time observations to make inferences on these parameters. The first step is to decide upon a criterion for measuring the degree of data fit. This is called the likelihood function. A widely used likelihood function is

\[ \sqrt{\frac{(2\pi)^{N-1}}{\det C_d}} \exp \left[ -\frac{1}{2} (d_c - d_{\text{obs}})^T C_d^{-1} (d_c - d_{\text{obs}}) \right] \]

where \( d_{\text{obs}} \) and \( d_c(m) \) are the observed and computed (i.e., predicted) data, and \( C_d \) is the covariance matrix of data uncertainties. The argument of this exponential defines an \( N \)-dimensional quadratic form, the level surfaces of which are ellipsoidal error surfaces. For any given degree of data fit, the corresponding level surface tells us the range of model parameters which “fit the data” at that tolerance. For simplicity, let us suppose that the data (travel times in this case) are subject to known independent Gaussian uncertainties. Then the solution of the inverse problem would seem to be: find slowness values \( \delta s \) such that

\[ \frac{1}{N} \sum_{i=1}^{N} \left( \frac{t_i^o(\delta s) - t_i^o}{\sigma_i} \right)^2 \]  

(3)

is approximately equal to 1, where \( N \) is the number of observations, \( t_i^o \) is the \( i \)-th observed travel time, \( t_i^o \) is the \( i \)-th travel time calculated through the current slowness model, and \( \sigma_i \) is the standard deviation of the \( i \)-th datum. Making Equation 3 approximately equal to one corresponds to fitting the data on average to one standard deviation. Of course, other choices could be made.

Let’s see how this works in practice. Figure 2 shows an example of such a tomography problem. The model consists of a homogeneous background with slowness equal to 2, and an embedded inclusion shaped like an “E” having slowness equal to 1.5. At the right we see the illumination of the rays. There are 10 sources, 20 receivers and \( 2^2 \) cells within which the slowness is assumed to be constant.

Next add uncorrelated Gaussian noise, so that the standard deviations \( \sigma_i \) are constant and known. The textbook solution to the problem is to do a singular value decomposition (SVD) of the Jacobian matrix. Then, build up the solution one singular vector at a time, starting with the largest singular values and moving down the spectrum, until the solution just manages to fit the data.

Figure 3 shows this truncated SVD solution as both a perspective (left) and contour (right) plot. Although we can see the inclusion in the contour plot, the question we have to address is whether exhibiting this model really solves an inverse problem. The model fits the data, undeniably, but are the complicated looking features really required to fit the data or are they merely numerical artifacts? Is this model realistic? Are there other realistic models that would fit the data just as well?

What’s the Point?

One conclusion to be drawn from this simple experiment is that data fit alone is never enough to judge the solution of an inverse calculation. At best we can determine ranges of parameter values that are consistent with the data, for example, by looking at the confidence intervals defined by the level surfaces of the likelihood function. But since unrealistic models can fit the data too, one must balance the extent to which a model fits the data against the extent to which it conforms to our knowledge about what makes a model realistic or not in a particular case. This knowledge takes many different forms. It might be as simple as saying that a particular parameter must be positive, or that it must lie in some interval. Or it might be more complex. We might have previous measurements of this parameter or measurements of a related parameter.

Now it could be argued that in the tomography calculation described above, it would make more sense to use a different kind of regularization than SVD truncation. A finite-difference penalty function would also regularize the inversion and assure a smooth solution (e.g., Constable et al. (1987) or Scales et al. (1990)). Certain this is a conservative strategy for problems where there is little a priori information. But for least squares,
Figure 2. Straight rays are traced between 10 sources and 20 receivers. The goal is to be able to characterize the unknown inclusion (left). On the right, the ray “illumination” plot is shown.

Figure 3. A perspective (left) and contour (right) plot of the truncated SVD solution which just fits the data in the sense that its normalized $\chi^2$ is 1.

such a regularization scheme is equivalent to choosing a particular a priori relationship among the model parameters: a model covariance matrix, in fact. For example, using an anisotropic Laplacian to regularize the calculation would be equivalent to making different assumptions about the degree of horizontal and vertical smoothness of the velocity model; probably a reasonable thing to do in sedimentary basins. So this begs the question of whether we could make a more systematic characterization of such covariance matrices from other geological or geophysical observations.

Further, even in this toy problem we have introduced a number of assumptions and approximations which must ultimately be quantified in order that the inferences made be meaningful. For example, what errors have been introduced by neglecting refraction, by assuming ray theory was valid in the first place, by discretizing the model into a certain number of cells of constant slowness, by using a least squares criterion to measure data fit, by assuming that the data errors were i.i.d. Gaussian random variables, and on and on.

Information and Uncertainty
Even from this simple example it is clear that to be able to assign meaningful uncertainties to the results of an inverse calculation one must quantify what is known about the parameters of the physical system before the experiment, and what is known about the uncertainties in the data.

Information About Models
In exploration seismology there is a large amount of a priori information that could be used to influence inverse calculations. Here, a priori refers to the assumption that this information is known independently of the data. Plausible geologic models can be based on rock
outcrops, models of sedimentological deposition, subsidence, etc. There are often in situ and laboratory measurements of rock properties that have a direct bearing on macroscopic seismic observations, such as porosity, permeability, crack orientation, etc. There are other, less quantitative, forms of information as well, the knowledge of experts for instance.

There is a simple conceptual model that can be used to visualize the application of this diverse a priori information. Suppose we have a black box into which we put all sorts of information about our problem. We can turn a crank or push a button and out pops a model that is consistent with whatever information that is put in. If this information consists of an expert’s belief that, say, a certain sand/shale sequence is 90% likely to appear in a given location, then we must ensure that 90% of the models generated by the black box have this particular sequence. One may repeat the process indefinitely, producing a collection of models that have one thing in common: they are all consistent with the information put into the black box.

Let us assume that a particular layered Earth description consists of the normal-incidence P-wave reflection coefficient \( r \) at 1000 different depth locations in some well-studied sedimentary basin. Suppose, further, that we know from in situ measurements that \( r \) in this particular sedimentary basin almost never exceeds 0.1. What does it mean for a model to be consistent with this information? We can push the button on the black box and generate models which satisfy this requirement. Figure 4 shows some examples.

The three models shown satisfy the hard constraint that \(|r| \leq 0.1\) but they look completely different. The question is, which is most consistent with our assumed prior information? And how do we measure this consistency? If we make a histogram of the in situ observations of \( r \) and it shows a nice bell-shaped curve, are we justified in assuming a Gaussian prior distribution? On the other hand, if we do not have histograms of \( r \) but only extreme values, so that all we really know is that \(|r| \leq 0.1\), are we justified in thinking of this information probabilistically?

If we accept for the moment that our information is best described probabilistically, then a plausible strategy for solving the inverse problem would be to generate a sequence of models according to the prior information and see which ones fit the data. In the case of the reflectivity sequence, imagine that we have surface seismic data to be inverted. So for each model generated by the black box, compute synthetic seismograms, compare them to the data and decide whether they fit the data well enough to be acceptable. If so, the models are saved; if not, they are rejected. Repeating this procedure many times results in a collection of models that are by definition a priori reasonable and fit the data. If the models in this collection all look alike, then the features the models show are well-constrained by the combination of data fit and a priori information. If, on the other hand, the models show a diversity of features, then these features cannot be well-resolved by the surface data.

**Should Hard Constraints Be “Softened”?’**

It is plausible that any probability distribution consistent with a given hard constraint is more informative than the constraint itself. To say, for instance, that any model with \(|r| \leq 0.1\) is feasible is certainly not the same thing as saying that all models with \(|r| \leq 0.1\) are equally likely. And while we could look for the most conservative or least favorable such probabilistic assignment, Backus (1968) makes an interesting argument against any such probabilistic replacement in high- or infinite-dimensional model spaces. His point can be illustrated with a simple example. Suppose that all we know about an \( n \)-dimensional model vector \( r \) (now no longer the reflection coefficient, but any model vector) is that its length is less than some particular value—unity.
for the sake of definiteness. In other words, suppose we know a priori that \( r \) is constrained to be within the \( n \)-dimensional unit ball \( B_n \). Backus considers various probabilistic replacements of this hard constraint; this is called "softening" the constraint. For instance, allow each component of \( r \) to be uniformly distributed on \([-1, 1]\). This particular choice turns out to be untenable since the expectation of \( r^2 \) goes to infinity as the dimension of the model space does, but this can be overcome without too much difficulty. We could choose a prior probability on \( r \) which is uniform on \( B_n \). Namely, the probability that \( r \) will lie in some small volume \( \delta V \in B_n \) shall be equal to \( \delta V \) divided by the volume of \( B_n \). Choosing this uniform prior on the ball, it is not difficult to show that the expectation of \( r^2 \) for an \( n \)-dimensional \( r \) is

\[
\langle r^2 \rangle = \frac{n}{n+2}
\]

which converges to 1 as \( n \) increases. Unfortunately this does not overcome Backus' basic objection, since the variance of \( r^2 \) goes as \( 1/n \) for large \( n \), and thus we seem to have introduced a piece of information that was not implied by the original constraint; namely that for large \( n \), the only likely vectors \( r \) will have length equal to one. The reason for this apparently strange behavior has to do with the way volumes behave in high-dimensional spaces. If we compute the volume of an \( n \)-dimensional shell of thickness \( \epsilon \) just inside an \( R \)-diameter ball we can see that:

\[
V_\epsilon \equiv V(R) - V(R - \epsilon) = C_n (R^n - (R - \epsilon)^n) = V(R) \left( 1 - \left( 1 - \frac{\epsilon}{R} \right)^n \right) \quad (4)
\]

where \( C_n \) depends only on the dimension. Now for \( \epsilon/R \ll 1 \) and \( n \gg 1 \) we have

\[
V_\epsilon \approx V(R) \left( 1 - e^{-\epsilon/R} \right).
\]

This says that as \( n \) gets large, nearly all of the volume of the ball is compressed into a thin shell just inside the radius.

But even this objection can be overcome with a different choice of probability distribution to soften the constraint. For example, choose \( r \equiv |r| \) to be uniformly distributed on \([0,1]\) and choose the \( n-1 \) spherical polar angles uniformly on their respective domains. This probability is uniform on \( |r| \), but non-uniform on the ball. However it is consistent with the constraint and has the property that the mean and variance of \( r^2 \) is independent of the dimension of the space. A different approach altogether, which I will not discuss, is the use of physical invariances associated with the model to derive so-called noninformative priors (Tarantola, 1987).

So, as Backus has said, we must be very careful in replacing a hard constraint with a probability distribution, especially in a high-dimensional model space. Apparently innocent choices may lead to unexpected behavior. One might well ask whether it was possible to measure the extra information, if any, implied by the softened version of a constraint. As we have seen, for specific choices it is possible to quantify certain aspects of this extra information. But perhaps it is best to not pose this question from the point of view of information, but rather to ask what are the consequences of various choices on the inferences we make. In other words, by computing the a posteriori resolution of the individual parameters (whether from a Bayesian viewpoint or via an approach such as confidence set inference (Stark, 1992)), we should be able to determine precisely the consequences of any given assumption about the information we use in an inverse calculation.

Information Naturally Suited to a Probabilistic Treatment

In the last section we looked at information that took the form of a hard constraint on the model vector or some function of the model vector, such as its length. Some information comes to us in a form that seems naturally expressed probabilistically. For example, Figure 3 shows an example of a sonic log,\(^\dagger\) which can be thought of as characterizing the short vertical wavelength (1 m or less) material fluctuations at a given location in the earth. Perhaps we can extract the statistical features of these data and use them to refine inferences made with other data sets from the same area, such as surface seismic data.

In order to focus on the statistical aspects of the data it is useful to subtract the long-wavelength trend, which is likely the result of more gradual, deterministic changes such as compaction. The trend and the resulting fluctuations are shown in Figures 6 and 7.

To go further, we need an expression for the ultimate (so-called a posterior) probability on the space of models as a function of the prior information and data uncertainties. It is

\[
\sigma(m) = kp_M(m)L(m) \quad (3)
\]

where \( \sigma \) is the posterior probability, which assimilates all the available information, \( p_M \) is what we know about the models a priori, \( L \) is the likelihood function (which depends implicitly on the data), and \( k \) is the normalization.

\(^\dagger\) That is, an in situ estimate of acoustic wave speed as a function of depth inferred from travel times between a source and receiver in a measuring tool lowered into a well. NB that the "data" are in fact the result of an inference. This is a common situation in which there is no fundamental distinction between the input and the output of an inverse calculation.
This equation is derived in the appendix for completely general kinds of information. For now, let us assume that the fluctuations of the well log are stationary and Gaussian, so that $\rho_M$ and $L$ are multi-dimensional Gaussian distributions. In this case, it is relatively straightforward to compute a covariance matrix associated with the observations. For example, Figure 8 shows the first 100 lags (1 meter per lag) of the correlation function for the fluctuating part of Figure 7, as well as an approximate covariance matrix. Then, the Bayesian prior probability $\sigma(m)$ is proportional to

$$\exp\left[-\frac{1}{2}(m-m_{\text{prior}})^T C^{-1}_M (m-m_{\text{prior}})\right]$$

(6)

where $C_M$ is the covariance matrix describing uncertainties in the models (as in Figure 8) and $m_{\text{prior}}$ is the center of this distribution. Examples of such Gaussian calculations and their relation to Tikhonov regularization are shown in Gouveia (1996) and Gouveia and Scales (1996).

But the question remains as to whether the assumptions of stationarity and Gaussianity are reasonable. These two effects can trade off against one another. Figure 9 shows the standard deviation of the log as a function of depth. The non-stationarity is apparent and is probably ubiquitous in practice. One simple technique to overcome it is to divide the log into windows within which stationarity can be safely assumed. There is some preliminary evidence on the question of Gaussianity in Figure 10. This shows the third order cumulant and bispectrum of a small window (75 meters) of the fluctuating part of the data. As described in Subba Rao and Gabr (1984), a rigorous test for Gaussianity involves showing that the bispectrum is zero at all frequencies by looking at the estimated values over an appropriate grid of points.

In any case, if one were to accept the Gaussian hypothesis, then it would be reasonable to compute a covariance matrix for the $a$ priori model distribution from widely available information. If the Gaussian hypothesis is rejected, then one needs to incorporate the higher order statistical information into the inverse calculation. How to do this is not obvious. If some number of moments of the unknown distribution can be estimated accurately from the data, then a plausibly conservative strategy would be to compute the maximum entropy distribution subject to these moments as constraints. If the first and second moments are used, this results in a Gaussian distribution; i.e., the Gaussian is the maximum-entropy distribution subject to fixed first and second moments. In Gouveia, Moraes & Scales (1996), we show examples of the estimation of 1D maximum-entropy prior distributions using up to four sample moments as constraints. Unfortunately, these techniques are not straightforward to apply for $n$-dimensional distributions, even assuming that one could accurately estimates moments beyond the second, and this is a big assumption.

Another possibility is to try to approximate directly the joint distribution function of the unknown process by making histograms of its low order marginal distributions. Here we rely on the fact that the $n$-dimensional distribution can be expressed in terms of the marginal distributions of the process. For example, the three dimensional joint distribution $f(x, y, z)$ can be written as $f(x|y, z)f(y, z)$, which can in turn be written as $f(x|y, z)f(y|z)f(z)$ and so on by induction for the general $n$-dimensional distribution. In principle, the marginal
distributions can be estimated by making histograms. Here is a simple example of how such a procedure might be used in practice. Using the fluctuating part of the well log, as shown in Figure 7, we begin by constructing histograms to estimate the marginal distributions of the process. The histogram of nearest-neighbor values along the log is shown in Figure 11. If the process were Markovian, this would be sufficient to completely characterize the fluctuations of the log. If not, then we would need to compute histograms of non-nearest-neighbor pairs, histograms of triples, and so on. The problem now is when to stop? One possibility is to use a statistical hypothesis test such as Kolmogorov-Smirnov to decide when models pseudo-randomly sampled from a given set of histograms are sufficiently close to the observations. Another possibility would be to argue on a priori grounds which marginals are needed to capture the important information in the process. But I have no idea how one would do this.

If we can compute a non-parametric estimate of the joint distribution function of some piece of information, then this information can readily be incorporated into a Monte Carlo importance sampling scheme of the type proposed by Mosegaard and Taranto (1995). In their approach, it is not necessary to assume that the prior information is expressed in terms of analytic functions; all that is required is a means of generating models in proportion to their a priori probability (i.e., importance sampling the prior distribution). These models are then accepted or rejected in a Metropolis-like procedure according to how well they fit the data, which guarantees that the collection of models that are generated do indeed sample the full a posteriori probability distribution. The key advance of of Mosegaard and Taranto (1995) is that we no longer need to generate models that are a priori unrealistic—we only compute the response (predicted data) of those models which have been sampled from the prior. By making the prior sufficiently informative, it may be possible to overcome the profound cost of global sampling in high dimensional spaces.
Uncertainties in the data

In the last section we looked at probabilistic characterizations of models of the Earth's upper crust from commonly available measurements. Next we look at estimating uncertainties in the seismic data themselves. This is less controversial than the assignment of prior probabilities, but no less important since a sine qua non for any inference calculation is a specification of just what it means to fit the data.

We are used to thinking of seismic experiments as being non-repeatable. While this is true in earthquake seismology, it is not necessarily true in controlled-source seismology. Most land reflection data recorded nowadays uses a vibroseis source, i.e., a swept continuous wave signal. By careful stacking and cross-correlation with the known input signal, such data can approximate the results of explosive sources with high signal to noise ratio. But usually the raw, unstacked data are not saved; once the stacking and correlation have been done, the raw data are usually discarded. Figure 12 shows an example in which the raw traces have been saved. There are 25 unstacked, uncorrelated vibroseis traces recorded by
the Reservoir Characterization Project of the Colorado School of Mines. Each trace corresponds to a separate vibroseis sweep at a single far-offset source location. All 23 traces were recorded within a span of about 10 minutes. On the left is shown the full 7 second sweep, while on the right the data have been windowed to just the first second, so as to exclude any source generated signal.

Since we now have repeated realizations of the experiment, it is possible to directly estimate a data covariance matrix. This is shown in Figure 13. Once we have the data covariance matrix $C_d$, then we can compute the Bayesian likelihood function

$$\sqrt{\frac{(2\pi)^{-N}}{\det C_d}} \exp \left[ -\frac{1}{2} (g(m) - d_{\text{obs}})^T C_d^{-1} (g(m) - d_{\text{obs}}) \right]$$

(7)

Well, almost. The problem is that we have been implicitly thinking about the data covariance matrix as representing random errors or noise, whereas there is often no clear distinction between random noise and unmodeled physics. If we neglect to include anisotropy in the forward model, for instance, then energy in the data due to anisotropy will be erroneously mapped into isotropic features.

Sometimes the distinction is largely a matter of taste. In the cross-hole tomography example, suppose we attempted to repeat the experiment in order to gather statistical information. Perhaps we cannot reposition the sources and receivers in exactly the right places, so small shifts in the—assumed known—source and receiver locations will introduce errors into the problem. Do we regard these as being essentially random fluctuations which broaden the distribution of the data, or do we look at the shifts in source and receiver locations as being model parameters (like station corrections) to be inferred from the data?

Tarantola (1987) shows that as long as the uncertainties are Gaussian, we can still use Equation 7 to describe the likelihood function, provided we replace the data covariance matrix $C_d$, with a combined covariance matrix

$$C_D = C_d + C_T$$

where $C_T$ represents the contribution due to theoretical errors.

More often than not, theoretical errors are neglected in real applications. And the reason is plain, for it is not usually possible to account for a particular kind of unmodeled physics without incurring the additional expense that led to the neglect of this physics in the first place. How can one quantify the effects of neglecting anisotropy without doing the complete calculation anisotropically from the beginning? In some cases it is possible to do just that. For example, looking at well logs, which exhibit heterogeneity on the 1 meter or smaller length scale, it is reasonable that multiple-scattering dispersion and attenuation will be a non-negligible effect in typical sedimentary rocks. The magnitude of the dispersive effects can be estimated by comparing the dispersion free ray-theoretic speed with the Backus long-wavelength effective speed: total dispersion of 5% or more is common (Scales, 1993). So in this case, one can estimate the data error associated with the neglect of dispersion or simply correct for it, so as to eliminate the error altogether (Scales & Van Vleck, 1996). Similar effects which are potentially as large or larger include azimuthal anisotropy due to aligned systems of cracks within reservoirs, long-wavelength effective transverse isotropy due to fine layering, anelasticity associated with fluid-filled pores, 3D wave propagation effects (at least in structurally complex areas), and many others. A careful treatment of these modeling errors is beyond the scope of this paper, but it seems clear that there are a number of such effects which must be taken into account in even the simplest problems of real exploration interest.
Conclusions

As in all such calculations, seismic inversion requires careful estimation of prior information and data uncertainties in order to be able to make quantitative inferences about the Earth. The prior information comes from diverse sources, some of which may be described as hard constraints on parameters or functions of parameters, and some which may best be described probabilistically. There are also sources of subjective prior information from experts. It is important that we assimilate such information consistently and quantitatively. In the event that all the uncertainties can be treated as being Gaussian random variables, this information conveniently takes the form of model and data covariance matrices. In that case, the computational formalism applied is superficially similar to Tikhonov regularization, or constrained least squares, although the goals and conclusions are potentially rather different. But in some cases such Gaussian assumptions may not be justified. In the general case, in which the probability distribution associated with the prior information can only be sampled point-wise, it may be necessary to use an importance sampling procedure based on Monte Carlo methods. But this begs the question of how one estimates such a general prior probability. I have given some suggestions as to how these probabilities may be estimated, but I suspect we are a long way from a definitive answer to this question.

Acknowledgements

This work was inspired by my many discussions over the past few years with Albert Tarantola. Also, I learned a lot, as well as how much more I need to learn, from several stimulating conversations in Aarhus and Golden with Philip Stark. I would like to thank Tom Davis and Steve Roche of the Reservoir Characterization Project at the Colorado School of Mines for the vibroseis data, and Wences Gouveia and Fernando Moraes for useful discussions. Wences Gouveia wrote the code that was used to compute Figure 10. This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, the Shell Foundation and the Army Research Office.

References


APPENDIX A: The Bayesian Posterior Probability

The Bayesian posterior probability density on the space of models must be the product of two terms: a term which involves the a priori probability on the space of models and a term which measures the extent of data fit:

\[ \sigma(m) = kp(m) L(m) \]  

where \( k \) is the normalization constant and \( L \) is the likelihood function, which depends implicitly on the data.

We now show how Equation (A1) follows logically from Bayes’ theorem provided we generalize our notion of “data” to allow for the possibility that the data might be specified by probability distributions (Tarantola, 1987). Following Duijndam (1987), we begin by using the notation common amongst Bayesians, then we show how this relates to the more standard inverse-theoretic notation in Tarantola (1987).

In this approach we assume that we have some prior joint distribution \( p_0(m, d) \). Further, we suppose that as the result of some observation, the marginal pdf of \( d \) changes to \( p_1(d) \). We regard \( p_1(d) \) as being the “data” in the sense that we often know the data only as a distribution, not exact numbers. In the special case where the data are exactly known, \( p_1(d) \) reduces to a delta function \( \delta(d - d_{\text{obs}}) \).

How do we use this new information in the solution of the inverse problem? The answer is based upon the following assumption: whereas the information on \( d \) has changed as a result of the experiment, there is no reason to think that the conditional degree of belief of \( m \) on \( d \) has. I.e.,

\[ p_1(m|d) = p_0(m|d). \]  

(A2)

From this one can derive the posterior marginal \( p_1(m) \):

\[ p_1(m) = \int_D p_1(m, d) \, dd \]  

(A3)

\[ = \int_D p_1(m|d)p_1(d) \, dd \]  

(A4)

\[ = \int_D p_0(m|d)p_1(d) \, dd \]  

(A5)

\[ = \int_D \frac{p_0(d|m)p_0(m)}{p_0(d)} p_1(d) \, dd \]  

(A6)

\[ = p_0(m) \int_D \frac{p_0(d|m)}{p_0(d)} p_1(d) \, dd, \]  

(A7)

where \( D \) denotes the data space.

Switching now to the inverse-theoretic notation, let us regard \( p_0(d) \) as being the prior distribution on data \( \mu_D(d) \): this is what we know about the data before we’ve actually done this particular experiment. Further, we identify \( p_1(m) \) as the posterior distribution on the space of models, \( p_1(d) \) as the data errors, \( p_0(m) \) as the prior distribution on the space of models, and \( p_0(d|m) \) as the modeling uncertainties.

\[ p_1(m) \equiv \sigma(m) \]  

(A8)

\[ p_1(d) \equiv \rho_D(d) \]  

(A9)

\[ p_0(m) \equiv \rho_M(m) \]  

(A10)

\[ p_0(d|m) \equiv \Theta(d|m), \]  

(A11)

then we arrive at essentially Equation (1.65) of Tarantola (1987), except that there is no explicit reference to the non-informative prior. The prior distributions simply represent what we know about the problem—models and data—before the data have been collected.
\[ \sigma(m) = \rho_M(m) \int_D \frac{P_D(d) \Theta(d|m)}{\mu_D(d)} \, dd \] \quad (A12)

An important special case occurs when the modeling errors are negligible, i.e., we have a perfect theory. Then the conditional distribution \( \Theta(d|m) \) reduces to a delta function \( \delta(d - g(m)) \) where \( g \) is the forward operator. In this case, the posterior is simply

\[ \sigma(m) = \rho_M(m) \left[ \frac{P_D(d)}{\mu_D(d)} \right]_{d = g(m)} \] \quad (A13)
Amplitude calculation for 3-D common offset $v(z)$ inversion

Meng Xu
Center for Wave Phenomena
Colorado School of Mines

ABSTRACT
The general 3-D inversion formula developed by Bleistein et al is specialized to a depth-dependent medium. An efficient ray tracer for this model is developed to calculate the necessary constituents of the ray theory. The traveltime and amplitude tables for the inversion are computed. The motivating application for this project is detection of small objects-10cm-in shallow water-10-20m.

Key words: Born inversion, three-dimensional, common offset, depth-dependent, amplitude

Introduction
This paper describes the background mathematical analysis for specialization of the Born/Kirchhoff inversion formalism [Cohen, Hagan and Bleistein, 1986; Bleistein, 1986], to a three dimensional (3-D) depth dependent background medium. Some numerical results that test the code I developed to implement this method are also presented.

The underlying motivating project has as its objective the detection of small objects (10-20cm) in shallow water (10-20m). This objective differs from the seismic inverse problem in two ways: first, the depth, time and frequencies differ from the seismic problem by approximately a factor of 100—not an important difference; second, the lateral extent of the scatterer is finite and small compared to the range from the source/receiver array or survey extent, whereas in the seismic problem, the lateral extent is often comparable to these other length scales. We have not, at this time, exploited this latter difference, so that the resulting code is applicable to both types of problems at both scales.

The survey we have in mind is traditional: a boat carries a towed array along parallel lines on the upper surface and periodically sets off an acoustic source, collecting the resulting backscattered data on the array. Data is then re-sorted in common (constant) offset data sets and processed by the program developed here to produce a reflector map. In addition, the output provides an estimate of the angularly dependent reflection coefficient at specular at the sample points on the scatterer, as well as an estimate of that incident specular angle with respect to the normal to the reflector. By processing the data for a suite of offsets, data for amplitude versus offset (AVO) or amplitude versus angle (AVA) analysis is generated for each point on the reflector.

To calculate the amplitude weights for a 3-D common offset $v(z)$ inversion program, I use the analytic formulas for 3-D ray data in a depth dependent propagation medium. These ray data are interpolated from ray coordinates to cartesian coordinates after ray tracing. From these ray data, I compute the Jacobian of the transformation from cartesian $(x,y,z)$ coordinates to ray coordinates $(\sigma, \alpha, \beta)$ and I compute the Beylkin determinant $h$. Both of these determinants are essential to the amplitude computation.

The numerical tests confirm that the computer code matches analytical values of the travel time and amplitude along the rays extremely well. Furthermore, for a model sphere in constant background, we find that the reflector map successfully produces an image of the sphere.

Inversion Formula
In this section, the inversion formulas of Bleistein(1986) are introduced. These formulas provide a tool for obtaining correct locations of interfaces as well as a model-consistent specular reflection coefficient and incidence angle. The inversion formulas have the form of aperture-
limited Fourier-like integrals. The integrand of these integrals contains a determinant that characterizes the viability of inverting a particular data set. This determinant is part of a Jacobian that depends on both the background propagation parameters and on the source-receiver configuration. As a result, the problem of extending the inversion formula to new recording geometries is reduced to a problem of computing the value of the form of this determinant that corresponds to a specific geometry.

The full 3-D Kirchhoff inversion formula (Bleistein, 1986) is

$$\beta(y) = \frac{1}{\pi^2} \int \frac{d^3 \xi \cdot |h(y, \xi)| \cdot \alpha(y, \xi)}{a(y, \xi) \cdot \nabla \cdot \phi(y, \xi)}$$

$$\int i \omega \ c \ e^{-i \omega \cdot \phi(y, \xi) \cdot u_{S}(x_{g}, x_{s}, \omega)}.$$  

(1)

In this equation, $\beta(y)$ is the reflectivity function for the imaging section; $y = (y_{1}, y_{2}, y_{3})$ is the 3-D position vector of the output or imaged point and $\xi = (\xi_{1}, 0)$ are the surface coordinates of the ray to $y$ from either the source or receiver. For a common offset inversion, we denote by $\tau_{x_{g}}$ and $\tau_{y}$ ($A_{2}$ and $A_{3}$) the amplitude and phase of the WKB-J Green’s functions at $y$ with initial point at the source. Thus, in the above equation

$$a(y, \xi) = A_{1} A_{2}, \quad \phi(y, \xi) = \tau_{x_{g}} + \tau_{y},$$  

(2)

with the latter being the total traveltime from source to $y$ to receiver. Furthermore, $u_{S}(x_{g}, x_{s}, \omega)$ is the data in $\omega$ (frequency) domain; $x_{s}$ and $x_{g}$ are the source and the receiver position coordinates; $h$ is the determinant characterizing the viability of inverting a particular data set with a specific geometry at the given output point. The expression of the spatial weighting in terms of this one determinant for any source-receiver configuration and background propagation speed is a major contribution of Beylkin’s approach to high-frequency inversion and is being referred as the "Beylkin determinant". The details of $h$ will be described later.

The reflectivity function provides a reflectivity map through a family of bandlimited delta functions that peak on the reflectors in the medium. The peak amplitude is proportional to the angularly dependent reflection coefficient at a specular angle $\theta_{s}$, where $2\theta_{s}$ is the angle between the ray directions from the source and receiver to the output point $y$:

$$\beta_{\text{peak}} \sim R(y, \theta_{s}) \cos \theta_{s} \pi c(y) \int_{-\infty}^{\infty} F(\omega) d\omega.$$  

(3)

That is, the peak value of $\beta$ for $y$ on $S$ is the geometrical optics reflection coefficient multiplied by $2 \cos \theta_{s} / c(y)$ and multiplied by $1/2\pi$ times the area under the filter in the $\omega$-domain.

For $y$ not on the reflector surface, the angle $\theta$ is defined by the equation,

$$\nabla_{y} \tau(y, x_{S}) \cdot \nabla_{y} \tau(y, x_{P}) = \frac{\cos 2\theta}{c^{2}(y)}.$$  

(4)

Then, by computing $\nabla \phi(y, \xi) \cdot \nabla \phi(y, \xi)$, we are able to show that

$$|\nabla \phi(y, \xi)| = \frac{2 \cos \theta}{c(y)}$$  

(5)

In particular, for $y$ on the reflector surface, this provides a means for estimating $\cos \theta_{s}$ through the introduction of another inversion operator, differing by one power of $|\nabla \phi|$:

$$\beta_{1}(y) = \frac{1}{\pi} \int d^{2} \xi \left| \frac{h(y, \xi)}{a(y, \xi) \cdot \nabla \phi(y, \xi)} \right|^{2}$$

$$\int i \omega \ c \ e^{-i \omega \cdot \phi(y, \xi) \cdot u_{S}(x_{g}, x_{s}, \omega)}.$$  

(6)

The addition of the extra divisor of $|\nabla \phi|$ introduces this divisor in the asymptotic amplitudes of the result. In particular:

$$\beta_{\text{peak}}(y) \sim R(y, \theta_{s}) \int_{-\infty}^{\infty} F(\omega) d\omega,$$  

(7)

The fact that these two operators differ by a factor of $2 \cos \theta / c(y)$ allows us to estimate $\cos \theta_{s}$ from the ratio of the outputs without ever having determined the specular source-receiver pair that produced the distinguished value of $\theta_{s}$. This, in turn allows us to determine $R(y, \theta_{s})$ from either output. With knowledge of $\theta_{s}$ and the background wavespeed, $c(y)$, it is conceivable, within the limits of the accuracy of the data, that we should be able to estimate the jump in the propagation speed across the reflector using the formula for the geometrical optics reflection coefficient. More generally, in a variable density medium, we determine the impedance jump from this analysis. By processing data for multiple offsets, we generate data for amplitude versus offset (AVO) analysis or amplitude versus angle (AVA) analysis.

Ray Tracing

Here, I describe the ray tracing in a depth-dependent medium in 3-D. For a more complete discussion of the underlying ray theory for determining solutions of the wave equation in the form $A \exp(i \omega \tau)$, refer to (Bleistein, 1984).

The ray equations

The general form of the ray equations

$$\frac{dx_{i}}{d\sigma} = p_{i}, \quad \frac{dp_{i}}{d\sigma} = p \frac{dp}{dx_{i}}, \quad i = 1, 2, 3,$$
provide the basis for ray theoretic modeling.

The slowness vector \( \mathbf{p} \equiv \nabla \tau \) points in the direction normal to the surfaces of constant \( \tau \). Surfaces of constant \( \tau \) are called “wavefronts”, and the \( p \) vector points in the direction tangent to the “raypaths”. These, in turn, are the spatial trajectories of the solution to (8).

We seek solutions of (8) for rays emanating from a single point, say \( x_0, y_0, z_0 \) in an arbitrary downward direction. Those directions are determined by the initial values of \( \alpha \) and \( \beta \) representing the azimuthal and polar angles, respectively, of the spherical coordinate system. See Fig. 1. Then, the initial ray direction \( \mathbf{p}_0 \) is given by

\[
p_0 = p_0 \mathbf{e}_1 \cos \alpha \sin \beta + p_0 \mathbf{e}_2 \sin \alpha \sin \beta + p_0 \mathbf{e}_3 \cos \beta.
\]  

(9)

**Specialization to a depth-dependent medium**

For a medium that has wavespeed variability only in the \( z \) direction, \( p_1 \) and \( p_2 \) are constants on each ray path; denote them by \( p_{10} \) and \( p_{20} \). Note that \( p_{10} \) and \( p_{20} \) are their respective initial values. The ray equations become

\[
\begin{align*}
\frac{dx_1}{d\sigma} &= p_{10}, & \frac{dx_3}{d\sigma} &= p_{20}, & \frac{dx_3}{d\sigma} &= p_3, \\
\frac{dp_1}{d\sigma} &= 0, & \frac{dp_2}{d\sigma} &= 0, & \frac{dp_3}{d\sigma} &= \frac{dp}{dx_3},
\end{align*}
\]

(10)

From the eikonal equation, we know the value of \( p_3 \)

\[
p_3 = \sqrt{p^2(x_3) - p_{10}^2 - p_{20}^2} = \sqrt{p^2(x_3) - p^2(x_{30}) \sin^2 \beta}.
\]

(11)

The third equation in (10) relates \( \sigma \) and \( x_3 \), so the ray equations may be rewritten in terms of \( x_3 \):

\[
\begin{align*}
\frac{dx_1}{dx_3} &= \frac{p_{10}}{p_3}, & \frac{dx_2}{dx_3} &= \frac{p_{20}}{p_3}, & \frac{dp_1}{dx_3} &= 0, & \frac{dp_2}{dx_3} &= 0, & \frac{dp_3}{dx_3} &= \frac{dp}{p_3}, & \frac{dt}{dx_3} &= \frac{p_3}{p_3}, & \frac{d\sigma}{dx_3} &= \frac{1}{p_3},
\end{align*}
\]

(12)

We solve for \( x_1, x_2, t, \) and \( \sigma \) as functions of \( x_3, \alpha, \) and \( \beta \). Let the initial depth be \( x_{30} \equiv x_0 \) and the final depth be \( x_3 \equiv x \). The solution of (13), then, is

\[
x - x_0 = \frac{\cos \alpha \sin \beta}{c(x_0)} \int_{x_0}^{x} \frac{ds'}{\sqrt{p^2(s') - p^2(x_0) \sin^2 \beta}},
\]

(13)

\[
\begin{align*}
y - y_0 &= \frac{\sin \alpha \sin \beta}{c(x_0)} \int_{x_0}^{x} \frac{ds'}{\sqrt{p^2(s') - p^2(x_0) \sin^2 \beta}}, \\
\tau - \tau_0 &= \int_{x_0}^{x} \frac{ds'}{c^2(s') \sqrt{p^2(s') - p^2(x_0) \sin^2 \beta}}, \\
\sigma - \sigma_0 &= \int_{x_0}^{x} \frac{ds'}{\sqrt{p^2(s') - p^2(x_0) \sin^2 \beta}}.
\end{align*}
\]

Recall that \( \alpha \) and \( \beta \) are the initial angles of the ray. When \( p_3 = 0 \), the ray is horizontally propagating. This is called a “turning point”. The equations above all have integrable singularities at the turning point, requiring special care in numerical computation, to be discussed below.

**The Jacobian J and ray amplitude**

Here, I discuss the solution of the transport equation

\[
2\nabla \tau \cdot \nabla A + A \nabla^2 \tau = 0,
\]

(14)

for the amplitude \( A \). This equation can also be written as an ordinary differential equation in the ray parameter \( \sigma \)

\[
\frac{d(A^2 J)}{d\sigma} = 0,
\]

(15)

with solution (Bleistein 1984, 8.3.12)

\[
A = \frac{\sqrt{\sin \beta}}{4\pi \sqrt{c(\xi) J(\sigma, \alpha, \beta)}}.
\]

(16)

In these equations, \( J \) is the Jacobian of the transformation from \( x \) to \( (\sigma, \alpha, \beta) \) via the solution of the ray equations,

\[
J = \left| \frac{\partial(x, y, z)}{\partial(\sigma, \alpha, \beta)} \right|.
\]

(17)

We calculate the ray theoretic amplitude \( A \) by computing \( J \).

Let us consider the solution of the ray equations in terms of \( \sigma \), rather than \( x \). The solution is a family of rays, distinguished from one another by the choice of the parameters \( \alpha \) and \( \beta \). Along each ray, the values of \( \tau, \sigma, x, y, \), \( z \), \( p_1 \), are known. \( p_1 \) and \( p_2 \) are constants on each ray; that is, independent of \( \sigma \):

\[
p_1 = \frac{1}{c(z_0)} \cos \alpha \sin \beta,
\]

(18)

\[
p_2 = \frac{1}{c(z_0)} \sin \alpha \sin \beta.
\]

With \( p_1 \) and \( p_2 \) independent of \( \sigma \), the equations for \( x \) and \( y \) are

\[
x = \xi_1 + \frac{\sigma}{c(z_0)} \cos \alpha \sin \beta,
\]

\[
y = \xi_2 + \frac{\sigma}{c(z_0)} \sin \alpha \sin \beta.
\]
\[ y = \xi_2 + \frac{\sigma}{c(z_0)} \sin \alpha \sin \beta. \quad (19) \]

and \( z \) is given in terms of \( \sigma \) by
\[ z = \int p_2 d\sigma. \quad (20) \]

I do not use equation (20) since we do the calculation on a uniform \( z \) grid, instead, we calculate \( \sigma \) for each \( z \) by using the third equation in (10). From (20) and (10), we can derive the nine terms of the determinant \( J \) in (17) as
\[
\begin{align*}
\frac{\partial x}{\partial \sigma} &= \frac{\sin \beta \cos \alpha}{c(z_0)}, \\
\frac{\partial y}{\partial \sigma} &= \frac{\sin \beta \sin \alpha}{c(z_0)}, \\
\frac{\partial z}{\partial \sigma} &= \frac{\sigma}{c(z_0)} - p_3, \\
\frac{\partial x}{\partial \alpha} &= -\frac{\sigma \sin \beta \sin \alpha}{c(z_0)}, \\
\frac{\partial y}{\partial \alpha} &= \frac{\sigma \sin \beta \cos \alpha}{c(z_0)}, \\
\frac{\partial z}{\partial \alpha} &= \frac{\sigma \cos \beta \cos \alpha}{c(z_0)}, \\
\frac{\partial x}{\partial \beta} &= \frac{\sigma \cos \beta \cos \alpha}{c(z_0)}, \\
\frac{\partial y}{\partial \beta} &= \frac{\sigma \cos \beta \sin \alpha}{c(z_0)}, \\
\frac{\partial z}{\partial \beta} &= \int \frac{-\sin \beta}{c(z_0)} d\sigma.
\end{align*}
\]

With these values, we can calculate \( J \) in (16) and then calculate \( A \) by using (17). Thus, we obtain \( A \) on a uniform grid in \( z \), but calculate it as if the independent parameter along the ray were \( \sigma \).

The Belkyin determinant \( h \)

The effects of source-receiver geometry are completely described by the Belkyin determinant
\[
h(y, \xi) = \det \begin{bmatrix} p_x + p_y \\ \frac{\partial (p_x + p_y)}{\partial \xi_1} \\ \frac{\partial (p_x + p_y)}{\partial \xi_2} \end{bmatrix}, \quad (22)
\]
in terms of the slowness vectors \( p_x \) and \( p_y \) and their derivatives with respect to the surface coordinates, \((\xi_1, \xi_2)\). These vector quantities are evaluated at the output point \( y \).

The determinant must be finite and nonzero for the identification of the cascaded model and inversion integral as an approximate Fourier transform. Thus, we could use the value of this matrix to characterize source-receiver configurations as providing invertible data by this formalism at an output point, \( y \). In particular, we require that this determinant be finite and nonzero for some range of \( \xi \) values at any \( y \) where the high frequency inversion is to be computed.

The general form of the Belkyin determinant can be written for all source-receiver geometries as
\[
h(y, \xi) = (p_x + p_y) \cdot (v_x + v_y) \times (w_x + w_y) \\
= (p_x + p_y) \cdot v_x \times w_y + (p_x + p_y) \cdot v_y \times w_x + (p_x + p_y) \cdot v_x \times w_x + (p_x + p_y) \cdot v_y \times w_y \quad (23)
\]
\[
(p_x + p_y) \cdot v_y \times w_y +
(p_x + p_y) \cdot v_x \times w_x,
\]
where,
\[
v_x \equiv \frac{\partial p_x}{\partial \xi_1}, \quad v_y \equiv \frac{\partial p_y}{\partial \xi_1},
\]
\[
w_x \equiv \frac{\partial p_x}{\partial \xi_2}, \quad w_y \equiv \frac{\partial p_y}{\partial \xi_2}.
\]
(24)

The first two terms can be recognized as being the Beylkin determinants for the common shot and common receiver geometries. We may write the Beylkin determinant as
\[
h(y, \xi) = 2 \cos^2 \theta \left[ h_s(y, \xi) + h_r(y, \xi) \right] +
(p_x + p_y) \cdot [v_x \times w_x + v_y \times w_y],
\]
where \(h_s(y, \xi) = p_x \cdot v_x \times w_x\) is the determinant for the common receiver case and \(h_r(y, \xi) = p_x \cdot v_y \times w_y\) is the determinant for the common source case. Unfortunately, the last term in (26) contains the cross products \(v_x \times w_y\) and \(v_y \times w_x\), which are not easily simplified.

Another way of writing the Beylkin determinant is
\[
h(y, \xi) = 2 \cos^2 \theta \left[ h_a(y, \xi) + h_b(y, \xi) \right]
+ h_c(y, \xi) + h_d(y, \xi),
\]
(26)
where the \(h_a, h_b, h_c\) and \(h_d\) are defined by
\[
h_a(y, \xi) \equiv p_x \cdot v_x \times w_x, \quad h_b(y, \xi) \equiv p_x \cdot v_y \times w_y,
\]
\[
h_c(y, \xi) \equiv p_x \cdot v_x \times w_y, \quad h_d(y, \xi) \equiv p_x \cdot v_y \times w_x.
\]
(27)

Unfortunately, we cannot write the general common-offset Beylkin determinant in a form that has a common multiplier of \(\cos^2 \theta\).

We already know the vectors \(p_x\) and \(p_y\) at the output point from the solution of the ray equations. Now we will compute their partial derivatives with respect to \(\xi_1\) and \(\xi_2\) as solutions of another set of ray equations.

By taking derivatives of (19) with respect to \(\xi_1\) and \(\xi_2\), respectively, and using (20), we obtain differential equations for \(\partial p_1/\partial \xi_1\) and \(\partial p_2/\partial \xi_1\). These equations are
\[
\frac{\partial p_1}{\partial \xi_1} = -\left(1 + p_x \frac{\partial \sigma}{\partial \xi_1}\right)/\sigma,
\]
\[
\frac{\partial p_2}{\partial \xi_1} = -p_y \frac{\partial \sigma}{\partial \xi_1}/\sigma,
\]
(28)
\[
\frac{\partial p_1}{\partial \xi_2} = -p_x \frac{\partial \sigma}{\partial \xi_2}/\sigma,
\]
\[
\frac{\partial p_2}{\partial \xi_2} = -p_y \frac{\partial \sigma}{\partial \xi_2}/\sigma,
\]
\[
\frac{\partial p_1}{\partial \xi_2} = -(1 - p x \frac{\partial \sigma}{\partial \xi_2})/\sigma.
\]

![Figure 2. Ray symmetry used in calculating h](image)

The only quantities that we do not know in the above equations are \(\partial \sigma/\partial \xi_1\) and \(\partial \sigma/\partial \xi_2\). It is difficult to implement the derivatives with respect to \(\xi_1\) and \(\xi_2\). However, using the symmetry of the \(v(x)\) medium, we find that \(d/dz = -d/dx\) and \(d/dz\), \(\sigma = -d/dy\) as shown schematically in Fig. 2. That is, when a ray is shifted horizontally, the quantity \(\sigma\) on that ray does not change. Also, one can see from Fig. 2, \(d\sigma/\partial \xi_1 = (\sigma_1 - \sigma_2)/\partial \xi_1 = -(\sigma_2 - \sigma_1)/\partial \xi_1\). Thus, \(\partial \sigma/\partial \xi_1\) and \(\partial \sigma/\partial \xi_2\) are determined from results already calculated.

From the eikonal equation—the last equation in (8), the derivatives of \(p_x\) with respect to \(\xi_1\) and \(\xi_2\) can also be found:
\[
\frac{\partial p_3}{\partial \xi_1} = -(p_x \frac{\partial p_1}{\partial \xi_1} + p_y \frac{\partial p_2}{\partial \xi_1})/p_3,
\]
\[
\frac{\partial p_3}{\partial \xi_2} = -(p_x \frac{\partial p_1}{\partial \xi_2} + p_y \frac{\partial p_2}{\partial \xi_2})/p_3.
\]
(29)

The derivatives for source and receiver are calculated separately using (28) and (29). There is a certain symmetry here that can be exploited since the difference between the source and the receiver is just a shift in \(x\). Once we have \(h\), together with the product of ray theoretic amplitudes, \(a\) in 1 and 2, the amplitude computation is complete.

**Numerical Implementation**

Ray data are generated along each raypath with a unique coordinate reference \((\alpha, \beta, \sigma)\). Transformation of ray data from this ray system to a uniform grid system is achieved by linear interpolation as shown in Fig. 3. Rays intersect with a \(z\)-plane. Four adjacent intersection points that surround a grid point are found. Linear interpolation is used to
determine the ray data at the grid point from the four intersection points.

This scheme is not accurate in the vicinity where rays are near turning because the distance between the intersection points will be too large. To deal with rays propagating nearly horizontally, that is, when $p_3$ is small, vertical interpolation is used. Vertical interpolation differs from the above scheme in that it use a vertical plane instead of a horizontal $z$-plane. The nearly horizontally propagating rays will intersect the vertical plane with a closer spacing leading to a more accurate interpolation than if we used a horizontal interpolation grid.

Now we have all the ray data we need on the uniform $(x, y, z)$ grid. Use these ray data to calculate the Jacobian $J$, ray amplitude $a$ and Beylkin determinant $h$. Combining these values gives the amplitude weight value at each grid point for that particular source point. This amplitude table along with the traveltine table will be used in the inversion, it is just a shift and interpolation for other source or receiver points as the ray data are invariant for lateral displacement.

Example: Constant Wavespeed

In this section, we present the specialization of the above results for a constant wavespeed medium where all calculations can be carried out explicitly. In this case, equation (17) reduces to

$$J = \frac{r^2}{c} \sin \beta,$$  \hfill (30)

and equation (16) becomes,

$$A = \frac{1}{4\pi r}.$$  \hfill (31)

For the special case of constant wavespeed, with a generally non-flat surface, some simplification of $h$ is possible. Application of Gaussian elimination produces the following simplified forms for $h_a$, $h_b$, $h_c$, and $h_d$ in (27)

$$h_a(y, \xi) \equiv \frac{-1}{c^2 r_y^2} \xi_y \cdot \frac{\partial x_y}{\partial \xi_y} \cdot \frac{\partial p_y}{\partial \xi_y},$$

$$h_b(y, \xi) \equiv \frac{-1}{c^2 r_y^2} \xi_y \cdot \frac{\partial p_y}{\partial \xi_y} \cdot \frac{\partial x_y}{\partial \xi_y},$$  \hfill (32)

$$h_c(y, \xi) \equiv \frac{-1}{c^2 r_y^2} \xi_y \cdot \frac{\partial x_y}{\partial \xi_y} \cdot \frac{\partial x_y}{\partial \xi_y},$$

$$h_d(y, \xi) \equiv \frac{-1}{c^2 r_y^2} \xi_y \cdot \frac{\partial x_y}{\partial \xi_y} \cdot \frac{\partial p_y}{\partial \xi_y}.$$

For the special case of constant wavespeed, with a flat recording surface, $h_a$, $h_b$, $h_c$, and $h_d$ become

$$h_a \equiv \frac{y_3}{c^2 r_y^2}, \quad h_b \equiv \frac{y_3}{c^2 r_y^2},$$

$$h_c \equiv \frac{y_3 \cos 2\theta}{c^2 r_y^2 r_y}, \quad h_d \equiv \frac{y_3}{c^2 r_y^2 r_y}.$$  \hfill (33)

Substituting these results into equation (26), the full expression for the Beylkin determinant may be written as

$$h(y, \xi) = 2 \cos^2 \theta \left[ \frac{(r_x + r_y)(r_y^2 + r_z^2)}{r_y^2 r_z^2} \right]$$  \hfill (34)

The corresponding formulas for $\beta(y)$ and $\beta_1(y)$ for 3D common-offset, constant-wavespeed, and with a flat
recording surface become

\[ \beta(y) = \frac{2\psi_y}{c} \int d^2 \xi \left[ \frac{(r_x + r_g)(r_x^2 + r_g^2)}{r_x^2 r_g^2} \right] \cos \theta \int i \omega \, d\omega \, e^{-i\omega(r_x + r_g)/c} \mathcal{U}_S(x_g, x_s, \omega), \]  

and

\[ \beta_1(y) = \frac{\psi_y}{c} \int d^2 \xi \left[ \frac{(r_x + r_g)(r_x^2 + r_g^2)}{r_x^2 r_g^2} \right] \int i \omega \, d\omega \, e^{-i\omega(r_x + r_g)/c} \mathcal{U}_S(x_g, x_s, \omega). \]  

Equation (36) exactly matches equation (30) in (Sullivan & Cohen, 1987), which was derived using a different approach.

Below is a list of comparison of the numerical results and the analytic results of a constant background. The grids are $40 \times 40 \times 40$ and the ray shooting grids in $\alpha$ and $\beta$ are $15 \times 15$, which is rather coarse. The numerical results have two or three digit accuracy:

<table>
<thead>
<tr>
<th>Depth</th>
<th>Analytic</th>
<th>Numerical</th>
<th>Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>$4\Delta z$</td>
<td>0.059402</td>
<td>0.062485</td>
<td>5.19%</td>
</tr>
<tr>
<td>$8\Delta z$</td>
<td>0.287655</td>
<td>0.291268</td>
<td>1.26%</td>
</tr>
<tr>
<td>$12\Delta z$</td>
<td>0.483947</td>
<td>0.490268</td>
<td>1.31%</td>
</tr>
<tr>
<td>$16\Delta z$</td>
<td>0.638489</td>
<td>0.644558</td>
<td>0.95%</td>
</tr>
<tr>
<td>$20\Delta z$</td>
<td>0.754078</td>
<td>0.759777</td>
<td>0.76%</td>
</tr>
<tr>
<td>$24\Delta z$</td>
<td>0.838724</td>
<td>0.843727</td>
<td>0.60%</td>
</tr>
<tr>
<td>$28\Delta z$</td>
<td>0.900615</td>
<td>0.905202</td>
<td>0.51%</td>
</tr>
<tr>
<td>$32\Delta z$</td>
<td>0.946287</td>
<td>0.950361</td>
<td>0.43%</td>
</tr>
<tr>
<td>$36\Delta z$</td>
<td>0.980478</td>
<td>0.983688</td>
<td>0.33%</td>
</tr>
<tr>
<td>$40\Delta z$</td>
<td>1.017178</td>
<td>1.020043</td>
<td>0.28%</td>
</tr>
</tbody>
</table>

These are the amplitude weights on one trace (with a 4-sample skip) at offset 10\Delta x. Notice that the accuracy increases with depth, this is because the polar angle $\beta$ associated with the rays increases with depth and thus has a closer ray spacing for interpolation. The computation of travel time was about 10 times more accurate.

**Inversion Examples**

The first example provides a simple test of positioning and amplitude accuracy. The model consists of constant velocity layers, as shown in Figure 4. Layer velocities are 2, 3, 6 and 10km/s from the top to the bottom. Figure 5 shows the model data with offset=3km. The inversion produces two outputs: one for $\beta$ and one for $\beta_1$. The amplitudes in the two outputs differ by the factor $\cos \theta$, where $\theta$ is the specular angle. For the first interface, the real values: R is 0.200, Rcos$\theta$ is 0.172 and $\theta$ is 30.90 degrees. The inversion results: R is 0.203, Rcos$\theta$ is 0.174 and $\theta$ is 30.84 degrees. The percentage errors are 1.5, 1.1 and 0.2 percent respectively. For the second interface, the exact values: R is 0.333, Rcos$\theta$ is 0.307 and $\theta$ is 22.76 degrees. The inversion results: R is 0.341, Rcos$\theta$ is 0.315 and $\theta$ is 22.71 degrees. The percentage errors are 2.4, 2.6 and 0.2 percent respectively. For the third interface, the exact values: R is 0.250, Rcos$\theta$ is 0.235 and $\theta$ is 19.94 degrees. The inversion results: R is 0.256, Rcos$\theta$ is 0.241 and $\theta$ is 19.89 degrees. The percentage errors are 2.3, 2.6 and 0.3 percent respectively.

Another example in Figure 7 and Figure 8 shows a synthetic dataset and its inversion results. The model is
a spherical object located in a constant velocity medium. The synthetic data is generated in 3D with an offset of approximately five times the diameter of the object. The imaging is done with GOCAD. In the inversion image, an isosurface is created for the peak amplitude so that the spherical object can be shown in 3D without the diffraction events. The reflection coefficient at the normal incident position (the top of the sphere) has an error of only 3.1 percent. The position around the top of the sphere is comparably accurate except in the neighborhood of the equator, where we have no specular returns. There, the sphere is filled in by diffraction return rather than by true specular points on the reflector. The lower half of the sphere is inaccurately positioned because the change in propagation speed across the upper sphere surface does not satisfy the criterion of being a depth dependent background velocity model. Thus, I used a constant velocity (1.5 km/s) down to the water/seabed interface and another constant below (2.0 km/s). Both are lower than the true velocity of 6 km/s in the sphere. Since the data were generated with the correct velocity, the specular returns from the lower part of the sphere came sooner than they would have at 2.0 km/s. Thus, in the imaging, the lower part of the sphere is pulled up, making a more elliptical surface."

Conclusions
I have described a method to calculate the amplitude weights for the 3-D common offset v(z) inversion by ray tracing. The traveltime and amplitude tables generated by this method are used in (1) to perform Kirchhoff inversion. Tests on synthetic data are in progress to improve the program and the algorithm.
Acknowledgements
I am grateful to professor N. Bleistein and J. Cohen, who have provided important comments and insights. I would also like to thank the sponsors of the Consortium Project at the Center for Wave Phenomena and Office of Naval Research for their financial support. Thanks for the computing facilities necessary for this research supported by the National Science Foundation under grant DMS-9506603 and Los Alamos National Laboratory.

References
3-D DMO operators in transversely isotropic media

Tariq Alkhalifah
PhD Candidate, Center for Wave Phenomena
Colorado School of Mines

ABSTRACT

The 3-D dip-moveout (DMO) operator in homogeneous transversely isotropic media with vertical symmetry axis (VTI media), unlike that in homogeneous isotropic media, has an out-of-plane (crossline) component. In general, this additional component has a shape that is opposite to that of the crossline component of the isotropic $v(z)$ saddle shaped operator. The width of the crossline component of the VTI operator, in general, is overall smaller than that associated with isotropic $v(z)$ media. When both typical anisotropy and inhomogeneity are combined, the net result is an operator with a smaller crossline component and a shape that is more influenced by the $v(z)$ behavior. The large cost of a 3-D DMO operator, as well as the generally small crossline components associated with the DMO operator in VTI media, suggests the possibility, as is usually done for isotropic $v(z)$ media, of just ignoring the contribution of the out-of-plane portion of the operator.

Key words: transverse isotropy, dip moveout, 3D

Introduction

The dip-moveout (DMO) operator (impulse response) for isotropic homogeneous media in the common-offset domain is two-dimensional, with the shape of an ellipse. The parameters of this ellipse are controlled by the source-receiver offset, $X$, and the normal moveout (NMO) corrected time, $t_n$. In 3-D homogeneous isotropic media, the DMO operator remains an ellipse in the vertical plane that contains the sources and receivers (Figure 1a). On the other hand, as Meinnardus and Schleicher (1993), Dietrich and Cohen (1992), and Artley et al. (1993) have shown, that the DMO operator in vertically inhomogeneous $[v(z)]$ isotropic media is the familiar saddle shape surface (Figure 1b).

In the inline plane (i.e., the vertical plane containing the source and receiver locations), the isotropic $v(z)$ operator can be well approximated by a squeezed version of the isotropic homogeneous ellipse (Hale and Artley, 1993). This approximation, however, ignores the crossline component of the $v(z)$ operator. In fact, the crossline component of the DMO operator is often ignored in practice in favor of efficiency and accuracy. It is believed that, in many cases, the $v(z)$ influence on the operator is counteracted by the presence of anisotropy (Gonzalez et al., 1992), so that better focusing is obtained by using the simple homogeneous operator rather than a $v(z)$ one. Alkhalifah (1996), however, showed that the influence of anisotropy on the inline component can exceed the influence of $v(z)$, requiring, in many cases, a more advanced treatment, namely use of an anisotropic $v(z)$ operator.

Although Figure 1b shows the full shape of the 3-D DMO operator, the perspective plot does not reveal the details needed for comparisons of the kinematics of DMO responses. Such details can be better seen by displaying the inline and crossline components of the operator, separately. Thus, for example, Figures 2 and 3 show inline and crossline cross-sections (through the apex) of the DMO responses shown in Figure 1 for isotropic homogeneous and $v(z)$ media, respectively. This format will be used for most of the comparisons discussed in this paper. All operators shown correspond to an offset, $X=1.5$ km/s, and a $t_n=1.75$ s. Variation in these parameters mainly influences the size of the operator, which is predictable but is beyond the scope of this note. For the homogeneous medium, the background velocity equals 2.0 km/s, whereas for the $v(z)$ media, velocity increases lin-
early with depth with a constant velocity gradient, \( k = 0.5 \) s\(^{-1}\) and an initial velocity at the surface, \( v_0 = 1.5 \) km/s. Both velocity models result in the same root-mean-square (rms) velocity of 2 km/s at \( t_0 = 1.75 \) s.

The crossline component of the isotropic homogeneous DMO operator, in Figure 2, is given by a single dot that represents the intersection of the crossline plane with the 2-D inline ellipse at its apex. The inline and crossline components of the \( v(z) \) isotropic operator in Figure 3 are largely as expected for the saddle shape shown in Figure 1b. The arrows in Figures 2 and 3 point to the reflector dip treated by this portion of the operator. Beyond a dip of about 75 degrees, the inline component shows a duplication that could not be clearly depicted in Figure 1b and was therefore not included in it.

Alkhalil and Tsvankin (1995) demonstrated that, for TI media with vertical symmetry axis (VTI media), just two parameters are sufficient for performing all time-related processing, such as normal moveout (NMO) correction (including non-hyperbolic moveout correction, if necessary), dip-moveout (DMO) correction, and prestack and poststack time migration. One of these two parameters, the short-spread NMO velocity for a horizontal reflector, is given by

\[
V_{\text{NMO}}(0) = V_p \sqrt{1 - 2\delta}.
\]

Taking \( V_p \) to be the P-wave velocity in the horizontal direction, the other parameter, \( \eta \), is given by

\[
\eta \equiv \frac{1}{2} \left( \frac{V_p}{V_{\text{NMO}}(0)} - 1 \right) = \frac{\epsilon - \delta}{1 + 2\delta},
\]

where \( V_p \) is the P-wave vertical velocity, and \( \epsilon \) and \( \delta \) are Thomason’s (1986) dimensionless anisotropy parameters. Moreover, Alkhalil and Tsvankin (1995) show that these two parameters, \( \eta \) and \( V_{\text{NMO}} \), are obtainable solely from surface seismic P-wave data, specifically from estimates of stacking velocity for reflections from interfaces having two distinct dips (the DMO method). The two-parameter representation and inversion also holds in \( v(z) \) media. For DMO applications in two dimensions, Alkhalil (1996) showed that \( \eta \) is the parameter responsible for the shape of the operator, whereas \( V_{\text{NMO}}(0) \) controls mainly the size of the operator, and a positive velocity gradient has the tendency to squeeze the operator.
and
\[ p_{yz} = \cos[\theta(p_y, t_y)] s(p_y, t_y), \]
where \( s \) is the slowness and \( \theta \) is the phase angle, both of which are calculated using ray tracing and tabulated as a function of ray parameter \( p \) and the traveltime \( t \). Then
\[ \lambda = \frac{\cos[\theta(p_0, t_0)] s(p_0, t_0)}{\cos[\theta(p_z, t_z)] s(p_z, t_z)} \frac{s(p_z, t_z)}{s(p_0, t_0)}. \]

Substituting equation (4) into the \( x \)- and \( y \)-components of equation (3) provides two of the six nonlinear equations needed to be solved.

The other four equations are given by
\[ 0 = \xi(p_z, t_z) \cos \phi_z - \xi(p_x, t_x) \cos \phi_x + 2h \]
(5)
\[ 0 = \xi(p_z, t_z) \sin \phi_z - \xi(p_x, t_x) \sin \phi_x \]
(6)
\[ 0 = \tau(p_0, t_0) - \tau(p_z, t_z) \]
(7)
\[ 0 = \tau(p_0, t_0) - \tau(p_y, t_y). \]
(8)

Equation (5) is the requirement that the surface distances, \( \xi \), along the inline component from each of the source and receiver to the specular reflection point (SRP) add up to equal the source-receiver offset, \( 2h \), and equation (6) is the requirement that the distances along the crossline component to SRP are equal for each of the source and receiver. Equations (7) and (8) imply that the vertical times, \( \tau \) (or depths) from each of the source, receiver, and zero-offset surface positions to SRP are the same. Like the slowness and phase angle, \( \xi \) and \( \tau \) are calculated using ray tracing and then stored in a table as a function of ray parameter \( p \) and the traveltime \( t \).

This system of equations, like its 2-D counterpart (Artley and Hale, 1994; Alkhalifah, 1996) can be solved using Newton-Raphson iterative method (e.g., Press et al., 1986).

3-D DMO Operators in VTI Media

Figure 4 shows the inline and crossline components of the 3-D DMO operator in a homogeneous VTI medium, with \( V_{nmo} = 2.0 \) km/s and \( \eta = 0.1 \). The inline component is the same as the 2-D operator shown by Alkhalifah (1996). Unlike the isotropic DMO operator, the anisotropic one has an out-of-plane component in a homogeneous medium. This out-of-plane component, moreover, is very different from that associated with isotropic \( v(z) \) media. The anisotropic operator is smaller in size than the \( v(z) \) one, and it is initially concaved upward. Based on the dip distribution along this operator, a reflection from an interface that is dipping at 75 degrees or more in crossline direction will be subjected to a simple time shift (almost no lateral migration). This time shift is needed to compensate for the difference between \( V_{nmo}(0) \), the velocity

Generating the 3-D DMO Operator

Following Artley et al. (1993), I construct the 3-D DMO operator by solving a system of six nonlinear equations to obtain six unknowns that include, among other things, the zero-offset time and surface position of the specular reflection point. Artley’s travel times are calculated and tabulated using an isotropic \( v(z) \) ray tracing. To construct the 3-D DMO operator for VTI media, modifications to Artley’s method include honoring Snell’s law for anisotropic media (recognizing the dependence of phase velocity on direction) as well as using anisotropic ray tracing to calculate travel times (Alkhalifah, 1996).

In either isotropic or anisotropic media, the zero-offset slowness vector \( p_0 \) is a scaled sum of the slownesses of the rays from the source \( p_s \) and receiver \( p_R \) to the specular point reflection. Thus,
\[ p_0 = \lambda(p_s + p_R). \]
(3)

Considering the \( z \)-component gives
\[ p_{0z} = \lambda(p_{sz} + p_{Rz}), \]
then
\[ \lambda = \frac{p_{0z}}{p_{sz} + p_{Rz}}. \]

Since
\[ p_{0z} = \cos[\theta(p_0, t_0)] s(p_0, t_0), \]
\[ p_{sz} = \cos[\theta(p_z, t_z)] s(p_z, t_z), \]
and
\[ p_{Rz} = \cos[\theta(p_R, t_R)] s(p_R, t_R), \]
used in the NMO correction, and $V_h$, the actual normal moveout velocity for such a dip. This somewhat explains the presence of the duplication in the crossline component.

The inline and crossline components of the DMO operator for a stronger anisotropy ($\eta=0.2$) are shown in Figure 5. The DMO operator is, in general, similar to that in Figure 4, with the features somewhat exaggerated: the inline component is further stretched with respect to the ellipse shape and exhibits triplication (Alkhalifah, 1996), and crossline component is larger with an even greater time shift for the reflector dip of 75 degrees (and larger) due to the larger difference between $V_h$ and $V_{hmo}(0)$.

The combined influence of anisotropy and inhomogeneity on the DMO operator is shown in Figure 6. The medium is characterized by $\eta = 0.1$ and $V_{hmo} = 1.5 + 0.6z$ km/s. Again, the inline component is the same as Alkhalifah (1996) and described in detail there. The crossline component has an overall concaved downward shape with a small concave upward region caused by the anisotropy. Nevertheless, the domination of $v(z)$ on the operator is apparent and, as a result, the general shape is largely a distorted saddle. Although, Alkhalifah (1996) demonstrated that, for a medium with the same inhomogeneity and anisotropy, the inline component (the 2-D operator) is dominated more by the anisotropy than by the vertical inhomogeneity (which results in an overall stretched operator rather than a squeezed one), the crossline component is dominated more by the inhomogeneity.

Note that the crossline component in Figure 6 is smaller than that in Figure 3, which implies that the anisotropy reduced the influence of $v(z)$ on the operator. In fact, the crossline component is interesting because, up to 30-degree reflector dip, the crossline component is practically a point, similar to that for the homogeneous isotropic operator. The influences of inhomogeneity and anisotropy do not fully cancel, as Gonzalez et al. (1992) suggested, but less harm will be introduced in the data when this crossline component is ignored than when it is ignored for isotropic $v(z)$ media.

A stronger anisotropy, given by $\eta=0.2$, gives a clearer picture of the triplication shown in Figure 6 (see Figure 7). Again the crossline component is smaller in size than the crossline component for purely isotropic $v(z)$ media.

**Cost Issues**

The computation cost for 3-D DMO operators is of an order magnitude higher than that for 2-D ones. Thus, for a simple Kirchhoff-type of DMO implementation (Deregowski, 1987), the cost of the 3-D DMO operator is about the cost of the 2-D operator multiplied by the
number of grid points needed to represent the crosline component of the 3-D DMO operator. In many cases, this additional cost may be considered to be unacceptable. Specifically, applying such a 3-D DMO followed by a post-stack time migration is comparable to the cost of a full 3-D prestack time migration.

Where seismic data are acquired along dip lines, 3-D DMO is not required, and therefore 2-D operators are dominant in practice. If ignoring the out-plane component is acceptable in v(z) isotropic media, it should be more so in both homogeneous and v(z) VTI media because, in VTI media, the operator generally has a smaller crosline component than does that for typical isotropic v(z) media.

On the other hand, if the dominant reflector-dip direction was not taken into consideration during acquisition, dips with large crosline components will require use of a 3-D operator, whether the medium is isotropic or anisotropic.

Discussion and Conclusions

The 3-D DMO operator in homogeneous VTI media, unlike that in homogeneous isotropic media, has an out-of-plane (crosline) component. In general, this additional component has a shape that is opposite to that of the crosline component of the isotropic v(z) saddle shaped operator. The crosline component of the VTI (η = 0.1) operator, as well as the inline component, are initially concaved upward (Figure 8a). The influence of a stronger anisotropy (η = 0.2) increases the breadth of the operator (Figure 8a). Nevertheless, the width of the crosline component of the VTI operator, in general, is overall smaller than that associated with isotropic v(z) media. When both typical anisotropy and inhomogeneity are combined, the net result is an operator with a smaller crosline component and a shape that is more influenced by the v(z) behavior (a saddle shape, see Figure 8b). In this case, the crosline component becomes less significant, with dips up to 30 degrees requiring practically no adjustment. Nevertheless, the inline component is controlled more by the anisotropy, causing the operator to stretch, rather than squeeze, relative to the elliptical operator for isotropic homogeneous media.

The large cost of a 3-D DMO operator, as well as the generally small crosline components associated with the DMO operator in VTI media, suggests the possibility, as is usually done for isotropic v(z) media, of just ignoring the contribution of the out-of-plane portion of the operator. Although on occasion it might be necessary to use a 3-D DMO operator, relying on a 2-D operator is an attractive option that preserves the efficiency of a poststack migration processing sequence.

Acknowledgments

I thank Ken Larner for helpful discussions at the critical stages of this study and his review of the manuscript. This research is partially supported by the Advanced Computational Technology Initiative (ACTI), subcontract number B316649, in conjunction with Lawrence Livermore National Laboratory and industry project partners. It is also partially supported by the Center for Wave Phenomena Consortium Project at the Colorado School of Mines. I also wish to acknowledge the financial support from KACST (Saudi Arabia).

REFERENCES


Figure 8. 3-D DMO operators for (a) a homogeneous VTI medium with velocity equal 2.0 km/s, and (b) a vertically inhomogeneous VTI medium with velocity given by $v(z)=1.5+0.6z$ km/s. For both models, $\eta=0.1$. The operator, again, corresponds to an offset of 1.5 km.

Figure 9. 3-D DMO operators for (a) a homogeneous VTI medium with velocity equal 2.0 km/s, and (b) a vertically inhomogeneous VTI medium with velocity given by $v(z)=1.3+0.6z$ km/s. For both models, $\eta$ now equals 0.2.
Multi-parameter DMO
in anisotropic media
using the generalized Radon transform

MAARTEN V. DE HOOP AND TARIQ ALKHALIFAH
Center for Wave Phenomena,
Colorado School of Mines,
Golden CO 80401-1887, USA.

April 10, 1996

Abstract

Dip moveout (DMO), as part of the transformation to zero-offset (TZO), is a process that transforms data collected at finite source-receiver offsets to data that would be collected at zero-offset. The kinematic or ray geometric principles underlying the DMO process have been well established. The aim of this paper is to extend the TZO and DMO processes to elastic media and develop dynamically correct transformations of wave field data. We accomplish this goal with the aid of the generalized Radon transform (GRT). We allow the media to be generically anisotropic.

Keywords: dip moveout, migration/demigration, generalized Radon transform.
1 Introduction

Dip moveout (DMO), as an intermediate step in the transformation to zero-offset (TZO), is a process that transforms data collected at finite source-receiver offsets to data that would be collected at zero-offset. In practice, true zero-offset data cannot be measured, but even if this were possible, the use of TZO applied to data from multiple offsets would enhance the signal-to-noise ratio in the data as in the conventional stack. In the past decade, the kinematic or ray-geometric principles underlying the DMO process have been well established. It is only recently, however, that the associated amplitude aspects have received attention. The aim of this paper is to extend the Kirchhoff-style TZO and DMO processes to elastic media, and to develop dynamically correct transformations of wavefield data. We accomplish this goal with the aid of the generalized Radon transform (GRT). We consider general, anisotropic media, and provide closed-form expressions for the constant-background-medium case.

For an overview of papers on DMO, we refer the reader to Hale [1]. The key idea of a wave-theoretic representation of TZO is to apply a migration/inversion to finite-offset data, followed by a modelling of or demigrating to, in the Born approximation, the zero-offset configuration; see Sullivan and Cohen [2] and Bleistein and Jorden [3]. In the acoustic case, Miller and Burridge [4] applied the GRT for this purpose. A different approach, in the context of a more general offset continuation, using differential equations rather than integral transforms, was followed by Fomel [5]. The extension to depth-varying media, as far as the kinematics is concerned, was accomplished by Artley and Hale [6], and an extension to transversely isotropic media with vertical symmetry axis can be found in Uren et al. [7] and Alkhalifah [8]. Certain true-amplitude aspects were discussed by Black et al. [9] and Bleistein and Cohen [10]. Here, we will account for the full radiation characteristics of the wavefields in anisotropic media, and derive an elastic DMO procedure. Our theory is based on the ray-Born approximation. Related work on DMO and Born inversion can be found in Liner [11].

In this paper, we consider single-mode scattering, in particular the single scattering associated with the fastest characteristics (qP). We also briefly indicate how to extend the theory to the multi-mode, or mode-converted, case.
2 The basic equations

Notation

First, we introduce some basic notation. Choose coordinates in the configuration according to

\[ \mathbf{x} = (x_1, x_2, x_3) = \text{Cartesian position vector}, \]
\[ s = (s_1, s_2, s_3) = \text{source point}, \]
\[ r = (r_1, r_2, r_3) = \text{receiver point}, \]
\[ t = \text{recording time}. \]

The medium is described by

\[ \rho(\mathbf{x}) = \text{density}, \]
\[ c_{ijkt}(\mathbf{x}) = \text{elastic stiffness tensor}, \]

while the wavefield is described by

\[ \mathbf{u}(\mathbf{x}, t) = (u_1(\mathbf{x}, t), u_2(\mathbf{x}, t), u_3(\mathbf{x}, t)) = \text{displacement vector}, \]

and generated by a source distribution given by

\[ \mathbf{f}(\mathbf{x}, t) = (f_1(\mathbf{x}, t), f_2(\mathbf{x}, t), f_3(\mathbf{x}, t)) = \text{body-force source density}. \]

In the remainder of the paper, we will employ the summation convention.

The displacement in an inhomogeneous, anisotropic medium satisfies the elastodynamic wave equation

\[ \rho \partial_t^2 u_i - \partial_j (c_{ijkt} \partial_t u_k) = f_i , \]  \hspace{1cm} (2.1)

with summation over repeated lower case indices, here and below. Let

\[ \mathbf{G}(\mathbf{x}, \mathbf{x}', t) = (G_{ip}(\mathbf{x}, \mathbf{x}', t)) \]  \hspace{1cm} (2.2)

be the causal Green's tensor, which satisfies (cf. Eq.(2.1))

\[ \rho \partial_t^2 G_{ip} - \partial_j (c_{ijkt} \partial_t G_{kp}) = \delta_{ip} \delta(\mathbf{x} - \mathbf{x}') \delta(t) , \quad G_{ip} = 0 \text{ for } t < 0 . \]  \hspace{1cm} (2.3)
Asymptotic ray theory

Here, we summarize the formulation of anisotropic ray theory for the evaluation of the Green's tensor (see Kendall et al. [12] for details). Let

\[ G_{ip}(x, x', t) = \sum_{N} A^{(N)}(x, x') \xi^{(N)}_i(x) \xi^{(N)}_p(x') \delta(t - \tau^{(N)}(x, x')) \]

(2.4) + terms smoother in \( t \).

In this equation, the arrival time \( \tau^{(N)} \) and the associated polarization vector \( \xi^{(N)} \) satisfy

(\( \rho \delta_{ik} - c_{ijkl}(\partial_k \tau^{(N)})(\partial_l \tau^{(N)}) \)) \( \xi_k^{(N)} = 0 \) (at all \( x \)),

(2.5)

which implies the eikonal equation

\[ \det(\rho \delta_{ik} - c_{ijkl}(\partial_k \tau)(\partial_l \tau)) = 0 \] (at all \( x \)).

(2.6)

The polarization vectors are assumed to be normalized so that \( \xi_i^{(N)} \xi_i^{(N)} = 1 \). Define the slowness vector \( \gamma^{(N)} \) by

\[ \gamma^{(N)}(x) = \nabla_{\xi} \tau^{(N)}(x, x') \]

(2.7)

Then, Eq.(2.6) constrains \( \gamma \) to lie on the sextic surface \( A(x) \) given by

\[ \det(\rho \delta_{ik} - c_{ijkl} \gamma^j \gamma^j) = 0 \].

(2.8)

\( A(x) \) consists of three sheets \( A^{(N)}(x), N = 1, 2, 3 \), each of which is a closed surface surrounding the origin. An individual sheet is described by (cf. Eq.(2.8))

\[ 2 \mathcal{H} = \rho - \xi_i c_{ijkl} \gamma^j \gamma^j \xi_k = 0 \].

(2.9)

The scalar amplitudes \( A \) must satisfy the transport equation

\[ \partial_j (c_{ijkl} \xi^{(N)}_i \xi^{(N)}_k (A^{(N)})^2 \partial_l \tau^{(N)}) = 0 \],

(2.10)

where \( N \), again, indicates the mode of propagation, that is, the sheet of the slowness surface on which the corresponding slowness vector lies.

The characteristic or group velocities \( v^{(N)} \) are normal to \( A^{(N)}(x) \) at \( \gamma^{(N)} \) and satisfy

\[ v^{(N)} \cdot \gamma^{(N)} = 1 ; \quad v^{(N)} = \frac{\nabla \gamma \mathcal{H}}{\gamma \cdot \nabla \gamma \mathcal{H}} \bigg|_{\mathcal{H}=0} \].

(2.11)

The normal or phase speeds are given by

\[ V^{(N)} = \frac{1}{|\gamma^{(N)}|} \].

(2.12)
The unit phase direction follows as
\[ \alpha^{(N)} = V^{(N)} \gamma^{(N)} . \]

From Eq. (2.11) it follows that
\[ V^{(N)} = |v^{(N)}| \cos \chi , \]

where \( \chi \) is the angle between \( v^{(N)} \) and \( \gamma^{(N)} \).

The amplitudes can be expressed in terms of certain Jacobians,
\[ A = \frac{1}{4 \pi |\rho(x) \rho(x')|^{1/2}} \quad \text{with} \quad M = \frac{|v(x')|V(x) \left| \frac{\partial x}{\partial q_1} \wedge \frac{\partial x}{\partial q_2} \right|}{\left| \frac{\partial \gamma}{\partial q_1} \wedge \frac{\partial \gamma}{\partial q_2} \right| x'} , \]

in which \( A \) and \( M \) carry the superscript \((N)\). Here, \((q_1, q_2)\) parametrize the rays originating from the source. We will substitute the phase angles, \( \alpha \), at the source for those parameters. The wedge denotes the vector cross product.

**Source and receiver Green's functions**

In the integral representation for the scattered field, we need the Green's functions originating both at the source and the receiver points. Further, the gradient of total travel times from the source to a scattering point to the receiver are required in preparation of the inverse GRT. We introduce these functions here.

Set
\[ \hat{G}(x, t) = G(x, s, t) , \quad \hat{G}(x, t) = G(r, x, t) . \]

Employing asymptotic ray theory in both Green's functions, we introduce the notation
\[ \hat{\tau}^{(N)}(x) = \tau^{(N)}(x, s) , \quad \hat{\tau}^{(M)}(x) = \tau^{(M)}(r, x) \]

and
\[ \hat{A}^{(N)}(x) = A^{(N)}(x, s) , \quad \hat{A}^{(M)}(x) = A^{(M)}(r, x) \]

in the case of scattering from incident mode \( N \) to outgoing mode \( M \).

According to Eq. (2.7), the slowness vectors at \( x \) are given by
\[ \hat{\gamma}^{(N)}(x) = \nabla_x \tau^{(N)}(x, s) , \quad \hat{\gamma}^{(M)}(x) = \nabla_x \tau^{(M)}(r, x) ; \]
the associated phase directions are given by

$$\hat{\alpha}^{(N)} = \frac{\tilde{\gamma}^{(N)}}{|\tilde{\gamma}^{(N)}|}, \quad \hat{\alpha}^{(M)} = \frac{\tilde{\gamma}^{(M)}}{|\tilde{\gamma}^{(M)}|}$$

(2.19)

and the normal (phase) speeds (cf. Eq.(2.12)) are given by

$$\tilde{\nu}^{(N)} = \frac{1}{|\tilde{\gamma}^{(N)}|}, \quad \tilde{\nu}^{(M)} = \frac{1}{|\tilde{\gamma}^{(M)}|}.$$  

(2.20)

The unit group directions are given by

$$\hat{\nu}^{(N)} = \frac{\tilde{\nu}^{(N)}}{|\tilde{\nu}^{(N)}|}, \quad \hat{\nu}^{(M)} = \frac{\tilde{\nu}^{(M)}}{|\tilde{\nu}^{(M)}|}.$$  

(2.21)

We also define the two-way travel time $T^{(N,M)}$ and its gradient,

$$T^{(N,M)}(r, y, s) \equiv \tau^{(N)}(y, s) + \tau^{(M)}(r, y), \quad \Gamma^{(N,M)}(r, x, s) \equiv \nabla_x T^{(N,M)}(r, x, s).$$  

(2.22)

From Eq.(2.18) we see that

$$\Gamma^{(N,M)}(r, x, s) = \tilde{\nu}^{(N)}(x) + \tilde{\nu}^{(M)}(x).$$  

(2.23)

The direction of $\Gamma^{(N,M)}$,

$$\nu \equiv \frac{\Gamma^{(N,M)}}{|\Gamma^{(N,M)}|},$$

will be the migration dip; it is the normal to the hypersurface of constant two-way travel time.

### 3 The single scattering equations

In this section, we introduce the ray-Born approximation representing the singly scattered waveform (see De Hoop et al. [13] and Burridge et al. [14]). For non-coincident, finite-offset, source-receiver pairs, we let $(r, s) \in \partial D \times \partial \mathcal{S}$; ideally, the boundaries $\partial D, \partial \mathcal{S} \sim S^2$ are closed surfaces surrounding a heterogeneous domain $\mathcal{D}$. For coinciding, zero-offset, source-receiver pairs, we employ the notation $s = r = S \in \partial \mathcal{S}$. We restrict ourselves to waves propagating in mode $N$.

**The wavefield at finite offset**

In the volume-scattering, ray-Born approximation, the multi-component ($q$ source and $p$ receiver) scattered displacement field is given by

$$u^{(1)}_{pq}(r, s, t) \simeq -\int_{\mathcal{D}} \tilde{\zeta}^{(N)}(r) \tilde{\zeta}^{(N)}(s) \Gamma^{(N,N)}(r, x, s) \times$$

$$\left(\mathbf{w}^{(N,N)}(x, \hat{\alpha}^{(N)}(x), \hat{\alpha}^{(N)}(x)))^T \mathbf{c}^{(1)}(x) \delta''(t - T^{(N,N)}(r, x, s)) \right) \, dx.$$  

(3.1)
where \( N \in \{1, 2, 3\} \),

\[
T^{(N \tilde{N})}(r, x, s) = \rho(x) \tilde{A}^{(\tilde{N})}(x) \tilde{A}^{(N)}(x)
\]  

(3.2)
contains the geometrical spreading, and

\[
w^{N\tilde{N}} = \left\{ s_{\hat{N}i} s_{\hat{N}i}, \frac{1}{2} \left[ \tilde{a}_{ij} \tilde{a}_{k\ell} + \tilde{a}_{k\ell} \tilde{a}_{ij} \right] \right\},
\]

\[
\tilde{a}^{N\tilde{N}}_{ij} = \frac{1}{2} V_o^{(N)} (\tilde{\xi}_{i}^{(N)} \tilde{z}_{j}^{(N)} + \tilde{\xi}_{j}^{(N)} \tilde{z}_{i}^{(N)}),
\]

\[
\tilde{a}^{\tilde{N}N}_{k\ell} = \frac{1}{2} V_o^{(N)} (\tilde{\xi}_{k}^{(N)} \tilde{z}_{\ell}^{(N)} + \tilde{\xi}_{\ell}^{(N)} \tilde{z}_{k}^{(N)}),
\]
describe the contrast-source radiation patterns, and

\[
c^{(N)} = \left\{ \rho^{(1)}, \frac{c_{ijkl}^{(1)}}{\rho V_o^{(N)} V_o^{(N)}} \right\},
\]

represents the relative medium perturbation (\( \rho^{(1)} \) is the absolute perturbation in density and \( c_{ijkl}^{(1)} \) is the absolute perturbation in stiffness with respect to the background medium). Here, \( V_o^{(N)} \) denotes the local phase velocity of mode \( N \) in the background medium for \( \gamma \) averaged over all phase directions. The hypersurface,

\[
\{ x \in D \mid T^{(N \tilde{N})}(r, x, s) = t \},
\]

over which the integral in Eq.(3.1) is taken, is denoted as the finite-offset isochron.

The wavefield at zero offset

To obtain the wavefield at zero offset, we substitute \( r = s = S \) in Eq.(3.1). To distinguish the zero-offset configuration from the finite-offset one, we use capital letters to indicate the components here. The rays emanating from the source or the receiver connect the same points; excluding the possibility of multi-pathing, those rays must coincide. Thus, we can omit \( \tilde{\gamma} \) and \( \tilde{x} \). In fact, we have

- \( \tilde{N} = \bar{N} \equiv N \);
- \( \tilde{\gamma} = \bar{\gamma} \equiv \gamma; \tilde{\alpha} = \bar{\alpha} \equiv \alpha \);
- \( \tilde{\xi} = \bar{\xi} \equiv \xi \).

Set the two-way travel time and its gradient to (cf. Eq.(2.22))

\[
T^{(N \bar{N})}(s, x, s) \equiv T^{(N)}(x, S), \quad \Gamma^{(N)}(x, S) = \nabla_x T^{(N)}(x, S) = 2 \gamma^{(N)}(x).
\]

(3.3)

It hence follows that

\[
\alpha^{(N)} \parallel \nabla_x T^{(N \bar{N})},
\]
expressing that the incident and scattered rays coincide. For the geometrical spreading, we have
\[ \mathcal{I}^{(N)}(S, x, S) \equiv \mathcal{I}^{(N)}(x, S). \] (3.4)

Also \( \tilde{a} = \hat{a} \equiv a \). Thus, we can write
\[ w^{(N)}(x, \alpha^{(N)}(x), \alpha^{(N)}(x)) \equiv w^{(N)}(x, \alpha^{(N)}(x)) \] (3.5)
for the radiation patterns. (Note that the first element of \( w^{(N)} \), the inner product of the polarization vector with itself, becomes unity.)

With \( u^{(1)}_{pq}(S, \tau) \equiv u^{(1)}_{pq}(S, S, \tau) \), we then obtain
\[ u^{(1)}_{pq}(S, \tau) \approx -\partial^2 \int_D \xi^{(N)}_p(S) \xi^{(N)}_q(S) \mathcal{I}^{(N)}(x, S) \times \]
\[ (w^{(N)}(x, \alpha^{(N)}(x)))^T c^{(1)}(x) \delta(\tau - T^{(N)}(x, S)) \, dx. \] (3.6)

The hypersurface,
\[ \{ x \in D \mid T^{(N)}(x, S) = \tau \}, \]
over which the integral in Eq.(3.6) is taken, is denoted as the zero-offset isochron. Here, \( \tau \) is the zero-offset time.

4 Transformation to zero offset

The idea of transformation to zero offset is to eliminate the medium perturbation from the equations representing the finite-offset (Eq.(3.1)) and zero-offset (Eq.(3.6)) scattered fields. This is achieved by cascading finite-offset GRT inversion with zero-offset GRT modelling.

Linearized inversion

The general expression for GRT inversion of the scattered field Eq.(3.1), an integration over the phase directions (associated with two source coordinates and two receiver coordinates), is given in De Hoop et al. [13] and Burridge et al. [14]. First, set
\[ \mathcal{F}^{(N)}(r, x, s) = \frac{\left| \mathcal{I}^{(N)}(r, x, s) \right|^3}{\mathcal{I}^{(N)}(r, x, s)}. \] (4.1)

Then
\[ c^{(1)}(x) \approx \frac{1}{8\pi^2} \int_{S \times \partial R} \left[ \xi^{(N)}(\nu(r, x, s)) \right]^{-1} w^{(N)}(x, \tilde{\alpha}^{(N)}(x), \tilde{\alpha}^{(N)}(x)) \times \]
\[ \mathcal{F}^{(N)}(r, x, s) \xi^{(N)}_p(r) u^{(1)}_{pq}(r, s, T^{(N)}(r, x, s)) \xi^{(N)}_q(s) \frac{\partial(\tilde{\alpha}, \tilde{\alpha})}{\partial(s, r)} \, dx \, dr. \] (4.2)
where $\nu$ is the migration dip. The scattering angle $\theta$ and azimuth $\psi$ associated with the dip, are defined through

$$\cos \theta = \hat{\alpha} \cdot \hat{\alpha}, \quad \text{while} \quad \psi = \text{third Euler angle around } \nu; \tag{4.3}$$

in Eq.(4.2),

$$\Lambda^{(\tilde{N} \tilde{K})}(\nu) = \int_{E_\theta(\nu)} \int_{E_\psi(\nu,\theta)} \mathbf{w}^{(\tilde{N} \tilde{N})}(\mathbf{w}^{(\tilde{N} \tilde{N})})^T \frac{\partial (\hat{\alpha}, \hat{\alpha})}{\partial (\nu, \theta, \psi)} \, d\psi \, d\theta \tag{4.4}$$

is used to carry out the inversion of amplitude versus scattering-angle/azimuth. Here, we have explicitly incorporated the limitations of the acquisition geometry by restricting the integrations over scattering angle and azimuth to finite domains, viz. $E_\theta$ and $E_\psi$ respectively. Note that the inversion integral is effectively carried out over diffraction hypersurfaces,

$$\{ (r, t) \in \partial R \times \mathbb{R}_{\geq 0} \mid t = T^{(\tilde{N} \tilde{N})}(r, x, s) \}. \tag{5}$$

**Jacobians**

For the inversion we have to evaluate the Jacobians introduced in Eqs.(4.2) and (4.4). The Jacobian occurring in the integrated normal matrix Eq.(4.4) has been derived by Burridge et al. [14] and is given by

$$\frac{\partial (\hat{\alpha}, \hat{\alpha})}{\partial (\nu, \theta, \psi)} = \frac{\sin \theta}{1 + (\tilde{\gamma} || \tilde{\gamma} || / \Gamma^2) (\tan \tilde{x} - \tan \tilde{x}) \sin \theta}, \tag{4.5}$$

where

$$\Gamma = \Gamma^{(\tilde{N} \tilde{N})}, \quad \cos \tilde{x} = \hat{n}_\| \cdot \hat{\alpha} \quad \text{and} \quad \cos \tilde{x} = \hat{n}_\parallel \cdot \hat{\alpha}. \tag{4.6}$$

Here, $\hat{n}_\| \hat{n}$ denote the normals to the slowness surface in the azimuth plane $\psi$ at the scattering point.

The Jacobian in Eq.(4.2) is directly related to dynamic ray theory. We have the factorization

$$\frac{\partial (\hat{\alpha}, \hat{\alpha})}{\partial (s, r)} \bigg|_x = \frac{\partial (\hat{\alpha})}{\partial (s)} \bigg|_x \frac{\partial (\hat{\alpha})}{\partial (r)} \bigg|_x. \tag{4.7}$$

In general, the factors can be expressed in terms of the dynamic ray amplitudes since, like the amplitudes, they follow from a variation of the anisotropic ray tracing equations (see Eq.(2.14)). For the source side:

$$\frac{\partial (s^\Sigma)}{\partial (\hat{\alpha})} \bigg|_x = \frac{1}{16\pi^2 \rho(s) \rho(x) \tilde{V}^{(\tilde{N})}(s)(\tilde{V}^{(\tilde{N})}(x))^3 (\tilde{A}^{(\tilde{N})}(x))^2}. \tag{4.8}$$

as long as $\partial S$ in the neighborhood of $s$ coincides with the wave front $\Sigma(x, \tau^{(\tilde{N})}(s, x))$ originating at $x$; $s^\Sigma$ denotes the coordinates on $\Sigma(x, \tau^{(\tilde{N})}(s, x))$. If this is not the case, we have to
correct for the ratio of the area on $\partial S$ to the area on the wave front $\Sigma(x, \tau^{(N)}(s, x))$ at $s$ onto which it is mapped by projection along the rays. This arises from the fact that $\partial S$ is not necessarily tangent to $\Sigma(x, \tau^{(N)}(s, x))$ at $s$. It amounts to dividing the previous Jacobian by the Jacobian

$$\frac{\partial(s^\Sigma)}{\partial(s)} = (\bar{\alpha}(s) \cdot \bar{\beta}(s)),$$

where

$$\bar{\beta}(s) = \text{normal to } \partial S \sim \partial D \text{ at the source}.$$

Note that $\bar{\alpha}(s)$ is the normal to the wave front at $s$. Similar expressions hold for the receiver side.

In the case of a homogeneous background medium (in which the rays are straight), we have the following simplifications

$$\frac{\partial(\bar{\alpha})}{\partial(\bar{n})} = \frac{\bar{n} \cdot \bar{\alpha}}{\bar{\kappa}}, \quad \frac{\partial(\bar{n})}{\partial(s)} \bigg|_y = \frac{\bar{n} \cdot \bar{\beta}}{|s - y|^2},$$

where

$$\bar{\kappa} = \text{Gaussian curvature of the slowness surface at } \bar{\gamma}^{(N)}.$$

If $\partial S$ coincides with $\Sigma(x, \tau^{(N)}(s, x))$, i.e., $s = s^\Sigma$, then $\bar{\beta} = \bar{\alpha}$ and

$$\frac{\partial(s)}{\partial(\bar{\alpha})} \cdot \lambda = \frac{\bar{\kappa}|s - x|^2}{\tilde{V}^2}.$$

In applications, the constant background is an effective medium rather than the actual one. For the process of transformation to zero offset, we will employ a constant background.

**Zero-offset modelling of the inverted data**

The Dirac distribution in the integral representation Eq.(3.6) can be used to reduce the dimensionality of the integration explicitly. Reconsider the zero-offset isochron defined by

$$\{ x \in \mathbb{R}^3 \mid T^{(N)}(x, S) = \tau \},$$

where $S$ denotes the zero-offset position and $\tau$ the zero-offset time. In a homogeneous medium, this surface is just the wave front or the group-velocity surface scaled by $\tau$, i.e.,

$$x - S = \frac{1}{2} v^{(N)} \tau.$$

We choose local, curvi-linear coordinates $\sigma$ such that $(\sigma_1, \sigma_2)$ span the group surface and $\sigma_3$ is the coordinate in the direction normal to it, parallel to the phase direction. The origin of the
coordinate system is chosen to be $S$. We write $\sigma_\perp = (\sigma_1, \sigma_2)$. Then, for fixed $\sigma_3$, $x(\sigma) = \frac{1}{2} s^\Sigma$ of the previous subsection. Let $\mathrm{d}\Sigma$ denote a surface element in the wave-front surface, then

$$
\mathrm{d}x = \frac{1}{|\nabla_x T^{(N)}|} \mathrm{d}\tau \, \mathrm{d}\Sigma(x), \quad \mathrm{d}\Sigma = g \, \mathrm{d}\sigma_1 \mathrm{d}\sigma_2 ,
$$

(4.12)

with

$$
\frac{1}{|\nabla_x T^{(N)}|} = \frac{1}{2} V^{(N)}, \quad g = |\partial_{\sigma_1} x \wedge \partial_{\sigma_2} x| .
$$

Choosing $\sigma_\perp$ to be equal to the zero-offset phase angles, $\alpha^{(N)}$, at the source, the Jacobian $g$ is directly related to the zero-offset ray amplitude, $g \sim |A^{(N)}|^{-2}$, and is given by Eq.(4.10) upon replacing $|s - x|$ with $\frac{1}{2} v \tau$, using the principle of reciprocity.

In Eq.(3.6) we will now use the property

$$
\int_D \cdots \delta(t - T^{(N)}(x, S)) \, \mathrm{d}x = \int_{T^{(N)}(., S) = \tau} \cdots \frac{1}{|\nabla_x T^{(N)}|} \, \mathrm{d}\Sigma(x) .
$$

The desired zero-offset experiment then becomes

$$
u^{(1)}_{PQ}(S, \tau) \simeq -\partial_\tau^2 \int_{T^{(N)}(., S) = \tau} \xi^{(N)}_p(S) \xi^{(N)}_Q(S) I^{(N)}(., S) \times

\left( w^{(N)}(., \alpha(\cdot)) \right)^T e^{(\cdot)}(\cdot) \frac{1}{2} V^{(N)}(\cdot) \big|_{x(\sigma)} \, \mathrm{d}\Sigma .
$$

(4.13)

Substituting Eq.(4.2) into Eq.(4.13) and interchanging the order of integrations, yields

$$
u^{(1)}_{PQ}(S, \tau) \simeq -\frac{1}{16 \pi^2} \int_{\partial S} \mathrm{d}S \int_{\mathbb{R}_2^0} \partial_\tau^2 \int_{T^{(N)}(., S) = \tau} \frac{\partial_i I^{(N)}(., S)}{\partial_i I^{(N)}(., S) = \tau} J^{(\tilde{N}, N)}(r, ., s) \times

\xi^{(N)}_p(S) \xi^{(N)}_Q(S) \hat{\xi}^{(N)}_p(r) u^{(1)}_{pq}(r, s, t) \xi^{(N)}_q(s) \times

\left( w^{(N)}(., \alpha^{(N)}(\cdot)) \right)^T \tilde{A}^{(\tilde{N}, N)}(\nu(r, ., s))^{-1} w^{(\tilde{N}, N)}(., \tilde{\alpha}^{(N)}(\cdot), \tilde{\alpha}^{(N)}(\cdot)) \times

\delta(t - T^{(\tilde{N}, N)}(r, ., s)) V^{(N)}(\cdot) \left. \frac{\partial (\tilde{\alpha}_1, \tilde{\alpha}_2)}{\partial (s, r)} \right|_{x(\sigma)} \, \mathrm{d}\Sigma \, \mathrm{d}t \, \mathrm{d}s \, \mathrm{d}r .
$$

(4.14)

This expression shows that for each point in the zero-offset isochron surface, the original field at all nonzero-offset isochron surfaces through that point contribute to the zero-offset field. We next carry out a stationary-phase analysis for the integrations over $(\sigma_1, \sigma_2)$.

5 Stationary phase analysis

We employ the one-sided Fourier representation

$$
\delta(t - T^{(\tilde{N}, N)}(r, ., s)) = \frac{1}{\pi} \text{Re} \int_{\mathbb{R}_2^0} \exp \left[ i \omega(t - T^{(\tilde{N}, N)}(r, ., s)) \right] \, \mathrm{d}\omega ,
$$

(5.1)
for the Dirac distribution in Eq.(4.14). Then we apply a two-dimensional stationary-phase analysis to the integration over \((\sigma_1, \sigma_2)\). The stationary points follow from the equations

\[
\partial_{\sigma_1} T^{(N)}(r, x(\sigma_1, \sigma_2, \tau), s) = 0, \quad (5.2)
\]

\[
\partial_{\sigma_2} T^{(N)}(r, x(\sigma_1, \sigma_2, \tau), s) = 0, \quad (5.3)
\]

where \(x(\sigma_1, \sigma_2, \tau)\) follows the zero-offset isochron of Eq.(4.11). The stationary points are denoted by \((\sigma^0_\pm, \tau)\). Then (cf. Eq.(4.14))

\[
u_{PQ}(S, \tau) \simeq -\frac{1}{8\pi} \int_{\partial S \times \partial R'} \int_{t \in \mathbb{R}_{\geq 0}} \partial^2 T^{(N)}(r, s) J^{(N)}(r, s) \times
\]

\[
\frac{\xi^{(N)}_p(S) \xi^{(N)}_q(S) \xi^{(N)}_p(r) U_{pq}^{(1)}(r, s, t) \xi^{(N)}_q(s)}{w^{(N)}(r, s, \alpha^{(N)}(r), \alpha^{(N)}(s), \alpha^{(N)}(r), \alpha^{(N)}(s))} \times
\]

\[
\frac{\delta^*(t - T^{(N)}(r, s))}{\sqrt{\det(\nabla_{\sigma_+} \nabla_{\sigma_-} T^{(N)})}} V^{(N)}(r, s, \alpha, \alpha') \frac{\partial(\alpha, \alpha')}{\partial(s, r)} |x(\sigma^0_\pm, \tau)| \int g(\sigma^0_\pm, \tau) dt ds dr, \quad (5.4)
\]

where

\[
U_{pq}^{(1)}(r, s, t) = \int_0^t u_{pq}^{(1)}(r, s, t') dt', \quad (5.5)
\]

and

\[
\delta^*(t) = \frac{1}{\pi} \Re \int_{\mathbb{R}_{\geq 0}} i \exp \left[ \frac{i\pi}{4} \text{sgn}(\nabla_{\sigma_+} \nabla_{\sigma_-} T^{(N)}) \right] \exp(it) d\omega. \quad (5.6)
\]

Here, sig denotes the signature of a matrix. Observe that the stationary point \(\sigma^0_\pm\) varies with \((s, r)\). Equation (5.4) shows that for a fixed zero-offset isochron, we have to integrate over all finite-offset isochrons that 'kiss', or come tangent to, the zero-offset one, see the subsection below.

The TZO in Eq.(5.4) contains in fact four mappings: one for the geometrical spreading, one for the polarizations, one for the radiation patterns (or reflection coefficients), and one for the coordinates. The bare TZO accounts for the final mapping only, see Deregowski and Rocca [15].

**The geometry of the stationary point set**

In Eqs.(5.2)-(5.3), we have

\[
\partial_{\sigma_1} T^{(N)}(r, x) = (\nabla_x T^{(N)}) \cdot (\partial_{\sigma_1} x) = (\nabla_x T^{(N)}) \cdot (\partial_{\sigma_1} (x - S)), \quad (5.7)
\]
where \( \mathbf{x} \) is given by Eq. (4.11). Note that \( \nabla_\mathbf{x} T^{(\vec{N}\vec{N})} \) is normal to the finite-offset isochron surface for fixed \((s, r)\), while \( \partial_{\sigma_1} \mathbf{x} \) and \( \partial_{\sigma_2} \mathbf{x} \) are tangent to the zero-offset isochron surface for fixed \((S, \tau)\). Hence, the condition of stationarity Eqs. (5.2)-(5.3) implies that the zero-offset isochron surface must be tangent to the finite-offset isochron surface.

The stationary geometry can be captured in a system of non-linear equations (see Artley et al. [16]). The first subsystem arises from the condition that the gradient of finite-offset travel time should be parallel to the slowness vector of the zero-offset ray. The second subsystem arises from the condition that the rays of the finite-offset configuration should connect at the stationary scattering point. Let us choose Cartesian coordinates such that the 1-direction coincides with the line connecting \( s \) with \( r \), and let \( h = \frac{1}{2} (r - s) \). Let the \((1, 2)\)-plane be defined by \((s, S, r)\); at stationarity, the points \( s, r, S, \mathbf{x} \) define a tetrahedron. Then, the first condition yields

\[
\frac{\Gamma_1^{(\vec{N}\vec{N})}}{\Gamma_3^{(\vec{N}\vec{N})}} = \frac{\gamma_1^{(N)}}{\gamma_3^{(N)}}, \quad \frac{\Gamma_2^{(\vec{N}\vec{N})}}{\Gamma_3^{(\vec{N}\vec{N})}} = \frac{\gamma_2^{(N)}}{\gamma_3^{(N)}}.
\]  

(5.8)

The second condition yields (see also Eq. (2.16))

\[
\tilde{v}_{1}^{(\vec{N})} \quad \tilde{T}^{(\vec{N})} - \tilde{v}_{1}^{(N)} \quad \tilde{T}^{(\vec{N})} = 2h,
\]

\[
\tilde{v}_{2,3}^{(\vec{N})} \quad \tilde{T}^{(\vec{N})} - \tilde{v}_{2,3}^{(N)} \quad \tilde{T}^{(\vec{N})} = 0,
\]

(5.9)

while the times are constrained by

\[
\tilde{T}^{(\vec{N})} + \tilde{T}^{(\vec{N})} = T^{(\vec{N}\vec{N})}.
\]

(5.10)

The zero-offset position and time follow from the third subsystem of equations, arising from the condition that the ray of the zero-offset configuration connects with the rays of the finite-offset configuration at the stationary scattering point:

\[
\tilde{v}_{1}^{(\vec{N})} \quad \tilde{T}^{(\vec{N})} - \frac{1}{2} \tilde{v}_{1}^{(N)} \quad \tilde{T}^{(\vec{N})} = S_1 + h,
\]

\[
\tilde{v}_{2}^{(\vec{N})} \quad \tilde{T}^{(\vec{N})} - \frac{1}{2} \tilde{v}_{2}^{(N)} \quad \tilde{T}^{(\vec{N})} = S_2,
\]

\[
\tilde{v}_{3}^{(\vec{N})} \quad \tilde{T}^{(\vec{N})} - \frac{1}{2} \tilde{v}_{3}^{(N)} \quad \tilde{T}^{(\vec{N})} = 0.
\]

(5.11)

As before, all the group velocities are taken in the background medium; the ray directions, appearing in Eqs. (5.9)-(5.11), coincide with the group velocities.
Eliminating the finite-offset times with the aid of Eq.(5.10) from equations (5.9), leads to

\[
\frac{\ddot{v}_3^{(N)}}{\ddot{v}_3^{(N)} + \ddot{v}_3^{(N)}} \ddot{v}_1^{(N)} - \frac{\ddot{v}_3^{(N)}}{\ddot{v}_3^{(N)} + \ddot{v}_3^{(N)}} \ddot{v}_1^{(N)} = \frac{2h}{T^{(N)N}},
\]

(5.12)

\[
\frac{\ddot{v}_2^{(N)}}{\ddot{v}_3^{(N)}} - \frac{\ddot{v}_2^{(N)}}{\ddot{v}_3^{(N)}} = 0.
\]

Eliminating the finite-offset times from equations (5.11), leads to

\[
S_1 = \frac{1}{2} \left( \frac{\ddot{v}_1^{(N)}}{\ddot{v}_3^{(N)}} v_3^{(N)} - v_1^{(N)} \right) \tau - h,
\]

\[
S_2 = \frac{1}{2} \left( \frac{\ddot{v}_2^{(N)}}{\ddot{v}_3^{(N)}} v_3^{(N)} - v_2^{(N)} \right) \tau,
\]

(5.13)

\[
\tau = \frac{2 \ddot{\gamma}_3^{(N)} \ddot{\gamma}_3^{(N)}}{v_3^{(N)} (\ddot{v}_3^{(N)} + \ddot{v}_3^{(N)})} T^{(N)N}.
\]

In Eqs.(5.8)-(5.13) we have an implicit relationship between group velocity and slowness vector in the background medium. Convenient variables are the two horizontal components of the three slowness vectors: \( \gamma_{1,2}^{(N)}, \gamma_{1,2}^{(N)}, \gamma_{1,2}^{(N)} \). The components \( \gamma_{1,2}^{(N)} \) parameterize the stationary point set; \( h \) and \( T^{(N)N} \) are given. Then, the four equations (5.8), (5.12) constitute a system with four unknowns, that has to be solved numerically. Equations (5.13) then give the zero-offset-point/zero-offset-time pair, \((S, \tau)\), at stationarity.

The Hessian can be written as

\[
\partial_{\sigma_1} \partial_{\sigma_1} T^{(N)N} = \nabla_x \nabla_x T^{(N)N} : (\partial_{\sigma_1} \nabla_x) (\partial_{\sigma_1} \nabla_x) + \nabla_x T^{(N)N} \cdot \partial_{\sigma_1} \partial_{\sigma_1} \nabla_x
\]

\[
= \left( \frac{1}{2} \tau \right)^2 |\nabla_x T^{(N)N}| \left( \frac{2}{\tau} \nabla \cdot \partial_{\sigma_1} \partial_{\sigma_1} \nabla_x + (\partial_{\sigma_1} \nabla_x) \cdot (\partial_{\sigma_1} \nabla_x) \frac{\nabla_x \nabla_x T^{(N)N}}{|\nabla_x T^{(N)N}|} \right),
\]

etc. Here, we have separated the Hessian into the obliquity factor, the Hessian attached to the zero-offset isochron surface, and the Hessian attached to the finite-offset isochron surface. Note that the obliquity factor, \( |\nabla_x T^{(N)N}| \), can be taken together with \( \mathcal{J}^{(N)N} \) and that the scaling \( \left( \frac{1}{2} \tau \right)^2 \) can be taken together with \( g \) (Eq.(5.4)).

It is noted, that given the zero-offset ray, only a restricted range of \((s, \tau)\) will allow for a stationary geometry to exist. We have indicated this restriction by replacing \( \partial S \times \partial R \) with \( \partial S \times \partial R' \) in Eq.(5.4). In Appendix A, we give the expressions for an isotropic medium in two dimensions, which establishes the correspondence to the conventional DMO procedure.
Kirchhoff-style Dip MoveOut

In this subsection, we review the concept of Dip MoveOut. For each point on a fixed finite-offset isochron surface (fixed \((s, r)\) and travel time \(T^N(r, s, t) = t\)) we construct the combination \((S, \tau)\) such that the zero-offset isochron surface is tangent to or touches the finite-offset one (cf. Eq. (5.4))

\[
\left( (r, s), T^N(r, x^0, s) = t \right) \rightarrow (S, \tau) \quad \text{for variable} \quad x^0 .
\]  

(5.15)

As \(x^0\) varies along the finite-offset isochron surface, the stationary dip \(\nu(r, x^0, s)\) varies according to the conditions of tangency. (In the previous subsection, we employed the horizontal components of the zero-offset slowness vector at the stationary point for this role.) Hence, the latter equation can be viewed as a mapping from dip to a zero-offset-point/zero-offset-time pair. Often, in this mapping, the midpoint \(y = \frac{1}{2}(r + s)\) is chosen to be the origin in \(S\)-space. It is also clear that each finite-offset time \(t\) maps on a set of zero-offset times \(\tau\). Mapping (5.15) illustrates that a particular sample of the original field contributes to the zero-offset field at various space-time combinations. In fact, in the implementation of transformation (5.4) one would loop over the finite-offset isochrons labelled by \((r, s, t)\), and accumulate the contributions from any finite-offset isochron to the zero-offset field at the associated values of \((S, \tau)\).

Observe now, that inversion formula (4.2) produces an image for each \((\theta, \psi)\) pair by integrating over the dip. This fact is commonly exploited in DMO, viz., by integrating the integrand of Eq. (5.4) over the midpoints \(y\) only, keeping the half-offset \(h = \frac{1}{2}(r - s)\) fixed. The result is a family of intermediate zero-offset fields \(u_P^{(1)}(S, \tau; h)\), the stack of which is the true zero-offset field,

\[
u_{PQ}^{(1)}(S, \tau) = \int_{\partial H} u_{PQ}^{(1)}(S, \tau; h) \, dh .
\]  

(5.16)

Here,

\[
u_{PQ}^{(1)}(S, \tau; h) = -\frac{1}{2\pi} \int_{\partial\Gamma(h)} \int_{t \in \mathbb{R}_{\geq 0}} \partial_r S_{PQpq}(S, \tau; y, t; h) \, U_{pq}^{(1)}(y + h, y - h, t) \, dt \, dy ,
\]  

(5.17)

where the smear kernel \(S_{PQpq}\) (Deregowski [15]) is given by (cf. Eq. (5.4))

\[
S_{PQpq}(S, \tau; y(r, s), t; h(r, s)) = \xi_p^{(N)}(S) \xi_q^{(N)}(S) \xi_p^{(N)}(r) \xi_q^{(N)}(s) \times
\]

\[
\left( \omega^{(N)}(\cdot, \alpha^{(N)}(\cdot)) \right)^T \left[ \Lambda^{(N)}(\nu(r, s)) \right]^{-1} \omega^{(N)}(\cdot, \alpha^{(N)}(\cdot)) \times
\]

\[
\delta^1(t - T^{(N)}(r, s)) \frac{\nabla^{(N)}(\cdot)}{\sqrt{\det(\nabla_{\sigma_-} \nabla_{\sigma_-} T^{(N)}(\cdot))}} \times
\]
\[ g \mathcal{T}^{(N)}(\ldots, S) \mathcal{J}^{(N)}(\ldots, s) \frac{\partial (\hat{\alpha}, \hat{\alpha})}{\partial (s, r)} \bigg|_{x(\sigma_0^0, \tau)} = \] (5.18)

using that

\[ \text{d} s \text{d} r = 4 \text{d} y \text{d} h. \]

Having solved for the slowness vectors at the stationary point, the remaining quantities are readily obtained.

The DMO impulse response is the zero-offset field obtained by substituting in Eq.(5.17) the non-physical field (unit data sample)

\[ U_{pq}^{(1)}(r, s, t) = \delta \left( r - [y^i + h^i] \right) \delta \left( s - [y^i - h^i] \right) \delta \left( t - t^i \right), \]

and is given by

\[ -\frac{1}{2\pi} \frac{\partial^2 S_{PQpq}}{\partial r^2} (S, \tau; y^i, t^i; h). \] (5.19)

The singular support of this function can be multi-valued. Equation (5.17) represents an integration of DMO impulse responses over midpoints and offset times.

In practical acquisition geometries, we have a preferred orientation in the configuration: \( s \) and \( r \) are constrained to a horizontal plane. This implies that there is a preferred migration dip, viz. the one parallel to vertical. The ‘Normal MoveOut’ now follows from the mapping \( t \rightarrow \tau \) in (5.15) at \( x^0 = x_n \) where the stationary dip \( \nu(r, x_n, s) = \nu_n \) coincides with the vertical direction (the isochron tangent plane is horizontal); we write

\[ ((r, s), t) \rightarrow (S_n, \tau_n) \text{ for any } y \text{ (fixing } x^0 = x_n) \]

and find \( t = t(\tau_n, h) \). In Eq.(5.17) we can now replace the integration variable \( t \) by \( \tau_n \), and set

\[ U_{n,pq}^{(1)}(y, h, \tau_n) \equiv U_{pq}^{(1)}(y + h, y - h, t(\tau_n, h)) \frac{\partial t}{\partial \tau_n}. \]

The field \( U_{n,pq}^{(1)} \) can be viewed as the NMO corrected finite-offset data \( U_{pq}^{(1)} \). The mapping Eq.(5.15) can be decomposed into a mapping based at \( x_n \) or \( \nu_n \) followed by a mapping from \( x_n \) or \( \nu_n \) to \( x^0 \) or \( \nu(r, x^0, s \). The second, correcting map,

\[ (S_n, \tau_n) \rightarrow (S, \tau) \]

represents the geometry underlying the DMO procedure.
6 Multi-mode scattering

To account for mode-converted waves, we replace $\hat{N}$ by $\hat{M}$. In a homogeneous background medium, for the zero-offset configuration, the incident and scattered ray directions would still coincide. However, the slowness vectors at the scattering point are no longer the same and may point in different directions. The zero-offset isochron surface,

$$\{x \in \mathbb{R}^3 \mid T^{(N,M)}(x, S) = \tau\},$$

is now given by

$$x - S = \left[1 + \frac{|\mathbf{v}^{(M)}|}{|\mathbf{v}^{(N)}|}\right]^{-1} \mathbf{v}^{(N)} \tau. \tag{6.1}$$

It is realistic that this surface may become multi-valued. We will not deal with that case here. If the zero-offset isochron surface is single-valued, the analysis of the previous sections applies, with the modification that

$$\nabla x T^{(N)} = \gamma^{(N)} + \gamma^{(M)} \gamma^{(N)},$$

in general.

7 Discussion

We have derived a DMO/TZO procedure that corrects for radiation characteristics and is valid in anisotropic, elastic media. We have used generalized Radon transforms, focussing on the Kirchhoff-style representation of the procedure.

Fundamentally, the TZO contains four mappings, one for the geometrical spreading, one for the polarizations, one for the radiation patterns, and one for the coordinates. The mapping of radiation patterns establishes a mapping between reflection coefficients from the finite-offset to the zero-offset configuration. The formalism is complete within the ray-Born scattering theory. We argue that the TZO should be correct at least within this approximation, but this does not assure that the procedure is valid beyond. Perhaps the various characteristics of the procedure (i.e., spreading, polarizations, radiation patterns and coordinates) have different sensitivities to the background model. It is still to be analyzed how large the contributions from the constituent mappings are going to be in a practical setting.

There are various ways to address numerical implementation of the TZO process. The Kirchhoff-style representation lends itself for a quasi-Monte Carlo discretisation [17]. Within certain limits, analytic solutions for the stationary-point set can be found, which significantly speeds up the computations.
Appendix A. The stationary phase analysis for an isotropic medium in the two dimensions

In an isotropic medium the group and phase directions coincide. Let the medium's velocity be $c^{(N)}$. Then

$$\alpha = \text{polar angle} \, , \quad v_1^{(N)} = c^{(N)} \cos \alpha \, , \quad v_3^{(N)} = c^{(N)} \sin \alpha \, ,$$

while

$$g = \frac{1}{2} c^{(N)} \tau^{(N)} \, .$$

Note that in a two-dimensional configuration $U^{(1)}$ is the Hilbert transformed scattered displacement field.

The time convolution

The time function with which the data has to be convolved (cf. Eq.(5.4)), in two dimensions, is given by

$$\frac{1}{\pi} \Re \int_{\mathbb{R}^2} \sqrt{\pi} \, \frac{1}{\omega} \exp \left[ - \frac{i\pi}{4} \right] \exp(i\omega t) \, d\omega = \frac{H(t)}{\sqrt{t}} \, . \quad (A.1)$$

The Hessian

In the Hessian, Eq.(5.14), we have the following substitutions. In an isotropic medium, we have

$$|\nabla_x T^{(N\bar{N})}| = \frac{2}{c^{(N)}} \cos \left( \frac{\theta}{2} \right) \, . \quad (A.2)$$

Further, note that

$$\partial_{\sigma_i} v^{(N)} = \partial_{\sigma_i} v^{(N)} = -v^{(N)} \, .$$

Since the phase and group directions coincide, at stationary, $\nu \parallel v^{(N)}$. Hence,

$$\nu \cdot \partial_{\sigma_i} \partial_{\sigma_j} v^{(N)} = -c^{(N)} \, . \quad (A.3)$$

The unit tangent vector to the isochrons at the stationary point is given by

$$c^{(N)} \partial_{\sigma_i} v^{(N)} = (-\sin \alpha, \cos \alpha) \, .$$

Hence,

$$[c^{(N)}]^{-1} (\partial_{\sigma_i} v^{(N)}) \cdot (\partial_{\sigma_j} v^{(N)}) = \frac{\frac{1}{2} \nabla_x \nabla_x T^{(N\bar{N})}}{|\nabla_x T^{(N\bar{N})}|} = \frac{1}{\tau^{(N\bar{N})}} \, , \quad (A.4)$$

where $\tau^{(N\bar{N})}$ is twice the radius of curvature of the finite-offset isochron at stationary. We obtain

$$\partial_{\sigma_i} \partial_{\sigma_j} T^{(N\bar{N})} = [\tau^{(N\bar{N})}]^2 \cos \left( \frac{\theta}{2} \right) \left[ \frac{1}{\tau^{(N)}} - \frac{1}{\tau^{(N\bar{N})}} \right] \, . \quad (A.5)$$
References


Amplitude preservation for offset continuation: Confirmation for Kirchhoff data

Sergey Fomel
Stanford Exploration Project, Stanford University

Norman Bleistein
Center for Wave Phenomena, Colorado School of Mines

ABSTRACT
Offset continuation (OC) is the operator that transforms common-offset seismic reflection data from one offset to another. In earlier papers by the first author, a partial differential equation in midpoint and offset was presented to achieve this transformation. The equation was derived from the kinematics of the continuation process. This is equivalent to proposing the wave equation from knowledge of the eikonal equation. While such a method will produce a PDE with the same traveltimes, it does not guarantee that the amplitude will be correctly propagated by the resulting second order partial differential equation. The second author (with J. K. Cohen) proposed a DMO operator for which a verification of amplitude preservation was proven for Kirchhoff data in the two-and-one-half dimensional case. It was observed by the first author that the solution of the OC partial differential equation produced the same DMO solution when specialized to continue data to zero offset. In a synthesis of these two approaches, we present here a proof that the solution of the OC partial differential equation does propagate amplitude properly at all offsets, at least to the same order of accuracy. That is, it provides a solution with the correct traveltime and correct leading order amplitude. “Correct amplitude” in this case means that the transformed amplitude exhibits the right geometrical spreading and reflection-surface-curvature effects for the new offset. The reflection coefficient of the original offset is preserved in this transformation. This result is more general than the earlier results in that it does not rely on the two-and-one-half dimensional assumption.

Key words: offset continuation (OC), OC partial differential equation, dip moveout (DMO), Kirchhoff approximation, eikonal equation, transport equation.

Introduction

Offset continuation (OC) is the operator that transforms common-offset seismic reflection data from one offset to another. Following the classic results of Deregowski and Rocca (1981), Bolondi et al. (1982; 1984) described OC as a continuous process of the gradual change of the offset by means of a partial differential equation. Being based on the small-offset small-dip approximation, Bolondi’s equation failed at large offsets or steep reflector dips. Nevertheless, the OC concept inspired a whole flow of research on dip moveout (DMO) correction (Hale, 1991). Since one can view DMO as a particular case of OC (continuation to zero offset), the offset continuation theory can serve as a natural basis for the DMO theory. Its immediate application is in interpolating data undersampled in the offset dimension.

Fomel (1994),(1995a) recently introduced a revised version of the OC differential equation and proved that it provides the correct kinematics of the continued wavefield for any offset and reflector dip under the assumption of constant effective propagation speed. Studying the laws of amplitude transformation shows that in 2.5-D media the amplitudes of the continued seismic gathers transform according to the rules of geometric seismics, except for the reflection coefficient, which remains
unchanged (Fomel, 1993a; Goldin & Fomel, 1995). The solution of the boundary problem on the OC equation for the DMO case (Fomel, 1993b) coincides in the high-frequency asymptotics limit with the amplitude-preserving DMO, also known as Born DMO (Liner, 1991; Bleistein, 1990). However, for the purposes of verifying that the amplitude is correct for any offset, this derivation is incomplete.

In this paper, we perform a direct test on the amplitude properties of the OC equation. We describe the input common-offset data by the Kirchhoff modeling integral, which represents the high-frequency approximation of a reflected (scattered) wavefield, recorded at the surface at non-zero offset (Bleistein, 1984). For reflected waves, the Kirchhoff approximation is accurate up to two orders in the high-frequency series (the ray series) for the differential operator applied to the solution, with the first order describing the phase function alone, and the second order describing the amplitude. We prove that both orders of accuracy are satisfied when the offset continuation equation is applied to Kirchhoff data. Thus, this differential equation is the “right” equation to two orders, producing the correct amplitude as well as the correct phase for offset continuation. That is, the geometrical spreading effects and curvature effects on the reflected data are properly transformed. The angularly dependent reflection coefficient of the original offset is preserved.

This proof relates the OC equation with “wave-equation” processing. It also provides additional confirmation of the fact that the true-amplitude OC and DMO operators do not depend on the reflector curvature and can properly handle reflections from arbitrarily shaped reflectors (Bleistein & Cohen, 1993). The latter result was specifically a 2.5D result, whereas the result here does not depend on the 2.5D assumption. That is, the result presented here remains valid when the reflector has out-of-plane variation.

Our method of proof is indirect. We first write the Kirchhoff representation for the reflected wave in a form that can be easily matched to the solution of the OC differential equation. We then present the analogs of the eikonal and transport equations for the OC equation and show that the amplitude and phase of the Kirchhoff representation satisfy those two equations.

The Kirchhoff Modeling Approximation

We introduce here the Kirchhoff approximate integral representation of the upward propagating response to a single reflector, with separated source and receiver point. We then show how the amplitude of this integrand is related to the zero-offset amplitude at the source receiver point on the ray making equal angles at the scattering point with the rays from the separated source and receiver. The Kirchhoff integral representation (Haddon & Suchen, 1981; Bleistein, 1984) describes the wavefield scattered from a single reflector. It is applicable in the situations where the high-frequency assumption is valid (the wavelength is smaller than the characteristic dimensions of the model) and corresponds in accuracy to the WKBJ approximation for reflected waves, including phase shifts through buried foci. The general form of the Kirchhoff modeling integral is

\[
U_S(r, s, \omega) = \int_\Sigma R(x; r, s) \frac{\partial}{\partial n} [U_I(s, x, \omega) G(x, r, \omega)] \, d\Sigma,
\]

where \(s = (s, 0, 0)\) and \(r = (r, 0, 0)\) stand for the source and the receiver location vectors at the surface of observation, \(x\) denotes a point on the reflector surface \(\Sigma\); \(R\) is the reflection coefficient at \(\Sigma\); \(n\) is the upward normal to the reflector at the point \(x\); \(U_I\) and \(G\) are the incident wavefield and Green's function, respectively, represented by their WKBJ approximation.

\[
U_I(s, x, \omega) = F(\omega) A_s(s, x) e^{i\omega \tau_s(s, x)}, \quad (2)
\]

\[
G(x, r, \omega) = A_r(x, r) e^{i\omega \tau_r(x, r)}. \quad (3)
\]

In this equation, \(\tau_s(s, x)\) and \(A_s(s, x)\) are the traveltime and the amplitude of the wave propagating from \(s\) to \(x\), \(\tau_r(x, r)\) and \(A_r(x, r)\) are the corresponding quantities for the wave propagating from \(x\) to \(r\), and \(F(\omega)\) is the spectrum of the input signal, assumed to be the transform of a bandlimited impulsive source. In the time domain, the Kirchhoff modeling integral transforms to

\[
U_S(r, s, t) = \int_\Sigma R(x; r, s) \, dx \quad (4)
\]

\[
\frac{\partial}{\partial s} [A_s(s, x) A_r(x, r) f(t - \tau_s(s, x) - \tau_r(x, r))],
\]

where \(f\) is the inverse temporal transform of \(F\). The reflection traveltime \(\tau_r\) corresponds physically to the diffraction from a point diffractor, located at the point \(x\) on the surface \(\Sigma\) and the amplitudes, \(A_s\) and \(A_r\), are point diffractor amplitudes, as well.

The main goal of this paper is to test the compliance of the representation (3) with the offset continuation differential equation. The OC equation contains the derivatives of the wavefield with respect to the parameters of observation \((s, r,\) and \(t)\). According to the rules of classic calculus, these derivatives can be taken under the integral sign in (3). Furthermore, since we do not assume the true-amplitude OC operator to affect the reflection coefficient \(R\), we can take the offset-dependence of this coefficient outside of the scope of consideration. Therefore, the only term to be considered as a trial solution to
Figure 1. Geometry of diffraction in a constant velocity medium: View in the reflection plane.

The OC equation is the kernel of the Kirchhoff integral, which is contained in the square brackets in (2) and (5) and has the form

$$ k(s, r, x, t) = A_s(s, r, x) f(t - \tau_s(s, r, x)) $$

where

$$ \tau_s(s, r, x) = \tau_a(s, x) + \tau_r(x, r) $$

$$ A_a(s, r, x) = A_a(s, x) A_r(x, r) $$

In a 3-D medium with a constant velocity $v$, the traveltimes and the amplitudes have the simple explicit expressions:

$$ \tau_a(s, x) = \frac{\rho_a(s, x)}{v} $$

$$ A_a(s, x) = \frac{1}{4\pi \rho_a(s, x)} $$

where $\rho_a$ and $\rho_r$ are the lengths of the incident and reflected rays, respectively (Figure 1).

We introduce a particular zero-offset amplitude, namely the amplitude along the zero offset ray that bisects the angle between the incident and reflected ray in this plane. See Figure 1. We denote the square of this amplitude by $A_0$. That is,

$$ A_0 = \frac{1}{(4\pi \rho_0)^2} $$

As follows from formulas (7) and (9), the amplitude transformation in DMO (continuation to zero offset) is characterized by the dimensionless ratio

$$ A_{sr} = \frac{\rho_b^2}{\rho_a \rho_r} $$

where $\rho_b$ is the length of the zero-offset ray (Figure 1).

As follows from the simple trigonometry of the triangles, formed by the incident and reflected rays (the law of cosines),

$$ \sqrt{\rho_a^2 + \rho_b^2 - 2 \rho_a \rho_b \cos \gamma} $$

$$ = \sqrt{\rho_a^2 + \rho_b^2 - 2 \rho_a \rho_b \cos 2\gamma} $$

where $\gamma$ is the reflection angle as shown in the figure.

After straightforward algebraic transformations of equation (13), we arrive to the explicit relationship between the ray lengths:

$$ \frac{(\rho_a + \rho_r) \rho_0}{2 \rho_a \rho_r} = \cos \gamma $$

Substituting (13) to (11) yields

$$ A_{sr} \frac{\tau_0}{A_0} = \frac{\tau_0}{\tau_{sr}} \cos \gamma $$

where $\tau_0$ is the zero-offset two-way traveltime ($\tau_0 = 2 \rho_0/v$).

What we have done here is rewrite the finite offset amplitude in the Kirchhoff integral in terms of a particular zero offset amplitude. That zero offset amplitude would arise as the geometrical spreading effect if there were a reflector whose dip was such that the finite offset pair would be specular at this scattering point. Of course, the zero offset ray would also be specular in this case.

The Offset Continuation Equation

In this section we introduce the offset continuation partial differential equation. We then develop its WKBJ or ray theoretic solution for phase and leading order amplitude. We explain how we verify that the traveltime and amplitude of the integrand of the Kirchhoff representation (5) satisfy the "eikonal" and "transport" equations of the OC partial differential equation. Here, we will make use of the relationship (14) derived from the Kirchhoff integral.

The offset continuation differential equation derived in earlier papers (Fomel, 1994; Fomel, 1995a) is

$$ h \left( \frac{\partial^2 P}{\partial y^2} \frac{\partial^2 P}{\partial h^2} \right) = t_n \frac{\partial^2 P}{\partial t_n \partial h} $$

In this equation, $h$ is the half-offset ($h = l/2$), $y$ is the midpoint ($y = (s + r)/2$) and $t_n$ is the NMO-corrected traveltime:

$$ t_n = \sqrt{t^2 - \frac{r^2}{v^2}} $$

Equation (15) describes the process of seismogram transformation in the time-midpoint-offset domain. One can obtain the high-frequency asymptotics of its solution by standard methods, as follows. We introduce a trial asymptotic solution of the form

$$ P(y, h, t_n) = A_n(y, h) f(t_n - \tau_n(y, h)) $$
We remind the reader that we have assumed that \( f \) is a “rapidly varying function”—think of a bandlimited delta function, for example. We substitute this solution into equation (15) and collect terms in order of derivatives of \( f \). This is the direct counterpart of collecting terms in powers of frequency when applying WKBJ in the frequency domain. From the leading asymptotic order (the second derivative of the function \( f \)), we obtain the “eikonal” equation describing the kinematics of the OC transformation:

\[
h \left( \left( \frac{\partial \tau_n}{\partial y} \right)^2 - \left( \frac{\partial \tau_n}{\partial h} \right)^2 \right) = -\tau_n \frac{\partial \tau_n}{\partial h}.
\]  

(18)

In this equation, we have replaced a multiplier of \( t_n \) by \( \tau_n \) on the right side of the equation. This is consistent with our assumption that \( f \) is a bandlimited delta function or some equivalent impulsive source. Analogously, collecting the terms containing the first derivative of \( f \) leads to the transport equation describing the transformation of the amplitudes:

\[
(\tau_n - 2h \frac{\partial \tau_n}{\partial h}) \frac{\partial A_n}{\partial h} + 2h \frac{\partial \tau_n}{\partial y} \frac{\partial A_n}{\partial y} + h A_n \left( \frac{\partial^2 \tau_n}{\partial y^2} - \frac{\partial^2 \tau_n}{\partial h^2} \right) = 0.
\]  

(19)

Rewriting the eikonal equation (18) in the time-source-receiver coordinate system:

\[
\left( \tau_{sr}^2 + \frac{\tau_{sr}}{c^2} \right) \left( \frac{\partial \tau_{sr}}{\partial r} - \frac{\partial \tau_{sr}}{\partial s} \right) = 2 l \tau_{sr} \left( \frac{\partial \tau_{sr}}{\partial r} \right) - \tau_{sr} \left( \frac{\partial \tau_{sr}}{\partial s} \right),
\]  

(20)

it is easy (for Mathematica) to verify that the explicit expression for the phase of the Kirchhoff integral kernel (6) satisfies it for any scattering point \( \mathbf{x} = (x_1, x_2, z) \). Here, \( \tau_{sr} \) is related to \( \tau_n \) as \( t \) is related to \( t_n \) in (16).

The general solution of the amplitude equation (20) has the form (Fomel, 1995a)

\[
A_n = A_0 \frac{\tau_0 \cos \gamma}{\tau_n} \left( \frac{1 + \rho_0 K}{\cos^2 \gamma - \rho_0 K} \right)^{1/2},
\]  

(21)

where \( K \) is the reflector curvature at the reflection point. Since the kernel (3) of the Kirchhoff integral (5) corresponds kinematically to the reflection from a point diffractor, we can obtain the solution of the amplitude equation for this case by formally setting the curvature \( K \) to infinity (setting the radius of curvature to zero). This leads to the relationship

\[
\frac{A_n}{A_0} = \frac{\tau_0}{\tau_n} \cos \gamma.
\]  

(22)

Again, we exploit the assumption that the signal \( f \) has the form of the delta-function. In this case, the amplitudes before and after the NMO correction are connected according to the known properties of the delta function, as follows:

\[
A_{sr} \delta (t - \tau_n(s, r, x)) = \left. \frac{\partial t_n}{\partial t} \right|_{t = \tau_n} A_n \delta (t_n - \tau_n(s, r, x))
\]  

(23)

\[
= A_n \delta (t_n - \tau_n(s, r, x)),
\]  

with

\[
A_n = \frac{\tau_{sr}}{\tau_n} A_{sr}.
\]  

(24)

Combining (24) with (22) leads to the equation

\[
\frac{A_{sr}}{A_0} = \frac{\tau_0}{\tau_{sr}} \cos \gamma,
\]  

(25)

exactly coincident with the previously found formula (14). As with the solution of the eikonal equation, we pass from an in-plane solution in two dimensions to a solution for a scattering point in three dimensions by replacing \( z^2 \) by \( z^2 + z_r^2 \).

Now, with the amplitude and phase of the ray theoretic solution of the OC differential equation (15) matching the amplitude and phase of the Kirchhoff representation, they are seen to be the same. Thus, the amplitude and phase of the Kirchhoff representation for arbitrary offset is seen to be the point diffractor WKBJ solution of the offset continuation differential equation. Hence, the Kirchhoff approximation is a solution of the OC differential equation when we hold the reflection coefficient constant. This means that the solution of the OC differential equation has all of the features of amplitude preservation as does the Kirchhoff representation, including geometrical spreading, curvature effects and phase shift effects. Furthermore, in the Kirchhoff representation and the solution of the OC partial differential equation by WKBJ, we have not used the 2.5D assumption. Hence, the preservation of amplitude is not restricted to cylindrical surfaces as was the true amplitude proof for DMO in (Bleistein & Cohen, 1995). This is what we sought to confirm.

**Discussion**

We have proved that the offset continuation equation correctly transforms common-offset seismic data modeled
by the Kirchhoff integral approximation. The kinematic and dynamic equivalence of the OC equation has been proved previously by different methods (Fomel, 1995a; Fomel, 1995b). However, connecting this equation with Kirchhoff modeling opens new insights into the theoretical basis of DMO and offset continuation:

(i) The Kirchhoff integral can serve as a link between the wave-equation theory, conventionally used in seismic data processing and the kinematically derived OC equation. Though the analysis in this paper follows the constant-velocity model, this link can be extended in principle to handle the case of a variable background velocity.

(ii) The OC equation operates on the kernel of the Kirchhoff integral, which is independent on the local dip and the curvature of the reflector. This proves that the true-amplitude OC and DMO operators can properly handle curved reflectors. Moreover, this result does not imply any special orientation of the reflector curvature matrix. Therefore, it does not require a commonly made 2.5-D assumption (Bleistein & Cohen, 1995; Fomel, 1995a). Implicitly, this proves the amplitude preservation property of the three-dimensional azimuth moveout (AMO) operator (Biondi & Chilinguirian, 1994; Fomel & Biondi, 1995), based on cascading the true-amplitude DMO and inverse DMO operators.

Acknowledgements

The first author gratefully acknowledges the support and encouragement of Fabio Rocca on the analysis of the offset continuation partial differential equation.

References


SUDREF: A distributed elastic reflectivity modeling algorithm

Wences Gouveia
Department of Geophysics
Center for Wave Phenomena
Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
From earthquake seismology to mineral exploration, accurate seismic modeling is one of the most important tools for studying the Earth's subsurface. Here, I present a distributed implementation of the elastic reflectivity method. In this method, the generation of the synthetic wave field is done by solving a sequence of boundary value problems. Since the computations are performed in the frequency domain, this approach presents a high degree of parallelism.

The distributed implementation (sudref) consists of a master-slave topology, in which all the required communication and synchronization procedures are provided by the PVM library. This algorithm is designed to take advantage of a heterogeneous computer network (i.e., with machines of different power or with machines with unequal processor loads), by partitioning the total work into tasks, the size of which is specified by the user.

Sudref has been benchmarked on an Ethernet connected network of 11 Pentium-based Linux workstations, a 4 processor SGI Power Challenge and a 16 processor IBM SP2. The distributed implementation shows good performance results. The elastic wavefield for a 400 layer model can be computed with an elapsed time of less than 1 hour on a network of eleven PC's, a speedup of a factor of 8.2 over the sequential time.

Key words: Seismic modeling, Distributed computing, Computational performance

Introduction
Due to the relevance of synthetic data generation in studies of the Earth's subsurface, many algorithms have been proposed, developed and tested for modeling waves in heterogeneous elastic media. Those algorithms differ in the complexity of the models they can handle, and also in the extent that the wavefield is modeled (for example, just the primary events as opposed to all models and multiples). Finite difference and finite element modeling techniques, are probably the method of choice if one is interested in synthesizing the complete wavefield in a fully heterogeneous medium. Geometric optics allows rapid computation of the wavefield, and also has the ability to generating single modes of propagation (for example, mode converted S-waves). It is however limited to smooth media, or media with large features compared to the dominant wavelength. In the situation where the medium presents some kind of symmetry, for example the elastic properties do not change laterally, analytic solutions of the wave equation for the complete wavefield, such as reflectivity (Fuchs and Müller, 1971, Kennet, 1983) may be used. This procedure casts the seismic modeling algorithm as the solution of a sequence of boundary value problems, in which the continuity of displacements and tractions across internal interfaces are used to propagate the wavefield from one layer to the next.

Here, I present a distributed implementation of an elastic isotropic reflectivity method called sudref. Since in this technique the computations are performed in the offset-temporal frequency domain, and the frequency components are independent of each other, it presents
a high degree of parallelism. Aiming at preserving full portability to different computer systems I developed the code in standard ANSI C and used the PVM library (Geist et al., 1994) for all required synchronization and message passing routines. That allowed me to benchmark this code in the following computer architectures:

- 4 processor SGI Power Challenge
- 1 Mbyte/s Ethernet connected network of 11 Pentium-based Linux workstations
- 16 processor IBM SP2

This work is structured as follows. First, I briefly describe the reflectivity method and illustrate how the algorithm works with an example. Next, I present the distributed implementation of this procedure, and discuss the simple load balancing approach used to make an efficient use of a heterogeneous network. Finally, I show benchmark results in which I measure the speedup of the implementation when several processors are used in the computation.

The Reflectivity Method

In this section I will introduce the reflectivity method in a schematic way, keeping the usage of mathematical expressions to a minimum. A detailed description of this procedure can be found in the excellent tutorial by Müller (1985). The geometry of the problem in cylindrical coordinates is shown in Figure 1. The elastic isotropic layered medium is completely characterized by the P-wave (\(c_p\)) and S-wave (\(c_s\)) velocities, the density (\(\rho\)) and the thickness (\(dz\)) for each layer. Consider for a moment a homogeneous medium above and below the source depth \(z_s\). In this case, the upgoing (\(z < z_s\)) and downgoing (\(z > z_s\)) displacement potentials due to a point force located at depth \(z_s\) with orientation \([F_1, F_2, F_3]\) in Cartesian coordinates are given by (Müller, 1985)

For \(z > z_s\):

\[
\Phi^u_s = \frac{1}{4\pi \rho_m} \int_0^\infty (e \cdot A_1; J_0(\omega r) + e_2 A_{s2} J_1(\omega r)) \exp[-i\omega a_m(z - z_s)] du
\]

\[
\Psi^u_s = \frac{1}{4\pi \rho_m} \int_0^\infty \frac{1}{j \omega} (e \cdot C_1; J_0(\omega r) + e_2 C_{s2} J_1(\omega r)) \exp[-i\omega b_m(z - z_s)] du
\]

\[
\chi^u_s = \frac{1}{4\pi \rho_m} \int_0^\infty \eta E_x J_1(\omega r) \exp[-i\omega b_m(z - z_s)] du
\]

\[
A_{s1} = u_s, \quad A_{s2} = \frac{u^2}{ia_m}, \quad C_{s1} = \frac{u^2}{b_m}, \quad C_{s2} = \frac{i}{u}
\]

\[
F_s = \frac{i}{\beta_s b_m}
\]

For \(z < z_s\):

\[
\Phi^u_s = \frac{1}{4\pi \rho_m} \int_0^\infty (e \cdot B_1; J_0(\omega r) + e_2 B_{s2} J_1(\omega r)) \exp[-i\omega a_m(z - z_s)] du
\]

\[
\Psi^u_s = \frac{1}{4\pi \rho_m} \int_0^\infty \frac{1}{j \omega} (e \cdot D_1; J_0(\omega r) + e_2 D_{s2} J_1(\omega r)) \exp[-i\omega b_m(z - z_s)] du
\]

\[
\chi^u_s = \frac{1}{4\pi \rho_m} \int_0^\infty \frac{1}{j \omega} (e \cdot F_1; J_0(\omega r) + e_2 F_{s2} J_1(\omega r)) \exp[-i\omega b_m(z - z_s)] du
\]

\[
B_{s1} = -u_s, \quad B_{s2} = \frac{u^2}{ia_m}, \quad D_{s1} = \frac{u^2}{b_m}, \quad D_{s2} = \frac{i}{u}
\]

\[
F_s = \frac{i}{\beta_s b_m}
\]

Here, \(\Phi\), \(\Psi\) and \(\chi\) are the scalar potentials in the frequency (\(\omega\)) domain for P-waves, SV-waves and SH-waves, respectively. \(u\) is the horizontal slowness, \(a_m\) and \(b_m\) are the vertical slownesses of layer \(m\), \(e_1 = F_3\), \(e_2 = F_1\cos \phi + F_3 \sin \phi\) and \(\eta = -F_1 \sin \phi + F_3 \cos \phi\), where \(\phi\) is the azimuthal angle. \(J_0()\) and \(J_1()\) are the Bessel functions of the first kind of zero-th and first-order, respectively. For simplicity, consider only the P-SV propagation modes. It is convenient to define the following source amplitude vectors

\[
S_{1.2} = \begin{bmatrix} A_{s1.2} \\ C_{s1.2} \end{bmatrix}, \quad S_{2.2} = \begin{bmatrix} B_{s1.2} \\ D_{s1.2} \end{bmatrix}
\]

The synthesis of the wavefield at the surface requires two steps. First the upgoing wave \(V_{1.2}\) at \(z = z_m\) should be computed. It is not difficult to show (Müller, 1985) that \(V_{1.2}\) are given by

\[
V_{1.2} = [I - R^- R^+]^{-1} (S^u_{1.2} + R^- S^v_{1.2})
\]

Here, \(R^-\) and \(R^+\) are the reflectivity matrices for the lower and upper stack of layers, respectively. Such matrices are derived recursively from considerations on only the continuity of the displacement and the traction fields across the interfaces. By the same approach it is possible to compute the transmissivity \(T^+\) of the upper
stack of layers and compute the potential at the surface, which is given by

$$ V_{1,2}^0 = T^* V_{1,2}. $$

The far-field displacement field at the surface is directly derived from the potentials (Aki and Richards, 1980),

$$ \begin{bmatrix} u_r \\ u_v \end{bmatrix} (r, \omega) = \omega^2 \sum_{i=1}^{N} c_i J_i \int_0^\infty J_i u V_{0} \, du. $$

(6)

Here, $u_r$ and $u_v$ represent the radial and vertical components of the displacement field, and the following definitions have been used:

$$ U = \begin{bmatrix} u & b_0 \\ a_0 & -u \end{bmatrix}, \quad J_i = \begin{bmatrix} -J_i(u \omega r) & 0 \\ 0 & iJ_i(u \omega r) \end{bmatrix}, $$

$$ J_i = \begin{bmatrix} J_0(u \omega r) & 0 \\ 0 & iJ_1(u \omega r) \end{bmatrix}. $$

(7)

Equation (6) is used to compute the $P-SV$ displacement field in the frequency domain. Of course, each frequency component of the wavefield is independent of each other, a feature that makes the reflectivity method a natural algorithm for parallel and distributed implementations.

**Example**

An elastic reflectivity code requires that the user supplies the thickness, $P$-wave and $S$-wave velocities and density, for each layer. Also, for this specific implementation, the quality factor $Q$ for each layer is also required, one for $P$-waves and another one, possibly different, for $S$-waves. Figure (2) illustrates depth profiles obtained from blocking well-logs acquired at the Sorrento field, southeast of Colorado. There are approximately 400 layers in this model. Using those profiles and a constant quality factor of 1000 for both $P$- and $S$-waves, I generated the synthetic shot gather shown in Figure (3). This is the vertical component of the displacement field acquired at the surface due to a vertical point source sitting at the same location of the first receiver of this gather. The computation of these seismograms took around 8 hours on a 90 MHz Pentium-based personal computer. In the next section, I describe the distributed implementation of this technique.

**SUDREF: A Distributed Implementation of the Reflectivity Method**

The distributed implementation of the reflectivity method is based on the simple idea of a master processor assigning frequency partitions to slave processors. Figure (4) illustrates the idea. As the slave processors finish the computation of their respective frequency partitions, they send the results back to the master processor. As a final step the master processor transforms the data to the time domain. As mentioned before, sudref is implemented in standard ANSI C, and all the message passing and synchronization routines are provided by the PVM library. The implementation depicted in Figure (4) is probably close to an optimum one in the case of a homogeneous network of computers, on which the only running program is the distributed reflectivity algorithm. This is not usually the case. The typical scenario is to have computers with distinct computational power and/or with different levels of CPU utilization connected to the same network. This characterizes a heterogeneous computer environment. In this situation some strategy of load balancing is required in order to allocate more work to computers of higher computational capacity. The load balancing used here is simple, but it provided satisfactory results. It consists of splitting the total modeling job.
marking the computer was completely dedicated to this application. The test model consists of approximately 400 layers and it is shown in Figure (2). The benchmarking procedure consists of measuring the elapsed time of the frequency loop and normalizing it by the elapsed time of the sequential run (see Figure 6). Notice that, since the frequency loop is the one executed in a distributed fashion, the sequential portion of the algorithm was not considered in the speedup measurements. This sequential portion consists basically of the transformation of the synthetic displacements from frequency to time domain, which represents a small fraction of the total execution time.

All the measurements were performed for 3 distinct task sizes. As mentioned before the size of the task is the number of frequency components that the master processor assigns to the slave processor at each time. A larger task size implies a smaller number of interactions (through message transfers) between master and slave, but also implies a larger message to be sent from slave to master. In the experiments presented here the size of this message ranges from 1.9 Kbytes (task size = 5 frequencies) to 5.6 Kbytes (task size = 15 frequencies). The task size can be regarded as a tuning parameter that can be used to yield a better performance of suqref on a specific hardware. The speedup measurements are discussed next, with a brief description of the computer architecture used in each one of the benchmarks.

Silicon Graphics Power Challenge

The basic configuration of this machine is described in Table (1). Figure (7) shows the speedup curves for suqref on the Power Challenge. The speedup is very close to linear. For this problem, slightly better performance results were obtained for a task size of 5.

Performance Measurements

The computational performance of suqref was evaluated on shared- and distributed-memory computer architectures, via speedup measurements. During the bench-
Pentium-Based Network of Linux Workstations

Here, I benchmarked *sudref* on a heterogeneous (due to different clock speeds) network of Pentium processors. Its basic configuration is described in Table (2). Figure (8) shows the speedup curves for *sudref* on this platform, which were obtained by averaging several runs. Alternative measurement procedures are of course possible for this case. When more than 4 processors are used in the computation, the speedup begins to depart from linear. Since the size of the messages exchanged by all processors do not exceed the bandwidth of the Ethernet, this reduction in performance is attributed to the larger overhead imposed to the master processor to manage a larger number of tasks. Recall that the slave processors remain idle while waiting for the frequency range assigned by the master processor. As the number of processors increase, so does this waiting time. Therefore the computational performance degrades. In this example the best speedups were obtained with a task size of 10.

IBM SP-2

Finally, I evaluated the performance of *sudref* on an 16-processor IBM SP2. Table (3) describes the basic configuration of this system. Due to the higher bandwidth provided by this system, as compared to the previous one, I was able to obtain speedup results very close to linear when up to 11 processors were used (see Figure (9)). After that the computational performance of the distributed code begins to degrade. Since we are far from exceeding the bandwidth of the high performance switch that connects the processors, the same comments made in the previous section apply here. When more processors are added to the computation, the overlap between communication and computation is smaller, reducing the overall performance of the application. Note
that in this case the results were independent of the task size.

How to get the code
The source code and installation instructions for sudref can be downloaded from the CWP Linux WWW server (http://landau.Mines.edu/). Sudref is written in ANSI-C and uses the PVM library version 3.3.x.

Conclusions
In this work I presented a distributed implementation of an elastic isotropic reflectivity code. Since the modeling of synthetic seismograms is performed in the frequency domain, this procedure is very attractive for parallel and distributed implementations. The distributed reflectivity code was benchmarked on 3 computer platforms, including shared- and distributed-memory systems. On a network of 11 PC's, the time required to solve a 400 layer problem dropped from 8 hours to less than 1 hour, a speedup higher than 8 over the sequential time. Improvements on those results are possible, by increasing the overlap between communication and computation.

Finally, the fact that sudref can be used on any computer platform that supports the C programming language and PVM makes the program very attractive for the computation of synthetic seismograms for a large number of layers.

Acknowledgments
I thank Prof. John Scales for the constant encouragement. Discussions with Alejandro Murillo and Alberto Villarreal on benchmarking and PVM tricks were very helpful. I am grateful to the IBM POWER Parallel Systems remote access program, which provided the dedicated time on the SP-2, and to the Reservoir Characterization Project for the well-log data used to build the test model for the benchmarks.

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, the Army Research Office and the National Science Foundation under grant DMS-9506603.

References
Distributed Seismic Unix: Master's thesis excerpt

Alejandro E. Murillo  
Department of Mathematical and Computer Sciences  
Center for Wave Phenomena  
Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT

This paper describes TCL/TK based software called DSU (Distributed Seismic Unix) designed to assist geophysicists in developing and executing sequences of Seismic Unix (SU) applications in clusters of workstations as well as on tightly coupled multiprocessor machines. SU is a collection of subroutine libraries, graphics tools, and fundamental seismic data processing applications that is freely available via the Internet from the Center for the Wave Phenomena (CWP) of Colorado School of Mines. SU is currently used at more than 500 different sites in 32 countries around the world. DSU is built on top of three publicly available software packages: SU itself; TCL/TK, which provides the necessary tools to build the graphical user interface (GUI); and PVM (Parallel Virtual Machine), which supports process management and communication. DSU is intended to handle tree-like graphs representing sequences of SU applications. Nodes of the graphs represent SU applications, while the arcs represent the way the data flow from the root node to the leaf nodes of the tree. Pipelining parallelism is obtained when executing single-branch tree sequences, while a higher degree of parallelism is obtained when executing sequences with several branches. DSU provides tools for creating, editing, setting parameters, saving in plain ASCII files and executing SU applications sequences over several types of multiprocessor environments.

Key words: Seismic Processing, Distributed computing, Computational performance, PVM, TCL/TK, SU

Introduction

The objective of a seismic study is to obtain a description or approximate picture of the subsurface. To achieve that, geophysicists produce vibrations (through controlled explosions or vibroseis trucks) at the surface of the earth to obtain measurements of wavefield reflected from the subsurface in the space-time domain. Because the desired subsurface image must be in the space-depth domain, several geophysical applications, that include signal processing and other transformations, have to be applied to obtain the desired image.

Seismic Unix (SC) (Cohen & Stockwell Jr., 1991) and (1996) is a collection of computer programs and subroutine libraries created to process seismic data. It also facilitates the creation and testing of new applications for seismic data processing.

SU applications typically read their input traces from the standard input file and write their output to the standard output file. In this way, the user can create sequences of SU applications by means of shell script files in which several applications (along with their corresponding parameters) are specified, each one separated from the next by the pipeline operator (|) provided by Unix shells. When the shell script is executed, the output from one application is communicated to the next one through a pipeline file. Figure 1 shows a listing of a Bourne shell script representing a sequence of four SU applications. In it, synthetic data are generated with the modeling application suspike; fields of the trace headers representing that data are modified using susew. Next, the seismic data are migrated using sugazmig and fi-
Figure 2. Example of a multi-branch SU application sequence as built using the DSU graphical tool. The objective of this particular one is to compare the response of two different migration algorithms: sumigps and sugazmig.

```bash
#!/bin/sh
#
# This shell script computes impulse response:
# It uses SUSPIKE to make test data
# Migrates with SUGAZMIG
# Displays output with SUXMOVIE

suspke nspk=3 |
sushw key=dt,d2 a=50000,.05 |
sugazmig tmin=0 vmin=1 |
suxmovie perc=99 title="Suspke data"
```

Figure 1. Example of a shell script representing a sequence of SU applications. This sequence is composed of four applications (separated by |): suspke, sushw, sugazmig and suxmove.

This script computes an impulse response using SUSPIKE, migrates the data with SUGAZMIG, and displays the output with SUXMOVIE.

Here I describe DSU (Distributed Seismic Unix) (Murillo, 1996), a tool and methodology to assist the geophysicist in executing sequences of SU applications in a multiprocessor environment. Additionally, DSU extends the capabilities of the regular SU software system by providing the user with capabilities to create multi-branch sequences of SU applications such as the one shown in Figure 2. Handling multi-branch sequences of SU applications is especially useful, as it allows the user to simultaneously execute several copies of a single branch sequence, but using different parameters on each branch; or to be able to quickly compare the answers of several algorithms that have the same objective. For instance, Figure 2 shows a sequence intended to compare the results of applying two different migration al-

Finally, the migrated data are displayed for analysis using suxmove.
algorithms, in this case supsmig and sugazmig, to a specific input data created with the susynlv application.

A prototype version of this system has been implemented using TCL/Tk for the GUI (Graphic User interface) aspects and the message-passing system PVM V3.9 (Parallel Virtual Machine) for the process management and communication issues. The system has been tested on a network of LINUX and Silicon Graphics workstations. For the sake of portability, PVM functions are not directly invoked from the application. Instead, functions from an intermediate subroutine library, created along with DSU, are explicitly invoked. In this way, we avoid strict dependency on a specific message-passing software (in this case PVM). Thus, the intermediate library facilitates the implementation of DSU with several message-passing systems by just implementing the intermediate library functions with the desired message-passing system. This is especially useful when going to a multiprocessor machine, where optimized message-passing libraries are usually provided.

The organization of this paper is as follows. Sections 2 and 3 describe the DSU system design and implementation. Section 4 contains a description of the graphical user interface tool and the facilities it provides. Section 5 explains the organization of the current implementation of DSU. Section 6 is a summary of the goals achieved and future research directions in this work.

**DSU System Design**

In DSU the user builds graphs that represent sequences of SU applications. Each node of the graph represents a specific SU application, while the arcs represent how the input seismic data flow through the applications. A shell script representing a sequence of SU applications is considered a single-branch graph-tree (each application has only one child application), while extended sequences of SU applications, such as the one shown in Figure 2, are considered multi-branch graph-trees.

In DSU there are four independent subsystems that are available to the user through the GUI. Figure 3 shows how these subsystems are inter-related. The implementation of these subsystems requires the use of external subroutine libraries for accomplishing the graphical user interface, process handling and communication aspects.

**The graphical user interface**

DSU must provide the user with a graphic tool to create and manipulate graphs representing SU application sequences. This subsystem, the user interface itself, accomplishes the functions of displaying and handling the graph representing the sequence of SU applications. It also provides all the facilities to enter parameters for the applications and context-sensitive help features. It must provide buttons and menus to allow the user to invoke the subsystems explained below.

The loading and saving tasks

This subsystem deals with the tasks of loading/saving sequences of applications from and to plain ASCII files. The format for representing a sequence in a plain file was adapted from the one used to represent regular SU application sequences: a shell script. This format is explained in detail below. The loading procedure is a lexical analyzer program that extracts a graph representing a sequence of SU applications from a plain ASCII file. On the other end, the saving procedure stores a given graph in a plain ASCII file using the format explained below.

The loading procedure is in charge of reading a sequence represented in a plain ASCII file and storing it in memory. This procedure is based on a lexical analyzer program whose job is to read the input file and
assemble the characters into tokens, and, in a subprogram in charge of handling those tokens, to assemble the sequence represented in the file. The lexical analyzer procedure developed here is based in the one described by (Rochkind, 1983).

A plain ASCII file, representing a sequence of SU applications, consists of a sequence of one or more SU applications (commands) separated with bars (|). Each application name is specified as explained below. Additionally, before starting the specification of the sequence of applications, these files might contain assignment commands, in which macros or variables are assigned with values. The loading procedure will process those assignments and store the values of the variables in a symbol table. The lexical analyzer subprogram will then be able to look up the value of those variables when a token of type $MACRO$ is specified.

The algorithm used for saving the graph in a plain ASCII file consists of recursively visiting and enumerating each node of the graph, starting from the root node. This algorithm is invoked the very first time with the root node, and the recursion will visit all the nodes.

**Interface with the message-passing software**

This subsystem involves the communication and process-management software to be used. This subsystem is in charge of handling all interaction of the user with that specific software. For example, software systems such as PVM allow the user to dynamically add or remove a machine to or from the so-called virtual machine. The user must be able to perform this kind of task through the DSU graphical tool, without interacting directly with PVM.

**Distributing and monitoring subsystem**

This subsystem handles the task of executing the sequence represented by the graph and a posteriori monitoring of that execution. It initiates the execution of each node (application) represented in the graph. It also communicates the respective parameters to each application. Similarly, this subsystem provides each spawned application with the process identification of its parent application [since we are handling tree-like graphs, each node (application) in the graph only has one parent] and the process identification of its children applications, so that the input data read (or generated) in the root node of the graph flow in the way indicated by the arcs of the graph.

Figure 4 shows how this subsystem works. When it is invoked, it combines the information about the hardware platform available and the information about each one of the applications included in the graph to attempt an efficient execution of the sequence on that platform. For example, over a heterogeneous network of computers, CPU-intensive applications would be assigned to computers with high processing capacity, while I/O intensive applications would be executed on disk or tape servers.

The task of initiating each application is achieved by visiting each node of the graph using a recursive pre-order visit of the nodes of the graph. Starting from the root node, each application is spawned and quickly provided with the parameter and parent/children nodes information.

**Format to represent a graph in a plain ASCII file**

To provide the user with an option to save a graph representing a sequence of SU applications for a posteriori use, we have defined a format to represent those graphs in a plain ASCII file. Since many sequences of SU appli-
with this specification, a list of the parameters provided for the application is specified using the following format:

\[
\text{parameter\_name=parameter\_value}
\]

Figure 5 shows how the sequence represented by the graph shown in Figure 2 is stored in a plain ASCII file. Note that the \text{hostname} field has been set up for the \text{sunmigs} and \text{sugazmig} applications, but since it is an optional field, it is not explicitly specified in every application. Also, by inspecting the \text{parent\_node} fields, it can be easily determined that the application \text{sushw} (node 2) has \text{sugazmig} (node 5) and \text{supsmig} (node 3) as children applications.

**DSU Implementation Details**

DSU is an event-driven graphic user interface whose main program is the implementation of the state diagram (Rumbaugh, 1991) shown in Figure 6. In summary, the DSU main program accomplishes some initialization tasks and then enters into a loop to accept commands for building sequences of SU applications. When requested by the user, it executes and monitors the execution of the created sequence.

DSU is implemented using TCL/TK for the GUI aspects and using PVM for the process management and communication aspects. TCL/TK is a free, extensible scripting language with some extensions that incorporate an X11 toolkit (called TK) that make it easy to build graphical user interfaces (Ousterhout, 1994). PVM provides facilities for spawning, communication and synchronization of processes over a network of heterogeneous machines. There are several message-passing systems available (Snair, 1996), (Gropp, 1994). (Carreiro & Gelernter, 1989), and (Flower, 1991). PVM was selected mainly because it provides spawning capabilities and also has been successfully ported to several shared and distributed memory supercomputers. For the sake of easy portability, however, an intermediate library is used to avoid embedding direct PVM subroutine calls in the code.

The GUI has a similar look and feel to the X-window interface of PVM (XPVM) (Geist, 1994); actually, some portions of the code were extracted from the XPVM code. The steps followed to obtain a preliminary implementation of DSU are the following:

1. A Graphical User Interface (GUI) was implemented using TCL/TK. Panels, menus and buttons were designed to provide the user with options to create, edit and handle sequences of SU applications.
(ii) The procedures associated with the tasks of loading and saving sequences of SU applications in plain ASCII files were implemented.

(iii) The procedure for running the sequence of applications in the network was developed. The work here involves spawning a process for each application in the sequence and distributing the respective parameters. It also involves the introduction of changes inside the code of the SU applications to achieve the necessary synchronization and data communication, once the applications have been spawned.

(iv) Finally, some functions were implemented to provide the GUI with a mechanism to report about the status of the execution of a sequence.

The graphical user interface (GUI)

As we mentioned above, the GUI was implemented using the scripting language TCL and the X-window toolkit TK. The GUI is composed of a panel of buttons, message widgets and a scrollable canvas where the graph representing a sequence of SU applications is displayed. As in a typical X-window based GUI interface, each button, menu item and mouse button has a callback procedure associated with it. The callback procedure is fired by X-windows whenever the associated widget is clicked by the user (e.g., a button is clicked on). Section II is a full description of the GUI.

Loading and saving procedures

The loading procedure is in charge of reading a sequence represented in a plain ASCII file and storing it in memory. This procedure is based on a lexical analyzer program whose job is to read the input file and assemble the characters into tokens, and in a subprogram in charge of handling those tokens to assemble the sequence represented in the file. The lexical analyzer procedure implemented here is based on the one described by Rochkind (1985).

Distributed execution

To implement the distributing task, we have used the spawning facilities provided by PVM. When using PVM as the parallel machine (virtual), each spawned process is identified with a unique integer number, called tid. These identifiers are used by the applications to specify the recipient of a message to be sent or the source (sender) of a received message. Therefore, an important task of the distributing procedure is to inform each process associated with an application of the sequence of the tid associated with its parent and children applications.

Starting from the root node, the distributing process performs a recursive pre-order visit of each node of the tree representing the sequence of applications. When a node is visited, the application associated with it is pvm-spawned and its tid (returned by the PVM spawning function) is saved to be returned later as the function return value. Then, each subsequence starting at this node is recursively spawned. Each recursive call will return a tid that corresponds to a child subprocess of this task.
When all the children tids are collected they are sent to the application along with the parent tid. To finish the visit to this node, the tid of this node is returned.

Monitoring task

Since DSU is an X-window graphical user interface, the interaction with it has to be done through events. Since it is not desirable to block the GUI with the monitoring process, several alternatives are possible. The first one is to initiate an independent monitoring process. The second alternative is have the user request a report of the progress once in a while, in which case the GUI can fire a subprocess that provides snapshot information. For simplicity, in the current implementation we followed a combination of these alternatives. It works in the following way. At interval times, after spawning all the applications of a sequence, DSU will fire a process in charge of checking for messages from the applications. Before finishing their execution, the applications must send a message to the GUI indicating their completion status. For messages indicating successful completion, the user is informed, but in cases of abnormal completion, the whole execution is aborted and the user is advised to check the log file for the application that abnormally ended. DSU keeps track of the number of applications spawned and when all of them have completed their execution, a message indicating the elapsed time taken by the execution is shown to the user.

DSU Graphical User Interface

DSU provides a graphical user interface (GUI) tool for creating, editing, setting parameters and executing a graph representing a sequence of SU applications on a multiprocessor environment. It allows the user to set up application parameters, input/output filenames, and other similar tasks related to the applications. It provides context-sensitive help about almost every component of SU. It also allows the user to interactively check the source code of selected main programs and subroutine library functions.

Figure 7 shows the DSU main interface. The upper part of the interface window is a message widget, where the current status is displayed. Beneath it, there is a panel of buttons that allows the user to build and handle the sequence of applications. The save button allows the user to save the current sequence of applications. It will allow the user to surf over the available file system to select the appropriate place to store the sequence. The sequence is saved in a plain ASCII file using the format that was described in section 3. The print button will display the information contained in each element (application) belonging to the subsequence starting at the current or active node on the screen. This information includes application name, parameter names and values. The load button allows the user to load a sequence of applications represented in a plain ASCII file.

Buttons in the second group are intended to handle single application nodes (single nodes of the graph). The fork and add buttons prompt the user for an application name to be added to the graph just after the current one. The current application is highlighted with a green color. A node of the graph becomes the current one if it was the last one inserted, if the user clicked the left mouse button on it or if one of its children nodes was removed. The fork button allows the user to create branches in the sequence. The delete button removes the current node from the graph. DSU ensures that nodes with multiple branches deriving from them cannot be removed.

The third group of buttons deals with the execution of the represented sequence. The execute button is actually a menu with two options. The first option allows the execution of the sequence with regular SU (under pipeline communication); since it is not possible to execute multi-branch sequences via pipe-file communication, only the main branch of a multi-branch sequence will be executed. This is useful when we want to test only the main branch of a sequence. The second option allows a distributed execution of the sequence. By selecting it, the user will fire the DSU distributing task. This task will initiate each application of the sequence on the machine best suited to its needs and will setup the communication among them accordingly to the arcs of the graph representing the sequence.

The other button associated with the execution of the graph (sequence) handles the interface with the message-passing system, PVM. As shown in Figure 7, the button labeled PVM is a menu with options to start, stop and reset PVM. Additionally, it allows the insertion into and removal of machines from the virtual machine.

The farthest right button of the interface is the help button. This button will prompt the user for the entity for which help is needed. There, one can request help about almost anything related with SU: library functions, main programs and others. Once the user has selected the entity, a scrollable window containing the help for that entity will be shown. When obtaining help for SU main applications, a see code button is provided to enable viewing of the source code for the application.
Mouse buttons

In addition to the buttons on the screen, the buttons on the mouse are an essential part of the user interface. Clicking once on the icon representing an application with the right button of the mouse will bring up a scrollable window containing the help information for that application. Figure 8 shows the help window for the ap-
application susynlv. This window provides the see code button, which allows the user to take a look at the actual source code of that application.

Clicking twice on the left button of the mouse on an application icon will bring up the window that allows the insertion of values for the parameters of the application represented by that icon. Figure 9 is an example of a window to get parameters for the susynlv SU application. This window has a hostname button that allows the specification of a hostname for that application; also, by using the buttons stdin and stdout, filenames can be entered to set up the file names for the standard input and output, respectively, of that application. Only the root node is allowed to have the standard input redirected; each other application in the sequence must expect its input from its parent node. Similarly, only leaf nodes of the graph (sequence) are allowed to have the standard output redirected. The options button allows the insertion of some parameters under the format: -option option.value; this syntax for parameters is understood for some X-Windows related applications that are part of SU, such as surfmove for example. Clicking the save button will save the parameter values entered; alternatively, hitting the Enter key on any parameter field will also automatically save all the parameter values entered too. Using void as a parameter value, indicates that the default value must be used.

Finally, clicking and holding the middle button of the mouse on an application icon will allow the user to move that icon to any position inside the drawing canvas.

A typical session with DSU

Figure 10 shows a typical session with the DSU graphical tool. The user usually loads or creates a sequence of applications and then spends most of the time providing parameters to the applications, adding to and removing single applications from the sequence and executing it. The user will occasionally save the current sequence with some specific parameters, then modify them and save again. Next, we discuss some typical actions when using DSU.

There are two ways to start the creation of a SU application graph. The first one is to use the load button to load a previously created sequence and work from there. The second one consists of clicking the add or fork buttons to bring up the panel containing all the available SU applications. Then by clicking twice on the
desired application name, that application gets added to the sequence just after the current one.

The GUI provides a button to select the type of execution that the user wants DSU to perform. Clicking this button will raise a menu with two options for execution: one is called Local and the other Distributed. The local option corresponds to executing the sequence as in a regular Unix-shell environment. The Distributed option will execute the sequence over the multiprocessor environment available.

**DSU Software Organization**

This section contains general information about DSU. It contains a short introduction to its organization and directions for installing it. It provides directions at a level appropriate for users with previous experience with PVM. We have also included directions to get you started on using DSU. Here we do not include directions about how to install SU, PVM or TCL/TK. The following World Wide Web sites contain comprehensive information about that:

**CWP/SU**
http://www.cwp.mines.edu/cwpcodes

**PVM**
http://www.epm.orl.gov/pvm/pvm_home.html

**TCL/TK**
http://www.sco.com/Technology/tcl/Tcl.html
Makefile.config: Modifications to
SCWPROOT/src/Makefile.conf
necessary to compile SU
applications under the DSU environment.
This file is included by DSU sub-
directory Makefiles.

A brief tour of the source directories

The DSU software is distributed in several subdirectories
grouped in a root directory called dsu. The directory
dsu must be placed in the subdirectory src of the SU
software so that the root directory of DSU can be re-
ferenced as: SCWPROOT/src/dsu. The root directory
of DSU is organized as follows:
dsupar: Contains the files describing the parameters
names and default values for each SU ap-
plication. There is a file for each SU ap-
plication. Additionally, a program tool that al-
 lows the extraction of the parameter names
of any SU application is also included here;
geometric: Contains Makefiles to compile SU graphi-
cal applications intended to be used with
DSU. It is composed of two subdirectories:
xplot, which contains a Makefile to
generate executables for the Xlib based X-
windows applications; and, psplot, which
contains a Makefile to generate the execut-
ables of postscript based graphical applica-
tions;
lib: Distributed Seismic Unix subroutine lib-
main: Contains a Makefile to generate executables
to be used with DSU, for the SU core ap-
lications;
main.dst: Contains PVM based applications that can
be used as any other SU application. It con-
tains the appropriate Makefile;
gui: Contains the graphical user interface
(GUI). PVM3 and TCL/TK must be avail-
able before compiling the GUI. The Make-
file in this directory must be custom-
ized before installing. The executable is
called dsu and is placed in the directory
SCWPROOT/bin.

Files in the DSU root directory

DSURC: Used to setup the environment
variables needed to run DSU.
Among this variables we have:
PVM_ROOT,
TCL_LIBRARY, DSULOG,
etc.

README: Directions to install DSU.

Requirements for installing the package

The only requirements for installing the package are:

(i) A machine running the UNIX operating system.
(ii) Seismic Unix V28
(iii) PVM 3.8 or higher
(iv) TCL/TK (7.4/4.0)

The package has been successfully installed on:

- Pentium workstations running LINUX
- Silicon Graphics workstations (Indigo II and Indy)
- IBM SP2

Brief summary of the steps to install DSU

(i) Install SU V28; PVM 3.8 or higher; TCL/TK
(7.4/4.0)
(ii) Obtain the tar file with dsu (dsurc.tar.gz). The

(iii) cd $CWPROOT/src

(iv) Create the root directory for DSU by untarring the distribution file in this way: (this creates the directory: $CWPROOT/src/dsu)

gunzip -c dsusrc.tar.gz | tar xvf -

(v) cd dsu (or cd $CWPROOT/src/dsu)

(vi) Make sure the following environment variables are set:

$CWPROOT (CWP SU main directory)
$PVM_ROOT (PVM main directory)
$PVM_ARCH (PVM main directory)

(vii) Customize the file Makefile.config. Since SU was previously installed, not much needs to be done here. Check the CTARGET for your machine.

(viii) Make sure the pointers to the TCL and TK libraries and include files are correct in the file: $CWPROOT/src/dsu/gui/Makefile

(ix) Type: make INSTALL

How to run dsu

To use DSU, follow these steps:

(i) Customize the file $CWPROOT/src/dsu/DSURC according to your site and ADD the following line to your .cshrc file:

source $CWPROOT/src/dsu/DSURC

(if you are not using csh or tclsh the customization of the corresponding files is similar). If TCL/TK are not installed in the default places, the environment variables TCL_LIBRARY and TK_LIBRARY must point to the places where they were actually installed;

(ii) Put the hostname of the machines you intend to use with PVM in the file $HOME/.dsu.hosts. DSU will set up a PVM virtual machine with the machine whose names are included in this file;

(iii) Make sure that $CWPROOT/bin is in your path;

(iv) Invoke DSU by typing: dsu

Summary and Future Work

Users of a well-designed distributed system should perceive a single, integrated computing facility even though it may be implemented on many computers in different locations (Couloris, 1994). Distributed Seismic Unix was designed with this idea in mind. The result has been an interface designed to simplify the task of creating, editing and executing sequences of SU applications in a multi-processor environment. In DSU, the user specifies the sequence of applications by drawing a graph whose nodes represent applications and whose arcs describe the way in which the data flow among the applications.

DSU is a notable enhancement to the capacities provided by the already useful SU software. The system has been designed to be portable and easy to extend.

The ability to handle multi-branch sequences of SU applications is one of the most important features provided with the system. This feature allows the user to simultaneously execute several copies of a single-branch sequence of applications, each one with different parameters.

Portability of the final product has been a top priority in both the design and implementation of DSU. Even though the current implementation of DSU utilizes PVM for process handing and the communications issues, an alternative message-passing system could be used with little effort, since an intermediate library was also developed to avoid embedding direct calls on PVM functions in the code.

Future research

There are several research directions that can be followed at this stage of the development of Distributed Seismic Unix. A major question to be explored is how the concept of active messages can help to dramatically improve both load balancing and the monitoring facilities in DSU.

Another aspect of chief importance is the design and incorporation of applications whose main purpose is to accelerate the execution of individual SU applications without explicitly parallelizing their algorithms. An example of this kind of application is the Fork-and-merge task. A Fork-and-merge task would take an SU application name as an argument and initiate many instances of that application. As traces arrive at the fork-and-merge application they are forwarded to the instances of the SU application in a round-robin fashion. The instances of the SU application send their output to the next application in the sequence. This kind of task may accelerate the execution of the provided SU application.

Acknowledgement

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, and by Los Alamos National Laboratory, as part of the ACTI project. I am also grateful to my advisor at CSM, Dr. Jean Bell, for her guidance in developing this work; to Dr. Jack Cohen and John
Stockwell Jr. for their great efforts making SU what it is today. And to John Scales for his interest in my work and his continual encouragement.

References


Welch, B. 1995. Practical programming in TCL and TK. Prentice Hall.

APPENDIX A: TCL/TK Overview

The information contained in this appendix was extracted entirely from the TCL/TK WWW home page. To obtain more detailed information about TCL/TK, please see this World Wide Web site:

http://www.sco.com/Technology/tcl/Tcl.html

This is a relatively short introduction to TCL/TK, which is designed to get you started in using TCL/TK. It is not a manual, or a comprehensive book on TK, or a book on programming, but attempts to introduce TCL/TK at a level which will be appropriate for most readers with some previous programming experience. Further information on TCL/TK can be found in the books by Ousterhout (Ousterhout, 1994) and (Welch, 1995). Some readers may also find that examining the source files (such as the widget demo program) supplied with the distribution, will be helpful. In addition, there is a newsgroup for exchange of information about TCL.

TCL/TK is a programming system developed by John Ousterhout at the University of California, Berkeley, which is easy to use, and which has very useful graphical interface facilities. TCL is the basic programming language, while TK is a ToolKit of widgets, which are graphical objects similar to those of other GUI toolkits, such as Xlib, Xview and Motif. Unlike many of the other toolkits, it is not necessary to use C or C++ in order to manipulate the widgets, and useful applications can be built very rapidly once some expertise of the TCL/TK system has been gained.

Some users will naturally wish to use the widgets with C or C++. The TCL/TK system can be configured to work co-operatively with other programming languages such as C or C++, and facilities to support this are described in section 3 of Ousterhout's book.

The TCL language is normally interpreted, so TCL applications will normally not run as fast as equivalent C programs. For a large class of applications this is not a disadvantage, however, since the speed of processing of modern computer systems is more than adequate. Where speed of processing is essential, use can be made of a TCL compiler, or processing can be carried out in a compiled language, such as C or C++, and the user interface written in TCL.

There are versions of TCL for different hardware systems, and for different operating systems, so TCL is to a large extent portable. However, it has already been noted that where TCL programs are used to access operating systems feature, portability will be sacrificed to convenience on the user hardware.
Distributed 3D finite difference modeling via domain decomposition

Alberto Villarreal and John A. Scales
Department of Geophysics and Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401, USA

ABSTRACT

Finite difference modeling of wave propagation in heterogeneous media is a useful technique in a number of disciplines, including seismology, laboratory ultrasonics, ocean acoustics, radar imaging, non-destructive evaluation, and others. However, the size of the models that can be treated by finite difference methods in three spatial dimensions has been rather limited except on supercomputers. We describe a finite difference domain-decomposition method for the 3D acoustic wave equation which is well-suited to distributed parallelization. We have implemented this algorithm using the PVM message passing library, and show here benchmarks on two different distributed memory architectures, the IBM SP2 and a network of low-cost PC's running the Linux operating system. We present performance measurements of this algorithm on both the low bandwidth PC network (10Mbs ethernet) and the high bandwidth SP2 cluster. These results demonstrate the feasibility of doing distributed finite difference acoustic modeling on networks of workstations, but point to the substantial efficiencies that can be expected as higher bandwidth networks become available.

Key words: Finite Difference Modeling, Domain Decomposition, Acoustic Wave Equation, Parallel Computing, Distributed Computing

Introduction

Finite difference and finite element methods are the most direct approaches to performing realistic seismic modeling in complex media and to provide a complete description of wavefields with essentially arbitrary variation of material properties. If a detailed and accurate modeling of the seismic experiment in a heterogeneous medium is required (including head waves, diffractions and surface waves), direct methods, such as finite differences, are the only possibility. However, the main limitation in this numerical technique is the great amount of computation and memory required on grids fine enough to meet the constraints of accuracy, stability, and minimum grid dispersion. In order to get some understanding of the scale of the computational demand, let us consider the requirements for a typical seismic model with dimensions 3 km \times 3 km \times 3 km, with velocities varying from 1.5 km/s to 4.5 km/s. If we consider the dominant wavelength to be 60 m (which corresponds to a source with a peak frequency of 25 Hz for the slowest wavespeed in the model), then the problem size in wavelengths is 50 \times 50 \times 50. Let us assume we perform acoustic forward modeling for 4 s on this model using a fourth-order explicit finite difference algorithm; then, for 12 grid points/wavelength, we will require a 600 \times 600 \times 600 grid and 7200 time steps. This makes a total of 1.5 \times 10^{12} grid point evaluations. Considering that the calculation to update each grid point requires about 25 floating point operations, then the complete calculation will take 38 \times 10^{12} floating point operations, which means about 3.8 \times 10^5 seconds (approximately 4 days) for each source running on a 100 Mflop workstation (assuming that approximately 1.2 Gigabytes of RAM is available).

In this work the problem of computational demand is
addressed using high order finite difference operators and parallel processing techniques. The traditional approach to parallelizing finite difference algorithms has been to use the SIMD paradigm with a fine-grained partition, as for example in Myczkowski (1991) and Fricke (1988).

Our approach is to perform a coarse spatial domain decomposition of the problem, dividing the finite difference grid into groups of adjacent points and to assign each one of these subdomains to a different workstation in a network of workstations. This approach permits us to adapt the size of the groups of points to the available hardware and/or other constraints of the model itself (such as boundary conditions, inhomogeneities, and nonuniform grids).

Considering the ubiquity and low cost of networked workstations, we want to explore the feasibility of performing 3D modeling in such an environment. The goal of this project is to develop scalable finite difference modeling algorithms on distributed memory networked workstations, as well as to study hardware performance-related issues involved in the execution of these algorithms. Issues such as the computation/communication ratio for different problem sizes and/or different number of processors could permit us to predict the behavior of the algorithms when different hardware is used. Basically, this will indicate to us the feasibility of using low cost hardware as a platform for developing and testing parallel software capable of scaling to faster hardware.

As a starting point, a distributed memory algorithm for 3D acoustic modeling is designed using a message-passing programming model and is implemented using the PVM library as the message passing interface. We show benchmarks of this algorithm on two different parallel systems. These results will be used to design a distributed elastic modeling algorithm, which is even more challenging in terms of computation and communication requirements. An important result we expect to develop using the results from these experiments is an analytical model describing parallel efficiency of the modeling algorithm as a function of hardware-dependent parameters, such as the speed of the network and the speed and the number of CPU's.

A portion of this paper is devoted to a brief introduction to parallel architectures and parallel programming models. Following this overview, details of the finite difference acoustic modeling are discussed. Next, a distributed implementation of this algorithm is presented, followed by benchmark results showing the parallel efficiency of the algorithm compared to single processor executions.

Parallel Architectures and Programming Models

Parallel Architectures

Multiprocessors, or shared memory MIMD (Multiple Instruction Multiple Data) computers, are perhaps the best known parallel computer architectures. The term MIMD means that each processor can execute a separate stream of instructions on its own data, and shared memory implies that all processors share access to a common (usually large) memory. Examples of this class of machine include the SGI Power Challenge, the Cray Y-MP and C-90, and many multiprocessor workstations.

Distributed memory MIMD computers comprise another important class of parallel computer systems. Distributed memory means that memory is distributed among the processors, rather than placed in a central location, as in multiprocessors. Here, each processing node executes its own program. This program may access local memory and may send and receive messages to or from other nodes. Messages are used to communicate with other nodes which means to read and/or write remote memories. The CM-5 and the Cray 3TD are examples of this type of MIMD machine.

A more specialized class of distributed memory architecture is the SIMD (Single Instruction Multiple Data). In these machines, all processors execute the same instruction stream on different pieces of data. This architecture is best suited to problems characterized by a high degree of regularity. The CM-2 is an example of this architecture, although MIMD machines such as the CM-5 can also emulate the SIMD characteristics.

Another class of MIMD parallel computer system is a cluster of networked workstations (NW). It can be built as a group of dedicated processors linked by a fast network or switch (as in the IBM SP2 cluster), or it can consist of a dynamically varying set of networked standalone workstations that perform long running computations during idle periods.

Modern processors commonly used in workstations today incorporate supercomputer technology to a certain degree, such as pipelining and other techniques, and in theory, aggregated memory and computing power in a cluster could approach supercomputer performance. Because NW's are created from standard systems, they are easy to scale. For example, as fast workstations become available, they can be added to the system just by plugging them into the network. However, a critical issue in these systems is the network. Applications with high synchronization and communication requirements are likely to run inefficiently in an NW if the network is slow.

There have been a number of parallel implementations of wave equation modeling algorithms. Most of
them have used the shared memory or SIMD architectures. Vafidis et al. (1992) have presented results of elastic 2D finite difference modeling in transversely isotropic media using a fully vectorized algorithm. Vector computation can generate important speed-ups in finite difference algorithms, although until recently this capability was present only in costly supercomputers. Other algorithms targeted to distributed memory architectures have been published, as for example in Myczkowski (1991), who describes an efficient 2D acoustic modeling code for a CM-2 system.

There have been several distributed implementations of geophysical algorithms in NW's. For example Black et al. (1991) showed a study of a 3-D depth migration on a network of RISC workstations, and Murillo (1993) and Almasi et al. (1992) showed performance results of a 2D Migration via a depth extrapolators algorithm implemented in a NW. Bunge and Baumgardner (1995) have performed thermal convection calculations for the Earth's mantle using a message-passing algorithm on a cluster of workstations. We know of no previous work showing the feasibility of 3D finite difference wave equation modeling in NW's.

Let us mention some of the factors which led us to consider NW's as our computer platform:

- Low cost and availability.
  Given the low cost per node of Unix workstations, these networks are becoming widely available. And often when these workstations are used for interactive work, they remain idle during nights and weekends and thus become available for performing long-running computations during those idle periods with no extra cost.
- Software portability.
  Given the high degree of standardization present in Unix workstations, it is possible to design portable software able to execute in a broad range of architectures. This means that the distributed application will take advantage of a heterogeneous network of workstations, rather than being constrained by a single architecture. Another important issue regarding portability is related to the portability of the message passing software itself. There are several freely available portable message-passing libraries. A popular example is Oak Ridge National Laboratory's PVM (Parallel Virtual Machine) (Geist et al., 1994), which we have selected to build our parallel system.
- Software scalability.
  Parallel software must be scalable (i.e. the size of the problem we are trying to solve, or the performance of the application, must increase when more resources become available) in order for the application to be cost-effective. And given that NW's are scalable, either by increasing the number of nodes or the bandwidth of the network, then it is possible to use a NW to design and test parallel systems with the idea of porting them to a larger NW or to an MPP system.

We must make clear, though, that by exploring the feasibility of solving a geophysical problem in a NW, we do not mean to replace high performance computational resources such as supercomputers, which because of lower latencies and higher bandwidths are more efficient for applications with large communication requirements. The NW can provide a low cost solution to many problems, and it can also improve the effectiveness of supercomputer resources by permitting them to be used for the most demanding applications while the NW can be used as a development and testing platform for parallel applications. However, advances in networking technology and processor performance are expanding the set of applications that can be executed efficiently on NW's.

Programming Models

A parallel programming model is a means to describe parallel algorithms for a certain computer architecture. For example, in the shared memory programming model, it is assumed that the data are in a memory which can be accessed by any processing element, and so it is well suited for shared memory machines. In the message passing programming model (which is a special case of distributed memory programming models), a parallel computation consists of several tasks executing concurrently, and every task can read and write its own memory, as well as remote memories (located in other processing nodes) by sending/receiving messages (a message is a set of bytes transferred from one computer to another computer), thus being appropriate for distributed memory architectures. However, with the advent of new programming tools such as message passing libraries and distributed shared memory systems, algorithms designed in a certain programming model are not restricted to execution in a particular architecture.

In this work, the finite difference algorithm will be designed using the message passing programming model. Implementing this algorithm making use of a portable message passing library, such as PVM, will permit us to execute the code in any machine in which PVM is available, including shared memory multiprocessors and Massively Parallel Computers (MPP).
Problem Statement: Finite Difference Modeling

A basic problem in seismology is the determination of the motion on the free surface and in the interior of a layered acoustic or elastic medium caused by an impulsive source. In what follows we will assume that the medium is acoustic and its density is constant.

Modeling the motion of the medium

In that case the equations of motion for the medium are

\[
\frac{\partial^2 p}{\partial t^2} + s(x_i) = c^2 \frac{\partial^2 p}{\partial x_i^2},
\]

(1)

where \( p \) is the pressure, \( x_i \) (\( i = 1, 2, 3 \)) are the three cartesian coordinates (later labeled \( x, y, \) and \( z \)), \( t \) is the time, and \( s(x_i) \) is the source term. The wave can be initiated using initial conditions, defined by \( p \) and \( \partial p/\partial t \) at \( t = 0 \), or using an explicit source function \( s(x_i) \). In this case an explicit source function will be used.

Discretization of the model

We divide the model into a grid of \( N_x \) by \( N_y \) by \( N_z \) points. We call \( \Delta x, \Delta y \) and \( \Delta z \) the distance between points in the grid in the \( x, y \) and \( z \) axes respectively, then we get \( x = n_x \Delta x, y = n_y \Delta y \) and \( z = n_z \Delta z \), where \( n_x = 1, 2, \ldots, N_x \), \( n_y = 1, 2, \ldots, N_y \) and \( n_z = 1, 2, \ldots, N_z \). Also if \( \Delta t \) is an increment in time, then \( t = k \Delta t \) where \( k \) is the time step with \( k = 1, 2, \ldots \).

Given this discretization, we can now describe a finite difference scheme to approximate the solution of the acoustic wave equation.

Numerical solution of the wave equation using finite differences

We substitute a \( q \)-th order central difference operator for the second space derivatives in the wave equation, and a second-order central difference operator for the second time derivatives, to obtain the discretized finite difference equation. A general framework for deriving higher-order finite difference schemes was proposed by Dablain (1986). He expresses the \( q \)-th-order centered finite difference operator for the second space derivative with respect to \( x_i \) as a weighted sum:

\[
\frac{\partial^2 p}{\partial x_i^2} \approx \frac{1}{\Delta x_i^2} \left[ w_0 P_r + \sum_{k=1}^{c/2} w_k (P_{r+k} + P_{r-k}) \right]
\]

(2)

where \( P_r \) are the values of the pressure \( p \) at the discrete position \( r \), in the \( i \)th direction. The weights \( w_0 \) are derived from the power series expansion of \( P_r \) and further simplification. The details of this calculation and values of coefficients \( w_0 \) will not be given here.

For the second time derivatives, we have the second-order central difference operator:

\[
\frac{\partial^2 p}{\partial t^2} \approx \frac{1}{\Delta t^2} \left[ P_r^{k+1} - 2P_r^k + P_r^{k-1} \right].
\]

(3)

Substituting these operators into the wave equation, we obtain an explicit formula for the values of the wavefield at time \( k + 1 \). As an example, for \( q = 2 \) we obtain the second-order (in space) scheme:

\[
P_{r+1, n, m} = r_{i, n, m} \left( P_{r-1, n, m} + P_{r+1, n, m} + P_{r, n-1, m} + P_{r, n+1, m} - 6P_{r, n, m} \right) + 2P_{r, n, m-1} - P_{r, n, m} + S_{i, n, m}
\]

(4)

where \( P_{r, n, m} \) is the value of the pressure at position \( (n_x \Delta x, n_y \Delta y, n_z \Delta z) \), and

\[
r_{i, n, m} = c_{i, n, m} \left( \frac{\Delta t^2}{\Delta x_i^2} \right)
\]

(5)

where \( c_{i, n, m} \) is the velocity at \( (n_x \Delta x, n_y \Delta y, n_z \Delta z) \) and where we have assumed that \( \Delta x = \Delta y = \Delta z \).

Constraints on the discretization of the model

There are several issues involved in the choice of the grid parameters in the discretization step. The first problem arises from the fact that in order to keep the model numerically stable, the ratio \( (\Delta t/h) \), where \( h \) is the maximum grid spacing, must be small. Specifically, the largest time step which must be used to guarantee stability is (Mitchell & Griffiths, 1980)

\[
\max (\Delta t) = \frac{\mu \max (\Delta x, \Delta y, \Delta z)}{\max (c_i)}
\]

where \( c_i \) (\( i = 1, 2, \ldots \)) are the acoustic wave speeds involved in the computation, and \( \mu \) is a constant that depends on the order of the method used.

The second problem, grid dispersion, comes from the fact that wave propagation in discrete systems is inherently dispersive. The way to minimize grid dispersion is to make the wavelengths long compared to the grid spacing. It has been observed empirically that using at least 10 grid points per the minimum wavelength involved in the computation for a second-order scheme, or at least 4 grid points per minimum wavelength for a fourth-order scheme will keep the grid dispersion at values sufficiently low so that it does not disturb the modeled wavefield in typical geophysical applications. More detailed quantitative analysis of the dispersion problem, including the relation between operator length and the required num-
number of grid points per shortest wavelength for a required accuracy, is given by Holberg (1987) and Marfurt (1984).

Finally, the accuracy of this numerical method, whose sources of error come from the replacement of continuum derivatives by finite difference operators, directly depends on the spatial and temporal sampling. Given that these values must be small in order to meet the criteria for stability and grid dispersion discussed above, then in typical geophysical applications the accuracy of the results will be acceptable.

Method

Sequential algorithm

The conventional sequential algorithm for extrapolating the wavefield looks as follows:

INPUT:
A model of material properties.

OUTPUT:
The wavefield at each time step

BEGIN
  - set up data structures, variables and constants.
  - read velocity and density models.
FOR every time step DO
  FOR every grid point DO
    - Update the wavefield.
  END
  - Update seismogram
END

Distributed algorithm design

Domain Decomposition

Domain decomposition involves assigning subdomains of a full computational grid to separate processors and solving the equations for each subdomain concurrently. Grids used to represent models in finite difference applications are natural candidates for simple domain decomposition techniques because of their regularity. This is especially evident in an explicit finite difference scheme, which can be finely partitioned to expose the maximum possible concurrency (i.e. defining a task for each grid point), because there are no data dependencies when computing the wavefield at different grid points for a given time step.

For our purposes, the grid can be partitioned in a coarser way, agglomerating a whole set of grid points (a subdomain) in a single task. Several possibilities exist for dividing the model into subdomains. Figure (1) shows two possible decompositions for a three-dimensional grid.

![Figure 1. Two possible domain decompositions for a three-dimensional grid. Each subdomain contains the same number of grid points in both decompositions. A one-dimensional decomposition is used in this work.](image)

The most natural decomposition from a coding point of view is a one-dimensional decomposition in the direction with the largest number of grid points as shown in Figure (2) and assigning each subdomain to a different processor. However, this decomposition, although adequate for a one-dimensional communication hardware (such as the Ethernet), seems to be inefficient for a nearly cubical grid, because the surface area of each subdomain (where the communication will take place) will be large relative to its volume (where the computation will take place) resulting in a low computation to communication ratio. A one-dimensional decomposition is used in this work for the present algorithm, and it will be tested in models with different geometry to investigate the effects of these differences on the performance. In the future we plan to permit the user to select the most appropriate decomposition for the input model.

Interprocessor communication

In the coarse-grained partition shown in Figure (2), each individual task (running in an individual processor) approximates the wavefield at each grid point in the subdomain at time \( k = 1 \), based on the values of the wavefield at time steps \( k \) and \( k - 1 \) at the same gridpoint and its adjacent neighbors. We immediately see that the wavefield can be updated simultaneously in all of the subdomains, except in those gridpoints at the borders since their neighbors are in a different layer. In this case the value of the wavefield at those neighboring gridpoints must be replicated in the contiguous layer, Figure (3), in order for the wavefield in that contiguous layer to be totally defined, and so to guarantee the continuity of the wavefield across the subdomains. This is the essence of the parallel algorithm. At each time step, the border layer of each subdomain must be interchanged between every two adjacent subdomains. And the thickness of this layer will depend on the order of the finite difference operator.
e.g., the thickness will be 1 grid point for a second-order operator, 2 grid points for a fourth-order operator, etc.

A basic algorithm

In summary, the parallel algorithm using the domain decomposition described can be represented as follows:

INPUT:
A velocity and density model.

OUTPUT:
The wavefield at each time step

BEGIN

- set up data structures, variables and constants.
- read velocity and density models.
- set up a partition of the grid according to the number of available CPU's.
- Create a task for each element (subdomain) of the partition and send to each task its correspondent density and velocity values.

FORALL tasks simultaneously DO

FOR every time step DO

FOR every grid point DO

- Compute wavefield

END

- Interchange border grid points with each adjacent layer.
- Update seismogram in each layer

END

- gather seismogram information from every task.

END.

This algorithm presents particular problems for parallel execution. It is a memory and communication intensive problem, especially if low-speed networks are to be used. An important amount of interprocessor communication is generated at every time step, considering that data must be interchanged between neighboring subdomains at every time step in order to maintain continuity of the extrapolated wavefield.
An improved algorithm

In order to minimize the impact of interprocessor communication on the performance of the algorithm, we must always look for opportunities for overlapping computation and communication. In this case, we can observe in the pseudocode that the data contained in the messages that are being interchanged between these tasks will be used only in the computation of the borders of each subdomain, and the computation of the inner grid points can take place while the messages are traveling across the network. Every task can update the border grid points, send these values to the other tasks, and start computing the inner grid points without waiting for the new data to arrive. After computing the inner grid points, the task must wait for the other tasks' information before updating again its border grid points. Depending on the speed of the network and processors, the new values could already be available in local memory by the time the task is ready to access them, thus greatly reducing the waiting time for each task. Notice here that the bigger the subdomain (i.e. the larger the number of grid points each task has to update, and so the longer that task is busy updating those grid points), the better the opportunities for a total overlapping between computation and communication.

Taking this overlapping procedure into account, the FORALL loop of an improved algorithm will look like this:

\[
\text{FORALL tasks simultaneously DO} \\
\text{FOR every time step DO} \\
\text{FOR every border grid point DO} \\
\quad \text{- Compute wavefield} \\
\quad \text{END} \\
\quad \text{- Send values of updated border grid points to each adjacent layer.} \\
\text{FOR every inner grid point DO} \\
\quad \text{- Compute wavefield} \\
\quad \text{END} \\
\quad \text{- Receive values of updated border grid points from each adjacent layer.} \\
\quad \text{- Update seismogram} \\
\text{END} \\
\text{END}
\]
with more emphasis on the efficiency of the communication pattern designed for the general problem of distributed implementation of finite differences. Specific hardware-dependent (pipelining, vectorization, etc.) or model parameter-dependent (moving grid) optimizations must be applied to this software when used in a production environment.

Regarding the performance of the message passing library, we used the freely available version of PVM for the same reason mentioned above. However, several optimized architecture-dependent implementations of PVM exist, and they should be used for production executions. A study about comparative performance of PVM versions on different hardware is given in Casanova (1991). Also the PVM group library was used to implement synchronization and communication aspects of the algorithm. This library provides facilities for handling groups of tasks and facilitates coding and debugging, although it also decreases the performance of the algorithm, especially when a slow network and a large number of machines are used. No load balancing was attempted at this point, mainly because the Pentium-based cluster has only a small degree of heterogeneity, while the IBM SP-2 cluster is totally homogeneous.

Performance results

Parallelization experiments using a one-dimensional decomposition were done to examine scaled speedup of the parallel algorithm, i.e. the speedup obtained using a scaled-size problem. In a scaled-size problem, the size of the problem (the total number of grid points in our case) increases as processors are added, thus keeping the amount of work done by each processor constant, independent of the number of processors added (Figure 4). In contrast, in fixed-size problems, as the number of processors increases the amount of work done by each processor decreases, making more difficult the task of evaluating the effects of the communication in the performance of the algorithm.

Our one-dimensional decomposition experiments were done with decomposition in the $z$ direction, although the same decomposition can be applied in the $x$ or $y$ directions, depending on the direction of the longest axis and/or specific constraints of the model. Subdomain sizes were fixed by the memory available in each processor. For the Pentium cluster every subdomain contained approximately 800,000 grid points, which in the sequential case (when running on one CPU) corresponds to a cubic grid with sides $93 \times 93 \times 93$, thus permitting one to scale the size of the complete grid to a 200 grid point cube when 10 processors are being used, as shown in Figure (4). Also a brick (parallelepiped) geometry was used, with the same 800,000 grid points in each subdomain, but distributed in such a way that the side of the brick where the decomposition is taking place (i.e. the $z$ direction) is twice as large as the other two sides (Figure (4)). For the IBM SP2 cluster, every subdomain contained approximately $4 \times 10^6$ grid points (a cubic grid with sides $159 \times 159 \times 159$ in the sequential case). Second-order and fourth-order finite difference schemes with cubic and brick model geometries each were used for simulations in each architecture. One to 8 processors were used for the experiments in both the Pentium network and the IBM SP2.

The wave propagation simulation parameters were held constant for all the experiments; in particular, holding the grid point spacing constant (for a scheme with a given order) allowed the physical dimension to increase in a given direction when the number of grid points increased in that direction. A homogeneous medium with acoustic wavespeed of 3,000 m/s was used for the simulations, and the wavefield propagation was simulated for 0.5 sec, with a time step of 0.67 msec (the computation time in this finite difference simulation is independent of the degree of heterogeneity of the medium, so our timing results will be equally applicable to models with arbitrary heterogeneity). A source wavelet with
a maximum frequency of 75 Mhz (which corresponds to a minimum wavelength of 40 m) was used to represent the source. The spatial grid point spacing for the second-order scheme was 4 m (in the three directions z, y and z), and 10 m for the fourth-order scheme. These parameters imply that the real size of the model we are working with when using 8 PC nodes is a 730 m cube when we use the second-order scheme, and a 1,800 m cube when we use the fourth-order scheme. These parameters satisfy the Courant-Friedrichs-Levy condition. Following are the speedup measurements we obtained by running the codes in several architectures.

### The Pentium network

Table (1) shows the scaled sizes for the models used in the experiments corresponding to a cubic geometry. Figures (5) and (6) show the results obtained for the simulations performed using the number of processors and sizes shown in Table (1), with a second-order and a fourth-order scheme, respectively.

<table>
<thead>
<tr>
<th>NODES</th>
<th>$N_x$</th>
<th>$N_y$</th>
<th>$N_z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>93</td>
<td>93</td>
<td>93</td>
</tr>
<tr>
<td>2</td>
<td>116</td>
<td>117</td>
<td>118</td>
</tr>
<tr>
<td>3</td>
<td>133</td>
<td>134</td>
<td>135</td>
</tr>
<tr>
<td>4</td>
<td>147</td>
<td>147</td>
<td>148</td>
</tr>
<tr>
<td>5</td>
<td>158</td>
<td>158</td>
<td>160</td>
</tr>
<tr>
<td>6</td>
<td>169</td>
<td>169</td>
<td>168</td>
</tr>
<tr>
<td>7</td>
<td>176</td>
<td>176</td>
<td>182</td>
</tr>
<tr>
<td>8</td>
<td>184</td>
<td>184</td>
<td>192</td>
</tr>
</tbody>
</table>

Table 1. Cubic geometry. Number of nodes and problem sizes (in grid points) for the experiments in the Pentium network.

![Figure 5. Cubic geometry with a second-order scheme. Measured wallclock times for the experiments using the Pentium network.](image)

![Figure 6. Cubic geometry with a fourth-order scheme. Measured wallclock times for the experiments using the Pentium network.](image)

where **sequential time** and **parallel time** correspond to the curves labeled **Sequential** and **Total**, respectively, in the figures. The sequential time shown in the figures was obtained using a 90 Mhz Pentium.

Notice that, as the size of the problem scales with the number of processors (keeping the computational load in every processor constant), the grid running in 8 CPU’s has eight times more grid points than the grid running in one CPU, although the elapsed time for this parallel simulation should be the same as the sequential time (the dashed line labeled **Sequential** in the figures) if the computation and communication overheads were negligible. This is not the case here, and we observe that the efficiency of the algorithm is decreased because of the time}

$$\text{parallel speedup} = \frac{\text{sequential time}}{\text{parallel time}}$$
spent in the extra computation and communication required by the distributed algorithm.

Table (2) shows the scaled sizes of the models with a brick geometry, and Figures (7) and (8), whose description is the same as the figures previously shown, reflect the results obtained for models with this kind of geometry.

The IBM SP-2

As this hardware has more memory than the PC’s used in the experiments shown in last section, a model with larger dimensions was used for the experiments here. The size of the model was scaled so that the percentage of the total available memory used in these simulations was the same as the one used in the Pentium network.

Table (3) and Table (4) show the scaled sizes for the models used in the experiments corresponding to a cubic and a brick geometry, respectively. Figures (9) and (10) show the results obtained for the simulations performed using the number of processors and sizes shown in Table (3), with a second-order and fourth-order scheme, respectively, while Figures (11) and (12) show the corresponding results using the data in Table (4).

Interpretation of results

Here are some of the interpretations of the results reported in this study:

- **Parallel efficiency.** Interprocessor communication overhead and redundant computations are the main limitations for obtaining high performances on workstations. The overhead increases with the number of processors used. Relatively high efficiencies are obtained in workstations when using 5 or fewer nodes, or when using high bandwidth networks.
<table>
<thead>
<tr>
<th>NODES</th>
<th>$N_x$</th>
<th>$N_y$</th>
<th>$N_z$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>126</td>
<td>126</td>
<td>252</td>
</tr>
<tr>
<td>2</td>
<td>159</td>
<td>158</td>
<td>318</td>
</tr>
<tr>
<td>3</td>
<td>181</td>
<td>182</td>
<td>363</td>
</tr>
<tr>
<td>4</td>
<td>201</td>
<td>201</td>
<td>396</td>
</tr>
<tr>
<td>5</td>
<td>216</td>
<td>217</td>
<td>425</td>
</tr>
<tr>
<td>6</td>
<td>229</td>
<td>229</td>
<td>456</td>
</tr>
<tr>
<td>7</td>
<td>242</td>
<td>243</td>
<td>476</td>
</tr>
<tr>
<td>8</td>
<td>252</td>
<td>252</td>
<td>504</td>
</tr>
</tbody>
</table>

Table 4. Brick geometry. Number of nodes and problem sizes (in grid points) for the experiments in the IBM SP2.

Figure 9. Cubic geometry with a second-order scheme. Measured wallclock times for the experiments using the IBM SP2.

Figure 10. Cubic geometry with a fourth-order scheme. Measured wallclock times for the experiments using the IBM SP2.

Figure 11. Brick geometry with a second-order scheme. Measured wallclock times for the experiments using the IBM SP2.

Figure 12. Brick geometry with a fourth-order scheme. Measured wallclock times for the experiments using the IBM SP2.

- **Optimization.** As we mentioned before, the algorithm was implemented avoiding hardware-dependent and model-dependent optimizations in order to get a better basis of comparison between the experiments in different hardware platforms. Thus the benchmarks presented should be regarded as lower bounds on efficiency. We have observed that optimization techniques available in high-end workstations, such as pipelining and cache-oriented programming, can greatly improve the speed of the code. The communication overhead can also be reduced using optimized versions of the message passing libraries, such as the enhanced PVMe library (IBM, 1995) for IBM, or forthcoming freely available versions of PVM, which will include improved communication capabilities.
Influence of model geometry and the order of scheme on the efficiency. Model geometries play an important role in the performance of this algorithm. The cubic geometry seems to generate the lower values of efficiency because the low volume/area ratio of each subdomain generates a low computation/communication ratio. This suggests that a higher order domain decomposition (two or three-dimensional) can be more efficient with this kind of geometry, as discussed in a previous section. The order of the finite difference scheme also affects the efficiency because of the increment in communication required by higher order methods. This is evident if we compare the high efficiencies shown in Figure (7), which corresponds to the brick geometry using a second order scheme, with the low efficiencies shown in Figure (6), corresponding to a cubic geometry using a fourth order scheme. However, using a high order scheme allows the grid to represent models with larger physical dimensions, and so the efficiency of the fourth order scheme in this case can be higher than the efficiency of the second order scheme regarding the physical dimensions of the model.

Size of the model. Considering that computation and communication are partially overlapped in this algorithm, as discussed before, we expect better efficiencies when the size of the subdomains increases. Larger subdomains will increase the amount of computation to update them and also the opportunity for the communication to occur concurrently with the computation. The only limitation for increasing the size of each subdomain is the amount of RAM available in each workstation. As more memory becomes available, larger problems can be handled.

Conclusions and future work

We have developed a distributed 3D acoustic finite difference modeling algorithm using a domain decomposition approach, and implemented this algorithm in ANSIG via the freely available PVM message passing library. We have presented benchmarks of this code on two different architectures, a cluster of network connected Linux PC’s, and an IBM SP2 cluster. Using a scaled problem (to keep the size of each machine’s sub-problem fixed) we have observed parallel efficiencies of from 46-76% on a cluster of 8 workstations (depending on the geometry of the model and the order of the finite difference used) and up to 94% on an 8 processor SP2. With this algorithm we are able to simulate 3D acoustic wave propagation for a model of $6.4 \times 10^5$ gridpoints on a network of 8 PC’s.

In order to achieve a large measure of portability we have taken advantage of no hardware-dependent optimization. In particular, we believe that considerable efficiencies could be achieved by taking advantage of pipelining and cache-oriented programming. We have limited the size of the subdomains to fit into the available memory of our workstations. As the resources of our workstation network increase (memory and bandwidth) we will be able to take advantage of these resources by adjusting the size of the subdomains. Our next goal is to extend this algorithm to the elastic wave equation, as well as other geophysical applications.

Acknowledgements

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, the Shell Foundation, the Army Research Office, and the National Science Foundation under grant DMS-9506603. We would especially like to thank IBM for access to their SP2 cluster at Poughkeepsie, NY.

References


Analytic study of the effective parameters for determination of the NMO velocity function in transversely isotropic media

Jack K. Cohen
Professor of Mathematics, Center for Wave Phenomena
Colorado School of Mines

ABSTRACT
In their studies of transversely isotropic media, Alkhalifah and Tsvankin have observed that, to high numerical accuracy, the normal moveout velocity for dipping reflectors as a function of ray parameter depends mainly on just two parameters, each of which can be determined from surface P-wave observations. They exploit this result to develop a processing sequence that culminates in a time migration algorithm that takes account of transverse isotropy.

Here, the two parameter Alkhalifah-Tsvankin numerical result is examined analytically. It is found that although there is, as these authors have already observed, some dependence on the remaining parameters of the problem, this dependence is weak in the practically important regime of weak to moderately strong transverse isotropy. The same is shown to be true for even strong transverse isotropy, in the special case when the phase angle—or equivalently—the ray parameter is small. In each of these regimes, an analytic solution is derived for the anisotropy parameter $\eta$, which has been shown to be important for time-related imaging.

Alkhalifah and Tsvankin have also established analytically that the dependence of the moveout velocity function on only two parameters is exactly true in the elliptic limit. Here, it is shown that this is also the case in the limit where the on-axis S-wave phase velocity is treated as negligible compared with the on-axis P-wave phase velocity.

The analytic results derived here are based on obtaining explicit formulas for both the P-wave phase velocity and normal moveout velocity in terms of the ray parameter. Various results of Tsvankin, Alkhalifah, and others are rederived and, in some cases, extended, in a uniform manner using these formulas. In particular, second-order expansions in the anisotropy parameters are derived for both the P-wave phase velocity function and normal moveout velocity function. Likewise expansions in powers of the ray parameter are derived for both these functions. These expansions are checked against the corresponding exact functions for several choices of the anisotropy parameters.

Key words: transverse isotropy, NMO, DMO, moveout function, analytic approximations, series expansions, weak anisotropy, moderate anisotropy, strong anisotropy

Introduction

Alkhalifah and Tsvankin (1995) have developed a seismic processing sequence based on P-wave surface observable quantities that culminates in a time migration code that takes account of vertical transverse isotropy. Such media are characterized by five parameters, here taken to be the Thomsen parameters (1986): the on-axis P- and S-phase
velocities, $c_p$ and $c_s$, and the anisotropy parameters, $\gamma$, $\delta$, and $\epsilon$. Focusing exclusively on $P$-wave data removes $\gamma$ from consideration and reduces the number of relevant parameters to four. Remarkably, Alkhalifah and Tsivankin show that with only two well selected parameters and using the ray parameter (instead of the phase angle) as running variable, it is possible to determine the normal moveout velocity function to good numerical accuracy. Moreover, the two parameters selected can be determined from $P$-wave surface data alone. With the moveout function determined, they develop a time-domain migration algorithm which has performed well on field data exhibiting transverse isotropy.

The two parameters selected by Alkhalifah and Tsivankin are the zero-dip moveout velocity, $V_{\text{zero}}(0)$, and a combination of the Thomsen anisotropy parameters denoted by $\eta$. These authors discuss how to determine these parameters from the $P$-wave surface data. The definition of these quantities in terms of the Thomsen parameters is given below.

When the Alkhalifah-Tsivankin principal parameters, $V_{\text{zero}}(0)$ and $\eta$, are adopted, the remaining two independent parameters can be taken to be the $c_s/c_p$ ratio and one of the Thomsen $P$-wave anisotropy parameters, say $\delta$. Alkhalifah and Tsivankin (1995) observe that, in general, the moveout function, expressed as a function of ray parameter, does exhibit some residual dependence on the two non-principal parameters or "residual parameters." However, by extensive numerical tests, they find that just choosing reasonable values for these non-observable parameters provides a numerical normal moveout function that is highly accurate even when measured against moveout functions constructed with significantly different values of the residual parameters.

There are special cases in which the two principal parameters suffice to exactly determine the moveout function. For example, in the elliptic limit, $\epsilon/\delta \rightarrow 1$, Alkhalifah and Tsivankin have shown that the moveout function expressed in terms of the ray parameter depends only on $V_{\text{zero}}(0)$. While the elliptic case is a common analytic simplification, it is of limited practical application. Here, another such theoretical limit is introduced: $c_s/c_p \rightarrow 0$, and it is shown that once again, the resulting moveout function depends only on the two principal parameters.

More important for treating field data is the weak isotropy limit. This limit is characterized by dropping all quadratic and higher-order dependence on $\delta$ and $\epsilon$. In this limit, Alkhalifah and Tsivankin show that the moveout function expressed in terms of the ray parameter again depends only on the two principal parameters. Here, this expansion is extended to the case where quadratic anisotropy parameters are retained, but cubic terms and higher-order terms are dropped. This regime is designated as "moderate transverse isotropy" and, of course, subsumes the weak transverse isotropy regime. In the moderate regime, it is shown that there is dependence on the residual parameters. Thus, the numerical evidence of some residual dependence is quantified analytically. However, the critical result is that the residual dependence is confined to terms that are (1) relatively small, and (2) vary smoothly with changes in the residual parameters. Thus, the two-parameter Alkhalifah-Tsivankin methodology is given analytic support in the most common physical situations. That is, these results support the notion that supplying crude reasonable values for the two-parameters that cannot be determined from surface $P$-wave data is sufficient to allow accurate determination of the moveout function and to justify the subsequent use of the time-migration algorithm based on it.

Analytic support for the two parameter methodology is also given for strong anisotropy in the case when the phase angle—or equivalently—the ray parameter is small. Here again, the residual parameter dependence is shown to be weak and smoothly varying, thus justifying the use of nominal values.

Finally, in both the moderate-anisotropy and the small-angle regimes, analytic approximations are derived for the anisotropy parameter $\eta$.

The analytic results derived here are based on obtaining explicit formulas for both the $P$-wave phase velocity and normal moveout velocity in terms of the ray parameter. Various results of Tsivankin, Alkhalifah, and others are rederived and, in some cases, extended, in a uniform manner using these formulas. In particular, second-order expansions in the anisotropy parameters are derived for both the $P$-wave phase velocity function and normal moveout velocity function. Likewise expansions in powers of the ray parameter are derived for both these functions. These expansions are checked against the corresponding exact functions for several choices of the anisotropy parameters.

Throughout this study, Mathematica was used extensively to derive and check results. In particular, the Mathematica package, Thomsen.m (Cohen, 1995), was used to convert equation (3) to Thomsen notation, in equation (3), and to obtain various results in the limit of weak transverse isotropy. Similarly, the use of Mathematica facilitated computing symbolic derivatives, series expansions, etc.
Phase Velocity as a Function of Ray Parameter

Begin with the formula for the P-wave phase velocity in a homogeneous, transversely isotropic medium with vertical axis of symmetry in terms of the phase angle, \( \theta \) (Tsvekin, 1994), equation (6):

\[
2pV^2(\theta) = (C_{11} + C_{44})\sin^2 \theta + (C_{33} + C_{44})\cos^2 \theta + [(C_{11} - C_{44})\sin^2 \theta - (C_{33} - C_{44})\cos^2 \theta]^2 \\
+ 4(C_{13} - C_{44})^2 \sin^2 \theta \cos^2 \theta \right)^{1/2}. 
\]  

(1)

Use the substitutions

\[
p = \sin \theta / V(\theta), \quad m = \cos \theta / V(\theta),
\]

(2)
to rewrite equation (1) as an equation for the slowness surface:

\[
2p = (C_{11} + C_{44})p^2 + (C_{33} + C_{44})m^2 + [(C_{11} - C_{44})p^2 - (C_{33} - C_{44})m^2]^2 \\
+ 4(C_{13} - C_{44})^2 p^2 m^2 \right)^{1/2}.
\]

To obtain a formula for \( V(p) \), follow the recipe given in Appendix A, (Alkhalifah & Tsvekin, 1995). Convert equation (3) to Thomsen notation and introduce the on-axis \( P \) and \( S \) velocities, \( c_P \) and \( c_S \). As a convenient parameter to characterize the \( c_S / c_P \) ratio define

\[
f = 1 - \frac{c_S}{c_P}.
\]

(4)

In this notation, the Thomsen-notation form of the slowness surface is

\[
\frac{2}{c_P^2} = (2 - f)m^2 + (2 + 2\epsilon - f)p^2 + \sqrt{4f(2\delta + f)m^2 p^2 + 4(p^2 - m^2) + 4c_P^2 p^2}.
\]

(5)

Next, solve equation (5) for \( m^2 \) and use equation (2) to write

\[
V(p) = \frac{1}{\sqrt{p^2 + m^2}}.
\]

(6)

This program produces \( V^2(p) \) in the form

\[
V^2(p) = c_P^2 \frac{A + \sqrt{B}}{2C},
\]

(7)

where,

\[
A = 2 - f - 2(\epsilon - f)\delta z, \\
B = f^2 - 4f(\epsilon - (2 - f)\delta z)z + 4[2f(1 - f)(\epsilon - \delta) + (\epsilon - f)^2]z^2, \\
C = 1 - 2\epsilon z - 2f(\epsilon - \delta)z^2.
\]

Here and below, the phase velocity is expressed in terms of the dimensionless parameter \( z \),

\[
z = \frac{c_P p^2}{V^2}.
\]

(9)

Before turning to some important special cases, observe the following consequences of the definitions in equations (2):

\[
\sin \theta = pV(p), \quad \cos \theta = mV(p) = \sqrt{1 - (pV)^2}, \\
p^2 + m^2 = \frac{1}{V^2}.
\]

(10)

These equations are used to translate between representations in phase angle and representations in the ray parameter.

Zero dip

Observe that for \( p = 0 \) (\( \zeta = 0 \)), the constants \( A, B, C \) reduce to

\[
A = 2 - f, \\
B = f^2, \\
C = 1,
\]

(11)

so that,

\[
V^2(0) = c_P^2 \frac{2 - f + f^2}{2} = c_P^2,
\]

(12)

leading to the expected result,

\[
V(0) = c_P.
\]

(13)

Elliptic anisotropy

In the elliptic case, \( \delta = \epsilon \), and the constants \( A, B, C \) reduce to

\[
A = 2 - f - 2(1 - f)\delta z, \\
B = (f + 2(1 - f)\delta z)^2, \\
C = 1 - 2\delta z,
\]

(14)

so that,

\[
V^2(p) = c_P^2 \frac{2}{2(1 - 2\delta z)} = \frac{c_P^2}{1 - 2\delta z}, \quad (\delta = \epsilon).
\]

(15)

In (Tsvekin, 1995), the equivalent elliptic limit value of \( V^2 \) in terms of the phase angle, \( \theta \), is given as

\[
V^2(\theta) = V_0^2 \cos^2 \theta + \frac{V^2_0 \sin^2 \theta}{V^2_0},
\]

(16)

where \( V_0 = V(0) \) and \( V_0 = V(\pi/2) \) in the phase-angle form of \( V \) given by equation (1). Likewise equation (15) is equivalent to

\[
V(\theta) = c_P \sqrt{1 + 2\delta \sin^2 \theta}, \quad (\delta = \epsilon)
\]

(17)

which is given in (Alkhalifah & Tsvekin, 1995).
Zero shear velocity

Another case in which the P-wave phase velocity takes a simple form is in the limit \( f = 1 \) (\( c_\delta = 0 \)). Here, expressions (8) reduce to

\[
\begin{align*}
A &= 1 - 2(\epsilon - \delta)z, \\
B &= [1 - 2(\epsilon - \delta)z]^2, \\
C &= 1 - 2\epsilon z - 2(\epsilon - \delta)z^2,
\end{align*}
\]  

(18)

so that,

\[
V^2(p) = c_p^2 \frac{1 - 2(\epsilon - \delta)z}{1 - 2\epsilon z - 2(\epsilon - \delta)z^2}, \quad (f = 1).
\]  

(19)

Weak and moderate transverse isotropy

In its strict interpretation, the weak transverse isotropy limit implies retaining only linear terms in \( \delta \) and \( \epsilon \). In this case, equation (7) gives

\[
V(p) \approx c_p \left[ 1 + \delta z + (\epsilon - \delta)z^2 \right].
\]  

(20)

Thomsen's (1986) equivalent expression in terms of phase angle is

\[
V(\theta) \approx c_p \left[ 1 + (\delta \cos^2 \theta + \epsilon \sin^2 \theta) \sin^2 \theta \right].
\]  

(21)

Retaining up to quadratic terms in \( \delta \) and \( \epsilon \) in equation (7) gives the "moderate transverse isotropy" expansion,

\[
V(p) \approx c_p \left[ 1 + \delta z + \left( \epsilon - \delta \right)(1 + 2\delta/f) + 3\delta^2/2 \right] z^2
\]

\[
+ \left( \epsilon - \delta \right)(5\delta + 2(\epsilon - 2\delta)/f)z^3
\]

\[
+ \left( \epsilon - \delta \right)^2(7/2 - 2/f)z^4.
\]  

(22)

The moderate transverse isotropy regime, in which this quadratically augmented expansion is valid, subsumes the ordinary weak transverse isotropic regime.

The left panels of Figure 1 compare the exact phase velocity to its weak and moderate approximations for the specific values \( f = 0.75 \), \( c_p = 1.0 \) and the values of the anisotropy parameters shown on the individual panels. Observe that, although the weak and moderate approximations are nominally valid on the basis of the size of \( \delta \) and \( \epsilon \), once these parameters are specified, the approximations deteriorate as \( z \) (or equivalently, the ray parameter, or the phase angle) increases. The complicated relation, \( z = c_p^2 p^2 = c_p^2 \sin^2 \theta/V^2(z) \), between \( z \) and the phase angle \( \theta \) precludes solving for \( z \) explicitly, so the corresponding phase angles have been shown as functions of \( z \) in the right panels of Figure 1. The figure shows that for moderately strong to strong anisotropy and, adopting an error tolerance of 2%, the validity of the weak regime typically extends at least up to 30° and that of the moderate regime to at least 45°.

The moderate-isotropy expansion for the square of the phase velocity given in equation (22) is

\[
V^2(p) \approx c_p^2 \left[ 1 + 2\delta z + (\epsilon - \delta)(1 + 2\delta/f) + 2\delta^2 \right] z^2
\]

\[
+ 4(\epsilon - \delta)(3\delta + (\epsilon - 2\delta)/f)z^3
\]

\[
+ 4(\epsilon - \delta)^2(2 - 1/f)z^4.
\]  

(23)

The equivalent expression in phase angle was given in Tsvankin (1994), and an argument based on it was made for the weak dependence on \( f \) (or \( c_\delta \)) of kinematic P-wave signatures in homogeneous, transversely isotropic media.

Small ray parameter

Equation (7) yields the small-\( p \) expansion,

\[
V(p) = c_p \left[ 1 + \delta z + (\epsilon - \delta)(1 + 2\delta/f) + 3\delta^2/2 \right] z^2
\]

\[
+ (\epsilon - \delta)(1 + 2\delta/f)(5\delta + 2(\epsilon - 2\delta)/f)
\]

\[
+ 5\delta^2/2 \right] z^3
\]

\[
+ (\epsilon - \delta)(1 + 2\delta/f)(7(\epsilon - \delta + 5\delta^2)/2)
\]

\[
- (2(\epsilon - \delta) - 7\delta(3\epsilon - 5\delta))/f
\]

\[
- 4(5\delta(\epsilon - \delta) - \epsilon^2)/f^2 + 35\delta^4/8 \right] z^4
\]

\[
+ O(z^5).
\]  

(24)

Remark: This expansion shows that the correct interpretation of "small \( p \)" is that \( z = (c_p p)^2 \ll 1 \). Since \( V \approx c_p \) for \( z \ll 1 \), equation (10) shows that this inequality for \( z \) can be written as \( \sin^2 \theta \ll 1 \). Thus, in dimensionless terms, "small \( p \)" represents a small phase angle \( \theta \).

Figure 2 compares the exact phase velocity to the successive small-\( z \) approximations of orders 1, 2, 3, and 4 for the same parameter values as in the previous section. Again adopting an error tolerance of 2%, the validity regime of the four term expansion in equation (24) typically extends up to 45° for moderate to strong anisotropy.

Normal Moveout Velocity as a Function of Ray Parameter

The derivation of the normal moveout velocity as a function of ray parameter follows from the corresponding derivation of its expression in terms of the ray angle given in Tsvankin (1995).

First, use equations (10) to find the following relation for \( dp/d\theta \):

\[
mV = \cos \theta = \frac{d \sin \theta}{d \theta} = (pV)' \frac{dp}{d\theta},
\]  

(25)

where the prime notation is used for \( p \)-differentiation. Thus,
Figure 1. For the specific values \( f = 0.75, \, c_p = 1.0, \) and the values of the anisotropy parameters shown, the left panels are plots of the exact phase velocity (dark), its moderate approximate (medium), and its weak approximate (light), as functions of \( z \). The right panels show the corresponding phase angles as functions of \( z \).

\[
\frac{dV}{d\theta} = V' \frac{dp}{d\theta} = \frac{mVV'}{(pV)}.
\]

(26)

and,

\[
\frac{d^2V}{d\theta^2} = \frac{mV}{(pV)} \left( \frac{mVV'}{(pV)} \right)' .
\]

(27)

Using these results in Tsvankin's equation (9), and simplifying using equations (10), yields the explicit formula (28) for the normal moveout velocity as a function of ray parameter.

In the Mathematica implementation of this result, a layer of square roots was avoided by writing \( V_{nmo}^2 \) in terms of \( W \equiv V^2 \). Also the dimensionless variable \( z \), defined in equation (9), was used resulting in equation (29).
Figure 2. For the specific values $f = 0.75, c_p = 1.0$, and the values of the anisotropy parameters shown, the left panels are plots of the exact phase velocity (dark), and its 4, 3, 2, and 1 term approximations shown successively lighter as functions of $z$. The right panels show the corresponding phase angles as functions of $z$.

\[
V_{nmo}^2(p) = \frac{(1 - p^2 V^2)VV'' + (3p^2 V^2 - 2)V'' + 2pV^3V'' + V^4}{(1 - p^2 V^2)V(pV)'}, \quad \text{where} \quad V' = \frac{dV}{dp}.
\]

\[
V_{nmo}^2(z) = c_p^2 \frac{2z(c_p^2 - zW)W\dot{W} + (c_p^2 + zW)W\ddot{W} - (3c_p^2 - 4zW)\dot{W}^2 + \dot{W}^3}{(c_p^2 - zW)W(zW)'}, \quad \text{where} \quad W = \frac{dW}{dz}.
\]
The Alkhalfah-Tsvankin Methodology

Before examining the simplifications of $V_{\text{nmo}}^2(p)$ corresponding to the special cases for $V$ discussed earlier, it is helpful to have in mind the methodology of Alkhalfah and Tsvankin (1995). These authors introduce the parameter,

$$\eta = \frac{\epsilon - \delta}{1 + 2\delta}, \quad (30)$$

and establish, both numerically, and by the successful processing of field seismic data, that $V_{\text{nmo}}^2(p)$ depends mainly on the principal parameters, $\eta$ and $V_{\text{nmo}}^2(0)$. They exploit this reduction to surface-observable parameters to develop a time-domain seismic processing sequence including, in particular, time-migration, that takes transverse anisotropy into account.

The only other parameters that $V_{\text{nmo}}^2(p)$ can depend on are $f = 1 - (c_s/c_P^2)$ and $\delta$ (or an equivalent pair such as the $c_s/c_P$ ratio and $\delta$). Thus, the Alkhalfah-Tsvankin methodology is tantamount to the assertion that $V_{\text{nmo}}^2(p)$ depends only weakly on the residual parameters, $f$ and $\delta$. Because of this, they can use nominal values for these parameters with little deleterious effect on the results. Typical choices are $f = 3/4$ (i.e., $c_P/c_S = 2$), and $\delta = 0$.

**Zero dip**

The two leading terms of equation (24) yield the approximations,

$$V' = c_P + c_P^3 \delta p^2 + O(p^4),$$

$$V'' = 2c_P^3 \delta p + O(p^3),$$

$$V''' = 2c_P^5 \delta + O(p^2). \quad (31)$$

Inserting these results into the general NMO equation (28) gives

$$V_{\text{nmo}}^2(p) = c_P^2 (1 + 2\delta) + O(p^2). \quad (32)$$

Setting $p = 0$ in this result establishes

$$V_{\text{nmo}}^2(0) = c_P^2 \sqrt{1 + 2\delta}. \quad (33)$$

Along with equation (30) for $\eta$, this expression completes the explicit definition of the two principal parameters in the Alkhalfah-Tsvankin methodology discussed above.

Trivially, in this first order small-p limit, $V_{\text{nmo}}(p) = V_{\text{nmo}}(0) + O(p^2)$, so to this order, $V_{\text{nmo}}(p)$ depends only on (one of) the Alkhalfah-Tsvankin principal parameters.

**Elliptic anisotropy**

Inserting the elliptic $P$-wave phase velocity as a function of ray parameter given in equation (15) into the equation (28) gives

$$V_{\text{nmo}}^2(p) = \frac{c_P^2 (1 + 2\delta)}{1 - (1 + 2\delta)c_P^2 p^2}. \quad (34)$$

In terms of $V_{\text{nmo}}(0)$, as given in equation (33), the elliptic moveout result is expressed by

$$V_{\text{nmo}}(p) = \frac{V_{\text{nmo}}(0)}{\sqrt{1 - p^2 V_{\text{nmo}}^2(0)}}, \quad (\delta = \epsilon). \quad (35)$$

in agreement with Alkhalfah and Tsvankin (1995). Note that in the elliptic case, the Alkhalfah-Tsvankin methodology introduces no error, since $V_{\text{nmo}}(p)$ depends only on (one of) the Alkhalfah-Tsvankin principal parameters.

In expressions for the moveout function, it is convenient to introduce the dimensionless quantity,

$$y = V_{\text{nmo}}^2(0)p^2, \quad (36)$$

and the notation,

$$V_{\text{nmo}}(y) \equiv \frac{V_{\text{nmo}}(0)}{\sqrt{1 - y}} = \frac{V_{\text{nmo}}(0)}{\sqrt{1 - y}} \quad (37)$$

for the elliptic result.

**Zero shear velocity**

Using equation (19) in equation (29) gives the result (38) for the normal moveout velocity function when the shear velocity vanishes (i.e., when $f = 1$). To express equation (38) in terms of the Alkhalfah-Tsvankin parameters, use equations (9), (30) and (36) to eliminate $z$, $\epsilon$, and $c_P$ via the equations,

$$z = \frac{y}{1 + 2\delta},$$

$$\epsilon = \delta + \eta (1 + 2\delta),$$

$$c_P = \frac{V_{\text{nmo}}(0)}{\sqrt{1 + 2\delta}.} \quad (39)$$

obtaining equation (40). Thus, the Alkhalfah-Tsvankin methodology is exactly verified in the special case of $f = 1$. In the next section, we turn from theoretically interesting limits to an approximation of practical interest.

**Weak and moderate transverse isotropy**

Using equation (20) in equation (29) gives

$$V_{\text{nmo}}^2(p) = \frac{c_P^2}{1 - z} \left( 1 + \frac{2(\delta + (\epsilon - \delta)(6 - 9z + 4z^2))}{1 - z} \right). \quad (41)$$

Again use equations (39) to introduce the Alkhalfah-Tsvankin principal parameters, converting equation (41) into

$$V_{\text{nmo}}^2(p) = V_{\text{nmo}}^2(y) \left( 1 + 2\eta F(y) \right), \quad (42)$$

with

$$F(y) = \frac{y(6 - 9y + 4y^2)}{1 - y}. \quad (43)$$
\[ V_{\text{nmo}}^2(p) = \frac{c_p^2(1 + 2\delta)(1 + 4(\epsilon - \delta)z - 6(\epsilon - \delta)(1 + 2\epsilon)z^2)}{(1 - (1 + 2\epsilon)z^2)(1 - (1 + 2\epsilon)z^2)(1 - (1 + 2\epsilon)z^2)}. \]  

(38)

\[ V_{\text{nmo}}^2(p) = V_{\text{nmo}}^2(0) \frac{1 + 2y(2 - 3y)\eta - 12y^2\eta^2}{(1 - 2y\eta)(1 - y - 2y\eta)(1 - 2y(2 - y)\eta + 4y^2\eta^2)}. \]  

(40)

in agreement with Alkhalihah and Ts GSM (1993). Equation (37) has been used to point out that the elliptic result, \( V_{\text{cl}}^2(y) \), is a factor of the weak-limit result. By its definition (36), the quantity \( y \) only involves a principal parameter and the ray parameter, so equation (42) shows that there is no residual parameter dependence at all in the weak limit.

Notice that the weak expansion is singular at \( y = 1 \). Indeed, this singularity is already present in the elliptic moveout function (37) (and so, already present even in the isotropic limit). Physically, this \( y \)-value corresponds to a steep dip related to the natural infinity of the moveout function for a vertical reflector. Mathematically, successive terms in the power series expansion in the anisotropy parameters are successively more singular at this critical value, thus invalidating this expansion when \( y \) approaches 1. Thus, applications of the weak anisotropy expansion must be limited by taking account both of the magnitudes of the anisotropy parameters and the size of \( y \) (which can be related to the size of the phase angle for given anisotropy parameters).

Equation (42) may be used to obtain the weak-limit estimate,

\[ \eta \approx \eta_0 = \frac{U}{2F(y)}, \quad U = \frac{V_{\text{nmo}}^2(p)}{V_{\text{cl}}^2(y)} - 1. \]  

(44)

At the next order in the anisotropy parameters (i.e., in the moderate-transverse isotropic regime),

\[ V_{\text{nmo}}^2(p) = V_{\text{cl}}^2(y) \left[ 1 + 2\eta F(y) \right] + 24\left( \frac{1}{f} - 1 \right)\eta^6 P_4(y) + \frac{12}{f} \eta^2 P_2(y) + 4\eta^2 Q(y), \]  

(45)

where \( F \) is defined in equation (43) and,

\[ P_0(y) = y(1 - 2y)(1 - y), \]

\[ P_2(y) = y^2(5 - 8y)(1 - y), \]

\[ Q(y) = \frac{y^2(26 - 68y + 63y^2 - 20y^3)}{(1 - y)^2}. \]  

(46)

The left panels of Figure 3 compare the exact moveout function to its weak and moderate approximations for the specific values \( f = 0.75 \), and \( c_p = 1.0 \). The resulting value of \( V_{\text{nmo}}(0) \) is 1.1. Just as in the case of the phase velocity, the weak and moderate approximations of the moveout function deteriorate as the phase angle (here monitored indirectly by \( y \)) increases and are valid for about the same phase angle ranges, as can be verified from the right panels of Figure 3.

Observe, that in the elliptic limit, all terms in equation (45) except the first vanish because of the factor of \( \eta \) multiplying them. Hence, consistency with the elliptic case is immediate. Thus, the elliptic limit and the isotropic limit are the same except for the implicit \( \delta \) in \( V_{\text{nmo}}(0) \).

The higher order terms introduce dependence on the residual parameters, \( \delta \) and \( f \). However, in addition to the fact that these terms are often small because they are quadratic in the anisotropy parameters, their importance is further mitigated both by their small size relative to the other terms and by their smooth dependence on the residual parameters. Since this is important analytic support for the Alkhalihah-Ts GSM methodology, these matters are now described in detail.

There are two quadratic-order terms involving the residual parameters. In the Alkhalihah-Ts GSM methodology, since these parameters cannot be determined from the P-wave surface seismic data they are just set to reasonable values. To see why this is justified, note first that the two terms depending on the residual parameters are small relative to the principal parameter terms. Indeed, the \( \eta \delta \) term is multiplied by \( 1/f - 1 \), which lies in the interval \([1/4, 2/3]\) when \( f \) is in the typical interval \([0.6, 0.8]\). The polynomial, \( P_4(y) \), also multiplying this term has absolute maximum value less than 0.1 on the interval \( 0 \leq y \leq 1 \). Thus, in the regime of practical parameter values, the coefficient of the \( \eta \delta \) term is less than 1/15 \( \approx 0.067 \) in absolute value. Similarly, the higher order \( \eta^2 \) term that is multiplied by \( 1/f \) is also multiplied by the polynomial \( P_2(y) \) which is less than 0.2 on \( 0 \leq y \leq 1 \), so the overall coefficient of this term is less than 1/3 in absolute value.

Furthermore, when the anisotropy is strong enough that the quadratic-order terms do influence the numerics, both of the terms just considered will be masked by the final \( \eta^2 \) contribution whose rational function coefficient, \( Q(y) \), grows rapidly as \( y \) increases from 0. Note that this last, and dominant, higher-order contribution depends only on the Alkhalihah-Ts GSM principal parameters, \( V_{\text{nmo}}(0) \) and \( \eta \).
Figure 3. For the specific values \( f = 0.75, \phi_p = 1.0 \), and the values of the anisotropy parameters shown, the left panels are plots of the exact moveout velocity (dark), its moderate approximate (medium), and its weak approximate (light), as functions of \( z \). The right panels show the corresponding phase angles as functions of \( z \).

Figure 4 illustrates the previous comments. Notice that for small \( y \)-values, all three terms coincide with the linear approximation, while for large \( y \)-values, the residual quadratic terms are swamped by the rapid growth of the principal parameter term (however, as previously noted, the whole expansion breaks down when \( y \) gets too close to 1). In the middle range of \( y \)-values, there are, indeed, visible deviations due to the residual dependencies for larger values of \( \eta \), but these deviations are small for practical values of the parameters.

Thus, even if these terms were neglected totally, their contribution would often be negligible. But in the Alkhalfah-Tsvankin methodology, they are not ignored, but rather implemented by using reasonable values for
Figure 4. With the typical parameter value $f = 0.75$ and the values of $\eta$ and $\delta$ and computed value of $\epsilon$ shown in the panels, each panel depicts $V_{\text{res}}(y)$ normalized by $V_{\text{res}}(y)$ plotted against $y = V_{\text{res}}(0)/y^2$ for the linear term alone (light), the linear term plus the portion of the quadratic term that depends only on the Alkilafiah-Tsvankin principal parameters (medium), and linear term plus full quadratic term (dark). The y-range in each plot corresponds to phase angles up to $60^\circ$.

the unknown residual parameters. If the errors made in estimating these parameters are denoted by $\Delta \delta$ and $\Delta f$, then the absolute error is $V_{\text{res}}(y)$ times

$$24 \left( \frac{1}{f} - 1 \right) \eta P_1(y) \Delta \delta - \frac{24}{f^2} \eta P_1(y) \Delta f - \frac{12}{f^2} \eta^2 P_2(y) \Delta f.$$

With the typical choices, $f = 3/4$ and $\delta = 0$, and noting the bounds on $P_1$ and $P_2$, this error factor is less than

$$0.8\eta | \Delta \delta | + 17.1\eta^2 | \Delta f |.$$
in absolute value. Considering that typically, $|\Delta \delta| < 0.2$ and $|\Delta f| < 0.1$, it seems that the error due to setting the residual parameters is rarely to be noticed.

In summary, the higher-order expansion supports the Alkhaliwash-Tsvanikin approach of using “best guess” values for the unobservable residual parameters $\delta$ and $f$ in the important regime of moderate anisotropy.

On the theoretical side, observe that if $f$ is set to one, then the $\delta$ contributions drop out completely along with the $f$ contributions; the residual parameter dependence is entirely removed. This is in agreement with the general result in equation (40), and expanding that result to second order in the anisotropy parameters gives the $f = 1$ limit of the present result.

The quadratic equation (45) can be used to improve the estimate for $\eta = \eta_0$ given in equation (44). We seek the solution of this quadratic that is close to $\eta_0$. Using only the terms free of dependence on the residual parameters, this solution is given by

$$\eta = \frac{2\eta_0}{1 + \sqrt{1 - 4\nu}} \approx \eta_0(1 + \nu), \quad \nu = \frac{-UQ}{P^2}, \quad (47)$$

where the final approximation is valid under the assumption that $\nu << 1$, and $U$ is defined in equation (44). In most cases, it is of some numerical benefit to include the $P_2$ term defined in equation (46). To do this requires only replacing $Q$ by $Q + \frac{1}{2} P_2$ in (47) and making a nominal choice of $f$. Similarly, inclusion of $P_2$ would be accomplished by replacing $F$ by $F + \frac{12\delta}{f} P_2$ in (44) and (47). However, the $P_2$ term is almost always negligible (and is altogether absent, if one makes the usual nominal choice $\delta = 0$). Numerical experiments indicate that this second-order approximation for $\eta$ affords an accurate starting value for obtaining a high-precision answer with only a few Newton iterations.

**Small ray parameter**

In the small-$p$ limit, it is possible to get analytical results even for strong anisotropy. At first, use just the first three terms of equation (24) in equation (28) to get

$$V_{nmo}^2(p) = c_2^2 \left[ (1 + 2\delta + (f - 8\delta f + 4\delta^2 f + 12\delta f - 24\delta^2 + 24\delta e) \frac{(cp)^2}{f} + O(cp^4) \right].$$

Expressing the result in terms of the Alkhaliwash-Tsvanikin principal parameters gives

$$V_{nmo}^2(p) = V_{nmo}^2(0) \left[ \left( 1 + \frac{1 + 2\eta \left( \frac{1 + 2\delta}{1 + 2\delta} \right) \left( \frac{1 + 2\eta}{1 + 2\eta} \right) \frac{1 + 2\delta}{1 + 2\delta} \right) + O(y^3) \right].$$

Numerical tests show that using this small-$p$ series for a dip of $15^\circ$ incurs about a 2% error in estimating $V_{nmo}^2(p)$ for typical values of $f$ and $\delta$.

Equation (49) gives analytic support to the Alkhaliwash-Tsvanikin methodology, since the only dependence on the residual parameters occurs in the ratio

$$g = \frac{1 + 2\delta/f}{1 + 2\delta}, \quad (50)$$

that multiplies $\eta$. Figure 5 shows plot of this function over the ranges of $f$ and $\delta$ that are relevant in practice. Observe that the function $g(\delta, f)$ varies slowly over these ranges, which justifies using reasonable values of $f$ and $\delta$ in place of the true values. In particular, the choice $\delta = 0$ leads to $g = 1$ and hence, with this choice, the choice of $f$ is irrelevant in the small-$p$ regime.

As in the moderate transverse isotropy regime, the analytic small-ray-parameter result allows the estimation of $\eta$ from surface observations. Indeed, if one has observations of $V_{nmo}(p)$ at $p = 0$ and some other (not too large) value $p = p_1$, the solution for $\eta$ implied by equation (49) is

$$\eta = \frac{1}{12g} \left( \frac{V_{nmo}^2(p_1) - V_{nmo}^2(0)}{V_{nmo}^2(0)} - 1 \right), \quad (51)$$

where, once again, nominal values of $\delta$ and $f$ can be used in the factor $g$ without incurring much error.

Finally, note that the full form of the series for $V_{nmo}^2(p)$ is given by

$$V_{nmo}^2(p) = V_{nmo}^2(0) \left( 1 + c_2^2 y + c_4^2 y^2 + c_6^2 y^3 + \cdots \right). \quad (52)$$

Using equation (24), which keeps eighth-order terms in the small-$p$ expansion of $V(p)$, yields

$$c_2 = 1 + 12g \eta, \quad c_4 = 1 + 6g(6 - 5g) \eta + \frac{60g^2}{f} \eta^2, \quad c_6 = 1 + 8g(9 - 15g) \eta^2 + \frac{224g^3}{f^2} \eta^3.$$

The left panels of Figure 6 compare the exact movement function to the successive small-$y$ approximations of orders 1, 2, 3, and 4 for the same parameter values as used previously. The resulting value of $V_{nmo}(0)$ is 1.1 and the resulting value for $g$ is 1.06. The corresponding right panels show the phase angles as functions of $y$.

Aside from the dependence on $g$ that was discussed above, the higher order terms in the small-$p$ (small-$y$) expansion of the movement velocity function have explicit dependences on $f$. The first of these occurs in the final term in $c_4$. However, in the present small-$p$ expansion, this term is multiplied not only by $p^1$, but also by $y^2$, which ameliorates its influence. Similarly, $c_6$ contains terms of order $y^3 p^0$ and $y^3 p^0$ which also give rise to the
same type of mild divergence from strict dependence on the two Alkhalifah-Tsvankin principal parameters.

If \( f \) is set to one, the residual parameter dependence is entirely removed, since it enters only through \( g \) (which equals one when \( f = 1 \)) and \( f \) itself—that is, there is no \( \delta \) dependence other than that contained in \( g \). Again, this is in agreement with the general result in equation (40); and, indeed, expanding that result for small-\( y \) gives the \( f = 1 \) limit of the present result.

As a further consistency check, observe that for small-\( p \) (small-\( y \)), the elliptic result in equation (35) expands to

\[
V_{ei}(y) = \frac{V_{nmo}^2(0)}{1 - y} = V_{nmo}^2(0)[1 + y + y^2 + \cdots]. \tag{54}
\]

This agrees with the small-\( p \) expansion above, since for the elliptic case, \( \eta = 0 \) and hence \( c_2, c_4, \) and \( c_6 \) in equation (53) all reduce to 1 and so, equation (52) reduces to equation (54). Thus, just as in the case of moderate transverse isotropy, to the order computed, the small-\( p \) expansion in the elliptic limit is identical to the small-\( p \) expansion in the isotropic limit except for the \( \delta \) implicit in \( V_{nmo}(0) \).

Conclusions

I derived (in some cases re-derived) formulas for the \( P \)-wave phase and normal moveout velocities as a function of ray parameter, with explicit results given for the cases of elliptic anisotropy, vanishing on-axis shear speed, weak to moderate transverse isotropy, and small to moderate ray parameter. The various expressions for the phase velocity, nominally expansions in the ray parameter \( p \), turn out to actually be expansions in the dimensionless parameter, \( z = \sqrt{p} \). Similarly, the normal moveout function expansions turn out to be expansions in \( y = V_{nmo}^2(0)p^2 \).

For some selected choices of the anisotropy parameters, both the phase velocity and moveout function expansions have been compared with the corresponding exact functions.

Analytic explanation is provided here for the observation in Alkhalifah and Tsvankin (1995), that to high numerical accuracy, the normal-moveout velocity function as a function of ray parameter depends mainly on its zero-dip value and the parameter \( \eta \), with only a weak dependence on the residual parameters \( f \) and \( \delta \). In particular, an expansion in the anisotropy parameters explained the weak residual dependence in the important physical case of weak to moderately strong transverse anisotropy and an expansion in ray parameter established the result even for strong transverse anisotropy for small to moderate phase angles. In establishing the weak residual dependences, it is crucial that the Alkhalifah-Tsvankin principal parameters be introduced in a consistent and uniform manner—here, equations (39) provide this transformation. The residual parameter dependence vanishes completely in several special cases: \( p = 0 \) (i.e., zero phase angle), \( \delta = \epsilon \) (elliptic case), and when \( f = 1 \) (\( c_8 = 0 \)).

Acknowledgement

I thank Tariq Alkhalifah for performing some preliminary numerical tests and for his helpful suggestions; Ken Larner for his encouragement and careful critiques of several drafts of the text; and especially Illya Tsvankin for his deep understanding of anisotropic issues at several crucial junctures in this research.

The support for this work was provided by the members of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, and by the United States Department of Energy (project "Velocity Ana-
For the specific values $f = 0.75$, $c_p = 1.0$, and the values of the anisotropy parameters shown, the left panels are plots of the exact moveout velocity (dark), and its 4, 3, 2, and 1 term approximations shown successively lighter as functions of $y$. The right panels show the corresponding phase angles as functions of $y$.

References


nomena, Golden, Colorado.
A convenient expression for the NMO velocity function in terms of ray parameter

Jack K. Cohen
Professor of Mathematics, Center for Wave Phenomena
Colorado School of Mines

ABSTRACT
Tsvankin (1995) derived an expression for the NMO velocity function as a function of phase angle that is valid in symmetry planes of anisotropic media. This explicit expression plays a key role in developing practical data processing algorithms for transversely isotropic media (Alkhalifah & Tsvankin, 1995). Many of the key results in this theory require expressing the NMO velocity function as a function of ray parameter instead of phase angle and Alkhalifah and Tsvankin achieve this end in their work by indirect means. Subsequently, Cohen (1996) transformed Tsvankin’s moveout velocity expression to an explicit function of ray parameter. Both these representations of the moveout function are in terms of the phase velocity, expressed respectively, as a function of phase angle and as a function of ray parameter.

Here it is shown that for determining the moveout function as a function of ray parameter, there is a simpler expression based on using the slowness surface instead of the phase velocity. This simpler expression follows directly from a form of the moveout function given in Hale, et al. (1992). This form of the moveout function also served as the starting point for Tsvankin’s derivation. The new expression for the moveout as a function of ray parameter implies the previously derived ray parameter expression. Thus all the results derived before can be derived (more directly) with the new expression. Because of this, it suffices to illustrate techniques with a simple example and the pure shear wave mode of a transversely isotropic medium is chosen for this purpose.

Key words: transverse isotropy, NMO, DMO, ray methods, WKBJ, moveout function, ray parameter

Introduction
Tsvankin (1995) derived an expression for the NMO velocity function in terms of the phase velocity expressed as a function of phase angle that is valid in symmetry planes of anisotropic media. Alkhalifah and Tsvankin (1995) pointed out the importance of representing the NMO velocity function as a function of ray parameter and obtained a number of significant results via an indirect use of this variable. Subsequently, Cohen (1996) directly transformed the Tsvankin expression for the moveout velocity obtaining an expression involving the phase velocity expressed as a function of ray parameter. By obtaining an explicit expression in ray parameter for the $P$-wave phase velocity in a vertically transverse isotropic medium, he was able to get various approximations for the ray parameter form of the moveout function in a vertically transverse isotropic medium and thus give analytic insight into the success of the velocity analysis methodology of Alkhalifah and Tsvankin.

Here, I show that for determining the moveout function as a function of ray parameter, there is actually a simpler expression based on using the slowness surface instead of the phase velocity. This simpler expression follows directly from an expression of the moveout function by Hale, et al. (1992) that also served as the starting point for Tsvankin’s derivation. The key step in the derivation of the new moveout formula is inverting the expression for the slowness surface for one component—
typically the vertical component. Moreover, inverting the slowness surface for $q = q(p)$ is the first step in getting practical results from the previous form of the moveout function as well, so this step is needed anyway.

The new formula was motivated by the observation that the Hale, et al. expression for the moveout function is expressed in terms of ray parameter. It is natural to pursue a direct derivation in terms of ray parameter instead of first obtaining an expression in terms of phase angle and then transforming back to ray parameter. The WKBJ or "ray" theory provides such a ray-parameter-based framework.

The new expression for the moveout as a function of ray parameter is shown to imply the previously derived ray parameter expression. Thus all the results derived before can be derived (more directly) with the new expression and, so, it suffices to illustrate techniques with a simple example and the pure shear wave mode of a transversely isotropic medium is chosen for this purpose.

The derivation of the new formula is obtained from the Hale et al. expression by applying the WKBJ or high frequency assumption to the elastic equations, thus obtaining Christoffel's matrix equation. The eigenvalues of the Christoffel matrix imply an eikonal or "dispersion" equation for the travel time appropriate to a specific propagation mode. Applying the "method of characteristics" to the eikonal equation yields the factors required to explicitly evaluate the quantities in the Hale et al. formula for the moveout velocity.

The focus here is deriving an analog of the moveout functions of Tsvankin (1995) and Cohen (1996) for the case of wave propagation in a symmetry plane of a homogeneous medium. However, the methodology exposed holds promise for deriving moveout functions for inhomogeneous media and for the non-symmetry plane case.

Throughout this study, *Mathematica* was used extensively to derive and check results. In particular, the *Mathematica* packages, *VTL.m* and *ATools.m*, were used to convert expressions to Thomsen notation, and to obtain various results in the limit of weak and moderate transverse isotropy. Similarly, *Mathematica* was used to compute symbolic derivatives, series expansions, etc.

The WKBJ Approximation

The Hale, et al. (1992) formula for the moveout function can be written:

$$ \nu_{\text{move}}^2(p) = \frac{1}{\tau} \lim_{\zeta \to 0} \frac{dx}{dp}, \quad (1) $$

where $\tau$ is the one-way travel time along the zero-offset specular ray to a dipping reflector, $x$ is the horizontal spatial variable, and $p$ is the horizontal ray parameter.

The derivation of equation (1) assumes that the rays stay in a plane, which implies, for an anisotropic medium, that the incidence plane is also a symmetry plane. It is interesting to note that equation (1) is closely related to equation (5) in the classic paper of Dix (1953).

The appearance of the ray parameter in the above expression suggests that we directly determine $\tau$ and $dz/dp$ from the ray or WKBJ approximation to the elasticity equations, whose form in $\omega$-$x$ domain is:

$$ (c_{ijkl} u_{k,i})_{,i} + \rho \omega^2 u_i = 0. \quad (2) $$

The WKBJ ansatz,

$$ u_i(x, \omega) \approx e^{i \tau(x)} U(x), \quad (3) $$

yields Christoffel’s matrix equation:

$$ (c_{ijkl} \tau \tau_{,i,j} - \rho \delta_{ik}) U_k = 0. \quad (4) $$

Higher order terms would yield a “transport” equation that would complete the determination of the WKBJ amplitude, but here we need only the kinematic results implied by equation (4). Defining the slowness vector, $p_i = \tau_{,i}$, obtain

$$ (M - \rho \mathbb{I}) U = 0, \quad M_{ik} = c_{ijkl} p_l p_j, \quad \mathbb{I}_{ik} = \delta_{ik}. \quad (5) $$

The symmetries of $c_{ijkl}$ imply that $M$ is a symmetric matrix, so its eigenvalues are real and its eigenvectors can be chosen to be an orthonormal basis in 3-space. Equation (3) implies that each eigenvalue,

$$ \lambda^{(p)} = \lambda(p, x), \quad (6) $$

defines an eikonal equation,

$$ \lambda(p, x) = \rho, \quad (7) $$

for the mode corresponding to that eigenvalue. The associated eigenvector $U$ defines the polarization of the mode.

Since $M$ is homogeneous of degree 2 in the $p_i$'s, the $\lambda$'s are also, and Euler's result for homogeneous functions yields the relation,

$$ \mathbf{p} \cdot \nabla_p \lambda = 2\lambda, \quad \nabla_p = (\partial_{p_1}, \partial_{p_2}, \partial_{p_3}). \quad (8) $$

Phase and Group Velocities

The phase and group velocities can also be defined in terms of the eigenvalues defined by equation (3). The group velocity is defined by

$$ \mathbf{g} = \nabla_k \omega, \quad \nabla_k = (\partial_k, \partial_{k_2}, \partial_{k_3}). \quad (9) $$

Here, $k$, the wave vector, is given by $k = \omega \mathbf{p}$. The matrix with components,

$$ M_{ik}^{(k)} = c_{ijkl} k_j k_l, \quad (10) $$
has the same eigenvectors as $M$, and the eigenvalues, $\lambda^k$, have the same functional form as the corresponding $\lambda^p$, and are homogeneous of degree 2 in the $k$ argument. Thus,
\begin{equation}
\lambda^k(k, x) = \omega^2 \lambda(p, x) = \rho \omega^2.
\end{equation}
(11)

Solving for $\omega$, gives
\begin{equation}
\omega = \sqrt{\frac{\lambda^k}{\rho}},
\end{equation}
(12)

hence, from equation (9)
\begin{equation}
g = \nabla_x \sqrt{\frac{\lambda^k}{\rho}} = \frac{1}{\omega} \nabla_p \sqrt{\frac{\omega^2 \lambda}{\rho}} = \nabla_p \sqrt{\frac{\lambda}{\rho}} = \frac{\nabla_p \lambda}{2 \sqrt{\rho \lambda}},
\end{equation}
(13)

where the abbreviation $\lambda = \lambda(p, x)$ has been used. The eikonal equation (7) implies the following two alternative forms for the group velocity, $g$:
\begin{equation}
g = \frac{1}{2} \nabla_p \lambda(p, x) = \frac{1}{2 \rho} \nabla_p \lambda(p, x).
\end{equation}
(14)

The phase velocity is defined by
\begin{equation}
v = \frac{\omega}{k}, \quad n \cdot n = 1,
\end{equation}
(15)

with magnitude, or phase speed,
\begin{equation}
v = \frac{\omega}{k}.
\end{equation}
(16)

Since $k = \omega p$, and $k = kn$, the unit normals $n$ are given in terms of the slowness $p$ as
\begin{equation}
n = \frac{\omega}{k} p = \nu p.
\end{equation}
(17)

The matrix with components,
\begin{equation}
M^{(n)} = c_{ijkl} n_i n_j n_k,
\end{equation}
(18)

again has the same eigenvectors as $M$, and again the eigenvalues, $\lambda^{(n)}$, have the same functional form as the corresponding $\lambda^{(p)}$, so
\begin{equation}
\lambda^{(n)}(n, x) = v^2 \lambda(p, x) = \rho v^2.
\end{equation}
(19)

Solving for $v$, gives
\begin{equation}
v = \sqrt{\frac{\lambda^{(n)}(n, x)}{\rho}} = \sqrt{\frac{\lambda(n, x)}{\rho}}.
\end{equation}
(20)

If we represent the unit vector $n$ in spherical polar coordinates, equation (20) gives the usual phase angle representation of the phase velocity. On the other hand, since $n \cdot n = 1$, equation (17) yields an alternative expression for phase velocity as a function of the slowness vector:
\begin{equation}
v = \frac{1}{\sqrt{p \cdot p}}.
\end{equation}
(21)

Remark: The definitions (9), (16) of group velocity and phase speed imply that $g = V_k(kv)$. This is a key result in Tsvankin’s derivation, but it will not be needed here.

**NMO Velocity as a Function of Ray Parameter**

First observe that
\begin{equation}
p \cdot g = \frac{p \cdot \nabla_p \lambda}{2 \lambda} = 1,
\end{equation}
(22)

where the final equality follows from equation (8). This well-known result will be critical to the derivation.

The eikonal equation, $\lambda = \rho$, is a first order partial differential equation for the traveltime $\tau$ and hence can be reduced to solving a system of ordinary differential equations by the “method of characteristics” (Courant & Hilbert, 1962). Denoting differentiation along the characteristics (or “rays”) with a dot, this system can be written:
\begin{align*}
\dot{x} &= \nabla_p \lambda = 2 \rho g, \quad x(0) = 0, \\
\dot{p} &= -\nabla_x \lambda, \quad p(0) = p_0, \\
\dot{\tau} &= \nabla_x \tau \cdot \dot{x} = 2 \rho \rho \cdot g = 2 \rho, \quad \tau(0) = 0.
\end{align*}
(23)

Here, equation (22) has been used to simplify the $\tau$ equation.

Remark: The ray equations imply that
\begin{equation}
\frac{d\tau}{d\xi} = \frac{2 \rho g}{2 \rho} = g.
\end{equation}
(24)

This could serve as an alternative definition of group velocity if one wished to avoid introducing the wave vector $k$.

Now confine attention to a symmetry plane where $p_z = 0$ and $x_2 = 0$, and since the ray problem is now two dimensional, introduce the special notations,
\begin{equation}
x = (x, 0, z), \quad p = (p, 0, q),
\end{equation}
(25)

here, $p = p_\xi$, is called the ray parameter and appears already in equation (1). The group velocity likewise simplifies and takes the form,
\begin{equation}
g = \frac{1}{2 \rho} \nabla_p \lambda = \frac{1}{2 \rho} (\lambda_\rho, 0, \lambda_q).
\end{equation}
(26)

Using $z$ as the parameter along rays, equations (23) simplify to
\begin{align*}
\frac{dz}{dz} &= \frac{\rho_z}{g_z}, \quad z(0) = 0, \\
\frac{dp}{dz} &= -\frac{\lambda_z}{2 \rho g_z}, \quad p(0) = p_0, \\
\frac{d\tau}{dz} &= \frac{1}{g_z}, \quad \tau(0) = 0.
\end{align*}
(27)

Here, in place of a differential equation for $q$, we conceive of solving the eikonal equation, $\lambda = \rho$, for $q = q(p, x, z)$. The implicit function theorem implies that
\begin{equation}
\frac{\lambda_p}{\lambda_q} = -q,
\end{equation}
(28)
where we use a prime to denote differentiation with respect to the ray parameter, \( p \). Thus,
\[
\frac{g_3}{g_2} = \frac{\lambda_p}{\lambda_q} = -q'.
\] (29)

This result and equation (22) imply that
\[
\frac{1}{g_3} = q - pq'.
\] (30)

Hence the ray equations become
\[
\begin{align*}
\frac{dx}{dz} &= -q', \quad x(0) = 0 \implies \frac{d}{dz}(\frac{dz}{dp}) = -q'', \quad x'(0) = 0, \\
\frac{dp}{dz} &= -\frac{\lambda_p}{\lambda_q}, \quad p(0) = p_0, \\
\frac{d\tau}{dz} &= q - pq', \quad \tau(0) = 0,
\end{align*}
\] (31)

with \( q \) determined by solving \( \lambda = \rho \). Notice that an analogous form of the ray equations (31) can be derived in a three-dimensional setting using the solution
\[
p_z = q(p_1, p_2, x, y, z) = \lambda = p.
\]

For a laterally homogenous medium, \( p \) is a constant along the rays, and the solutions of the ray equations can be expressed as ordinary integrations over \( z \). Here, we specialize further to the case of a completely homogenous medium, in which case, \( q = q(p) \) is constant on rays and explicit solutions are immediately available:
\[
x = -q'z, \quad x' = -q''z, \quad \tau = (q - pq')z.
\] (32)

Thus for a homogeneous medium, the Hale formula (1) gives
\[
V_{nmo}^2(p) = \frac{q''}{pq' - q},
\] (33)

where again, \( q = q(p) \) is determined by \( \lambda = \rho \) for each mode. The derivation of (33) is the key result of this paper and applications are shown in the next section.

As contrasted to the methodologies in Tsvankin (1995) and Cohen (1996), the moveout function is here obtained in terms of \( q(p) \), instead of in terms of the phase speed, \( v(\theta) \) or \( v(p) \). The expression for the moveout function in terms of \( v(p) \) follows from solving equation (21) for \( q \) to obtain
\[
q(p) = \sqrt{1/v^2(p) - p^2}.
\] (34)

Inserting this expression into equation (33), gives the result in equation (35) which agrees with the moveout function derived in Cohen (1996).

Application
As a simple example of the results obtained above, consider the pure shear mode in a VTI medium (i.e., a medium that is transversely isotropic with a vertical axis of symmetry). For corresponding quasi-\( P \) wave results based on the equivalent, but more complicated, equation (35), see Cohen (1996).

The moveout function
For a VTI medium, the \( \mathbf{M} \)-matrix is shown in equation (36). For the pure shear mode, the eigenvalue, expressed in terms of the isotropic speed \( c_s \) and the Thomson parameter \( \gamma \) is
\[
\lambda_S = \rho c_s^3[1 + 2(1 + \gamma) p^2 + q^2].
\] (37)

with corresponding \( q \), given by
\[
q(p) = \frac{1}{c_s} \sqrt{1 - (1 + 2\gamma) c_s^2 p^2}.
\] (38)

This solution only involves solving the linear equation (37) in the quantity \( q^2 \). For the other modes, the quasi-\( P \) and quasi-\( S \), the corresponding equation is a quadratic in \( q^2 \) and so, can again be solved explicitly. Indeed, the solution of this quadratic for \( q^2 \) was a step in the phase velocity-based methodologies (Alkhalifeh & Tsvankin, 1995), (Cohen, 1996).

With \( q(p) \) given by equation (38), equation (33) gives
\[
V_{nmo}^2(p) = \frac{c_s^2(1 + 2\gamma)}{1 - (1 + 2\gamma) c_s^2 p^2}.
\] (39)

Introducing the zero-dip moveout,
\[
V_{nmo}(0) = c_s \sqrt{1 + 2\gamma},
\] (40)

and the dimensionless variable,
\[
y = V_{nmo}^2(0)p^2,
\] (41)

one obtains the result in Alkhalifeh-Tsvankin format as
\[
V_{nmo}(p) = \frac{V_{nmo}(0)}{\sqrt{1 - y}}.
\] (42)

Notice that \( V_{nmo}(p) \) has been obtained without first determining the phase speed.

Phase and group velocity
Although the phase speed is not required for the determination of the moveout function, the \( q \)-function does also determine the phase speed as a function of ray parameter. Indeed, equation (34) gives \( v(p) \) as
\[
v(p) = \frac{1}{\sqrt{p^2 + q^2(p)}}.
\] (43)

For example, using the \( q(p) \) derived above for the shear mode in a transversely isotropic medium gives
\[
V_S(p) = \frac{c_s}{\sqrt{1 - 2\gamma c_s^2 p^2}}.
\] (44)
\[ V_{n,m}^2(p) = \frac{(1 - p^2v^2)\nu'\nu + (3p^2v^2 - 2)v'^2 - 2pv^2v' + v^4}{(1 - p^2v^2)\nu(p\nu)'}, \]  

(35)

\[ M_{V'TI} = \begin{pmatrix} 
    (c_{11} + c_{33})p_1^2 + c_{33}p_3^2 & (c_{11} + c_{33})p_1p_3 & (c_{11} + c_{33})p_1p_3 \\
    (c_{11} + c_{33})p_1p_3 & c_{33}p_3^2 + c_{11}p_1^2 + c_{33}p_3^2 & (c_{11} + c_{33})p_2p_3 \\
    (c_{11} + c_{33})p_1p_3 & (c_{11} + c_{33})p_2p_3 & c_{33}p_3^2 + c_{33}p_3^2 + c_{33}p_3^2 
\end{pmatrix} \]  

(36)

Alternatively, to obtain the phase speed as a function of angle, equation (20) allows direct use of the eigenvalue. Indeed, writing the S-wave eigenvalue shown above for the formulation in terms of the unit normal gives

\[ \lambda_S(n) = \rho c_S^2 [(1 + 2\gamma)n_1^4 + n_2^4]; \]  

(45)

so

\[ u_S(\theta) = c_S\sqrt{1 + 2\gamma \sin^2 \theta}. \]  

(46)

Finally, equation (26) allows writing the group velocity in terms of partial derivatives of the eigenvalue in equation (37) as

\[ g = c_S^2 ((1 + 2\gamma)p, 0, 0), \]  

(47)

or using the solution, \( g(p) \), as

\[ g(p) = ((1 + 2\gamma)c_S^2 p, 0, c_S \sqrt{1 - (1 + 2\gamma)c_S^2 p^2}). \]  

(48)

The substitution, \( p = \sin \theta / u_S(\theta) \), produces

\[ g(\theta) = \frac{c_S}{\sqrt{1 + 2\gamma \sin^2 \theta}} ((1 + 2\gamma) \sin \theta, 0, \cos \theta). \]  

(49)

**Conclusions**

A new and simpler expression for the moveout velocity in a symmetry plane as a function of ray parameter has been derived from first principles. Its use has been illustrated with an example from transversely isotropic media.

**Acknowledgement**

Support for this work was provided by the members of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, and by the United States Department of Energy (project “Velocity Analysis, Parameter Estimation, and Constraints on Lithology for Transversely Isotropic Sediments” within the framework of the Advanced Computational Technology Initiative).

**References**


Entropy, information and inference

Wences Gouveia, Fernando S. de Moraes and John A. Scales
Department of Geophysics and Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401, USA

ABSTRACT
The goal of inversion is to make quantitative inferences about physical systems from observations. Since such inferences are invariably uncertain, owing to uncertainties in the data, in the specification of the model and in the accuracy of the forward modeling operator, a conservative guiding principle would seem to be: draw the least informative conclusions that are consistent with the available information. This is especially important when it comes to the specification of those prejudices we hold as to what makes a model reasonable or not. These prejudices can be based on theory or observation but must enter the inference process since otherwise we may just as easily compute unreasonable models that are consistent with the data. A possible strategy for conservatively quantifying prior information is to find that distribution which has the least information content (or maximum entropy) among all distributions that agree with the available information. In order to carefully lay out the logic of this approach and its potential pitfalls, we will give an overview of the foundations of Bayesian entropy-based methods in inference. Starting with the thermodynamic definition of entropy, which is fundamentally linked to the extent to which one can specify the state of a system, we show how this leads naturally to statements about the uncertainty of certain experiments involving the state of thermodynamic systems. It is a short leap from this to the information-theoretic use of entropy developed by Shannon. We will see that a certain set of postulates, which we believe any reasonable measure of information must satisfy, leads inevitably to Shannon’s definition of entropy, but that in order make the concept rigorous and useful in inversion we must abandon the notion of absolute states of information and instead focus on measuring the information of one state relative to another (reference) state. We compute Bayesian probability distributions by maximizing the entropy of the distribution subject to constraints on some number of estimated moments. Using only first and second moments of the prior information we get a Gaussian approximation to the prior distribution. Using higher-order moments affords a better approximation to some data, especially data that have an asymmetrical distribution. We will show examples of one-dimensional calculations using moments up to fourth order. The application of these methods to n-dimensional problems is a significant computational challenge; we show an approach that eliminates the need of doing multi-dimensional quadrature.

Key words: Entropy, Information, Statistical inference, Bayesian prior distributions, Inverse theory

Introduction

The goal of inversion is to make quantitative inferences about physical systems from observations. An inference is a conclusion drawn from certain premises. Inasmuch as today’s conclusions form tomorrow’s premises, conclusions and premises can both be viewed as relative states of information. Information is the unifying concept behind all inverse calculations. When we contemplate such a calculation we must begin by surveying the information
available to us from which we can draw conclusions. This information comes in many forms and describes what we know about the errors in the data, the distribution of reasonable model parameters and the accuracy of the forward modeling procedure. Some of this information is absolutely certain and results from purely logical argument. The fact that density is positive is not a claim subject to experimental verification. A second level of information comes as the result of a theoretical argument. The fact that the speed of light in vacuo must be constant, for example. But by far, the preponderance of information we have at our disposal is the result of direct observation and measurement of physical systems. These observations can result in subjective information—expertise—or quantitative information—data, sometimes with no clear distinction between the two. We believe that these two forms of information should enter inference calculations in a unified fashion.

Thus the motivation for quantifying states of information is clear. We need to be able to assess the amount of information contained in different premises, which premises are usually expressed as functionals on the space of models and data. The statement that, for instance, the P-wave velocity $V_p$ of a certain piece of the earth is positive is less informative than the statement that the $V_p$ velocity lies within some finite range $[0 < z \leq V_p \leq y < \infty]$. The first is absolutely true but conveys little information, while the second is rather more informative, but possibly wrong. When we attempt to quantify the state of information that results from an inference or inverse calculation, we are, in effect, measuring the resolution of the calculation. This quantification takes the form of probabilistic statements about hypotheses, expressing a degree of belief in their truth. In the above example, we may wish to know the smallest range of P-wave velocity that contains the true value with a given degree of confidence.

The modern theory of information can be traced to the pioneering work of Shannon who defined the information of a communication channel in terms of the probabilities associated with the different symbols being transmitted. Actually Shannon dealt not with information directly, but with entropy, the negative of information: the more information, the less entropy, and vice versa. Unfortunately, in attempting to understand these ideas and, perhaps, make use of them in our work on inversion, we quickly discovered a number of troubling technical issues. E.g., most of the textbooks define entropy (or information) in a way that is unacceptable for our purposes since the definition fails to satisfy a basic invariance property. (It cannot matter what coordinate system we use to make our inferences.) Further, it appears that in using a limiting argument to pass from a discrete to a continuous measure of information, certain infinities crop up that are swept under the rug. As we shall see, fixing the first problem also fixes the second. Additionally, in so doing we achieve a measure of information that is finite for finite (and normalizable) states of information. We will also show how reasonable results can be achieved in the case of infinite states of information by a suitable limiting procedure.

The price we pay for this mathematical consistency is the replacement of the notion of information (or entropy) with one of relative information: we cannot speak of the content of a given state of information; we can only speak of the information of one state of information relative to another. But this is, after all, what we need for inverse calculations since we always begin with some state of information and attempt to refine it.

This point of view is hardly new. Here is Kolmogorov writing 40 years ago:

Furthermore, I insist that the fundamental concept, which admits of generalization to perfectly arbitrary continuous information and signals, is not directly the entropy concept but the concept of the quantity of information $I(q; \xi)$ in the random object $\xi$ relative to the object $q$. It is well known that the quantity $h(\xi)$ [continuous entropy] has no direct real interpretation and is not even invariant with respect to coordinate transformations in the space of the $\xi$'s. For an infinitely-dimensional distribution, the analog of $h(\xi)$ is nonexistent, in general (Kolmogorov, 1956).

And yet this principle of relative information does not seem to be well understood in practice, Tarantola's book (1987), being a notable exception.

Our original motivation for looking into entropy was the ability to use the principle of maximum entropy to calculate conservative Bayesian priors from observations. For example, if we take in-situ measurements of some property and attempt to use those measurements to build a Bayesian $a$ priori distribution, then we are faced with the task of constructing the joint n-dimensional distribution of a process of unknown complexity. If we are willing to assume that the process is Gaussian, then computing the prior reduces to the classical problem of estimating the mean and covariance from sampled data. But if we wish to get beyond the Gaussian assumption, then we must somehow take into account higher order information. In principle, one can estimate some number of higher order marginal distributions via histograms, but without some underlying model of the process, we need an external criterion to decide when to stop this procedure. In Scales & Tarantola (1994) we considered using statistical hypothesis testing to answer this question, but that approach is not without limitations too, and could be prohibitively expensive beyond second or third order
Entropy, information and inference 269

marginals without additional approximations. Thus we were lead to consider whether a properly conservative estimation of prior distributions could be achieved via the principle of maximum entropy. In other words, given sampled data, estimate some number of sample moments of the unknown distribution and then calculate a prior distribution that is least informative subject to these sample moments as constraints. Once we have the prior distribution we can proceed to solve the full Bayesian inference calculation. In theory this estimation of the prior distribution is a straightforward application of the maximum entropy method. Indeed we show examples of this approach applied to one-dimensional random processes. However, the challenges in higher dimensions are formidable.

But we are getting ahead of ourselves. To lay a proper foundation for the study of entropy-based methods we begin at the beginning. In the first section we discuss the other entropy, i.e., that used in thermodynamics and statistical mechanics. By considering a simple example of an irreversible thermodynamic process we will see that even the classical entropy can be thought of in terms of the uncertainty of the microstate of a system. This connection between the macroscopic (thermodynamic) and the microscopic (statistical mechanics) view of a system allows one to derive an expression for the entropy in terms of the probability distribution for the coordinates and momenta of a microscopic state of the system in phase space $\Gamma$. It was Shannon’s great insight that this quantitative measure of uncertainty could be applied generally to problems involving information. In the second section we discuss Shannon’s definition and show by a number of simple examples that it does indeed capture our qualitative understanding as a measure of information. This alone may seem hardly compelling. However, it can be shown that by asking what properties any reasonable measure of information should have, we will be drawn inexorably to Shannon’s definition, or something very like it. The hedge in the last sentence has to do with the fact that, as we have already mentioned, Shannon’s definition of entropy fails to be invariant with respect to coordinate transformations. The remedy for this is mathematically quite straightforward, but the interpretation requires that we alter fundamentally our interpretation of the procedures by which we obtain and measure states of information. Using this revised definition of information clears up a number of related mathematical difficulties and allows a single theoretical framework for continuous and discrete states of information.

In an appendix we give a review of these claims and consider several different derivations of the basic equations. In addition, we show a number of worked examples of how to do these calculations. For example, it makes no sense to speak of the entropy of a distribution (even a normalizable one) relative to a constant of information on an unbounded domain since no such reference state of information is normalizable. Fortunately we get the right answer via a limiting procedure, treating only normalizable distributions at each stage.

With this theoretical foundation, we are then in a position to discuss the principle of maximum entropy. Since entropy is a measure of uncertainty, maximizing the entropy, subject to certain constraints determined by our information, represents a fundamental statement of conservatism: in the presence of uncertainty we should make the least informative inferences (i.e., the safest ones) consistent with our information. So the maximum entropy calculation is actually a constrained optimization calculation for an unknown probability distribution. In principle this can be solved by the method of Lagrange multipliers. And for one-dimensional problems this is straightforward. The situation is much less clear in higher dimensions. We will conclude with some recent work we have been doing on the numerical extension of these ideas to $N$-dimensional calculations and show how these techniques might be used to incorporate complicated prior information in the Bayesian inverse calculations.

Entropy in Thermodynamics and Statistical Mechanics

The concept of entropy in thermodynamics is intimately related to the question of reversibility. The second law of thermodynamics says that a process whose only net result is to take heat from a reservoir and convert it to work is impossible. Carnot used this principle to show that the amount of work obtained for any two reversible engines extracting an amount of heat $Q_1$ from a reservoir at temperature $T_1$ and delivering it to a reservoir at temperature $T_2$ must be the same independent of the design of the engine.

Carnot considered an idealized kind of heat engine which operates using a gas-filled piston to extract heat $Q_1$ from a reservoir at temperature $T_1$ during an isothermal expansion. The gas is then allowed to expand further but now isolated from the heat reservoir so that the temperature drops to $T_2$ with no heat being transferred. Next the gas is put in contact with a heat reservoir of temperature $T_2$ and compressed isothermally delivering a heat $Q_2$ to the reservoir. Finally the gas is removed from contact with the $T_2$ reservoir and compressed adiabatically (no heat transfer) raising the temperature from $T_2$ to $T_1$. If the piston is frictionless and all the operations are carried on very slowly, then this idealized engine is
reversible. The gas ends up in exactly the state it began with some useful work having been delivered.

Since Carnot's results are independent of the type of reversible engine used, it is possible to study a particular kind of engine and derive universal laws. For Carnot's engine using an ideal gas (i.e., a gas whose equation of state is \( PV = NkT \), where \( P \) is the pressure, \( V \) is the volume, \( N \) is the number of particles, \( k \) is Boltzmann's constant, and \( T \) is the temperature), it is possible to evaluate the expressions for \( Q_1 \) and \( Q_2 \) in terms of \( T_1, T_2 \) and the volumes of the piston at various stages of the engine's cycle. It then follows that

\[
\frac{Q_1}{T_1} = \frac{Q_2}{T_2}.
\]  

(1)

The derivation can be found in most thermodynamics books, but an especially clear one is given by Feynman (Feynman et al. (1963), Volume I, section 44). The universality of Carnot's result means that although Equation (1) has been proved for an ideal gas, it must be true for any reversible engine. (Feynman calls this purely logical deduction of a universal principle one of the most beautiful pieces of reasoning in physics.) Since there can be no net change in \( \frac{Q}{T} \) during any reversible cycle, this ratio is special enough to warrant a new name: entropy. So in a reversible cycle the total entropy of the system does not change. In an irreversible cycle the total entropy of the system always increases.

Let us consider a specific example of an irreversible transformation. Suppose we have a box divided in half by a partition with a quantity of gas in one half of the box. We remove the partition and let the gas expand to fill the whole box while keeping the temperature constant. Clearly this is an irreversible change since the gas molecules will never arrange themselves so as to spontaneously occupy only half of the box. The entropy change in this transformation is

\[
\Delta S = \int \frac{dQ}{T}.
\]  

(2)

Since the temperature does not change, we must add a little heat \( dQ \) to equal the work done by the gas in the expansion \( PdV \). So

\[
\Delta S = \int_{V_1}^{V_2} \frac{P}{T} dV
\]  

(3)

which, for an ideal gas, equals

\[
\int_{V_1}^{V_2} \frac{NkT}{V} dV.
\]  

(4)

Thus,

\[
\Delta S = Nk \ln \frac{V_2}{V_1}.
\]  

(5)

The entropy has increased by an amount \( Nk \ln \frac{V_2}{V_1} \) even though the temperature and energy of the system are the same. All we have done is given the molecules more room to move around. Or, anticipating a bit, we have increased the number of possible states of the system; that is, the number of ways in which the molecules could be arranged. The entropy is the logarithm of this number of states.

Now that we have slipped into thinking about our system microscopically, it is useful to imagine the state of any system of \( N \) particles as being specified by a point in a \( 6N \)-dimensional phase space \( \Gamma \) comprising the values of all \( 3N \) coordinates \( q \) and \( 3N \) momenta \( p \). The time evolution of the system is then described by a trajectory \( \{q(t), p(t)\} \) in phase space which is uniquely determined by \( 6N \) initial values of the coordinates and momenta \( \{q(t_0), p(t_0)\} \). So since it is clear from the gas expansion example that the density of states available to a system is closely related to the classical entropy, it will be useful to define a probability density \( \rho \) on the phase space telling us the probability that our system will be found in any element \( dqdp \) by

\[
P_{\{q_1, p_1\} \in dqdp} \equiv \rho_{\{q, p\}} dqdp
\]  

(6)

The precise form of the ensemble function \( \rho \) depends on the the type of problem being treated. If the system is isolated, then the Hamiltonian \( H \) is constant and any state on the surface \( H = E \) is equally likely. Thus this micro-canonical ensemble must be described by a delta function: \( \rho = \delta(H(q, p) - E) \). However, for purposes of Carnot's theory, we need the ensemble associated with a system at constant constant temperature, not constant energy. This is called the canonical ensemble and the expression for \( \rho \) in this case turns out to be the Boltzmann distribution (Weiner, 1983)

\[
\rho_{\{q, p\}} = \exp(-H(q, p)/kT).
\]

The key idea is this: Our expression above for the change of entropy (Equation (2)) involves the change in heat or energy at constant temperature. In the canonical ensemble this would just be the time rate of change of the expected value of the Hamiltonian. The details can be found in (Weiner, 1983). The upshot is that by making this connection between the macroscopic thermodynamic picture and the microscopic statistical one, it is possible to deduce the following expression for the entropy of a system in the canonical ensemble:

\[
S = -k \int \rho_{\{q, p\}} \ln \rho_{\{q, p\}} dqdp
\]  

(7)

Anticipating further our journey into information theory, we can imagine the thought experiment of observing our system at any instant of time. The possible outcomes of this experiment are the points in the phase
space accessible to the system. And so the entropy of the system must be a measure of the uncertainty associated with this experiment. If the trajectory of the system is confined to a relatively small volume of $\Gamma$ it will have a relatively small entropy. If the trajectory wanders widely in $\Gamma$ it will have a correspondingly greater entropy.

**Classical Measures of Information**

Consider an experiment with $N$ possible outcomes each occurring with a probability $p_i$. In analogy with the statistical mechanical definition of entropy just discussed, (Shannon, 1948) introduced the following definition of the entropy for such discrete probabilities:

$$H(p) = - \sum p_i \log p_i. \quad (8)$$

Following Shannon, three postulates should be satisfied by $H(p)$ or any other measure of information. Those are:

(i) Continuity;
(ii) Monotonicity, and
(iii) Composition Law.

We elaborate on these postulates in Appendix A, for now a qualitative understanding is sufficient. The first postulate requires that we should not gain a large amount of information by making a small change to the probabilities. The second postulate, monotonicity, refers to the information associated with a collection of independent, equally likely events. It is clear that in such a case the uncertainty must increase monotonically with the number of possible outcomes. The third postulate requires that it should not matter how one regroups the events of a given set. The entropy of the set should stay the same.

Later we will have to modify the definition of $H$ given in Equation 8 in an important way, but for now let us just consider its basic meaning. If one of the outcomes of the experiment is absolutely certain, then we represent the probability density by a Kronecker delta: $p_i = \delta_{iq}$ where $q$ is the certain event. In this case there is no uncertainty and $H(p) = 0$. If there are two equally likely outcomes then $H(p) = -1/2 \log(1/2) = 1/2 \log(1/2) = \log 2$. Whereas if one of the events has a probability 1/10 and one has probability 9/10 then the entropy is $H(p) = -1/10 \log(1/10) - 9/10 \log(9/10) = \log 10 - .9 \log 9$, which is about half that in the equally likely case. For $M$ equally likely events $H(p) = \log M$. And in the limit that $M$ goes to infinity, then the uncertainty must too.

The usefulness of the definition 8 depends on the definition of the probability $p$. If the $p$ is a 1D distribution associated with the frequency of outcomes of the possible events, then $p$ is not affected by the correlation of the points. E.g., if we sample 10 points pseudo-randomly from a probability distribution with two equally likely outcomes 0 and 1, we might see something like the following

$$0010110110.$$

As luck would have it there are 5 0's and 5 1's. Now sort these outcomes in increasing order

$$0000011111.$$

There are still 5 0's and 5 1's but we certainly would not regard the latter experiment as representing the same degree of uncertainty as the former. Similarly, if we sample 1000 points independently from a Gaussian we'll see a nice bell-shaped curve. But the frequencies of the binned events is independent of their order. So, once again, sorting them into monotonic order will not change the entropy. Later we will see how it is possible to include the dependency of events in the definition of $p$ and achieve a completely general definition of entropy.

**Extension to Continuous Probabilities.**

The extension of Equation (8) definition to continuous probabilities is necessary since in many scientific problems we deal with continuous variables. Several approaches can be found in the literature (Jaynes, 1963, Rietzch (1977)), and others). Here we describe the one given by Rietzch. Assume that $x$ is a continuous variable that lies in some interval $[a,b]$. Let $p(x)dx$ denote the probability that a value of $x$ be in the interval $[x,x+\Delta x]$. A "natural" way to define the entropy of $x$ would be to subdivide the interval $[a,b]$ into $n$ sub-intervals $[x_{i-1},x_i]$ of length $\Delta x_i$; $p(x_i)\Delta x_i$ is the probability that $x$ is in this interval. Therefore, the entropy of this discretized probability distribution reads

$$H[p(x)] = - \sum_i p(x_i)\Delta x_i \log p(x_i)\Delta x_i =$$

$$- \sum_i p(x_i)\Delta x_i \log p(x_i) - \sum_i p(x_i)\Delta x_i \log \Delta x_i. \quad (9)$$

Introducing the variable $q(x_i)$ and the constant $\delta$ such that

$$\Delta x_i = \frac{\delta}{q(x_i)}, \sum_i q(x_i)\Delta x_i = n\delta = 1,$$

we have for Equation (9):

$$H[p(x)] = - \sum_i p(x_i)\Delta x_i \log p(x_i).$$
\[-\sum_i p(x_i) \Delta x_i \log \frac{\delta}{q(x_i)} \]
\[= - \sum_i p(x_i) \Delta x_i \log p(x_i) \]
\[= - \sum_i p(x_i) \Delta x_i \log \delta \]
\[= + \sum_i p(x_i) \Delta x_i \log q(x_i) \]
\[= - \sum_i p(x_i) \Delta x_i \log \frac{p(x_i)}{q(x_i)} + \log n, \quad (11) \]
in which we used the fact that \(\sum p(x_i) \Delta x_i = 1\). Notice that as \(n \to \infty\), Equation (11) diverges because of \(\log n\). Nonetheless this term is neglected in Rietcsch's derivation, and in the limit, the final result reads:

\[H[p(x); q(x)] = - \int_a^b p(x) \log \frac{p(x)}{q(x)} \, dx. \quad (12)\]
The above expression defines what is known as relative entropy, \(q(x)\) is Jayne's invariant measure. It is not complicated to show that \(q(x)\) is sufficient to make the function \(H[p(x); q(x)]\) invariant to coordinate transformations. To do so consider changing coordinates from \(x\) to \(y\) in Equation (12). This yields:

\[H[p(x); q(x)] = - \int_a^{b'} p(y) \frac{dy}{dx} \log \frac{p(y)}{q(y)} \, dx \quad (13)\]
where \(a'\) and \(b'\) are the corresponding limits of integration in the \(y\) coordinates. Equations (12) and (13) are of course identical. As it will be shown later, the quantity \(q(x)\) can also be interpreted as a reference (or prior) state of information. In this view, \(H[p(x); q(x)]\) quantifies the relative entropy of the probability function \(p(x)\) with respect to \(q(x)\).

However, we are not entirely comfortable with Rietcsch's derivation. Other approaches to the extension from discrete to continuous entropy resort to similar arguments; and while the renormalization of infinite integrals by subtraction is not unheard of, we were motivated to see whether this was really necessary. Let us therefore consider the following definition for discrete entropy:

\[H[p_i; q_i] = - \sum_i p_i \log \frac{p_i}{q_i}, \quad (14)\]
Here, \(q_i\) is a discrete probability characterizing a reference state of information. The extension of Equation (14) to the continuous case is clean and straightforward:

\[H[p(x); q(x)] = - \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)} \, dx, \quad (15)\]
which is finite.

From here on we will adopt Equation (14) as the definition of relative entropy in the discrete case, and, as commonly done, the last expression of Equation (16) as the definition of relative entropy in the continuous case. \(q(x)\), or \(q_1\), represents a state of information against which we make comparisons. Finally it is worth mentioning that the negative of the quantity \(H[p(x); q(x)]\), known as cross-entropy, was first defined by Kullback (1959) as the directed divergence. This quantity defines the amount of information of the probability density \(p(x)\) with respect to \(q(x)\). See also (Shore & Johnson, 1981).

Reference Priors
Let us look more closely at the reference state of information. Some argue that \(q(x)\) represents a fundamental state of knowledge involving the variable \(x\). Tarantola (1987) calls it null state of information, which implies that \(q(x)\) must be completely non-informative. Jaynes (1968) refers to such distributions as ignorance priors.

The issue of assigning non-informative prior distributions goes back to the time of Bayes and Laplace. They both used uniform priors, but even then it was clearly unsatisfactory since that assumption leads to contradictions upon a change of variables. Specifically, if one considers the uniform distribution for \(x\) as the fundamental state of knowledge and changes the variables to a different system of coordinates \(y\), the reference prior will not be the same. But if we begin with a uniform prior for \(y\), it will yield a different reference prior for \(x\), which makes no sense. Because of that, Jaynes (1968) proposed the use of group theory to define invariant priors to represent our ignorance about a given parameter.

This can be illustrated with a simple example from Jaynes (1968). If we are completely ignorant about a location and scale parameters \((\mu, \sigma)\, e.g.,\, the\ mean and standard deviation), a change of variables

\[\mu' = \mu + b \quad \text{and} \quad \sigma' = a \sigma \quad (16)\]
takes our problem into a completely equivalent one. That is, a change in variable does not contain any information about the unknown quantities. By the usual transformation equation, any distribution \(f(\mu, \sigma)\) transforms as \(g(\mu', \sigma') = a^{-1} f(\mu, \sigma)\), \(\quad (18)\).
where \( a^{-1} \) is the Jacobian of the transformation. So, if we require that the prior distribution be invariant with respect to a change of coordinates, then we must have
\[
g = f.
\]
By substituting this into (18), we get the functional equation
\[
f(\mu, \sigma) = a f(\mu + b, a\sigma).
\]
(20)
The solution for this is Jeffrey’s prior, which is given by
\[
f(\mu, \sigma) = \frac{c}{\sigma},
\]
where \( c \) is an arbitrary constant.

Thus, we can outline a general procedure for determining the invariant measure as given by

(i) Find the relevant group of transformations for the problem. In the above example, it is given by all transformations if the form of Equations (16) and (17) with \( 0 < a < \infty \) and \( -\infty < b < \infty \).

(ii) By the combination of Equations type (18) and (19) get the functional equation representing the invariance.

(iii) Solve the functional equation to find the invariant prior.

Usually, physical reasoning is used for defining an appropriate group of transformations relevant to the problem at hand and leading to a functional equation whose solution exists.

Example 19 in Chapter 1 of (Tarantola, 1987) provides an illustration of this sort of reasoning. This example describes the problem of assigning a non-informative distribution for the velocity of a non-relativistic particle. Then, if \( r \) is the Cartesian vector describing the particle’s position at time \( t \), the magnitude of the particle’s velocity is given by
\[
v = \frac{\|dr\|}{dt}.
\]
The same quantity can be represented through the Galilean transformation
\[
r' = r_0 + \alpha r,
\]
\[
t' = t_0 + bt.
\]
Using the postulate of space-time homogeneity we have the invariance of mathematical form. Thus, the particle’s velocity in the new coordinate system is
\[
v' = \frac{\|dr'\|}{dt'},
\]
which leads to \( v' = \alpha v \), where \( \alpha = \frac{b}{t} \). To avoid the contradiction described in the previous section, a non-informative distribution for both systems also should be the same, which is expressed by Equation (19). Consequently, we find a functional equation similar to Equation (20), which is given by \( f(v) = \alpha f(\alpha v) \), the solution of which is \( f(v) = \frac{c}{\alpha} \).

Some Examples of Entropy Calculations

In this section we digress briefly to show an analytic calculation of relative entropy. We shall calculate the relative entropy \( H[p(x); q(x)] \) of one uncorrelated Gaussian with respect to another. By examining the limiting cases of zero and infinite variance, we can derive the relative entropy associated with states of perfect knowledge and total uncertainty. Let \( p \) and \( q \) be two Gaussian distributions defined by
\[
p(x) = \sqrt{\frac{\sigma^2}{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]
\[
q(x) = \sqrt{\frac{\sigma^2}{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}},
\]
where \( n \) is the dimension of the vector variable \( x \), and \( \sigma_p \) and \( \sigma_p \) are the respective standard deviations. In the limit that \( \sigma \) goes to zero, these distributions converge to delta functions, corresponding to states of perfect knowledge. In the limit that \( \sigma \) goes to infinity, they converge to (zero) uniform probability distributions, corresponding to states of complete uncertainty. This is illustrated in Figure (1).

Given the definition of relative entropy it is straightforward to show that:
\[
H[p(x); q(x)] = - \int p(x) \log p(x) dx - \int p(x) \log q(x) dx
\]
\[
= - \int p(x) \left( \log \frac{1}{2\pi \sigma^2} - \frac{1}{2\sigma^2} (x^2) \right) dx
\]
positive (relative entropy always negative) and that this information is a minimum when the two distributions are the same, and a maximum when the measured state \( p \) is infinitely precise, for a fixed reference state. When this state is more informative than the one defined by \( p(x) \), the relative entropy provides inconsistent results, and an alternative information measure should be used in this case.

**Principle of maximum entropy**

Convinced that entropy is a suitable measure for the uncertainty of a probability distribution, Jaynes (1957) showed that a useful tool for conservatively assigning probabilities was to maximize the entropy of the unknown distribution subject to constraints on its moments.

Mathematically this variational problem can be expressed by maximizing Equation (14) subject to the normalization of the distribution

\[
\sum_{i=1}^{N} p(x_i) = 1,
\]

and to other constraints given in the form of expectations

\[
\sum_{i=1}^{N} w_k(x_i) p(x_i) = \mu_k, \quad k = 1, ..., K,
\]

where \( \mu_k \) is a numerical value computed from the available data.

This is equivalent to the unconstrained problem, given by

\[
S(p; \lambda, q) = -\sum_{i=1}^{N} p(x_i) \ln \frac{p(x_i)}{q(x_i)}
\]

\[
-(\lambda_0 - 1) \left[ \sum_{i=1}^{N} p(x_i) - 1 \right]
\]

\[
-\sum_{k=1}^{K} \lambda_k \left[ \sum_{i=1}^{N} w_k(x_i) p(x_i) - \mu_k \right],
\]

where the \( \lambda \) are the Lagrange multipliers associated with the constraints. Note that the term \( (\lambda_0 - 1) \) is just a redefinition of the zero-order Lagrange multiplier introduced for convenience. If we take the first variation of the functional \( S(p; \lambda, q) \) with respect to the probabilities, we get that \( \delta S(p; \lambda, q) \) equals

\[
\sum_{i=1}^{N} \left[ \frac{\partial H}{\partial p(x_i)} - (\lambda_0 - 1) - \sum_{k=1}^{K} \lambda_k w_k(x_i) \right] \delta p(x_i),
\]

with

\[
\frac{\partial H}{\partial p(x_i)} = -\left[ \ln \frac{p(x_i)}{q(x_i)} + 1 \right].
\]
The solution to the problem can be found in the usual way by letting $\delta S(p) = 0$, which yields

$$p(x_i) = q(x_i) \exp \left[ -\lambda_0 - \sum_{k=1}^{K} \lambda_k w_k(x_i) \right], \quad (30)$$

or

$$p(x_i) = Z^{-1} q(x_i) \exp \left[ -\sum_{k=1}^{K} \lambda_k w_k(x_i) \right], \quad (31)$$

with

$$Z \equiv \exp(\lambda_0) = \sum_{i=1}^{N} q(x_i) \exp \left[ -\sum_{k=1}^{K} \lambda_k w_k(x_i) \right]. \quad (32)$$

However, for a complete solution we still need to find the values for the other Lagrange multipliers and to specify the prior probability $q$.

Maximization of the continuous entropy functional is essentially the same as in the discrete case presented before. As usual we require the distribution be normalized

$$\int_{\mathbb{R}} p(x) \, dx = 1. \quad (33)$$

And there are other constraints in the form of expectations $\langle w_k(x) \rangle$

$$\int_{\mathbb{R}} w_k(x) p(x) \, dx = \mu_k, \quad k = 1, \ldots, K, \quad (34)$$

where $\mu_k$ is a numerical value known from the available data. The solution for this problem obtained in the same way as before is given by

$$p(x) = q(x) \exp \left[ -\lambda_0 - \sum_{k=1}^{K} \lambda_k w_k(x) \right], \quad (35)$$

or

$$p(x) = Z^{-1} q(x) \exp \left[ -\sum_{k=1}^{K} \lambda_k w_k(x) \right], \quad (36)$$

with

$$Z \equiv \exp(\lambda_0) = \int_{\mathbb{R}} q(x) \exp \left[ -\sum_{k=1}^{K} \lambda_k w_k(x) \right] \, dx. \quad (37)$$

Like the discrete case, the complete solution requires the determination of the Lagrange multipliers and the reference prior $q(x)$, which we investigate next.

### Computation of the Lagrange Multipliers

The specification of the maximum-entropy probability requires the determination of the Lagrange multipliers. In this section we present two approaches for such computation. The first one relies on optimization techniques for the determination of these parameters, and it is fully described by Mead and Papanicolaou (1984). The second is an extension of the technique proposed in Jumarie (1990) for the multidimensional case.

#### The Optimization Approach of Mead and Papanicolaou

Consider the one-dimensional case. As discussed before, the maximum-entropy distribution when the moments $\langle x^i \rangle = \mu_i, i = 1, \ldots, n$ and the prior distribution $q(x)$ are known, is given by:

$$p(x) = Z^{-1} q(x) \exp \left( -\sum_{i=1}^{n} \lambda_i x_i \right), \quad (38)$$

where

$$Z = \int q(x) \exp \left( -\sum_{i=1}^{n} \lambda_i x_i \right). \quad (39)$$

Mead and Papanicolaou (1984) defined the following potential function:

$$\Gamma(\lambda_1, \lambda_2, \ldots, \lambda_n) = \log Z + \sum_{i=1}^{n} \lambda_i \mu_i. \quad (40)$$

The desired set of Lagrange multipliers are the stationary points of the potential $\Gamma$, being the solution of the linear system of equations:

$$\frac{\partial \Gamma}{\partial \lambda_i} = 0 \Rightarrow \langle x_i \rangle = \mu_i, \quad i = 1, \ldots, n. \quad (41)$$

Therefore, the computation of the Lagrange multipliers can be formulated as an optimization problem, which can be solved by Newton's method. If we denote the vector of the Lagrange multipliers by $\lambda$ and the gradient of $\Gamma$ by $r$, we can write the iteration equation for Newton's method as

$$\lambda^{(n+1)} = \lambda^{(n)} - H^{-1} r. \quad (41)$$

From Equation (41), each component of $r$ is given by

$$r_i = \mu_i - \langle m^i \rangle, \quad (41)$$

which is the residual between the input sample moment and the corresponding expected value over the estimated pdf at the $n$-th iteration.

### An Alternative Approach

In the multidimensional case the maximum-entropy distribution is given by:

$$p(x) = q(x) \exp \left[ -\lambda_0 - \sum_{i=1}^{n} \lambda_i w_i(x) \right], \quad (42)$$
once the prior distribution \( g(x) \) and \( (w_i(x)), i = 1, \ldots, n \) are available. \( w(x) \) represents the expression for the moment used in the maximum-entropy computation. For example, if we are dealing with the \( m \)-th order moment about zero, \( w(x) \) is given by:

\[
w_m(x) = \prod_{i=1}^{n} x_i
\] (43)

The algorithm proposed here is a multidimensional version of the procedure described in Jumarie (1990), which is based solely on the integration by parts of the following moment integrals:

\[
\langle w_i(x) \rangle = \int w_i(x) p(x) \, dx, \quad i = 1, \ldots, n.
\] (44)

To simplify the description of the procedure, consider that the problem at hand is to find the maximum-entropy distribution when \( g(x) \) and moments up to the second-order are known, and given by:

\[
\langle x_i \rangle = \int x_i p(x) \, dx, \quad w(x) = x_i,
\]

\[
\langle x_i x_j \rangle = \int x_i x_j p(x) \, dx, \quad w(x) = x_i x_j,
\] (45)

valid for all \( i, j = 1, \ldots, n \). Cumbersome but straightforward algebra shows that if we perform an integration by parts in the first expression of Equation (46) in the variable \( x_{i+1} \), we obtain:

\[
\langle x_i \rangle = \lambda_{i+1} \langle x_i x_{i+1} \rangle + \sum_{j=1}^{n} \lambda_{i+1} \langle x_i x_{j}, x_{i+1} x_j \rangle
\]

\[
+ \lambda_{i+1} \langle x_i x_{i+1}^2 \rangle, \quad i = 1, \ldots, n - 1.
\] (46)

If we do the same with the second expression of Equation (46), now however with the integration by parts being on the variable \( x_k \), with \( k \neq i \) and \( k \neq j \), we obtain:

\[
\langle x_i x_j \rangle = \lambda_k \langle x_i x_j x_k \rangle + \sum_{i=1}^{n} \lambda_k \langle x_i x_j x_k, x_i \rangle
\]

\[
+ \lambda_k \langle x_i x_j x_k^2 \rangle, \quad i = 1, \ldots, n.
\] (47)

Equations (46) and (47) form a linear system of equations that in principle could be solved for the Lagrange multipliers. Notice that no numerical integration is required in this procedure as it is in the method previously described. The price we pay is the need for computation of higher moments. In this case, a second-order multidimensional moment problem, we have to compute up to the fourth-order moments. These moments should be obtained from the data available to the maximum-entropy problem. This is still an open topic, but our conclusion up to this point is that higher-order moments (higher than 2) are sensitive to the number of data samples (population size); therefore accurate estimates are difficult to obtain. Once one overcomes this complication, the algorithm formulated here can be used for the construction of the maximum-entropy multidimensional probability function.

**Numerical Calculation of 1D Priors Via Maximum Entropy**

To illustrate the algorithm of Mead and Papanicolaou, we will describe several calculations involving different distributions. In these examples, various probability densities are used to generate data sets with one thousand samples each (Figure 3). Then, sample moments up to the fourth-order (Figure 4) are computed and input into a routine that computes the Lagrange multipliers. The resulting maximum entropy probability densities are shown in Figure 5. The reference prior for all examples is uniform on \([x_l, x_u]\), where \( x_l \) and \( x_u \) are respectively the minimum and maximum values computed from each set of samples. We stopped the iteration of the algorithm when the moments of the maximum entropy distribution agreed with the sample moments to \(10^{-6}\) or better. In the computations performed, it took an average of six to seven iterations for the convergence of the algorithm. An alternative model was computed for comparison, using only the first two moments. This corresponds to truncated normal distributions, shown in Figure 6. When Figures 3 and 6 are compared, they show how the inclusion of the third and fourth sample moments into the computations improves the fit between the true and the maximum entropy distributions.

**Conclusions**

Maximizing the entropy of an unknown function subject to certain constraints is a powerful technique for finding the most featureless function consistent with those constraints. When applied to inference, this principle allows one to calculate the most conservative probabilistic description of prior information consistent with calculable features of that information such as sample moments. In order to make this procedure satisfy basic consistency requirements, it is necessary to abandon the notion of absolute states of information (or their negative, entropy), and instead think in terms of the information of one state relative to another. To some, the latter (reference) state of information has deep significance derived from physical invariance properties; while to others, this reference state is simply what we know about a problem before we have done an experiment to acquire new information.

We have set forth the basic principles of entropy and maximum entropy in a general way and shown that any
reasonable specification of an information measure will lead to essentially the same definition. We have shown that some of the renormalization procedures used to connect the discrete and continuous information measures are unnecessary and result from a failure to properly define the reference state of information. Finally, we have shown some applications of the principle of maximum entropy to the problem of computing a priori distributions from synthetic data as well as a technique for computing multi-dimensional maximum entropy distributions.

Acknowledgements

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, the sponsors of the Gravity and Magnetics Project, both at the Colorado School of Mines, the Shell Foundation and the Army Research Office. In addition, the second author acknowledges the support of the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq, Brazil).
Figure 5. Comparison between the exact probability density (solid line) and the distribution computed using entropy maximization subjected to moment constraints up to the fourth-order (dashed line). The reference prior is uniform on [min, max], which are computed from the samples.

References


Shore, J. E., & Johnson, R. W. 1981. Properties of


APPENDIX A: Postulates of Information

A1 Derivation of entropy

After Shannon's original derivation of the information entropy, several works were published giving alternative derivations. The majority of these derivations begin from a list of conditions that an adequate measure of information must satisfy, from which the entropy functional form was invariably achieved in a unique way. This rather informal approach led to criticisms that it could only indicate the plausibility of the entropy formula, but it did not rule out some other forms. Because of that, Shore & Johnson (1980) presented their derivation beginning from four axioms of consistency that any information-based method must satisfy. Their results pointed not only to the mathematical expression for entropy, but also to the principle of maximum entropy and minimum cross-entropy, which is the symmetric of relative entropy. In this section we will discuss the original arguments leading to the functional form of entropy given by Shannon (1948).

Consider a set of propositions $A = \{A_1, \ldots, A_N\}$, with individual probabilities $P(A_i) = p_i$. Assuming that there exists a measure of total uncertainty $H$ in a probability distribution, it must satisfy the following conditions:

(i) Continuity: $H$ must be a continuous function of the individual probabilities $p_i$, which implies that a small change in the probability distribution corresponds to a small change in entropy;

(ii) Monotonicity: for the case when all propositions are equally likely ($p_i = \frac{1}{N}$), $H\left(\frac{1}{N}, \ldots, \frac{1}{N}\right)$ is a monotonic increasing function of $N$;
Figure A1. Schematic figure illustrating the composition law. 
(a) example given in Shannon (1948) and (b) example given in Jaynes (1957) for equally likely probabilities.

(iii) Composition law: consider the case where the propositions in $A$ are grouped into $M$ other propositions each with $m_i$ elements, defining another set of propositions $B$. More precisely, $B = \{B_1, \ldots, B_M\}$, where $B_i = \{A_1, \ldots, A_{m_i}\}$ with $p_i = 1 - \sum_{j=1}^{m_i} m_j$ (with $m_0 = 0$) and $q_i = p_i + m_i - 1$. The measure $H$ computed from the probabilities directly assigned to $A$ will be the same as the weighted sum of individual measures $H$ computed from probabilities assigned to $B$ and to $A$ conditional to $B$. Thus, if we denote $P(B_i) = q_i$, we have

$$H(p_1, \ldots, p_N) = H(q_1, \ldots, q_M) + q_i H \left( \frac{p_1}{q_i}, \ldots, \frac{p_N}{q_i} \right).$$

The normalization of the probabilities in the right-hand side of Equation (A1) is considered since, after observing that $B_j$ is true with probability $q_j$, the probability that $A_i \notin B_j$ is true is given by $p_i / q_i$. One example of an application of this condition is illustrated in Figure A1 a. For this example

$$H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2} H(1) + \frac{1}{2} H(\frac{2}{3}, \frac{1}{3})$$

but $H(1) = 0$.

To find the mathematical form for the measure $H$, consider the example depicted in Figure A1 b, which begins with a set $A$ of equally likely propositions. When all the probabilities are equal, whatever turns out to be the final entropy function will be completely defined by the number of possibilities (or propositions) $N$. Thus, to simplify the notation we may define $h(N) \equiv H \left( \frac{1}{N}, \ldots, \frac{1}{N} \right)$. Following the guidelines in Condition (iii), we can define the set $B$, in this example, as composed by three propositions $\{B_1, B_2, B_3\}$, where

$$B_2 = \{A_1, A_5, A_6, A_7\} \text{ and } B_3 = \{A_8, A_9\}.$$  

The probability for each $B_i$ will be $q_i = \frac{3}{9}, q_2 = \frac{4}{9}$ and $q_3 = \frac{2}{9}$, respectively. With the above determinations, we can just write the expression equivalent to Equation (A1), for the proposed example, as given by

$$h(9) = H \left( \frac{3}{9}, \frac{4}{9}, \frac{2}{9} \right) + \frac{3}{9} h(3) + \frac{4}{9} h(4) + \frac{2}{9} h(2) \quad (A2)$$

or

$$H \left( \frac{3}{9}, \frac{4}{9}, \frac{2}{9} \right) = h(9) - \left[ \frac{3}{9} h(3) + \frac{4}{9} h(4) + \frac{2}{9} h(2) \right] \quad (A3)$$

The Equation (A3) is really the form that reflects the term that we are interested in, which is the entropy for general probabilities. To find an expression for the measure $h$, we can generalize Equation (A2) by letting $m_i$ be the number of propositions $A_i$ in each set $B_i$ and the total number of propositions in $B$ ($3 \to r$) be arbitrary, which gives

$$h(\Sigma_i m_i) = H(p_1, \ldots, p_r) + \sum_{i=1}^{r} p_i h(m_i) \quad (A4)$$

If we now let $m_i$ be just a constant $m$ we can reduce the above equation to

$$h(r m) = h(r) + h(m) \quad (A5)$$

since $p_i = \frac{m_i}{m_i}$, where $M = \Sigma_i m_i$. This equation has a solution given by

$$h(m) = K \log m \quad (A6)$$

where $K$ is a multiplicative constant corresponding to the choice of basis for the logarithmic function. If we substitute (A6) back in Equation (A4), using $K = 1$, we get

$$\log M = H(p_1, \ldots, p_r) + \sum_{i=1}^{r} p_i \log m_i$$

$$H(p_1, \ldots, p_r) = \sum_{i=1}^{r} \frac{m_i}{M} \log M - \sum_{i=1}^{r} \frac{m_i}{M} \log m_i$$

$$H(p_1, \ldots, p_n) = - \sum_{i=1}^{r} p_i \log p_i.$$  

A2 Uniqueness

To show that (A6) is the unique solution of Equation (A5), Shannon (1948) used the fact that it can be extended by induction to yield

$$h(r \circ \ldots \circ p) = h(r) + h(m) + h(o) + h(p) + \cdots \quad (A7)$$
or when all arguments are equal
\[ h(r^n) = kh(r). \] (A8)

Now, using that for \( t \) and \( s \) integer, not less than 2, and
\( n \) arbitrarily large, we can find an \( m \) such that
\[ s^m \leq t^n < s^{m+1}. \]

Taking the logarithm of the above inequality and dividing by \( n \log s \), we get
\[ \frac{m}{n} \leq \frac{\log t}{\log s} < \frac{m-1}{n} \]
or
\[ \left| \frac{m}{n} - \frac{\log t}{\log s} \right| < \varepsilon, \]
where \( \varepsilon \) is an arbitrarily small positive constant. By the
monotonicity condition of \( h \) we can write
\[ h(s^m) \leq h(t^n) < h(s^{m+1}) \]
or
\[ m h(s) \leq n h(t) < (m+1) h(s). \]

Similarly, we may divide last inequality by \( n h(s) \) to get
\[ \frac{m}{n} \leq \frac{h(t)}{h(s)} < \frac{m+1}{n} \quad \text{or} \quad \left| \frac{m}{n} - \frac{h(t)}{h(s)} \right| < \varepsilon, \]
which imply that
\[ \left| \frac{h(t)}{h(s)} - \frac{\log t}{\log s} \right| \leq 2\varepsilon. \]
The above expression is uniquely satisfied for
\[ h(t) = -K \log t. \] (A9)

### A3 Khinchin's derivation

Following Khinchin (1957), the entropy functional can also be uniquely defined from the following conditions

(i) The uncertainty associated with a finite complete scheme \( A \) takes its largest value if all the events are equally likely, which means
\[ H(p_1, \ldots, p_n) \leq H \left( \frac{1}{n}, \ldots, \frac{1}{n} \right). \]

(ii) Addition of an impossible event to a scheme does not change the amount of uncertainty.

(iii) The uncertainty in the product \( AB \) of two schemes \( A \) and \( B \) is equal to the uncertainty in scheme \( A \) increased by the uncertainty remaining in scheme \( B \) after the realization of scheme \( A \). This means that
\[ H(AB) = H(A) + H_A(B), \]
where \( H_A \) is the conditional entropy. Below we show that the conditional entropy corresponds to a weighted sum

in Shannon's original derivation discussed in the previous section.

Next, Khinchin considers that schemes \( A \) and \( B \) are divided into \( n \) and \( m \) events respectively. Thus, the product \( AB \) will have \( nm \) events with probabilities given by
\[ AB = \left( \begin{array}{cccc}
A_1B_1 & A_1B_2 & \cdots & A_1B_m \\
p_1q_1 & p_1q_2 & \cdots & p_1q_m \\
\vdots & \vdots & \ddots & \vdots \\
p_nq_1 & p_nq_2 & \cdots & p_nq_m
\end{array} \right). \] (A10)
The components of the expression given in item (iii) above will be defined as follows:

\[ H(AB) = H(p_1q_1; p_1q_2; \ldots, p_nq_m), \]
\[ H(A) = H(p_1; \ldots, p_n) \quad \text{and} \]
\[ H_A(B) = \sum_{i=1}^n p_i H(q_1; \ldots, q_i). \] (A11)

We can see that this will lead to the same equations of the previous section by explicitly writing the values for the probability for the events in (A10). In doing so, we get
\[ A_1B = \left( \begin{array}{cccc}
A_1B_1 & A_1B_2 & A_1B_3 & A_1B_4 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
A_1B_5 & A_1B_6 & A_1B_7 & A_1B_8 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{array} \right), \]
\[ A_2B = \left( \begin{array}{cccc}
A_2B_1 & A_2B_2 & A_2B_3 & A_2B_4 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
A_2B_5 & A_2B_6 & A_2B_7 & A_2B_8 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{array} \right), \]
\[ A_3B = \left( \begin{array}{cccc}
A_3B_1 & A_3B_2 & A_3B_3 & A_3B_4 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
A_3B_5 & A_3B_6 & A_3B_7 & A_3B_8 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{array} \right). \]

If we substitute those probability assignments in Equations (A11), using the condition 2, we get exactly Equation (A2), since
\[ H(p_1q_1; p_1q_2; \ldots, p_nq_m) \rightarrow h(9) \]
and
\[ H(q_1; \ldots, q_i) \rightarrow h(n_i). \]

### A4 Frequency connection

Consider the classic problem of a random experiment on the variable \( X \), which can take the values \( x_k \), for \( k = 1, \ldots, K \). Thus in \( N \) trials we have \( K^N \) possible outcomes yielding a set of observed frequencies for each possible value of \( X \) given by
This experiment can be repeated many times. Thus, we may wonder: what set of frequencies \( f_i \) can be realized in the greatest number of ways? A possible way to find this is by maximizing the multiplicity (subjected to any set of linear constraints), which is given by

\[
W(f_1, \ldots, f_K) = \frac{N!}{(Nf_1)! \cdots (Nf_K)!}. \tag{A13}
\]

Or alternatively by maximizing any monotonic function of \( W(f_1, \ldots, f_K) \). Jaynes (1982) shows that when \( N \to \infty, n_i \to \infty \) so that \( f_i \to \text{constant} \), using Stirling’s approximation for the factorial

\[
N^{-1} \log W(f_1, \ldots, f_K) \to - \sum_{i=1}^{K} f_i \log f_i.
\]

This means that for \( N \) large, by maximizing the entropy we also find the set of frequencies that can be realized in greatest number of ways.

A natural question that may arise concerns how far, in an entropy sense, other possible distributions are from the one of maximum entropy. That is, a certain fraction \( F \) of all runs of the experiment will yield distributions of frequencies with entropies in the range

\[
H_{\text{max}} - \Delta H \leq H(f_1, \ldots, f_K) \leq H_{\text{max}}.
\]

This is summarized by the Jaynes’ Entropy Concenation Theorem (Jaynes, 1982), which says that \( 2N\Delta H \) is asymptotically \( \nu \), distributed with \( \nu = K - M - 1 \) degrees of freedom, where \( K \) is the number of probabilities (frequencies) \( M \) is the number of linear constraints in the maximization problem (same as the number of multipliers (excluding \( t_0 \)) and the \( 1 \) results from the normalization constraint \( \sum f_i = 1 \). Thus, in terms of the upper tail area \( (1 - F) \), \( \Delta H \) is given by

\[
2N\Delta H = \chi^2(1 - F). \tag{A14}
\]

**APPENDIX B: Examples**

The principle of maximum entropy provides the most conservative probability distribution (i.e., the least informative) \( p(x) \) that is consistent with moments of the underlying process. In this section, we illustrate via examples the computation of such distributions when moments up to the second order are available. We will work mainly in the multidimensional continuous case. The unidimensional case can be derived from those results in a straightforward way. It is our intention to be specific about which form of the reference state of information given by \( q(x) \) is used in the entropy computations.

As mentioned before, the general expression for the maximum entropy probability density is

\[
p(x) = -\kappa q(x) \exp \left( -\lambda_0 - \lambda_1 x_1 - \lambda_{ij} x_i x_j - \cdots \right) \tag{B1}
\]

where the \( \lambda \) are Lagrange multipliers and \( \kappa \) is a normalization constant that from now on will be incorporated into \( \lambda_0 \). Here, we use Einstein’s summation convention such that \( \lambda_{ij} x_i x_j = \sum_i \sum_j \lambda_{ij} x_i x_j \). Direct substitution of Equation (B1) into the definition of \( H \) provides the expression of the entropy for this probability distribution function:

\[
H[p(x); q(x)] = \lambda_0 + \lambda_1(x_1) + \lambda_{ij}(x_i x_j) + \lambda_{ijk}(x_i x_j x_k) \ldots \tag{B2}
\]

Here, \( (\cdot) \) means the expectation with respect to \( p(x) \). Equation (B2) is helpful for the calculation of the relative entropy of the probability distributions studied in this section.

**B1 Zero-th Order Moment**

Consider the case in which the new state of information described by \( p(x) \) consists of only the zero-th order moment of \( p(x) \). That is, the only “new” information available to the problem is:

\[
\int_V p(x) dx = 1 \tag{B3}
\]

In this case, the maximum entropy distribution is given by:

\[
p(x) = q(x) \exp(-\lambda_0).
\]

Of course because \( q(x) \) is normalized the constant \( \lambda_0 \) is equal to zero. The relative entropy as given by Equation (B2) is equal to zero as expected since the probability function \( p(x) \) is equal to the reference state of information \( q(x) \).

**B2 First-order moments**

Now assume that the zero-th and the first order moments are available for the maximum entropy computation:

\[
\int_V p(x) dx = 1 \tag{B5}
\]

\[
\int_V x_i p(x) dx = \langle x_i \rangle, \quad i = 1, \ldots, n. \tag{B6}
\]

Here, \( n \) is the dimension of the process. We will assume that the prior probability is uniform within the interval \([0, X]\), as given by the following expression:

\[
q(x) = \kappa \mathcal{R} [0, X] = \left\{ \begin{array}{ll}
\kappa & \text{if } 0 \leq x_i \leq X_i \\
0 & \text{otherwise}
\end{array} \right.
\]

(\( B7 \)}
where $\kappa$ is a normalization constant. The maximum entropy distribution associated with Equations (B6) and (B5) is given by:

$$ p(x) = \mathcal{R}[0, x|x \exp [-\lambda_0 - \lambda_i x_i]], \quad (B8) $$

where $\kappa$ was incorporated into $\lambda_0$. So in this case, the maximum entropy distribution is a truncated exponential distribution. This distribution arises in the situation when $n$ independent exponential random variables are restricted to be in a given region of $\mathbb{R}^n$, specified by $\mathcal{R}[0, X]$. The normalization constant $\lambda_0$ is given by:

$$ \lambda_0 = \sum_{i=1}^{n} \log \left\{ \frac{1}{\lambda_i} \left[ 1 - \exp (-\lambda_i X_i) \right] \right\} . \quad (B9) $$

And the mean of this distribution is given by:

$$ \langle x_i \rangle = \frac{1}{\lambda_i} \left[ 1 - (1 + \lambda_i X_i) \exp (-\lambda_i X_i) \right] \quad (B10) $$

Notice that the mean and the unknown Lagrange multipliers are related in a nonlinear way. Iterative numerical procedures are likely to be the method of choice for the determination of those parameters. Once these parameters are computed, the relative entropy of the distribution given in Equation (B8), can be computed from Equation (B2), yielding:

$$ H[p(x); q(x)] = \sum_{i=1}^{n} \log \left\{ \frac{1}{\lambda_i} \left[ 1 - \exp (-\lambda_i X_i) \right] \right\} + \frac{1 - (1 + \lambda_i X_i) \exp (-\lambda_i X_i)}{1 - \exp (-\lambda_i X_i)} . \quad (B11) $$

In the situation of a less informative prior $q(x)$, i.e. for large values of the components of $X$, Equations (B9), (B10) and (B11) can be approximated by

$$ \lambda_0 = \sum_{i=1}^{n} \log \left\{ \frac{1}{\lambda_i} \right\} , \quad (B12) $$

$$ \langle x_i \rangle = \frac{1}{\lambda_i} \quad (B13) $$

$$ H[p(x); q(x)] = \sum_{i=1}^{n} \log \left\{ \langle x_i \rangle \right\} + n. \quad (B14) $$

Clearly in this case, the maximum entropy distribution (Equation (B8)) can be approximated by an exponential distribution. This example illustrates that, if only the first moments $\langle x_i \rangle$ are used and when the prior distribution $q(x)$ assumes the form of Equation (B7) the probability density function derived from the maximization of the entropy is a truncated exponential. Results will change if a different reference state of information $q(x)$ is used.

### B3 Second-order moments

Consider next the situation where moments up to the second order are available for the maximum entropy calculation:

$$ \int p(x) dx = 1. \quad (B15) $$

$$ \int x_i p(x) dx = \langle x_i \rangle, \quad i = 1, \ldots, n. \quad (B16) $$

$$ \int x_i x_j p(x) dx = \langle x_i x_j \rangle, \quad i, j = 1, \ldots, n. \quad (B17) $$

The probability distribution that maximizes the relative entropy functional is given by:

$$ p(x) = q(x) \exp (-\lambda_0 - \lambda_i x_i - \lambda_{ij} x_i x_j). \quad (B18) $$

Assuming that the reference state of information is given by an uniform probability distribution function such as

$$ q(x) = \kappa \mathcal{R}[-X, X] = \left\{ \begin{array}{ll} 1 & \text{if } -X_i \leq x_i \leq X_i \\ 0 & \text{otherwise.} \end{array} \right. \quad (B19) $$

where $\kappa$ is a normalization constant. Therefore, we have for the maximum entropy distribution

$$ p(x) = \mathcal{R}[-X, X] \exp (-\lambda_0 - \lambda_i x_i - \lambda_{ij} x_i x_j). \quad (B20) $$

As before $\kappa$ was absorbed into $\lambda_0$. This equation defines a truncated Gaussian probability distribution. In this case, it is complicated to derive explicit expressions relating the Lagrange multipliers and the parameters (mean and covariances) of this distribution, as was done in the previous section, due to the complexity of the integrations. A diagonalization of the matrix $\lambda_{ij}$ facilitates this matter but still the algebra involved is rather cumbersome in the multidimensional case. In the one-dimensional case, the problem is more tractable, and for the probability distribution given by

$$ p(x_i) = \mathcal{R}[-X_i, X_i] \exp (-\lambda_0 - \lambda_i x_i - \lambda_{ii} x_i^2), \quad (B21) $$

the following expressions can be derived:

$$ \lambda_0 = \log \Lambda, \quad (B22) $$

$$ \langle x_i \rangle = \frac{1}{2\lambda_i} \left\{ \lambda_i \left[ \exp (-\lambda_i X_i - \lambda_{ii} X_i^2) \right] - \exp (-\lambda_i X_i - \lambda_{ii} X_i^2) \right\} \quad (B23) $$

and

$$ \langle x_i^2 \rangle = \frac{1}{4\lambda_i^2} \left\{ \left[ \lambda_i - 2\lambda_{ii} X_i \right] \exp (-\lambda_i X_i - \lambda_{ii} X_i^2) - \left( \lambda_i + 2\lambda_{ii} X_i \right) \exp (-\lambda_i X_i - \lambda_{ii} X_i^2) + \lambda_i^2 + 2\lambda_{ii} \right\} \quad (B24) $$

where $\Lambda$ is given by

$$ \Lambda = \frac{\sqrt{\pi}}{2\sqrt{\lambda_i}} \exp \left( \frac{\lambda_i^2}{4\lambda_i^2} \right) \left\{ \text{Erf} \left[ \sqrt{\lambda_i} \left( \frac{\lambda_i}{2\lambda_i} + X_i \right) \right] \right\} $$
\[
- \text{Erf} \left[ \sqrt{\frac{1}{2\lambda_{11}}} \left( \frac{\lambda_{11}}{2\lambda_{11}} - X_1 \right) \right], \quad (B25)
\]

where \( \text{Erf} \) is the error function. As in the exponential case the Lagrange multipliers and the moments of the distribution are related in a nonlinear fashion. Again iterative procedure should probably be used to solve for the Lagrange multipliers in Equations (B23) and (B24).

Once they are computed, the entropy of the distribution expressed in Equation (B21) is derived from Equation (B2). The result follows
\[
H[p(x); R[-X_1, X_1]] = - \log \frac{1}{\Lambda} + \frac{\lambda_{11}}{2\lambda_{11}} \left( \frac{1}{\Lambda} \left( -\lambda_1 X_1 - \lambda_{11} X_1^2 \right) - \lambda_1 \right) \\
+ \frac{1}{4\lambda_{11}} \left( \frac{1}{\Lambda} \left( \lambda_1 - 2\lambda_{11} X_1 \right) \exp \left( -\lambda_1 X_1 - \lambda_{11} X_1^2 \right) \right) \\
- \left( \lambda_1 + 2\lambda_{11} X_1 \right) \exp \left( -\lambda_1 X_1 - \lambda_{11} X_1^2 \right) \\
+ \lambda_1^2 \left( 2\lambda_{11} \right). 
\]  

(B26)

As the components of \( \mathbf{X} \) increase, Equation (B20) approaches a multidimensional Gaussian probability distribution given by
\[
p(x) = \sqrt{\frac{\pi^{n}}{2 \det C}} \exp \left[ -\frac{1}{2} (x_1 - x_0) C_{ij}^{-1} (x_j - x_{0j}) \right]. \quad (B27)
\]

where \( n \) is the dimension of the variable \( x \), \( x_0 \) is the \( i \)-th component of the mean vector, \( C_{ij}^{-1} \) is the \( ij \)-th element of the inverse covariance matrix and \( \det C \) its determinant. In this case, a direct comparison between Equations (B20) and (B27) yields the following relationships between the Lagrange multipliers and the mean and covariances of the Gaussian distribution
\[
\lambda_0 = \frac{n}{2} \log(2\pi) + \frac{1}{2} \log \det C + \frac{1}{2} x_0 C_{ij}^{-1} x_{0j}, \\
\lambda_i = -x_{0j} C_{ij}^{-1}, \quad \text{and} \\
\lambda_{ij} = \frac{1}{2} C_{ij}^{-1}. 
\]  

(B28)

Use in Equation (B2) of the Lagrange multipliers defined in Equation (B28) yields
\[
H[p(x); q(x)] = \lambda_0 + \lambda_i (x_i) + \lambda_{ij} (x_i x_j)
\]
\[
= \frac{n}{2} \log(2\pi) + \frac{1}{2} \log \det C + \frac{1}{2} x_0 C_{ij}^{-1} x_{0j} - \\
x_0 C_{ij}^{-1} x_{0j} - \frac{1}{2} C_{ij}^{-1} (x_i x_j) \\
= \frac{n}{2} \log(2\pi) + \frac{1}{2} \log \det C + \frac{1}{2} C_{ij}^{-1} ((x_i x_j) - (x_i)(x_j)) \quad (B29)
\]

Considering the above equation that:
\[
C_{ij} = ((x_i - x_0)(x_j - x_{0j})) = (x_i)(x_j) - (x_i)(x_j), \quad (B30)
\]

we finally obtain:
\[
H[p(x); q(x)] = \frac{n}{2} (\log(2\pi) + \frac{1}{2} \log \det C + \frac{1}{2} C_{ij}^{-1} (x_i - x_0)^2) \\
= \frac{n}{2} \log(2\pi) + \frac{1}{2} \log \det C + \frac{n}{2}. \quad (B31)
\]

This equation is an approximate expression for the relative entropy of the maximum entropy distribution computed in the situation where up to the second order moments are known and the prior reference state of information is given by Equation (B19), when the components of \( \mathbf{X} \) are large. It is left to the interested reader to check that Equation (B26) reduces to Equation (B31) in the one-dimensional case when \( x_i \) gets large enough.

Figure (B1) illustrates the behavior of Equation (B31) for a multidimensional Gaussian process of dimension 200 as a function of its correlation length. As expected for smaller correlation lengths the relative entropy should be larger due to the more erratic pattern of the random sequence. As the correlation length increases the information content of the sequence increases, and so the relative entropy decreases. Notice that it is possible to have a negative relative entropy as Equation (B31) indicates. Tarantola (1987) worked out an example similar to the one just discussed, however with the reference state of information described by a multidimensional Gaussian distribution. In that case, we have for for the distributions \( p(x) \) and \( q(x) \) the following expressions:
\[
p(x) = \sqrt{\frac{(2\pi)^n}{\det C_p}} \exp \left[ -\frac{1}{2} (x_i - x_0^p)^2 C_p^{-1} (x_i - x_0^p) \right] \quad (B30)
\]

and
\[ q(x) = \sqrt{\frac{(2\pi)^n}{\det C^q}} \exp \left[ -\frac{1}{2} (x_i - x_i^q)^T C_i^q \left( x_j - x_j^q \right) \right]. \]  

(B33)

Where \( C^p \) and \( C^q \) are the covariance matrices, and \( x_i^q \) and \( x_j^q \) the \( i \)-th component of the mean vector of the probability distributions \( p(x) \) and \( q(x) \) respectively. Computation of the relative entropy given by Equation (B2) is straightforward but cumbersome. The main steps of this computation are shown in Equation (B34)

\[
H[p(x); q(x)] = - \int_V p(x) \log \left\{ \sqrt{\frac{\det C^q}{\det C^p}} \right\}
\exp \left[ -\frac{1}{2} (x_i - x_i^q)^T C_i^q \left( x_j - x_j^q \right) \right] \, dx
\]

\[
= - \int_V p(x) \left[ \frac{1}{2} \left( x_i^q C_i^q \left( x_j - x_j^q \right) + 2 x_i^q C_i^q \left( x_j - x_j^q \right) \right) \right] \, dx
\]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

\[
= -x \left[ C_i^q \left( x_j - x_j^q \right) \right] - \frac{1}{2} \frac{1}{2} x_i^q C_i^q \left( x_j - x_j^q \right) + x_i^q C_i^q \left( x_j - x_j^q \right) \]

(B34)

where

\[
\kappa = \log \sqrt{\frac{\det C^q}{\det C^p}}.
\]

In the above derivation the following relationships were used:

\[
C_{ij} = \langle x_i x_j \rangle - \langle x_i \rangle \langle x_j \rangle , \text{ and}
\]

\[
A_{ij} B_{ij} = \text{trace}(AB).
\]

(B35)

The last expression holds if \( A \) and \( B \) are symmetric matrices. Equation (B34) corrects a minor error in the derivation shown in Tarantola (1987). Figure (B2) illustrates, for a random process of dimension 100, the relative entropy derived in Equation (B34) as a function of the ratio of the correlation lengths of the probability functions \( p(x) \) and \( q(x) \). Recall that a larger correlation length implies a more informative distribution.

Note the similarity between Figures (2) and (B2). This is expected since both Figures display the relative entropy \( H[p(x); q(x)] \) in the situation where \( p(x) \) ranges from less informative (ratio of correlation lengths less than 1) to more informative (ratio of correlation lengths greater than 1) than \( q(x) \). Again one would expect that the relative entropy would decrease smoothly as \( p(x) \) gets progressively more informative and not the behavior shown in Figure (B2), for the case in which the ratio of correlation lengths is smaller than 1.

As a final remark notice that the distribution that maximizes the entropy when just the first and second moments are specified is not necessarily Gaussian as often said, although it was in the above example. As should be clear from these examples, such computation depends on the prior probability \( q(x) \).
Variation of P-wave reflectivity with offset and azimuth in anisotropic media

Andreas Rüger
Department of Geophysics
Center for Wave Phenomena
Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
Recent results on symmetry-plane P-wave reflection coefficients in azimuthally anisotropic media are extended to observations at arbitrary azimuth, large incidence angles and to lower symmetry systems. The approach presented here is based on analysis of linearized reflection and transmission coefficients in azimuthally anisotropic media. The approximate P-wave reflection coefficient in transversely isotropic media with a horizontal axis of symmetry (HTI) (typical for a system of parallel vertical cracks embedded in an isotropic matrix) shows that the AVO gradient varies as a function of the squared cosine of the azimuthal angle. This change can be inverted for the symmetry-plane directions and a combination of the shear-wave splitting parameter $\gamma$ and the anisotropy coefficient $\delta^{(V)}$. The parameters $\gamma$ and $\delta^{(V)}$ cause the same functional form of azimuthal change of the AVO gradient and cannot be extracted individually.

Azimuthal variations in the reflection coefficient at large angles of incidence (greater than 20°) can serve as an indication of azimuthal anisotropy even if no azimuthal AVO-gradient change can be observed. In this case, the azimuthal variation of the reflection coefficient at large incidence angles is primarily dependent on the anisotropy coefficient $e^{(V)}$ (closely related to Thomsen's parameter $\epsilon$) and provides enough information to detect the two symmetry-plane directions.

The reflection coefficient study is also extended to media of orthorhombic symmetry which are believed to be more realistic models for fractured reservoirs. The study shows that the orthorhombic and HTI reflection coefficients are very similar and that the azimuthal variation in the orthorhombic P-wave reflection response is a function of the shear-wave splitting parameter $\gamma$ and two anisotropy parameters describing P-wave anisotropy for near-vertical propagation in the symmetry planes.

Key words: Reflection coefficients, P-wave AVO, azimuthal anisotropy

Introduction
Fracturing and directional horizontal stress fields can cause azimuthal anisotropy in the subsurface. Knowledge of the direction of open fractures, the degree of fracturing and compartmentalization is critical in understanding the flow of fluids or gas through the reservoir and in making decisions on drilling locations and optimization of reservoir productivity.

Most upper-crustal media and many naturally fractured hydrocarbon reservoirs show azimuthal anisotropy of various types and strength (Crampin, 1985). The most simple, “first-order” model to describe azimuthal anisotropy is transverse isotropy with a horizontal axis of symmetry (HTI). The most common physical reason for a medium of HTI symmetry is a system of parallel penny-shaped vertical cracks embedded in an isotropic matrix (Thomsen, 1995). Any deviation from this model lowers the symmetry of the system. Media of orthorhombic symmetry that are believed to represent
more realistic, azimuthally anisotropic models can, for example, be caused by a combination of horizontal layering and vertically aligned cracks. Likewise, a system of two orthogonal but not necessary identical crack systems or two identical crack systems at oblique angles lead to orthorhombic symmetry (Winterstein, 1990). More general combinations of fracture systems yield monoclinic or triclinic symmetries.

While seismic signatures of compressional waves in transversely isotropic media with a vertical axis of symmetry (VTI) have been an active study area in the last decade, research on HTI media has been predominantly focused on the propagation of shear waves. Shear waves are very sensitive to the direction and amount of the fracturing and several algorithms [mainly through analysis of shear-wave birefringence (Alford, 1986; Thomsen, 1988)] are nowadays used to extract this information from multicomponent data (Martin & Davis, 1987; Lynn, 1995; Micheletti, 1995). While much attention has been devoted to the study of shear-wave splitting in HTI models, the dependence of compressional wave data on the azimuthal anisotropy is much less understood and P-wave* data are rarely used to directly detect and characterize fractured zones or “weak spots”. Certainly, a method for the direct identification of fracturing using P-wave data would be highly beneficial. The exploration community has become very sophisticated in the processing and acquisition of P-wave data; additionally, compressional data generally have better quality and are cheaper as compared to shear-wave data.

Just recently, it became better known that P-wave AVO signatures are sensitive to fractures or cracks. Mallick and Frazer (1991), for example, show on synthetic data that the P-wave AVO response from a fractured layer is azimuthally dependent. Similar azimuthal variations in the P-wave reflection response have been observed on field data by Lynn et al. (1995) and Johnson (1995).

To elucidate the relation between the P-wave reflection coefficients and the medium parameters, Rüger (1996b) derived approximate reflection coefficients for the symmetry planes of HTI media. The reflection coefficient was hereby linearized with respect to the relative changes in isotropy parameters across the reflecting boundary and new anisotropy parameters $\varepsilon^{(V)}$, $\delta^{(V)}$, and $\gamma^{(V)}$, similar to Thomsen’s (1986) coefficients. Based on his results and a new exact description of normal-moveout velocities in HTI media (Tsarkin, 1996c), Rüger and Tsarkin (1995) proposed a new AVO al-

* I am omitting the qualifiers in “quasi P*-wave and “quasi S*-wave

gorithm to improve the characterization of fractured reservoirs using P-wave data.

This paper is a continuation of the work by Rüger (1996b). Specifically, the fracture-detection algorithm based on the symmetry-plane P-wave reflection coefficients shown in Rüger (1996b) can be extended to observations at arbitrary azimuth, to variations of reflection coefficients at large incidence angles and to lower symmetry systems. While the symmetry-plane coefficients describe the magnitude of azimuthal change and show how to relate it to the medium parameters, the newly derived expressions help to apply the fracture detection algorithm for any set of azimuths and characterize the functional type of the azimuthal variation caused by the different anisotropy parameters.

A second extension investigates the reflection response in the vertical symmetry planes of orthorhombic media. Finally, for completeness and further reference, the full system of symmetry-plane reflection and transmission coefficients for pure modes in HTI media is given. Although scattering coefficients for converted modes have been derived as well, they are not shown here due to their algebraic complexity.

**Symmetry planes of HTI media: analogies and notation**

HTI media have two vertical symmetry planes. The plane formed by the symmetry axis and the vertical is called the “symmetry-axis” plane (Rüger, 1996b; Tsarkin, 1996c), while the plane perpendicular to the symmetry axis is denoted the “isotropy” plane. If the HTI symmetry is caused by vertically aligned cracks, the isotropy plane coincides with the fracture plane.

The equivalence between HTI and VTI media is the key observation to derive the polarization vectors, the phase velocities and the approximate and exact reflection coefficient in the symmetry-axis plane: the Christoffel equation that in general differs for VTI and HTI media is identical for waves propagating in the symmetry-axis plane of HTI media and vertical symmetry planes of VTI models. Consequently, all equations describing kinematic properties and polarizations are identical for transversely isotropic media with vertical symmetry axis and the symmetry-axis plane of HTI media (Rüger, 1996b; Tsarkin, 1996c).

The observed analogy also leads to the introduction of a new set of dimensionless anisotropy parameters similar to Thomsen’s (1986) coefficients. Specifically, HTI media are conveniently described by the vertical P-wave velocity $\alpha$, the vertical velocity of the shear wave polarized parallel to the isotropy plane $\beta (\neq \beta')$, and three
anisotropy parameters $\delta^{(V)}$, $\epsilon^{(V)}$ and $\gamma^{(V)}$. Anisotropy coefficients $\delta^{(V)}$, $\epsilon^{(V)}$ and $\gamma^{(V)}$ are Thomsen parameters defined with respect to vertical in the same way as in VTI media, i.e., they are different from the generic coefficients defined with respect to the horizontal symmetry axis. The set of new parameters naturally evolves from the analysis of the Christoffel system. Additionally, they represent the effective parameters describing the seismic signatures in HTI media. For example, $\alpha$ is the normal-moveout velocity recovered from seismic surveys aligned with the isotropy plane; $\delta^{(V)}$, on the other hand, determines the difference between the NMO velocity in this plane and in the symmetry-axis plane (Tsvankin, 1996c).

In some contexts, it is convenient to additionally use $\beta^\perp$, the slow vertical shear-wave velocity and the symmetry-axis velocities $V_{p0}$ and $V_{s0}$. Table 1 summarizes the coefficients used in the paper and relates them to the generic Thomsen parameters. Also included in this table is Thomsen’s parameter $\gamma$ that is directly related to $\gamma^{(V)}$, the $\gamma$-parameter of the equivalent VTI model. The former is chosen in the parameterization of the reflection coefficients below because of its importance in the classical shear-wave splitting analysis.

### Approximate scattering coefficients for HTI media

The continuity of displacement and traction between elastic layers in welded contact leads to a system of equations that can be inverted for the reflection and transmission coefficients. The forward problem thus consists in solving a system of two, four or six boundary conditions, depending on the symmetry and the number of the wave modes above and below a reflecting interface. The inverse problem of estimating medium parameters from the angular changes of the reflection response represents a much more difficult problem. Koeoef (1955) went through the laborious exercise of numerically investigating reflection coefficients for many different sets of elastic, isotropic parameters. Obviously, extending this approach to azimuthally anisotropic media is not feasible. Instead, it is more helpful to derive approximate scattering coefficients (Borthfeld, 1961; Richards & Frasier, 1976; Aki & Richards, 1980; Banik, 1987; Thomsen, 1993; Ursin & Haugen, 1996) and learn about the influence of the anisotropy parameters on the reflection response.

The derivation of $P$-wave reflection coefficients in azimuthally anisotropic media involves the study of six wave modes generated by the incident $P$-wave. Probably the first results on this topic have been performed (but not explicitly published) by Corrigan (1990), who investigated the small-angle response at isotropic/orthorhombic boundaries. Using a Born-scattering approach, he derived the dependence of the initial slope of the reflection coefficient (also denoted as AVO gradient) $B$ on source-receiver azimuth of the form

$$B = B^{(0)} + B^{(1)} \cos 2\phi,$$

consistent with observations on synthetic data by Mallick (1991; 1995).

To verify and extend Corrigan’s result to interfaces between two HTI media and higher angular terms, I used a perturbation technique similar to that applied by Thomsen (1993) for (azimuthally isotropic) VTI media. This approach considers a perturbation from an isotropic background medium and requires analytic expressions for $P$-wave polarization vectors, phase velocities and vertical slownesses as functions of the incidence phase angle $i$ and azimuthal phase angle $\phi$ (Figure 1) for weakly anisotropic HTI media.

A derivation of the polarization vectors without investigating the eigenvector problem of the HTI Christoffel matrix is shown in Appendix A. It only requires to apply the analogy of HTI and VTI media and some basic geometry to show that the $P$-wave polarization vector \( \mathbf{d} \) in HTI media can be expressed through the incidence phase angle and azimuthal phase angle as

$$\mathbf{d}(i, \phi) = \begin{pmatrix} l(i, \phi) \sin i \cos \phi \\ m(i, \phi) \sin i \sin \phi \\ m(i, \phi) \cos i \end{pmatrix},$$

where $m$ and $l$ are given as

$$l(i, \phi) = 1 - f \sin^2 i \cos^2 \phi \times \left[ \delta^{(V)} + 2(\epsilon^{(V)} + \delta^{(V)}) \sin^2 i \cos^2 \phi \right],$$

$$m(i, \phi) = f \left( 1 - \sin^2 i \cos^2 \phi \right) \times \left[ \delta^{(V)} + 2(\epsilon^{(V)} + \delta^{(V)}) \sin^2 i \cos^2 \phi \right],$$

with $f = \alpha^2/(\alpha^2 - \beta^2)$.

Phase velocities in HTI media as function of the angle with the symmetry axis and $\epsilon^{(V)}$ and $\delta^{(V)}$ are given
Table 1. Anisotropy parameters used to study HTI media and their relation to the generic Thomsen parameters. The $c_{ij}$ representation assumes that the symmetry axis is pointing in the $x_1$-direction. Parameter $f$ in the equation for $\delta^{(V)}$ is given by $f = 1 - V_{SO}^2/V_{Po}^2$.

<table>
<thead>
<tr>
<th>$c_{ij}$ notation</th>
<th>generic Thomsen notation (exact)</th>
<th>(weak-anisotropy)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha$</td>
<td>$\sqrt{\cos^2 \rho}$</td>
<td>$V_{Po} \sqrt{1 - 2\epsilon}$</td>
</tr>
<tr>
<td>$\beta$</td>
<td>$\sqrt{\cos^2 \rho}$</td>
<td>$V_{SO} \sqrt{1 - 2\gamma}$</td>
</tr>
<tr>
<td>$\beta^*$</td>
<td>$\sqrt{\cos^2 \rho}$</td>
<td>$V_{SO}$</td>
</tr>
</tbody>
</table>

\[
\delta^{(V)} = \frac{(c_{13} + \rho c_{44})^2 - (c_{11} - c_{12})^2}{2(c_{13} + \rho c_{44})^2} \delta^{(P)} \quad \epsilon^{(V)} = \frac{c_{44} - c_{11} + \rho c_{44}}{2c_{13}} \quad \gamma^{(V)} = \frac{c_{44} - c_{11} - \rho c_{44}}{2c_{13}}
\]

in Tsvankin (1996c). Expressed using incidence and azimuthal angle, the $P$-wave phase velocity for weak anisotropy yields:

\[
V_{Po}(i, \phi) = \alpha \left[ 1 + \delta^{(V)} \sin^2 i \cos^2 \phi + (\epsilon^{(V)} - \delta^{(V)}) \sin^2 i \cos^2 \phi \right].
\]  

(4)

Knowledge of the polarizations and phase velocities allows us to set up the perturbation equations for the reflection and transmission coefficients of the waves scattered at HTI-HTI interfaces with the same orientation of the symmetry axis above and below the interface. From the derivation shown in Appendix B, it follows that the compressional plane-wave reflection coefficient has the following dependence on the incidence (polar) and azimuthal phase angles:

\[
R_P(i, \phi) = \frac{1}{2} \frac{\Delta Z}{Z} \left[ \frac{\Delta C}{C} \right] \sin^2 i \tan^2 \phi + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \frac{\Delta \epsilon^{(V)}}{\epsilon^{(V)}} \right) \sin^2 i \cos^2 \phi + \frac{1}{2} \left( \frac{\Delta \beta}{\beta} - \left( \frac{2 \beta}{\alpha} \right)^2 \frac{\Delta \rho}{\rho} \right) \sin^2 i 
\]

(3)

where $Z = \rho \alpha$ is the vertical $P$-wave impedance and $G = \rho \beta^2$ denotes the vertical shear modulus. The changes in the elastic parameters are expressed as relative differences. The vertical $P$-wave velocities in the upper and lower layer, for example, can be written as functions of the average velocity $\bar{\alpha} = 1/2(\alpha_2 + \alpha_1)$ and the difference $\Delta \alpha = \alpha_2 - \alpha_1$:

\[
\begin{align*}
\alpha_1 & = \bar{\alpha} \left( 1 - \frac{1}{2} \frac{\Delta \alpha}{\bar{\alpha}} \right), \\
\alpha_2 & = \bar{\alpha} \left( 1 + \frac{1}{2} \frac{\Delta \alpha}{\bar{\alpha}} \right).
\end{align*}
\]

Corresponding expressions are defined for the shear modulus, the density and the $P$-wave impedance.

The simplicity of approximation (5) is striking, especially because no angular terms have been neglected; simple trigonometrical relations describe the anisotropic contribution as a function of azimuth. For azimuth $\phi = 90^\circ$, equation (5) reduces to the approximate reflection coefficient for the isotropy plane in HTI media (Rüger, 1996b):

\[
R_P(i, \pi/2) = \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \left( \frac{2 \beta}{\alpha} \right)^2 \frac{\Delta \rho}{\rho} \right) \sin^2 i + \frac{1}{2} \frac{\Delta \rho}{\rho} \sin^2 i \tan^2 \phi
\]

(6)

The velocities of waves propagating in the isotropy plane are angle-independent and $R_P(i, \pi/2)$ has the same form as the isotropic approximation implied by Wright (1986) and first published by Thomsen (1990). As the other approximations used in this study, equation (6) is linearized in small relative differences (a useful assumption in many exploration contests) and it is usually accurate enough for incidence angles not too close to the critical angle. Most importantly, it tells us what parameter combinations can be inverted for by investigating the intercept, gradient and higher angle terms of the reflection coefficient and hence forms the basis of conventional (isotropic) AVO analysis.

For the second vertical symmetry plane (the symmetry-axis plane) at azimuth $\phi = 0$, the linearized reflection coefficient has the following form:

\[
R_P(i, 0) = \frac{1}{2} \frac{\Delta Z}{Z} + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} - \left( \frac{2 \beta}{\alpha} \right)^2 \frac{\Delta C}{C} - 2 \frac{\Delta \gamma}{\gamma} + \Delta \delta^{(V)} \right) \sin^2 i
\]
\[
\frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \Delta \epsilon^{(V)} \right) \sin^2 i \tan^2 i . 
\]

As mentioned above, the same result can be derived by taking the known VTI approximate reflection coefficient and exploiting the analogy between VTI and HTI media. Comparing expressions (6) and (7) leads to the observation that the AVO-gradient term is varying between the isotropy plane and the symmetry-axis plane and that the difference in the shear-wave splitting parameter and the coefficient \( \delta^{(V)} \) determine the magnitude of this variation (Rüger, 1996b; Rüger & Tsvankin, 1995). Similar observations can be made by investigating the symmetry-plane shear-wave reflection response in azimuthally anisotropic media. Symmetry-plane shear-wave reflection and transmission coefficients for HTI media have been used to study the viability of shear-wave AVO (Yardley & Crampin, 1991; Rüger, 1996a) and are included, for completeness, in Appendix C.

This paper is primarily focused on the azimuthal dependence of the reflection coefficient. Before studying equation (5) in more detail, it is instructive to first compare some numerically computed exact reflection coefficients with their linearized approximations. Hereby, it is not of most importance to achieve a high numerical accuracy (in that case, more complicated approximations such as those shown in Ursin and Haugen (1996) should be used), but to see if the simple analytic approximation (5) of an otherwise incomprensibly complex reflection coefficient helps to quickly analyze the influence of anisotropy on the reflection signature. Equation (5) is linearized in nine small quantities \( \frac{\Delta \alpha}{\alpha}, \frac{\Delta \epsilon}{\epsilon}, \) etc. A total of 45 unknown quadratic terms are dropped in the derivation and it is not clear how the accuracy of the approximation depends on the medium parameters. In particular, it is of great interest to study the accuracy of equation (5) for different values of the anisotropy parameters \( \epsilon^{(V)}, \delta^{(V)} \) and \( \gamma \). Figure 2 shows the reflection coefficients evaluated at boundaries between isotropic and HTI media for azimuths of 0°, 30°, 60° and 90° and incidence angles up to 40 degrees. The parameters of the models used in this test are listed in Table 2. To perform a representative test of equation (5) and to study the sensitivity of the approximation with respect to the individual parameters, most examples shown use only one nonzero anisotropic parameter. This is done for test purposes only; in general, the anisotropy coefficients \( \epsilon^{(V)}, \delta^{(V)} \) and \( \gamma \) are not independent if the anisotropy is caused by vertical cracks.

The first test shown in Figure 2a corresponds to a reflecting HTI medium with a 10% shear-wave splitting parameter. The exact solutions (solid lines with decreasing thickness away from the symmetry-axis plane) and the approximations (dashed) are in very good agreement.

<table>
<thead>
<tr>
<th>( \frac{\Delta \alpha}{\alpha} )</th>
<th>( \frac{\Delta \epsilon}{\epsilon} )</th>
<th>( \frac{\Delta \delta}{\delta} )</th>
<th>( \epsilon^{(V)} )</th>
<th>( \delta^{(V)} )</th>
<th>( \gamma )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model a</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>0</td>
<td>0.1</td>
</tr>
<tr>
<td>Model b</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Model c</td>
<td>0.1</td>
<td>0.1</td>
<td>0</td>
<td>-0.1</td>
<td>0</td>
</tr>
<tr>
<td>Model d</td>
<td>0.1</td>
<td>0.1</td>
<td>-0.05</td>
<td>-0.05</td>
<td>0.15</td>
</tr>
</tbody>
</table>

Table 2. Models used to test the accuracy of equation (5) and the sensitivity of the approximate reflection coefficients. The upper medium is isotropic, the lower medium has the following isotropic parameters: compressional vertical velocity \( \alpha_2 = 2.5 \), faster shear-wave vertical velocity \( \beta_2 = 1.5 \) and density \( \rho_2 = 2.7 \).

for all azimuths. The approximations practically coincide with the exact curves for 0° and 30° azimuths with the symmetry axis, while a small deviation is visible for azimuths 60° and 90° and large incidence angles. The second model has a nonzero coefficient \( \delta^{(V)} \) in the reflecting medium. Equation (3) correctly predicts the split of the reflection coefficient curves. The slope of the exact curves for 0° and 30° azimuths is slightly overcorrected by the approximation and leads to deviations for large angles of incidence. Model c has a nonzero value of parameter \( \epsilon^{(V)} \). According to equation (5), there should be no significant change in the AVO gradient (incidence angles < 20°) and indeed, the exact solution does not show any split in the reflection coefficient curves for low angles of incidence. Note that inaccuracies for 0° and 30° azimuths are smaller than for 90° azimuth. The latter approximation coincides with the classical isotropic linearization of the reflection coefficient that is used successfully in conventional AVO analysis. Therefore, in this case, the accuracy of the anisotropic approximation is higher than that for the isotropic one. Model d has three nonzero anisotropy parameters, including a shear-wave splitting coefficient of 15%. The accuracy of the approximation is acceptable for small values of incidence angles and also shows the correct trend for higher incidence angles. Deviations for large angles of incidence and 0° and 30° azimuths are caused by the approximation of shear-wave velocity \( \beta^{-} \) through \( \beta \) and \( \gamma \) and hence are caused by the design of the approximation. For a better accuracy at small azimuths, the approximation (3) should be rewritten in terms of \( \beta^{-} \).

**Functional type of the azimuthal variation**

Rather than simply studying the reflection coefficient as a function of the incidence angle for several azimuths, equation (5) additionally suggests to fix the incidence
Figure 2. Reflection coefficients for an isotropic layer overlying an HTI medium. Shown are the exact solution (solid lines) and the weak elastic, weak anisotropic approximation (dashed) [equation (5)] for 0°, 30°, 60° and 90° azimuth. Table 2 lists the model parameters.
angle and analyze the reflection coefficient as a function of azimuth. These curves are shown in Figure 3, for the models given in Table 2. The exact reflection coefficients are shown as solid lines for 10°, 20°, 30° and 40° incidence angles while the approximations are shown as dashed curves. Clearly, the approximations are very accurate for all azimuths and 10° and 20° incidence angles. For higher angles, the approximations start to deviate from the exact result, but correctly predict the functional type of the azimuthal change. Specifically, observing the azimuthal change of reflection coefficients for a fixed incidence phase angle clearly shows that the symmetry-plane directions at 0° and 90° azimuth coincides with the extrema of the reflection-coefficient curves.

Study of azimuthal changes of reflection coefficients such as shown in Figure 3 can help to find the orientation of the natural coordinate system of the subsurface without any prior knowledge of the medium parameters. Moreover, because the location of the extrema is the same for all incidence angles, the analysis of stacked data for several azimuths is enough to invert for the symmetry-plane directions, provided that the stack is performed with the proper (azimuthally varying) normal-moveout velocity. Amplitude analysis on stacked data, however, is not sufficient to decide which azimuth corresponds to the
symmetry-axis plane and which to the isotropy plane. For example, the maximum of the reflection curves occurs at $0^\circ$ azimuth in Model a, while it is located at $90^\circ$ azimuth in Model b.

The curves generated for Model c show several features that are not observed for the other examples. First, no azimuthal variation in the reflection coefficient is visible for $10^\circ$ incidence angle and only a very slight variation at $20^\circ$ incidence angle. Additionally, the functional type of the variation is different for large angles of incidence. As predicted by the $\cos^3 \phi$ term in equation (5), the anisotropic part of the reflection coefficient rapidly decays away from the symmetry-axis plane while the azimuthal change is much less pronounced close to the isotropy-plane (Figure 6). As for the other examples, while the high-incidence-angle approximations start to deviate from the exact solution, the functional type of the azimuthal variation is nonetheless accurately approximated.

This discussion of the amplitude variations with offset and azimuth is still primarily qualitative. To relate the observed changes more closely to the medium parameters, let us now consider the implications of equation (5) to AVO-with-azimuth analysis.

Studies on AVO inversions based on the symmetry-plane coefficients (6) and (7) in Rüger and Tsvankin (1995) did not make use of the smooth azimuthal variation of the gradient and higher-angle $(\sin^2 \phi \tan^2 \phi)$ terms between the two vertical symmetry planes. Moreover, while it is clear that $\Delta \gamma$ and $\Delta \delta^{(V)}$ are responsible for the magnitude of the azimuthal variation, it has not been obvious whether both coefficients cause a different rate of azimuthal change and hence can be inverted separately. Finally, in situations with small or non-existing AVO gradient variations, it is uncertain if the higher-angle term provides usable and extractable information about the subsurface. Below, I will use equation (5) to elucidate the character of the azimuthally varying reflection coefficient and discuss the feasibility of parameter estimation from variations of the AVO gradient and the higher-angle reflection amplitudes.

The AVO gradient term

Equation (5) shows that azimuthal anisotropy in HTI media causes a different azimuthal dependence for the AVO gradient and the higher-angle term.

Small angular variations of $R_P(i, \phi)$ are described by the AVO gradient $B$ composed of the azimuthally invariant part $B^{\text{iso}}$ and the anisotropic contribution $B^{\text{ani}}$ multiplied with the squared cosine of the azimuthal angle $\phi$ with the symmetry axis. If the symmetry-axis orientation is unknown, $\phi$ should be formally expressed by the difference between the azimuthal direction $\phi_k$ of the $k$-th observed azimuth and the direction of the symmetry-axis plane $\phi_{\text{sym}}$. The AVO gradient measured at azimuth $\phi_k$ thus can be written as

$$B(\phi_k) = B^{\text{iso}} + B^{\text{ani}} \cos^2(\phi_k - \phi_{\text{sym}}).$$

(8)

This equation is nonlinear with three unknowns ($B^{\text{iso}}$, $B^{\text{ani}}$, and $\phi_{\text{sym}}$). If the direction of the symmetry axis is known (for example from $S$-wave surveys), the equation becomes linear and two independent measurements suffice to solve for $B^{\text{iso}}$ and $B^{\text{ani}}$. The special case of inverting AVO measurements in the isotropy and symmetry-axis planes has been discussed in Rüger and Tsvankin (1995). In more realistic settings with noisy data, many more azimuths should be sampled and a least-squares approach could be used to find the optimum values of the unknowns in equation (8). Due to the nonlinearity of equation (8), the solution will not be unique and will yield a set of two orthogonal symmetry-plane orientations. However, a simple rough estimate of the sign of $B^{\text{ani}}$ or a-priori knowledge of the approximate symmetry-axis direction is enough to identify which direction corresponds to the isotropy and the symmetry-axis plane.

If $B(\phi_k)$ is known for several $\phi_k$'s, a graphical approach to determine the symmetry-plane directions and the magnitude of the variations can be used. If $B$ is plotted as a function of $\phi_k$, the locations of the extrema correspond exactly with the symmetry-plane orientations at $0^\circ$ and $90^\circ$ azimuth (and multiples of $90^\circ$). The values of $B$ at the extrema yield estimates of $B^{\text{iso}}$ and $(B^{\text{iso}}+B^{\text{ani}})$.

The dashed lines in Figure 4 shows the approximate AVO gradient based on equation (8). The "exact" gradient (solid line) is computed by simply averaging the slopes of the exact reflection coefficients for the first $20^\circ$ incidence angles. Limiting the analysis to a smaller range of incidence angles would have produced a better match between the two curves, but it is unlikely that enough traces are available within this small angular range in a field data example. The curves shown in Figure 4 are evaluated for the models shown in Table 2. The extrema denoting the symmetry-plane directions can be easily picked, even though a-priori information is needed to distinguish between the symmetry-axis and the isotropy plane. Table 3 compares the magnitude of the observed AVO-gradient change with the the value of

$$B^{\text{ani}} = 1/2 \left[ \Delta \delta^{(V)} + 2 \left( \frac{25}{a} \right)^2 \Delta \gamma \right],$$

(9)

that would be obtained using equation (5).

Let us now assume that the solid curve is observed in a field experiment and that geologic data indicates that $0^\circ$ and $90^\circ$ azimuth corresponds to the symmetry-axis
plane and the isotropy-plane, respectively, and that an approximate value of \( \frac{\beta}{\alpha} \) is known. In this case, it is possible to invert for the parameter combination of \( \delta^{(V)} \) and \( \gamma \) in the lower medium. For example, if it is known that \( \gamma \) is the only nonzero anisotropy parameter for Model a, the interpretation of the measured value of \( B_{\text{observed}}^{\text{ani}} \) using equation (9) yields an estimated value of \( \gamma_{\text{estimated}} = 0.09 \) as compared to the exact value of \( \gamma_{\text{exact}} = 0.1 \). Similarly, for Model b, we would obtain a value of \( \delta^{(V)}_{\text{estimated}} = -0.11 \) as compared to the exact solution \( \delta^{(V)}_{\text{exact}} = -0.1 \). The azimuthal variation of the AVO gradient in Model c is very small; additionally, the azimuthal variation observed in Figure 4c correctly indicates that this change has a \( \cos^2 \phi \)-dependence and is caused by a nonzero value of \( \epsilon^{(V)} \) in the higher-angle term. Finally, if the value of \( \delta^{(V)} \) is known for Model d (for example from normal-moveout analysis), \( \gamma \) would be have an estimated value of \( \gamma_{\text{estimated}} = 0.124 \) as compared to the exact value of \( \gamma_{\text{exact}} = 0.15 \).

If \( B^{\text{iso}} + B^{\text{ani}} \) and \( B^{\text{iso}} \) are of the same sign and \( B^{\text{ani}} < B \), an alternative graphical interpretation is possible and the value of \( B(\phi_k) \) can be assigned to the length of a radius vector at different azimuths \( \phi_k \). The tip of this vector then delineates a curve that closely resembles an ellipse with semi-axis aligned with the symmetry-plane directions. The deviation \( \Delta B = (B(\phi_k) - B(\text{ellipse} \phi_k)) \) from the ellipse can be expressed.
in the form:

\[ \Delta B(B_k) \approx -1/2 \left( \frac{B_{\text{ani}}^{\text{sym}}}{{B}} \right)^2 \times \sin^2(\phi_k - \phi_{\text{sym}}) \cos^2(\phi_k - \phi_{\text{sym}}). \]  

Equation (10) shows that the mismatch between the ellipse and the approximate AVO gradient predicted in equation (5) is a function of \( B_{\text{ani}}^{\text{sym}} \) and \( B_{\text{iso}} \) and obtains its maximum value if the angle \( (\phi_k - \phi_{\text{sym}}) = 45^\circ \). Under the above assumptions, fitting the azimuthally varying AVO gradient with an ellipse is therefore an accurate algorithm to determine the fracture direction and the parameter combinations \( B_{\text{iso}} \) and \( B_{\text{ani}} \). Specifically, the semi-axes will have the lengths \( B_{\text{iso}} \) and \( B_{\text{iso}} + B_{\text{ani}} \) and point towards the symmetry-directions of the medium. Figure 5 shows one example for this approach.

There are several alternative ways to represent \( R_P(t, \phi) \). For example, equation (5) can be shown to be consistent with Corrigan’s (1990) approximation [equation (1)], with

\[ B^{(0)} = \frac{\Delta \alpha}{\alpha} - \left( 2 \frac{\Delta \phi}{\alpha} \right)^2 \Delta G \Delta \gamma + 1/2 \Delta \delta^{(V)} \]

\[ B^{(i)} = \left( 2 \frac{\Delta \phi}{\alpha} \right)^2 \Delta \gamma + 1/2 \Delta \delta^{(V)}. \]  

(11)

In exploration situations, it may happen that \( \delta^{(V)} \), the fracture-perpendicular shear velocity is known instead of \( \beta \). To achieve a higher accuracy, the AVO gradient in HTI media should then be rewritten through \( \beta^{(V)} \):

\[ B(\phi) = 1/2 \left( \frac{\Delta \alpha}{\alpha} - \left( 2 \frac{\beta^{(V)}}{\alpha} \right)^2 \Delta G \Delta \gamma + \Delta \delta^{(V)} \cos^2 \phi - 2 \left( 2 \frac{\beta^{(V)}}{\alpha} \right)^2 \Delta \gamma \sin^2 \phi \right). \]  

(12)

Finally, one of the most important lessons learned from the studies of the approximate AVO gradient is that the difference in the shear-wave splitting parameters \( \Delta \gamma \) and \( \Delta \delta^{(V)} \) cannot be determined separately from AVO gradient measurements in HTI media because both parameters influence the azimuthal AVO-gradient signature in the same way. Only in situations where \( \delta^{(V)} \) and \( \gamma \) are related (such as for tight formations) and \( \beta^{(V)}/\alpha \) is known can the shear wave splitting parameter be determined directly from the azimuthal dependence of the AVO response. In all other cases, \( \delta^{(V)} \) needs to be determined by independent measurements. One possibility is to relate differences in stacking velocities in the symmetry-axis plane or for a known azimuth to the vertical velocities measured in VSP experiments to \( \delta^{(V)} \). Another approach is to extract \( \delta^{(V)} \) from the azimuthal difference in NMO velocities (Tsvankin, 1996c).

Azimuthal variation of the higher-angle term

For incidence angles \( i > 20^\circ \), the AVO gradient term and the higher-angle term will influence the reflection coefficient. For azimuths close to the symmetry axis and large incidence angles, the coefficient \( \delta^{(V)} \) will have a significant impact on the azimuthal variations.

Here, we will discuss the interesting case of reflection coefficients with small or even nonexistent AVO gradient variations, i.e., the term \( \Delta \delta^{(V)} + 2 \left( \frac{\Delta \phi}{\alpha} \right)^2 \Delta \gamma \) being negligible small. In this case, equation (5) still predicts an azimuthal change in the reflection coefficient at large angles of incidence for nonzero values of \( \Delta \epsilon^{(V)} \) and \( \Delta \delta^{(V)} \). Note that both parameters have a different influence on the azimuthal change of the reflection coefficient. \( \Delta \delta^{(V)} \) has its major impact on the large-angle reflection response at an azimuth of \( 45^\circ \). \( \Delta \epsilon^{(V)} \), the parameter that
Figure 6. The influence of the anisotropy parameters on the reflection response depends on the azimuthal angle with the symmetry-axis plane. The AVO gradient changes as a function of the squared cosine of azimuth. The higher-angle \((\sin^2 \Theta \tan^2 \Theta)\) term contains two parameters with different azimuthal dependences of \(\cos^2 \phi\) and \(\sin^2 \phi \cos^2 \phi\).

Azimuthal variations of the transmission coefficient

A complete treatment of AVO in vertically stratified media and borehole studies in azimuthally anisotropic media include the investigation of transmission phenomena. The derivations shown in Appendix B yield the following linearized P-wave transmission coefficient at interfaces between two weakly anisotropic HTI media:

\[
T_P(i, \phi) = 1 - \frac{1}{2} \frac{\Delta Z}{Z} + \frac{1}{2} \left\{ \frac{\Delta \alpha}{\alpha} + \Delta \delta^{(V)} \cos^2 \phi \right\} \sin^2 i + \frac{1}{2} \left\{ \frac{\Delta \alpha}{\alpha} + \Delta \epsilon^{(V)} \cos^2 \phi \right\} \sin^2 i \tan^2 i + \left( \Delta \epsilon^{(V)} - \Delta \delta^{(V)} \right) \cos^2 \phi \cdot \sin^4 i. \tag{13}
\]

Unlike the linearized reflection coefficient, \(T_P(i, \phi)\) shows no dependence on the shear-wave splitting parameter or the shear-wave vertical velocity. The difference in \(\delta^{(V)}\) is the only term responsible for the azimuthal variations in the gradient term. For higher angles of incidence, two terms \((\sin^2 i \tan^2 i)\) with complicated azimuthal dependence of the anisotropic parameters become increasingly important.

Orthorhombic media

HTI media are useful models to study the first-order influence of azimuthal anisotropy. More realistic models can be described by orthorhombic symmetry systems. Wave propagation in orthorhombic media is rather complex and illustrations and analysis of wavefronts and slowness surfaces in orthorhombic media can, for example, be found in (Musgrave, 1970; Helbig, 1994; Schoenberg & Helbig, 1995).

Although this investigation is valid for orthorhombic media of any origin, it is instructive to visualize an orthorhombic model as a combination of a VTI model with a system of vertical cracks such as seen in Figure 7. Also sketched in Figure 7 are the two vertical symmetry planes (symmetry plane \([x_1, x_2]\) and symmetry plane \([x_2, x_3]\)). Incident waves confined to one of these symmetry planes generate just four out of the six wave modes that are created by waves incident outside of the \([x_1, x_2]\) and \([x_2, x_3]\) plane.

Let us now consider a P-wave incident in the \([x_2, x_3]\) plane at oblique angles with vertical. Because all four scattered modes have polarization directions parallel to
The crack planes (Figure 7), they essentially do not sense the cracks and propagate exactly as in VTI media. Consequently, reflection and transmission in the \([x_2, x_3]\) symmetry plane of orthorhombic media is analogous to the known VTI/VTI energy-partitioning problem.

The reflection problem looks more complex in the second symmetry plane \([x_1, x_3]\). Incident and generated waves are probing both the cracks and the horizontal layering and it is less straightforward to predict the structure of the scattering coefficients. However, recent studies on the analogy of HTI and VTI symmetry planes show that vertically fractured rocks can be described by an equivalent VTI medium, and it may be expected that the combination of vertical and horizontal layering effectively yields a system of VTI symmetry in the \([x_1, x_3]\) plane. Based on these observations, we suggest that VTI equations are entirely valid to describe the reflection and transmission coefficients in both vertical symmetry planes of orthorhombic media.

**Analysis of symmetry-plane Christoffel systems**

A more rigorous mathematical analysis of symmetry-plane reflection responses is possible by studying the corresponding Christoffel equations for orthorhombic media in the same way as shown in Rüger (1996b) for HTI models. This study (that is not repeated here) proves that the previous observations have been correct and that symmetry-plane propagation in orthorhombic media can be completely described by known VTI equations. In other words, all observations based on the equivalence of VTI and HTI media (Rüger, 1996b; Tsvankin, 1996c) remain valid for symmetry-plane propagation in orthorhombic models:

- Equations describing kinematic properties and polarizations are identical for VTI and orthorhombic models if stated in \(c_{ij}\)-tensor notation (and appropriate substitutions are made).
- Equations stated in Thomsen notation are identical, but the generic Thomsen parameters of the equivalent VTI model have to be replaced by parameters defined with respect to the vertical \((x_3)\) axis.
- Boundary value problems such as continuity of stress and strain that are expressed in Cartesian coordinates are identical because no rotation of coordinates is necessary.

Application of these rules helps to describe seismic signatures in the \([x_1, x_3]\) plane of rocks with orthorhombic symmetry by means of already known VTI equations. Published VTI phase-velocity equations in \(c_{ij}\)-notation, for example, exactly expresses phase velocity in orthorhombic media. Additionally, the exact reflection and transmission coefficients for boundaries between VTI, HTI, orthorhombic and any of its combination can be computed by simply using Graebner’s (1992) algorithm for VTI systems, provided that the natural coordinate frame has the same orientation on both sides of the interface.

As indicated above, a new set of Thomsen’s parameters has to be introduced in the same way as previously shown for HTI media. Anticipating that a similar procedure will be necessary for the \([x_2, x_3]\) symmetry plane, a convenient notation for the equivalent VTI coefficients is (Tsvankin, 1996a)

\[
\delta^{(2)} = \frac{(c_{13} + c_{53})^2 - (c_{15} - c_{55})^2}{2 c_{11} (c_{15} - c_{55})},
\]

\[
\varepsilon^{(2)} = \frac{c_{11} - c_{55}}{2 c_{13}}.
\]

Following the simple recipe of replacing \(\delta\) and \(\varepsilon\) with \(\delta^{(2)}\) and \(\varepsilon^{(2)}\) yields the approximate \(P\)-wave reflection coefficient in the \([x_1, x_3]\) symmetry-plane:

\[
R_{p}^{[x_1, x_3]}(\iota) = \frac{1}{2} \frac{\Delta Z}{Z} + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} - \left( \frac{2 \beta^2}{\alpha^2} \right) \frac{\Delta \sigma}{\sigma} + \Delta \delta^{(2)} \right) \sin^2 \iota + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \Delta \varepsilon^{(2)} \right) \sin^2 \iota \tan^2 \iota, \tag{14}
\]

with \(\beta^2 = \sqrt{c_{55}/\rho}\). At the expense of reducing numerical accuracy, equation (14) can also be written as a function of the vertical shear-wave velocity \(\beta = \sqrt{c_{44}/\rho}\) and the
well-known shear-wave splitting parameter $\gamma = \frac{c_{44} - c_{66}}{2c_{66}}$:

$$R_p^{[z_1,z_2]}(i) = \frac{1}{2} \frac{\Delta Z}{Z} + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} - \left(2\frac{\beta}{\alpha}\right)^2 \frac{\Delta G}{G} - 2\Delta \gamma + \Delta \delta^{(2)} \right) \sin^2 i + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \Delta \epsilon^{(2)} \right) \sin^2 i \tan^2 i. \quad (13)$$

Basically the same observations can be made by comparing the $[z_2,z_3]$-plane Christoffel equations of orthorhombic and VTI media, except that the equivalence is only achieved by substituting

$$c_{11} \rightarrow c_{22}, \quad c_{33} \rightarrow c_{44}, \quad c_{13} \rightarrow c_{23}.$$ Hence, the $[z_2,z_3]$ plane of orthorhombic media can be described by a second equivalent VTI medium and it is possible to introduce the Thomsen parameters $\epsilon^{(2)}$ and $\delta^{(1)}$ for the $[z_2,z_3]$ plane:

$$\delta^{(1)} = \frac{(c_{23} + c_{44})^2 - (c_{33} - c_{44})^2}{2c_{33}(c_{33} - c_{44})}$$

$$\epsilon^{(2)} = \frac{c_{22} - c_{33}}{2c_{33}}.$$ Finally, substitution of $\delta^{(1)}$ and $\epsilon^{(2)}$ yields the approximate reflection coefficient for the $[z_2,z_3]$ symmetry plane:

$$R_p^{[z_2,z_3]}(i) = \frac{1}{2} \frac{\Delta Z}{Z} + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} - \left(2\frac{\beta}{\alpha}\right)^2 \frac{\Delta G}{G} + \Delta \delta^{(1)} \right) \sin^2 i + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \Delta \epsilon^{(2)} \right) \sin^2 i \tan^2 i. \quad (16)$$

The introduction of the set of new Thomsen parameters $\delta^{(1)}, \delta^{(2)}, \epsilon^{(1)}, \epsilon^{(2)}$ has advantages not only for the study of reflection coefficient, but is generally useful in describing seismic signatures within and outside of the symmetry planes (Tsvankin, 1996a). To completely describe orthorhombic symmetry systems, we need a total of seven anisotropy coefficients together with two vertical velocities to replace the nine independent coefficients in the orthorhombic stiffness matrix. One choice of anisotropy parameters is shown in Table 4, together with their relation to the generic Thomsen parameters and parameters $\delta^{(V)}$ and $\epsilon^{(V)}$ introduced for HTI models. A more detailed discussion of effective parameters in orthorhombic media is out of the scope of this paper and is presented in Tsvankin (1996a).

**Azimuthal variation of orthorhombic reflection coefficients**

The discussion above shows that the approximate reflection coefficients in the two vertical symmetry planes of orthorhombic media are identical to the HTI symmetry-axis plane reflection coefficient. This result makes it possible to use the difference in $P$-wave reflection coefficients in the two vertical symmetry planes of orthorhombic media in the inversion for the shear-wave splitting parameter. To eliminate the “isotropic quantities” in equations (15) and (16), I follow the approach suggested by Rüger and Tsvankin (1995) of taking the difference between the two coefficients:

$$R_p^{[z_1,z_2]} - R_p^{[z_2,z_3]} = \left( \frac{\beta}{\alpha} \right)^2 \Delta \gamma_2 + \frac{1}{2} \left( \Delta \delta^{(2)} - \Delta \delta^{(1)} \right) \sin^2 i + \frac{1}{2} \left( \Delta \epsilon^{(2)} - \Delta \epsilon^{(1)} \right) \sin^2 i \tan^2 i. \quad (17)$$

For an isotropic overburden, equation (17) shows that the difference in the AVO gradient depends on just three anisotropic parameters - the shear-wave splitting parameter $\gamma_2$ and two new parameters $\delta^{(2)}$ and $\delta^{(1)}$ describing the anisotropy for near-vertical $P$-wave propagation in the symmetry planes. Tsvankin (1996a) shows that it is possible to obtain the $\delta^{(1)}$ and $\delta^{(2)}$ parameters from shortspread $P$-wave moveout analysis in the vertical symmetry planes combined with an estimate of the vertical velocity; alternatively, these parameters can be estimated by the difference between the stacking velocities and the vertical velocities measured in a VSP experiment.

The approximate (low angle) reflection coefficient for an isotropic overburden and a reflecting orthorhombic layer has been derived by Corrigan (1990). Represented

<table>
<thead>
<tr>
<th>Orthorhombic</th>
<th>VTI</th>
<th>HTI</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\delta^{(2)} = \frac{(c_{22} - c_{33})^2 - (c_{33} - c_{44})^2}{2c_{33}(c_{22} - c_{33})}$</td>
<td>$\delta^{(V)}$</td>
<td>$\delta^{(V)}$</td>
</tr>
<tr>
<td>$\delta^{(1)} = \frac{(c_{22} - c_{33})^2 - (c_{11} - c_{33})^2}{2c_{11}(c_{22} - c_{33})}$</td>
<td>$\delta^{(V)}$</td>
<td>$0$</td>
</tr>
<tr>
<td>$\gamma^{(2)} = \frac{c_{22} - c_{33}}{2c_{33}}$</td>
<td>$\gamma^{(V)}$</td>
<td>$-\gamma^{(V)}$</td>
</tr>
<tr>
<td>$\gamma^{(1)} = \frac{c_{22} - c_{33}}{2c_{33}}$</td>
<td>$\gamma^{(V)}$</td>
<td>$0$</td>
</tr>
</tbody>
</table>

**Table 4.** Seven anisotropy parameters (plus two vertical velocities) completely describe orthorhombic anisotropy. Their relations to the generic (VTI) Thomsen coefficients and to HTI parameters $\delta^{(V)}$ and $\epsilon^{(V)}$ are shown in the second and third column. The relation for $\delta^{(2)}$ and $\gamma^{(2)}$ are approximate, the other relations are valid for any strength of anisotropy.
in the new parameterization, his solution implies the following form of the approximate low-angle reflection coefficient for arbitrary azimuth

\[ R_P(i, \phi) = \frac{1}{2} \frac{\Delta Z}{Z} + \frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} - \left( \frac{2 \beta}{\alpha} \right)^2 \frac{\Delta G}{G} + \left[ \delta_1^{(2)} + 2 \left( \frac{2 \beta}{\alpha} \right)^2 \gamma_2 \right] \cos^2 \phi + \delta_2^{(1)} \sin^2 \phi \right) \sin^2 i. \] (18)

Discussion and conclusions

Analysis of reflection coefficients in azimuthally anisotropic media shows that the magnitude and the functional type of the azimuthal variation of P-wave reflectivity can be inverted for the anisotropy parameters and the symmetry-plane directions of the subsurface. The examples shown in this paper show that the derived approximate reflection coefficients closely describe the exact reflection response. However, rather than achieving a high numerical accuracy with the new approximations, it is more important in this study to establish a physical foundation for (exact) numerical inversion algorithms and create simple dependencies that interpreters can use to quickly evaluate the importance of anisotropy for their project. To obtain a reasonably simple representation of the otherwise incomprehensibly complex reflection coefficients, a perturbation approach has been used in the derivation. The main assumption in this reflection-coefficient study - small jumps in the elastic parameters across the reflecting interface and weak anisotropy - are geologically reasonable and proved useful in many exploration contexts. Although the approximations may help to detect pronounced anomalies, they fail in any quantitative analysis at interfaces with a large contrast in the elastic parameters and strong anisotropy. The HTI approximations shown will also be inaccurate if the natural coordinate systems of the incidence and reflected layer differ significantly in orientation, i.e., when the axis of symmetry changes abruptly across the boundary.

Within the limits of these assumptions, the approximate P-wave reflection coefficient for HTI interfaces predicts a change of the AVO gradient as a function of the squared cosine of azimuth. This change can be inverted for the symmetry-plane directions and a combination of the shear-wave splitting parameter \( \gamma \) and the anisotropy coefficient \( \delta^{(V)} \). Coefficients \( \gamma \) and \( \delta^{(V)} \) cause the same functional form of the azimuthal variation in the AVO gradient and cannot be extracted individually.

The two symmetry-plane directions can even be extracted in situations where no azimuthal AVO-gradient change can be observed. In this case, the azimuthal variation at large incidence angles is primarily dependent on the anisotropy coefficient \( \delta^{(V)} \) (closely related to Thomsen’s parameter \( \epsilon \)) and provides enough information to detect the natural coordinate system of the subsurface. Certainly, realistic models of fractured reservoirs deviate from simple HTI media, and it is interesting whether the reflection response in media of lower symmetry differs significantly from that in the HTI case. This study shows that although wave propagation is significantly more complex in orthorhombic media, the approximate reflection coefficient basically has the same form as in HTI models. The main observation of the study of reflection coefficients in orthorhombic media is that the difference between the P-wave reflection response in the vertical symmetry-planes is a function of the shear-wave splitting parameter \( \gamma \) and two new parameters describing the anisotropy for near-vertical P-wave propagation in the symmetry planes.

The practical implementation of the newly proposed AVO inversion faces challenges similar or even greater than for conventional AVO analysis. Prerequisite for any useful AVO-with-azimuth investigation is a proper processing sequence that preserves azimuthally varying amplitude signatures. Moreover, a profound understanding of lateral inhomogeneities is required. Wave propagation in an anisotropic overburden has a significant impact on the amplitude variations with offset and azimuth (Tsvankin, 1995; Rüger & Tsvankin, 1995) and need to be included in any meaningful AVO analysis. Other difficulties include thin layering, curved reflectors or dipping symmetry-axes. These issues are not within the scope of this investigation; however, they will be addressed in a forthcoming paper on AVO for fractured reservoirs.

Acknowledgments

Thanks to Ilya Tsvankin for guidance, discussions and support. Vladimir Grechka reviewed this manuscript and Björn Rommel examined Appendix A. Dennis Corrigan was kind enough to share his programs, derivations and time. Jack Cohen’s Mathematica-package ATools.m was very helpful in the derivation of the linearized reflection coefficient. Thanks also to Leon Thomsen for pointing out the original author of approximation (6) and to members of the A(Anisotropy)-Team at the Center for Wave Phenomena (CWP) for useful discussions. The support for this work was provided by the members of the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP, Colorado School of Mines, and by the United States Department of Energy (project “Velocity
Analysis, Parameter Estimation, and Constraints on Lithology for Transversely Isotropic Sediments within the framework of the Advanced Computational Technology Initiative. Partial funding for this project has also been provided by a scholarship from Phillips Petroleum Co. and by a grant from Coleman Energy & Environmental Systems—Blackhawk Geoscience Division.

References

Urns, B., & Haugan, G. 1996. Weak-contrast approximation of the elastic scattering matrix in anisotropic media. Proceedings of workshop on seismic waves in laterally inhomogeneous media, Castle of Trest,
APPENDIX A: Polarization vectors in HTI media

One of the key elements to solve reflection and transmission coefficients in anisotropic media is the derivation of polarization vectors. Equations describing the direction of compressional wave polarizations in VTI media have been published by Rommel (1994) and Tsvankin (1996b). Denoting the phase and polarization angles with the symmetry axis as $\theta$ and $\gamma$, respectively, one can show that

\[
\begin{align*}
\cos \gamma &= \left(1 - f \sin^2 \theta \left[\delta + 2(\epsilon - \delta) \sin^2 \theta \right]\right) \cos \theta \\
\sin \gamma &= \left(1 + f \cos^2 \theta \left[\delta + 2(\epsilon - \delta) \sin^2 \theta \right]\right) \sin \theta
\end{align*}
\]

with $f = \alpha^2/(\alpha^2 - \beta^2)$.

Polarization vectors in HTI media can be derived using the analogy between VTI and HTI media (Rüger, 1996b). First, replace Thomsen’s (1986) parameters $\epsilon$ and $\delta$ with $\epsilon^{(V)}$ and $\delta^{(V)}$, the parameters of the equivalent VTI model. This substitution yields the polarization vector for quasi P-waves in the vertical plane containing the symmetry axis of HTI media and, as seen below, provides sufficient information to determine the polarization for P-waves with any incidence and azimuthal angle.

Equation A1, expressed in terms of angles $\gamma$ and $\theta'$ with respect to the horizontal symmetry axis in HTI models yields

\[
\begin{align*}
\sin \gamma' &= \left(1 - f \cos^2 \theta' \left[\delta^{(V)} + 2(\epsilon^{(V)} - \delta^{(V)}) \cos^2 \theta' \right]\right) \sin \theta' \\
\cos \gamma' &= \left(1 + f \sin^2 \theta' \left[\delta^{(V)} + 2(\epsilon^{(V)} - \delta^{(V)}) \cos^2 \theta' \right]\right) \cos \theta'
\end{align*}
\]

where $f = \alpha^2/(\alpha^2 - \beta^2)$ as in equation (A1), but $\alpha$ and $\beta$ denote the vertical compressional and the vertical fracture-parallel shear-wave velocity, respectively.

Physical properties in transversely isotropic media are invariant for a fixed angle with the symmetry axis, hence it is perfectly valid to generalize this equation and allow the polarization vector to rotate about the horizontal axis. The next step in the derivation of the polarization vectors is to relate the angles $\gamma$ and $\theta'$ with the symmetry axis to their corresponding incidence angle and azimuth. From simple geometry (Figure A1), we can find a relation between the cosine of an arbitrary angle $\alpha$ with the horizontal axis and its associated incidence polar angle $i$ and azimuth $\phi$:

\[
\cos \alpha = \sin i \cos \phi \tag{A3}
\]

The cartesian components of the unit polarization vector $\hat{d} = (d_1, d_2, d_3)^T$ and the unit wave vector $\hat{n} = (n_1, n_2, n_3)^T$ normal to the wavefront can be written as

\[
\hat{d} = \begin{pmatrix} \sin i \cos \phi \\ \sin i \sin \phi \\ \cos i \end{pmatrix} \quad \hat{n} = \begin{pmatrix} \sin \theta' \\ \cos \theta' \end{pmatrix} \tag{A4}
\]

Here, angles of the polarization vector and the phase vector with the vertical axis are called $i$ and $\phi$, respectively; similar notation applies to the azimuthal angles.

Using equations (A2) and (A3), one can relate the first components to obtain the simple expression $d_1 = n_1$. Corresponding equations for the two remaining components can be found using geometrical observations. First note that in transversely isotropic media, the polarization vector $\hat{n}$ lies in the plane formed by the symmetry axis and the phase vector. The two triangles shown in Figure A2 thus have the same angle $\psi$ and we find

\[
\begin{align*}
d_2 &= \frac{A}{B} = \sin \gamma' \cos \psi \\
&= \frac{n_2}{\sin \theta'} \\
&= \frac{m}{n_2} \\
d_2 &= \frac{B}{C} = \sin \gamma' \sin \psi \\
&= \frac{n_3}{\sin \theta'} \\
&= \frac{m}{n_3} \tag{A5}
\end{align*}
\]

Summarizing, we see that the P-wave polarization vec-
**P-wave reflectivity in anisotropic media** 303

The main idea of the derivation is to replace the exact boundary value problem

\[ \mathbf{A} \mathbf{R} = \mathbf{b}, \quad (B1) \]

where \( \mathbf{A} \) is the boundary equation matrix formed by the scattered (reflected and transmitted) wave types, \( \mathbf{R} \) is the vector of reflection and transmission coefficients, and \( \mathbf{b} \) is composed of the contribution of the incident wave to the boundary conditions, with a perturbation from the corresponding expression for the interface between two identical, isotropic and homogeneous media. Denoting the unperturbed quantities with the superscript \( u \), Thomsen (1993) showed that

\[ \Delta \mathbf{R} = (\mathbf{A}^u)^{-1} (\Delta \mathbf{b} - \Delta \mathbf{A} \mathbf{R}^u). \quad (B2) \]

Here, the operator \( \Delta \) is defined as \( \Delta \equiv \frac{\partial}{\partial x} \) with the vector \( d_j \) of small deviations from the two identical homogeneous, isotropic media given by

\[ d_j = \left( \frac{\Delta \alpha}{\alpha}, \frac{\Delta \beta}{\beta}, \frac{\Delta \rho}{\rho}, \delta_1^{(V)}, \delta_2^{(V)}, \delta_1^{(P)}, \delta_2^{(P)}, \varepsilon_1^{(V)}, \varepsilon_2^{(V)}, \gamma_1, \gamma_2 \right)^T. \quad (B3) \]

The 6 x 6 matrix \( \mathbf{A}^u \) needs to be inverted analytically. This can be achieved by relating \( \mathbf{A}^u \) to the eigenvectors of the transformed wave equation (Ursin & Haugen, 1996). To find the elements \( \Delta \mathbf{b} \) and \( \Delta \mathbf{A} \), it is necessary to use the weak anisotropy approximations to polarization vectors and plane velocities shown in the main text [equations (3) and (4)], while the approximate refracted angles can be derived using Snell’s law. The two elements of \( \Delta \mathbf{R} \) of most interest are the P-wave reflection and transmission coefficients shown in equation (5) and (13).

**APPENDIX C: Approximate symmetry-axis plane scattering coefficients for HTI/HTI interfaces**

The following equations are accurate for small discontinuities in the elastic properties across the boundary and weak anisotropy. The medium considered has the same parameterization as introduced in the main text. \( R_P \) and \( T_P \) are the P-wave reflection and transmission coefficients. Similar notation is used for the shear-wave scattering coefficients of \( S^\parallel \) and \( S^\perp \), polarized parallel and perpendicular to the isotropy plane, respectively.

To facilitate the comparison with the reflection coefficient

\[ R_P, T_P \]

in the previous approximation, the approximations are
shown as functions of $\beta$, the vertical velocity of the shear mode polarized parallel to the isotropy plane.

For cases where the best possible linear approximation of the symmetry-axis scattering coefficient is desired, the coefficient should be rewritten as function of the shear velocity $\beta^*$ ($\approx \beta(1 - \gamma)$).

A detailed discussion of the $P$-wave approximations is given in the main text. The shear-wave reflection coefficients are thoroughly investigated in Rüger (1996a).

\[
R_p (i) = \frac{1}{2} \frac{\Delta Z}{Z} - \\
\frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} - \left( \frac{2 \beta}{\beta} \right)^2 \left( \frac{\Delta G}{G} - 2 \Delta \gamma \right) + \Delta \delta^{(V)} \right) \sin^2 i + \\
\frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \Delta \delta^{(V)} \right) \sin^2 i \tan^2 i \\
T_p (i) = 1 - \frac{1}{2} \frac{\Delta Z}{Z} + \\
\left( \frac{\Delta \delta^{(V)} - \Delta \delta^{(V')}}{\Delta \delta^{(V)}} \right) \sin^4 i + \\
\frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \Delta \delta^{(V)} \right) \sin^2 i + \\
\frac{1}{2} \left( \frac{\Delta \alpha}{\alpha} + \Delta \delta^{(V)} \right) \sin^2 i \tan^2 i \\
R_s^{-} (j) = -\frac{1}{2} \left( \frac{\Delta \beta}{\beta} + \frac{\Delta \beta}{\beta} - \Delta \gamma \right) + \\
\left\{ \frac{1}{2} \left( \frac{\Delta \beta}{\beta} - \Delta \gamma \right) + 2 \frac{\Delta \rho}{\rho} \right. \\
\frac{1}{2} \left( \frac{\Delta \beta}{\beta} - \Delta \gamma \right) \left( \Delta \delta^{(V)} - \Delta \delta^{(V')} \right) \} \sin^2 j \\
- \frac{1}{2} \left( \frac{\Delta \beta}{\beta} - \Delta \gamma \right) \sin^2 j \tan^2 j \\
T_s^{-} (j) = 1 - \frac{1}{2} \left( \frac{\Delta \beta}{\beta} + \frac{\Delta \beta}{\beta} - \Delta \gamma \right) + \\
\left( \frac{\Delta \beta}{\beta} \right)^2 \left( \Delta \delta^{(V)} - \Delta \delta^{(V')} \right) \sin^4 j + \\
\frac{1}{2} \left\{ \frac{\Delta \beta}{\beta} - \Delta \gamma + \frac{1}{2} \left( \frac{\alpha}{\beta} \right)^2 \times \\
\left( \Delta \delta^{(V)} - \Delta \delta^{(V')} \right) \} \sin^2 j + \\
\left( \frac{\Delta \beta}{\beta} - \Delta \gamma \right) \sin^2 j \tan^2 j \\
R_s^{+} (j) = -\frac{1}{2} \left( \frac{\Delta \beta}{\beta} + \frac{\Delta \beta}{\beta} \right) + 
\]
Effective parameters and $P$-wave velocity for orthorhombic anisotropy

Ilya Tsvankin

Center for Wave Phenomena, Department of Geophysics
Colorado School of Mines

ABSTRACT

Although orthorhombic (or orthotropic) symmetry is believed to be common for fractured reservoirs, the difficulties in dealing with nine independent elastic constants have precluded this model from being used in seismology. A new notation introduced in this work is designed to make seismic inversion and processing for orthorhombic media more practical by simplifying the description of a wide range of seismic signatures. Taking advantage of the fact that the Christoffel equation has the same form in the symmetry planes of orthorhombic and transversely isotropic (TI) media, the stiffness coefficients can be replaced by two vertical ($P$ and $S$) velocities and seven dimensionless parameters that represent an extension of Thomsen’s (1986) anisotropy coefficients to orthorhombic models. By design, this notation preserves all attractive features of Thomsen parameters in treating wave propagation and performing 2-D processing in the symmetry planes of orthorhombic media.

The new notation has proved equally useful in describing seismic signatures outside the symmetry planes. Linearization of $P$-wave phase velocity in the dimensionless anisotropic coefficients leads to a concise weak-anisotropy approximation that provides good accuracy even for models with pronounced polar and azimuthal velocity variations. This approximation can be efficiently used to build analytic solutions for such signatures as group velocity, dip-dependent normal-moveout velocity, polarization vector, radiation pattern etc. One of the most important advantages of the new notation is the reduction in the number of parameters responsible for $P$-wave velocities. All kinematic signatures of $P$-waves in orthorhombic media depend just on the vertical velocity $V_{P0}$ and five anisotropic parameters, with $V_{P0}$ being no more than a scaling coefficient in homogeneous media; this conclusion holds even for orthorhombic models with strong velocity anisotropy. Furthermore, the influence of anisotropy on $P$-wave reflection traveltimes is largely determined just by the two pairs of the coefficients $c$ and $\delta$ defined in the vertical symmetry planes.

Key words: azimuthal anisotropy, orthorhombic symmetry, phase velocity, seismic signatures

Introduction

The vast majority of existing studies of seismic anisotropy are limited to transversely isotropic (TI) models with different orientation of the symmetry axis. It has been shown in the literature (e.g., Sayers, 1994a) that TI (or hexagonal) symmetry adequately describes the elastic properties of shales, which represent the major source of anisotropy in sedimentary basins. Most shale formations are horizontally layered, yielding an effective transversely isotropic medium with a vertical symmetry axis (the so-called VTI media). In some cases, shale layers may be dipping (e.g., near flanks of salt domes), which
leads to a tilt of the symmetry axis away from vertical (Tsvankin, 1990c).

However, the transversely isotropic model becomes much more restrictive when applied to the description of cracked media that contain small (compared to the predominant wavelength) fractures or microcracks. Indeed, only the simplest fractured model with a single system of parallel vertical circular ("penny-shaped") cracks embedded in an isotropic matrix exhibits transverse isotropy with a horizontal symmetry axis (HTI media). Deviations from the circular crack shape, misalignment of the crack planes, the addition of a second crack system or the presence of anisotropy in the matrix make the effective medium either orthorhombic or lead to an even lower symmetry. It is believed that one of the most common reasons for orthorhombic anisotropy in sedimentary basins is a combination of parallel vertical cracks and vertical transverse isotropy in the the background medium, as illustrated by Figure 1 (Wild and Crampin, 1991; Schoenberg and Helbig, 1996). Orthorhombic symmetry can also be caused by two or three mutually orthogonal crack systems or two identical systems of cracks making an arbitrary angle with each other. Hence, orthorhombic anisotropy is expected to be rather typical for realistic fractured media.

Wave propagation in orthorhombic media has been studied in a number of publications including Musgrave (1970), Tsvankin and Chesnokov (1990a,b), Wild and Crampin (1991), Brown et al. (1991), Sayers (1994b) and Schoenberg and Helbig (1996). Advanced numerical tools, such as finite-difference or reflectivity methods generalized for anisotropic media, make the modeling of seismic wavefields in orthorhombic models a relatively straightforward (albeit a costly) procedure. However, extending inversion and processing techniques to orthorhombic media is a much more difficult endeavor. Seismic methods of fracture detection, extensively developed since the early 1980's, are based on the analysis of the travel times and reflection coefficients of split shear waves at near-vertical incidence (Crampin, 1985; Thomsen, 1988). If azimuthal anisotropy is caused by a single system of penny-shaped vertical cracks, the relative time delay or the difference in the normal-incidence reflection amplitude between the shear modes provide a direct estimate of the crack density, while the polarization of the fast S-wave determines the crack orientation.

These well-known shear-wave splitting methods remain valid in orthorhombic media formed by vertical penny-shaped cracks embedded in a VTI background. However, even in this relatively simple model, vertically travelling shear waves can provide us with a single anisotropic parameter (the splitting coefficient) that cannot be used, for instance, to separate empty and gas-filled (or fluid-filled) cracks. In more complicated orthorhombic models, such as ones containing two crack systems or non-circular cracks, interpretation of the shear-wave splitting parameter becomes ambiguous, unless some additional data are available. Also, it is difficult even to distinguish between HTI and orthorhombic media using polarizations and time delays of split shear waves at near-vertical incidence (Wild and Crampin, 1991). It may be possible to use point shear-wave singularities to identify orthorhombic models (Bush and Crampin, 1991), but distortions of the wavefronts and polarizations make the wavefield in the vicinity of singular points extremely complicated.

Therefore, seismic characterization of media with orthorhobic symmetry should involve analysis of azimuthally-dependent P-wave signatures and, if possible, shear data at oblique incidence angles. Recently, Lynn et al. (1995) presented a field-data example showing the feasibility of detecting azimuthal anomalies of P-wave velocities and amplitudes over a typical fractured reservoir. Interpretation of such anomalies and development of seismic inversion and processing methods for orthorhombic models is impossible without a good understanding of the dependence of seismic signatures on the anisotropy parameters. Since general orthorhombic media are described by nine independent stiffness components, the complexity of this problem cannot be overestimated.

This work is aimed at simplifying seismic treatment

* For brevity, the qualifiers in "quasi-P-wave" and "quasi-S-wave" will be omitted.
of orthorhombic media by identifying the combinations of the stiffness coefficients responsible for a wide range of seismic signatures. The notation that enables us to reach this goal is based on the same principle as Thomsen (1986) parameters for VTI media. The unified description of vertical transverse isotropy and orthorhombic media provides a convenient bridge between the two models and makes it possible to take full advantage of the recent developments in the analytic description of seismic signatures in VTI media (Tsiankin, 1996a). This approach has already been successfully applied by Rüger (1996) and Tsiankin (1996b) to TI models with a horizontal symmetry axis. For instance, Tsiankin (1996b) showed that the anisotropic parameter that governs azimuthally-dependent P-wave normal-moveout (NMO) velocity in HTI media represents the same combination of the stiffness coefficients as Thomsen’s parameter δ originally suggested for vertical transverse isotropy. Also, Thomsen-style parameters have been used in physical-modeling studies by Brown et al. (1991) and Cheadle et al. (1991) to characterize anisotropy in the symmetry planes of an orthorhombic material (phenolic laminate).

After introducing the new notation, I give an analytic description of seismic signatures in the symmetry planes of orthorhombic media by using a limited analogy with vertical transverse isotropy. Linearization of the exact phase-velocity equation in the dimensionless anisotropic parameters leads to a simple weak-anisotropy approximation for P-wave phase velocity outside the symmetry planes. Further analytic and numerical study of the phase-velocity function shows that P-wave kinematic signatures in orthorhombic media are determined by the vertical velocity and only five anisotropic coefficients instead of the nine stiffnesses in the conventional notation.

Orthorhombic Stiffness Tensor

The stiffness tensor $c_{ijkl}$ for orthorhombic media can be represented in a compressed two-index notation (the so-called “Voigt recipe”) as follows:

$$ C_{ijkl} = \begin{pmatrix}
  c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\
  c_{12} & c_{22} & c_{23} & 0 & 0 & 0 \\
  c_{13} & c_{23} & c_{33} & 0 & 0 & 0 \\
  0 & 0 & 0 & c_{44} & 0 & 0 \\
  0 & 0 & 0 & 0 & c_{55} & 0 \\
  0 & 0 & 0 & 0 & 0 & c_{66}
\end{pmatrix} \quad (1) $$

This tensor has the same vanishing components as the one for transversely isotropic (TI) media with the symmetry axis aligned with one of the coordinate axes. However, the TI symmetry leads to several relations between the non-zero elements that reduce the number of independent coefficients. For instance, for vertical transverse isotropy (VTI media) the symmetry direction coincides with the $\mathbf{x}_3$ axis, and the $\mathbf{x}_1$ and $\mathbf{x}_2$ directions are equivalent; thus, in VTI media

$$ c_{11} = c_{22} ; \quad c_{12} = c_{23} ; \quad c_{44} = c_{55} ; \quad c_{12} = c_{11} - 2c_{66} \quad (2) $$

Relations (2) no longer hold in orthorhombic media, and all nine stiffnesses in tensor (1) are generally independent. For certain subsets of orthorhombic models (such as media due to vertical cracks in a VTI background), the number of independent stiffnesses can be reduced (Schoenberg and Helbig, 1996), but at this point we do not restrict ourselves to any specific type of orthorhombic anisotropy.

The phase velocity $V$ and the displacement vector $\mathbf{U}$ of plane waves in arbitrary homogeneous anisotropic media satisfy the Christoffel equation (e.g., Musgrave, 1970):

$$ [G_{ik} - \rho V^2 \delta_{ik}] U_k = 0, \quad (3) $$

where $\rho$ is the density, $\delta_{ik}$ is the Kronecker’s symbolic $\delta$, and $G_{ik}$ is the so-called Christoffel matrix given by

$$ G_{ik} = c_{ijkl} n_j n_l, \quad (4) $$

$\mathbf{n}$ is the unit vector in the slowness direction.

The Christoffel equation (3) describes a standard eigenvalue($\rho V^2$)-eigenvector($\mathbf{U}$) problem for the symmetric matrix $G$, with the eigenvalues determined by

$$ \det [G_{ik} - \rho V^2 \delta_{ik}] = 0. \quad (5) $$

Solving the cubic equation (5) at any slowness direction $\mathbf{n}$ yields three positive values of the squared phase velocity $V^2$ (in compliance with stability conditions), which correspond to the quasi-P-wave and two quasi-S-waves. For certain orientations of the slowness vector the velocities of the split S-waves coincide with each other, which leads to the so-called shear-wave singularities. After the eigenvalues have been determined, the associated displacement (polarization) vectors $U_k$ for each mode can be found from equation (3). As follows from the properties of the eigenvalue-eigenvector problem for symmetric matrices, the polarization vectors of the three modes are always orthogonal to each other. However, none of the polarization vectors is necessarily parallel or perpendicular to the slowness direction, and in general there are no “pure” P or S-waves in anisotropic media.

Using equations (1) and (4) and the Voigt recipe for the two-index notation, we obtain the components of the Christoffel matrix for orthorhombic media as

$$ G_{11} = c_{11} n_1^2 + c_{66} n_2^2 + c_{55} n_3^2 ; \quad (6) $$

$$ G_{22} = c_{66} n_1^2 + c_{22} n_2^2 + c_{44} n_3^2 ; \quad (7) $$
Wave Propagation in the Symmetry Planes

Orthorhombic media have three mutually orthogonal planes of mirror symmetry that coincide with the coordinate planes \([x_1, x_2], [x_1, x_3],\) and \([x_2, x_3]\) (Figure 1). Due to the relative simplicity of the basic equations, symmetry-plane analysis provides important insights into the wave propagation in orthorhombic media. Also, we will use the properties of seismic waves in the symmetry planes to introduce a new notation broadly useful in treating orthorhombic models.

Anticipating future applications in reflection seismology, we will define the vertical and horizontal directions by assuming that the symmetry plane \([x_1, x_2]\) is horizontal. Evidently, information obtained from reflection seismic experiments is mostly related to the properties of vertical planes (including two vertical symmetry planes), especially to near-vertical phase- and group-velocity variations. If none of the symmetry planes is horizontal, the treatment of reflected waves becomes more complicated; however, all the equations derived below for a homogenous orthorhombic medium remain entirely valid.

For a wave propagating in the \([x_1, x_3]\) plane, the projection of the slowness vector on the \(x_2\)-axis vanishes \((n_2 = 0)\), and the Christoffel matrix \(G_{ik}\) (equations (6)-(11)) has only four non-zero components \((G_{11}, G_{22}, G_{32}, G_{13})\). The Christoffel equation (3) in the \([x_1, x_3]\) plane takes the form shown in equation (12).

Equation (12) coincides with the corresponding Christoffel equation for the transversely isotropic model with the symmetry axis in the \(x_3\)-direction (VTI medium). In the orthorhombic model, however, the stiffness components \(c_{44}\) and \(c_{45}\) are no longer equal to each other, which leads to shear-wave splitting at vertical incidence (this will be discussed in more detail below). This analogy with VTI media has already been mentioned in the literature (Musgrave, 1970; Cheadle et al., 1991; Schoenberg and Helbig, 1996); here, however, we will use it in a systematic fashion to describe wave propagation in the symmetry planes and to introduce new dimensionless anisotropy parameters for orthorhombic anisotropy.
\[
\begin{bmatrix}
\sigma_1 n_1^2 + c_{55} n_3^2 - \rho V^2 \\
\sigma_0 n_1^2 + c_{44} n_3^2 - \rho V^2 \\
(c_{13} + c_{55}) n_1 n_3 \\
\end{bmatrix}
\begin{bmatrix}
U_1 \\
U_2 \\
U_3 \\
\end{bmatrix}
= 0.
\] (12)

\[
\begin{bmatrix}
\sigma_1 n_1^2 + c_{55} n_3^2 - \rho V^2 \\
\sigma_0 n_1^2 + c_{44} n_3^2 - \rho V^2 \\
(c_{13} + c_{55}) n_1 n_3 \\
\end{bmatrix}
\begin{bmatrix}
U_1 \\
U_2 \\
U_3 \\
\end{bmatrix}
= 0.
\] (13)

\[
\begin{bmatrix}
\sigma_1 \sin^2 \theta + c_{55} \cos^2 \theta - \rho V^2 \\
\sigma_0 \sin^2 \theta + c_{44} \cos^2 \theta - \rho V^2 \\
(c_{13} + c_{55}) \sin \theta \cos \theta \\
\end{bmatrix}
\begin{bmatrix}
U_1 \\
U_2 \\
U_3 \\
\end{bmatrix}
= 0.
\] (14)

A significant influence on the distribution of energy along the wavefront within the symmetry planes (Tsvankin and Chesnokov, 1990a).

Equation (12) also has a solution corresponding to the shear wave polarized orthogonally to the \([x_1, x_3]\) plane (\(U_1 = U_3 = 0\), \(U_2 \neq 0\)).

\[
c_{56} \sin^2 \theta + c_{44} \cos^2 \theta - \rho V^2 = 0.
\] (15)

Equation (15) describes a "pure" shear mode (the polarization vector is normal to the slowness direction) with an elliptical slowness surface and an elliptical wavefront in the \([x_1, x_3]\) plane. The vertical phase (and group) velocity of this SH-wave is given by \(\sqrt{c_{55}/\rho}\), the horizontal velocity is equal to \(\sqrt{c_{44}/\rho}\). As in the case of the in-plane polarized waves, equation (15) is identical to the phase-velocity equation for the elliptical SH-wave in VTI media. However, since in orthorhombic media \(c_{13}\) is not equal to \(c_{55}\), the vertical velocities of the shear waves are different, which leads to shear-wave splitting at vertical incidence. In fact, the two shear waves in the \([x_1, x_3]\) plane of an orthorhombic medium are fully decoupled because they depend on different sets of elastic constants (compare equations (14) and (15)).

Similar conclusions can be drawn for wave propagation in the second vertical plane of mirror symmetry. Substituting \(n_1 = 0\) into the Christoffel matrix (equations (6)-(11)), we obtain equation (16) from the Christoffel equation (3) in the \([x_2, x_3]\) plane.

Equation (16) splits into two separate equations for the waves polarized within and perpendicular to the \([x_2, x_3]\) plane. For the \(P\)-wave and the in-plane polarized shear wave (\(SV\)-mode) we have equation (17).

Introducing the in-plane phase angle with vertical \((n_2 = \sin \theta; n_3 = \cos \theta)\) yields equation (18), which becomes identical to the corresponding Christoffel equation (14) for VTI media and the \([x_1, x_3]\) plane of orthorhombic media if we replace \(c_{22}\) with \(c_{11}\), \(c_{44}\) with \(c_{55}\), and \(c_{23}\) with \(c_{13}\) (and, of course, \(U_2\) with \(U_1\)). These replacements amount just to interchanging the indices 1 and 2 in the appropriate components of the stiffness tensor \(c_{ijkl}\).

Therefore, to obtain the phase velocity of both in-plane polarized modes in the \([x_2, x_3]\) plane as a function of the phase angle with vertical, it is sufficient to make the above substitutions of the elastic constants in the known phase-velocity equations for VTI media (or for the \([x_1, x_3]\) plane of orthorhombic media). The \(P\)-wave velocity in the horizontal \((x_2)\) direction is equal to \(\sqrt{c_{22}/\rho}\), while the vertical and horizontal velocity of the in-plane polarized \(S\)-wave in the \([x_2, x_3]\) plane is given by \(\sqrt{c_{44}/\rho}\). At vertical incidence, the shear wave polarized within the \([x_2, x_3]\) plane (in the \(x_2\) direction) plays the role of the transversely polarized \(S\)-wave (\(SH\)-wave) with respect to wave propagation in the \([x_1, x_3]\) plane (see equation (15)).

For the shear wave polarized orthogonally to the propagation plane \((U_2 = U_3 = 0; U_1 \neq 0)\), we get an expression analogous to equation (15):

\[
c_{56} \sin^2 \theta + c_{44} \cos^2 \theta - \rho V^2 = 0.
\] (19)

Equation (19) describes an elliptical slowness curve (leading to an elliptical wavefront) of the \(SH\)-wave in the \([x_2, x_3]\) plane, with the elliptical axes determined by the vertical \((\sqrt{c_{55}/\rho})\) and horizontal \((\sqrt{c_{44}/\rho})\) velocities.

It can be demonstrated exactly in the same fashion that the Christoffel equation in the third \([x_1, x_3]\) symmetry plane becomes identical to the VTI equations with the appropriate substitution of elastic constants. For instance, the phase-velocity equation for the waves polarized within the \([x_1, x_2]\) plane has the form

\[
(c_{22} \sin^2 \theta + c_{56} \cos^2 \theta - \rho V^2)
\]

\[
(c_{56} \sin^2 \theta + c_{11} \cos^2 \theta - \rho V^2)
\]

\[
-(c_{12} + c_{56})^2 \sin^2 \theta \cos^2 \theta = 0,
\] (20)

where \(\theta\) is the phase angle with the \(z\) axis defined by \(n_1 = \cos \theta; n_2 = \sin \theta\). Equation (20) will coincide with the corresponding phase-velocity expression for VTI me-
\[
\begin{bmatrix}
c_{66} n_3^2 + c_{33} n_2^2 - \rho V^2 \\
0 \\
0 \\
\end{bmatrix}
\begin{bmatrix}
c_{33} n_2^2 + c_{44} n_2^2 - \rho V^2 \\
(c_{23} + c_{44}) n_2 n_3 \\
c_{44} n_2^2 - c_{33} n_3^2 - \rho V^2 \\
\end{bmatrix}
\begin{bmatrix}
U_1 \\
U_2 \\
U_3 \\
\end{bmatrix} = 0.
\]

(16)

\[
\begin{bmatrix}
c_{33} n_2^2 + c_{44} n_2^2 - \rho V^2 \\
(c_{23} - c_{44}) n_2 n_3 \\
c_{44} n_2^2 + c_{33} n_3^2 - \rho V^2 \\
\end{bmatrix}
\begin{bmatrix}
U_2 \\
U_3 \\
\end{bmatrix} = 0,
\]

(17)

\[
\begin{bmatrix}
c_{22} \sin^2 \theta + c_{44} \cos^2 \theta - \rho V^2 \\
(c_{22} + c_{44}) \sin \theta \cos \theta \\
c_{44} \sin^2 \theta + c_{22} \cos^2 \theta - \rho V^2 \\
\end{bmatrix}
\begin{bmatrix}
U_2 \\
U_3 \\
\end{bmatrix} = 0.
\]

(18)

dia [see equation (14)] if we make the following substitutions:

\[
\begin{align*}
c_{11} &= c_{33}; \\
c_{22} &= c_{11}; \\
c_{66} &= c_{33}; \\
c_{12} &= c_{13}.
\end{align*}
\]

(21)

**Effective Parameters for Orthorhombic Anisotropy**

The results of the previous section justify direct application of the known VTI equations to the symmetry planes of orthorhombic media (with the above-mentioned exception of body-wave amplitudes). To take full advantage of the analogy with VTI media, I suggest to replace the stiffness coefficients \(c_{ij}\) with effective anisotropic parameters (combinations of stiffnesses) that concisely characterize a wide range of seismic signatures for orthorhombic anisotropy. Indeed, since seismic signatures for transversely isotropic media are especially convenient to describe by dimensionless Thomsen (1986) parameters (Tsvankin, 1996a), it is natural to develop a Thomsen-style notation for the symmetry planes of orthorhombic media. As we will see below, these parameters turn out to be equally helpful in studying wave propagation outside the symmetry planes.

First, following Thomsen’s recipe, we have to define two vertical velocities (for \(P\) and \(S\)-waves) of the reference isotropic model. In the case of orthorhombic media, we can choose either of the two velocities of split shear waves at vertical incidence. Here, the preference is given to the \(S\)-wave polarized in the \(z_1\) direction to make the new notation for the \("P\!\!-\!SV\"") waves in the \([x_1, x_2]\) plane exactly identical to Thomsen’s notation in the VTI case. Thus, the “isotropic” velocities are defined by

\[
\begin{align*}
V_{P0} &= \sqrt{\frac{c_{33}}{\rho}}, \\
V_{S0} &= \sqrt{\frac{c_{66}}{\rho}}.
\end{align*}
\]

(22)

(23)

Since the Christoffel equation (14) for the waves polarized in the \([x_1, x_2]\) plane is identical to the corresponding equation for vertical transverse isotropy, we introduce the dimensionless coefficients \(\epsilon^{(2)}\) and \(\delta^{(2)}\) [the superscript (2) denotes the \([x_1, x_2]\) symmetry plane that is normal to the \(x_2\) axis] through the same equations as those used by Thomsen (1986) for VTI media:

\[
\epsilon^{(2)} = \frac{c_{11} - c_{33}}{2c_{33}},
\]

(24)

\[
\delta^{(2)} = \frac{(c_{13} + c_{55})^2 - (c_{23} - c_{55})^2}{2c_{33}(c_{23} - c_{55})}.
\]

(25)

Note that in presenting the definition of \(\delta\) for VTI media, it is common to replace the coefficient \(c_{55}\) with \(c_{44}\), but these two coefficients are different in orthorhombic media. The analogy with vertical transverse isotropy implies that we can obtain kinematic signatures and polarizations of the \(P\!\!-\!S\!\!V\)-waves in the \([x_1, x_2]\) plane of orthorhombic media just by substituting \(V_{P0}, V_{S0}, \epsilon^{(2)},\) and \(\delta^{(2)}\) into the known equations for VTI media expressed through Thomsen parameters \(V_{P0}, V_{S0}, \epsilon,\) and \(\delta,\) respectively.

For instance, the parameters \(\epsilon^{(2)}\) and \(\delta^{(2)}\) govern the velocity variations of the in-plane polarized waves exactly in the same way as Thomsen’s \(\epsilon\) and \(\delta\) determine the \(P\!\!-\!S\!\!V\) velocity anisotropy in VTI media. Using the exact phase-velocity equation for VTI media expressed through Thomsen parameters (Tsvankin, 1996a) and replacing \(\epsilon\) and \(\delta\) with \(\epsilon^{(2)}\) and \(\delta^{(2)},\) respectively, we find the phase velocities of the \(P\!\!-\!S\!\!V\)-waves in the \([x_1, x_2]\) plane as

\[
\frac{V^2(\theta)}{V_{P0}^2} = 1 + \epsilon^{(2)} \sin^2 \theta - \frac{f}{2}
\]

\[
\pm \frac{f}{2} \sqrt{\left(1 + 2\epsilon^{(2)} \sin^2 \theta\right)^2 - \frac{2(\epsilon^{(2)} - \delta^{(2)}) \sin^2 2\theta}{f}}.
\]

(26)

\[
f = 1 - \frac{V_{S0}^2}{V_{P0}^2}.
\]

(27)

The plus sign in front of the radical corresponds to the \(P\)-wave, while the minus sign - to the \(SV\)-wave. In the weak-anisotropy approximation, the \(P\)-wave phase velocity in the \([x_1, x_2]\) plane can be linearized in the small
parameters $\varepsilon^{(2)}$ and $\delta^{(2)}$ to obtain the well-known Thomsen’s (1986) expression:

$$V_P(\theta) = V_{P0} \left(1 + \delta^{(2)} \sin^2 \theta \cos^2 \theta + \varepsilon^{(2)} \sin^2 \theta \right).$$  \hspace{1cm} (28)

Phase-velocity equations (26) and (28) demonstrate how we can use the new effective anisotropic parameters to take advantage of the analogy between the symmetry planes of orthorhombic media and vertical transverse isotropy. A more detailed description of some other seismic signatures in the planes of mirror symmetry will be given in the next section.

Next, using the equivalence between equation (15) and the corresponding expression for the $SH$-wave in VTI media, we introduce another dimensionless anisotropic parameter, $\gamma^{(2)}$:

$$\gamma^{(2)} \equiv \frac{c_{66} - c_{14}}{2c_{14}}.$$  \hspace{1cm} (29)

The coefficient $\gamma^{(2)}$ is identical to Thomsen’s parameter $\gamma$ for VTI media: it is responsible for the velocity variations of the elliptical $SH$-wave (pure $S$-wave polarized normal to the propagation $[x_1; x_3]$ plane).

Note that the coefficients $\delta^{(2)}$, $\gamma^{(2)}$, and $\gamma^{(2)}$ coincide with the parameters $\varepsilon^{(V)}$, $\delta^{(V)}$, and $\gamma^{(V)}$ (respectively) introduced by Tsvankin (1996b) and Rüger (1996) for transversely isotropic media with the symmetry axis along the $x_1$ direction. This HTI model, as discussed in more detail below, may be considered as a special case of the more general orthorhombic media described here.

Note that the vertical velocity of the $SH$-wave is equal to $V_{S1} = \sqrt{c_{44}/\rho}$ which, in contrast to the VTI case, is different from the vertical velocity of the $SV$-wave.

In principle, it would be convenient to specify the parameters $\varepsilon$, $\delta$ and $\gamma$ in the same fashion for the other two symmetry planes as well. However, in this case some of the anisotropy coefficients would not be independent, and the new notation would suffer from redundancy. In defining the anisotropic parameters, we put an emphasis on simplifying seismic signatures in the two vertical planes of symmetry. To describe $P - SV$-waves in the $[x_2, x_3]$ symmetry plane (the plane normal to the $x_1$ axis), we introduce the parameters $\varepsilon^{(1)}$ and $\delta^{(1)}$ defined analogously to $\varepsilon^{(2)}$ and $\delta^{(2)}$. As shown in the previous section, the Christoffel equation in the $[x_2, x_3]$ plane of orthorhombic media becomes identical to that in VTI media after the following substitutions:

$$c_{22} = c_{11}; \quad c_{44} = c_{55}; \quad c_{23} = c_{13}.$$  \hspace{1cm} (30)

Replacing the stiffnesses in the definitions of $\varepsilon^{(2)}$, $\delta^{(2)}$ and $\gamma^{(2)}$ [equations (24), (25), and (29)] in accordance with relations (30) yields

$$\varepsilon^{(1)} \equiv \frac{c_{22} - c_{33}}{2c_{33}},$$  \hspace{1cm} (31)

$$\gamma^{(1)} \equiv \frac{c_{66} - c_{55}}{2c_{55}}.$$  \hspace{1cm} (33)

Now, for instance, we can obtain the phase velocities of the $P$-$SV$-waves in the $[x_2, x_3]$ plane from equation (26) by substituting $\varepsilon^{(1)}$ for $\varepsilon^{(2)}$ and $\delta^{(1)}$ for $\delta^{(2)}$ (also, $f$ should contain $V_{S1}$ instead of $V_{S0}$).

The two vertical velocities and six anisotropy parameters introduced above replace eight original stiffness coefficients: $c_{11}$, $c_{22}$, $c_{33}$, $c_{44}$, $c_{55}$, $c_{66}$, $c_{23}$, and $c_{12}$. The only remaining independent stiffness $c_{12}$ is responsible for the velocity variations of the $P - SV$-waves in the $[x_1; x_2]$ plane. We replace $c_{12}$ with a dimensionless anisotropic parameter analogous to the $\delta$ coefficients in the vertical planes of symmetry:

$$\delta^{(3)} \equiv \frac{(c_{12} - c_{66})^2 - (c_{11} - c_{66})^2}{2c_{11}(c_{11} - c_{66})}.$$  \hspace{1cm} (34)

From the equivalence between the $[x_1; x_2]$ plane and VTI media described by relations (21), it is clear that $\delta^{(3)}$ plays the role of Thomsen’s $\delta$ in TI equations written for the symmetry axis coinciding with the $x_1$ direction.

After introducing the new parameters, it is convenient to list all of them along with their brief descriptions:

- $V_{P0}$ – the vertical velocity of the $P$-wave;
- $V_{S0}$ – the vertical velocity of the $S$-wave polarized in the $x_1$ direction;
- $\varepsilon^{(2)}$ – the VTI parameter $\varepsilon$ in the symmetry plane $[x_1; x_3]$ normal to the $x_2$ axis (close to the fractional difference between the $P$-wave velocities in the $x_1$ and $x_3$ directions);
- $\delta^{(2)}$ – the VTI parameter $\delta$ in the $[x_1; x_3]$ plane (responsible for near-vertical $P$-wave velocity variations, also influences $SV$-wave velocity anisotropy);
- $\gamma^{(2)}$ – the VTI parameter $\gamma$ in the $[x_1; x_3]$ plane (close to the fractional difference between the $SH$-wave velocities in the $x_1$ and $x_3$ directions);
- $\varepsilon^{(1)}$ – the VTI parameter $\varepsilon$ in the $[x_2; x_3]$ plane;
- $\delta^{(1)}$ – the VTI parameter $\delta$ in the $[x_2; x_3]$ plane;
- $\gamma^{(1)}$ – the VTI parameter $\gamma$ in the $[x_2; x_3]$ plane;
- $\delta^{(3)}$ – the VTI parameter $\delta$ in the $[x_1; x_2]$ plane ($\delta$ plays the role of the symmetry axis).

While the new parameters are uniquely determined by the nine independent stiffness coefficients of orthorhombic media, the inverse transition is unique only for the coefficients associated with the velocities along the coordinate axes ($c_{11}$, $c_{22}$, $c_{33}$, $c_{44}$, $c_{55}$, and $c_{66}$). To obtain the other three coefficients ($c_{12}$, $c_{23}$, $c_{32}$) from the corresponding $\delta$ values, it is necessary to specify the sign of the sums $(c_{12} + c_{23})$ [equation (25)], $(c_{22} + c_{44})$ [equation (32)], and $(c_{12} + c_{66})$ [equation (34)]. As discussed by Tsvankin,
(1996a), exactly the same problem arises with Thomsen parameters in transversely isotropic media. However, since the stability condition requires the coefficients $c_{55}$, $c_{44}$, and $c_{66}$ to be always positive, the sums under consideration can be negative only for extremely uncommon large negative values of $c_{13}$, $c_{23}$ or $c_{12}$. Therefore, for practical purposes of seismic processing and interpretation we can assume that $(c_{13} + c_{55})$, $(c_{23} + c_{44})$, and $(c_{12} + c_{66})$ are positive. That would correspond to one of the conditions of the so-called “mild anisotropy” as specified by Schoenberg and Helbig (1996) and ensure the absence of anomalous body-wave polarizations in the symmetry planes (Helbig and Schoenberg, 1987).

Both vertical and horizontal transverse isotropy can be considered as degenerated special cases of orthorhombic media. An orthorhombic medium reduces to the VTI model if the properties of all vertical planes are identical, while the velocity of each mode in the $[x_1, x_2]$ plane (the so-called “isotropy plane”) is constant (although the velocities of the two $S$-waves are generally different). Hence, the VTI constraints, conventionally presented through the stiffness coefficients [see relations (2)], can be rewritten in terms of the new parameters as

- $\epsilon^{(2)} = \epsilon^{(2)} = \epsilon$,
- $\delta^{(2)} = \delta^{(2)} = \delta$,
- $\gamma^{(2)} = \gamma^{(2)} = \gamma$,
- $\delta^{(3)} = 0$,

where $\epsilon$, $\delta$, and $\gamma$ are Thomsen’s anisotropy coefficients for VTI media. These four relations reduce the number of independent parameters from nine in orthorhombic media to five for vertical transverse isotropy.

Another special case is transverse isotropy with a horizontal axis of symmetry. If the symmetry axis is oriented along the $x_3$ direction, then the axes $x_2$ and $x_3$ form the isotropy plane, and

- $\epsilon^{(3)} = 0$;
- $\delta^{(3)} = 0$;
- $\gamma^{(3)} = 0$.

The parameters $\epsilon^{(2)}$, $\delta^{(2)}$ and $\gamma^{(2)}$ in this case coincide with the coefficients $\epsilon^{(V)}$, $\delta^{(V)}$ and $\gamma^{(V)}$ (respectively) introduced by Tsvankin (1996b) and Rüger (1996). Wave propagation in HTI media is fully described by these three dimensionless anisotropic coefficients and two vertical velocities $V_{P0}$ and $V_{S0}$. The last anisotropic parameter, $\delta^{(3)}$, in this model becomes the generic Thomsen’s coefficient $\delta$ that can be found, if necessary, from the following equation:

$$\delta^{(2)} = \delta^{(3)} \left( 1 + 2\epsilon^{(2)} \right) \left( 1 + 2\epsilon^{(2)} / f \right) + 2\epsilon^{(2)} \left( 1 + \epsilon^{(2)} / f \right),$$

where $f$ is defined in equation (27). The above expression for $\delta^{(2)}$ follows from the relations $c_{12} = c_{13}$ and $c_{55} = c_{66}$ for this HTI medium and can be obtained directly from the equation for the corresponding parameter $\delta^{(V)}$ in the paper by Tsvankin (1996b, Appendix A).

If the symmetry axis is parallel to the $x_2$ direction, then

- $\epsilon^{(2)} = 0$;
- $\delta^{(2)} = 0$;
- $\gamma^{(2)} = 0$.

In this case, the parameters $\epsilon^{(1)}$, $\delta^{(1)}$ and $\gamma^{(1)}$ play the role of the HTI coefficients $\epsilon^{(V)}$, $\delta^{(V)}$ and $\gamma^{(V)}$, respectively. The last anisotropic parameter, $\delta^{(3)}$, is simply equal to $\delta^{(1)}$.

While the nine parameters introduced above are sufficient to characterize general orthorhombic media, one may need to use different combinations of these coefficients in specific applications. For instance, shear-wave splitting at vertical incidence, extensively studied in a number of publications (e.g., Crampin 1985, Thomsen 1988), is conventionally described by the fractional difference between the parameters $c_{44}$ and $c_{55}$:

$$\gamma^{(S)} \equiv \frac{c_{44} - c_{55}}{2c_{55}} = \frac{\gamma^{(1)} - \gamma^{(2)}}{1 + 2\gamma^{(3)}}. \quad (33)$$

This definition of the splitting parameter $\gamma^{(S)}$ makes it identical to the generic Thomsen’s parameter $\gamma$ for transversely isotropic media with a horizontal symmetry axis that coincides with the $x_3$ direction. In terms of the shear-wave vertical velocities ($V_{S0}$ and $V_{S1}$), $\gamma^{(S)}$ takes the following form:

$$\gamma^{(S)} = \frac{V_{S1}^2 - V_{S0}^2}{2V_{S0}^2} \approx \frac{V_{S1} - V_{S0}}{V_{S0}}.$$

Therefore, $\gamma^{(S)}$ represents a direct measure of the time delay between two split shear waves at vertical incidence.

Application of the New Notation in the Symmetry Planes

By design, the new parameters provide a convenient way of describing seismic signatures in the symmetry planes of orthorhombic media using the known equations for VTI media expressed through Thomsen parameters. A detailed description of Thomsen notation and its advantages for transverse isotropy can be found in Thomsen (1986) and Tsvankin (1996a). In addition to simplifying analytic expressions for various signatures, Thomsen parameters provide an easy insight into the character and magnitude of the anisotropy. The dimensionless coefficients $\epsilon$, $\delta$, and $\gamma$ go to zero if the medium is isotropic and, therefore, can be used to characterize the
strength of the anisotropy in VTI media. Likewise, the coefficients \( e^{(2)} \), \( \delta^{(2)} \), and \( \gamma^{(2)} \) quantify the magnitude of velocity anisotropy in orthorhombic media, both within and outside the symmetry planes. The parameters \( e^{(2)} \) and \( \epsilon^{(2)} \) are close to the fractional difference between vertical and horizontal P-wave velocities in the planes \([x_1, x_3]\) and \([x_2, x_3]\) (respectively) and, therefore, yield an overall measure of the “P-wave anisotropy” in these planes. Similarly, the coefficients \( \gamma^{(2)} \) and \( \gamma^{(1)} \) govern the magnitude of the velocity variation of the elliptical SH-wave in the vertical symmetry planes.

One of the most important advantages of Thomsen notation is the reduction in the number of parameters responsible for P-wave kinematic signatures. The exact P-wave phase-velocity equation for VTI media [analogous to equation (26)] contains four independent parameters \((V_P, V_S, \epsilon, \delta)\), but the influence of the shear-wave vertical velocity \(V_S\) is practically negligible, even for strong anisotropy (Tsankin and Thomsen, 1994). As discussed by Tsankin (1996a), the stiffness coefficient that determines the shear-wave vertical velocity \((C_{44} = C_{55} \text{ in VTI media})\) does not contribute to the P-wave velocity equations, but only through the parameter \(\delta\). The analogy with VTI media implies that P-wave kinematic signatures in each symmetry plane are determined by just three independent coefficients. In the vertical symmetry planes the needed parameters are the vertical P-wave velocity \(V_P\) and a pair of the anisotropic coefficients in {\(e^{(2)}\) and \(\delta^{(2)}\) (\([x_1, x_3]\) plane) or \(e^{(1)}\) and \(\delta^{(1)}\) (\([x_2, x_3]\) plane).}

If a medium is homogeneous, the vertical velocity \(V_P\) becomes no more than the scaling coefficient for P-wave velocities and traveltime, and the P-wave anisotropy in the vertical symmetry planes is governed entirely by the corresponding parameters \(\epsilon\) and \(\delta\). As in VTI media, the coefficients \(e^{(2)}\) and \(\delta^{(2)}\) are responsible for phase and group velocity in different ranges of phase angles, which is extremely convenient for purposes of seismic processing and inversion. Specifically, the coefficients \(\epsilon^{(2)}\) (plane \([x_1, x_3]\) and \(\delta^{(2)}\) (plane \([x_2, x_3]\)) determine near-vertical P-wave velocity variations [see equation (28)], as well as the anisotropic term in the expression for normal-moveout velocity from horizontal reflectors (discussed in more detail below). Also, the new notation simplifies the elliptical condition in both vertical symmetry planes: the P-wave anisotropy in the \([x_1, x_3]\) plane is elliptical if \(\epsilon^{(2)} = \delta^{(2)}\) (the SV-wave velocity in this case is constant). Similarly, elliptical anisotropy in the \([x_2, x_3]\) plane requires that \(\epsilon^{(1)}\) is equal to \(\delta^{(1)}\).

To obtain any kinematic signature (e.g., phase and group velocity, normal-moveout from horizontal reflectors etc.), polarization vector, and reflection coefficients of the \(P\) and \(SV\)-waves in the \([x_1, x_3]\) plane of orthorhombic media, it is sufficient to substitute \(V_{P_0}, V_{S_0}, \epsilon^{(2)}\) and \(\delta^{(2)}\) into VTI equations expressed through \(V_P, V_S, \epsilon\) and \(\delta\) respectively. The same analogy with vertical transverse isotropy holds for the \([x_2, x_3]\) plane, if we use VTI equations with the coefficients \(V_{P_0}, V_{S_1}, \epsilon^{(1)}\) and \(\delta^{(1)}\). We have already taken advantage of this limited equivalence between orthorhombic and VTI media to obtain the exact and approximate phase velocities in the \([x_1, x_3]\) symmetry plane [equations (26) and (28)]. Below, I briefly describe adaptation of other VTI signatures to the symmetry planes of orthorhombic media using the new anisotropic parameters. The reflection coefficients in the symmetry planes of orthorhombic media will be discussed in detail in a forthcoming paper by Andreas Rüger.

**Phase and group velocity**

The exact phase velocity of the \(P - SV\)-waves in the \([x_1, x_3]\) plane and the weak-anisotropy approximation for the P-wave phase velocity were expressed through the new parameters in equations (26) and (28), respectively (analogous expressions hold in the \([x_2, x_3]\) plane). Next, we can adapt Thomsen’s (1986) weak-anisotropy approximation for the phase velocity of the \(SV\)-wave in VTI media. In the \([x_1, x_3]\) plane, the \(SV\)-wave phase velocity linearized in the anisotropic parameters takes the form

\[
V_{SV}(\theta) \{[x_1, x_3]\ \text{plane}\} = V_{S_0} \left(1 + \sigma^{(2)} \sin^2 \theta \cos^2 \theta\right),
\]

(36)

where \(\sigma^{(2)}\) is the following combination of Thomsen parameters introduced by Tsankin and Thomsen (1994):

\[
\sigma^{(2)} \equiv \left(\frac{V_{P_0}}{V_{S_0}}\right)^2 (\epsilon^{(2)} - \delta^{(2)}).
\]

(37)

The parameter \(\sigma^{(2)}\) largely determines the kinematic signatures of the \(SV\)-wave in the \([x_1, x_3]\) symmetry plane, although the influence of the individual values of the \(V_{P_0}/V_{S_0}\) ratio, \(\epsilon^{(2)}\), and \(\delta^{(2)}\) on \(SV\) propagation becomes more pronounced with increasing anisotropic coefficients. Similarly, in the \([x_2, x_3]\) plane

\[
V_{SV}(\theta) \{[x_2, x_3]\ \text{plane}\} = V_{S_1} \left(1 + \sigma^{(1)} \sin^2 \theta \cos^2 \theta\right),
\]

(38)

\[
\sigma^{(1)} \equiv \left(\frac{V_{P_0}}{V_{S_1}}\right)^2 (\epsilon^{(1)} - \delta^{(1)}).
\]

Although the weak-anisotropy approximations for phase velocity become less accurate with increasing velocity anisotropy, they provide a valuable analytic insight into a wide range of seismic signatures (Tsankin, 1996a). It is also possible to improve the accuracy of equations (28), (36), and (38) by adding terms quadratic in the anisotropic parameters. For instance, Tsankin (1996a) presented the quadratic weak-anisotropy approximation for the P-wave velocity in VTI media that can be
directly applied in the symmetry planes of orthorhombic media.

Using the analogy with the \(SH\)-wave in VTI media (Thomsen, 1986), we can obtain the exact phase velocity of the elliptical \(SH\)-wave in the \([x_1, x_3]\) plane of orthorhombic media in the following form:

\[
V_{SH}(\theta) \{ [x_1, x_3] \text{ plane} \} = V_S \sqrt{\frac{1 + 2\gamma (\theta)^2 \sin^2 \theta}{1 + 2\gamma (\theta)^2}}
\]

\[= V_S 0 \sqrt{\frac{1 + 2\gamma (\theta)^2}{1 + 2\gamma (\theta)^2}} \sqrt{1 + 2\gamma (\theta)^2 \sin^2 \theta}.
\]  (39)

The vertical velocity of the \(SH\)-wave in the \([x_2, x_3]\) plane is equal to \(V_S\), and the phase velocity is given by

\[V_{SH}(\theta) \{ [x_2, x_3] \text{ plane} \} = V_S 0 \sqrt{1 + 2\gamma (\theta)^2 \sin^2 \theta}.
\]  (40)

Since the group-velocity vector for any slowness direction within symmetry planes cannot have out-of-plane components, the group velocity \(v_g\) in the symmetry planes of orthorhombic media represents the same function of phase velocity as in VTI media (e.g., Berryman, 1979):

\[v_g = V \sqrt{\frac{1}{1} + \left(1 \frac{dV}{d\theta} \right)^2},
\]  (41)

\[\theta\] is the phase angle with one of the coordinate axes. Hence, to compute the group velocity in any of the symmetry planes, we just have to substitute the appropriate phase-velocity function into equation (41) (e.g., from equation (25) for the \(P - SV\)-waves in the \([x_1, x_3]\) plane).

The anisotropic terms containing the first derivative of phase velocity is squared, and in the linearized weak-anisotropy approximation the group velocity expressed through the phase angle coincides with the phase velocity. However, the group velocity corresponds to the energy propagating at the group angle \(\psi\) that is given in VTI media by (Berryman, 1979)

\[\tan \psi = \frac{\tan \theta - \frac{V}{V} \frac{dV}{d\theta}}{1 - \tan \theta \frac{dV}{d\theta}}.
\]  (42)

Again, equation (42) can be used to calculate the group angle \(\psi\) in the symmetry planes of orthorhombic media by substituting the corresponding phase-velocity functions. To find the group angle with the \(x_2\) axis in the vertical symmetry planes, the phase angle \(\theta\) in equation (42) should be computed with respect to the vertical \((x_3)\) axis as well. Also, the weak-anisotropy approximations for the group angles of all three modes in VTI media (Thomsen, 1986) remain valid in the symmetry planes of orthorhombic media if we use the appropriate anisotropic coefficients. For instance, the \(P\)-wave's group angle with vertical in the \([x_1, x_3]\) plane can be represented as

\[\tan \psi = \tan \theta [1 + 2\gamma (\theta)^2] + 4\gamma (\theta)^2 \sin^2 \theta.
\]  (43)

An obvious modification of equation (43) can be used in the \([x_2, x_3]\) symmetry plane. Note that the difference between the group and phase angles [see equations (42) and (43)] contains terms linear in the anisotropic coefficients and, therefore, cannot be ignored in the linearized weak-anisotropy approximation. For wave propagation along the coordinate axis of orthorhombic media the derivative of phase velocity vanishes, and the group- and phase-velocity vectors for any wave type coincide with each other.

**Polarization vector**

For a given slowness direction in an arbitrary anisotropic medium, the polarization vectors of the three plane waves (the eigenvectors of the Christoffel equation) are always mutually orthogonal. This orthogonality of the plane-wave polarizations does not hold, however, for point-source radiation because three body waves recorded at any fixed point in space correspond to different slowness directions. Also, the polarization vector of the plane \(P\)-wave is not necessarily aligned with either phase (slowness) or group (ray) vector. Likewise, in general the split shear waves are not polarized orthogonally to the phase- or group-velocity vector (this explains the qualifier “quasi” usually added to the names of body waves in anisotropic media). In some directions, however, the medium behaves isotropically and the polarization vector of the \(P\)-wave is parallel to the slowness vector; Helbig (1993) suggested to call such directions “longitudinal.” In the symmetry planes of orthorhombic media the longitudinal directions include (but are not limited to) the coordinate axes (Schoenberg and Helbig, 1996).

As mentioned above, the polarization vector in the symmetry planes of orthorhombic media is determined by the Christoffel equation that has the same form as in transversely isotropic media. Using the known expression for the polarization vector in transversely isotropic media (e.g., Helbig and Schoenberg, 1987; Tsvankin, 1996a), we can write the polarization angle of the plane \(P\)-wave in the \([x_1, x_3]\) plane of orthorhombic media as

\[\tan \gamma = \frac{U_1}{U_3} = \frac{\sin \theta \cos \theta (c_{13} + c_{33})}{\rho V^2 - c_{13} \sin^2 \theta - c_{33} \cos^2 \theta},
\]  (44)

where \(V\) is the phase velocity of the \(P\)-wave. Equation (44) is valid for any strength of the anisotropy and determines the \(SV\)-wave polarization angle as well (the polarization vectors of the \(P\)- and \(SV\)-waves are orthogonal). As discussed above, the second shear wave in the \([x_1, x_3]\) plane represents a pure shear \((SH)\) mode polarized in the \(z_3\) direction. With the appropriate substitutions of the elastic constants [see relations (30)], equa-
tion (44) can be used to obtain the polarization angles in the \([x_2, x_3]\) symmetry plane.

Adapting the results of Rommel (1994) and Tsvankin (1996a) obtained for TI media, we find the following weak-anisotropy approximation for equation (44):

\[
\tan \gamma = \tan \theta \left(1 + B \left[2\delta^{(2)} + 4 \left(\epsilon^{(2)} - \delta^{(2)}\right) \sin^2 \theta\right]\right),
\]

\[
B \equiv \frac{1}{2f} = \frac{1}{2(1 - V^2_{S_0}/V^2_{P_0})},
\]  

As discussed by Tsvankin (1996a), comparison of the weak-anisotropy approximations for the group angle (43) and the polarization angle (45) shows that for weak and moderate anisotropy, the \(P\)-wave polarization vector lies closer to the corresponding ray (group vector) than to the phase vector. To obtain body-wave polarizations in the far field of a point source, it is necessary to find the phase angle \(\theta\) corresponding to a given group (source-receiver) direction and substitute it into equations (44) and (45). Near-field polarizations, however, depend on the relative amplitudes of several terms of the ray series expansion and, therefore, can be influenced by azimuthal velocity variations.

Moveout from horizontal reflectors

Suppose a common-midpoint (CMP) line is parallel to the axis \(x_1\) of a horizontal orthorhombic layer. Then the phase- and group-velocity vectors of the waves reflected from the bottom of the layer stay in the \([x_1, x_3]\) symmetry plane, and the normal-moveout (NMO) velocity can be obtained from the known VTI equations (e.g., Thomsen, 1986; Tsvankin, 1996a). Using the substitutions described above, we find for the pure (non-converted) modes

\[
V^{(2)}_{nmo} [P\text{-wave}] = V_{P0} \sqrt{1 + 2\delta^{(2)}},
\]  

\[
V^{(2)}_{nmo} [SV\text{-wave}] = V_{S0} \sqrt{1 + 2\sigma^{(2)}},
\]  

\[
V^{(2)}_{nmo} [SH\text{-wave}] = V_{S0} \sqrt{1 + 2\gamma^{(2)}},
\]

where \(\sigma^{(2)}\) was defined in equation (37).

Similarly, the \(P\)-wave NMO velocity on a line parallel to the \(x_2\) axis is given by

\[
V^{(1)}_{nmo} [P\text{-wave}] = V_{P0} \sqrt{1 + 2\delta^{(1)}}.
\]  

Moveout velocity is one of the most important parameters in reflection data processing, and the simplicity of the above expressions is a good illustration of the advantages of our notation. Note that equations (46)-(49) remain valid for any strength of velocity anisotropy. Evidently, if the vertical velocity is known, short-spread \(P\)-wave moveout in the vertical symmetry planes makes it possible to obtain both \(\delta\) coefficients. The behavior of NMO velocity outside the symmetry planes will be discussed in a forthcoming paper.

To find normal-moveout velocity in symmetry planes of multilayered laterally homogeneous orthorhombic models, it is sufficient just to apply the conventional Dix equation. This conclusion follows from the general NMO formula for any symmetry plane in anisotropic media given by Alkhalifah and Tsvankin (1995). Of course, for a stack of orthorhombic layers to have a throughgoing vertical symmetry plane, the horizontal coordinate axes in all layers have to be either aligned or rotated by 90° from layer to layer.

For relatively large spreadlengths (exceeding the reflector depth), reflection moveout becomes nonhyperbolic and can be described by the following expression developed by Tsvankin and Thomsen (1994) for vertical transverse isotropy:

\[
t^2(x) = t^2_0 + A_2 x^2 + \frac{A_4 x^4}{1 + A x^2},
\]

where \(t_0\) is the zero-offset reflection traveltime, \(A_2 = 1/V^2_{nmo}\), \(A_4\) is the quartic moveout coefficient, and

\[
A = \frac{A_4}{V^2_{P0} - A_2},
\]

where \(V_{P0}\) is the horizontal velocity. The normalization of the quartic moveout term makes equation (37) numerically accurate for \(P\)-wave reflection moveout on long spreads (2-3 times, and more, the reflector depth), even in VTI models with pronounced velocity anisotropy (Tsvankin and Thomsen, 1994). Equation (50) works well not only for a single horizontal layer but also for stratified VTI media, provided the appropriate coefficients \(A_2, A_4,\) and \(A\) are used.

The analogy with vertical transverse isotropy implies that the nonhyperbolic moveout equation (50) can be used without any modifications on CMP lines parallel to one of the horizontal coordinate axes of an orthorhombic layer (or in the symmetry planes of a layered orthorhombic medium). For instance, for a line coinciding with the \(x_1\)-axis we have to make the following substitutions in equation (50) (for the \(P\)-wave)

\[
A_2 = \frac{1}{V^2_{P0} (1 + 2\delta^{(2)})},
\]

\[
V_{P0} = V_{P0} \sqrt{1 + 2\epsilon^{(2)}},
\]

and

\[
A_4 = -\frac{2\left(\epsilon^{(2)} - \delta^{(2)}\right)}{t^2_0 V^2_{P0} (1 + 2\delta^{(2)})/f},
\]

with \(f\) given by equation (27). While equation (50) as a whole is an approximation for long-spread moveout, the above expression for the quartic coefficient \(A_4\) is exact for any strength of the anisotropy.
Dipping reflectors and 2-D processing

The equivalence between orthorhombic media and vertical transverse isotropy breaks down if the phase- or group-velocity vectors of seismic waves deviate from the vertical symmetry planes. Evidently, 3-D seismic processing in orthorhombic media cannot be carried out by just adapting the known VTI algorithms. However, if the subsurface can be approximated by a 2-D structure with the strike parallel to one of the horizontal coordinate axes of an orthorhombic model, the incident and reflected rays on a line normal to the strike (the dip line) will be confined to one of the vertical symmetry planes (or more complicated models it may be possible to suppress out-of-plane events). Normal-moveout velocity of in-plane dipping events can be described by the following equation valid for pure modes in any symmetry plane of arbitrary anisotropic media (Tsvankin, 1995):

$$V_{nmo}(\phi) = \frac{V(\phi)}{\cos \phi} \sqrt{1 + \frac{\frac{1}{v(\phi)} \frac{d^2v}{d\theta^2}}{1 - \frac{\tan \phi}{\tan \phi} \frac{dV}{d\phi} \frac{d\phi}{d\theta}}} = 0,$$  \hspace{1cm} (51)

where $\theta$ is the phase angle with vertical, and $\phi$ is the dip angle of the reflector. If the medium is isotropic, the derivatives of phase velocity vanish, and equation (51) reduces to the well-known cosine-of-dip dependence obtained by Levin (1971):

$$V_{nmo}(\phi) = \frac{V}{\cos \phi}.$$  \hspace{1cm} (52)

Equation (51) can be considerably simplified by using the weak-anisotropy approximation. Adapting the result obtained by Tsvankin (1995) for vertical transverse isotropy in the limit of small $\epsilon$ and $\delta$, we get the following expression for the P-wave NMO velocity in the $[x_1, x_3]$ plane of orthorhombic media ($|\epsilon^{(2)}| \ll 1$, $|\delta^{(2)}| \ll 1$):

$$\frac{V_{nmo}^{(2)}(\phi)}{V_{nmo}(0)} = 1 + \delta^{(2)} \sin^2 \phi \hspace{1cm} (53)$$

$$-3(\epsilon^{(2)} - \delta^{(2)}) \sin^2 \phi (2 - \sin^2 \phi).$$

Analysis of equation (54) shows that the difference $\epsilon^{(2)} - \delta^{(2)}$ determines, to a large degree, the angular behavior of the P-wave NMO velocity and the errors in conventional (i.e., isotropic) dip-moveout processing (Tsvankin, 1995).

In reflection data processing, NMO velocity is usually treated as a function of the ray parameter $p(\phi)$ that can be determined from the slope of reflections recorded on zero-offset (or CMP-stacked) seismic sections. For vertical transverse isotropy, P-wave NMO velocity from dipping reflectors [equation (51)] expressed as a function of $p$ is controlled by just two parameters: the NMO velocity from a horizontal reflector and an anisotropic coefficient defined as $\eta \equiv (\epsilon - \delta)/(1 + 2\delta)$ (Alkhalifah and Tsvankin, 1995). The parameter $\eta$ goes to zero for any elliptically anisotropic model ($\epsilon = \delta$) and, therefore, can be considered as a measure of "anellipticity" of a particular TI medium. Alkhalifah and Tsvankin (1995) have shown that all time-related processing steps (NMO, DMO, prestack and poststack time migration) in homogeneous or vertically inhomogeneous VTI media are dependent on just two effective parameters: $\eta$ and the zero-dip NMO velocity $V_{nmo}(0)$.

This conclusion remains entirely valid for P-wave reflections confined to one of the vertical symmetry planes of orthorhombic media. For instance, the kinematics of 2-D time processing in the $[x_1, x_3]$ symmetry plane is governed by the corresponding P-wave NMO velocity from a horizontal reflector [equation (46)] and a parameter $\eta^{(2)}$ given by

$$\eta^{(2)} \equiv \frac{c^{(2)} - \delta^{(2)}}{1 + 2\delta^{(2)}}.$$  \hspace{1cm} (54)

As an example, the nonhyperbolic moveout equation (50) in the $[x_1, x_3]$ plane can be rewritten as a function of just these two effective parameters and the vertical time $t_0$ (the contribution of $V_{S0}$ to the quartic term $A_4$ can be ignored):

$$t^2(x) = t_0^2 + \frac{x^2}{V_{nmo}(0)^2} - \frac{2\eta^{(2)} x^4}{[V_{nmo}(0)]^2} \left\{ t_0^2 [V_{nmo}(0)]^2 + (1 + 2\eta^{(2)} x^2) \right\}.$$  \hspace{1cm} (55)

If $\eta^{(2)} = 0$, the P-wave anisotropy in the $[x_1, x_3]$ plane is elliptical and the reflection moveout is purely hyperbolic.

As demonstrated by Alkhalifah and Tsvankin (1995), both parameters needed for P-wave time processing in VTI media can be obtained from surface P-wave data using NMO velocities from two reflectors with different dips. This algorithm holds for parameter estimation using in-plane dipping events in the vertical symmetry planes of orthorhombic media. For instance, if the strike of the structure is parallel to the $x_2$ axis, NMO velocities of reflections from horizontal and dipping planes recorded on the dip line (parallel to the $x_2$ axis) can be used to recover the zero-dip NMO velocity and the anisotropic parameter $\eta^{(2)}$ from equation (51). In principle, the coefficient $\eta^{(2)}$ can be estimated using the long-spread (nonhyperbolic) P-wave moveout from horizontal reflectors [see equation (56)], although with a much lower accuracy (Tsvankin and Thomsen, 1995). Once the two effective parameters have been determined, 2-D time processing in the $[x_1, x_3]$ symmetry plane can be carried out using the methodology developed for vertical transverse isotropy. To perform 2-D depth processing in either of the
vertical symmetry planes, it is also necessary to know the
P-wave vertical velocity.

In general, all 2-D data-processing steps (NMO, DMO, time and depth migration) in the vertical symmetry
planes of orthorhombic media can be performed by
VTI algorithms. Body-wave amplitudes, however, will
be influenced by azimuthal velocity variations (with the
exception of the plane-wave reflection coefficients), which
may cause distortions in the so-called "true-amplitude"
DMO and migration codes.

Phase Velocity Outside the Symmetry Planes

Exact phase-velocity function

While the new notation has obvious advantages in de-
scribing symmetry-plane seismic signatures, the big ques-
tion is whether it can be useful in treating wave propa-
gation outside the symmetry planes. First, let us represent
the exact phase velocity in an arbitrary slowness direc-
tion as a function of the new parameters. Solving the
Christoffel equation (5) for phase velocity leads to the
following cubic equation valid for any homogeneous an-
isotropic medium:

\[ x^3 + ax^2 + bx + c = 0, \quad (56) \]

where \( x = \rho V^2 \) and

\[ a = - (G_{11} + G_{33} + G_{33}), \quad (57) \]

\[ b = G_{11}G_{22} + G_{11}G_{33} + G_{22}G_{33} - G_{13}^2 - G_{23}^2 - G_{33}^2, \quad (58) \]

\[ c = G_{11}G_{22}G_{33} + G_{22}G_{33}G_{11} + G_{33}G_{11}G_{22} - 2G_{11}G_{22}G_{33} - 2G_{11}G_{33}G_{22} \quad (59) \]

Through a change of variables \( (x = y - a/3) \), we can
eliminate the quadratic term in equation (56) and reduce
it to the following form (e.g., Schoenberg and Helbig,
1996):

\[ y^3 + dy + q = 0, \quad (60) \]

with the coefficients

\[ d = - \frac{a^2}{3} + b, \quad (61) \]

\[ q = 2 \left( \frac{a}{3} \right)^3 - \frac{ab}{3} + c. \quad (52) \]

Using the fact that the matrix \( G_{ik} \) is symmetric, it
can be shown that the coefficient \( d \) is negative (Schoen-
berg and Helbig, 1996). For roots of equation (60) to be
real, the following combination of \( d \) and \( q \) should be non-
positive:

\[ Q = \left( \frac{d}{3} \right)^3 + \left( \frac{q}{2} \right)^2 \leq 0. \]

Then the three solutions of equation (60) can be re-
presented as

\[ y_{1,2,3} = 2 \sqrt{-\frac{d}{3} \cos \left( \frac{\nu}{3} + k \frac{2\pi}{3} \right)}, \quad (63) \]

where \( k = 0, 1, 2, \) and

\[ \cos \nu = - \frac{q}{2 \sqrt{(-d/3)^3}}; \quad 0 \leq \nu \leq \pi. \]

The phase velocity is found from

\[ \rho V^2 = y - a/3. \quad (64) \]

The largest root \( (k = 0) \) of equation (63) yields the
phase velocity of the \( P \)-wave, while the other two roots
\( (k = 1, 2) \) correspond to the split shear waves. For
equation (60) to have three distinct roots, \( Q \) should be neg-
ative; if \( Q = 0 \), two of the roots are identical. In the case
of the Christoffel equation, the two smaller roots may
coincide with each other, which leads to shear-wave sin-
gularities in certain slowness directions. In principle, the
two bigger roots (for the \( P \)-wave and one of the \( S \)-waves)
may also become identical (then the third root is differ-
ent), but this is an extremely unusual occurrence even for
lower anisotropic symmetries.

Equation (63) is valid for an arbitrary anisotropic
medium provided the appropriate Christoffel matrix is
substituted into equations (57)-(59). Here, we are inter-
ested in evaluating the phase velocity in orthorhombic
media as a function of the new anisotropic parameters
introduced above. Representing the stiffness coefficients
through the new parameters [using equations (22)-(25),
(29), (31)-(34)] and substituting the results into the ex-
pressions for the components of the Christoffel matrix
[equations (5)-(11)], we find

\[ \frac{G_{11}}{\rho} = n_1^2 V_{P0}^2 (1 + 2\varepsilon^{(2)}) + n_2^2 V_{S0}^2 (1 + 2\gamma^{(1)}) + n_3^2 V_{S0}^2; \quad (65) \]

\[ \frac{G_{22}}{\rho} = n_1^2 V_{S0}^2 (1 + 2\gamma^{(1)}) + n_2^2 V_{P0}^2 (1 + 2\varepsilon^{(1)}) + n_3^2 V_{P0}^2 \quad + 2 n_3^2 V_{S0}^2 \quad \frac{1 + 2\gamma^{(1)}}{1 + 2\gamma^{(2)}}; \quad (66) \]

\[ \frac{G_{33}}{\rho} = n_1^2 V_{S0}^2 + n_2^2 V_{S0}^2 \frac{1 + 2\gamma^{(1)}}{1 + 2\gamma^{(2)}} + n_3^2 V_{P0}^2; \quad (67) \]

\[ \frac{G_{12}}{\rho} = n_1 n_2 V_{P0}^2 (1 + 2\varepsilon^{(2)}) D \sqrt{1 - 2\delta^{(3)}/D}; \quad (68) \]

\[ D = 1 - \frac{V_{S0}^2}{V_{P0}^2} \quad \frac{1 + 2\gamma^{(1)}}{1 + 2\varepsilon^{(2)}}. \]
\[
\frac{G_{12}}{\rho} = n_1 n_3 V_{P0}^2 \sqrt{1 + 2\delta^{(2)}} / f;
\]

\[
\frac{G_{23}}{\rho} = n_2 n_3 V_{P0}^2 E \sqrt{1 + 2\delta^{(1)}} / E;
\]

\[
E = 1 - \frac{V_{P0}^2 \gamma^{(1)} + 2 \gamma^{(2)}}{V_{P0}^2 \gamma^{(1)} + 2 \gamma^{(2)}}.
\]

In equations (68)-(70) it is assumed that \( c_{12} + c_{22} \geq 0, c_{13} + c_{33} \geq 0, \) and \( c_{23} + c_{44} \geq 0, \) which corresponds to the conditions of “mild” anisotropy (Schoenberg and Helbig, 1996) discussed above. After calculating the components of the Christoffel matrix at a given slowness direction, we can compute the coefficients of the cubic equation (55) and obtain the phase velocity [equation (64)] using the trigonometric solution (63). Thus, equations (65)-(70) make it possible to use the new notation in calculating the exact phase velocity outside the symmetry planes of orthorhombic media (alternatively, we can simply express the stiffness coefficients through the new parameters). Although the Christoffel matrix becomes somewhat more complicated when represented through the new coefficients, it makes little (if any) difference in the numerical implementation.

Since Thomsen notation does not provide any tangible simplification of the exact phase-velocity equations for transversely isotropic media (Tsukanov, 1996a), our notation can hardly be expected to accomplish this task for orthorhombic models. However, as demonstrated by the symmetry-plane analysis in the previous section, the new notation may be helpful in developing concise weak-anisotropy approximations for phase and group velocity, reducing the number of parameters responsible for P-wave velocities for any strength of the anisotropy, and simplifying the exact analytic expressions for reflection seismic signatures (such as NMO velocity). Hence, our next step here is to transform the phase-velocity equations in the limit of weak anisotropy.

**Weak-anisotropy approximation for phase velocity**

Since the dimensionless anisotropic parameters introduced above should be small in weakly anisotropic media, the new notation is particularly suitable for developing weak-anisotropy approximations by expanding seismic signatures in the powers of \( \epsilon^{(1)}, \delta^{(1)}, \) and \( \delta^{(2)} \) (usually only the linear terms are retained). For the planes of symmetry it is sufficient just to adapt the known expressions for weak transverse isotropy, as described in the previous sections. The most straightforward way to obtain phase-velocity approximations outside the symmetry planes is to linearize equations (63) and (64) in the anisotropic coefficients. The derivation, given in Appendix A, leads to the following result for the P-wave phase velocity [equation (A15)]:

\[
V_p^2 = V_{P0}^2 \left[ 1 + 2n_1 \epsilon^{(2)} + 2n_2 \epsilon^{(1)} + 2n_3 \delta^{(2)} + 2n_2 n_3 \delta^{(1)} + 2n_3 n_3 \delta^{(2)} + (2\epsilon^{(2)} + \delta^{(2)}) \right].
\]

It is convenient to replace the directional cosines of the slowness (or phase-velocity) vector by the polar (\( \theta \)) and azimuthal (\( \phi \)) phase angles:

\[
n_1 = \sin \theta \cos \phi, \quad n_2 = \sin \theta \sin \phi, \quad n_3 = \cos \theta.
\]

Then, taking the square root of equation (71), we obtain the phase velocity exactly in the same form as in VTI media [equation (28)], but with azimuthally-dependent coefficients \( \epsilon \) and \( \delta \):

\[
V_p(\theta, \phi) = V_{P0} \left[ 1 + \delta(\phi) \sin^2 \theta \cos^2 \theta + \epsilon(\phi) \sin^4 \theta \right]^{1/2};
\]

\[
\delta(\phi) = \delta^{(2)} \cos^2 \phi + \delta^{(1)} \sin^2 \phi;
\]

\[
\epsilon(\phi) = \epsilon^{(2)} \cos \phi \sin^2 \phi + \epsilon^{(1)} \sin^4 \phi + (2\epsilon^{(2)} + \delta^{(2)}) \sin^2 \phi \cos^2 \phi.
\]

In both vertical symmetry planes, equation (72) reduces to Thomsen’s (1986) weak-anisotropy approximation for transversely isotropic media. Indeed, in the \( \{x_1, x_3\} \) plane (corresponding to \( \phi = 0 \)) the anisotropic coefficients in equation (72) become \( \delta(\phi = 0) = \delta^{(2)} \) and \( \epsilon(\phi = 0) = \epsilon^{(2)} \), while in the \( \{x_2, x_3\} \) plane \( \delta(\phi = 90^\circ) = \delta^{(1)} \) and \( \epsilon(\phi = 90^\circ) = \epsilon^{(1)} \).

The structure of equation (72) is similar to the expansion of P-wave phase velocity in a series of spherical harmonics developed by Sayers (1994a). Here, however, instead of describing the medium by perturbations of the stiffness coefficients (as was done by Sayers), we have obtained a concise approximation in terms of the dimensionless anisotropic parameters \( \epsilon^{(1)} \) and \( \delta^{(1)} \).

**Parameters responsible for P-wave velocity**

It should be emphasized that equation (72) does not contain any of the three parameters \( (V_{P0}, \gamma^{(1)}, \gamma^{(2)}) \) that describe the shear-wave velocities in the directions of the coordinate axes. Evidently, kinematic signatures of P-waves in weakly anisotropic orthorhombic media depend on just five anisotropic coefficients \( (\epsilon^{(1)}, \delta^{(1)}, \epsilon^{(2)}, \delta^{(2)}, \delta^{(3)}) \) and the vertical velocity. The equivalence between the symmetry planes of orthorhombic media and transverse isotropy, discussed in detail above, implies that these six parameters are sufficient to describe P-wave phase velocity in all three symmetry planes, even for strong velocity anisotropy. An important question, to be addressed next, is whether P-wave phase velocity depends only on
the same six parameters outside the symmetry planes and in the presence of significant velocity variations.

In the numerical examples below, I examine the influence of the parameters $V_{S0}$, $\gamma^{(1)}$, and $\gamma^{(2)}$ on the exact phase velocity of $P$-wave for orthorhombic models with moderate and strong velocity anisotropy. It is sufficient to consider the velocity function in a single octant, for instance, the one corresponding to $0 \leq \theta \leq 90^\circ$, $0 \leq \phi \leq 90^\circ$. Figure 2 shows the dependence of $P$-wave phase-velocity variations in four vertical planes on the shear-wave vertical velocity $V_{S0}$ (or the $V_{P0}/V_{S0}$ ratio), with all other model parameters being fixed. Due to the equivalence between the $[x_1, x_2]$ symmetry plane and VTI media, the $P$-wave velocity at the zero azimuth ($\phi = 0$) should be practically independent of $V_{S0}$ provided the parameters $V_{P0}$, $\epsilon^{(2)}$ and $\delta^{(2)}$ are held constant (Tsvankin, 1996). As demonstrated by Figure 2, the influence of $V_{S0}$ within a wide range of the $V_{P0}/V_{S0}$ ratios is negligible not only at $\phi = 0$, but outside the symmetry planes as well. This conclusion holds for two models with more pronounced phase-velocity variations shown in Figure 3: note that even for a medium with uncommonly strong velocity anisotropy (Figure 3b) the difference between the phase-velocity curves corresponding to the two extreme values of the $V_{P0}/V_{S0}$ ratio remains barely visible. The contribution of $V_{S0}$ to the $P$-wave phase velocity becomes somewhat more noticeable only for models with close values of the $P$ and shear-wave velocities in one of the coordinate directions. However, these models, which “almost” violate the conditions of “mild anisotropy,” can be regarded as unlikely for earth’s subsurface. It should be emphasized that if the conventional notation is used, the influence of the stiffness coefficients $c_{44}$, $c_{55}$, and $c_{66}$ on $P$-wave velocity cannot be ignored. For instance, $c_{55}$ does make a contribution to the value of $\delta^{(2)}$ [equation (25)]. As in VTI media, our notation makes it possible to reduce the number of parameters responsible for $P$-wave velocity by combining the stiffness coefficients $c_{12}$ and $c_{13}$ in the single parameter $\delta^{(2)}$; the same holds for the stiffnesses $c_{44}$ and $c_{66}$ as well.

The $P$-wave phase velocity at two extreme values of the parameter $\gamma^{(1)}$ is shown in Figure 4. Since none of the elastic constants governing $P$-wave propagation in the $[x_1, x_2]$ plane depends on $\gamma^{(1)}$ (with $V_{S0}$ being fixed), the plot starts at an azimuth of $20^\circ$. Although the influence of $\gamma^{(1)}$ slightly increases in the vicinity of the $[x_2, x_3]$ plane ($\phi = 90^\circ$), the contribution of this parameter can be safely ignored at all azimuthal angles. Similar plots showing the independence of $P$-wave velocity of the coefficient $\gamma^{(2)}$ are omitted.

Thus, $P$-wave phase-velocity variations in orthorhombic media are governed by just five anisotropic parameters ($\epsilon^{(1)}$, $\delta^{(1)}$, $\epsilon^{(2)}$, $\delta^{(2)}$, and $\delta^{(3)}$), with the vertical $P$-wave phase velocity being no more than a scaling coefficient (in homogeneous media). The 3-D phase-velocity (or slowness) function fully determines the group-velocity (ray) vector and, therefore, all other kinematic signatures (e.g., reflection traveltime). This means
that the five anisotropic parameters listed above and the vertical velocity are sufficient to describe all kinematic properties of P-waves in orthorhombic media.

Furthermore, the influence of orthorhombic anisotropy with a horizontal symmetry plane on reflection traveltimes is largely determined by the four anisotropic parameters defined in the vertical planes of symmetry ($\varepsilon^{(1)}, \delta^{(1)}, c^{(2)},$ and $\delta^{(2)}$). Since the contribution of the fifth coefficient, $\delta^{(3)}$, is mostly limited to near-horizontal velocity variations outside the vertical symmetry planes [see equation (72)], it cannot be expected to make a tangible contribution to reflection traveltimes for typical orthorhombic models.

**Accuracy of the weak-anisotropy approximation**

Now that we have proved that the weak-anisotropy approximation (72) contains all parameters responsible for P-wave velocity in orthorhombic media, we can study the dependence of the error of equation (72) on these anisotropic coefficients. As illustrated by Figures 5 and 6, equation (72) provides an excellent approximation for the exact phase velocity in media with moderate velocity anisotropy, both within and outside the symmetry planes. The model in Figure 6, taken from from Schoenberg and Helbig (1996), corresponds to an effective orthorhombic medium formed by parallel vertical cracks embedded in a background medium with the VTI symmetry.

Even in the model with uncommonly strong velocity anisotropy shown in Figure 7, equation (72) deviates from the exact solution by no more than 10%. Also, note that the weak-anisotropy approximation does not deteriorate outside the symmetry planes: in fact, it is in the symmetry planes where we observe the maximum error for some models. Although the accuracy of the weak-anisotropy approximation becomes lower with increasing anisotropic coefficients, the error remains relatively small for polar angles $\theta$ up to about 70°. Higher errors of equation (72) near horizontal are not surprising since our notation is designed to provide a better approximation for near-vertical velocity variations. To increase the accuracy of the weak-anisotropy approximation at near-horizontal directions (which may be important in cross-hole studies), the “isotropic” parameter should represent one of the horizontal velocities rather than the vertical velocity $V_P$.

Equation (72) can be used to build analytic weak-anisotropy solutions for such signatures as group velocity, polarization angle, point-source radiation pattern, etc. However, since these solutions require multiple additional linearizations and involve the derivatives of phase
velocity, their accuracy may be much lower than that of the phase-velocity expression itself.

Discussion and Conclusions

Analytic description of wave propagation in orthorhombic media can be significantly simplified by replacing the stiffness coefficients with two “isotropic” vertical velocities and a set of anisotropy parameters similar to the coefficients $\epsilon$, $\delta$, and $\gamma$ suggested by Thomsen (1986) for vertical transverse isotropy. The definitions of the anisotropy parameters introduced here are based on the analogous form of the Christoffel equation in the symmetry planes of orthorhombic and transversely isotropic media. This analogy implies that all kinematic signatures of body waves, plane-wave polarizations, and reflection coefficients in the symmetry planes of orthorhombic media are given by the same equations (with the appropriate substitutions of $a_{ij}$'s) as for vertical transverse isotropy. Since Thomsen notation provides particularly concise expressions for seismic signatures in VTI media (Tsankin, 1996a), the new parameters can be conveniently used to obtain velocities and polarizations in the symmetry planes of orthorhombic media by adapting the known VTI results. This approach yields simple analytic expressions for such symmetry-plane signatures as phase, group, and normal-moveout velocity, nonhyperbolic (long-spread) reflection moveout, and plane-wave polarization angle. The equivalence with VTI media can even be used to get analytic expressions for reflection moveout in throughgoing symmetry planes of layered orthorhombic media.

All 2-D seismic processing steps in the vertical symmetry planes can be carried out using the algorithms developed for vertical transverse isotropy. To perform 2-D P-wave time processing (NMO, DMO, time migration) in either symmetry plane, it is necessary to obtain the corresponding anisotropy coefficient $\eta$ introduced by Alkhalifah and Tsvankin (1998). The conclusion by Alkhalifah and Tsvankin that the parameter $\eta$ and the zero-dip NMO velocity are sufficient for time processing in VTI
media remains entirely valid in the symmetry planes of orthorhombic media.

It should be mentioned, however, that the equivalence between the symmetry planes of orthorhombic media and vertical transverse isotropy does not apply to point-source radiation and body-wave amplitudes in general. Although out-of-plane phenomena cannot influence the kinematics of wave propagation in the planes of symmetry, radiation patterns strongly depend on azimuthal velocity variations and, therefore, should be studied using analytic and numerical methods developed for azimuthally anisotropic media (Tsvankin and Chesnokov, 1990a; Gajewski, 1993). The influence of out-of-plane velocity variations on body-wave amplitudes also means that near-field polarizations in the symmetry planes of orthorhombic media may be different from those in VTI media. Indeed, polarization in the “nongeometrical” region close to the source is generally nonlinear and depends on the relative amplitudes of at least two first terms of the ray-series expansion (Tsvankin and Chesnokov, 1990a). In contrast, far-field polarizations are adequately described by the geometrical-seismics (plane-wave) approximation that can be obtained by analogy with vertical transverse isotropy.

The dimensionless anisotropic coefficients conveniently characterize the magnitude of anisotropy and represent a natural tool for developing weak-anisotropy approximations outside the symmetry planes of orthorhombic media. The linearized weak-anisotropy approximation for P-wave phase velocity in terms of the new parameters has the same form as for vertical transverse isotropy, but with azimuthally-dependent coefficients ε and δ. This expression provides sufficient accuracy even in models with pronounced velocity anisotropy and can be used to develop weak-anisotropy solutions for the group (ray) angle, polarization vector etc.

The phase-velocity approximation discussed above also shows that the kinematics of P-wave in weakly anisotropic orthorhombic models is independent of the three parameters responsible for shear-wave velocities along the coordinate axes (the influence of c44, c55, and c66 is absorbed by the δ coefficients). Although the error of the weak-anisotropy approximation grows with increasing anisotropy, the exact phase velocity is still governed by only five anisotropic parameters (ε(1), δ(1), ε(2), δ(2), and δ(3)) and the vertical P-wave velocity. This means that all kinematic signatures of P-waves in orthorhombic media depend on six parameters (rather than nine in the conventional notation), one of which – the vertical P-wave velocity – represents just a scaling coefficient in homogeneous media. Furthermore, reflection travel times for orthorhombic media with a horizontal symmetry plane are weakly dependent on the anisotropy coefficient δ(3). In each symmetry plane, the kinematics of P-waves is determined by three parameters – a pair of the corresponding coefficients ε and δ and the P-wave velocity along one of the in-plane coordinate axes. Clearly, the new parameters capture the combinations of the stiffness coefficients responsible for P-wave kinematic signatures in orthorhombic media. This result leads to considerable simplifications in P-wave inversion algorithms for orthorhombic media that should be devised to target these anisotropic coefficients instead of the original stiffness components.

It is important to emphasize that the notation introduced here is convenient to use in orthorhombic media with any strength of the anisotropy. The most important advantages of the new anisotropic coefficients, such as the reduction in the number of parameters responsible for P-wave velocity and the simple expressions for seismic signatures in the symmetry planes, remain valid even in strongly anisotropic models. In a companion paper (Tsvankin, 1996d) shows that the exact NMO velocity outside the symmetry planes of a horizontal orthorhombic layer represents a simple function of the two δ coefficients defined in the vertical symmetry planes. Hence, concise weak-anisotropy approximations can be regarded as just another advantage of this notation, as is the case with Thomsen notation in VTI media (Tsvankin, 1996a).

In addition, the new parameters make up a unified framework for an analytic description of seismic signatures in orthorhombic, VTI and HTI models. Both vertical and horizontal transverse isotropy can be characterized by specific subsets of the dimensionless anisotropic parameters defined here for the more complicated orthorhombic model. For instance, the parameters ε(1), δ(1), ε(2), δ(2), and δ(3) introduced for HTI media by Tsvankin (1996b) and Rüger (1996) are equivalent to the set of the anisotropic coefficients in one of the vertical symmetry planes (depending on the orientation of the symmetry axis). This uniformity of notation simplifies comparative analysis of seismic signatures and transition between different anisotropic models in seismic inversion. On the whole, the new parameters provide an analytic basis for extending inversion and processing algorithms to media with orthorhombic symmetry, which is believed to be typical for realistic fractured reservoirs.

Acknowledgments
I am grateful to Dirk Gajewski (University of Hamburg), Francis Muir (Stanford), Patrick Rasolofoson (IFP), Leon Thomsen (Amoco) and members of the A(nisotropy)-Team at the Center for Wave Phenomena
(CWP) for useful discussions. Special thanks to Jack Cohen (CSM) for his helpful suggestions and for taking up the painful task of converting this paper into the two-column format. The support for this work was provided by the members of the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP, Colorado School of Mines, and by the United States Department of Energy (project "Velocity Analysis, Parameter Estimation, and Constraints on Lithology for Transversely Isotropic Sediments" within the framework of the Advanced Computational Technology Initiative).

*REFERENCES*


Helbig, K., Longitudinal directions in media with arbitrary anisotropy: Geophysics, 58, 680-691.


Tsvankin, I., 1996a, P-wave signatures and notation for transversely isotropic media: An overview: Geophysics, 61, 467-483.


Tsvankin, I., 1996d, Analytic description of group and moveout velocity in azimuthally anisotropic media: CWP-203, this volume.


**APPENDIX A:** The weak-anisotropy approximation for P-wave phase velocity

The exact P-wave phase velocity in orthorhombic media can be obtained from equations (63) (with $k = 0$) and (64) of the main text:
\[ \rho V^2 = 2 \sqrt{\frac{-d}{3}} \cos \left( \frac{\nu}{3} \right) - \frac{a}{3}, \]  
(A1)

where

\[ \cos \nu = -\frac{q}{2\sqrt{-d/3}}; \quad 0 \leq \nu \leq \pi; \]

d and q are defined in equations (61) and (62), a is given by equation (57).

Our goal is to obtain the linearized weak-anisotropy approximation for \( V^2 \) by expanding \( p, \nu, \) and \( a \) in the anisotropic parameters \( \epsilon^{(i)}, \delta^{(i)}, \) and \( \gamma^{(i)} \) and dropping quadratic and higher-order terms. The cosine function in equation (A1) can be represented as

\[ \cos \left( \frac{\nu}{3} \right) \approx 1 - \frac{\nu^2}{18} + \frac{1}{4!} \frac{\nu^4}{81} - \ldots. \]

Straightforward algebraic transformations show that in the absence of anisotropy \( \nu = 0 \). This implies that \( \nu \), after being expanded in the anisotropic parameters, is composed of linear and higher-order anisotropic terms with no purely “isotropic” contribution. Therefore, \( \nu^2 \) contains only quadratic or higher-order terms in the anisotropic coefficients, and in the linearized weak-anisotropy approximation \( \cos(\nu/3) \) can be replaced with unity. Then the phase-velocity equation (A1) reduces to

\[ \rho V^2 \approx 2 \sqrt{\frac{-d}{3}} - \frac{a}{3}. \]  
(A2)

The next step is to obtain the weak-anisotropy approximation for the term \( \sqrt{-d/3} \). Using equations (57), (58), and (61) of the main text, \( d \) can be expressed in the following way through the components of the Christoffel matrix:

\[ d = -\frac{1}{3} (G_{11}^2 + G_{22}^2 + G_{33}^2 + 3G_{12}^2 + 3G_{13}^2 + 3G_{23}^2) - G_{11}G_{22} - G_{11}G_{33} - G_{22}G_{33}, \]  
(A3)

Substituting the expressions for \( G_{ik} \) given in the main text [equations (65)-(70)] into equation (A3) and retaining the terms linear in the anisotropic coefficients, we find

\[ -3d = n_2^2 A_1 + n_2^2 A_2 + n_2^2 A_3 + n_2^2 n_3^2 A_4 + n_2^2 n_3^2 A_5 + n_2^2 n_3^2 A_6, \]  
(A4)

where

\[ A_1 = V_{p0}^2 (1 + 4\epsilon^{(1)} + \gamma^{(8)}) + V_{s0}^2 (1 + 2\gamma^{(1)}); \]  
(A5)

\[ A_2 = V_{p0}^2 (1 + 4\epsilon^{(1)} + \gamma^{(8)}) + V_{s0}^2 (1 + 2\gamma^{(8)} + 2\gamma^{(1)}); \]

\[ -2V_{p0}^2 V_{s0}^2 \left( 1 + 2\gamma^{(1)} + \gamma^{(8)} + \gamma^{(1)} \right); \]

\[ A_3 = V_{p0}^2 + V_{s0}^2 (1 + 2\gamma^{(1)}) - 2V_{p0}^2 V_{s0}^2 (1 + \gamma^{(1)}); \]

\[ A_4 = -V_{p0}^2 (1 + 2\epsilon^{(1)} + 2\epsilon^{(2)} + \gamma^{(1)} + 2\gamma^{(8)}); \]

\[ + 2V_{p0}^2 V_{s0}^2 \left( 1 + \epsilon^{(1)} + \epsilon^{(2)} + \epsilon^{(3)} + 2\gamma^{(8)} \right); \]

\[ + 3V_{p0}^2 f^2 \left\{ 1 + \frac{2}{f} \left[ 2\gamma^{(2)} - \frac{2V_{p0}^2}{V_{p0}} \gamma^{(1)} + \delta^{(1)} \right] \right\} \]  
(A8)

\[ A_5 = -V_{p0}^2 (1 + 2\epsilon^{(1)} + \gamma^{(8)}); \]

\[ + 2V_{p0}^2 V_{s0}^2 \left( 1 + \epsilon^{(2)} - \gamma^{(8)} - \gamma^{(1)} \right); \]

\[ + 3V_{p0}^2 f^2 \left( 1 + \frac{2\delta^{(2)}}{f} \right); \]

(A9)

\[ A_6 = -V_{p0}^2 (1 + 2\epsilon^{(1)} + 3\gamma^{(8)}); \]

\[ + 2V_{p0}^2 V_{s0}^2 \left( 1 + \epsilon^{(1)} + \gamma^{(8)} + \gamma^{(1)} \right); \]

\[ + 3V_{p0}^2 f^2 \left\{ 1 + \frac{2}{f} \left[ \frac{2V_{p0}^2}{V_{p0}} \gamma^{(8)} + \delta^{(1)} \right] \right\}. \]  
(A10)

It is convenient to transform equation (A4) by separating the isotropic and anisotropic components:

\[ -3d = (V_{p0}^2 - V_{s0}^2)^2 + AN (V_{p0}^2 - V_{s0}^2), \]  
(A11)

where \( AN \) is an anisotropic term given by

\[ AN = n_2^4 B_1 + n_2^4 B_2 + n_2^4 B_3 + n_2^4 B_4 + n_2^2 n_3^2 B_5 + n_2^2 n_3^2 B_6. \]  
(A12)

The coefficients \( B_i \) are obtained from equations (A5)-(A10) as

\[ B_1 = 4V_{p0}^2 \epsilon^{(2)} - 2V_{s0}^2 \gamma^{(1)}; \]

\[ B_2 = 4V_{p0}^2 \epsilon^{(1)} - 2V_{s0}^2 (\gamma^{(8)} + \gamma^{(1)}); \]

\[ B_3 = -2V_{s0}^2 \gamma^{(8)}; \]

\[ B_4 = 2V_{p0}^2 (5\epsilon^{(2)} - \epsilon^{(1)} + 3\epsilon^{(3)}) - 2V_{s0}^2 (\gamma^{(8)} + 2\gamma^{(1)}); \]

\[ B_5 = 2V_{p0}^2 (-\epsilon^{(2)} + 3\epsilon^{(3)}) - 2V_{s0}^2 (\gamma^{(8)} + \gamma^{(1)}); \]

\[ B_6 = 2V_{p0}^2 (-\epsilon^{(1)} + 3\epsilon^{(3)}) - 2V_{s0}^2 (2\gamma^{(8)} + \gamma^{(1)}). \]

Taking the square root of equation (A11) and linearizing in the anisotropic coefficients yields

\[ \sqrt{-3d} = V_{p0}^2 - V_{s0}^2 + \frac{1}{2} AN. \]  
(A13)

To obtain phase velocity from equation (A2), it is also necessary to find the weak-anisotropy approximation for \( a = -(G_{11} + G_{22} + G_{33}) \) [equation (57)]. Using equations (65)-(70) for the components of the Christoffel matrix \( G_{ik} \), we get the following linearized expression:
\[-a = V_{P0}^2 + 2V_{S0}^2 + 2n_2^2 \left( V_{P0}^2 \epsilon^{(2)} + V_{S0}^2 \gamma^{(1)} \right) \\
+ 2n_2^2 \left( V_{P0}^2 \epsilon^{(1)} + V_{S0}^2 \left( \gamma^{(5)} + \gamma^{(1)} \right) \right) \\
+ 2n_2^2 V_{S0}^2 \gamma^{(S)}. \quad (A14)\]

Now we substitute equations (A13), (A12), and (A14) into the phase-velocity expression (A2) to obtain a concise approximation for \( P \)-wave phase velocity:

\[V_P^2 = V_{P0}^2 \left[ 1 + 2n_1^2 \epsilon^{(2)} + 2n_2^2 \epsilon^{(1)} + 2n_1^2 n_2^2 \delta^{(2)} \\
+ 2n_2^2 \delta^{(1)} + 2n_1^2 n_2^2 \left( 2\epsilon^{(2)} + \delta^{(3)} \right) \right]. \quad (A15)\]
Analytic description of group and moveout velocity in azimuthally anisotropic media

Ilya Tsvankin
Center for Wave Phenomena, Department of Geophysics
Colorado School of Mines

ABSTRACT

Analytic description of seismic signatures in azimuthally anisotropic media is of primary importance in characterization of fractured reservoirs. The formalism developed here provides a convenient way for modeling and inverting group and normal-moveout (NMO) velocity in media with orthorhombic and lower symmetry. Group velocity in arbitrary anisotropic media can be expressed directly through the phase-velocity function and its derivatives; furthermore, after an appropriate coordinate rotation two components of the group-velocity vector can be obtained from the same equations as for vertical transverse isotropy (VTI media). The general group-velocity expressions were combined with the weak-anisotropy approximation for phase velocity to obtain simple linearized equations for the group angles in orthorhombic media.

The relation between group and phase velocity was also used to derive the exact normal-moveout velocity for an azimuthally anisotropic layer with a horizontal symmetry plane (e.g., the medium can be orthorhombic or monoclinic). Normal moveout for any pure mode depends on a single function of the azimuthal phase angle that governs near-vertical phase-velocity variations. For the P-wave in orthorhombic media, the azimuthal dependence of NMO velocity represents an ellipse with the semi-axes in the vertical symmetry planes. Therefore, three or more azimuthal P-wave NMO measurements can be inverted for the orientation of the symmetry planes and the normal-moveout velocities within them. If the vertical velocity is known, NMO velocity is sufficient to obtain the values of two anisotropic parameters equivalent to Thomsen’s coefficient $\delta$.

Normal moveout from dipping reflectors was studied under the assumption of weak anisotropy and only for common-midpoint (CMP) lines in the dip plane of the reflector. NMO velocity and all other kinematic P-wave signatures in any vertical plane of weakly anisotropic media with orthorhombic symmetry are described by the same equations as vertical transverse isotropy. This implies that all 2-D P-wave processing algorithms developed for VTI media remain entirely valid in orthorhombic media, as long as velocity variations are small. The analogy with vertical transverse isotropy was used to obtain concise expressions for the dip-dependent P-wave NMO velocity in terms of the azimuthally-varying anisotropic parameters $\epsilon$ and $\delta$.

Key words: azimuthal anisotropy, group velocity, normal-moveout velocity

Introduction

Azimuthally anisotropic models are used to describe fractured reservoirs that contain one or more systems of vertical cracks. It was realized more than a decade ago that seismic methods can provide valuable information about the orientation and physical properties of crack systems (Crampin, 1985). Although most existing investigations
of fractured media are limited to the analysis of split shear waves at near-vertical incidence (e.g., Thomsen, 1988), some recent experimental studies have demonstrated the sensitivity of P-wave \* data to the presence of azimuthal anisotropy (Lynn et al., 1995). Clearly, interpretation of the azimuthal variations in such commonly used reflection signatures as normal-moveout (NMO) velocity, small-angle reflection coefficient etc. is impossible without an analytic insight into the influence of azimuthal anisotropy on seismic wavefields. The dependence of seismic signatures on the anisotropy parameters is relatively well understood for the transversely isotropic model with a vertical symmetry axis (VTI media) that does not exhibit azimuthal anisotropy. A summary of recent advances in the analytic description of seismic velocities and amplitudes in VTI models can be found in an overview paper by Tsvankin (1996a). Some of the results obtained for vertical transverse isotropy can be directly applied to TI models with a horizontal symmetry axis (HTI media) used to describe the simplest fractured formations that contain parallel vertical penny-shaped cracks in a purely isotropic background. Evidently, such signatures as phase and group velocity, polarization vector, and point-source radiation pattern can be expressed in the same form for any homogeneous TI model, whether the symmetry axis is vertical or horizontal. The situation with reflection seismic signatures, however, is much more complicated. For instance, normal moveout from horizontal interfaces in HTI media is influenced by azimuthal velocity variations and cannot be described by the known VTI equations. The exact expression for azimuthally dependent normal-moveout (NMO) velocity of any pure mode in an HTI layer was recently presented by Tsvankin (1996b). Rüger (1996) and Rüger and Tsvankin (1995) obtained the weak-anisotropy approximations for the reflection coefficients in HTI media and suggested a method of fracture characterization based on the azimuthal variation in the P-wave reflectivity. Horizontal transverse isotropy, however, is a relatively restrictive model that cannot be used to characterize fractured reservoirs with two vertical crack systems, non-aligned cracks, or an anisotropic background medium. Realistic fractured media may well have orthorhombic (Wild and Crampin, 1991), monoclinic or even the lowest, triclinic symmetry. Orthorhombic models have three mutually orthogonal symmetry planes, in which the Christoffel equation has the same form as in transversely isotropic media (Musgrave, 1970; Schoenberg and Helbig, 1996). Therefore, body-wave velocities and polarizations in the symmetry planes of orthorhombic media are described by the same equations as for VTI media. Tsvankin (1996c) used this limited analogy with VTI media to introduce dimensionless anisotropic parameters for orthorhombic media defined similarly to Thomsen's (1986) VTI coefficients $\epsilon$, $\delta$, and $\gamma$. This notation allowed Tsvankin (1996c) to give a systematic description of kinematic signatures and polarizations in the symmetry planes of orthorhombic media by adapting the VTI equations represented through Thomsen parameters. The dimensionless parameters also proved to be well-suited for developing a concise weak-anisotropy approximation for phase velocity outside the symmetry planes of orthorhombic models (Tsvankin, 1996c). Even more importantly, Tsvankin shows that the new notation makes it possible to reduce the number of independent parameters responsible for P-wave kinematic signatures from nine (in the conventional notation) to six. However, these developments represent no more than the first steps in the analytic description of seismic signatures in azimuthally anisotropic media. Here, I present concise analytic expressions for the group-velocity vector and normal-moveout velocity valid outside the symmetry planes of anisotropic models. First, the exact group-velocity vector for arbitrary anisotropic media is represented as a simple function of phase velocity and its derivatives with respect to the phase angles. The group-velocity expression is then used to obtain an exact equation for normal-moveout velocity from horizontal reflectors in anisotropic media with a horizontal symmetry plane. These general results are applied to study group and NMO velocity in media with orthorhombic symmetry. To gain analytic insight into the behavior of group velocity, the exact expression is transformed into a simple weak-anisotropy approximation that describes the orientation of the group-velocity vector. The azimuthal dependence of the exact P-wave NMO velocity in orthorhombic media is shown to represent an ellipse in the horizontal plane with the semi-axes in the vertical symmetry planes. Also, the NMO velocity from dipping reflectors outside the symmetry planes is studied as a function of the dimensionless anisotropic parameters introduced by Tsvankin (1996c).

**General treatment for azimuthal anisotropy**

**Group velocity**

The group-velocity vector defines the direction of energy propagation (seismic rays) and, therefore, is of primary importance in seismic traveltime methods. While the

* The qualifiers in "quasi-P-wave" and "quasi-S-wave" will be omitted.
Anisotropic group and NMO velocity

\[ \vec{V}_G = \frac{\partial(kV)}{\partial k_x} \vec{z} + \frac{\partial(kV)}{\partial k_y} \vec{y} + \frac{\partial(kV)}{\partial k_z} \vec{z}, \]

where \( V \) is the phase velocity, \( k \) is the wave vector, which is parallel to the phase-velocity vector and has the magnitude \( k = \omega/V \) (\( \omega \) is the angular frequency), and \( \vec{z}, \vec{y}, \) and \( \vec{z} \) are the unit coordinate vectors. Differentiation with respect to each component of the wave vector has to be performed with the other two components held constant. Since both group-velocity components in the \([x, z]\) plane (\( V_{Gx} \) and \( V_{Gz} \)) are calculated for \( k_y = 0 \), they are independent from out-of-plane phase-velocity variations. As further confirmed by the derivation in Appendix A, \( V_{Gx} \) and \( V_{Gz} \) are given by the well-known expressions for the group-velocity vector in VTI media (Berryman, 1979) or in any symmetry plane in arbitrary anisotropic media (Tsvankin, 1995):

\[ V_{Gx} = \frac{\partial(kV)}{\partial k_x} = V \sin \theta + \frac{\partial V}{\partial \phi}_{\phi=\text{const}} \cos \theta. \]

\[ V_{Gz} = \frac{\partial(kV)}{\partial k_z} = V \cos \theta - \frac{\partial V}{\partial \phi}_{\phi=\text{const}} \sin \theta. \]

The transverse component of the group-velocity vector \( V_{Gz} \) depends solely on azimuthal phase-velocity variations and exists only outside the symmetry planes. As shown in Appendix A, \( V_{Gz} \) is determined by the first derivative of phase velocity with respect to the azimuthal phase angle \( \phi \):

\[ V_{Gz} = \frac{\partial(kV)}{\partial k_z} = \frac{1}{\sin \theta} \frac{\partial V}{\partial \phi}_{\phi=\text{const}}. \]

Equations (2)-(4) conveniently express the group-velocity vector in arbitrary anisotropic media through 3-D variations in the phase-velocity function. This representation has proved helpful both in deriving the exact normal-moveout velocity in azimuthally anisotropic models and in gaining insight into the behavior of group velocity in orthorhombic media.

Normal moveout from horizontal reflectors

Here, I present a general expression for normal-moveout velocity from horizontal reflectors in a homogeneous azimuthally anisotropic layer. Let us consider a common-midpoint (CMP) line that makes the azimuthal angle \( \phi = \text{const} \) with axis \( x_1 \) (Figure 2). The only assumption made about the character of the anisotropy is that the horizontal plane (reflector) represents a plane of symmetry of the incidence medium (i.e., the model can be HTI, orthorhombic, or monoclinic). In this case, as discussed by Tsvankin (1996b), the group-velocity vectors (or rays) of incident and reflected waves are confined to the vertical incidence (sagittal) plane that contains the common-midpoint (CMP) line. The symmetry with respect to the
reflector also implies that the incidence and reflection angles are equal to each other, and the specular reflection point always coincides with the projection of the CMP on the reflector. The phase-velocity vectors corresponding to CMP reflections, however, deviate from the incidence plane, and the azimuthal angle $\phi$ of the phase-velocity vector associated with the reflected ray is different from the angle $\alpha$.

The derivation starts with the general equations for normal-moveout velocity described in detail by Tsvankin (1996b). NMO velocity is conventionally defined as

$$V_{nmo}^2 = \lim_{z \to 0} \frac{d(z^2)}{d(t^2)},$$

(5)

where $z$ is the source-receiver offset in the CMP gather and $t$ is the corresponding two-way traveltime.

Transformation of equation (5) in terms of one-way traveltime yields (Hale et al., 1992)

$$V_{nmo}^2 = \frac{2}{t_0} \lim_{h \to 0} \left[ \frac{d}{dh} \left( \frac{dt}{dh} \right) \right],$$

(6)

where $t$ is the one-way traveltime from the zero-offset reflection point to the receiver, $h = x/2$ is half the source-receiver offset, and $t_0$ is the two-way traveltime along the zero-offset ray. Equation (6) is based on the assumption that the specular reflection point is independent of offset, which is true in our case. It should be mentioned, however, that reflection-point dispersal cannot change the NMO velocity anyway because its influence is limited to higher-order terms of the traveltime series (Hubral and Krey, 1980, Appendix D; Tsvankin, 1995).

Next, it is convenient to introduce the ray parameter (horizontal slowness) $p$ using the fact that the derivative $dt/dh$ in equation (6) is equal to the projection of the slowness vector on the CMP line. Further steps of the derivation, described in detail in Appendix B, involve representing the half-offset $h$ through the group angle and using group-velocity equations (2)-(4) to express the right-hand side of equation (6) through phase velocity and phase angles. For a medium with a horizontal symmetry plane, NMO velocity from a horizontal reflector is shown to be fully determined by a single anisotropic function $a(\phi)$ dependent on the second derivative of phase velocity with respect to the polar phase angle $\theta$:

$$a(\phi) = \frac{1}{2V_0} \left. \frac{\partial^2 V}{\partial \theta^2} \right|_{\theta=0},$$

(7)

where $V_0$ is the vertical velocity, and the derivative is evaluated at a constant azimuthal phase angle $\phi$.

The final equation for NMO velocity expressed through $a(\phi)$ is given in Appendix B:

$$V_{nmo}^2 = V_0^2 \left[ 1 + 2a(\phi) + \frac{a''(\phi)^2}{1 + 2a(\phi)} \right],$$

(8)

where $\phi$ is the azimuthal phase angle corresponding to near-vertical group-velocity vectors which lie in the incidence plane $\alpha = \text{const}$. As demonstrated in Appendix B, the angle $\phi$ in equation (8) is related to the CMP azimuth $\alpha$ in the following way:

$$\tan(\alpha - \phi) = \frac{a'(\phi)}{1 + 2a(\phi)}.$$

(9)

The combination of equations (8) and (9) represents a general recipe for computing NMO velocity from horizontal reflectors in azimuthally anisotropic media with a horizontal symmetry plane. It should be emphasized that this result is valid for all pure modes and for models with any strength of the velocity anisotropy. To find an explicit expression for NMO velocity of a specific wave type, it is necessary to obtain a single anisotropic function – the second derivative of phase velocity at vertical incidence – and then solve equation (8) for the angle $\phi$. Below, we use equations (8) and (9) to obtain a simple exact expression for P-wave NMO velocity in orthorhombic media.

For reflection surveys within vertical symmetry planes, the phase-velocity vector does not deviate from the incidence plane ($a'(\phi) = 0, \phi = \alpha$), and equation (8) reduces to a simple expression

$$V_{nmo}^2 = V_0^2 \left[ 1 + 2a(\alpha) \right] = V_0^2 \left[ 1 + \frac{1}{V_0^2} \left. \frac{\partial^2 V}{\partial \theta^2} \right|_{\theta=0} \right],$$

(10)

which is identical to the symmetry-plane NMO equation.
of Tsvankin (1995) for the special case of horizontal reflectors.

For \( P \)-waves in the simplest VTI model, \( a = \text{const} \) for all azimuths, and equation (10) yields the well-known expression given by Thomsen (1986):

\[
V_{\text{nmo}}^2 = V_0^2 (1 + 2\delta),
\]

where \( \delta = a \) is the anisotropic coefficient introduced by Thomsen (1986).

In the limit of weak anisotropy, we can ignore the difference between the azimuthal phase angle \( \phi \) and the CDP azimuth \( \alpha \) since out-of-plane phenomena can influence only terms quadratic in the anisotropy parameters (this is illustrated below for orthorhombic media). Also, the term \( [a' (\phi)]^2 \), quadratic in the anisotropy coefficients, drops out of the NMO expression. As a result, for weak anisotropy equation (8) reduces to the symmetry-plane expression (10) that can be simplified further by carrying out linearization in the anisotropic term \( \frac{\partial^2 V}{\partial \phi^2} \bigg|_{\phi=0} \):

\[
V_{\text{nmo}} \approx V_0 [1 + a(\alpha)] = V_0 \left[ 1 + \frac{1}{2V_0} \frac{\partial^2 V}{\partial \phi^2} \bigg|_{\phi=0} \right].
\]

**Application to orthorhombic anisotropy**

**Description of the orthorhombic model**

Orthorhombic anisotropy describes several typical fractured models including those containing a system of parallel vertical cracks in a VTI background medium or two orthogonal crack systems (Figure 3). Media with orthorhombic symmetry have three mutually orthogonal planes of mirror symmetry: for the model with a single crack system shown in Figure 3, the vertical symmetry planes are defined by the directions parallel and normal to the cracks. The velocities and polarizations in the symmetry planes of orthorhombic media are given by the same equations as for vertical transverse isotropy. (Body-wave amplitudes in the symmetry planes, however, are influenced by the azimuthal velocity variations and require a special treatment.) Tsvankin (1996c) has taken advantage of the limited equivalence between orthorhombic and VTI media to introduce the following dimensionless anisotropic parameters defined similarly to the well-known Thomsen's (1986) coefficients \( \epsilon, \delta \) and \( \gamma \) for vertical transverse isotropy:

- \( V_{P0} \) – the vertical velocity of the \( P \)-wave:

\[
V_{P0} \equiv \sqrt{\frac{c_{33}}{\rho}}.
\]

- \( V_{S0} \) – the vertical velocity of the \( S \)-wave polarized in the \( x_1 \) direction:

\[
V_{S0} \equiv \sqrt{\frac{c_{55}}{\rho}}.
\]
\[ \delta^{(2)} = \frac{(c_{12} + c_{66})^2 - (c_{11} - c_{66})^2}{2c_{11} (c_{11} - c_{66})}. \] (21)

Therefore, the two vertical velocities and seven dimensionless anisotropic coefficients can replace the nine independent stiffness components of the orthorhombic model. Due to the equivalence with VTI media, the new parameters can be conveniently used to describe seismic velocities and polarizations in the symmetry planes of orthorhombic media using the known VTI equations expressed through Thomsen parameters (Tsvankin, 1996c). For instance, the exact normal-moveout velocities in the \([x_1, x_3]\) symmetry plane can be found by analogy with VTI media as

\[ v_{nmo}^{(2)} \] (P-wave) = \[ V_{p0} \sqrt{1 + 2\delta^{(2)}}, \] (22)

\[ v_{nmo}^{(2)} \] (SV-wave) = \[ V_{s0} \sqrt{1 + 2\sigma^{(2)}}, \] (23)

\[ \sigma^{(2)} = \left( \frac{V_{p0}}{V_{s0}} \right)^2 (\epsilon^{(2)} - \delta^{(2)}); \]

\[ v_{nmo}^{(2)} \] (SH-wave) = \[ V_{s0} \sqrt{1 + 2\gamma^{(2)}}. \] (24)

Numerical predictions for NMO velocities in the symmetry planes of orthorhombic media (not based on the formalism developed here) were verified in the physical-modeling experiments of Brown et al. (1991).

Group velocity in the symmetry planes of orthorhombic media represents the same function of phase velocity as in VTI media. The horizontal \((V_{C2})\) and vertical \((V_{G2})\) components of the group-velocity vector in the vertical symmetry planes are given by equations (2) and (3), while the transverse component \((V_{C3})\) goes to zero. As an example of the analogy with VTI media, the weak-anisotropy approximation for the P-wave group angle with vertical in the \([x_1, x_3]\) symmetry plane (\(\psi\)) takes the following form:

\[ \tan \psi = \tan \theta [1 + 2\delta^{(2)} + 4(\epsilon^{(2)} - \delta^{(2)}) \sin^2 \theta], \] (25)

which is identical to the expression presented by Thomsen (1986) for VTI media in terms of the coefficients \(\epsilon\) and \(\delta\).

Advantages of the new notation, however, are not limited to the symmetry planes. All kinematic signatures of P-waves in orthorhombic models are determined by the vertical velocity (a scaling coefficient in homogeneous media) and only five anisotropy parameters: \(\epsilon^{(1)}, \delta^{(1)}, \epsilon^{(2)}, \delta^{(2)},\) and \(\delta^{(3)}\) (Tsvankin, 1996c), as compared to nine stiffnesses in the conventional notation. Also, Tsvankin (1996c) presented a concise weak-anisotropy approximation for P-wave phase velocity in terms of these five relevant coefficients. Below, we use these results and the general formalism introduced in the previous section to describe the group and normal-moveout velocity outside the symmetry planes of orthorhombic media.

Weak-anisotropy approximation for group velocity

Due to the complexity of the phase-velocity function, equations (2)-(4) do not provide an easy insight into the dependence of group velocity on the anisotropy parameters. Here, I transform the expression for P-wave group-velocity in orthorhombic media under the assumption of weak anisotropy. The weak-anisotropy approximation for P-wave phase velocity, linearized in the dimensionless anisotropic coefficients, was given by Tsvankin (1996c):

\[ V_p(\theta, \phi) = V_{p0} [1 + \delta(\phi) \sin^2 \theta \cos^2 \theta + \epsilon(\phi) \sin^2 \theta]; \] (26)

\[ \delta(\phi) = \delta^{(3)} \cos^2 \phi + \delta^{(1)} \sin^2 \phi, \]

\[ \epsilon(\phi) = \epsilon^{(2)} \cos \phi + \epsilon^{(3)} \sin \phi \]

\[ + (\epsilon^{(2)} + \delta^{(3)}) \sin^2 \phi \cos^2 \phi. \]

Equation (26) has exactly the same form as the Thomsen's (1986) weak-anisotropy approximation for vertical transverse isotropy, but the parameters \(\epsilon\) and \(\delta\) in orthorhombic media become azimuthally dependent. In both vertical symmetry planes, equation (26) reduces to the VTI expression that includes the appropriate pair of the parameters \(\epsilon\) and \(\delta\) \((\epsilon^{(2)}\) and \(\delta^{(2)}\) in the \([x_1, x_3]\) plane and \(\epsilon^{(1)}\) and \(\delta^{(1)}\) in the \([x_2, x_3]\) plane). Substituting equation (26) into the group-velocity expressions (2)-(4), we can obtain an explicit approximation for group velocity in terms of the anisotropy parameters.

The magnitude of the group-velocity vector can be found from equations (2)-(4):

\[ V_G = V \sqrt{1 + \left( \frac{1}{V} \frac{dV}{d\phi} \right)^2 + \left( \frac{1}{V \sin \theta} \frac{dV}{d\theta} \right)^2}. \] (27)

Clearly, all anisotropic terms in equation (27) are quadratic in the anisotropy parameters. This means that in the linearized weak-anisotropy approximation the absolute values of phase and group velocity are identical, as in VTI media. However, the difference between the directions of the group- and phase-velocity vectors cannot be ignored even in the weak-anisotropy approximation. It is convenient to describe the orientation of the group-velocity vector in terms of the “in-plane” and “out-of-plane” group angles. The in-plane group angle \(\psi_1\) (Figure 1) is defined as

\[ \tan \psi_1 = \frac{V_{C2}}{V_{G2}}. \] (28)

Since the components \(V_{C2}\) and \(V_{G2}\) represent the same functions of the phase-velocity variations in the \([x_2, x_3]\) plane as in VTI media, the angle \(\psi_1\) can be expressed through phase velocity using the known VTI equations. In the weak-anisotropy limit, P-wave phase
velocity in each vertical \((x, z)\) plane is given by equation (26) that has the same form as for vertical transverse isotropy. Therefore, to obtain the weak-anisotropy approximation for the angle \(\psi_1\), it is sufficient to substitute \(\epsilon(\phi)\) and \(\delta(\phi)\) into the VTI equation (25):
\[
\tan \psi_1 = \tan \phi \left[ \frac{1 + 2\delta(\phi) + 4 \left[ \epsilon(\phi) - \delta(\phi) \right] \sin^2 \theta}{1 - 2\delta(\phi) \sin^2 \phi} \right].
\] (29)

In the \((x_1, x_3)\) symmetry plane \(\delta = \delta^{(2)}\), \(\epsilon = \epsilon^{(2)}\), and equation (29) reduces to equation (25).

The second group angle, \(\psi_2\), describes the deviation of the group-velocity vector from the vertical \((x, z)\) plane \(\phi = \text{const} \) (Figure 1):
\[
\tan \psi_2 = \frac{V_{Gy}}{V_{PO}}. \] (30)

In the linearized weak-anisotropy approximation,
\[
\tan \psi_2 = \frac{V_{Gy}}{V_{PO}}. \] (31)

Substituting the phase-velocity expression (26) into equation (4) for \(V_{Gy}\), and dividing by \(V_{PO}\) [equation (31)], we find
\[
\tan \psi_2 = \sin \theta \sin 2\phi \left[ \frac{\left( \delta^{(1)} - \delta^{(2)} \right) \cos \theta}{1 - 2\delta^{(2)} \sin^2 \phi + \delta^{(3)} \cos 2\phi} \right].
\] (32)

If the medium is azimuthally isotropic (VTI), then \(\delta^{(1)} = \delta^{(2)}, \epsilon^{(1)} = \epsilon^{(2)}, \delta^{(3)} = 0\), and the angle \(\psi_2\) is identically zero for all phase directions. Also, \(\psi_2 = 0\) in both vertical symmetry planes of orthorhombic media corresponding to \(\phi = 0\) and \(\phi = 90^\circ\). Evidently, the group-velocity vector deviates from the vertical plane that contains the phase vector only outside the symmetry planes of azimuthally anisotropic media.

Equations (29) and (32) can be efficiently used in seismic tomography to relate the traveltimes in orthorhombic media to the anisotropy parameters. For instance, the equations for the group angles can be combined with the weak-anisotropy approximation for the magnitude of the group-velocity vector [given just by the phase-velocity equation (26)] to calculate explicit analytic expressions for the Frechét derivatives of the traveltime needed in the singular value decomposition of the tomographic inverse problem.

**Exact NMO velocity from horizontal reflectors**

While NMO velocities in the symmetry planes of orthorhombic media can be easily obtained by analogy with vertical transverse isotropy [equations (22)-(24)], the azimuthal dependence of normal moveout requires a special treatment. The derivation in Appendix D based on the

Anisotropic group and NMO velocity

333

general NMO equations (8) and (9) leads to the following exact expression for P-wave NMO velocity in a single orthorhombic layer:
\[
V_{nmo}^2 = V_{PO}^2 \frac{(1 + 2\delta^{(1)}) (1 + 2\delta^{(2)})}{1 + 2\delta^{(2)} \sin^2 \alpha + 2\delta^{(1)} \cos^2 \alpha}.
\] (33)

For CMP lines within the vertical symmetry planes \(\alpha = 0^\circ\) and \(\alpha = 90^\circ\), equation (33) yields the NMO expressions obtained in Tsvankin (1996c) using the analogy with VTI media. For instance, the NMO velocity at an azimuth of \(\alpha = 0^\circ\) is given by
\[
V_{nmo}^2 = V_{PO}^2 (1 + 2\delta^{(2)}),
\]
in agreement with equation (22) for the \((x_1, x_3)\) symmetry plane.

Normal-moveout velocities for vertical and horizontal transverse isotropy can be found as special cases of equation (33). In VTI media, the \(\delta\) coefficients in the vertical symmetry planes are identical \(\delta^{(1)} = \delta^{(2)} = \delta\), and equation (33) reduces to the well-known Thomsen's (1986) expression:
\[
V_{nmo}^2 = V_{PO}^2 (1 + 2\delta).
\]

If the medium is transversely isotropic with a horizontal symmetry axis pointing in the \(z\) direction, \([x_2, x_3]\) represents the isotropy plane, and \(\delta^{(1)} = 0\). In this case, equation (33) becomes
\[
V_{nmo}^2 = V_{PO}^2 \frac{1 + 2\delta^{(2)}}{1 + 2\delta^{(2)} \sin^2 \alpha},
\] (34)

which is equivalent to the NMO expression for HTI media presented by Tsvankin (1996b) who denoted the \(\delta\) coefficient in the vertical plane that contains the symmetry axis by \(\delta^{(V)}\). It should be mentioned that any orthorhombic medium with \(\delta^{(1)} = 0\) or \(\delta^{(2)} = 0\) is fully equivalent to horizontal transverse isotropy in terms of the azimuthally-dependent P-wave NMO velocity.

In the weak-anisotropy approximation \(\left| \frac{\delta^{(1)}}{\delta^{(2)}} \right| \ll 1\), \(\left| \frac{\delta^{(2)}}{\delta^{(1)}} \right| \ll 1\), we can drop the terms quadratic in the anisotropic coefficients in equation (33) to obtain
\[
V_{nmo} = V_{PO} \left( 1 + \delta^{(2)} \cos^2 \alpha - \delta^{(1)} \sin^2 \alpha \right).
\] (35)

This expression can also be derived from the general weak-anisotropy approximation for NMO velocity [equation (12)] by substituting the second derivative of the P-wave phase velocity from equation (C15). Since equation (12), based on the assumption that the phase-velocity vector does not deviate from the incidence plane, coincides with equation (35), it is clear that out-of-plane phenomena can change only quadratic and higher-order terms in the anisotropic coefficients.

It seems that the NMO velocity from horizontal reflectors in orthorhombic media is a function of four in-
dependent parameters: the vertical velocity $V_{p0}$, the azimuthal angle $\alpha$ between the survey line and the $z_1$ axis, and the anisotropic coefficients $\delta^{(2)}$ and $\delta^{(1)}$. However, a simple transformation of equation (33) shows that $V_{p0}$, $\delta^{(1)}$, and $\delta^{(2)}$ contribute to normal moveout only through two effective parameters – NMO velocities in the $[x_1, x_3]$, $[V_{nmo}], (V_{nmo})_{1,1}$ and $[x_2, x_3]$, $[V_{nmo}]_{2,2}$ planes. Substituting

$$V_{nmo}^{(2)} = V_{nmo}^{(2)} (1 + 2\delta^{(2)})$$

and

$$V_{nmo}^{(1)} = V_{nmo}^{(1)} (1 + 2\delta^{(1)})$$

into equation (33), we find

$$V_{nmo}^2 = \frac{[V_{nmo}^{(1)}]^2 [V_{nmo}^{(2)}]^2}{[V_{nmo}^{(1)}]^2 \sin^2 \alpha + [V_{nmo}^{(2)}]^2 \cos^2 \alpha},$$

or

$$\frac{1}{V_{nmo}^2} = \frac{\cos^2 \alpha}{[V_{nmo}^{(1)}]^2} + \frac{\sin^2 \alpha}{[V_{nmo}^{(2)}]^2}.\quad (37)$$

Equation (37) shows that $P$-wave NMO velocity in orthorhombic media represents an ellipse in the horizontal plane with the semi-axes in the vertical symmetry planes (Figure 4). Azimuthal moveout measurements (a minimum of three) can be used to recover the NMO velocity in the symmetry planes and their orientation, but moveout data are not sufficient to obtain the vertical velocity or either of the $\delta$ coefficients individually.

Normal-moveout velocity in layered orthorhombic media can be found by applying the Dix (1955) rms averaging formula to the interval NMO velocities (for the same azimuth) described by equations (33) or (36). Although the Dix equation is not exact outside the symmetry planes, the results of Al-Dajani and Tsvankin (1995) indicate that it provides sufficient accuracy for azimuthally anisotropic models likely to be encountered in the subsurface.

### Approximate NMO velocity from dipping reflectors

Here, I extend the results of the previous sections to normal moveout from dipping reflectors in orthorhombic media. The discussion will be limited to $P$-wave NMO velocity in the dip plane of the reflector under the assumption of weak anisotropy. If the medium is isotropic, and the CMP gather lies in the dip plane of the reflector, the incident and reflected rays remain in the vertical incidence plane. In azimuthally anisotropic media, the rays and the corresponding phase-velocity vectors diverge from the incidence plane, but these out-of-plane phenomena can be ignored in weakly anisotropic media. Since the phase- and group-velocity vectors can be assumed to lie in the incidence plane, we can use the NMO equation derived by Tsvankin (1995) for symmetry planes of anisotropic media:

$$V_{nmo}(\gamma) = \frac{V(\gamma)}{\cos \gamma} \sqrt{1 - \frac{1}{V(\gamma)} \frac{dV}{d\phi} \frac{d\phi}{d\gamma} \frac{d\gamma}{d\phi}},$$

(38)

where $\gamma$ is the dip angle of the reflector.

For isotropic media, the derivatives of phase velocity vanish, and equation (38) reduces to the well-known cosine-of-dip expression given by Levin (1971):

$$V_{nmo}(\gamma) = \frac{V}{\cos \gamma} = \frac{V_{nmo}(0)}{\cos \gamma} \cdot (39)$$

To obtain the dip-dependence of NMO velocity at any azimuth in weakly anisotropic orthorhombic media, we need to substitute the phase-velocity function into equation (38) and carry out linearization in the anisotropic coefficients. However, since the weak-anisotropy approximation for $P$-wave phase velocity [equation (26)] is exactly the same function of the angle $\theta$ as in VTI media, we can just use the result obtained by Tsvankin (1995) for vertical transverse isotropy:

$$V_{nmo}(\gamma) = \frac{V_{nmo}(0)}{\cos \gamma} \left(1 + \delta(\phi) \sin^2 \gamma + 3 \epsilon(\phi) \sin^2 \gamma (2 - \sin^2 \gamma) \right),$$

(40)

where

$$V_{nmo}(0) = V_{p0} \left[1 + 2\delta(\phi)\right]$$

is the zero-dip NMO velocity at the azimuth $\phi$ in the weak-anisotropy approximation. The term multiplied with $V_{nmo}(0)/\cos \gamma$ represents the correction to the isotropic relation between the NMO velocities for the hori-
Anisotropic group and NMO velocity

horizontal and dipping reflector [equation (39)]. Since conventional dip-moveout (DMO) algorithms are based on the isotropic cosine-of-dip dependence, equation (40) gives analytic insight into the errors in 2-D DMO correction in orthorhombic media.

However, dip-moveout algorithms and reflection seismic processing in general operates with the ray parameter corresponding to zero-offset reflection, rather than with the (unknown) dip angle. To find the DMO error in terms of the ray parameter, it is sufficient just to adapt the expression obtained by Alkhalifah and Tsvankin (1995) for weak transverse isotropy:

$$V_{nmo}(p) = \frac{V_{nmo}(0)}{\sqrt{1 - y}} \left[ 1 + \left( \epsilon(\phi) - \delta(\phi) \right) f(y) \right],$$

(42)

where

$$f(y) \equiv y \left( 4y^2 - 9y + 6 \right), \quad y \equiv p^2 V_{nmo}^2(0),$$

and $p = \sin \gamma / V(\gamma, \phi)$. The term that involves the difference $\epsilon(\phi) - \delta(\phi)$ is the anisotropic correction factor in the dependence of NMO velocity on the ray parameter. Note that if at some particular azimuth $\epsilon(\phi) = \delta(\phi)$, $P$-wave anisotropy in the vertical plane $\phi = \text{const}$ is elliptical (in the weak-anisotropy limit), and the isotropic DMO equation is entirely valid.

Cohen (1996) provided more accurate analytic expressions for $V_{nmo}(p)$ that include terms quadratic in the anisotropy coefficients. However, to extend the weak-anisotropy approximations in this section beyond the linear terms, we have to account for out-of-plane phenomena and add quadratic terms to our phase-velocity expression (which may be different from those for VTI media).

Nonetheless, within the limits of the linearized weak-anisotropy approximation, the kinematic properties of $P$-waves in any vertical plane of orthorhombic media can be described by the known VTI equations. For instance, the conclusion by Alkhalifah and Tsvankin (1995) that $P$-wave time processing in VTI media depends just on the zero-dip NMO velocity $V_{nmo}$ and the anisotropic parameter $\eta = (\epsilon - \delta)/(1 + 2\delta)$, remains entirely valid for 2-D processing in weakly anisotropic orthorhombic media. Both effective parameters in orthorhombic media become azimuthally dependent, with the value of $V_{nmo}(0)$ given by equation (41) and $\eta \equiv \epsilon(\phi) - \delta(\phi)$. In general, all anisotropic 2-D DMO and migration methods, developed for vertical transverse isotropy, are fully applicable for 2-D processing in weakly anisotropic orthorhombic media.

Discussion and conclusions

The analytic description of the group-velocity vector and normal-moveout velocity given here is valid for anisotropic media with orthorhombic and lower symmetry. Group velocity for arbitrary anisotropic media can be conveniently expressed through phase velocity and its derivatives with respect to the phase angles. This representation simplifies the analysis of the group-velocity function and its application in the derivation of normal-moveout velocity. The group-velocity components in the vertical plane that contains the phase-velocity vector can be evaluated by analogy with VTI media. The transverse component of the group-velocity vector, normal to this vertical phase plane, depends on the derivative of phase velocity with respect to the azimuthal phase angle and appears only outside the symmetry planes of azimuthally anisotropic media.

The exact group-velocity expression was transformed into a much simpler weak-anisotropy approximation for orthorhombic media using the phase-velocity equations given by Tsvankin (1996c). In the limit of weak anisotropy, the group and phase velocity are equal to each other, but the angle between the group- and phase-velocity vector is linear in the anisotropic coefficients. The analytic approximation for the group-velocity function can be used in traveltime tomography to perform singular-value decomposition or to build fast ray-tracing algorithms in weakly anisotropic orthorhombic media.

The group-velocity expression was also applied to obtain the normal-moveout velocity from a horizontal reflector in azimuthally anisotropic media with a horizontal symmetry plane. This NMO equation is valid for all pure modes in orthorhombic or monoclinic media with any strength of velocity anisotropy. The influence of anisotropy on NMO velocity is fully determined by a single function—the azimuthally-dependent second derivative of phase velocity with respect to the polar phase angle (evaluated at vertical incidence). Application of this equation, however, requires solving an equation for the azimuthal phase angle corresponding to small-offset CMP reflections.

For $P$-waves in orthorhombic media, the general NMO equation was reduced to a simple function of the vertical velocity, the azimuth of one of the symmetry planes, and the anisotropic parameters $\delta^{(1)}$ and $\delta^{(2)}$ introduced by Tsvankin (1996c). However, moveout data are not sufficient to resolve the vertical velocity and the $\delta$ coefficients individually because these parameters contribute to short-spread $P$-wave moveout only through the values of the two NMO velocities in the vertical symmetry planes. The azimuthal dependence of normal-moveout velocity represents an ellipse in the horizontal plane, with the semi-axes in the symmetry-plane directions. Therefore, a minimum of three azimuthal measurements is required to recover the orientation of the sym-
metry planes and reconstruct the NMO velocity at all azimuths.

The treatment of moveout from dipping reflectors was limited to weakly anisotropic orthorhombic models and to azimuthal CMP directions normal to the strike of the reflector. Under these assumptions, out-of-plane phenomena can be ignored, and the NMO velocity can be found from the symmetry-plane equation given by Tsvankin (1995). Since the weak-anisotropy approximation for phase velocity (Tsvankin, 1996c) in orthorhombic media has exactly the same form as for vertical transverse isotropy, the dip-dependent NMO velocity was obtained by just adapting the known VTI expression. This kinematic analogy with VTI media also implies that all 2-D seismic processing algorithms developed for vertical transverse isotropy can be applied in weakly anisotropic orthorhombic media.

The results for orthorhombic media were expressed through the dimensionless anisotropic parameters defined by Tsvankin (1996c) similarly to Thomsen's (1986) coefficients for VTI media. This notation makes it possible to reduce the number of parameters responsible for P-wave kinematic signatures and simplify expressions for both group and normal-moveout velocity.

The analytic treatment developed in this paper provides new insights into the dependence of seismic signatures on the parameters of azimuthally anisotropic medium. The expressions for group and normal-moveout velocity given here can be used in the inversion and processing of seismic data acquired over fractured formations.

Acknowledgments

I am grateful to members of the “A(nisotropy)-Team” at the Center for Wave Phenomena (CWP), especially Jack Cohen and Vladimir Grechka, for useful discussions. The support for this work was provided by the members of the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP, Colorado School of Mines, and by the United States Department of Energy (project “Velocity Analysis, Parameter Estimation, and Constraints on Lithology for Transversely Isotropic Sediments” within the framework of the Advanced Computational Technology Initiative).

REFERENCES

Cohen, J.K., 1996, Analytic study of the effective parameters for determination of the NMO velocity function in transversely isotropic media: CWP report, this volume.
Tsvankin, I., 1996a, P-wave signatures and notation for transversely isotropic media: An overview: Geophysics, 61, 467-483.
Tsvankin, I., 1996c, Effective parameters and P-wave velocity for orthorhombic anisotropy: CWP report, this volume.
APPENDIX A: Group velocity in azimuthally anisotropic media

Here, the group-velocity vector in arbitrary anisotropic media is expressed as a function of phase velocity $V(\theta, \phi)$ represented through the spherical polar $(\theta)$ and azimuthal $(\phi)$ phase angles. For purposes of this derivation, it is convenient to use an auxiliary Cartesian coordinate system $[x, y, z]$ rotated by the angle $\phi$ around the $x_3$ axis of the original system $[x_1, x_2, x_3]$ (Figure 1). We start with the general expression for group velocity introduced in the main text (e.g., Berryman, 1979):

$$
\vec{V}_c = \frac{\partial (k V)}{\partial k_x} \hat{x} + \frac{\partial (k V)}{\partial k_y} \hat{y} + \frac{\partial (k V)}{\partial k_z} \hat{z},
$$

(A1)

where $V$ is the phase velocity and $\vec{k}$ is the wave vector. First, we obtain the two components of the group-velocity vector confined to the vertical plane $y = 0$ ($V_{c_x}$ and $V_{c_z}$); in the following, we will call them the "in-plane" components. From equation (A1) we have

$$
V_{c_x} = \frac{\partial (k V)}{\partial k_x} = V \frac{\partial k}{\partial k_x} + k \frac{\partial V}{\partial k_x}.
$$

(A2)

To evaluate $V_{c_z}$, both $V$ and $k_z$ can be represented as functions of the phase angle $\theta$ with the vertical ($z$) axis. Keeping the components $k_y$ and $k_z$ constant and substituting $k = \sqrt{k_x^2 + k_y^2 + k_z^2}$, we get

$$
\frac{\partial k}{\partial k_z} \bigg|_{k_y, k_z = \text{const}} = \frac{k_x}{k} = \sin \theta.
$$

(A3)

The second term in the right-hand side of equation (A2) takes the form

$$
k \frac{\partial V}{\partial k_x} \bigg|_{k_y, k_z = \text{const}} = k \left( \frac{\partial V}{\partial \theta} \bigg|_{\phi = \text{const}} \right) / \left( \frac{\partial k_x}{\partial \theta} \bigg|_{k_z = \text{const}} \right)
$$

Since $k_z = k_x \tan \theta$,

$$
\frac{\partial k_x}{\partial \theta} \bigg|_{k_z = \text{const}} = \frac{k_z}{\cos^2 \theta}.
$$

(A4)

Taking into account that $k_z/k = \cos \theta$, we obtain

$$
k \frac{\partial V}{\partial k_x} = \frac{\partial V}{\partial \theta} \bigg|_{\phi = \text{const}} \cos \theta.
$$

(A5)

Substituting equations (A3) and (A5) into equation (A2) yields

$$
V_{c_z} = V \sin \theta + \frac{\partial V}{\partial \theta} \bigg|_{\phi = \text{const}} \cos \theta.
$$

(A6)

Similarly, the second in-plane component of the group-velocity vector is given by

$$
V_{c_y} = V \cos \theta - \frac{\partial V}{\partial \phi} \bigg|_{\theta = \text{const}} \sin \theta.
$$

(A7)

Equations (A6) and (A7) for $V_{c_x}$ and $V_{c_y}$ are identical to the expressions for the vertical and horizontal components of the group-velocity vector in transversely isotropic media in any symmetry plane in anisotropic media (Tsvankin, 1995).

For the transverse component of the group-velocity vector we have

$$
V_{c_y} = \frac{\partial (k V)}{\partial k_y} \bigg|_{k_z, k_x = \text{const}} = V \frac{k_y}{k} + k \frac{\partial V}{\partial k_y}.
$$

(A8)

The first term in the right-hand side of equation (A8) vanishes because the transverse component of the wave vector $(k_y)$ in the $xz$ plane is zero. Both the phase velocity $V$ and $k_y$ in the second term are convenient to express through the azimuthal phase angle $\phi$ defined with respect to the plane $y = 0$:

$$
\frac{\partial V}{\partial k_y} = \left( \frac{\partial V}{\partial \phi} \right) / \left( \frac{\partial k_y}{\partial \phi} \right).
$$

(A9)

Expressing $k_y$ through the second horizontal component $k_z$, we have

$$
k_y = k_z \tan \phi.
$$
Therefore,
\[ \frac{\partial k_y}{\partial \phi} \bigg|_{k_x, k_z = \text{const}} = \frac{k_x}{\cos^2 \phi}. \]

Since the derivative should be taken at \( \phi = 0 \),
\[ \frac{\partial k_y}{\partial \phi} \bigg|_{k_x, k_z = \text{const}} = k_x. \]  \hfill (A10)

When computing the derivative of phase velocity with respect to \( \phi \), we have to take into account that both the polar angle \( \theta \) and the azimuthal angle \( \phi \) change with \( k_y \) and \( \phi \). Therefore, we need to evaluate
\[ \frac{\partial V}{\partial \phi} \bigg|_{k_x, k_z = \text{const}} = \frac{\partial V}{\partial \theta} \bigg|_{\theta = \text{const}} \frac{\partial \theta}{\partial \phi} \bigg|_{k_x, k_z = \text{const}} + \frac{\partial V}{\partial \phi} \bigg|_{\theta = \text{const}} \frac{\partial \theta}{\partial \phi} \bigg|_{k_x, k_z = \text{const}}. \] \hfill (A11)

From simple geometry,
\[ \tan \theta = \frac{k_x}{k_z \cos \phi}, \]
and
\[ \frac{\partial \theta}{\partial \phi} \bigg|_{k_x, k_z = \text{const}} = \frac{k_z \cos^2 \theta}{k_z \cos^2 \phi} \sin \phi. \]

Substituting \( \phi = 0 \), we find that
\[ \frac{\partial \theta}{\partial \phi} \bigg|_{k_x, k_z = \text{const}} = 0. \]

If the \( y \)-axis points in the positive \( \phi \)-direction (counterclockwise from the \( x_1 \)-axis),
\[ \frac{\partial \phi}{\partial \phi} = 1, \]
and equation (A11) reduces to
\[ \frac{\partial V}{\partial \phi} \bigg|_{k_x, k_z = \text{const}} = \frac{\partial V}{\partial \phi} \bigg|_{\theta = \text{const}}. \] \hfill (A12)

Using equations (A10) and (A12), we can rewrite equation (A9) as
\[ \frac{\partial V}{\partial k_y} = \frac{1}{k_x} \frac{\partial V}{\partial \phi} \bigg|_{\theta = \text{const}}. \] \hfill (A13)

Finally, for the transverse component of the group vector [equation (A8)] we have
\[ V_{Gy} = \frac{k}{k_x} \frac{\partial V}{\partial \phi} \bigg|_{\theta = \text{const}} = \frac{1}{\sin \theta} \frac{\partial V}{\partial \phi} \bigg|_{\theta = \text{const}}. \] \hfill (A14)

The \( y \)-axis in equation (A14) points in the direction of increasing \( \phi \), i.e., counterclockwise from the \( x_1 \)-axis of the original coordinate system.

**APPENDIX B: NMO velocity in azimuthally anisotropic media**

Here, I give a derivation of the normal-moveout velocity from horizontal reflectors in azimuthally anisotropic layer using the results of Appendix A. Moveout is studied on a common-midpoint (CMP) line that makes the azimuthal angle \( \phi \) with the axis \( x_1 \) (Figure B1). The reflector is assumed to represent a symmetry plane of the medium, so the group-velocity vectors (or rays) of incident and reflected waves are confined to the vertical incidence (sagittal) plane that contains the common-midpoint (CMP) line.

The basic equations for NMO velocity are introduced and discussed in the main text. We start the derivation by rewriting the NMO equation (6) as a function of the ray parameter (horizontal slowness) \( p \). Indeed, the derivative of
Anisotropic group and NMO velocity

Figure B1. The geometry of the group- and phase-velocity vectors for the reflection from the bottom of a horizontal azimuthally anisotropic layer. SO and OR are the incident and reflected group-velocity vectors (rays) confined to the vertical incidence plane. The orientation of the phase-velocity vector OD (associated with the ray OR) is described by the polar ($\theta$) and azimuthal ($\phi$) phase angles.

the one-way traveltime $t$ with respect to half-offset $h$ in equation (6) is simply equal to the projection of the slowness vector on the CMP line ($p_n$). Also note that since the reflector represents a symmetry plane, the phase- and group-velocity vectors are identical for the zero-offset reflection (this may not be the case for horizontal reflectors in general anisotropic media). Thus, the half-offset $h$ can be represented as $h = (V_0 t_0 / 2) \tan \beta$, where $V_0$ is the vertical phase and group velocity and $\beta$ is the group angle with the vertical axis (Figure B1). Then equation (6) becomes

$$V_{nmo}^2 = V_0 \left. \frac{d \tan \beta}{dp_n} \right|_{\beta=0}.$$  \hspace{1cm} (B1)

The next step in the evaluation of the NMO velocity involves relating the group angle $\beta$ to the phase-velocity vector. For the simplest azimuthally anisotropic model - transversely isotropic media with a horizontal symmetry axis - the phase-velocity vector is always confined to the plane formed by the symmetry axis and the group-velocity vector. Also, the group-velocity vector in HTI media is a simple function of the phase velocity and the phase angle with the symmetry axis. Tsvankin (1996b) took advantage of these convenient features of the HTI model in his derivation of the exact NMO velocity using equation (B1). Here we have to deal with more complex azimuthally anisotropic models, in which the orientation of the phase-velocity vector for each reflected ray is determined by the general relation between phase and group velocity.

We will represent the group angle $\beta$ as a function of the polar phase angle $\theta$ of the corresponding phase-velocity vector. As the group-velocity vector of the reflected wave (reflected ray) tilts in the incidence plane, deviating from vertical with increasing offset, the position of the corresponding phase-velocity vector can be described by a certain function $\phi(\theta)$ that depends on the azimuth of the CMP line. Thus, the derivative in equation (B1) can be rewritten as

$$V_{nmo}^2 = V_0 \left[ \left( \frac{d \tan \beta}{d \theta} \right) \right] \left( \frac{dp_n}{d \theta} \right) \bigg|_{\beta=0}.$$  \hspace{1cm} (B2)

To find an explicit expression for the group angle $\beta$, we use equations (A6), (A7), and (A14) obtained in Appendix A:

$$V_{Gx} = V \sin \theta \cos \theta \left. \frac{\partial V}{\partial \theta} \right|_{\omega=\text{const}},$$  \hspace{1cm} (B3)
\[ V_{Cz} = V \cos \theta \sin \theta \frac{\partial V}{\partial \phi} \bigg|_{s=const}, \]  
\[ V_{Cy} = \frac{1}{\sin \theta} \frac{\partial V}{\partial \phi} \bigg|_{s=const}. \]

Here the group-velocity vector \( \vec{V}_g \) is represented in an auxiliary Cartesian coordinate system \( [x, y, z] \) with the horizontal axes rotated by the angle \( \phi \) with respect to the original coordinate system. (The \( y \)-axis in equation \((A14)\) points in the direction of increasing \( \phi \), i.e., counterclockwise from the \( z \)-axis of the original coordinate system.) Then

\[
\tan \beta = \frac{\sqrt{V_{Gz}^2 + V_{Gy}^2}}{V_{Gz}}. 
\]

Differentiating with respect to \( \theta \) yields

\[
\frac{d \tan \beta}{d\theta} = \frac{V_{Gz} \frac{dV_{Gz}}{ds} + V_{Gy} \frac{dV_{Gy}}{ds}}{V_{Gz} \sqrt{V_{Gz}^2 + V_{Gy}^2}} - \frac{\sqrt{V_{Gz}^2 + V_{Gy}^2}}{V_{Gz}} \frac{dV_{Gz}}{ds}. 
\]  
\[ (B6) \]

The second term in the right-hand side of equation \((B6)\) vanishes at vertical incidence \((\theta = 0)\) since both horizontal group-velocity components are zero, and

\[
\frac{d \tan \beta}{d\theta} \bigg|_{s=0} = \frac{V_{Gz} \frac{dV_{Gz}}{ds} + V_{Gy} \frac{dV_{Gy}}{ds}}{V_{Gz} \sqrt{V_{Gz}^2 + V_{Gy}^2}} \bigg|_{s=0}. 
\]  
\[ (B7) \]

Next, we represent the components of the group-velocity vector at small polar angles \( \theta \) using the expansion of phase velocity in powers of \( \sin^2 \theta \):

\[ V(\theta, \phi) = V_0 \left[ 1 + a(\phi) \sin^2 \theta \right], \]  
\[ \text{where} \quad a(\phi) = \frac{1}{2V_0} \frac{\partial^2 V}{\partial (\sin \theta)^2} \bigg|_{s=0} = \frac{1}{2V_0} \frac{\partial^2 V}{\partial \phi^2} \bigg|_{s=0}. \]  
\[ (B8) \]

It will be clear from the derivation below that higher-order terms omitted in equation \((B8)\) have no influence on NMO velocity from horizontal reflectors. The derivatives of the phase velocity needed to evaluate the group-velocity components are given by

\[
\frac{\partial V}{\partial \theta} \bigg|_{s=const} = V_0 a(\phi) \sin 2\theta, 
\]  
\[ (B9) \]

\[
\frac{\partial V}{\partial \phi} \bigg|_{s=const} = V_0 a'(\phi) \sin \theta, 
\]  
\[ (B10) \]

\[
a'(\phi) = \frac{\partial a(\phi)}{\partial \phi}. 
\]

Substituting equations \((B8), (B9), \) and \((B10)\) into the expressions for the group velocity \((\text{equations} \ (B3) - (B5))\), we obtain

\[ V_{Cz} = V_0 \cos \theta \left[ 1 + a(\phi) \sin^2 \theta \right] - V_0 a(\phi) \sin \theta \sin 2\theta, \]  
\[ \text{(B11)} \]

\[ V_{Cy} = V_0 \sin \theta \left[ 1 + a(\phi) \sin^2 \theta \right] - V_0 a(\phi) \cos \theta \sin 2\theta, \]  
\[ \text{(B12)} \]

\[ V_{Gy} = V_0 a'(\phi) \sin \theta. \]  
\[ \text{(B13)} \]

Note that all terms involving the azimuthal angle \( \phi \) in equations \((B11) - (B13)\) can be represented as

\[ F(\theta, \phi) = g(\theta) f(\phi), \]

with \( g(\theta) = 0 \) at \( \theta = 0 \). Hence, although \( \phi \) is not defined at vertical incidence \((\theta = 0)\), all terms that contain \( \phi \) vanish at \( \theta = 0 \) anyway. When calculating the derivatives of the horizontal components of the group-velocity vector needed in equation \((B6)\), we have to take into account that \( \phi \) does change with \( \theta \) to keep the corresponding group-velocity vector in the incidence plane. However, the derivatives of the group-velocity components with respect to \( \theta \) at vertical incidence do not contain the value of \( \frac{d\phi}{d\theta} \). Indeed, from the definition of the derivative,
\[
\frac{dF(\theta, \phi)}{d\theta} \bigg|_{\theta=0} = \lim_{\epsilon \to 0} \left\{ f(\phi) \right\}.
\]  
(B14)

Using equations (B12) - (B14), we find
\[
\frac{dV_G}{d\theta} \bigg|_{\theta=0} = V_0 \left[ 1 + 2a(\phi) \right],
\]  
(B15)
\[
\frac{dV_G}{d\phi} \bigg|_{\phi=0} = V_0 a'(\phi),
\]  
(B16)
where both \( a(\phi) \) and \( a'(\phi) \) actually represent the limits taken at \( \theta \to 0 \).

Substituting the expressions for the group-velocity components and their derivatives [equations (B11), (B12), (B13), (B15), (B16)] into equation (B7) and cancelling \( \sin \theta \) in the numerator and denominator yields
\[
\frac{d\tan \beta}{d\theta} \bigg|_{\theta=0} = \frac{V_0^2 \left[ 1 + 2a(\phi) \right]^2 + V_0^2 \left[ a'(\phi) \right]^2}{V_0 \sqrt{V_0^2 \left[ 1 + 2a(\phi) \right]^2 + V_0^2 \left[ a'(\phi) \right]^2}} = \sqrt{\left[ 1 + 2a(\phi) \right]^2 + \left[ a'(\phi) \right]^2}.
\]  
(B17)

Now it is clear that even if we had included the \( \sin^4 \theta \) term (and any higher-order terms) in the expansion of the phase-velocity function [equation (B8)], they would have vanished in equation (B17).

The second derivative needed to obtain the NMO velocity from equation (B2) is much easier to evaluate since it involves just the phase-velocity function itself. The projection of the slowness vector on the CMP line is just
\[
\rho_n = \frac{\sin \theta}{V} \cos(\alpha - \phi),
\]
where \( \alpha \) is the azimuth of the CMP line. Differentiating with respect to \( \theta \), we have
\[
\frac{d\rho_n}{d\theta} \bigg|_{\theta=0} = \frac{\cos(\alpha - \phi)}{V_0}.
\]  
(B18)

Combining equations (B17) and (B18), we obtain the NMO velocity [equation (B2)] as
\[
V_{nmo}^2 = V_0^2 \frac{\sqrt{\left[ 1 + 2a(\phi) \right]^2 + \left[ a'(\phi) \right]^2}}{\cos(\alpha - \phi)}.
\]  
(B19)

The last step in the derivation is to express the azimuthal phase angle \( \phi \) through the model parameters. As shown above, the value of \( \phi \) in equation (B19) actually represents the limit taken for \( \theta \to 0 \). Since the group vector lies in the plane \( \alpha = \text{const} \),
\[
\tan(\alpha - \phi) = \frac{V_G}{V_Gz},
\]
and using equations (B12) and (B13) in the limit \( \theta \to 0 \), we find the following equation for \( \phi \):
\[
\tan(\alpha - \phi) = \frac{a'(\phi)}{1 + 2a(\phi)}.
\]  
(B20)

Using equation (B20), NMO velocity from equation (B19) can be rewritten as
\[
V_{nmo}^2 = V_0^2 \left\{ 1 + 2a(\phi) + \frac{[a'(\phi)]^2}{1 + 2a(\phi)} \right\},
\]  
(B21)
with \( \phi \) defined by equation (B20).

**APPENDIX C: Representation of near-vertical phase-velocity variations**

To calculate normal-moveout velocity from horizontal reflectors, it is necessary to know the behavior of phase velocity near vertical (at small polar angles \( \theta \)). Since the exact expression for phase velocity in azimuthally anisotropic media is a complicated function of the phase angles \( \theta \) and \( \phi \), it is convenient to expand phase velocity within each vertical plane \( (\phi = \text{const}) \) in a Taylor (Maclaurin) series at \( \theta = 0 \). The continuity of the derivatives of the phase-velocity function ensures the existence of this series expansion at small polar angles.

If a medium has a horizontal plane of symmetry, phase velocity in any vertical plane is symmetric with respect to the vertical direction. Therefore, the series expansion of phase velocity can have only even terms in the polar angle.
\( V(\theta, \sigma)_{\sigma=\text{const.}} = V_0 + \frac{1}{2} \frac{\partial^2 V}{\partial (\sin \theta)^2} \bigg|_{\theta=0} \sin^2 \theta + \ldots \)  

(C1)

Series expansion (C1) is truncated after the quadratic term because higher-order terms are not needed in the derivation of NMO velocity (Appendix B).

Evidently,

\[
\frac{\partial^2 V}{\partial (\sin \theta)^2} \bigg|_{\theta=0} = \frac{\partial^2 V}{\partial \theta^2} \bigg|_{\theta=0},
\]

(C2)

and to describe velocity variations near vertical, we have to find the second derivative of the phase-velocity function with respect to the polar angle \( \theta \).

The squared phase velocity in any anisotropic medium satisfies a cubic equation

\[ x^2 + ax^2 + bx + c = 0, \]

where \( x = \rho v^2 \) and the coefficients represent the following functions of the components of the Christoffel matrix \( G_{ik} \) (e.g., Tsvankin, 1996c):

\[
a = -(G_{11} + G_{22} + G_{33}),
\]

(C4)

\[
b = G_{11} G_{22} + G_{12} G_{33} + G_{22} G_{13} - G_{13}^2 - G_{23}^2 - G_{33}^2,
\]

(C5)

\[
c = G_{11} G_{22} G_{33} + G_{11} G_{33} G_{12} + G_{12} G_{22} G_{13} - G_{12} G_{33} G_{23} - 2G_{13} G_{23} G_{33}.
\]

(C6)

Differentiating equation (C3) twice with respect to the polar angle \( \theta \) (the azimuthal phase angle \( \phi \) is held constant) and taking into account that \( \theta V/\partial \theta = 0 \) at normal incidence, we obtain

\[
V''\bigg|_{\theta=0} = -\frac{1}{2V_0} \frac{a'' V_0^4 + b'' V_0^2 + c''}{3V_0^2 + 2aV_0^2 + b} \bigg|_{\theta=0},
\]

(C7)

where the \( V'', a'' \) etc. are second partial derivatives with respect to \( \theta \).

Equation (C7) is valid for any anisotropic medium with a horizontal symmetry plane. Next, we obtain the exact value of \( V''\big|_{\theta=0} \) for the \( P \)-wave in orthorhombic media. The components of the Christoffel matrix for orthorhombic anisotropy are given by (e.g., Schoenberg and Helbig, 1996)

\[
G_{11} = c_{11} n_1^2 + c_{66} n_2^2 + c_{55} n_3^2;
\]

(C8)

\[
G_{22} = c_{66} n_1^2 + c_{22} n_2^2 + c_{44} n_3^2;
\]

(C9)

\[
G_{33} = c_{55} n_1^2 + c_{44} n_2^2 + c_{33} n_3^2;
\]

(C10)

\[
G_{12} = (c_{12} + c_{66}) n_1 n_2;
\]

(C11)

\[
G_{13} = (c_{13} + c_{55}) n_1 n_3;
\]

(C12)

\[
G_{23} = (c_{23} + c_{44}) n_2 n_3.
\]

(C13)

with the directional cosines defined through the phase angles as

\[
n_1 = \sin \theta \cos \phi, \quad n_2 = \sin \theta \sin \phi, \quad n_3 = \cos \theta.
\]

Substituting equations (C8)-(C13) into the expressions for \( a, b, \) and \( c \) [equations (C4)-(C6)] and carrying out the differentiation required in equation (C7) yields

\[
V''\big|_{\theta=0} \text{[P-wave]} = 2V_0 \left\{ \frac{(c_{13} + c_{55})^2 - (c_{33} - c_{55})^2}{2c_{33}(c_{23} - c_{55})} \cos^2 \phi \right. \\
\left. - \frac{(c_{23} + c_{44})^2 - (c_{33} - c_{44})^2}{2c_{33}(c_{23} - c_{44})} \sin^2 \phi \right\}.
\]

(C14)

Using equations (16) and (19) of the main text, the combinations of stiffnesses in equation (C14) can be identified
as the anisotropy parameters \( \delta^{(2)} \) and \( \delta^{(1)} \). This leads to the final form of the expression for \( V''\rvert_{\theta=0} \), which will be used in Appendix D to derive the P-wave normal-moveout velocity in orthorhombic media.

\[
V''\rvert_{\theta=0} [P\text{-wave}] = 2V_0 (\delta^{(2)} \cos^2 \phi + \delta^{(1)} \sin^2 \phi).
\]  

(C15)

**APPENDIX D: Explicit NMO expression for orthorhombic media**

To obtain an explicit expression for NMO velocity as a function of the azimuth of the CMP line, it is necessary to solve equation (B20) for the azimuthal phase angle \( \phi \). Evidently, this requires specifying the azimuthally-dependent phase-velocity term \( a(\phi) \). For \( P \)-waves in orthorhombic media, the exact expression for \( a(\phi) \) was given in Appendix C [equation (C15)]:

\[
a(\phi) = \delta^{(2)} \cos^2 \phi + \delta^{(1)} \sin^2 \phi.
\]  

(D1)

Then equation (B21) takes the following form:

\[
V_{nmo}^2 = V_0^2 \left\{ 1 + 2\delta^{(2)} \cos^2 \phi + 2\delta^{(1)} \sin^2 \phi + \frac{\sin^2 2\phi (\delta^{(1)} - \delta^{(2)})^2}{1 + 2\delta^{(2)} \cos^2 \phi + 2\delta^{(1)} \sin^2 \phi} \right\},
\]  

(D2)

with \( \phi \) defined [from equation (B20)] as

\[
\tan(\alpha - \phi) = \frac{\sin 2\phi (\delta^{(1)} - \delta^{(2)})}{1 + 2\delta^{(2)} \cos^2 \phi + 2\delta^{(1)} \sin^2 \phi}.
\]  

(D3)

While equation (D3) is easy to solve for \( \alpha \), obtaining \( \phi \) as a function of \( \alpha \) is far less straightforward. First, we rewrite equation (D3) as

\[
\frac{\tan \alpha - \tan \phi}{1 - \tan \alpha \tan \phi} = R,
\]  

(D4)

\[
R = \frac{\sin 2\phi (\delta^{(1)} - \delta^{(2)})}{1 + 2\delta^{(2)} \cos^2 \phi + 2\delta^{(1)} \sin^2 \phi}.
\]  

From equation (D4) we find

\[
\tan \alpha = \frac{R + \tan \phi}{1 - R \tan \phi},
\]  

and

\[
\cos^2 \alpha = \cos^2 \phi \frac{(1 - R \tan \phi)^2}{1 + R^2}.
\]  

(D5)

Substituting the value of \( R \) into equation (D5) yields

\[
\cos^2 \alpha = \cos^2 \phi \frac{(1 + 2\delta^{(2)})^2}{1 + 4\delta^{(2)} \cos^2 \phi (1 + \delta^{(2)}) + 4\delta^{(1)} \sin^2 \phi (1 + \delta^{(1)})}.
\]

Thus, the above transformations allowed us to reduce equation (D3) to a simple linear equation for \( \cos^2 \phi \) with the solution

\[
\cos^2 \phi = \cos^2 \alpha \frac{(1 + 2\delta^{(1)})^2}{(1 + 2\delta^{(2)})^2 + 4 \cos^2 \alpha (\delta^{(1)}(1 + \delta^{(1)}) - \delta^{(2)}(1 + \delta^{(2)}))}.
\]  

(D6)

Now we can substitute \( \cos^2 \phi \) from equation (D6) into equation (D2) to obtain the NMO velocity as an explicit function of the azimuth of the CMP line. First, we find

\[
1 + 2\delta^{(2)} \cos^2 \phi + 2\delta^{(1)} \sin^2 \phi = \frac{(1 + 2\delta^{(2)})(1 + 2\delta^{(1)}) (1 + 2\delta^{(2)} \sin^2 \phi + 2\delta^{(1)} \cos^2 \alpha)}{(1 + 2\delta^{(2)})^2 + 4 \cos^2 \alpha (\delta^{(1)}(1 + \delta^{(1)}) - \delta^{(2)}(1 + \delta^{(2)}))}.
\]  

(D7)

Then we use equation (D6) to get
\[
\sin^2 2\phi (\delta^{(1)} - \delta^{(2)})^2 = \\
\left\{ \frac{\sin 2\alpha (1 + 2\delta^{(2)}) (1 + 2\delta^{(1)}) (\delta^{(1)} - \delta^{(2)})}{(1 + 2\delta^{(2)})^2 + 4 \cos^2 \alpha (\delta^{(1)} + \delta^{(2)}) (\delta^{(1)} - \delta^{(2)})(1 + \delta^{(2)})} \right\}^2.
\] (D8)

Substitution of equations (D7) and (D8) into equation (D2) leads, after algebraic transformations, to a simple final result:

\[
V_{\text{noo}} = V_{\tilde{\phi} \phi} \frac{(1 + 2\delta^{(1)}) (1 + 2\delta^{(2)})}{1 + 2\delta^{(2)} \sin^2 \alpha + 2\delta^{(1)} \cos^2 \alpha}.
\] (D9)
Nonhyperbolic reflection moveout for horizontal transverse isotropy

AbdulFattah Al-Dajani  
*Center for Wave Phenomena, Colorado School of Mines*

Ilya Tsvankin  
*Department of Geophysics, Colorado School of Mines*

**ABSTRACT**

The transversely isotropic model with a horizontal axis of symmetry (HTI) has been extensively used in studies of shear-wave splitting to describe fractured formations with a single system of parallel vertical penny-shaped cracks. Here, we present an analytic description of long-spread reflection moveout in horizontally-layered HTI media with arbitrary strength of anisotropy.

The hyperbolic moveout equation parameterized by the exact normal-moveout (NMO) velocity provides sufficient accuracy for P-waves on conventional-length spreads close to the reflector depth. However, the influence of anisotropy leads to the deviation of the moveout curve from a hyperbola with increasing spreadlength, even in a single-layer model. To account for nonhyperbolic moveout, we have derived an exact expression for the azimuthally-dependent quartic term of the Taylor series traveltime expansion \( \frac{6P(z^2)}{z^2} \), valid for any pure mode in an HTI layer. The quartic moveout coefficient and the normal-moveout velocity are then substituted into the nonhyperbolic moveout equation, originally designed by Tsvankin and Thomsen for vertical transverse isotropy (VTI media). Numerical examples for media with both moderate and uncommonly strong nonhyperbolic moveout show that this equation accurately describes azimuthally-dependent P-wave reflection traveltimes in an HTI layer, even on spreadlengths twice as large as the reflector depth.

In multilayered HTI media, the NMO velocity and the quartic moveout coefficient reflect the combined influence of layering and azimuthal anisotropy. We show that the conventional Dix equation for NMO velocity remains entirely valid for any azimuth in HTI media if the *group-velocity* vectors (rays) on the CMP gather do not deviate from the vertical incidence plane. Although this condition is not exactly satisfied in realistic HTI media, rms averaging of the interval NMO velocities provides a good approximation for short-spread moveout. Therefore, the NMO velocity in any layer of interest can be obtained by the conventional Dix differentiation and inverted for the anisotropy parameters.

Furthermore, the quartic moveout coefficient for multilayered HTI media can also be calculated with a good accuracy using the known averaging equations for vertical transverse isotropy. This allows us to extend the nonhyperbolic moveout equation to horizontally stratified media composed of any combination of isotropic, VTI, and HTI layers. In addition to providing analytic insight into the behavior of nonhyperbolic moveout, these results can be used in modeling and inversion of reflection traveltimes in azimuthally anisotropic media.

**Key words:** horizontal transverse isotropy (HTI), nonhyperbolic reflection moveout, normal-moveout velocity.
Introduction

Recent experimental studies (Lynn et al., 1995) have shown that $P$-wave reflection moveout and amplitude-variation-with-offset (AVO) response may be strongly influenced by the presence of azimuthal anisotropy. However, the current understanding of seismic signatures in azimuthally anisotropic media is hardly sufficient for the inversion and processing of seismic data, even if the medium is horizontally homogeneous. This work is devoted to an analytic description of long-spread reflection moveout in the transversely isotropic model with a horizontal axis of symmetry (HTI media) — the simplest type of azimuthally anisotropic media associated with a system of parallel vertical cracks embedded in an isotropic matrix (Cramin, 1983; Thomsen, 1988). Weak-anisotropy approximations for reflection moveout in HTI media were discussed by Thomsen (1988), Sena (1991) and Li and Cramin (1993); the latter paper also treats reflection traveltimes in an orthorhombic layer.

Recently, Tsvankin (1996b) presented an exact equation for azimuthally-dependent normal-moveout velocity valid for pure modes in a single HTI layer. He also showed that all kinematic signatures including normal moveout (as well as plane-wave polarizations) in the symmetry plane of HTI media that contains the symmetry axis ("the symmetry-axis" plane) are given by the same equations as for transversely isotropic media with a vertical symmetry axis. The analogy between HTI and VTI media allowed Tsvankin (1996b) and Rüger (1996) to introduce Thomsen's (1986) parameters for HTI media using exactly the same expressions as for vertical transverse isotropy. This notation proved to be much more convenient in describing reflection signatures than the generic Thomsen coefficients defined with respect to the symmetry axis. For instance, the $P$-wave NMO velocity in HTI media depends on the vertical velocity, orientation of the symmetry axis, and a single anisotropic coefficient — the parameter $\delta(V)$ expressed through the stiffnesses in the same way as Thomsen's coefficient $\delta$ for VTI media (Tsvankin, 1996b).

Despite all these developments, some important issues pertaining to moveout analysis for horizontal transverse isotropy remained unresolved. Among them is the behavior of long-spread (nonhyperbolic) moveout in the presence of azimuthal anisotropy and the feasibility of obtaining interval NMO velocities for vertically inhomogeneous HTI models with strong velocity anisotropy. Both problems have to be examined outside the vertical symmetry planes of HTI media, since reflection moveout within the symmetry planes is identical to that in VTI media.

It is well known that reflection moveout in anisotropic media is generally nonhyperbolic, unless anisotropy is elliptical. Hake et al. (1984) derived the quartic Taylor series term $A_4$ of $t^2 - \alpha^2 t^2$ reflection-moveout curves for pure modes in TI media with a vertical axis of symmetry. Tsvankin and Thomsen (1994) obtained the coefficient $A_4$ for converted $P - SV$ waves and represented the quartic terms of the pure modes in a more compact form using Thomsen's (1986) notation. They also developed a nonhyperbolic moveout equation for layered VTI media based on the exact quadratic (NMO velocity) and quartic moveout coefficients that converges at infinitely large horizontal offsets as well. In TI media with a vertical symmetry axis, this equation remains close to the exact $P$-wave moveout on uncommonly long spreads that may be three times as large as the reflector depth. The moveout expression of Tsvankin and Thomsen (1994) will serve as a basis for our study of nonhyperbolic reflection moveout in HTI media.

The influence of nonhyperbolic moveout can hamper the estimation of normal-moveout velocity using conventional hyperbolic semblance analysis (Alkhalifah, 1996). Tsvankin (1996b) gives a numerical example showing that in a single HTI layer the hyperbolic moveout equation parameterized by the exact NMO velocity provides sufficient accuracy on the spreadlength equal to the reflector depth. In layered media, however, the magnitude of nonhyperbolic moveout may increase due to vertical velocity variations. Also, the Dix equation, conventionally used to obtain interval NMO velocities, is no longer strictly valid outside the symmetry planes of azimuthally anisotropic media.

Here, we derive a concise expression for the quartic moveout coefficient valid for all pure modes in an HTI layer with any strength of the anisotropy. This azimuthally dependent quartic moveout term is used to extend the nonhyperbolic equation of Tsvankin and Thomsen (1994) to horizontal transverse isotropy. The quadratic and quartic moveout coefficients in multilayered HTI media are obtained by the same averaging equations as for vertical transverse isotropy. To substantiate applicability of the Dix (1955) equation for normal-moveout velocity in HTI media, we show that the Dix rms averaging procedure is valid for any azimuthally anisotropic medium if the group-velocity vector does not deviate from the incidence plane. Although this assumption is seldom satisfied exactly in multilayered HTI media, the group-velocity vector usually lies much closer to the incidence plane than the phase vector, and the Dix equation proves to be a valid approximation for conventional-spread moveout. Extensive numerical testing demonstrates excellent accuracy of our nonhyperbolic moveout equation,
even for media with significant depth-varying azimuthal anisotropy and pronounced nonhyperbolic moveout.

Description of the HTI Model and Notation

In this section, we will briefly describe the main features of the HTI model and a convenient Thomsen-style notation for horizontal transverse isotropy introduced by Tsvankin (1996b) and Rüger (1996). Proper understanding of wave propagation in the two mutually orthogonal vertical symmetry planes (Figure 1) is extremely important in the analysis of seismic signatures in HTI media. In the plane normal to the symmetry axis (the isotropy plane), body-wave velocities are independent of direction, and the influence of anisotropy manifests itself only through the different velocities of the two \( S \)-waves. (The split shear waves in HTI media will be denoted as "\( S_{\perp} \)" and "\( S_{\parallel} \), with the \( S_{\perp} \)-wave polarization vector confined to the plane formed by the symmetry axis and the slowness vector of the \( S_{\parallel} \)-wave polarized within the isotropy plane.)

In the second vertical symmetry plane that contains the axis of symmetry (the "symmetry-axis plane"), the velocities do change with propagation angle, but the Christoffel equation has exactly the same form as in transversely isotropic media with a vertical symmetry axis. This means that phase velocity and polarization vector represent the same functions of the stiffness coefficients and phase angle with vertical as in VTI media. Since phase velocity determines group (ray) velocity and group angle, all kinematic signatures in the symmetry-axis plane, including normal-moveout velocity and longspread (nonhyperbolic) reflection moveout, are given by the known VTI equations.

Taking advantage of this limited equivalence, Tsvankin (1996b) and Rüger (1996) introduced the Thomsen parameters of the "equivalent" VTI model through the same equations as those used by Thomsen (1986) for actual VTI media. For the HTI model with the symmetry axis in the \( x_1 \) direction, these parameters are defined through the stiffness coefficients \( c_{ij} \) and density \( \rho \) as

\[
\begin{align*}
V_{P_{\text{vert}}} &= \frac{\sqrt{c_{33}}}{\rho} , \\
V_{S_{\perp \text{vert}}} &= \frac{\sqrt{c_{55}}}{\rho} , \\
\epsilon^{(V)} &= \frac{c_{11} - c_{33}}{2c_{33}} , \\
\delta^{(V)} &= \frac{(c_{13} + c_{55})^2 - (c_{33} - c_{55})^2}{2c_{33}(c_{33} - c_{55})} , \\
\gamma^{(V)} &= \frac{c_{66} - c_{44}}{2c_{44}} ,
\end{align*}
\]

where \( V_{P_{\text{vert}}} \) and \( V_{S_{\perp \text{vert}}} \) are the vertical velocities of the \( P \)- and \( S_{\perp} \)-wave, respectively (note that \( c_{55} = c_{66} \)). The vertical velocity of the (fast) shear wave \( S_{\parallel} \) is determined as

\[
V_{S_{\parallel \text{vert}}} = \frac{\sqrt{\epsilon^{(V)}}}{\rho} = \frac{V_{S_{\perp \text{vert}}}}{\sqrt{1 + \frac{2\gamma^{(V)}}{\epsilon^{(V)}}}}.
\]

This notation makes it possible to obtain the kinematic signatures and polarizations in the symmetry-axis plane of HTI media just by adapting the corresponding equations for vertical transverse isotropy expressed through Thomsen parameters. The convenient features of Thomsen notation in the analytic description of seismic wavefields in HTI media were summarized by Tsvankin (1996a). Furthermore, the Thomsen coefficients of the equivalent VTI model control the moveout outside the symmetry planes of HTI media, where the analogy with VTI media is no longer valid.

The exact expression for the phase velocity in HTI media in terms of the parameters \( \epsilon^{(V)} \), \( \delta^{(V)} \), and \( \gamma^{(V)} \) was presented in Tsvankin (1996b). For \( P \)- and \( S_{\perp} \)-waves, the phase velocity is given by

\[
\begin{align*}
\frac{V^2(\theta)}{V_{P_{\text{vert}}}^2} &= 1 + \epsilon^{(V)} \cos^2 \theta - \frac{\delta^{(V)}}{2} + \frac{\gamma^{(V)}}{2} \sqrt{1 + \frac{2\gamma^{(V)} \cos^2 \theta}{f^{(V)}}} - 2 \frac{\epsilon^{(V)} - \delta^{(V)} \sin^2 \theta}{f^{(V)}} ,
\end{align*}
\]

where the plus sign corresponds to the \( P \)-wave, and the minus — to the \( S_{\perp} \)-wave; \( f^{(V)} = 1 - (V_{S_{\perp \text{vert}}} / V_{P_{\text{vert}}})^2 \); \( \theta \) is the phase angle with the horizontal symmetry axis.

Alternatively, HTI media can be characterized by the "generic" Thomsen parameters defined with respect...
to the symmetry axis. However, since the symmetry axis is horizontal, these parameters (especially the coefficient $\delta$) are not well-suited to describe reflection seismic signatures, which are largely dependent on near-vertical velocity variations. For reference, the relations between the two sets of Thomsen parameters, described by Tsvankin (1996b) and Rüger (1996), are reproduced in Appendix A.

Note that the values of the parameters $\epsilon^V$, $\delta^V$, and $\gamma^V$ are quite different from those typical for actual VTI media. While the coefficients $\epsilon$ and $\gamma$ for vertical transverse isotropy are non-negative, in HTI media $\epsilon^V \leq 0$ and $\gamma^V \leq 0$. Also, for typical ratios of the vertical velocities ($V_{V,\perp}/V_{V,\parallel} \leq .707$), the parameter $\delta^V \leq 0$, which is possible but not typical for actual VTI formations. However, similar to VTI media, the difference $\epsilon^V - \delta^V$ that characterizes the "anisotropy" of the medium is usually positive in HTI media.

Analytic Approximations of Reflection Moveout

In the practice of seismic data processing, reflection moveout on common-midpoint (CMP) gathers is conventionally approximated by a hyperbolic equation:

$$t^2 = t_0^2 + \frac{z^2}{V_{nmo}^2},$$

where $t$ is the reflection traveltime at the source-receiver offset $z$, $t_0$ is the two-way zero-offset traveltime, and $V_{nmo}$ is the normal-moveout (stacking) velocity defined in the zero-spread limit.

Equation (3) is strictly valid only for a homogeneous isotropic (or elliptical anisotropic) layer. The presence of layering and/or anisotropy leads to increasing deviations of the moveout curve from the short-spread hyperbola (3). However, for vertical transverse isotropy the hyperbolic moveout equation for $P$-waves usually provides sufficient accuracy on conventional-length spreads close to the reflector depth (Tsvankin and Thomsen, 1994).

Nonhyperbolic moveout on longer spreads can be described by a three-term Taylor series expansion (Taran and Koehler, 1969):

$$t^2 = t_0^2 + A_2 z^2 + A_4 z^4,$$

where $A_2 = 1/V_{nmo}^2$, and $A_4$ is the quartic moveout coefficient. The parameter $A_4$ for pure modes in horizontally layered VTI media was given by Hake et al. (1984) and represented in a more compact form by Tsvankin and Thomsen (1994). Due to the influence of the $z^4$ term, the quartic equation (4) becomes divergent with increasing offset and can be replaced by a more accurate non-

Figure 2. Common-midpoint reflections on a line that makes the angle $\alpha$ with the symmetry plane of an HTI layer. Since the model has a horizontal symmetry plane, the incident and reflected rays of pure modes lie in the vertical incidence (sagittal) plane (after Tsvankin, 1996b).

The hyperbolic moveout equation developed by Tsvankin and Thomsen (1994):

$$t^2 = t_0^2 + A_2 z^2 + \frac{A_4 z^4}{1 + A_2 z^2},$$

where $A = A_4/(1/V_{hor}^2 - 1/V_{nmo}^2)$; $V_{hor}$ is the horizontal velocity. The denominator of the nonhyperbolic term ensures the convergence of this approximation at infinitely large horizontal offsets. As a result, equation (5) provides an accurate description of $P$-wave traveltimes on long CMP spreads (2-3 times as large as the reflector depth), even for models with pronounced nonhyperbolic moveout.

Although equation (5) was originally designed for vertical transverse isotropy, it could be used in arbitrary anisotropic media if the appropriate coefficients $A_2$, $A_4$, and $A$ were found. Our goal is to extend this nonhyperbolic moveout approximation to single- and multilayered HTI media. As discussed in the previous section, for a CMP line parallel to the symmetry axis no generalization is necessary, since the moveout in the symmetry-axis plane can be obtained directly from the original VTI equation (5) by substituting the Thomsen coefficients of the equivalent VTI model. Clearly, in the isotropy plane long-spread reflection moveout of any given mode is not influenced by the anisotropy at all. The analogy with VTI media also holds for throughgoing vertical symmetry planes of multilayered models containing VTI and HTI layers.

However, for CMP lines outside the vertical symmetry planes of HTI media it is necessary to obtain the azimuthally-dependent parameters of equation (5). Be-
Moveout in a Single HTI Layer

Normal-moveout and horizontal velocity

The quadratic moveout coefficient $A_2$ (or the NMO velocity) in a single HTI layer was found by Tsvankin (1996b). He showed that the exact normal-moveout velocity of any pure mode on a CMP line that makes the angle $\alpha$ with the symmetry axis (Figure 2) is given by

$$V_{nmo}^2 = V_{vert}^2 \frac{1 + \frac{\partial^2 V}{\partial \theta^2}}{1 + \sin^2 \alpha \frac{\partial^2 V}{\partial \theta^2}} \bigg|_{\theta=90^\circ},$$  

(6)

where $V$ is the phase velocity as a function of the phase angle $\theta$ with the symmetry axis; the phase velocity and its second derivative are evaluated at the vertical phase (and group) direction. For a given orientation of the symmetry axis, the influence of anisotropy in equation (6) is concentrated in a single velocity term $\frac{\partial^2 V}{\partial \theta^2}$ at 90° that represents the anisotropic correction to the NMO velocity in the symmetry-axis plane:

$$\frac{1}{V} \frac{d^2 V}{d \theta^2} \bigg|_{\theta=90^\circ} = \frac{V_{nmo}^2(\alpha = 0)}{V_{vert}^2} - 1.$$

This term can be expressed through the anisotropic parameters in the following way (Tsvankin, 1996b):

$$\frac{1}{V} \frac{d^2 V}{d \theta^2} \bigg|_{\theta=90^\circ} = \left[ P-\text{wave} \right] = 2\sigma^{(V)},$$

$$\frac{1}{V} \frac{d^2 V}{d \theta^2} \bigg|_{\theta=90^\circ} = \left[ S^-\text{wave} \right] = 2\epsilon^{(V)},$$

$$\frac{1}{V} \frac{d^2 V}{d \theta^2} \bigg|_{\theta=90^\circ} = \left[ S^+\text{wave} \right] = 2\gamma^{(V)},$$

where

$$\sigma^{(V)} = \left( \frac{V_{vert}}{V_{nmo}} \right)^2 (\epsilon^{(V)} - \delta^{(V)}),$$

and the anisotropic coefficients $\delta^{(V)}$, $\epsilon^{(V)}$ and $\gamma^{(V)}$ were introduced in equation (1). Therefore, the azimuthal dependence of NMO velocity for horizontal transverse isotropy is governed by the Thomsen parameters of the equivalent VTI medium rather than the generic coefficients defined with respect to the symmetry axis.

Normal-moveout velocity of the $P$-wave depends just on the parameter $\delta^{(V)}$:

$$V_{nmo}^2 = V_{vert}^2 \frac{1 + 2\delta^{(V)}}{1 + 2\delta^{(V)} \sin^2 \alpha}.$$

(7)

To obtain the term A in the nonhyperbolic moveout equation (5), we also have to find the azimuthally dependent horizontal group velocity ($V_{hor}$) that controls reflection moveout at large offsets approaching infinity. Since the influence of small errors in $V_{hor}$ for spreadlengths feasible in reflection surveys is not significant, we will ignore the difference between phase and group velocity and calculate $V_{hor}$ as the phase velocity evaluated at the azimuth of the CMP line. For the $P$- and $S^+$-waves we find the horizontal velocity by substituting $\theta = \alpha$ into equation (2).

Quartic moveout coefficient

Application of the nonhyperbolic moveout equation (5) also requires knowledge of the quartic moveout coefficient $A_4$. An exact expression for $A_4$ in a single HTI layer is derived in Appendix B:

$$A_4 = \cos^4 \alpha \left[ -\frac{\frac{\partial^2 V}{\partial \theta^2} + 2\left( \frac{\partial^2 V}{\partial \theta^2} \frac{\partial^2 V}{\partial \theta^2} \right)^2 + \frac{\partial^2 V}{\partial \theta^2}}{12\left( \frac{\partial^2 V}{\partial \theta^2} \right)^3} \right] \bigg|_{\theta=90^\circ}.$$  

(8)

As before, $\alpha$ is the angle between the CMP line and the symmetry axis; the phase velocity $V$ and its derivatives are evaluated at the vertical phase (and group) direction ($\theta = 90^\circ$).

Equation (8) is valid for HTI models with any strength of anisotropy and can be used for any pure-mode reflection ($P$, $S^-$, $S^+$). The most interesting feature of equation (8) is the simplicity of the azimuthal dependence of the quartic moveout coefficient. The maximum absolute value of $A_4$ corresponds to the symmetry-axis plane ($\alpha = 0^\circ$), while in the isotropy plane ($\alpha = 90^\circ$) $A_4 = 0$ and moveout is purely hyperbolic. The form of the angular dependence ($\cos^4 \alpha$) implies that the quartic coefficient rapidly decreases with azimuth away from the symmetry-axis plane.

The term in square brackets represents the quartic coefficient in the symmetry-axis plane ($\alpha = 0^\circ$), where all moveout parameters are given by the same equations as for vertical transverse isotropy. Although an equivalent expression was obtained by Hake et al. (1984), here $A_4$ is represented for the first time as a simple function of phase velocity and its derivatives. In addition to the axis orientation, the influence of anisotropy on the quartic moveout coefficient is absorbed by two velocity terms: $\frac{\partial^2 V}{\partial \theta^2}$ at 90° and $\frac{\partial^2 V}{\partial \theta^2}$ at 90°.

To find the quartic coefficient for the $P$-wave, we evaluate the derivatives of phase velocity using equation (2):

$$\frac{1}{V} \frac{d^2 V}{d \theta^2} \bigg|_{\theta=90^\circ} = 2\delta^{(V)},$$  

(9)
\[
\frac{1}{V} \frac{d^2V}{d\theta^2}_{\theta=90^\circ} = 24(\epsilon^{(V)} - \delta^{(V)})(1 + \frac{2\delta^{(V)}}{f^{(V)}}) - 2\delta^{(V)}(4 + 6\delta^{(V)}). \tag{10}
\]

Substituting equations (9) and (10) into equation (8), we obtain an explicit expression for the \(P\)-wave quartic moveout coefficient:

\[
A_4^{(P)} = \cos^4 \alpha \left[ \frac{-2(\epsilon^{(V)} - \delta^{(V)})(1 + 2\delta^{(V)}/f^{(V)})}{t^2_t V^2_{p,vert}(1 + 2\delta^{(V)})^2} \right]. \tag{11}
\]

As expected from the analogy between the symmetry-axis plane of HTI media and vertical transverse isotropy, the expression in brackets [\(A_4^{(\alpha = 0^\circ)}\)] in equation (11) is identical to the \(P\)-wave quartic coefficient for VTI media given in Tsvankin and Thomsen (1994). Alkhalifah and Tsvankin (1995) showed that the contribution of the shear-wave vertical velocity to the \(P\)-wave quartic coefficient in VTI media can be ignored, and \(A_4^{(P)}\) can be represented as a function of just two effective parameters: normal-moveout \(V_{nmo}\) and the anisotropic parameter \(\eta = (\epsilon - \delta)/(1 + 2\delta)\). Therefore, this conclusion remains entirely valid for the \(P\)-wave quartic moveout coefficient in HTI media, with \(V_{nmo}\) calculated in the symmetry-axis plane and \(\eta = \eta^{(V)} = (\epsilon^{(V)} - \delta^{(V)})/(1 + 2\delta^{(V)})\).

Similarly, for the \(S^\perp\)-wave equation (2) yields:

\[
\frac{1}{V} \frac{d^2V}{d\theta^2}_{\theta=90^\circ} = 2\sigma^{(V)}, \tag{12}
\]

and

\[
\frac{1}{V} \frac{d^2V}{d\theta^2}_{\theta=90^\circ} = -24\sigma^{(V)}(1 + \frac{2\delta^{(V)}}{f^{(V)}}) - 2\sigma^{(V)}(4 + 6\sigma^{(V)}), \tag{13}
\]

takes the following form:

\[
A_4^{(S^\perp)} = \cos^4 \alpha \left[ \frac{2\sigma^{(V)}(1 + 2\sigma^{(V)}/f^{(V)})}{t^2_t V^2_{S^\perp,vert}(1 + 2\sigma^{(V)})^2} \right]. \tag{14}
\]

Again, the expression in brackets in equation (14) is identical to the \(S^\perp\)-wave quartic coefficient in VTI media given in Tsvankin and Thomsen (1994), which represents a useful check of our results.

Similar to the quadratic moveout coefficient (NMO velocity), the azimuthal dependence of the quartic moveout coefficient for horizontal transverse isotropy is governed by the Thomsen parameters of the equivalent VTI medium rather than the generic coefficients.

The anisotropy for the shear wave \(S^\perp\) is elliptical, and the quartic coefficient in the symmetry-axis plane is equal to zero. According to equation (8), this means that the quartic moveout term for the \(S^\perp\)-wave vanishes at all other azimuthal directions as well, and \(S^\perp\)-wave moveout in a single HTI layer is purely hyperbolic*.

Thus, the last two sections provide the expressions for the NMO velocity, the quartic moveout coefficient, and the horizontal velocity needed to construct the non-hyperbolic moveout equation (5) for a single HTI layer.

**Moveout in Multilayered Media**

In multilayered anisotropic media, both the quadratic and quartic moveout coefficients reflect the combined influence of layering and anisotropy. On conventional-length spreads, the hyperbolic moveout equation (3) can be expected to provide an adequate description of moveout, but the NMO velocity should be averaged over the stack of layers. In isotropic and VTI media, this averaging is performed by means of the conventional isotropic DIX (1955) equation (Hake et al., 1984). Furthermore, Alkhalifah and Tsvankin (1995) showed that the DIX equation remains valid in symmetry planes of any anisotropic medium, if the interval NMO velocities are evaluated at the ray-parameter value of the zero-offset ray. In Appendix C we extend this generalized DIX equation to arbitrary CMP directions in azimuthally anisotropic media under the assumption that the group-velocity vector does not deviate from the vertical incidence plane (the orientation of the phase-velocity vector in this case has no influence on the results). For horizontal interfaces, the ray parameter of the zero-offset ray is always zero, and our generalized NMO equation (C4) reduces to the conventional DIX formula:

\[
V_{nmo}^2 = \frac{1}{t_0} \sum_{i=1}^{N} V_{s_i}^2 \Delta t_i, \tag{15}
\]

where \(t_0\) is the two-way zero-offset time to reflector \(N\), \(V_{s_i}\) is the NMO velocity for each individual layer \(i\), and \(\Delta t_i\) is the two-way zero-offset time in layer \(i\). The interval NMO velocity \(V_{s_i}\) for any wave type in HTI media is given by equation (6).

Therefore, rms averaging of the interval velocities should provide sufficient accuracy in azimuthally anisotropic media if the group-velocity vector (ray) is confined to the incidence plane for the whole raypath. This assumption is strictly satisfied only in a single HTI layer, as well as in any other homogeneous layer with a horizontal symmetry plane. However, incident and reflected rays can be expected to lie close to the incidence plane even in stratified models with a similar character.

* Using equation (A4) from Appendix A to evaluate the \(S^\perp\)-phase velocity and its derivatives needed in equation (8) yields \(A_4^{(S^\perp)} = 0\).
of azimuthal velocity variations in all layers (e.g., in HTI media with a uniform crack orientation). In principle, the Dix equation can be extended to account for arbitrary ray trajectories in azimuthally anisotropic media, but the discussion of this more general averaging expression is outside of the scope of this paper. Later on, we will study the applicability of equation (15) to multilayered HTI models.

To use the nonhyperbolic moveout equation (5) in multilayered media, we also need to account for the influence of layering on the quartic moveout term. The exact coefficient \( A_4 \) for pure modes in VTI media was presented by Hake et al. (1984):

\[
A_4 = \frac{(\sum_i^N V_{z_1}^2 \Delta t_i)^2 - t_0 \sum_i^N V_{z_1}^2 \Delta t_i}{4(\sum_i^N V_{z_1}^2 \Delta t_i)^4} \nonumber \\
+ \frac{t_0 \sum_i^N A_{4} V_{z_1}^2 \Delta t_i^3}{(\sum_i^N V_{z_1}^2 \Delta t_i)^4}, \tag{16}
\]

where \( A_{4i} \) is the quartic moveout coefficient for layer \( i \).

The first term in the right-hand side of equation (16) has the same form as for isotropic media (Al-Chalabi, 1974), but it contains the interval NMO velocities which are different from the true vertical velocities in the presence of anisotropy. The second term goes to zero in isotropic or elliptically anisotropic media and, therefore, represents a purely anisotropic contribution to the quartic moveout coefficient. Tsvankin and Thomsen (1994) showed that the nonhyperbolic moveout equation (5) with the quartic term given by equation (16) accurately describes \( P \)-wave reflection moveout in multilayered VTI media.

In stratified HTI media, both phase- and group-velocity vectors deviate from the incidence plane, which violates the main assumptions behind the VTI averaging [equation (16)]. However, we can still expect equation (16) to provide reasonable accuracy in HTI media if we use the exact expressions for the interval values \( V_{z_1} \) [equation (6)] and \( A_{4i} \) [equation (8)] that honor the azimuthal dependence of the moveout coefficients. Therefore, in the numerical examples below both the quadratic and quartic moveout coefficients in layered HTI media are calculated using the same averaging equations as for VTI media, but with the exact interval values derived for horizontal transverse isotropy.

The effective horizontal velocity (\( V_{\text{hor}} \)) contained in the term \( A \) of the nonhyperbolic moveout equation (5) can be computed in several different ways (Alkhalifah, 1996). In VTI media, it is usually sufficient to apply a simple rms average (Tsvankin and Thomsen, 1994):

\[
V_{\text{hor}}^2 = \frac{1}{t_0} \sum_i^N V_{\text{hor}}^2 \Delta t_i, \tag{17}
\]

where \( V_{\text{hor}} \) is the interval horizontal velocity in layer \( i \). The interval horizontal velocity \( V_{\text{hor}} \) in HTI media is sufficiently approximated by equation (2) evaluated at the azimuth of the CMP line. Equation (17) is used in most numerical examples in this work.

However, it is important to mention that the nonhyperbolic moveout equation (5) with the horizontal velocity determined by equation (17) is not well-suited for isotropic layering and, consequently, for the vicinity of the isotropy plane in HTI media. In isotropic or elliptically anisotropic media the interval values of \( V_{\text{rms}} \) and \( V_{\text{hor}} \) are identical, and the effective NMO velocity [equation (15)] becomes equal to the effective horizontal velocity [equation (17)]. As a result, the coefficient \( A \) in the denominator of the nonhyperbolic moveout term goes to infinity, and equation (5) reduces to the hyperbolic equation (3). This means that the rms averaging in equation (17) does not allow us to account for nonhyperbolic moveout in layered isotropic (or elliptically anisotropic) media.

Therefore, it is desirable to find an alternative way to compute the effective horizontal velocity. In theory, the effective horizontal velocity should be equal to the maximum horizontal velocity of the medium. However, application of the maximum horizontal velocity in equation (5) may generate erroneous results in the presence of thin high-velocity layers embedded in a medium with a much lower velocity (Alkhalifah, 1996). Another altern-
ative to avoid these difficulties is to use a different averaging equation that makes the effective horizontal velocity closer to the maximum $V_{\text{hor}}$:

$$V_{\text{hor}}^2 = \frac{1}{t_0^2} \sum_{i}^{N} V_{\text{hor},i}^2 \Delta t_i.$$  \hspace{1cm} (18)

The approximations made in this section allowed us to apply the concise averaging equations developed for vertical transverse isotropy at the expense of partly ignoring out-of-plane phenomena in multilayered HTI media. The accuracy of these approximations will be studied in the next section.

**Numerical Study of P-Wave Moveout in HTI Media**

Here, we present a numerical study of P-wave reflection moveout in order to test the accuracy of the hyperbolic [equation (3)] and nonhyperbolic [equation (5)] moveout equations in HTI media. For single-layer models, we need to find out whether the exact NMO velocity [equation (7)] makes the hyperbolic moveout equation sufficiently accurate on conventional-length spreads and whether the new equation (11) for the quartic moveout coefficient allows the nonhyperbolic equation (5) to match the reflection traveltime at large offsets feasible for reflection surveys. For multilayered media, it is also necessary to verify the validity of the VTI averaging expressions for the NMO velocity [equation (15)] and for the quartic moveout coefficient [equation (16)] outside the vertical symmetry planes of HTI media.

The exact traveltimes were computed using a 3-D anisotropic ray-tracing code developed by Gajewski and Psencik (1987). The moveout velocity on finite spreads was obtained by least-squares fitting of a hyperbolic moveout equation to the calculated traveltimes, i.e.,

$$V_{\text{mo}}^2 = \frac{\sum_{i}^{N} x_i^2}{\sum_{i}^{N} t_i^2 - \sum_{i}^{N} t_0^2},$$  \hspace{1cm} (19)

where $x_i$ is the offset of the $i$-th trace, $t_i$ is the corresponding two-way reflection traveltime, $t_0$ is the two-way vertical traveltime, and $N$ is the number of traces. In our least-squares fitting procedure, the vertical time $t_0$ was fixed at the correct value.

**Single HTI layer**

In HTI media, it is sufficient to study reflection moveout in a single quadrant (Figure 3). First, we test the accuracy of the P-wave hyperbolic moveout equation parameterized by the exact NMO velocity [equation (7)]. As demonstrated in Figure 4, the moveout velocity obtained from the exact traveltimes using equation (19) is close to the analytic NMO value [equation (7)] on conventional-length spreads ($X/D \leq 1$) (Figure 4a,c). Predictably, the difference between the two velocities increases on longer spreads (Figure 4b,d) due to the anisotropy-induced deviations of the moveout curve from a hyperbola. It should be mentioned that Model 2 is characterized by an uncommonly high magnitude of nonhyperbolic moveout.

Figure 4 shows that the difference between the finite-spread and NMO velocity reaches its maximum in the symmetry-axis plane (azimuth = 0°) and goes to zero in the isotropic plane (azimuth = 90°). Clearly, despite the influence of out-of-plane phenomena, the magnitude of nonhyperbolic moveout outside the symmetry planes is smaller than in the direction of the symmetry axis. This result could be expected from the azimuthal dependence of the quartic moveout coefficient (11). As discussed above, reflection moveout in the symmetry-axis plane is described by the VTI equations extensively studied in the literature (Hake et al., 1984; Tsvankin and Thomsen, 1994; Alkhalifah, 1996).

The inadequacy of the hyperbolic moveout equation (3) on long spreads is illustrated in more detail by Figure 5a,c showing the time residuals after the conventional hyperbolic moveout correction. Deviation of the hyperbolic curve from the exact traveltimes is much more pronounced for Model 2 that has an extremely large quartic moveout term (Figure 5c). For such uncommonly strong anisotropy, conventional hyperbolic velocity analysis does not work well even on spreadlengths close to the reflector depth. In contrast, the nonhyperbolic moveout equation (5), that includes the exact quadratic and quartic moveout coefficients, provides excellent accuracy for both HTI models and for the whole range of offsets shown in Figure 5b,d. Application of the nonhyperbolic moveout equation reduces the residual moveout at large offsets (i.e., twice as large as the reflector depth) by a factor of ten compared to the residuals after the conventional hyperbolic correction. The excellent accuracy of the nonhyperbolic moveout equation is further illustrated by Figure 6 that shows a detailed comparison between the exact traveltimes and both moveout approximations at four azimuthal directions including the symmetry-axis and isotropy planes.

**Multilayered HTI media**

For the single-layer HTI model the group-velocity vectors (rays) of the incident and reflected waves are always confined to the vertical incidence plane. This is not the case for multilayered HTI media, in which both group- and phase-velocity vectors deviate from the incidence plane to satisfy Snell's law at each interface. As discussed above,
the conventional Dix equation is no longer strictly valid if the group-velocity vector does not lie in the incidence plane for the whole raypath.

However, the effective normal-moveout velocity calculated by the Dix averaging of the exact interval values (equation (15)) is sufficiently close to the moveout velocity determined from hyperbolic moveout analysis on conventional-length spreads close to the reflector depth (Figure 7). The maximum deviation of the NMO velocity from the finite-spread value is observed for the CMP line in the throughgoing symmetry-axis plane of the medium ($\alpha = 0^\circ$). As mentioned above, the Dix equation (15) is entirely valid for the NMO velocity in the symmetry-axis plane, and the difference between the two velocities at $\alpha = 0^\circ$ is caused by the influence of nonhyperbolic moveout. Outside the symmetry planes, the Dix equation (15) itself may become inaccurate since the group-velocity vector deviates from the incidence plane. However, Figure 7 shows that the Dix averaging provides a good approximation for NMO velocity at all azimuths, and the smaller magnitude of nonhyperbolic moveout away from the symmetry-axis plane makes the NMO equation (15) even more accurate in out-of-plane directions.

A detailed comparison between the exact traveltimes and the hyperbolic moveout equation (3) in the same three-layer model is shown in Figures 8a and 9. Clearly, the hyperbolic moveout equation based on the exact interval NMO velocities averaged by the Dix formula (15) provides a good approximation to the traveltimes on spreads that do not exceed the reflector depth. This conclusion holds even for the model with depth-varying orientation of the symmetry axis from Figure 10a.
Figure 5. Comparison between the exact traveltimes and moveout approximations in an HTI layer. The gray curves are the exact refraction traveltimes in two models as functions of the offset-to-depth ratio for azimuths α of 0°, 30°, 45°, 60°, and 90°. (a) and (b) = Model 1 from Figure 4; (c) and (d) = Model 2 from Figure 4. The black curves in (a) and (c) are the time residuals after the conventional hyperbolic moveout correction using equation (3), while the black curves in (b) and (d) are the residuals after the nonhyperbolic moveout correction using equation (5).

Different character of the azimuthal velocity variations from layer to layer in this model causes a more significant deviation of the group-velocity vector from the incidence plane. Nonetheless, our results indicate that this deviation in typical HTI models is not large enough to cause measurable errors in the Dix equation. Note, however, that the hyperbolic moveout equation breaks down if we disregard the azimuthal dependence of the interval NMO velocities described by equation (7). Application of any single value of NMO velocity would lead to misalignment of reflection events and poor stacking quality in certain ranges of azimuthal angles.

As in the homogeneous model, the error of the hyperbolic moveout equation increases with offset, this time due to the combined influence of anisotropy and layering (Figures 7-10). To describe long-spread moveout in layered HTI media, we use equation (5) with the effective values of the moveout coefficients given by equations (15), (16), and (17). Despite the approximate character of the averaging equation (16) for the quartic moveout term, the nonhyperbolic moveout equation (5) provides excellent accuracy for multilayered HTI media as demonstrated in Figures 8b and 9. The nonhyperbolic moveout curve [equation (5)] is sufficiently close to the exact traveltimes even in strongly anisotropic models with depth-varying orientation of the symmetry axis (Figure 10b). For the models we have considered, the residual moveout at large offsets (twice the reflector depth) after the nonhyperbolic moveout correction is about one order of magnitude lower compared to the residual after the hyperbolic correction using equation (3).

A stack of isotropic layers over an HTI layer

In the previous examples, we examined reflection moveout in models consisting only of azimuthally anisotropic layers with the HTI symmetry. However, fractured reservoirs are often overlain by an isotropic overburden, possibly with a vertical velocity gradient. It seems that the presence of isotropic layers should help to mitigate out-of-plane phenomena since phase- and group-velocity vectors in isotropic media coincide with each other. While
Figure 6. Accuracy of the moveout approximations for Model 2 at four different CMP azimuths: 0° (a and b), 30° (c and d), 60° (e and f), and 90° (g and h). The exact traveltimes on all plots are shown by the gray curves. For comparison, the plots on the left (a, c, e, and g) also display the traveltimes calculated from the hyperbolic moveout equation (3) (dark gray curves), while the plots on the right (b, d, f, and h) show the traveltimes calculated from the nonhyperbolic moveout equation (5) (black curves). The model parameters are given in Figure 4. The nonhyperbolic moveout equation accurately approximates the exact traveltimes. The maximum error of both approximations corresponds to the symmetry-axis plane (azimuth 0°), while in the isotropy plane (azimuth 90°) the reflection moveout is purely hyperbolic and the error goes to zero.
Figure 7. Accuracy of the Dix equation for P-wave normal-moveout velocity in layered HTI media. The solid curve is the moveout velocity as a function of azimuth determined by fitting a hyperbola to the exact $t^2 - x^2$ curves [equation (19)] on two different spreadlengths: $X/D = 1$ (a), and $X/D = 2$ (b); the dashed curve is the normal-moveout (zero-spread) velocity from the Dix equation (15). The velocities are calculated for the reflection from the bottom of the third layer in a three-layer model with the same orientation of the symmetry axis in all layers. For layer 1, $\epsilon (V) = -0.143 (\epsilon = 0.2)$, $\delta (V) = -0.184 (\delta = 0.1)$, $V_p_{vert} = 2.662 \text{ km/s} \ (V_p = 2.25 \text{ km/s})$, and the depth of the bottom $d_1 = 0.7 \text{ km}$; for layer 2 $\epsilon (V) = -0.045 (\epsilon = 0.05)$, $\delta (V) = -0.203 (\delta = -0.15)$, $V_p_{vert} = 2.622 \text{ km/s} \ (V_p = 2.35 \text{ km/s})$, and the $d_2 = 1.0 \text{ km}$; for layer 3 $\epsilon (V) = -0.143 (\epsilon = 0.2)$, $\delta (V) = -0.318 (\delta = -0.2)$, $V_p_{vert} = 2.988 \text{ km/s} \ (V_p = 2.5 \text{ km/s})$, and $d_3 = 1.5 \text{ km}$. Note that the difference between the two velocities reaches its maximum in the symmetry-axis plane (azimuth 0°), and goes to zero in the isotropy plane (azimuth 90°).

Figure 8. Comparison between the exact traveltimes and the moveout approximations for the three-layer HTI model from Figure 7. The gray curves are the exact reflection traveltimes from all three interfaces for azimuths $\alpha$ of 0°, 30°, 45°, 60°, and 90°. The black curves in (a) are the time residuals after the conventional hyperbolic moveout correction using equation (3) with the NMO velocity from equation (15), while the black curves in (b) are the residuals after the nonhyperbolic moveout correction using equation (5) and the effective coefficients from equations (15), (16), and (17).

for the phase-velocity vector that may indeed be the case. Deviation of the group-velocity vector from the incidence plane may even increase if we replace some of the HTI layers in the model by isotropic ones. To keep the group-velocity vector close to the incidence plane, the character of azimuthal velocity variations should be similar throughout the section. In the limit of a single HTI layer (a uniform character of azimuthal anisotropy), the group-velocity vector is strictly confined to the incidence plane. During the transmission from an HTI layer into a purely isotropic layer, the phase-velocity vector stays in the same vertical plane, and the group-velocity vector has to coincide with the phase-velocity vector after the transmission. Since in the HTI layer the azimuths of the group- and phase-velocity vectors may be substantially different, this may lead to a significant azimuthal rotation of the group vector at the boundary and potential errors in the averaging equations (15) and (16).

Nevertheless, as demonstrated by Figure 12, our conclusions drawn for layered HTI models remain essentially valid in this case as well. The model in Figure 12 includes five isotropic layers above a layer with HTI sym-
Figure 9. Accuracy of the moveout approximations for the three-layer HTI model from Figure 7 at four different CMP azimuths: $0^\circ$ (a and b), $30^\circ$ (c and d), $60^\circ$ (e and f), and $90^\circ$ (g and h). The exact traveltimes on all plots are shown by the gray curves. For comparison, the plots on the left (a, c, e, and g) also display the traveltimes calculated from the hyperbolic moveout equation (3) (dark gray curves), while the plots on the right (b, d, f, and h) show the traveltimes calculated from the nonhyperbolic moveout equation (5) (black curves). The nonhyperbolic moveout equation accurately approximates the exact traveltimes.
Figure 10. Comparison between the exact traveltimes and the moveout approximations for a three-layer HTI model with varying orientation of the symmetry axis. The azimuth of the symmetry axis in the second layer is rotated counterclockwise by 60° with respect to the axes in the first and third layer. The gray curves are the exact reflection traveltimes from all three interfaces for azimuths of 0°, 30°, 45°, 60°, and 90°. (The azimuthal angles are calculated with respect to the symmetry-axis direction in the first layer; the azimuth of the axis in the second layer is 60°.) The black curves in (a) are the time residuals after the conventional hyperbolic moveout correction using equation (9) with the NMO velocity from equation (15), while the black curves in (b) are the residuals after the nonhyperbolic moveout correction using equation (5) and the effective coefficients from equations (15), (16), and (17). Except for the axis orientation in the second layer, the model parameters are the same as in Figure 7.

metry. The hyperbolic moveout equation parameterized by the NMO velocity (15) gives an adequate description of the moveout on conventional spreads, while the nonhyperbolic moveout equation (5) is close to the reflection traveltimes at large horizontal offsets. The accurate result for the azimuthally anisotropic sixth layer was ensured by using the exact interval expressions (7) and (11) for the quadratic and quartic moveout coefficients.

However, as mentioned earlier, the rms averaging of the horizontal velocity that we used before [equation (17)] completely eliminates the nonhyperbolic moveout term in equation (5) in the case of vertically inhomogeneous isotropic media. To properly account for nonhyperbolic moveout associated with isotropic layering, the effective horizontal velocity for the first five interfaces was chosen to be equal to the maximum horizontal velocity of the medium above the reflector. The effective $v_{\text{eff}}$ for the reflection from the bottom of the HTI layer was obtained by fourth-power averaging [equation (18)]. Note the significant improvement in accuracy that we gained by using the nonhyperbolic moveout equation instead of the hyperbolic one (Figure 12), especially for the deepest event (reflection from the bottom of the HTI layer). This difference becomes even more apparent on longer spreads and for models with more pronounced anisotropy.

Discussion and Conclusions

We have presented an analytic description of nonhyperbolic (long-spread) reflection moveout in layered TI media with a horizontal symmetry axis. For a single-layer HTI model, hyperbolic moveout equation parameterized by the exact NMO velocity given in Tsvankin (1996b) provides a good approximation for $P$-wave traveltimes on conventional-length CMP spreads close to the reflector depth. However, the accuracy of the hyperbolic equation rapidly decreases with offset due to the influence of anisotropy-induced nonhyperbolic moveout. While reflection traveltimes (as well as all other kinematic signatures of body waves) in the symmetry-axis plane of HTI media can be found by analogy with vertical transverse isotropy, description of nonhyperbolic moveout outside the symmetry planes requires a special treatment.

Here, we have derived an exact expression for the azimuthally-dependent quartic moveout coefficient $A_4$ valid for any pure mode in a homogeneous HTI layer. The equation for $A_4$ has an extremely simple form, with a single trigonometric function ($\cos^4 \alpha$, where $\alpha$ is the angle with the symmetry axis) multiplied with the quartic coefficient in the symmetry-axis plane. Therefore, the magnitude of nonhyperbolic moveout rapidly decreases with azimuth away from the symmetry-axis plane, where it can be obtained by analogy with VTI media. To account for deviations from hyperbolic moveout on long spreads (2-3 times as large as the reflector depth), we have substituted the exact azimuthally-dependent values of the NMO velocity and the quartic moveout coefficient into the nonhyperbolic moveout equation developed by Tsvankin and Thomsen (1994) for VTI media. Numerical examples show that this equation provides excellent accuracy for $P$-waves recorded in all azimuthal directions.
over an HTI layer, even for models with significant velocity anisotropy and pronounced nonhyperbolic moveout.

In multilayered HTI media, the moveout coefficients reflect the combined influence of layering, polar and azimuthal anisotropy. By extending the generalized Dix (1955) equation of Alkhalifah and Tsvankin (1995) to off-symmetry directions in azimuthally anisotropic media, we show that the NMO velocity in a stack of horizontal HTI layers can be obtained by the conventional Dix rms averaging procedure if the group-velocity vector (ray) is confined to the incidence plane. Although the rays do diverge from the incidence plane on CMP lines in off-symmetry directions of multilayered HTI models, the magnitude of these deviations usually is not sufficient to cause measurable errors in the Dix equation, especially for models with a similar character of the azimuthal velocity variations in all layers (e.g., media with uniform orientation of cracks). To determine the quartic moveout term in stratified HTI media, we use the same averaging equations as for vertical transverse isotropy (Hale et al., 1984; Tsvankin and Thomssen, 1994), but with the exact interval values of the NMO velocity and the quartic moveout coefficient for each HTI layer. These effective values of the NMO velocity and the quartic moveout coefficient are used in the same nonhyperbolic moveout equation as in the single-layer model. Extensive numerical testing for stratified HTI media with both uniform and depth-
varying orientation of the symmetry axis (as well as for models composed of HTI and isotropic layers) demonstrates sufficient accuracy of our nonhyperbolic approximation in the description of long-spread reflection moveout.

The nonhyperbolic moveout equation discussed above can serve as a replacement for ray tracing in modeling of long-spread moveout in layered HTI media. Also, analytic results of this work have important applications in the inversion of reflection data for horizontal transverse isotropy. As proved by our numerical study, normal-moveout velocity in typical HTI models can be obtained from hyperbolic semblance analysis on conventional-length spreads that do not exceed the reflector depth. For models with unusually pronounced deviations from hyperbolic moveout, NMO velocity can be determined by means of the more accurate nonhyperbolic semblance analysis (Alkhalifah, 1996). Subsequent Dix differentiation of the moveout-velocity function yields the interval values of the NMO velocity, which can be inverted for the anisotropic parameters of the HTI medium. The azimuthally-dependent P-wave NMO velocity in a given layer provides enough information to determine the azimuth of the symmetry axis (crack orientation), the vertical velocity, and the anisotropic coefficient $\delta^\text{V}$. In the special cases described by Tsvankin (1996b), the results of the NMO-velocity inversion are sufficient to constrain the crack density – a parameter of considerable interest in seismic characterization of fractured reservoirs. In principle, with the analytic formalism presented in this paper, we can use the nonhyperbolic portion of the moveout curve to determine the coefficient $\delta^\text{V}$ and use it to estimate the crack density from P-wave surface seismic data. However, this procedure suffers from the trade-off between the quadratic and quartic moveout terms (Tsvankin and Thomsen, 1993). The accuracy and stability of the inversion of P-wave moveout data will be discussed in a sequel paper.

Acknowledgments

We are grateful to Ken Larner and Tariq Alkhalifah (both of CSM) for useful discussions and to Dirk Gajewski (University of Hamburg) for providing his 3D ray-tracing code that was extensively used in our numerical study. We would like to thank all members of the “Anisotropy-Team” at the Center for Wave Phenomena (CWP) for useful critiques. AbdulFattah Al-Dajani would like to thank the Saudi Arabian Oil Company (Saudi Aramco) for the financial support. Al-Dajani is especially grateful to Mahmoud Abdul-Baqi and John Ward of Saudi Aramco for making the scholarship at CSM possible.

References


Tsvankin, I., 1996a, P-wave signatures and notation for transversely isotropic media: Geophysics, 61, 467-483.


Tsvankin, I., and Thomsen, L., 1994, Nonhyperbolic re-
APPENDIX A: Two Sets of Thomsen Parameters for HTI Media

The HTI model is usually characterized by the following stiffness tensor $c_{ijkl}$ that corresponds to the coordinate frame in which $z_1$ represents the symmetry axis (Figure 1):

\[
C_{HTI} = \begin{pmatrix}
c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\
c_{12} & c_{33} & c_{34} & 0 & 0 & 0 \\
c_{13} & c_{34} & c_{22} & 0 & 0 & 0 \\
0 & 0 & 0 & c_{44} & 0 & 0 \\
0 & 0 & 0 & 0 & c_{55} & 0 \\
0 & 0 & 0 & 0 & 0 & c_{55}
\end{pmatrix}
\]

The "generic" Thomsen's (1986) parameters defined with respect to the symmetry axis are given by (Tsvankin, 1996b, Rüger, 1996):

\[
\begin{aligned}
\epsilon &= \frac{c_{33} - c_{11}}{2c_{11}}, \\
\delta &= \frac{(c_{12} + c_{13})^2 - (c_{11} - c_{33})^2}{2c_{11}(c_{11} - c_{33})}, \\
\gamma &= \frac{c_{44} - c_{55}}{2c_{55}}, \\
V_{PP} &= \sqrt{\frac{c_{11}}{\rho}}, \\
V_{SS} &= \sqrt{\frac{c_{55}}{\rho}},
\end{aligned}
\]

where $\rho$ is the medium density, and the velocities $V_{PP}$, $V_{SS}$, and $V_{SS}$ correspond to the symmetry-axis direction.

The parameters of the equivalent VTI medium introduced in equation (1) of the main text are related to the generic parameters by the following equations (Tsvankin, 1996b, Rüger, 1996):

\[
\begin{aligned}
\epsilon^{(V)} &= -\frac{1}{1 + 2\epsilon}, \\
\delta^{(V)} &= \frac{\delta - 2\epsilon}{(1 + 2\epsilon)(1 + 2\gamma f)}, \\
\gamma^{(V)} &= \frac{1}{1 + 2\gamma}, \\
V_{P_{vert}} &= V_{PP}\sqrt{1 - 2\epsilon}, \\
V_{S_{vert}} &= V_{SS}, \\
V_{SS_{vert}} &= V_{SS}\sqrt{1 + 2\gamma},
\end{aligned}
\]

where $f = 1 - \left(\frac{V_{SS}}{V_{PP}}\right)^2$.

Figure A1. For a homogeneous HTI layer, the specular reflection point for any offset coincides with the zero-offset reflection point, and there is no reflection-point dispersal on CMP gathers. $y$ denotes the CMP location and $h$ is half the source-receiver offset.

The exact expression for the phase velocity of the $P$- and $S^+$-waves in terms of the generic Thomsen parameters was presented in Tsvankin (1996a):

\[
\frac{v^2(\theta)}{V_{PP}} = 1 + \epsilon \sin^2 \theta - \frac{f}{2} \pm \frac{1}{2} \sqrt{\left(1 + 2\epsilon \sin^2 \theta\right)^2 - 2(1-\epsilon)\sin^2 2\theta},
\]

where the plus sign corresponds to the $P$-wave, and the minus – to the $S^+$-wave; $\theta$ is the phase angle with the (horizontal) symmetry axis. Phase-velocity equation (A3) is rewritten in terms of the parameters of the equivalent VTI model in the main text [equation (2)].

For the $S^+$-wave, the anisotropy is elliptical, and the phase velocity is given by (Tsvankin 1996b):

\[
V(\theta) [S^+\text{-wave}] = V_{S_{vert}}\sqrt{1 + 2\gamma^{(V)} \cos^2 \theta}.
\]

APPENDIX B: Quartic Moveout Coefficient in HTI Media

Here, the approach suggested by Tsvankin (1996b) in his derivation of the NMO velocity in HTI media is extended to obtain the quartic coefficient $A_4$ of the Taylor series expansion of the squared traveltime $[t^2(x^2)]$. First, we find an expression for $A_4$ in terms of the one-way traveltime from the zero-offset reflection point. Since a horizontal reflector coincides with a symmetry plane in HTI media, the group-velocity (ray) vector of any pure reflected wave in an HTI layer represents the mirror image of the incident ray with respect to the horizontal plane (Figures A1 and B1). This means that there is no reflection-point dispersal on CMP gathers above a homogeneous HTI layer, and we can represent the two-way traveltime along the specular raypath as the sum of the traveltimes from the zero-offset reflection point to the source and receiver (Figure A1). The one-way traveltime to the source
or receiver can be expanded in a Taylor series in the powers of the half-offset $h$. This expansion was first used by Haile et al. (1992) to express the normal-moveout velocity from dipping reflectors through the ray parameter.

Here, we are interested in deriving the quartic moveout coefficient, so we will keep the quartic and lower-order terms in the Taylor series:

$$t(y + h) = t(y) + h \frac{dt}{dx} + \frac{h^2}{2} \frac{d^2t}{dx^2} + \frac{h^3}{6} \frac{d^3t}{dx^3} + \frac{h^4}{24} \frac{d^4t}{dx^4} + \cdots$$

$$t(y - h) = t(y) - h \frac{dt}{dx} = \frac{h^2}{2} \frac{d^2t}{dx^2} + \frac{h^3}{6} \frac{d^3t}{dx^3} + \frac{h^4}{24} \frac{d^4t}{dx^4} + \cdots$$

where all derivatives are evaluated at the CMP location $y$.

Summing the two series expansions above, we obtain:

$$t_h = t(y + h) + t(y - h) = t_0 + h^2 \frac{d^2t}{dx^2} + \frac{h^4}{12} \frac{d^4t}{dx^4},$$

(B2)

where $t_0$ is the two-way zero-offset travelt ime. Note that in the case of a horizontal reflector beneath HTI media, both the phase- and group-velocity (ray) vectors of the zero-offset reflection are vertical.

Squaring both sides of equation (B2) and ignoring terms of higher order than $h^4$ yields

$$t_h^2 = t_0^2 + 2 t_0 d^2 t \frac{dx^2}{2} h^2 + \left[ \frac{d^2 t}{dx^2} \right]^2 \frac{h^4}{6}.$$

(B3)

Comparing equation (B3) with the Taylor’s series expansion of the squared reflection travelt ime $[t^2(x^2)]$

$$t_h^2 = t_0^2 + A_2 (2 h)^2 + A_4 (2 h)^4,$$

we find the quadratic ($A_2$) and the quartic moveout coefficients ($A_4$) as

$$A_2 = \lim_{h \to 0} \left[ \frac{t_0 d^2 t}{2 h^2} \right],$$

(B5)

and

$$A_4 = \frac{1}{16} \lim_{h \to 0} \left[ \frac{d^2 t}{h^2} \right]^2 + \frac{t_0 d^4 t}{6 h^4}. $$

(B6)

Tsvankin (1996b) used equation (B5) to derive the normal-moveout velocity ($V_{NM} = \frac{1}{A_2}$) in a single HTI layer as a function of phase velocity and the symmetry-axis azimuth. Here, we apply equation (B6) to obtain the quartic moveout coefficient $A_4$. Equations (B3) and (B6) are valid for arbitrary anisotropic media if the specular reflection point does not change with offset on CMP.

Figure B1. The group- and phase-velocity vectors for the reflected waves in a homogeneous HTI layer. The incident (SO) and reflected (OR) group-velocity vectors (rays) lie in the vertical incidence plane and are symmetric with respect to the horizontal plane. The phase-velocity vector (direction OD) of the reflected ray OR is confined to the plane formed by OR and the axis of symmetry. Triangle ROB defines a plane normal to the symmetry axis (after Tsvankin, 1996b).

gathers. It turns out, however, that reflection-point dispersal does not influence the value of normal-moveout velocity ($A_2$) (Hubral and Krey, 1980, Appendix D; Tsvankin, 1995), and equation (B3) can be used for both horizontal and dipping reflection events in media with any symmetry. Equation (B6) for the quartic term is more restrictive and can be applied only in the absence of reflection-point dispersal (which is the case for our model).

We consider a CMP line that makes the azimuthal angle $\alpha$ with the symmetry axis (Figure B1). Since the derivative $dt/dh$ represents the apparent slowness on the CMP gather, it equal to the projection of the slowness vector on the CMP line:

$$p_h = \frac{dt}{dh},$$

and the quartic moveout coefficient [equation (B6)] can be rewritten as

$$A_4 = \frac{1}{16} \lim_{h \to 0} \left[ \frac{dp_h}{dh} \right]^2 + \frac{t_0}{96} \lim_{h \to 0} \left[ \frac{d^2 p_h}{dh^2} \right].$$

(B7)

Introducing the group angle $\beta$ in the incidence plane (Figure B1) and taking into account that $h = h_0 \tan \beta$ and $z_0 = V_{vert} t_0 / 2$ ($V_{vert}$ is the vertical velocity), we find:

$$A_4 = \frac{1}{4V^2_{vert} \tan^2 \beta} \lim_{\beta \to 0} \left[ \frac{dp_h}{d \tan \beta} \right]^2.$$
Nonhyperbolic reflection moveout for HTI

\[
\frac{1}{12\theta_0^2 V_{vert}^3} \lim_{\beta \to 0^+} \left[ \frac{d^3 p_h}{d (\tan \beta)^3} \right]. \tag{B8}
\]

It is convenient to represent \( \beta \) and \( p_h \) as functions of the phase angle \( \theta \) with the symmetry axis (Figure B1). Although the rays stay within the vertical incidence plane, the influence of azimuthal anisotropy moves the phase-velocity (slowness) vectors of the incident and reflected waves out of plane. Still, the phase-velocity vector in transversely isotropic media always lies in the plane formed by the symmetry axis and the group-velocity vector (Figure B1). Rewriting the derivatives in equation (B8) in terms of \( \theta \), we obtain:

\[
\frac{dp_h}{d (\tan \beta)} = \frac{dp_h}{d \theta} \frac{d \theta}{d (\tan \beta)}, \tag{B9}
\]

and

\[
\frac{d^3 p_h}{d (\tan \beta)^3} = \frac{d^3 p_h}{d \theta^3} \left( \frac{d \theta}{d (\tan \beta)} \right)^3 + 3 \left( \frac{d^2 p_h}{d \theta^2} \right) \frac{d \theta}{d (\tan \beta)} \frac{d^2 \theta}{d (\tan \beta)^2} \frac{d \theta}{d (\tan \beta)} + \frac{d^2 p_h}{d (\tan \beta)^2} \left( \frac{d \theta}{d (\tan \beta)} \right)^2 \tag{B10}
\]

where

\[
\frac{d \theta}{d (\tan \beta)} = \frac{d \theta}{d \theta} \left( \frac{d \theta}{d (\tan \beta)} \right) \frac{d \theta}{d (\tan \beta)}
\]

and similarly

\[
\frac{d^2 \theta}{d (\tan \beta)^2} = \frac{d^2 \theta}{d (\tan \beta)^2} \left( \frac{d \theta}{d (\tan \beta)} \right)^2 \frac{d \theta}{d (\tan \beta)}
\]

Substituting equations (B9) and (B10) into equation (B8) yields:

\[
A_4 = \frac{1}{12\theta_0^2 V_{vert}^3} \lim_{\theta \to 90^\circ} \left[ \frac{dp_h}{d \theta} \frac{d \theta}{d (\tan \beta)} \right]^4 + \frac{1}{12\theta_0^2 V_{vert}^3} \lim_{\theta \to 90^\circ} \left[ \frac{dp_h}{d \theta} \frac{d \theta}{d (\tan \beta)} \right]^3 + \frac{1}{12\theta_0^2 V_{vert}^3} \lim_{\theta \to 90^\circ} \left[ \frac{dp_h}{d \theta} \frac{d \theta}{d (\tan \beta)} \right] \tag{B11}
\]

Next, it is necessary to compute the derivatives in equation (B11). From simple trigonometry (Figure B1), we can relate the angle \( \beta \) to the group angle \( \psi \) of ray OR with the symmetry axis as suggested in Tsivkin (1996b):

\[
\sin \beta = \frac{\cos \psi}{\cos \alpha},
\]

and

\[
\tan \beta = \frac{1}{\tan \psi} \sqrt{1 - \sin^2 \alpha \sin^2 \psi}.
\]

The group angle \( \psi \) can be expressed through the phase angle \( \theta \) and phase velocity \( V(\theta) \) as (Thomsen, 1986)

\[
\tan \psi = \frac{\tan \theta + \frac{d V}{d \theta}}{1 - \tan \theta \frac{d V}{d \theta}}. \tag{B12}
\]

Applying the chain rule again, we have

\[
\frac{d \theta}{d \tan \beta} = \frac{d \psi}{d \theta} \frac{d \theta}{d \psi}.
\]

Therefore,

\[
\frac{d \theta}{d \tan \beta} = -\frac{\sin^2 \psi (1 - \sin^2 \alpha)^2}{\cos^2 \alpha} \left[ \frac{1 + \left( \frac{d V(\theta)}{d \theta} \right)^2}{1 + \left( \frac{d V(\theta)}{d \psi} \right)^2} \right]. \tag{B13}
\]

By representing \( \tan \beta \) as a function of \( \psi \) and \( \theta \), we were able to evaluate \( \frac{d \theta}{d \tan \beta} \), which also enables us to find the higher-order derivatives in equation (B11). Evaluating the derivatives at \( \theta = \psi = 90^\circ \), we obtain:

\[
\frac{d \theta}{d \tan \beta} \bigg|_{\theta=\psi=90^\circ} = -\cos \alpha \left[ \frac{1}{1 + \left( \frac{d V(\theta)}{d \psi} \right)^2} \right], \tag{B14}
\]

and

\[
\frac{d^2 \theta}{d (\tan \beta)^2} \bigg|_{\theta=\psi=90^\circ} = \cos \alpha (2 + \sin^2 \alpha) \left[ \frac{1 + \left( \frac{d V(\theta)}{d \psi} \right)^2}{1 + \left( \frac{d V(\theta)}{d \psi} \right)^2} \right] \tag{B15}
\]

The derivative \( \frac{d^2 \theta}{d (\tan \beta)^2} \bigg|_{\theta=\psi=90^\circ} \) turns out to be unnecessary in equation (B11) since \( \frac{d^2 \theta}{d \psi^2} \bigg|_{\theta=\psi=90^\circ} = 0 \).

To evaluate \( A_4 \), we also need to find the derivatives of the projection of the slowness vector on the CMP line \( p_h \) with respect to the phase angle \( \theta \). As suggested by Tsivkin (1996b), the slowness vector (which is parallel to OD in Figure B1) can be decomposed into two vectors parallel to sides OC and CD of triangle OCD. Then we find the horizontal projection of the slowness component parallel to CD using

\[
\cos(LRCB) = \frac{\tan \beta \sin \alpha}{\sqrt{1 + \tan^2 \beta \sin^2 \alpha}} = \frac{\tan \alpha}{\tan \psi},
\]

Summing up the projections of both components onto the CMP line yields

\[
p_h = \frac{1}{V} (\cos \theta \cos \alpha + \sin \theta \sin \alpha \tan \alpha / \tan \psi), \tag{B16}
\]

with \( \tan \psi \) determined by equation (B12).
The derivatives of equation (B16) with respect to \( \theta \), evaluated at \( \theta = \psi = 90^\circ \), are given by

\[
\frac{dp_n}{d\theta} \bigg|_{\theta=\psi=90^\circ} = \frac{1}{V_{\text{vert}} \cos \alpha} \left( \sin^2 \alpha \left( \frac{1}{V} \frac{d^2V}{d\theta^2} \right) \right)_{\theta=90^\circ},
\]

(B17)

\[
\frac{dp_n}{d\theta} \bigg|_{\theta=\psi=90^\circ} = 0,
\]

(B18)

and

\[
\frac{dp_n}{d\theta^2} \bigg|_{\theta=\psi=90^\circ} = \frac{\cos \alpha}{V_{\text{vert}}} \left( 1 + \frac{3}{V} \frac{d^2V}{d\theta^2} |_{\theta=90^\circ} \right) \left( \frac{1}{V} \frac{d^2V}{d\theta^2} \right) + \frac{\sin^2 \alpha}{V_{\text{vert}} \cos \alpha} \left( 1 - \frac{1}{V} \frac{d^2V}{d\theta^2} |_{\theta=90^\circ} \right).
\]

(B19)

Substituting equations (B14), (B15), (B17), (B18), and (B19) into equation (B11) leads, after algebraic transformations, to a concise final result:

\[
A_4 = \cos \alpha \left[ \frac{\frac{d^2V}{d\theta^2} + 3 \left( \frac{1}{V} \frac{d^2V}{d\theta^2} \right)^2 + \frac{d^2V}{d\theta^2}}{12V_{\text{vert}} (1 + \frac{1}{V} \frac{d^2V}{d\theta^2})^2} \right]_{\theta=90^\circ}
\]

(B20)

Note that the first and third derivatives of the phase velocity in the vertical direction (\( \theta = 90^\circ \)) go to zero. Equation (B20) is valid for any pure mode \((P, S^-, S^-)\) in HTI media with arbitrary strength of anisotropy.

APPENDIX C: Dix Equation for Layered Azimuthally Anisotropic Media

In their derivation of the generalized Dix equation for anisotropic media, Alkhalfah and Tsvankin (1993) assumed that the phase and group-velocity vectors of incident and reflected waves are confined to the sagittal (incidence) plane. This implies that the incidence plane (i.e., the vertical plane that contains the CMP line) should represent a symmetry plane of the medium and the dip plane of the reflector. Here, we show that the generalized Dix equation retains the same form outside the symmetry planes of azimuthally anisotropic media, if the group-velocity vector does not deviate from the incidence plane for the whole ray path. Although this assumption cannot be satisfied exactly for multilayered azimuthally anisotropic models, it is helpful in gaining insight into the influence of azimuthal anisotropy on the accuracy of the Dix equation.

We consider CMP reflections from either a horizontal or a dipping interface overlain by a horizontally layered arbitrary anisotropic medium (Figure C1).

Normal-moveout velocity in the CMP geometry represents the following function of the one-way traveltime \( t \) from the zero-offset reflection point (equation (B5)):

\[
V_{\text{nm}} = \frac{2}{t_0} \lim_{h \to 0} \frac{d}{dh} \left( \frac{dt}{dh} \right)^{-1} = \frac{2}{t_0} \lim_{h \to 0} \frac{dh}{dp_n},
\]

(C1)

where \( h = \sigma/2 \) is half the source-receiver offset (\( h \) is positive in the down-dip direction), \( t_0 \) is the two-way zero-offset traveltime, and \( p_n \) is the projection of the slowness vector on the CMP line.

The ray parameter \( p \) (horizontal slowness), as well as \( p_n \), remains constant along any ray above the reflector since the overburden is horizontally homogeneous. In the case considered by Alkhalfah and Tsvankin (1995), the slowness vector did not deviate from the incidence plane, and \( p_n \) was equal to \( p \). However, as shown below, the difference between \( p_n \) and \( p \) has no influence on the form of the generalized Dix equation provided the group-velocity vector (ray) stays within the incidence plane.

Since we assume that the whole raypath is confined to the incidence plane, the half-offset \( h \) can be written as
\[ h = \left( \sum_{i=1}^{n} x^{(i)} - x_0 \right) , \]

where \( x^{(i)} \) is the horizontal displacement of the ray in layer \( i \), and \( x_0 \) is the total horizontal displacement of the zero-offset ray with respect to the reflection point (Figure C1). Substituting \( h \) into equation (C1) yields

\[ V_{nmo}^2 = \frac{2}{t_0} \lim_{H \to 0} \sum_{i=1}^{n} \frac{d(x^{(i)})}{dp_h} , \]

To identify the interval values of NMO velocity in equation (C2), let us draw an imaginary reflector through the intersection of the zero-offset ray with the bottom of layer \( i \). The normal to the reflector is chosen to coincide with the slowness (phase) vector that corresponds to the segment of the zero-offset ray in this layer. Note that since the slowness vector associated with the zero-offset ray is allowed to be out of plane, the dip plane of the imaginary reflector is generally different from the incidence plane. Next, we assume that the intersection of the zero-offset ray with the top of layer \( i \) represents a common-midpoint location. Then the segment of the zero-offset ray in layer \( i \) will coincide with the raypath of the zero-offset CMP reflection from the imaginary interface. In accordance with equation (C1), NMO velocity from the imaginary reflector at this CMP location will be given by

\[ V_{nmo}^2 (s^{(i)}) = \frac{1}{t_0} \lim_{x^{(i)} \to x_0^{(i)}} \frac{d(2x^{(i)})}{dp_h} , \]

where \( t_0^{(i)} \) is the two-way traveltime along the zero-offset ray in layer \( i \), \( x_0^{(i)} \) is the horizontal displacement of the zero-offset ray in layer \( i \), and \( s^{(i)} \) is the slowness vector associated with the zero-offset ray. Now it is obvious that the summation in equation (C2) is carried out over the NMO velocities from reflectors normal to the “zero-offset” slowness vectors in each layer. Substituting equation (C3) into equation (C2) yields

\[ V_{nmo}^2 = \frac{1}{t_0} \sum_{i=1}^{n} t_0^{(i)} [V_{nmo}^2 (s^{(i)})]^2 . \]

Although this expression looks identical to the conventional Dix equation, interval NMO velocities in equation (C4) correspond to reflectors with different dips determined by the orientation of the slowness vector associated with the zero-offset ray. In contrast with the symmetry-plane Dix equation obtained by Alkhalifah and Tsvankin (1993), equation (C4) is influenced by the 3-D character of wave propagation since the normals to these reflectors are not necessarily confined to the incidence plane (Figure C1).
Analytic insight into shear-wave AVO for fractured reservoirs

Andreas Rüger
Department of Geophysics
Center for Wave Phenomena
Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
It is well established that azimuthal anisotropy caused by fracture systems in reservoir rocks has a first-order influence on kinematic properties of shear waves. Analysis of time differences between split shear waves at near-vertical incidence can yield information on the orientation of the crack system and estimates of crack density. However, this technique may have a limited vertical resolution and may not be usable if the reservoir is too thin to generate measurable time delays.

Study of the normal-incidence reflection coefficients of the two split shear modes can be used to overcome this difficulty. Here, extensions to shear-wave amplitude variation with offset (AVO) analysis are investigated using concise linearized reflection coefficients in the two vertical symmetry planes of azimuthally anisotropic media. The derived approximations explicitly describe the influence of medium parameters on the AVO gradients and the higher angle terms of the shear-wave reflection coefficients. The contribution of anisotropy in the subsurface is hereby characterized by the shear-wave splitting parameter $\gamma$ and anisotropy parameters similar to Thomsen's (1986) coefficients.

The important result of this study is that AVO gradients of shear waves propagating in the vertical symmetry planes of fractured media are sensitive to the shear-wave splitting parameter and to parameter combination $(\epsilon(V) - \delta(V))$ important for time processing in anisotropic media. Analytic insight developed in this study naturally leads to an inversion algorithm for the isotropy and anisotropy parameters of the fractured medium.

Key words: AVO, shear waves, anisotropy, HTI media

Introduction
Shear-wave surveys are routinely acquired and interpreted for the presence of azimuthal anisotropy (Michelena, 1995; Lynn, 1995; Kendall, 1995). The main objective of these studies is to relate the anisotropy (determined by conventional shear-wave splitting analysis) to the fracture parameters of the medium. Anisotropic compressional stresses required to fracture rocks are relatively small and even flat-lying strata are commonly fractured, with the most famous example probably being the Austin Chalk in Texas (Mueller, 1991).

Although natural fractures can occur in many forms and settings, depending on the orientation and the magnitude of paleo- and current stress regimes, the most plausible reason for the observed anisotropy is the presence of aligned cracks in an isotropic matrix (Crampin, 1983; Crampin & Atkinson, 1983). A first-order model that describes the presence of vertical stress-aligned cracks is the transversely isotropic medium with a horizontal axis of symmetry (HTI) (Thomsen, 1988). Most of the derivations and insights developed in this paper are restricted to HTI models, but extensions of this work to orthorhombic symmetry systems are discussed and presented in more detail in Rüger (1996b). Azimuthal anisotropy has a first-order influence on vertically incident shear waves causing them to be split
into two components traveling with different velocities. The parameter of crack systems of great interest in exploration is the crack density, which is approximated by the fractional difference between the velocities of split shear waves at vertical incidence [Thomsen's (1986) coefficient $\gamma$]. Estimates of anisotropy using the time delay between the fast and slow shear waves yield robust measurements of $\gamma$ averaged over the propagation path of the shear waves within the medium. Thus, analysis of shear-wave kinematic signatures via the classical rotation analysis (Alford, 1986) helps to estimate the crack density (for HTI media), albeit with a low vertical resolution.

For situations where a detailed measure of local anisotropy (highly resolved in depth/time) is desired, or where the rotation analysis fails due to insignificant time delays, Thomsen (1988), Lynn and Thomsen (1990) and Yardley et. al. (1991) suggest using the differences in normal-incidence reflection coefficients and the amplitude-variation-with-offset signature between the fast and slow shear waves to characterize the target horizon.

The reflection response of plane waves at interfaces between anisotropic media has been thoroughly investigated in the literature (Musgrave, 1970; Henneke, 1972; Keith & Crampin, 1977). The computation of the reflection coefficient is straightforward and generally requires a numerical inversion of a 6x6 matrix. The entries of the matrix include polarizations and tractions of three different wave modes in both media that have to be obtained by the eigenvector analysis of the Christoffel equation.

Yardley et. al. (1991) used the reflectivity method to compute the reflection coefficients as a function of incidence angle for various azimuthally anisotropic two-layer models. In their studies, they modeled layers of HTI symmetry that contain parallel vertical cracks normal and oblique to the survey line. They qualitatively described and discussed changes of the reflection response for layers with water- and gas-filled cracks and observed a dependence of the AVO response on both the crack density and crack content. Yardley et. al. (1991) also emphasized severe difficulties in measuring shear-wave amplitudes if the acquisition (survey) line is not parallel or perpendicular to the crack orientation.

The results obtained by Yardley et. al. (1991) motivate a more detailed analysis of the AVO problem for shear waves in HTI media. To overcome the algebraic complexity of the reflection/transmission problem in anisotropic media, I hereby follow the approach devised by Richards and Frasier (1976), Aki and Richard (1980) and Shuey (1985) for isotropic media to derive concise analytic approximations of the reflection and transmission coefficients. The resulting equations, which are linear in the relative changes of the elastic parameters across the reflecting boundary, reveal the dependence of the normal incidence reflection coefficient and the AVO gradient on the anisotropy in the medium.

The analytic equations for shear-wave reflection coefficients derived in this paper are valid for small changes in the elastic properties across interfaces between isotropic and HTI layers. The solutions are also valid for interfaces between two HTI media if the principal axes of symmetry coincide in both layers and for interfaces between HTI layers and transversely isotropic layers with a vertical axis of symmetry (VTI). A typical geologic model for the VTI/HTI boundary is a massive shale layer overlying a fractured reservoir. In all situations described in the paper, the survey line has to be either near-parallel or near-perpendicular to the crack orientation. The orientation of the symmetry axis can be obtained in many exploration contexts by in-situ stress measurements (Lynn & Thomsen, 1990), wireline data, geologic studies and $P$-wave AVO and NMO analysis (Rüger, 1996a; Rüger & Tsvankin, 1995). If the incidence and symmetry planes do not coincide, three wave modes are coupled and the polarization of the individual wave modes is more complicated, hampering a reliable extraction of the amplitudes.

Models with horizontal transverse isotropy represent simple first-order approximations of fractured media. Extension of this work to subsurfaces with (more complicated and more realistic) orthorhombic symmetry systems is briefly discussed.

**Shear-wave propagation in HTI media**

Wave propagation in anisotropic media is a complex physical phenomenon. Before discussing the feasibility of fracture characterization using dynamic signatures of
shear waves, it is useful to recall basic properties of waves propagating in HTI media and introduce notation. Let us assume that the symmetry axis of the HTI model is parallel to the $x_1$ direction (Figure 1). The $[x_1, x_3]$-plane that contains the symmetry axis is hereafter referred to as the “symmetry-axis plane.” Waves confined to the plane normal to the symmetry axis have the same velocity for all incidence angles; hence, this plane is the so-called “isotropy plane” or “fracture plane.”

Analysis of seismic waves (Červený, 1972) shows that for a given slowness, three plane-wave modes with mutually orthogonal polarizations can propagate in anisotropic media. If the medium is HTI, one mode ($S^{\parallel}$) is polarized within the isotropy plane (the $[x_2, x_3]$ plane). This mode is truly transverse since it is polarized perpendicular to the propagation direction. The two remaining modes have quasi-longitudinal ($qP$) and quasi-transverse ($S^\perp$) polarizations with respect to the slowness direction $p$ (see Figure 2). Polarization directions are defined according to the convention in Aki and Richards (1980): the positive polarization direction is such that the horizontal component points in the direction of horizontal slowness.

The minimum values of the phase velocities for the $S^{\parallel}$ and $qP$ mode are in the symmetry direction ($x_1$ axis), while the maximum occurs in the isotropy plane. By contrast, the $S^\perp$-shear mode must exhibit at least one extremum as the propagation direction changes from the symmetry-axis directions towards the isotropy plane because the $S^\perp$ phase velocity is identical in both directions. Unlike the $S^{\parallel}$ and $qP$ modes, the $S^\perp$ wave may exhibit triplicated wavefronts even in homogeneous anisotropic media. Illustrations and qualitative description of wavefront triplications can be found, for example, in Rüger (1994).

The discussion in this paper is mainly focused on the reflections of shear waves traveling in the two vertical symmetry planes of HTI models. Figures 3a and 3b illustrate how $S^{\parallel}$ and $S^\perp$ waves propagate within the isotropy plane and reflect from the interface. The HTI medium has a vertical $P$-wave velocity $V_{P0} = 2.6$ km/s and a vertical shear velocity $V_{S0} = 2$ km/s. The anisotropy simulated in this model is described by generic Tomsen’s (1986) parameters $\epsilon = 0.15, \delta = -0.1$ and $\gamma = 0.1$. The black lines emitted from the source at the origin denote seismic rays. The continuous white lines are obtained by simply connecting points of 1s traveltime of the individual rays and thus represent wavefronts. The $S^\perp$ mode shows a perfectly circular wavefront typical for isotropic wave propagation. As expected, the $S^{\parallel}$ wavefront is also circular, but it is more advanced than the $S^\perp$ wavefront and already reflects from the interface. The different speed of the two shear waves is a manifestation of the well-known phenomenon of shear-wave splitting that indicates that the subsurface model is not isotropic.

Rays and wavefronts of shear waves propagating within the symmetry-axis plane are shown in Figures 3c ($S^{\parallel}$) and 3d ($S^\perp$ mode). For near-vertical propagation, the $S^{\parallel}$ wavefront is more advanced than that of the $S^\perp$ wave. Additionally, a closer view at the figures reveals that several features of these figures differ significantly from the familiar behavior of rays and wavefronts in isotropic homogeneous media:

- The wavefronts are not circular. In particular, the $S^{\parallel}$-wavefront is most advanced in vertical direction. This can be seen by comparison with the circular $S^{\parallel}$-wavefront in the isotropy plane (Figure 3a): Both wavefronts coincide at vertical incidence, but the symmetry-axis plane $S^{\parallel}$-wavefront is less advanced for large incidence angles.
- The $S^\perp$-wavefront clearly shows an unusual angular shape although the medium is homogeneous.
- The ray density is not uniform. Moreover, the angular distribution of ray density differs for $S^{\parallel}$- and $S^\perp$-waves. While the $S^{\parallel}$-wave shows a uniform ray-distribution, a concentration of $S^\perp$-rays is seen for ray angles of about 45°.

Figures 3a-d also demonstrate the dynamic properties of the wavefields. More specifically, the ray density is related to the phenomena of energy focusing and defocusing (Tsvankin & Chesnokov, 1990; Tsvankin, 1995). Energy increases in regions with a high concentration of rays (or, equivalently, a high concentration of group velocity vectors). Conversely, a low concentration of rays indicates energy defocusing. This phenomenon needs to be taken into account for reliable AVO analysis in anisotropic media (Tsvankin, 1995). Variations of amplitudes
with offset are not only caused by the changes in the reflection coefficients with angle, but also by focusing and defocusing of energy during the wave-propagation in the anisotropic overburden. After deriving equations for reflection coefficients in HTI media, I therefore first consider the simpler case of an isotropic layer overlying a fractured layer. In this case, the wave-propagation effects are limited to spherical divergence of energy. The more complex case of an HTI layer overlying an isotropic layer or the contact between two anisotropic layers is discussed later in the text.

**Linearized reflection coefficients**

Physical constraints on the continuity of the displacement and traction between layers in welded contact yield a system of equations that can be solved for the reflection and transmission coefficients. Depending on the type of symmetry above and below the interface and the incident-wave mode, two, four or six boundary conditions have to be solved for the amplitude coefficients of the reflected and transmitted waves.

Mathematically, the equation to be solved can be expressed as

$$ \mathbf{A} \mathbf{R} = \mathbf{b}, \quad (1) $$

where \( \mathbf{A} \) is the boundary equation matrix formed by the scattered (reflected and transmitted) waves, \( \mathbf{R} \) is the vector of reflection and transmission coefficients, and \( \mathbf{b} \) is composed of the contribution of the incident wave to the boundary conditions (Aki & Richards, 1980; Thomsen, 1993).

Solutions for the plane-wave reflection and transmission coefficients are algebraically very involved and require an inversion of the boundary matrix \( \mathbf{A} \). The complexity of the energy-partitioning problem at planar interfaces obscures any insight into the dependence of the reflection coefficient on the medium parameters and incidence angle. To overcome this difficulty, I follow Richards and Frasiers (1976) and Aki and Richards (1980) approach of linearizing the reflection coefficient with respect to relative changes in elastic parameters across the interface. Extensions of this approach to AVO-analysis have been developed by Shuey (1983) and Wright (1986) (for isotropic media) and for transversely isotropy media by Banik (1987), Thomsen (1993) and Rüger (1996a). In most exploration contexts, the two layers under consideration have similar properties and it is convenient to express the elastic properties as functions of the average density \( \bar{\rho} = 1/2 (\rho_2 + \rho_1) \) and the difference \( \Delta \rho = \rho_2 - \rho_1 \):

$$ \rho_1 = \bar{\rho} \left(1 - \frac{1}{2} \frac{\Delta \rho}{\bar{\rho}}\right), $$

$$ \rho_2 = \bar{\rho} \left(1 + \frac{1}{2} \frac{\Delta \rho}{\bar{\rho}}\right). $$

Analogous expressions are defined for the shear and compressional velocities \( \beta \) and \( \alpha \), respectively.

**Isotropic layers**

For isotropic layers, the solution for \( \mathbf{R} \) in equation (1), linearized with respect to small elastic contrast coefficients is fairly straightforward (Thomsen, 1993). The approximate (linearized) reflection coefficient for the mode polarized perpendicular to the vertical propagation plane (here denoted as the SH-mode to follow the standard notation), for example, yields

$$ R_{33}^{\text{iso}} = -1/2 \left(\frac{\Delta \rho}{\bar{\rho}} + \frac{\Delta \beta}{\bar{\beta}}\right) + 1/2 \frac{\Delta \beta}{\bar{\beta}} \sin^2 j + 1/2 \frac{\Delta \beta}{\bar{\beta}} \sin^2 j \tan^2 j, \quad (2) $$

or, equivalently

$$ R_{33}^{\text{iso}} = -1/2 \left(\frac{\Delta \rho}{\bar{\rho}} + \frac{\Delta \beta}{\bar{\beta}}\right) + 1/2 \frac{\Delta \beta}{\bar{\beta}} \tan^2 j, \quad (3) $$

with \( j \) denoting the incidence shear-wave angle. The disadvantage of this equation is that the approximation is not exact for normal incidence. Differences between vertical incidence reflection coefficients of different wave modes will later be attributed to (small) anisotropic parameters and any inaccuracy due to the isotropic approximaton can hamper the analysis and inversion for anisotropy coefficients. I therefore suggest to introduce the shear-wave impedance \( Z = \rho \beta \) and to use the exact normal-incidence reflection coefficient:

$$ R_{33}^{\text{iso}} = -1/2 \frac{\Delta Z}{Z} + 1/2 \frac{\Delta \beta}{\bar{\beta}} \tan^2 j, \quad (4) $$

From the physical point of view, it is clear that the normal-incidence reflection coefficient at isotropic boundaries is the same for all shear-modes because the polarization vector is always parallel to the reflecting interface. For oblique incidence angles, \( SV \) waves will generate reflected and transmitted \( P \)-waves in addition to shear waves and it is not surprising that the angular variation of the \( SV \) reflection coefficient differs from that in equation (4) (Thomsen, 1988; Thomsen, 1993):

$$ R_{33}^{\text{SV}} = -1/2 \frac{\Delta Z}{Z} + \left(\frac{\beta}{2} \frac{\Delta \beta}{\bar{\beta}} + 2 \frac{\Delta \rho}{\bar{\rho}}\right) \sin^2 j - 1/2 \frac{\Delta \beta}{\bar{\beta}} \sin^2 j \tan^2 j. \quad (5) $$
Figure 3. Propagation and reflection of shear waves traveling within the two vertical symmetry planes of an HTI medium ($V_p = 2.6\text{km/s}, V_s = 2\text{km/s}, \epsilon = 0.15, \delta = -0.1, \gamma = 0.1$). a.) and b.) show \(S^\parallel\) and \(S^\perp\) waves propagating in the isotropy plane, respectively; c.) and d.) denote \(S^\parallel\) and \(S^\perp\) waves traveling in the symmetry-axis plane. Shown in black are the seismic raypaths, the white curves denote the wavefronts at 1s traveltime.

It is well known that the coupling between transverse and longitudinal waves for oblique incidence angles manifests itself in the dependence of the \(P\)-wave AVO gradient on the shear-wave velocity (Shuey, 1985). Surprisingly, the opposite is not the case and the \(SV\)-wave AVO gradient is independent of \(\Delta\omega\), under the assumption of small relative changes in the elastic parameters across the reflecting boundary\(^*\).

Equation (5), as well as most reflection-coefficient approximations in this paper, show a simple decomposition of the form

\[
R(j) = A + B \sin^2 j + C \sin^2 j \tan^2 j.
\]  

(6)

The advantage of representing the reflection coefficient in this form has been thoroughly discussed in the literature (Castagna & Backus, 1993). Here it is appropriate to make some observations for the particular case of shear-wave amplitude studies. The normal-incidence term \(A\), for example, can hardly be obtained because of the difficulty to extract absolute amplitudes in data processing. However, as observed by Thomsen (1988) and Lynn and Thomsen (1990), comparison of \(A\) for both shear modes in models of HTI symmetry can yield valuable insight into medium parameters.

The AVO-gradient term \(B\) is of most interest in conventional \(P\)-wave AVO studies, as well as in shear-wave amplitude-variation-with-offset analysis. Note that this term dominates the angular variation in the reflection response for small angles of incidence (< 30°) recorded in most seismic surveys. Small incidence angles are particularly important for shear-wave studies because the first critical angle is often small. Even moderate incidence angles of about 30° can generate evanescent energy at the free surface that induces complicated phase and amplitude distortions. Thus, for shear waves, it is often difficult to obtain reliable estimates of the C-term in equation (6).

\(^*\) The approximate \(P\)-\(P\)-transmission coefficient is also independent of \(\Delta\omega\). Apparently, \(P\)-wave AVO is based on a rare physical gratitude.
The simplicity of the linearized equations helps to quickly identify the contributions of different physical parameters to the reflection response. For instance, we are reminded that only relative differences in elastic parameters influence the reflected seismic signature. For the particular case of the SH reflection coefficient, equation (3) shows that the relative differences in density and shear-wave velocity have the same influence on the normal-incidence reflection coefficient. Additionally, we recognize that $\frac{\Delta \rho}{\rho}$ is the only term responsible for small- and large-angle variations of the SH-wave reflection response. Thus, if amplitudes of SH-waves incident at large angles (> 20°) can be reliably recovered, it is advantageous to plot them as a function of the squared tangent of the incidence angle [see equation (3)].

**Vertical transverse isotropy**

The first step away from simple, isotropic media is to consider vertical transverse isotropy (VTI), the symmetry system that can be caused by massive shales and thinly-layered media. The three wave modes propagating in VTI media are the quasi P wave, the quasi SV wave and the truly transverse SH wave. Velocities in transversely isotropic media depend on the angle with the symmetry axis and every velocity parameter needs to be associated with a certain propagation direction. In the following, shear-wave velocity $\beta$ is the vertical shear-velocity which is identical for both shear modes in VTI media: the compressional velocity $\alpha$ in the equations below also correspond to vertical incidence. Seismic signatures in VTI media are invariant with respect to rotations about the vertical symmetry axis; however, while for VTI media the reflection response is still azimuthally invariant, the variation in velocity for different angles with vertical has an impact on the angular variation of the reflection coefficient.

Thomsen's (1986) anisotropy parameters can be used to describe the departure from isotropy. Using $\delta_1, \epsilon_1, \gamma_1$, and $\delta_2, \epsilon_2, \gamma_2$ to describe the anisotropy in the first and second layer, it is possible to derive linearized reflection coefficients (Thomsen, 1993; Rüger, 1996a) for VTI media by adding the appropriate anisotropic contributions to equations (4) and (5). The reflection coefficient for the SH mode, for example, can be written as

$$R_{SH}^{VTI} = R_{SH}^{iso} + 1/2 (\gamma_2 - \gamma_1) \tan^2 j,$$

where $j$ is the incident phase angle. As seen in equation (7), anisotropy contributes to the angular variation of the reflection coefficient. Ignoring the presence of anisotropy yields an erroneous AVO interpretation whenever there exists a difference between the anisotropy parameters $\gamma_1$ and $\gamma_2$. Finally, as for the isotropic AVO-contribution present in the expression for $R_{SH}^{iso}$, just one parameter ($\gamma$) is responsible for small and large-angle variations caused by the vertical transverse isotropy.

The SH-phase velocity in VTI media (Thomsen, 1986) displays a similar $\sin^2 j$-term multiplied by the anisotropy parameter $\gamma$ as the $R_{SH}^{iso}$ reflection coefficient. In the limit of weak anisotropy, the angular variation of the SV-wave phase velocity $\sigma$ is governed by

$$\sigma = \left( \frac{\beta}{\alpha} \right)^2 (\epsilon - \delta) \sin^2 j,$$

Unlike the isotropic SV-wave AVO gradient, equation (8) depends on the average compressional velocity. For a typical ratio of $\alpha/\beta \approx 2$, the difference in $(\epsilon - \delta)$ is of the same importance to the AVO gradient as the relative difference in density and vertical shear velocity [recall equation (5) for the dependence of $R_{SV}^{iso}$ on $\Delta \alpha$ and $\Delta \beta$]. For large angles of incidence and nonzero values of $\Delta \beta$, the $\sin^2 \tan^2$-term in $R_{SV}^{iso}$ will give an increasingly important contribution to $R_{SV}^{VTI}$. Note that there is no small-large-angle term in the anisotropic addition in equation (8).

As discussed in Tsventik and Thomsen (1994) and Tsventik (1995), the same $\left( \frac{\beta}{\alpha} \right)^2 (\epsilon - \delta)$-term dominates the radiation pattern of SV-waves in VTI media and needs to be accounted for in AVO analysis.

**Horizontal transverse isotropy**

The reflection coefficients for isotropic and VTI subsurface models are essentially known from previously published equations (Aki & Richards, 1980; Thomsen, 1993), although the specific form used in this paper was helpful in elucidating the physics of the energy-partitioning problem of shear waves. By contrast, azimuthally varying shear-wave AVO signatures for models of HTI symmetry have not yet been analyzed analytically and a few published studies have been based entirely on numerical experiments. Clearly, without an understanding of the relations between the medium parameters and the AVO response, it is hardly possible to make quantitative estimates of the influence of anisotropy on the recorded seismic signatures and, even more importantly, develop inversion procedures similar to conventional isotropic AVO analysis. Although the additional dependence on azimuth

---

$^\dagger$ As discussed above, this shear-wave and the compressional wave are not truly transverse and longitudinal in VTI and HTI media, but I omit the qualifier quasi for brevity.
Figure 4. Sketch of a shear-wave experiment in the presence of aligned near-vertical cracks (plane view). Arrows show the polarization directions for \( S^H \) and \( S^\perp \) waves. Amplitude analysis will be carried out along the strike \( (x_2) \) and the symmetry-axis direction \( (x_1) \).

makes the HTI reflection problem more involved, it can also yield more information about the anisotropy of the subsurface. Thus, it is reasonable to investigate the viability of extracting the subsurface parameters from the azimuthally-dependent shear-wave AVO response in HTI media.

Strike-line survey

The survey geometry in Figure 4 is similar to that of shear-wave experiments by Lynn and Thomsen (1990) and Lynn et al. (1993). As described in the previous chapter, the \( S^H \)-wave polarization is confined to the fracture plane (or isotropy plane) whereas the \( S^\perp \) wave is polarized perpendicular to it. Let us first consider an AVO experiment carried out along the fracture strike. It is hereby assumed that both layers are fractured in the same direction, but with different fracture intensity. If the source is oriented in the strike direction, a \( S^\perp \) shear wave polarized within the fracture plane is generated. Within this plane, the shear-wave velocity is independent of the propagation angle and the approximate reflection coefficient is identical to the isotropic reflection coefficient for shear waves polarized within the propagation plane:

\[
R_{S1}^{\text{strike}} = R_{S1}^{\text{iso}},
\]

with \( \alpha \) and \( \beta \) in \( R_{S1}^{\text{iso}} \) replaced by

\[
\alpha = \sqrt{\frac{C_{33}}{\rho}},
\]

\[
\beta = \sqrt{\frac{C_{44}}{\rho}},
\]

i.e., the compressional and shear velocities of waves polarized within the fracture plane.

For a survey aligned with fracture strike, the \( S^\perp \) wave is polarized perpendicular to the vertical propagation plane. Therefore, the corresponding approximate reflection coefficient has the form of equation (4):

\[
R_{S1}^{\text{strike}} = R_{S1}^{\text{iso}},
\]

where \( \beta \) in \( R_{S1}^{\text{iso}} \) needs to be replaced by the correct shear velocity \( \beta^\perp \). The latter can be expressed as:

\[
\beta^\perp = \sqrt{\frac{C_{44}}{\rho}}.
\]

Sample computations of exact and approximate reflection coefficients of shear waves propagating in the isotropy plane are shown in Figure 5. In this experiment, the upper medium is isotropic. The parameters of the lower medium are equivalent to the coefficients for a saturated, fractured model \((\gamma = 0.115)\) shown in Thomsen (1995), Table 1. In this experiment, a relative change of 20 percent in the vertical compressional velocity, vertical \( S^H \)-wave velocity and density across the reflecting boundary are assumed. The approximate \( S^H \) reflection coefficient is shown in Figure 5a, together with the exact solution. Small deviations from the exact solution can be observed only for incident angles close to 20°. Figure 5b shows the exact \( S^\perp \) reflection response and the approximations using equations (2) and (4). The former approximation is slightly shifted vertically with respect to the exact solution due to the approximate normal-incidence term. Equation (4), on the other hand, is exact for vertical incidence. To enable a better comparison between gradient and intercepts of the two AVO responses, both \( S^\perp \) and \( S^H \) reflection coefficients are combined in Figure 5c.

Differences in slope and normal incidence terms can be easily explained using the insight developed in the above discussion: The \( R_{S1}^{\text{strike}} \) normal-incidence term is more negative as compared to the \( S^\perp \) coefficient due to the larger shear-wave impedance contrast \((\rho\beta > \rho\beta^\perp)\). The \( R_{S1}^{\text{strike}} \) gradient, on the other hand, is influenced by two times the sum of relative changes in \( \beta \) and \( \rho \) as compared to \( 1/2 \alpha \beta^\perp \) in the \( R_{S1}^{\text{iso}} \) gradient.

What if anisotropy is ignored and both shear waves travel with the same velocities \( \beta \) in the lower medium? Certainly, both reflection coefficients are also different in the isotropic case, however, we would have missed the difference in normal-incidence reflection coefficient and the \( R_{S1}^{\text{strike}} \) gradient is overestimated because \( \Delta \frac{\alpha}{\beta} > \Delta \frac{\beta^\perp}{\beta} \).
Symmetry-line survey

Let us now consider a survey line acquired along the axis of symmetry (the $x_1$-axis in Figure 4). As discussed in the last section, wave propagation in the symmetry-axis plane is more complicated than in the fracture plane. Fortunately, the derivations necessary for this study are rather straightforward if one applies the analogy between elastic wave propagation in the $[x_1, x_3]$ plane of HTI and VTI models, first recognized by Rüger (1996a) and Tsvankin (1996c).

Propagation of body waves in elastic media is governed by the Christoffel matrix, here shown in the $c_{ijkl}$-tensor notation with slowness components $p_i$:

$$\det [c_{ijkl} p_i p_j - \rho] = 0 \quad (13)$$

Using the Voigt recipe to compact the indices of the fourth-order tensor $c_{ijkl}$ (Musgrave, 1970), equation (13) applied to wave propagation in the $[x_1, x_3]$ plane of HTI media with the symmetry axis pointing in the $x_1$ direction takes the form

$$0 = \begin{vmatrix} c_{11} p_1^2 + c_{44} p_3^2 - \rho \\ c_{11} p_1^2 + c_{55} p_3^2 - \rho \\ c_{13} + c_{55} p_1 p_3 \\ c_{13} + c_{55} p_1 p_3 \\ c_{13} + c_{55} p_3^2 - \rho \end{vmatrix} \times \begin{vmatrix} c_{11} & c_{44} & p_1 & p_3 \\ c_{55} & c_{55} & p_3 & p_3 \\ c_{13} & c_{55} & p_1 & p_3 \\ c_{13} & c_{55} & p_3 & p_3 \end{vmatrix} \quad (14)$$

Equation (14) coincides with the corresponding equation for the VTI medium with the symmetry axis pointing in the $x_3$ direction. However, it is crucial to understand that in HTI media $c_{55} = c_{44}$, whereas VTI symmetry implies that $c_{55}$ equals $c_{44}$.

The second term of equation (14) provides the solution for the velocities and the polarization direction of $P$- and shear waves ($S^+$-wave for HTI models and $SV$-wave for VTI media) with polarization vectors confined to the $[x_1, x_3]$-plane. Thus, all equations describing velocities, traveltime, polarization and stresses in the $[x_1, x_3]$-plane are identical for media with the symmetry axis pointing in either the $x_1$ (HTI) or the $x_3$ (VTI) direction.

In other words, for any HTI model there exists an “equivalent” VTI model that has the same kinematic properties and polarizations of $P$ and $S^+$ waves in the $[x_1, x_3]$-plane. Thus, $P$- and $S^+$-wave propagation in the $[x_1, x_3]$-plane of HTI media can be described by the known VTI equations using the elastic stiffness components $c_{ij}$, or, alternatively, using Thomsen’s parameters $\epsilon^{(V)}$ and $\delta^{(V)}$ defined for the equivalent VTI model (Rüger, 1996a; Tsvankin, 1996c):

$$\gamma^{(V)} \equiv \frac{c_{55} - c_{44}}{2c_{44}}$$

$$\epsilon^{(V)} \equiv \frac{c_{55} - c_{44}}{2c_{33}}$$

$$\delta^{(V)} \equiv \frac{(c_{13} + c_{55})^2 - (c_{33} - c_{35})^2}{2c_{33}(c_{33} - c_{55})} \quad (13)$$

Figure 5. Reflection coefficients of shear waves propagating in the isotropy plane. The black curves denote the exact solutions. Approximations are shown as dashed lines. The upper medium is isotropic, the lower medium has HTI symmetry with the shear-wave splitting parameter $\gamma = 0.115$; the vertical $P$ and $S^+$-wave velocities of the lower medium are 2.69 km/s and 1.4 km/s; its density is 2.07 g/cm$^3$. A relative change of 20 percent in the vertical compressional velocity, vertical $S^+$-wave velocity and density across the reflecting boundary are assumed. To better recognize differences between the $S^+$- and $S^-$ coefficients, the first two plots are combined and shown in the lower figure.
The coefficients $\gamma^{(V)}$ and $\delta^{(V)}$ can be expressed through the generic Thomsen parameters $\gamma$, $\epsilon$ and $\delta$ as defined with respect to the horizontal symmetry axis in the following way:

$$
\gamma^{(V)} = -\frac{\gamma}{1 + 2\gamma},
$$

$$
\epsilon^{(V)} = -\frac{\epsilon}{1 + 2\epsilon},
$$

$$
\delta^{(V)} = \frac{\delta - 2\epsilon (1 + \frac{\gamma}{\epsilon})}{(1 + 2\epsilon) (1 + \frac{\gamma}{\epsilon})},
$$

where

$$
f \equiv 1 - (\beta_0/\alpha_0)^2,
$$

as introduced in Tsvankin (1996b). Both $\alpha_0$ and $\beta_0$ are measured along the horizontal symmetry axis.

Recognizing this analogy between VTI and HTI models has broad implications and can be used to solve the reflection/transmission problem in the symmetry-axis plane. Note that the boundary conditions are identical for the VTI and the HTI model since in both cases, the incident waves are reflected from the $[x_1, x_2]$ plane resulting in the identical expressions for the tractions and displacements at the boundary. Consequently, the exact analytic solution for the VTI reflection problem (Graebner, 1992) is valid for reflection coefficients in HTI media, as long as the incident wave is confined to the symmetry-axis plane. Moreover, this solution is also valid for interfaces between VTI and HTI media, no matter whether the HTI medium is above or below the boundary.

This limited analogy between VTI and HTI media can be applied to obtain the approximate reflection coefficient $R_{S^{\perp}}^{\text{sym}}$ from equation (8):

$$
R_{S^{\perp}}^{\text{sym}} = R_{S^{\perp}}^{\text{iso}} + 1/2 \left( \frac{\delta}{\beta^{\perp}} \right)^2 \times
$$

$$
(\epsilon^{(V)} - \epsilon^{(V)} + \delta^{(V)} - \delta^{(V)}) \sin^2 j,
$$

with $\beta$ in $R_{S^{\perp}}^{\text{iso}}$ denoting $\beta^{\perp}$, the vertical velocity of the shear wave polarized in the symmetry-axis plane.

Very similar arguments allow to obtain the reflection coefficient for the $S^{\parallel}$ wave. The only additional information required is that in VTI media, $c_{44}$ equals $c_{55}$, whereas in HTI media, $c_{44}$ equals $c_{66}$. The approximate reflection coefficient then yields:

$$
R_{S^{\parallel}}^{\text{sym}} = R_{S^{\parallel}}^{\text{iso}} + 1/2 (\gamma^{(V)}^{(V)} - \gamma^{(V)}^{(V)}) \tan^2 j.
$$

Here, the $\beta$-term in $R_{S^{\parallel}}^{\text{iso}}$ corresponds to the vertical $S^{\parallel}$ wave-velocity. Figure 6 shows the symmetry-axis-plane reflection coefficients for the same model as in Figure 5. Again, the accuracy of the approximations is sufficiently high for both shear modes. The approximation for the $S^{\parallel}$ virtually coincides with the exact solution. Also note

**Figure 6.** Symmetry-axis plane reflection coefficients for both shear modes. Exact solutions are shown as black lines. The model is identical to the one used to generate the reflection coefficients in Figure 5. The anisotropy parameters of the lower medium are $\gamma^{(V)} = -0.093 (\gamma = 0.115)$, $\epsilon^{(V)} = -0.109$ and $\delta^{(V)} = -0.194$. The upper medium is purely isotropic.
the significant differences in the gradient and intercept of the reflection responses of the two individual shear modes. Ignoring anisotropy (and assuming that \( \beta \) is the shear velocity in the lower medium) would result in overestimated AVO gradients and a zero difference in normal-incidence reflection coefficients.

**Dependence on the shear-wave splitting parameter**

The application of the developed analytic results requires finding a set of parameters suitable for comparison of the individual shear-wave reflection coefficients. Obviously, for anisotropic media, \( \beta(= \beta^\bot) \neq \beta^\perp \). The parameter of most interest in reservoir characterization is the crack density that is proportional to the product of the number of cracks per unit volume and their mean cubed diameter. Since the crack density for penny-shaped cracks is close to the shear-wave splitting parameter \( \gamma \) (Tsvankin, 1996c), I will retain \( \gamma \) in the expressions for both shear waves. From the definition of anisotropy parameter \( \gamma \), it follows that

\[
\beta^\perp = \frac{\beta}{\sqrt{1 + 2\gamma}} \approx \beta (1 - \gamma),
\]

\[
\frac{\Delta \beta^\perp}{\beta^\perp} \approx \frac{\Delta \beta}{\beta} + \gamma_1 - \gamma_2,
\]  

(19)

accurate to first order in \( \gamma \). \( \frac{\Delta \beta}{\beta} \) will denote the relative change in fast shear-wave velocity for the remainder of this paper. To better understand the accuracy of this approximation, I compute \( \frac{\Delta \beta}{\beta} \) as a function of \( \gamma \) (the upper medium is isotropic) for two values of \( \frac{\Delta \beta}{\beta} \) (Figure 7). Shown are the exact solutions (solid lines) and the approximations based on equation (19). For increasing \( \gamma \), the contrast in \( \beta^\perp \) decreases because the velocity for shear waves polarized perpendicular to the fractures decreases. The approximations show little deviation from the exact result. However, even this small deviation can have an observable influence on the reflection coefficient. Note that the reflection response is formed by small differences in the elastic constants across a boundary and that inaccuracies of the size observed in Figure 7 can be significant.

Using the relation between the vertical shear velocities shown in equation (19), both reflection coefficients for shear waves traveling in the fracture plane [equations (17) and (18)] can be conveniently expressed as follows:

\[
R^{\text{strike}}_S = -1/2 \frac{\Delta Z}{Z} \sin^2 j \tan^2 j
\]

\[+ \left( \frac{7}{2} \frac{\Delta \beta}{\beta} + 2 \frac{\Delta \rho}{\rho} \right) \sin^2 j \]

\[+ 1/2 \left( \frac{\Delta \beta}{\beta} + \gamma_1 - \gamma_2 \right) \tan^2 j. \]  

(20)

The difference between \( R^{\text{strike}}_S \) and \( R^{\text{sym}}_S \) at normal incidence is equal to half the difference in anisotropy across the reflecting interface (Thomsen, 1988). This information can be combined with an estimate of the AVO gradient extracted from the \( S^\perp \) data set to yield \( \frac{\Delta \beta}{\beta} \). The last parameter of interest, \( \frac{\Delta \rho}{\rho} \), can be calculated from the \( S^\parallel \) AVO gradient.

Shear-wave AVO analysis along the strike direction has the potential of uniquely determining the difference in the elastic parameters across a reflecting boundary. In a noisy environment, it may be desirable to further constrain the inversion by adding additional information. A similar analysis, for example, is possible with the higher-angle \( \sin^2 j \tan^2 j \) term, but it is unlikely that this coefficient can be reliably extracted from field data; polarization and amplitude distortions for postcritically incident shear waves can prohibit accurate AVO analysis. Another option is AVO analysis of shear waves propagating normal to the fractures. Clearly, for vertical propagation, there is no distinction between wave propagation in the symmetry-axis plane or the fracture plane. However, the AVO gradients in the symmetry-axis plane are different from those in the fracture plane and these variations can be helpful for an inversion of medium parameters. Taking into account that for weak anisotropy \( \gamma(\nu) \approx -\gamma \) [equation (16)], we find

\[
R^{\text{sym}}_S = -1/2 \frac{\Delta Z}{Z}
\]
\[ R_{S^2} = -\frac{1}{2} \left( \frac{\Delta \beta}{\beta} + \gamma_1 - \gamma_2 \right) \tan^2 j + \frac{1}{2} \left\{ \frac{1}{2} \left( \frac{\Delta \beta}{\beta} + \gamma_1 - \gamma_2 \right) + 2 \Delta \delta \right\} \]

\[ \frac{1}{2} \left( \frac{\Delta \beta}{\beta} + \gamma_1 - \gamma_2 \right) \left( \epsilon_2^{(V)} - \epsilon_1^{(V)} + \delta_1^{(V)} - \delta_2^{(V)} \right) \sin^2 j - \frac{1}{2} \left( \frac{\Delta \beta}{\beta} + \gamma_1 - \gamma_2 \right) \sin^2 j \tan^2 j \]

If the difference \( \gamma_1 - \gamma_2 \) is extracted from the normal-incidence reflection coefficients, the AVO gradient of the \( S^1 \)-wave then provides an additional measure of \( \Delta \beta \). Unlike the \( S^1 \) wave that solely depends on \( \Delta \beta \) and \( \gamma \), the \( S^2 \)-wave is additionally influenced by the difference of coefficients \( \delta^{(V)} \) and \( \epsilon^{(V)} \), as demonstrated by Tsankin (1996c), is close to anisotropy coefficient \( \eta \) (Alkhalifah & Tsankin, 1995) needed for time processing in transversely isotropic media.

Certainly, the additional linearizations used to derive equations (20) and (21) will change the accuracy of the approximation. To study the quality of the approximation and to observe the variations of the reflection coefficients from one symmetry plane to the other, Figure 8 shows the \( S^1 \) and \( S^2 \)-reflection coefficients for the same model as used for Figures 5 and 6. As discussed above, the approximation for the \( S^2 \) reflection coefficients slightly deviates from the exact solution at normal incidence due to the inaccurate value of the relative change in \( \Delta \gamma \).

The reflection coefficient curves look rather similar for both \( S^1 \) and \( S^2 \) waves in Figure 8, an observation consistent with equations (20) and (21). Specifically, in the weak anisotropy limit, both \( R_{S^1} \) and \( R_{S^2} \) have the same AVO gradient. The difference in AVO gradient between \( R_{S^1} \) and \( R_{S^2} \), on the other hand, is a function of \( \Delta \gamma \) and \( \Delta \sigma \).

Reflection coefficients computed for 4 different combinations of isotropic/HTI interfaces are shown in Figure 9, with model parameters given in Table 1. The first model has only one nonzero anisotropy parameter \( \delta_1^{(V)} \). Equations (20) and (21) correctly predict an identical, moderately negative AVO gradient for \( R_{S^1} \) and \( R_{S^2} \), and, after some algebra, a more negative \( R_{S^1} \) gradient as compared to \( R_{S^2} \). The accuracy of the approximations is sufficiently high, as it is the case for the other examples. Model b has a zero value of \( \epsilon_1^{(V)} \), typical for tight formations such as coal layers. Although the \( S^1 \) reflection curves are vertically shifted due to the different normal-incidence reflection coefficients, it is still possible to observe very similar \( R_{S^1} \) and \( R_{S^2} \) gradients. The \( R_{S^2} \) gradient, on the other hand, is more negative than the \( R_{S^1} \) gradient. The same observation can be made for Model c, with nonzero anisotropy parameter in the lower medium. Finally, a model of elliptical anisotropy \( (d_2^{(V)} = \epsilon_2^2) \) is used in example d.

Note that approximations (20) and (21) use the \( S^1 \)-wave vertical velocity as parameter, hence the difference in accuracy between the \( S^1 \) and \( S^2 \) coefficients is not related to the physics, but is a matter of design of the approximations. It is straightforward to rewrite equations (20) and (21) as functions of \( \Delta \beta \) if this parameter is better known as compared to \( \Delta \beta \).

The examples shown in Figure 9 demonstrate that the derived approximations are accurate for positive and negative jumps in relative difference of the isotropic parameters across the reflecting boundary and nonzero values of anisotropy parameters. Moreover, rather than just studying the AVO gradients for each wavetype independently, equations (20) and (21) motivate to additionally explore the differences in AVO gradients for waves.
Propagating in the vertical symmetry planes of HTI media. Keeping just the lowest-order term and denoting \( \Delta \sigma^{(V)} = \left( \frac{2}{3} \right)^2 (\Delta \varepsilon^{(V)} - \Delta \sigma^{(V)}) \), we find the following differences in the reflection responses:

\[
R_{S_2}^{\text{sym}} - R_{S_2}^{\text{strike}} = -1/2 \Delta \gamma \\
R_{S_2}^{\text{sym}} - R_{S_2}^{\text{strike}} = -1/2 \Delta \gamma + \left( 3 \frac{\Delta \sigma^3}{\beta} + 2 \frac{\Delta \rho}{\rho} + 1/2 \Delta \gamma \right) \sin^2 j \\
R_{S_2}^{\text{sym}} - R_{S_2}^{\text{strike}} = 1/2 \Delta \gamma + \left( -7/2 \Delta \gamma + 1/2 \Delta \sigma^{(V)} \right) \sin^2 j \\
R_{S_2}^{\text{sym}} - R_{S_2}^{\text{strike}} = - \left( 3 \frac{\Delta \sigma^3}{\beta} + 2 \frac{\Delta \rho}{\rho} + 1/2 \Delta \gamma \right) \sin^2 j \\
R_{S_2}^{\text{sym}} - R_{S_2}^{\text{strike}} = \left( 3 \frac{\Delta \sigma^3}{\beta} - \Delta \gamma \right) + 2 \frac{\Delta \rho}{\rho} + 1/2 \Delta \sigma^{(V)} \sin^2 j \\
R_{S_2}^{\text{sym}} - R_{S_2}^{\text{strike}} = -1/2 \Delta \gamma - \left( 3 \frac{\Delta \sigma^3}{\beta} - \Delta \gamma \right) + 2 \frac{\Delta \rho}{\rho} + 1/2 \Delta \sigma^{(V)} \sin^2 j \quad (22)
\]

Four different configurations allow to obtain the shear-wave splitting parameter from the difference in normal incidence reflection coefficient. Additionally, 5 equations for the differences in the AVO gradients can be solved for the 4 unknowns \( \frac{\Delta \sigma^3}{\beta} \), \( \frac{\Delta \sigma^3}{\beta} \), \( \Delta \gamma \) and \( \Delta \sigma^{(V)} \).

Off-symmetry-plane polarization

The reflection coefficients discussed so far correspond to the two vertical symmetry planes of HTI media. In principle, it is possible to derive off-symmetry reflection responses for the individual wave types in the same manner as outlined by (Thomsen, 1993) and (Rüger, 1996a), but it is doubtful if the resulting expressions would be usable in practice. Interpretation of shear-wave data for waves with polarizations vectors outside the two vertical symmetry planes is very complicated even in the case of a reflection from an isotropic/isotropic interface: shear waves with nonzero polarization components perpendicular and parallel to the incidence plane will get rotated upon reflection due to the different reflection coefficients for the two components.

A simple isotropic layer on top of a reflecting azimuthally anisotropic medium will cause a rotation of the polarization vector even if the shear-wave polarization is aligned or perpendicular to the incidence plane. Consider the shear-wave displacement vector shown in plane view in Figure 10. This vector can be decomposed in two components parallel and perpendicular to fracture strike. Upon incidence on the reflecting HTI boundary, both components are scaled by individual reflection coefficients and SV/SH conversions are generated. Although the reflected shear wave is not split in a fast and slow component, the difference in reflection coefficient will effectively yield a rotation of the polarization vector. The resulting amplitude will be a composite of the reflection response for the parallel and perpendicular components. Although this amplitude and the rotation angle contain information about the anisotropy in the lower layer, this information is hard to extract and thus severely hampers any reliable AVO study.

Propagation in anisotropic overburden

The goal of AVO analysis is to investigate the angular variation of reflection coefficients rather than directly studying the change of amplitudes with offset. As mentioned above, a substantial amplitude distortion may be associated with wave propagation in anisotropic layers above the reflector. To obtain the reflection coefficients at the target horizon, it is necessary to correct for the amplitude variation associated with the wave propagation between the reflector and the surface. Specifically, the presence of anisotropy in the overlying medium leads to wavefront focusing phenomena, i.e., distortions of the amplitude distribution along the wavefronts of the incident and reflected waves (Tsankin, 1995). One example herefore can be seen in Figure 3d, where the areas of high

<table>
<thead>
<tr>
<th>Model</th>
<th>Model a</th>
<th>Model b</th>
<th>Model c</th>
<th>Model d</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \frac{\Delta \sigma^3}{\beta} )</td>
<td>-0.1</td>
<td>0.1</td>
<td>-0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>( \frac{\Delta \sigma^3}{\beta} )</td>
<td>-0.15</td>
<td>-0.15</td>
<td>-0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>( \frac{\Delta \sigma^3}{\beta} )</td>
<td>-0.15</td>
<td>-0.15</td>
<td>-0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>( c_2^{(V)} )</td>
<td>0</td>
<td>0</td>
<td>-0.11</td>
<td>-0.11</td>
</tr>
<tr>
<td>( c_2^{(V)} )</td>
<td>-0.15</td>
<td>-0.057</td>
<td>0.18</td>
<td>-0.11</td>
</tr>
<tr>
<td>( \gamma )</td>
<td>0</td>
<td>0.15</td>
<td>0.15</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Table 1. Model parameters for the reflection-coefficient experiments shown in Figure 9. The vertical P and S-velocity velocities of the lower medium are 2.69 km/s and 1.4 km/s; its density is 2.07 g/cm³. The upper medium is purely isotropic.
The presence of anisotropy in the overburden may be quite typical for fractured reservoirs. Examples of a VTI-HTI boundary include shale-sandstone interfaces such as those discussed by Haugen (1996); also, it is quite common that fracturing extends throughout the section, making the overburden azimuthally anisotropic.

$P$- and $S$-wave radiation patterns from point sources in VTI media derived in Tsvankin (1993) show that anisotropy has a comparable influence on both energy focusing and the reflection coefficient. The angular amp-
Constraining the inversion for anisotropy parameters

Analysis of AVO intercepts and gradients of the individual shear waves can be combined with a study of the differences in the reflection coefficients in the symmetry planes as indicated in equations (22). Certainly, for high-quality surveys these data are sufficient to solve for the change in isotropic medium parameters as well as for anisotropy parameters γ and σ in the subsurface. In more realistic situations, amplitude extraction from shear-wave data is a difficult task and it is useful to combine all additionally available data for a robust inversion of the medium parameters.

The conventional method of determining γ from surface data is based on analyzing the differences in traveltimes between vertically traveling $S^\perp$ and $S^\parallel$ waves (Crambin, 1985) and comparison of normal-incidence reflection coefficients. More recently, Rüger and Tsvankin (1995) described an algorithm to obtain estimates of γ and δ(γ) from P-wave AVO analysis.

An alternative approach to constrain the inversion is to use the azimuthal variations in the normal-moveout velocities in HTI media. Normal-moveout velocities from horizontal reflectors in HTI media have been derived by Tsvankin (1996c). Here, I show the linearized expressions for shear-wave NMO velocities in the vertical symmetry planes:

$$v_{\alpha_{\text{str}}}(P - \text{wave}) = \alpha$$

$$v_{\alpha_{\text{str}}}(S^\perp - \text{wave}) = \beta^\perp$$

$$v_{\alpha_{\text{str}}}(S^\parallel - \text{wave}) = \beta,$$

and

$$v_{\alpha_{\text{sym}}}(P - \text{wave}) \approx \alpha (1 + \delta(\gamma))$$

$$v_{\alpha_{\text{sym}}}(S^\perp - \text{wave}) \approx \beta^\perp (1 + \sigma(\gamma))$$

$$v_{\alpha_{\text{sym}}}(S^\parallel - \text{wave}) \approx \beta (1 - \gamma).$$

Equations (25) and (26) indicate that the azimuthal variation of NMO velocity can be inverted for anisotropy parameters. For example, a simple method to estimate γ is to study the $S^\parallel$-wave NMO velocity perpendicular and parallel to fracture strike (Tsvankin, 1996c). The difference in NMO velocity is close to the shear-wave splitting parameter γ. Unlike the purely hyperbolic moveout of $S^\parallel$-wave reflections from horizontal interfaces, the moveout of P and S waves deviates from a hyperbola even for a single horizontal layer in homogeneous HTI media. However, the deviations in the symmetry-axis plane are small for most realistic values of anisotropy and conventional spread lengths (Tsvankin & Thomsen, 1994).

If the medium is stratified and composed of transversely isotropic layers, the azimuthal NMO analysis is still possible in through-going symmetry-planes using the generalized Dix equation presented by Alkhalifah and Tsvankin (1995). Thus, even if the vertical resolution is smaller for the study of kinematic signatures, it is cer-
tainly beneficial to combine moveout and AVO analysis to obtain a more stable inversion.

The most likely application of the developed methodology is in seismic investigations of fractured reservoirs. The derived equations, however, are valid for HTI media of any origin. If the symmetry is created by thin layering or thin vertical cracks, only two of the three anisotropy parameters are independent, further simplifying the inversion.

Orthorhombic media

HTI media are useful models to study the first-order influence of azimuthal anisotropy. For cracked media, any deviation from vertical penny-shaped cracks embedded in an isotropic matrix requires the investigation of models with lower symmetry. Media of orthorhombic symmetry, for example, are believed to describe more realistic models of fractured reservoirs such as layers with swarms of elongated cracks. The orthorhombic symmetry system includes azimuthally anisotropic strata caused by a combination of horizontal layering and vertically aligned cracks. Orthorhombic symmetry can also be created by a system of two orthogonal but not necessary identical crack systems or two indistinguishable crack systems at oblique angle (Winterstein, 1990).

The derivations of HTI and orthorhombic media reflection coefficients are very similar. Analysis of symmetry-plane Christoffel systems yields the scattering coefficients and naturally leads to the introduction of a new effective parameterization (Tsukanov, 1996a). For completeness, I state the [z1, x3]-symmetry-plane reflection coefficients for the shear wave polarized (for vertical incidence and symmetry-plane propagation) in the [x1, x3]-plane (S'), and in the [x1, x2]-plane (S):

\[
R_{S_{1}}^{[x_{1}, x_{3}]} = R_{S_{H}}^{\text{iso}} + \frac{1}{2} (\gamma_{3}^{(2)} - \gamma_{1}^{(2)}) \tan^2 j
\]

\[
R_{S_{z}}^{[x_{1}, x_{3}]} = R_{S_{V}}^{\text{iso}} + \frac{1}{2} \left( \frac{\alpha}{\beta} \right)^2 \times (\epsilon_{3}^{(2)} - \epsilon_{1}^{(2)} + \delta_{3}^{(2)} - \delta_{1}^{(2)}) \sin^2 j, \tag{27}
\]

with shear-wave velocities \( \beta = \sqrt{c_{33}/\rho} \) in \( R_{S_{1}}^{[x_{1}, x_{3}]} \) and \( \beta' = \sqrt{c_{33}/\rho} \) in \( R_{S_{z}}^{[x_{1}, x_{3}]} \). Parameters \( \delta^{(2)} \) and \( \epsilon^{(2)} \) are responsible for near-vertical and near-horizontal P-wave propagation, respectively, exactly as \( \delta \) and \( \epsilon \) in VTI models. \( \gamma^{(2)} \) determines the velocity variation of the \( S' \) wave polarized normal to the [x1, x3]-plane.

For shear waves propagating in the [x2, x3]-plane, the

\[
\delta^{(2)} = \frac{(c_{44} - c_{55})^2}{2c_{23}(c_{44} - c_{55})} \delta \quad \delta^{(V)}
\]

\[
\delta^{(1)} = \frac{(c_{44} - c_{55})^2}{2c_{23}(c_{44} - c_{55})} \delta \quad 0
\]

\[
\epsilon^{(2)} = \frac{c_{44} - c_{55}}{2c_{23}} \epsilon \quad \epsilon^{(V)}
\]

\[
\epsilon^{(1)} = \frac{c_{44} - c_{55}}{2c_{23}} \epsilon \quad 0
\]

\[
\gamma^{(2)} = \frac{c_{44} - c_{55}}{2c_{44}} \gamma \quad \gamma^{(V)}
\]

\[
\gamma^{(1)} = \frac{c_{44} - c_{55}}{2c_{44}} \gamma \quad 0
\]

Table 2. Anisotropy parameters describing orthorhombic anisotropy. Their relations to generic (VTI) Thomsen coefficients and to HTI parameters \( \delta^{(V)}, \gamma^{(V)} \) and \( \epsilon^{(V)} \) are shown in the second and third column.

solutions for the reflection coefficients are as follows:

\[
R_{S_{1}}^{[x_{2}, x_{3}]} = R_{S_{H}}^{\text{iso}} + 1/2 \left( \frac{\alpha}{\beta} \right)^2 \times (\epsilon_{2}^{(1)} - \epsilon_{1}^{(1)} + \delta_{2}^{(1)} - \delta_{1}^{(1)}) \sin^2 j.
\]

\[
R_{S_{z}}^{[x_{2}, x_{3}]} = R_{S_{V}}^{\text{iso}} + 1/2 (\gamma_{2}^{(1)} - \gamma_{1}^{(1)}) \tan^2 j. \tag{28}
\]

As before, the \( S' \)-wave coefficients are shown with velocity \( \beta \) whereas the \( S_{z} \)-wave vertical shear-velocity is \( \beta' \). A definition of the new effective coefficients and their relations to the generic Thomsen parameters (VTI) and HTI parameters is given in Table 2. Note that the conventionally studied shear-wave splitting parameter \( \gamma^{(e)} = \frac{c_{44} - c_{55}}{2c_{44}} = \frac{\epsilon^{(1)} - \delta^{(2)}}{1 + 2\gamma^{(2)}} \) (Tsukanov, 1996a).

Conclusions

If the shear-wave AVO gradients can be reliably extracted from prestack shear-wave amplitudes, they can supplement the fracture-detection algorithm based on the normal-incidence shear-wave reflection coefficients (Thomsen, 1988; Mueller, 1991).

This paper elucidates the physics of energy-partitioning of shear waves in azimuthally anisotropic media. For the two vertical symmetry planes in HTI symmetry systems, concise analytic expressions of the AVO gradients predict that differences in the shear-wave splitting parameter and anisotropy parameters \( \delta^{(V)} \) across a reflecting boundary have a first-order influence on shear-wave AVO gradients. AVO gradients for \( S' \) waves traveling in the symmetry-axis plane and \( S_{z} \) waves propagating along fracture strike are shown to be approximately identical while difference in \( S' \)-strike and \( S_{z} \)-symmetry axis AVO gradients depends on a weighted sum of \( \gamma \).
and $\sigma^{(V)}$. Rather than just studying the AVO gradients for each wavetype independently, this analysis suggests to additionally explore the differences in AVO gradients for waves propagating in the vertical symmetry planes of HTI media.

Obviously, amplitude analysis of shear-wave data is a difficult task in realistic settings. To stabilize the inversion algorithm based on shear-wave AVO analysis, I suggest combining the amplitude analysis with a study of the azimuthal variations in normal-moveout velocities. Also, since the magnitude of the splitting parameter is relatively small, even a substantial percentage error in $\gamma$ may prevent this algorithm from detecting anomalies in the crack density.

Probably the most restricting assumption in this study is that the symmetry-axis direction has to be known prior to the shear-wave acquisition. If the excited shear-wave polarization is oblique to the natural coordinate system of the subsurface, the reflected shearwave amplitudes and polarization will be distorted and severely hampers the AVO analysis. Fortunately, for fractured reservoirs, the crack orientation and thus the symmetry axis direction can often be obtained from stress measurements, geologic information and borehole data (e.g., tiltmeter, breakouts) and shear-wave polarization analysis.

Finally, as for $P$-wave studies, an essential part of the shear-wave AVO is the correction for energy focusing and defocusing for waves propagating in the anisotropic overburden.

Acknowledgments

Thanks to Ilya Tsvankin, John Stockwell and Konstantin Osypov for the review of this paper and to members of the A(nisotropy)-Team at the Center for Wave Phenomena (CWP) for useful discussions. The support for this work was provided by the members of the Consortium Project on Seismic Inverse Methods for Complex Structures at CWP, Colorado School of Mines, and by the United States Department of Energy (project "Velocity Analysis, Parameter Estimation, and Constraints on Lithology for Transversely Isotropic Sediments" within the framework of the Advanced Computational Technology Initiative). Partial funding for this project has been provided by a scholarship from Phillips Petroleum Co.

References


Castagna, & Backus, Eds. 1993. In Offset Dependent Reflectivity: Theory and methods. SEG.


Day, San Francisco.
Rüger, A. 1996b. Variation of P-wave reflectivity with offset and azimuth in anisotropic media. CWP-203 (This volume).
Thomsen, L. 1993. Weak anisotropic reflections. In Offset Dependent Reflectivity (Castagna and Backus, Eds.), SEG, Tulsa.
Lyapunov exponents and localization in randomly layered elastic media

John A. Scales
Department of Geophysics and Center for Wave Phenomena
Colorado School of Mines, Golden, CO 80401, USA

Erik S. Van Vleck
Department of Mathematical and Computer Sciences
Colorado School of Mines, Golden, CO 80401, USA

ABSTRACT
Wave propagation in a randomly heterogeneous elastic medium is often modeled as a random dynamical system associated with a lattice of coupled springs and masses. The scattering effects of the heterogeneities on a propagating pulse are characterized by the frequency dependent localization length—effectively the "skin depth" for multiple scattering attenuation. On a one-dimensional (1D) lattice of length \( N \), the localization length is the reciprocal of the positive Lyapunov exponent of an \( N \)-fold product of \( 2 \times 2 \) random matrices. In this case, all propagating solutions are either exponentially growing or decaying. In higher dimensions the situation is more complicated, but assuming quasi-1D propagation, the localization length can still be characterized in terms of the Lyapunov spectrum of products of random matrices. We describe a robust numerical procedure for estimating these exponents for 1D or quasi-1D propagation in randomly layered media. In addition, we provide uncertainty estimates for the exponents.

Key words: Localization, Lyapunov Exponents, Random Media

Background: Wave Propagation in Disordered Media and Localization

Broadly speaking, the effects of multiple scattering on a wave propagating in a disordered medium are:

- The path length is increased.
- The pulse is dispersed.
- The pulse is attenuated as energy is shifted from the direct arrival into the multiple scattering coda.

These ideas are well understood and can be explained via perturbation theory for weakly disordered media. (For an overview of the mathematical results see (Asch et al., 1991).) Randomly layered media are common in a number of applications including seismology, where the layering is due to the natural process of sedimentation, and non-destructive evaluation, where the layering could be the result of lamination of composite materials. As an example, consider the model shown in Figure 1. Here we have chosen a pseudo-random sequence (to be described later) of layer thicknesses and spring constants (i.e., elastic stiffnesses). We can scale these to reflect typical values for different applications. For example, if we take the minimum layer thickness to be 1 m, assume a constant density, and scale the spring constants to correspond to acoustic speeds in the range of 2400-4000 m/s, we can model the lithology of the upper crust of the earth. Figure 2 shows a 2D finite difference wave propagation simulation for such an earth model; at a given depth, the plot shows the energy recorded at that depth as a function of time. The first energy seen at any given depth is the primary downgoing pulse. Coherent events sloping downward to the right (such as the primary arrival) correspond to downgoing waves, while events sloping downward to the left correspond to upgoing waves. In a simulation such as this it is possible to follow the envelope of the pulse as it propagates into the medium. Figure 3 shows a log-linear plot of the amplitude of this pulse. Since the medium is perfectly elastic, the
Figure 1. A model of a randomly laminated medium obtained by choosing pseudo-random sequences of spring constants and layer thicknesses. These spring constants and layer thicknesses can then be scaled to reflect typical elastic wave speeds for a given application.

Figure 2. We take the model shown in Figure 1 and scale the spring constants and layer thicknesses to reflect values typical of the earth's near surface. Here is shown a 2D acoustic finite-difference simulation of a pulse propagating into such a medium. The pulse is generated by applying a band-limited point source (with a central frequency of 60 Hz) just below the upper surface. This figure shows the acoustic response at each layer as a function of time. The first event seen at each depth is the direct wave, with all the other events being upgoing or downgoing multiply scattered energy.

where $T$ is the transmission coefficient and $L$ is the propagation distance. An approximation to the localization length for finite $L$ is

\[ \ell(f) \approx \frac{-L}{\ln |T|}. \]  

The frequency dependent localization length would be an extremely useful quantity to have in many applications involving wave propagation in highly heterogeneous media since it provides the means to scale the effects of complex microstructure. For example, given the exponential nature of the decay, it is natural to think of the multiple scattering as a relaxation mechanism. Hence there is a $Q$ (quality factor) equal to $\pi f \ell / v(f)$, where $v$ is the phase velocity (Futterman, 1962). Then from Kramers-Krönig, the velocity dispersion is

\[ \frac{1}{v(f)} - \frac{1}{v(\infty)} = \mathcal{H} \left( \frac{1}{2\pi f \ell} \right). \]
Lyapunov exponents and localization

Equations of Motion

Let us begin with the simplest case, waves propagating in a heterogeneous 1D medium. Later, we will show how to generalize these results to higher dimensional systems. The 1D wave equation is

$$\rho(z) \frac{\partial^2 U}{\partial t^2} = \frac{\partial}{\partial z} \left( K(z) \frac{\partial U}{\partial z} \right)$$

where $U$ is the displacement, $\rho$ the density and $K$ the elastic modulus. To obtain a discrete approximation to the wave equation we could either apply a finite difference approximation to the spatial derivatives in Equation 5, or replace the continuum with a discrete spring/mass system and appeal to Newton's second law of motion.

The discrete approach has the advantage that the problem can be characterized in terms of the eigenmodes of the lattice. The localization itself is seen to arise from the effects of individual heterogeneities (lattice defects), which give rise to eigenmodes whose envelopes decay exponentially away from the defect sites. Further, the analysis of localization can be studied via the asymptotic properties of products of (small) random matrices. We will see, for example, that the localization length $\ell$ can be identified with the reciprocal of the smallest positive Lyapunov exponent of this product of random matrices.

In 1D this is straightforward since there is only one positive Lyapunov exponent. In the more general quasi-1D case this exponent will control the dominant features of decay, but the other positive exponents could play a role in understanding the fine details.

There are $N$ point masses $m_\ell$ connected by $N - 1$ Hooke's law springs of stiffness $k_\ell$. The end masses are fixed, as shown in Figure 4 and the longitudinal displacement of the $\ell$-th mass $x_\ell$ is measured relative to its equilibrium position.

The kinetic energy of the system is $\frac{1}{2} \sum_{\ell=0}^{N} m_\ell \dot{x}_\ell^2$ and the potential energy (assuming only nearest neighbor interactions) is $\frac{1}{2} \sum_{\ell=0}^{N} k_{\ell-1} (x_{\ell-1} - x_\ell)^2$. The lattice equations of motion are

$$m_\ell \ddot{x}_\ell - k_{\ell-1} (x_{\ell-1} - x_\ell) - k_\ell (x_{\ell-1} - x_\ell) = 0.$$  

Assuming a sinusoidal time dependence, the equation for the spatial part of the motion is
\[ -m_i \omega^2 z_i = k_{i+1} z_{i+1} - (k_{i+1} + k_i) z_i + k_i z_{i-1} \]  
where \( z_i \equiv z_i e^{i \omega t} \), or

\[ (T + \omega^2 M) z = 0 \]  
where \( M \) is the mass matrix \( M = \text{diag}(m_i) \) and \( T \) is the tridiagonal matrix of spring constants: \( T(\ell, \ell) = -(k_{\ell-1} + k_\ell) \), \( T(\ell, \ell+1) = k_{\ell-1} \), and \( T(\ell, \ell+1) = k_\ell \).

**Free oscillations of the 1D lattice**

For the ordered monoatomic 1D lattice: \( m_i = m \), \( k_i = k \) for all \( \ell \), so the equation for the free oscillations reduces to

\[ (T + \omega^2 I) z = 0 \]  
Since the spring constants are all equal to \( k \), \( T \) reduces to \( k \) times

\[
\begin{pmatrix}
-2 & 1 & 0 & 0 & \ldots \\
1 & -2 & 1 & 0 & \ldots \\
0 & 1 & -2 & 1 & \ldots \\
\vdots & & & \ddots & \vdots \\
0 & \ldots & 0 & 1 & -2
\end{pmatrix}
\]  
(10)

The eigenvalues of this matrix are

\[ \omega_i^2 = -2 + 2 \cos \left( \frac{\ell \pi}{N} \right) \]  
while the eigenvectors are

\[ z_i = \{ \sin(\ell \pi/N), \sin(2\ell \pi/N), \ldots, \sin((N-1)\ell \pi/N) \} \]  
(12)

Another way of looking at the free oscillations of the lattice is to recast the eigenvalue problem in Equation 7 as a recursion relation:

\[ z_{i+1} = \left( 2 - \frac{m_\omega^2}{k} \right) z_i - z_{i-1} \]  
(13)

which can be re-written as a one-step mapping

\[
\begin{pmatrix}
z_i \\
z_{i-1}
\end{pmatrix}
= \begin{pmatrix}
0 & 1 \\
-1 & 2 - \frac{m_\omega^2}{k}
\end{pmatrix}
\begin{pmatrix}
z_{i-1} \\
z_i
\end{pmatrix}.
\]  
(14)

Defining the matrix

\[ B = \begin{pmatrix}
0 & 1 \\
-1 & 2a
\end{pmatrix} \]  
(15)

where \( a = 1 - \frac{m_\omega^2}{k} \), it follows by induction that

\[
\begin{pmatrix}
z_i \\
z_{i-1}
\end{pmatrix} = B^\ell \begin{pmatrix}
z_0 \\
z_1
\end{pmatrix}
\]  
(16)

where \( B^\ell \) is the \( \ell \)-th power of the matrix \( B \).

The eigenvalues of \( B \) are \( \lambda_{1,2} = a \pm \sqrt{a^2 - 1} \). If \( |a| > 1 \) (\( \omega^2 > 4k/m \)) then the \( z_i \) solutions are exponentially growing or decaying and cannot satisfy the zero displacement boundary conditions. If \( |a| < 1 \) (\( \omega^2 < 4k/m \)), then the solutions are oscillatory and can be made to satisfy the boundary conditions. Taking \( z_0 = 0 \), then

\[
\begin{pmatrix}
z_{N-1} \\
z_N
\end{pmatrix} = \begin{pmatrix}
0 & 1 \\
-1 & 2a
\end{pmatrix}^N \begin{pmatrix}
0 \\
1
\end{pmatrix}
\]  
(17)

Obviously this equation cannot produce an eigenvector for arbitrary values of \( a \). For example, if we are trying to compute normal modes whose displacement vanishes at both ends of the lattice, then only using \( a \) associated with an eigenvalue will result in \( z_N \) being zero.

**Localized States on Lattices**

For the homogeneous lattice all the eigenmodes are nonzero everywhere except at the nodal points. In that sense they have global extent. Suppose we disturb the regularity of the lattice by perturbing either a spring constant or a mass. Rayleigh's principle ([Rayleigh, 1945], Volume I, Sec. 88) says that if a single mass is reduced (or a spring constant increased) then all the frequencies are unchanged or increased, but not by more than the distance to the nearest unperturbed frequency. Similarly, if a single mass is increased (or a spring constant decreased) then all the frequencies are unchanged or decreased, but not by more than the distance to the nearest unperturbed frequency.

In spite of this apparent symmetry between the effects of increasing versus decreasing masses (or decreasing versus increasing spring constants), there is a profound difference in practice. Because the low frequency limit is zero (at least for the monoatomic chain), decreasing a spring constant, and hence one or more frequencies, cannot result in new frequencies outside the band of frequencies allowed by the dispersion relation for the homogeneous or perfect lattice. On the other hand, increasing a spring constant can result in perturbing an eigenfrequency beyond the maximum allowed by the homo-
Lyapunov exponents and localization 389

Figure 5. Perturb a single site by increasing its spring constant (or decreasing its mass). Rayleigh’s principle says that this will result in increasing (or leaving unchanged) the frequencies. Increasing a single spring constant sufficiently results in a single eigenvalue being pushed out of the band of frequencies allowed by the homogeneous dispersion relation. This must therefore be associated with a localized mode. Perturbing two lattice sites in this case pushes two frequencies outside the allowed band, resulting in two localized modes.

Products of Random Matrices

The first systematic study of the randomly disordered 1D chain was made by Freeman Dyson (1952). Dyson developed a method for calculating the distribution of eigenfrequencies in the $N \to \infty$ limit as a continued fraction. If the oscillators themselves are distributed according to an exponential law, Dyson’s result is analytic. This work was extended by Helmut Schmidt (1957) who developed the recursive method of computing the eigenfunctions.

The eigenvector recursion formula extends readily to arbitrarily disordered chains. In this case the equations of motion are

$$z_{t+1} = \frac{k_{\ell+1} + k_\ell - m_\ell \omega^2}{k_{\ell+1}} z_t - \frac{k_\ell}{k_{\ell+1}} z_{t-1},$$  \hspace{1cm} \text{(18)}$$

which we can rewrite as

$$\begin{pmatrix} z_{t+1} \\ z_{t-1} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -\frac{k_\ell}{k_{\ell+1}} & \frac{k_{\ell+1}}{k_{\ell+1}} \end{pmatrix} \begin{pmatrix} z_t \\ z_{t-1} \end{pmatrix},$$ \begin{pmatrix} z_{t+1} \\ z_{t-1} \end{pmatrix} = B_t \cdot B_{t-1} \cdots B_1 \left( \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \right) \equiv P_t \left( \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \right).$$ \begin{pmatrix} z_{t+1} \\ z_{t-1} \end{pmatrix} = B_t \cdot B_{t-1} \cdots B_1 \left( \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \right) \equiv P_t \left( \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \right).$$ \begin{pmatrix} z_{t+1} \\ z_{t-1} \end{pmatrix} = B_t \cdot B_{t-1} \cdots B_1 \left( \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \right) \equiv P_t \left( \begin{pmatrix} z_0 \\ z_1 \end{pmatrix} \right).$$

The physics of randomly disordered lattices began to be studied intensively in the 1940s and 1950s. A pioneering series of papers by Lifshitz (Lifshitz, 1943a, (Lifshitz, 1943b), (Lifshitz, 1944)) was published in Russia during the Second World War. This phase of research culminated in the seminal paper by Anderson (1958). Anderson’s model was of lattices—regular or irregular—of electron spins (or other entities) each of whose energy was a random variable. Then, provided the interparticle potential decayed sufficiently fast as a function of distance (faster than $r^{-3}$), and the disorder was stronger than some critical value, Anderson was able to show that the electron wave function was localized in space asymptotically with time. For a recent review of localization theory and experiment see Kramer & MacKinnon (1993).
In this way we have managed to reduce the study of wave propagation in disordered 1D media to application of the theory of products of random matrices (PRM). We note that for constant spring constants and varying masses the coefficient matrix is symplectic, i.e. \( J^T B_j J = B_t \) for all \( e \) where
\[
J = \begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix}
\] (22)

Also in this case, the determinant of each transfer matrix is 1 and hence the determinant of the product \( P_N \) is also 1. That the product matrices are unimodular assures that the sum of the two LEs is zero. In general, for varying spring constants, the matrices \( B_t \) are not symplectic, nor are the determinants equal to 1. But since the determinant of the product \( P_N \) is the product of the determinants of the individual \( B_t \), the determinant of \( P_N \) equals \( k_i / k_N \).

In the 1D numerical experiments described below, with pseudo-random \( k_i \), we have always found there to be one positive and one negative LE.

Using the PRM approach it is clear that the growth or decay of solutions to the equations of motion is governed by exponential growth and decay rates of the product matrix \( P_N \). The logarithm of the eigenvalues of a symmetrized version of this matrix (described below) are the Lyapunov exponents. Now the localization length is defined in terms of the exponential decay of eigenfunctions. But it seems reasonable to associate this localization length with the reciprocal of the positive (1D) or smallest positive (quasi-1D) LE. This conjecture is well supported by numerical evidence (cf. (Crisanti et al., 1993), (Kramer & MacKinnon, 1993), and the simulations below).

So there remain two issues that must be addressed. First, what can be said about asymptotic properties of products of random matrices, such as occur in Equation 21? Secondly, how can the Lyapunov spectrum of \( P_N \) be accurately and efficiently computed?

**Asymptotic Properties of Products of Random Matrices**

The first basic result is due to Furstenberg (1963), who showed that if the matrices are independent and non-singular in a certain sense, the limit
\[
\lambda_l = \lim_{N \to \infty} \frac{1}{N} \ln \| P_N \| \tag{23}
\]
exists almost surely. Further, the maximum Lyapunov exponent \( \lambda_1 \) is a non-random quantity; in other words
\[
\lambda_1 = \lim_{N \to \infty} \frac{1}{N} \langle \ln \| P_N \| \rangle \tag{24}
\]
where the angled brackets refer to the average under the distribution associated with the \( B_t \), which are assumed to be independent and identically distributed. The Lyapunov exponent measures the growth rate associated with typical vectors \( z \)
\[
\lambda_l = \lim_{N \to \infty} \frac{1}{N} \ln \| P_N z \| \tag{25}
\]
although this does not preclude different growth rates for improbable choices of \( z \).

These results were generalized by Virster (1979), who showed that if the matrices \( B_t \) were of the form \( B_t = B(\xi) \), where \( \xi \) is a stationary ergodic Markov chain and \( B \) is a matrix function on the state space of the chain, then \( \lambda_1 \) exists and is positive almost surely. Virster's results, which are equivalent to Furstenberg's in the special case of independent matrices, thus apply to a rather large class of disordered systems including, for example, exponentially correlated Gaussian systems.

**Calculation of the Lyapunov Spectrum**

To determine the Lyapunov exponents of our linear mapping we will employ a method based upon performing a QR decomposition at each iteration. The discrete QR algorithm to be described below is the most widely used technique to approximate LEs. Other methods based upon performing a singular value decomposition are also possible (see Abarbanel et al. (1992) and Geist et al. (1990)). We consider the rates of growth and decay for the linear mapping
\[
P_{N+1} = B_N P_N, \quad P_N \in \mathbb{R}^{n \times n}, \quad N = 0, 1, \ldots, Y_0 = I \tag{26}
\]

**Definition (Oseledec, 1968):** Let
\[
P_{N+1} = B_N \cdots B_0, \quad N = 1, \ldots,
\]
be a fundamental solution of (26) (with \( (P^0)^T P^0 = I \)). Then, the following symmetric positive definite matrix exists
\[
\Lambda = \lim_{N \to \infty} \langle P_N (P_N^T P_N)^{-1} \rangle
\]
the logarithms of the eigenvalues of which are called **Lyapunov exponents**, and are denoted by \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \). The \( \lambda_i \)'s do not depend on the initial condition matrix \( P^0 \) almost surely.

The theorem of Oseledec (Oseledec, 1968) leads to an equivalent characterization of LEs. Let \( \lambda^{(1)} > \lambda^{(2)} > \cdots \) be the LEs of (26) not repeated by multiplicity. Let \( E^{(1)} \) be the invariant subspace of \( \mathbb{R}^n \) corresponding to the eigenvalues of \( \Lambda \) whose logarithm is less than or equal to \( \lambda^{(1)} \), so that \( \mathbb{R}^n = E^{(1)} \oplus E^{(2)} \oplus \cdots \). Let \( p_k \in E^{(k)} \setminus E^{(k+1)} \), then one has
\[
\lambda^{(k)} = \limsup_{N \to \infty} \frac{1}{N} \log \| P_N p_k \|
\]
The system (26) is called regular (Lyapunov, 1949) if
\[
\sum_{k=1}^{p} \lambda_k = \lim_{N \to \infty} \frac{1}{N} \log(|\det(B_{N-1} \cdots B_0)|)
\]
\[
= \lim_{N \to \infty} \frac{1}{N} \log(|\det(P_N)|).
\]
If (26) is regular and upper triangular, one has (Lyapunov, 1949)
\[
\lambda_k = \lim_{N \to \infty} \frac{1}{N} \sum_{j=0}^{N-1} \log(|B_{j}^{(k)}|), \quad k = 1, \ldots, p.
\]

When the \( B_j \) are not upper triangular, we will successively compute QR factorizations of the transfer matrices. Given orthogonal \( P_0 : P_0^T P_0 = I \), let \( Q_0 = P_0 \).

Set \( Z_{N+1} = B_N Q_N \), \( N = 0, 1, \ldots \) and then decompose \( Z_{N+1} = R_{N+1} \), where \( R_{N+1} \) is upper triangular with positive diagonal entries. Since we obtain \( Q_{N+1} R_{N+1} = B_N Q_N \), then \( Q_{N+1} R_{N+1} = B_N Q_N \), which is upper triangular. Thus, we can obtain the LEs as
\[
\lambda_k = \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \log(|R_{j}^{(k)}|)
\]
\[
= \lim_{N \to \infty} \frac{1}{N} \sum_{j=1}^{N} \log(|B_{j}^{(k)}|).
\]

Following the work in (Dieci et al., 1995) and (Dieci & Van Vleck, 1995) we will estimate the error in our calculation by fitting the error as \( \frac{a}{N} + c \). If \( \lambda_k(N) \) is the \( k \)th finite time Lyapunov exponent at iterate \( N \), then simple linear regression gives us the following values for \( \epsilon \) and \( K \):
\[
K = \frac{\sum_{n=1}^{N} \lambda_k(n) - \frac{1}{N} \sum_{n=1}^{N} \lambda_k(n)}{\sum_{n=1}^{N} \frac{1}{n^2} - \frac{1}{N} \sum_{n=1}^{N} \frac{1}{n}}
\]
and
\[
\epsilon = \frac{1}{N} \left( \sum_{n=1}^{N} \lambda_k(n) - K \sum_{n=1}^{N} \frac{1}{n} \right) - \lambda_k^{*}
\]
where \( \lambda_k^{*} \) is the exact Lyapunov exponent. Since we do not know the exact Lyapunov exponent, we use the best available approximation, \( \lambda_k(N) \), instead of \( \lambda_k^{*} \).

To recap, we have described a technique for estimating the exponential growth of decay of products of random transfer matrices by computing the Lyapunov spectrum of this matrix product via a discrete QR method. Error estimates follow from linear regression of the finite time exponents. Now we will illustrate applications of these methods to wave propagation problems of the sort described in the first section.

Simulations - One Dimension

To do the simulations we generate pseudo-random sequences of spring constants and layer thicknesses. These are chosen in an uncorrelated fashion from uniform distributions and then correlated by applying a running average of a given length.

Each layer is described by its thickness and spring constant and the medium is described by a sequence of thicknesses and the corresponding spring constants in each layer. Thus, the statistical properties of our simulated medium are described in terms of maximum and minimum values and correlation lengths for both the thicknesses of the layers and the spring constants.

If the layer thicknesses were all equal to 1, say, then the maximum frequency of propagation on the lattice would correspond to a wavelength of 1. By putting groups of homogeneous spring constants between the “scatterers” we can simulate wavelengths smaller than the distance between the scatterers.

Once the medium is determined, the task is to compute the positive Lyapunov exponents. The Lyapunov exponents are defined as limits, although our medium will have a finite thickness. To obtain better approximations one could average over several starting vectors or, more easily, simply consider the medium as periodic and average over several periods.

In Figures 6 – 12 we illustrate some of the numerical results we have obtained. All plots are in log-log scale and plot the frequency \( \omega/(2\pi) \) on the horizontal axis against the computed localization length on the vertical axis.

We use the computed upper bound for the Lyapunov exponent to compute a lower bound on the localization length; this is denoted by \( \text{LESerror} \) in the figures. The simulations are for media with a total thickness of 1000 and the Lyapunov exponent was obtained by averaging over 10 “periods.” In all of the simulations the spring constants had a maximum value of 100 and a minimum value of 1. The correlation length for the spring constants and the thicknesses was allowed to vary, as were the maximum and minimum thickness of the layers.

The maximum localization length computable in this way is the length of the lattice, in this case 1000. Thus, values of \( 1/\lambda_1 \) greater than 1000 correspond to weak attenuation. On the other hand, the minimum meaningful localization length is the lattice spacing. We see that at sufficiently high frequencies the waves are localized. However, as the medium becomes more and more smooth, the range of frequencies that are localized goes to zero.

Figures 6–8 illustrate the lack of monotonicity that occurs for layers of unit thickness and small correlation lengths for the spring constants. All figures provide com-
comparisons of the localization length and the spectrum of the tridiagonal matrix $-T$ (see Equations 7 and 8). The local minima in the localization length appear to occur where there are gaps in the spectrum of $-T$. (Since the frequency domain equations of motion involve $\omega^2$ we are actually showing the square roots of the eigenvalues of $-T$.)

Next we show results for media composed of random "layers" of varying thickness. In other words, to make a layer of thickness $n$ we put $n$ springs of a given spring constant together. In Figures 9 and 10 for thickness between 5 and 10, and 10 and 20 respectively, the localization length $\ell(f)$ is basically monotone, with regions of different behavior of $\ell(f)$ depending on whether $\ell$ is large or small compared to the layer thickness.

For Figures 11 and 12 we have thicknesses between 10 and 20. Also, the spring constant correlation lengths (thickness correlation lengths) are 1 and 4 (4 and 16), respectively. As the thicknesses and/or correlation lengths increase the localization length approaches that of a uniform medium, as expected.

The model used for the calculations in Figure 12, which show a localization length greater than or equal to 1000 for almost all frequencies, is the same as in Figure 1. However the results of the finite difference calculation in Figures 2 and 3 are for a 2D simulation, corresponding to the propagation of a point source in a layered 2D medium. For comparison with the 1D LE results we also carried out the 1D time-domain finite difference calculation, corresponding to a vertically incident plane wave rather than a point source. The results are kinematically identical to Figure 2, but amplitudes will be different due to the geometrical spreading. Figure 13 shows the corresponding 1D decay with distance of the time-domain pulse. This linear-log plot is well-fit by a straight line whose slope corresponds to a localization length of 1000. This is consistent with the theory since in 1D there is only 1 positive LE and so the decay must be a pure exponential; it also agrees quantitatively with the direct LE calculation. On the other hand, in the 2D finite difference results shown in Figure 3, there are clearly two different length scales involved. Legandijk et al. (1985) argue that for strong scattering the first exponential decay length is associated with the scattering mean free path, while the second is the characteristic length associated with absorption.

**Simulations - Two Dimensions**

The extension of our approach to higher dimensional lattices presents some challenges. The theory of the products of random matrices (PRM) still applies to some extent, although with more complicated matrices. In two-dimensions an approach similar to that considered above may be employed. Consider again Equation (8) where now $T$ represents the two-dimensional analogue of the Laplacian operator with random spring coefficients. For $z = \{z_{i,j}\}^{(1)}(N_1,N_2)$ we impose "corkscrew" boundary conditions in one direction. The corkscrew boundary conditions allow us to linearly order the variables as

$$z_{1,1}, \ldots, z_{1,N_2}, z_{2,1}, \ldots, z_{2,N_2}, \ldots$$

e tc., and write the evolution as a mapping whose Lyapunov exponents may be calculated. We note here that periodic boundary conditions may also be employed and this has been done for the Schrödinger equation (corresponding to constant spring constants and random masses) in Crisanti et al. (1993). For the case of periodic boundary conditions, random masses and constant spring constants, the matrices that are obtained are symplectic, in contrast with the matrices obtained below.

Let

$$z^{(1)} = (z_{1,1}, \ldots, z_{1,N_2}, z_{2,1}, \ldots, z_{2,N_2})^T,$$
$$z^{(2)} = (z_{1,2}, \ldots, z_{1,N_2}, z_{2,1}, \ldots, z_{2,N_2})^T,$$

etc., then for $n = i + (j-1)N_2$, $1 \leq i \leq N_1$, $1 \leq j \leq N_2$ we have

$$z^{(n+1)} = A_n z^{(n)}$$

where

$$A_n = \begin{pmatrix} 0 & I \\
-a_n & 0, \ldots, 0, -b_n, c_n, -d_n, 0, \ldots, 0 \end{pmatrix}.$$  

The coefficients of the matrix $A_n$ are given in terms of the frequency and the spring constants as:

$$a_n = k_n/k_{N_2-n},$$
$$b_n = k_{N_2-n}/k_{N_2-n},$$
$$c_n = (k_n + k_{N_2-n-1} + k_{N_2+n-1} - (\omega^2)/k_{N_2+n}),$$
$$d_n = k_{N_2+n}/k_{N_2-n}$$

(29)

modulo the boundary conditions. Notice that for each of the matrices $A_n$ the trace is zero and the determinant is $\pm a_n$.

In Figures 14, 15 and 16 we illustrate the results of numerical experiments obtained for various values of the width $N_2$ computed over 5 periods. Note that the Lyapunov exponents are ordered according to

$$\lambda_1 \geq \lambda_2 \geq \ldots \lambda_{N_2} \geq \ldots \lambda_{2N_2}.$$  

The pseudo-random models used have the same statistical properties as in Figures 7 and 8, respectively.
Figure 6. $\frac{1}{\lambda_1}$ vs. $\xi^2$, where $\lambda_1$ is the positive Lyapunov exponent, the spring correlation length = 1, thickness correlation length = 1, maximum thickness = 1, minimum thickness = 1

Figure 7. $\frac{1}{\lambda_1}$ vs. $\xi^2$, spring correlation length = 2, thickness correlation length = 1, maximum thickness = 1, minimum thickness = 1

Figure 16 was computed for the model shown in Figure 1 and therefore affords a direct comparison with the 1D LE calculation in Figure 12 and the time-domain finite difference results in Figures 3 (2D) and 13 (1D). In 2D, as we would have expected from the finite difference results, we see localization lengths less than 1000 for nearly all frequencies. In addition we see clear evidence of nonmonotonicity of the frequency dependence of the localization length.

That the Lyapunov exponents are nonrandom in the quasi-1D case follows under reasonable assumptions from the work of Osceledce (1968). However, in the quasi-1D case there is no guarantee that the first $N_2$ finite time Lyapunov exponents are positive even if $det(P_\nu) \geq 1$. As a consequence there appear to be two alternative defini-
Figure 8. $\frac{1}{\lambda_1}$ vs. $\frac{1}{\lambda_2}$, spring correlation length = 4, thickness correlation length = 1, maximum thickness = 1, minimum thickness = 1

Figure 9. $\frac{1}{\lambda_1}$ vs. $\frac{1}{\lambda_2}$, spring correlation length = 1, thickness correlation length = 1, maximum thickness = 10, minimum thickness = 5

tions of the localization length. The first is to define the localization length as the reciprocal of the smallest positive exponent. The second is to define the localization length as the $\frac{1}{\lambda N_2}$ if $\lambda N_2$ is positive and infinite otherwise. We have employed the second definition in our simulations.

Finally, to decrease the computation time we compute only the first $N_2$ exponents and not all $2N_2$ exponents (see Dieci & Van Vleck (1993)). Note in Figures 14 and 15 that it appears that the localization lengths are converging as the width $N_2$ is being increased.

Conclusions
 Classical waves scattering from random heterogeneities in elastic solids may excite vibrational modes that are loc-
Figure 10. $\frac{1}{\lambda} \text{ vs. } \frac{1}{\lambda}$, spring correlation length = 1, thickness correlation length = 1, maximum thickness = 20, minimum thickness = $10^1$.

Figure 11. $\frac{1}{\lambda} \text{ vs. } \frac{1}{\lambda}$, spring correlation length = 1, thickness correlation length = 4, maximum thickness = 20, minimum thickness = $10^1$.

alized about the heterogeneities. This causes propagating energy to be converted into localized fluctuations and is manifested in the dispersion and attenuation of the waves. With a sufficiently strong random distribution of heterogeneities, it is possible for a propagating wave to become trapped altogether. The frequency-dependent length scale on which this attenuation occurs—the localization length—is the fundamental quantity characterizing this phenomenon. For quasi-1D wave propagation, i.e. propagation that can be described via products of transfer matrices, the localization length can be estimated from the Lyapunov spectrum of the product of random matrices. We have developed an algorithm for computing the positive, finite-time Lyapunov exponents (along with error estimates) associated with layered elastic media and applied this algorithm to a number
of pseudo-randomly generated models. The results show overall agreement with coarse estimates of localization made from time-domain finite difference calculations, but provide a more detailed picture of scattering attenuation.

**Acknowledgements**

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena and the National Science Foundation under grant DMS-
Figure 14. $\frac{1}{\lambda_{N_2}}$ vs. $\frac{1}{\tau_t}$, spring correlation length = 2, thickness correlation length = 1, maximum thickness = 1, minimum thickness = 1

Figure 15. $\frac{1}{\lambda_{N_2}}$ vs. $\frac{1}{\tau_t}$, spring correlation length = 4, thickness correlation length = 1, maximum thickness = 1, minimum thickness = 1

9506003. In addition JS acknowledges the support of the Shell Foundation and EVV acknowledges the support of the National Science Foundation under grant DMS-9505049.

References


of random matrices. Springer.


Generalized Bremmer series with rational approximation for the scattering of waves in inhomogeneous media

Mattheus J.N. van Stralen*, Maarten V. de Hoop†

and

Hans Blok*

April 12, 1996

Abstract

The Bremmer series solution of the wave equation in generally inhomogeneous media, requires the introduction of pseudo-differential operators. In this paper, sparse matrix representations of these pseudo-differential operators are derived. We focus on designing sparse matrices, keeping the accuracy high at the cost of ignoring any critical scattering-angle phenomena. Such matrix representations follow from rational approximations of the vertical slowness and the transverse Laplace operator symbols, and of the vertical derivative, as they appear in the parabolic equation method. Sparse matrix representations lead to a fast algorithm. An optimization procedure is followed to minimize the errors, in the high frequency limit, for a given discretisation rate. The Bremmer series solver consists of three steps: directional decomposition into up- and downgoing waves, one-way propagation, and interaction of the counter-propagating constituents. Each of these steps is represented by a sparse matrix equation. The resulting algorithm provides an improvement of the parabolic equation method, in particular for transient wave phenomena, and extends the latter method, systematically, for backscattered waves.

*Laboratory of Electromagnetic Research, Faculty of Electrical Engineering, Delft University of Technology, P.O. Box 5031, 2600 GA Delft, the Netherlands.
†Center for Wave Phenomena, Colorado School of Mines, Golden CO 80401-1887, USA.
1 Introduction

Directional wave field decomposition is a tool for analyzing and computing the propagation of waves in configurations with a certain directionality, such as the waveguiding structure in Figure 1. The method consists of three main steps: (i) decomposing the field into two constituents, propagating upward or downward along a preferred direction, (ii) computing the interaction of the counterpropagating constituents and (iii) recomposing the constituents into observables at the positions of interest. The method is beneficial because it can be computationally efficient and it can be used to separate different propagation phenomena, which is of importance in the interpretation and inversion of measurements.

In the Bremmer series approach to modelling, we encounter pseudo-differential operators in the directional (de)composition, in the downward and upward propagation or continuation, and in the reflections and transmissions due to variations in medium properties in the preferred direction (De Hoop [1]). For the numerical implementation, we employ a total rational-approximation approach to find, upon discretisation, sparse matrix representations of these pseudo-differential operators. The rational approximation has its roots in the parabolic equation (PE) method (Claber [2], and Tappert [3]), and has been extended and explored by Ma [4], Greene [5], Halpern and Trefethen [6], and Collins [7]. The rational approximation should be carried out in a delicate way, to ensure conservation of acoustic power flow, see also Collins and Westwood [8]. The Bremmer series generates systematically the backscattered field, a topic investigated by Collins and Evans [9], and Collins [10]. For fixed sampling rates, allowing the numerical grid to be coarse, we consider optimizations of the matrix representations for the three steps, (de)composition, propagation, and interaction, such that the numerical dispersion is minimized. The idea of optimization was exploited by Collins [11], and Cederberg et al. [12]; the numerical dispersion was carefully analysed by Trefethen [13], Beaumont et al. [14], and Holberg [15, 16].

The improvement in accuracy and efficiency, and extensions of rational approximation techniques remain to get significant attention. Recent advances in the application to exploration seismics can be found in Graves and Clayton [17], and Rühl et al. [18]. They controlled the errors at large scattering angles by initiating the propagation with a transversely homogeneous background phase shift. In the field of integrated optics, the PE method was introduced by Fleck, Morris and Feit [19]. Various extensions of the PE-style method have been developed since then. These are now known as Beam Propagation Methods (BPMs). Examples of present BPMs are the Method of Lines [20], the Mode Expansion Method [21] and wide-angle methods based upon higher-order rational approximations [22, 23, 24]. For a recent overview we refer to Hoekstra [25]. For the developments in ocean acoustics, we refer the reader to Collins [7, 26]. A recent overview of PE methods in underwater acoustics is given by Lee and Pierce [27].

The discretisation of the ‘one-way’ wave equation, the propagation step, is based on the third-order Thiele-type-approximation of the left vertical wave-slowness symbol. We enforce the associated vertical slowness operator to be self-adjoint (‘energy-conserving’), as the exact one is self-adjoint in the real $L^2$. This implies that we depart from the usual principal symbol analysis. The one-way wave equation is thus approximated by a partial differential equation. By discretising the transverse derivatives (the Laplace operator) according to a rational approximation of its spectrum, the partial differential equation is transformed into a system of ordinary differential equations. The solution of this system is formally written as a product integral. The exponent in the discretised version of the product integral is then replaced by its $(n,n)$-Padé approximation. Such a procedure guarantees numerical stability. We will pay most attention to the $(1,1)$-Padé approximation, which yields the Crank-Nicholson implicit finite-difference scheme in the preferred direction. The resulting algebraic equations, which now involve sparse matrices, can be solved rapidly with standard procedures available in various software libraries. In two dimensions, direct matrix inversions are carried out; in three dimensions, iterative techniques should be applied.
The vertical phase and group slownesses associated with the ultimate system of algebraic equations can be evaluated and are used to analyze the numerical artifacts introduced by the various rational approximations. The accuracy of the vertical group slowness as compared with the exact vertical slowness is indicative for the numerical anisotropy; the difference between the vertical phase and group slownesses is indicative for the numerical dissipation. To arrive at an optimal system of algebraic equations, fixing the bandwidth of the wave field and the sampling rate, a simultaneous optimization of the Thiele-type approximation, the transverse finite-difference, and the vertical finite-difference representations is carried out. Since the optimization will depend on the medium's wave speed, the optimal parameter set will vary with frequency and position as well. The optimization procedure is repeated for the composition, decomposition, reflection and transmission operators. With the various matrix representations, the Bremmer coupling series solution of the wave equation can now be computed.

The outline of the paper is as follows. In the next section, a summary of the method of directional decomposition, leading to a coupled system of one-way wave equations is given. In Section 3, the concept of generalized slowness surface is introduced, which is used in representations of the Green's functions of the one-way wave equations. In Section 4, the Bremmer coupling series and its numerical implementation are discussed. The remainder of the paper is dedicated to the derivation and optimization of approximate, sparse, matrix representations. In Section 5, the one-way wave equations are discretized for the purpose of solving for the one-way Green's functions. In Section 6, transverse, transparent boundaries are introduced. In Section 7, the optimization procedure for the one-way propagation is explained. Sections 8 and 9 contain the discretisations and optimizations of the (de)composition and interaction operators. Finally, our algorithm is illustrated in Section 10 by various numerical examples.
2 Directional wave field decomposition

For the details on the derivation of the Bremmer coupling series solution of the acoustics wave equation, we refer the reader to De Hoop [1]. Here, we restrict ourselves to a summary of the method.

Notation, transformations

In each subdomain of the configuration where the acoustic properties vary continuously with position, the acoustic wave field satisfies the hyperbolic system of partial differential equations

\begin{align}
\partial_t p + \rho \partial_t v_k &= f_k, \\
\kappa \partial_t p - \partial_r v_r &= q,
\end{align}

where \(p\) = acoustic pressure [Pa],
\(v_r\) = particle velocity [m/s],
\(\rho\) = volume density of mass [kg/m\(^3\)],
\(\kappa\) = compressibility [Pa\(^{-1}\)],
\(q\) = volume source density of injection rate [s\(^{-1}\)],
\(f_k\) = volume source density of force [N/m\(^3\)],

where \((x_1, x_2, x_3)\) are the right-handed, orthogonal, Cartesian coordinates, \(t\) is the time, and the subscript notation and the summation convention for Cartesian tensors are employed. We assume that the coefficient \(\kappa\) is smooth, and that \(\rho\) is constant, for the purpose of wave field decomposition. Furthermore, we assume that these functions are constant outside a compact domain. This provision enables us to formulate the acoustic wave propagation, when necessary, as a scattering problem in a homogeneous embedding. The smoothness entails that the singularities of the wave field (in particular the ones in the neighborhood of the wave arrival) arise from the ones in the signatures of the source distributions. Further, causality of the wave motion is enforced. This implies that if the sources that generate the wave field are switched on at the instant \(t = 0\), the wave field quantities satisfy the initial conditions

\begin{align}
p(x_m, t) &= 0 \text{ for } t < 0 \text{ and all } x_m , \\
v_r(x_m, t) &= 0 \text{ for } t < 0 \text{ and all } x_m.
\end{align}

Due to the time invariance of the medium, the causality of the wave motion can also be taken into account by carrying out a one-sided Laplace transformation with respect to time and requiring that the transform-domain wave quantities are bounded functions of position in all space when the time Laplace-transform parameter \(s\), which is in general complex, lies in the right half \(\text{Re}\{s\} > 0\) of the complex \(s\)-plane. The limiting case of sinusoidal oscillations of angular frequency \(\omega \in \mathbb{R}\) is covered by considering the limiting case \(s \to \iota \omega\), in which \(\iota\) is the imaginary unit, the limit being taken via \(\text{Re}\{s\} > 0\).
To show the notation, we give the expression for the acoustic pressure

$$\tilde{p}(x_m, \omega) = \int_{t=0}^{\infty} \exp(-i\omega t)p(x_m, t) \, dt, \quad \text{Im}\{\omega\} < 0.$$  \hfill (2.5)

Under this transformation, assuming zero initial conditions, we have $\partial_t \rightarrow i\omega$. The transformed system of first-order equations follows from Eqs. (2.1) and (2.2) as

$$\partial_k \tilde{p} - i\omega \rho \tilde{v}_k = \tilde{f}_k,$$  \hspace{1cm} (2.6)

$$i\omega \kappa \tilde{p} + \partial_r \tilde{v}_r = \tilde{q}.$$  \hspace{1cm} (2.7)

The change of the wave field in space along a direction of preference can now be expressed in terms of the changes of the wave field in the plane perpendicular to it. The direction of preference is taken along the $x_3$-axis (or 'vertical' axis) and the remaining ('transverse' or 'horizontal') coordinates are denoted by $x_\mu$, $\mu = 1, 2$. Our numerical examples, however, will be restricted to two dimensions.

**The reduced system of equations**

The procedure requires a separate handling of the horizontal components of the particle velocity. From Eqs. (2.6) and (2.7) we obtain

$$\dot{\tilde{v}}_\mu = i\rho^{-1}\omega^{-1}(\partial_\mu \tilde{p} - \tilde{f}_\mu),$$  \hspace{1cm} (2.8)

leaving, upon substitution, the matrix differential equation

$$(\partial_3 \delta_{l,j} + i\omega \hat{A}_{l,j})\hat{F}_j = \hat{N}_l, \quad \hat{A}_{l,j} = \hat{A}_{l,j}(x_\mu, D_\nu; x_3), \quad D_\nu \equiv \frac{i}{\omega} \partial_\nu,$$  \hspace{1cm} (2.9)

in which the elements of the acoustic field matrix are given by

$$\hat{F}_l = \hat{\tilde{p}}, \quad \hat{F}_3 = \hat{\tilde{v}}_3,$$  \hspace{1cm} (2.10)

the elements of the acoustic system's operator matrix by

$$\hat{A}_{1,1} = \hat{A}_{2,2} = 0,$$  \hspace{1cm} (2.11)

$$\hat{A}_{1,2} = \rho,$$  \hspace{1cm} (2.12)

$$\hat{A}_{2,1} = -D_\nu (\rho^{-1} D_\nu) + \kappa,$$  \hspace{1cm} (2.13)

and the elements of the notional source matrix by

$$\hat{N}_1 = \hat{f}_3, \quad \hat{N}_2 = D_\nu (\rho^{-1} \hat{f}_\nu) + \hat{q}.$$  \hspace{1cm} (2.14)

It is observed that the right-hand side of Eq. (2.8) and $\hat{A}_{l,j}$ contain spatial derivatives $D_\nu$ with respect to the horizontal coordinates only. $D_\nu$ has the interpretation of horizontal slowness operator. Further, it is noted that $\hat{A}_{1,2}$ is simply a multiplicative operator.

To ensure that the medium is smooth, we employ equivalent medium averaging at any point over a box that is twice the spatial sample size (see Coates and Schoenberg [28]).
The coupled system of one-way wave equations

To distinguish up- and downgoing constituents in the wave field, we shall construct an appropriate linear operator \( \hat{L}_{I,J} \) with

\[
\hat{F}_I = \hat{L}_{I,J} \hat{W}_J ,
\]

that, with the aid of the commutation relation

\[
(\partial_3 \hat{L}_{I,J}) = [\partial_3, \hat{L}_{I,J}]
\]

([.,.] denotes the commutator), transforms Eq.(2.9) into

\[
\hat{L}_{I,J} (\partial_3 \delta_{J,M} + i\omega \hat{A}_{J,M}) \hat{W}_M = -(\partial_3 \hat{L}_{I,J}) \hat{W}_J + \hat{N}_J ,
\]

as to make \( \hat{A}_{J,M} \), satisfying

\[
\hat{A}_{I,J} \hat{L}_{J,M} = \hat{L}_{I,J} \hat{A}_{J,M} ,
\]

a diagonal matrix of operators. We denote \( \hat{L}_{I,J} \) as the composition operator and \( \hat{W}_M \) as the wave matrix. The expression in parentheses on the left-hand side of Eq.(2.17) represents the two so-called one-way wave operators. The first term on the right-hand side of Eq.(2.17) is representative for the scattering due to variations of the medium properties in the vertical direction. The scattering due to variations of the medium properties in the horizontal directions is contained in \( \hat{A}_{J,M} \) and, implicitly, in \( \hat{L}_{I,J} \).

To investigate whether solutions of Eq.(2.18) exist, we introduce the column matrix operators \( \hat{L}_I^{(\pm)} \) according to

\[
\hat{L}_I^{(\pm)} = \hat{L}_{I.1} , \quad \hat{L}_I^{(\mp)} = \hat{L}_{I.2} .
\]

Upon writing the diagonal elements of \( \hat{A}_{J,M} \) as

\[
\hat{\Lambda}_{1.1} = \hat{\Gamma}_1^{(\pm)} , \quad \hat{\Lambda}_{2.2} = \hat{\Gamma}_2^{(\mp)} ,
\]

Eq.(2.18) decomposes into the two systems of equations

\[
\hat{\Lambda}_{I,J} \hat{L}_J^{(\pm)} = \hat{L}_I^{(\pm)} \hat{\Gamma}_J^{(\pm)} .
\]

By analogy with the case where the medium is translationally invariant in the horizontal directions, we shall denote \( \hat{\Gamma}_1^{(\pm)} \) as the vertical slowness operators. Notice that the operators \( \hat{L}_I^{(\pm)} \) compose the acoustic pressure and that the operators \( \hat{L}_2^{(\pm)} \) compose the vertical particle velocity, whereas the elements of \( \hat{W}_M \) may be physically 'non-observable'. Through mutual elimination, the equations for \( \hat{L}_I^{(\pm)} \) and \( \hat{L}_2^{(\pm)} \) can be decoupled as follows:

\[
\hat{\Lambda}_{1.2} \hat{\Lambda}_{2.1} \hat{L}_I^{(\pm)} = \hat{L}_I^{(\pm)} \hat{\Gamma}_2^{(\pm)} \hat{\Gamma}_1^{(\pm)} ,
\]

\[
\hat{\Lambda}_{2.1} \hat{\Lambda}_{1.2} \hat{L}_2^{(\pm)} = \hat{L}_2^{(\pm)} \hat{\Gamma}_1^{(\pm)} \hat{\Gamma}_2^{(\pm)} .
\]

The partial differential operators on the left-hand sides differ from one another in case the volume density of mass does vary in the horizontal directions.
To ensure that non-trivial solutions of Eqs. (2.22)-(2.23) exist, one equation must imply the other. To construct a formal solution, an Ansatz is introduced in the form of a commutation relation for one of the components \( \hat{L}^{(\pm)}_j \) that restricts the freedom in the choice for the other component. In the acoustic-pressure normalization analog one assumes that \( \hat{L}^{(\pm)}_2 \) can be chosen such that

\[
[\hat{L}^{(\pm)}_2, \hat{A}_{2,1} \hat{A}_{1,2}] = 0 .
\]  

(2.24)

In view of Eq. (2.23) the \( \tilde{\Gamma}^{(\pm)} \) must then satisfy

\[
\hat{A}_{2,1} \hat{A}_{1,2} - \tilde{\Gamma}^{(\pm)} \hat{\Gamma}^{(\pm)} = 0 .
\]  

(2.25)

The commutation relation for \( \hat{L}^{(\pm)}_1 \) follows as \([\hat{A}_{1,2}^{-1} \hat{L}^{(\pm)}_1, \hat{A}_{2,1} \hat{A}_{1,2}] = 0 \) and a possible solution of Eq. (2.21) is

\[
\hat{L}^{(\pm)}_2 = \hat{\Gamma}^{(\pm)}, \hat{L}^{(\pm)}_1 = \hat{A}_{1,2} .
\]  

(2.26)

Since \( \hat{L}^{(\pm)}_2 \) as given by Eq. (2.26) satisfies Eq. (2.24), the Ansatz is justified. The solutions of Eq. (2.25) are written as

\[
\hat{\Gamma}^{(+)} = -\hat{\Gamma}^{(-)} = \hat{\Gamma} = \hat{A}^{1/2} , \quad \hat{A} = \hat{A}_{2,1} \hat{A}_{1,2} .
\]  

(2.27)

Thus, the composition operator becomes

\[
\hat{L} = \begin{pmatrix}
\hat{A}_{1,2}^{1/2} & \hat{A}_{1,2}^{1/2} \\
\hat{\Gamma} & -\hat{\Gamma}
\end{pmatrix} .
\]  

(2.28)

Note that we have decomposed the pressure (up to a multiplication by density) viz. according to \( \hat{F}_1 = \hat{F}_1^+ + \hat{F}_1^- , \hat{F}_1^+ = \hat{A}_{1,2} \hat{W}_2 \) and \( \hat{F}_1^- = \hat{A}_{1,2} \hat{\bar{W}}_2 \). In terms of the inverse vertical slowness operator, \( \hat{\Gamma}^{-1} = \hat{A}^{-1/2} \), the decomposition operator then follows as

\[
\hat{L}^{-1} = \frac{1}{2} \begin{pmatrix}
\hat{A}_{1,2}^{-1/2} & \hat{\Gamma}^{-1} \\
\hat{A}_{1,2}^{-1/2} & -\hat{\Gamma}^{-1}
\end{pmatrix} .
\]  

(2.29)

Using the decomposition operator, Eq. (2.17) transforms into

\[
(\partial_2 \delta_{I,M} + i \omega \hat{A}_{I,M}) \hat{W}_M = -(\hat{L}^{-1})_{I,M}(\partial_2 \hat{L}_{M,K}) \hat{W}_K + (\hat{L}^{-1})_{I,M} \hat{N}_M ,
\]  

(2.30)

which can be interpreted as a coupled system of one-way wave equations. The coupling between the counter-propagating components, \( \hat{W}_M \), is apparent in the first source-like term on the right-hand side. We have

\[
-L^{-1}(\partial_2 \hat{L}) = \begin{pmatrix}
\hat{T} & \hat{R} \\
\hat{R} & \hat{T}
\end{pmatrix} ,
\]  

(2.31)

in which \( \hat{T} \) and \( \hat{R} \) represent the transmission and reflection operators, respectively.
\section{The one-way wave propagator}

We now subject the left-hand side of Eq.\,(2.30) to a further investigation. In it, we recognize the operators

\[ \partial_3 + i\omega \hat{\Gamma}^{(\pm)} : \mathcal{L}(\mathbb{R}, H^r(\mathbb{R}^2)) \to \mathcal{L}(\mathbb{R}, H^{r-1}(\mathbb{R}^2)), \]

(3.1)

where \( \mathcal{L}(\mathbb{R}, H^r(\mathbb{R}^2)) \) denotes the space of maps \( \mathbb{R} \to H^r(\mathbb{R}^2) \) and \( H^r \) is a reserved notation for Sobolev spaces (see De Hoop [1]). The operators in Eq.\,(3.1) are the exact one-way wave operators.

To arrive at the coupled system of integral equations that is equivalent to Eq.\,(2.30) and that can be solved in terms of a Neumann expansion, we have to invert the operator occurring on the left-hand side. The one-sided elementary kernels \( \hat{G}^{(\pm)}(x_\mu, x_\nu; x'_\mu, x'_\nu) \) associated with the operators

\[ \hat{G}^{(\pm)} = (\partial_3 + i\omega \hat{\Gamma}^{(\pm)})^{-1} : \mathcal{L}(\mathbb{R}, H^r(\mathbb{R}^2)) \to \mathcal{L}(\mathbb{R}, H^{r}(\mathbb{R}^2)) \]

in three-dimensional space are the so-called Green's functions. The latter satisfy the equations

\[ \partial_3 \hat{G}^{(\pm)} + i\omega \hat{\Gamma}^{(\pm)} \hat{G}^{(\pm)} = \delta(x_\nu - x'_\nu) \delta(x_3 - x'_3), \]

(3.2)

together with the condition of causality.

Now, consider the case \( \hat{G} = \hat{G}^{(\mp)} \), \( \hat{G} = \hat{G}^{(\mp)} \) and \( \hat{\Gamma} = \hat{\Gamma}^{(\mp)} \). The operator \( \hat{G} \) acts on a test field \( \hat{u} \) as

\[ \langle \hat{G}\hat{u}\rangle(x_\mu, x_3) = \int_{\zeta \in \mathbb{R}} \int_{x'_\mu, x'_3, \zeta} \hat{G}(x_\mu, x_3; x'_\mu, \zeta) \hat{u}(x'_\mu, \zeta) \text{d}x'_\mu \text{d}x'_3 \text{d}\zeta. \]

(3.3)

Let us define the initial-value problem of determining a function \( \hat{U}(x_\mu, x_3; \zeta) \) satisfying

\[ (\partial_3 + i\omega \hat{\Gamma}) \hat{U} = 0 \text{ for } x_3 \geq \zeta, \quad \hat{U}(x_\mu, \zeta; \zeta) = \hat{u}(x_\mu, \zeta). \]

(3.4)

Then it is observed that

\[ \langle \hat{G}\hat{u}\rangle(x_\mu, x_3) = \int_{\zeta = -\infty}^{x_3} \hat{U}(x_\mu, x_3; \zeta) \text{d}\zeta. \]

(3.5)

\section{The product integral}

We observe that the vertical slowness operators at different levels of \( x_3 \) do not necessarily commute with one another due to the heterogeneity of the medium. Thus we arrive at a ‘time’-ordered product integral representation \cite{29} of the one-sided propagators (cf.\,Eq.\,(3.4)) associated with the one-way wave equations, where ‘time’ refers to the vertical coordinate \( x_3 \),

\[ \hat{U}^{(\pm)}(\cdot, x_3) = \pm H(\mp [x'_3 - x_3]) \left\{ \prod_{\zeta = x'_3}^{x_3} \exp[-i\omega \hat{\Gamma}^{(\pm)}(\cdot, \zeta) \text{d}\zeta] \right\} \hat{u}(\cdot, x'_3). \]

(3.6)

In this expression, the operator ordering is initiated by \( \exp[-i\omega \hat{\Gamma}(\cdot, x'_3) \text{d}\zeta] \) acting on \( \hat{u}(\cdot, x'_3) \) followed by applying \( \exp[-i\omega \hat{\Gamma}(\cdot, \zeta) \text{d}\zeta] \) to the result, successively for increasing \( \zeta \).

If the medium in the interval \([x'_3, x_3]\) were weakly varying in the vertical direction, the Trotter product formula can be applied to the product integral in Eq.\,(3.6). This results in the Hamiltonian path integral.
representations for the Green’s functions,
\[
\hat{G}^{(\pm)}(x_\nu, x_\mu; x'_\nu, x'_\mu) = \pm H(\mp [x'_3 - x_3]) \int P \mathcal{D}(x''_\nu, \alpha''_\nu) \\
\exp \left[-i\omega \int_{\zeta=x'_3}^{x_3} d\xi \{\alpha''_\nu(d\xi x''_\nu) - \delta^{(\pm)}(x''_\mu, \zeta, \alpha''_\nu)\} \right],
\]  \tag{3.7}
\]

\(P\) being a set of paths \((x''_\nu(\zeta, \alpha''_\nu(\zeta))\) in (horizontal) phase space satisfying
\[
x''_\nu(\zeta=x'_3) = x'_\nu, \quad x''_\nu(\zeta=x_3) = x_\nu.
\]  \tag{3.8}

Omitting the Heaviside function in the expression for the \(\hat{G}^{(\pm)}\) yields the kernel \(g^{(\pm)}\) of the so-called phase shift operator. In Eq.(3.7), \(\delta^{(\pm)}\) is the left symbol of \(\Gamma^{(\pm)}\),
\[
\hat{\Gamma}^{(\pm)}(x_\mu, D_\nu; x_3) \exp(-i\omega\alpha_\nu x_\nu) = \delta^{(\pm)}(x_\mu, x_3, \alpha_\nu) \exp(-i\omega\alpha_\nu x_\nu).
\]  \tag{3.9}

We restrict ourselves to causal solutions by imposing that \(\hat{G}^{(\pm)}\) remain bounded as \(|x_3| \to \infty\).

The path integral in Eq.(3.7) is to be interpreted as the lattice multi-variate integral
\[
\hat{G}^{(\pm)}(x_\nu, x_\mu; x'_\nu, x'_\mu) = \pm H(\mp [x'_3 - x_3]) \lim_{M \to \infty} \int \prod_{i=1}^{M} (\omega/2\pi)^2 d^2 \alpha^{(i)} \prod_{j=1}^{M-1} d^2 x^{(j)} \\\n\exp \left[-i\omega \sum_{k=1}^{M} \{\alpha^{(k)}_\nu(x^{(k)}_\nu - x^{(k-1)}_\nu) + \delta^{(\pm)}(x^{(k)}_\mu, \zeta_k - \frac{1}{2} M^{-1} \Delta z_3, \alpha^{(k)}_\mu) M^{-1} \Delta z_3\} \right].
\]  \tag{3.10}

with
\[
x^{(0)}_\nu = x'_\nu, \quad x^{(M)}_\nu = x_\nu,
\]  \tag{3.11}

and
\[
\Delta z_3 = x_3 - x'_3.
\]  \tag{3.12}

All the integrations are taken over the interval \((-\infty, \infty), \Delta z_3\) is the step size in \(\zeta\), and \((x^{(j)}_\nu, \alpha^{(j)}_\nu)\) are the coordinates of a path at the discrete values \(\zeta_j\) of \(\zeta\) as \(j = 1, \cdots, M\).

If \(\Delta z_3\) is sufficiently small, the path integral reduces to
\[
\hat{G}^{(\pm)}(x_\mu, x_3; x'_\nu, x'_3) \simeq \pm H(\mp [x'_3 - x_3]) \int (\omega/2\pi)^2 d\alpha''_\nu \\hat{G}^{(\pm)}
\exp \left[-i\omega \{\alpha''_\nu(x'_\nu - x'_3) + \delta^{(\pm)}(x_\mu, x_3 - \frac{1}{2} \Delta z_3, \alpha''_\nu) \Delta z_3\} \right].
\]  \tag{3.13}

The generalized slowness surface

In the acoustic pressure normalization analog, the key differential operator (cf. Eq.(2.25)) is given by
\[
\hat{A} = -D_\nu D_\nu - \rho^{-1}(D_\nu \rho) D_\nu - \rho^{-2}(D_\nu \rho)(D_\nu \rho) - \rho^{-1}(D_\nu D_\nu \rho) + \kappa \rho,
\]  \tag{3.14}
with left symbol \( \hat{a} = \hat{a}(x_{\mu}, \alpha_{\nu}) \),

\[
\hat{a} = \hat{a}_{2} + \hat{a}_{1} + \hat{a}_{0},
\]

(3.15)
in which

\[
\hat{a}_{2} = -\sigma_{\nu}^{2} + \kappa \rho,
\]

\[
\hat{a}_{1} = \rho^{-1}(D_{\nu} \rho) \alpha_{\nu},
\]

\[
\hat{a}_{0} = \rho^{-1}(D_{\nu}^{2} \rho) - \rho^{-2}(D_{\nu} \rho)^{2}.
\]

In our configuration, we have \( \hat{a}_{1} = \hat{a}_{0} = 0 \), since we assume the density to be constant. The symbol lies in the space \( S^{2} \) (see De Hoop [1]). Note that

\[
c = (\kappa \rho)^{-1/2}
\]
is the wave speed in the medium.

Using the composition rule for symbols of pseudo-differential operators, the operator equation (2.25) is transformed into an equation for the corresponding left symbols,

\[
-(\omega/2\pi)^{2} \int_{x'_{\nu} \in \mathbb{R}^{2}} \int_{\alpha'_{\nu} \in \mathbb{R}^{2}} 
\]

\[
\hat{\gamma}(x_{\mu}, \alpha'_{\lambda}) \exp[i\omega(x_{\sigma} - x'_{\sigma})(\alpha_{\sigma} - \alpha'_{\sigma})] \hat{\gamma}(x'_{\lambda}, \alpha_{\nu}) \, d\alpha'_{\nu} \, dx'_{\nu} \, dx'_{\lambda} + \hat{a}(x_{\mu}, \alpha_{\nu}) = 0.
\]

(3.17)

This equation defines the generalized slowness surface and has solutions \( \hat{\gamma}(\pm) \in S^{1} \). The two branches are \( \hat{\gamma}(\pm)(x_{\mu}, \alpha_{\lambda}) \) such that

\[
\text{Im}\{\hat{\gamma}(\pm)(x_{\mu}, \alpha_{\lambda})\} \leq 0 \quad \text{and} \quad \text{Im}\{\hat{\gamma}(\mp)(x_{\mu}, \alpha_{\lambda})\} \geq 0.
\]

Due to the up/down symmetry of the medium we have \( \hat{\gamma}(\pm) = -\hat{\gamma}(\mp) \). Note that as \( \omega \to \infty \) the composition of symbols tends to an ordinary multiplication, and the solution of Eq.(3.17) reduces to the principal parts of the symbols. The principal part of the vertical slowness symbol corresponds to the vertical gradient of travel time, in accordance with the eikonal equation.

### 4 The generalized Bremmer coupling series

**The coupled system of integral equations**

Applying the operators with kernels Eq.(3.7) to Eq.(2.30) we obtain a coupled system of integral equations. In operator form, they are given by

\[
(\delta_{I,J} - \hat{K}_{I,J}) \hat{W}_{J} = \hat{W}_{I}^{(0)},
\]

(4.1)
in which \( \hat{W}^{(0)} \) denotes the incident, downward travelling field. In our configuration the domain of heterogeneity will be restricted to the slab \( (0, x^{\text{init}}) \), see Figure 1, and the excitation of the waves will be specified through an initial condition at the level \( x_{3} = 0 \), viz.

\[
\hat{W}_{1}^{(0)}(x_{\mu}, x_{3}) = \int_{x'_{\nu} \in \mathbb{R}^{1}} \hat{G}(x_{\mu}, x_{3}; x'_{\nu}, 0) \hat{W}_{1}(x'_{\nu}, 0) \, dx'_{\nu} \, dx'_{\lambda},
\]

(4.2)

\[
\hat{W}_{2}^{(0)}(x_{\mu}, x_{3}) = 0,
\]

(4.3)
in the range of interest, $x_3 \in [0, x_3^{exit}]$; the second equation reflects the assumption that there is no excitation below the heterogeneous slab. The integral operators in Eq.(4.1) are given by

$$\hat{W}_1(x_\mu, x_3) = \int_{\zeta=0}^{x_3^{exit}} \int_{x'_3 \in \mathbb{R}} \hat{g}^{(+)}(x_\mu, x_3; x'_\mu, \zeta)(\hat{T}\hat{W}_1)(x'_\mu, \zeta) \, dx'_\mu \, dx'_3 \, d\zeta, \quad (4.4)$$

$$\hat{W}_2(x_\mu, x_3) = \int_{\zeta=0}^{x_3^{exit}} \int_{x'_3 \in \mathbb{R}} \hat{g}^{(\pm)}(x_\mu, x_3; x'_\mu, \zeta)(\hat{R}\hat{W}_2)(x'_\mu, \zeta) \, dx'_\mu \, dx'_3 \, d\zeta, \quad (4.5)$$

$$\hat{W}_1(x_\mu, x_3) = \int_{\zeta=0}^{x_3^{exit}} \int_{x'_3 \in \mathbb{R}} \hat{g}^{(-)}(x_\mu, x_3; x'_\mu, \zeta)(\hat{R}\hat{W}_1)(x'_\mu, \zeta) \, dx'_\mu \, dx'_3 \, d\zeta, \quad (4.6)$$

$$\hat{W}_2(x_\mu, x_3) = \int_{\zeta=0}^{x_3^{exit}} \int_{x'_3 \in \mathbb{R}} \hat{g}^{(-)}(x_\mu, x_3; x'_\mu, \zeta)(\hat{T}\hat{W}_2)(x'_\mu, \zeta) \, dx'_\mu \, dx'_3 \, d\zeta. \quad (4.7)$$

They describe the interaction between the counter-propagating constituent waves.

We can represent the action of the one-sided Green’s kernels by product integrals, viz.,

$$\hat{W}_1^{(0)}(\cdot, x_3) = \left\{ \prod_{\zeta'=0}^{x_3^{exit}} \exp[-i\omega\hat{G}^{(+)}(\cdot, \zeta') d\zeta'] \right\} \hat{W}_1(\cdot, 0), \quad (4.8)$$

while

$$\hat{W}_1^{(0)}(\cdot, x_3) = \int_{0}^{x_3^{exit}} \left\{ \prod_{\zeta'=\zeta}^{x_3^{exit}} \exp[-i\omega\hat{G}^{(-)}(\cdot, \zeta') d\zeta'] \right\} \hat{W}_1(\cdot, \zeta) d\zeta, \quad (4.9)$$

and so on.

**Bremmer series**

If $\omega = -is$ (and $\alpha_s = i\sigma_s \in i\mathbb{R}$) with $s$ real and sufficiently large, the Neumann expansion can be employed to invert $(\delta_{j,j} - \hat{K}_{j,j})$ in Eq.(4.1). Such a procedure leads to the Bremmer coupling series,

$$\hat{W}_j = \sum_{j=0}^{\infty}(\hat{K}^j)_{j,j} \hat{W}_j^{(0)} = \hat{W}_j^{(0)} + (\hat{K}_1)_{j,j} \hat{W}_j^{(0)} + (\hat{K}_2^2)_{j,j} \hat{W}_j^{(0)} + \cdots. \quad (4.10)$$

To emphasize the physical nature of the expansion, we write

$$\hat{W}_j = \sum_{j=0}^{\infty} \hat{W}_j^{(j)}; \quad (4.11)$$

in which

$$\hat{W}_j^{(j)} = \hat{K}_{j,j} \hat{W}_j^{(j-1)} \quad \text{for} \quad j \geq 1, \quad (4.12)$$

can be interpreted as the $j$-times reflected or scattered wave. This equation indicates that the solution of Eq.(4.1) can be found with the aid of a recursive scheme.
A recursive scheme

To arrive at a recursive scheme, consider the \( j \)-times reflected constituent wave. We split the interval \([0, x_3^{\text{exit}}]\) into \( M \) thin slabs with thickness \( \Delta x_3 \). Set

\[
\hat{W}_1^{(j)}(\cdot, k) = \hat{I}_1^{(j)}(\cdot, k) + \hat{I}_1^{(j)}(\cdot, k),
\]

\[
\hat{W}_2^{(j)}(\cdot, k) = \hat{I}_2^{(j)}(\cdot, k) - \hat{I}_2^{(j)}(\cdot, k),
\]

\( j = 1, 2, \cdots \) and \( k = 0, 1, \cdots, M \), where (cf. Eq. (4.12))

\[
\hat{I}_1^{(j)}(\cdot, k) = (\hat{R}_1 \hat{W}_1^{(j-1)})(\cdot, k \Delta x_3),
\]

\[
\hat{I}_2^{(j)}(\cdot, k) = (\hat{R}_1 \hat{W}_2^{(j-1)})(\cdot, k \Delta x_3).
\]

Upon comparison with Eq. (4.9) we find that

\[
\hat{I}_1^{(j)}(\cdot, k) = \int_{\zeta=0}^{\Delta x_3} \left\{ \prod_{\zeta'=\zeta}^{\Delta x_3} \exp[-i \omega \hat{T}^{(-)}(\cdot, \zeta')] \right\} \hat{X}_1^{(j)}(\cdot, \zeta) \, d\zeta,
\]

with

\[
\hat{X}_1^{(j)}(\cdot, \zeta) = (\hat{T} \hat{W}_1^{(j-1)})(\cdot, \zeta).
\]

Similarly,

\[
\hat{I}_1^{(j)}(\cdot, k) = \int_{\zeta=0}^{\Delta x_3} \left\{ \prod_{\zeta'=\zeta}^{\Delta x_3} \exp[-i \omega \hat{T}^{(+)}(\cdot, \zeta')] \right\} \hat{X}_1^{(j)}(\cdot, \zeta) \, d\zeta,
\]

\[
\hat{I}_2^{(j)}(\cdot, k) = -\int_{\zeta=0}^{\Delta x_3} \left\{ \prod_{\zeta'=\zeta}^{\Delta x_3} \exp[-i \omega \hat{T}^{(-)}(\cdot, \zeta')] \right\} \hat{X}_2^{(j)}(\cdot, \zeta) \, d\zeta,
\]

\[
\hat{I}_2^{(j)}(\cdot, k) = -\int_{\zeta=0}^{\Delta x_3} \left\{ \prod_{\zeta'=\zeta}^{\Delta x_3} \exp[-i \omega \hat{T}^{(-)}(\cdot, \zeta')] \right\} \hat{X}_2^{(j)}(\cdot, \zeta) \, d\zeta,
\]

with

\[
\hat{X}_1^{(j)}(\cdot, \zeta) = (\hat{R} \hat{W}_1^{(j-1)})(\cdot, \zeta),
\]

\[
\hat{X}_2^{(j)}(\cdot, \zeta) = (\hat{R} \hat{W}_2^{(j-1)})(\cdot, \zeta),
\]

\[
\hat{X}_2^{(j)}(\cdot, \zeta) = (\hat{T} \hat{W}_2^{(j-1)})(\cdot, \zeta).
\]

In a configuration with constant density, we have \( \hat{R} = -\hat{T} \) and hence

\[
\hat{X}_2^{(j)}(\cdot, \zeta) = -\hat{X}_1^{(j)}(\cdot, \zeta) \quad \text{and} \quad \hat{X}_1^{(j)}(\cdot, \zeta) = -\hat{X}_2^{(j)}(\cdot, \zeta).
\]
To construct the recursion scheme, we carry out the following steps. Let $\hat{P}$ denote the thin slab *propagator* (cf. Eq. (3.13))

$$
\hat{P}(\cdot, k) = \left\{ \prod_{\zeta' = (k-1)\Delta x_3}^{k\Delta x_3} \exp[-i\omega \hat{T}^{(-)}(\cdot, \zeta') \, d\zeta'] \right\}.
$$

(4.25)

Then, using the semi-group property,

$$
\hat{T}^{(j)}_{1,1}(\cdot, k) = \left\{ \prod_{\zeta' = (k-1)\Delta x_3}^{k\Delta x_3} \exp[-i\omega \hat{T}^{(+)\zeta}(\cdot, \zeta') \, d\zeta'] \right\}

\int_{\zeta = 0}^{(k-1)\Delta x_3} \left\{ \prod_{\zeta' = \zeta}^{(k-1)\Delta x_3} \exp[-i\omega \hat{T}^{(+)\zeta'}(\cdot, \zeta') \, d\zeta'] \right\} \hat{X}^{(j)}_{1,1}(\cdot, \zeta) \, d\zeta,

+ \int_{\zeta = (k-1)\Delta x_3}^{k\Delta x_3} \left\{ \prod_{\zeta' = \zeta}^{k\Delta x_3} \exp[-i\omega \hat{T}^{(+)\zeta'}(\cdot, \zeta') \, d\zeta'] \right\} \hat{X}^{(j)}_{1,1}(\cdot, \zeta) \, d\zeta,
$$

(4.26)

which can be written as

$$
\hat{T}^{(j)}_{1,1}(\cdot, k) = \hat{P}(\cdot, k) \hat{T}^{(j)}_{1,1}(\cdot, k - 1) + \hat{Q}^{(j)}_{1,1}(\cdot, k),
$$

(4.27)

where

$$
\hat{Q}^{(j)}_{1,1}(\cdot, k) = \int_{\zeta = (k-1)\Delta x_3}^{k\Delta x_3} \left\{ \prod_{\zeta' = \zeta}^{k\Delta x_3} \exp[-i\omega \hat{T}^{(+)\zeta'}(\cdot, \zeta') \, d\zeta'] \right\} \hat{X}^{(j)}_{1,1}(\cdot, \zeta) \, d\zeta.
$$

(4.28)

Recursion relations similar to the one in Eq. (4.27) can be found for the other elements of $\hat{T}$, viz.,

$$
\hat{T}^{(j)}_{1,j}(\cdot, k) = \hat{P}(\cdot, k) \hat{T}^{(j)}_{1,j}(\cdot, k - 1) + \hat{Q}^{(j)}_{1,j}(\cdot, k) \quad \text{for} \quad k = 1, 2, \ldots, M
$$

$$
\hat{T}^{(j)}_{2,j}(\cdot, k) = \hat{P}(\cdot, k - 1) \hat{T}^{(j)}_{2,j}(\cdot, k + 1) + \hat{Q}^{(j)}_{2,j}(\cdot, k) \quad \text{for} \quad k = M - 1, M - 2, \ldots, 0.
$$

(4.29)

Here

$$
\hat{Q}^{(j)}_{1,j}(\cdot, k) = \int_{\zeta = (k-1)\Delta x_3}^{k\Delta x_3} \left\{ \prod_{\zeta' = \zeta}^{k\Delta x_3} \exp[-i\omega \hat{T}^{(+)\zeta'}(\cdot, \zeta') \, d\zeta'] \right\} \hat{X}^{(j)}_{1,j}(\cdot, \zeta) \, d\zeta,
$$

(4.30)

$$
\hat{Q}^{(j)}_{2,j}(\cdot, k) = \int_{\zeta = (k-1)\Delta x_3}^{k\Delta x_3} \left\{ \prod_{\zeta' = k\Delta x_3}^{\zeta} \exp[i\omega \hat{T}^{(-)\zeta'}(\cdot, \zeta') \, d\zeta'] \right\} \hat{X}^{(j)}_{2,j}(\cdot, \zeta) \, d\zeta.
$$

(4.31)
The initial values for the recursive scheme (4.29) are given by
\[ J^{(j)}_{\mu,0} = 0 , \quad J^{(j)}_{\mu,M} = 0 , \] (4.32)
again, for \( j = 1, 2, \cdots \).

**Numerical issues**

The implementation of the recursive scheme is as follows. It is initiated by the calculation of the incident field, \( \hat{W}^{(0)}_1 \), according to
\[ \hat{W}^{(0)}_1(\cdot, k \Delta z_3) = \hat{P}(\cdot, k) \hat{W}^{(0)}_1(\cdot, (k-1) \Delta z_3) \] for \( k = 1, 2, \cdots, M \),
(4.34)
see Figure 2, with initial condition
\[ \hat{W}^{(0)}_1(\cdot, 0) = \hat{W}_1(\cdot, 0) , \]
(4.35)
according to Eq.(4.2). During the forward propagation, at each of the discrete levels, \( \hat{X}^{(j)}_{J,3} \) are computed and stored; \( \hat{X}^{(j)}_{J,2} \) are set to zero. The procedure is continued by the backward propagation defined by the second recursion in Eq.(4.29). At each of the discrete levels, \( \hat{W}^{(j)}_3 \) is computed (Eq.(4.14)) and used to calculate \( \hat{X}^{(j)}_{J,2} \); the latter quantity is stored as before. The scheme continues to switch from backward to forward propagation based on the first recursion in Eq.(4.29), and so on.

To evaluate the elements of \( \hat{Q}^{(j)} \), we apply the trapezoidal rule. Then
\[ \hat{Q}^{(j)}_1(\cdot, k) \simeq \frac{1}{2} \Delta z_3 \left[ \hat{X}^{(j)}_{1,2}(\cdot, k \Delta z_3) + \hat{P}(\cdot, k) \hat{X}^{(j)}_{1,2}(\cdot, (k-1) \Delta z_3) \right] , \] (4.36)
\[ \hat{Q}^{(j)}_2(\cdot, k) \simeq \frac{1}{2} \Delta z_3 \left[ \hat{X}^{(j)}_{2,2}(\cdot, k \Delta z_3) + \hat{P}(\cdot, k) \hat{X}^{(j)}_{2,2}(\cdot, (k+1) \Delta z_3) \right] , \] (4.37)
which formulas are accurate \( O((\Delta z_3)^2) \).

The whole numerical scheme is elucidated in a flow chart, Figure 3. In the remainder of this paper, we will derive approximate, sparse matrix representations for \( \hat{\Gamma}, P, \hat{\Gamma}^{-1} \) and \( \hat{T} \). We will constrain ourselves to two dimensions, i.e., \( \mu, \nu, \cdots = 1 \).

5 **Sparse matrix representation of the propagator**

Consider the homogeneous one-way wave equation (cf. Eq.(2.30))
\[ \partial_3 \hat{W}^{(0)}_1 + i \omega \hat{T} \hat{W}^{(0)}_1 = 0 , \quad x_3 \in (0, z^{\text{crit}}_3) \]
(5.1)
which is satisfied by the leading-order term of the Bremmer series. In this section, we will derive a sparse matrix representation for its propagator, \( \hat{P}(\cdot, k) \). In the following approximations, we will maintain the hyperbolicity of the time-domain-equivalent equations, but we will deform the post-critical regime.
Approximating the vertical slowness leading to a partial differential equation

The principal part of the left vertical wave-slowness symbol is given by (cf. Eq. (3.17))

\[ \tilde{\gamma}_{pp}(x_1, x_3, \alpha_1) = [(c^{-2} - \alpha_1^2)^{1/2}]. \]

(5.2)

This principal part is equal to the vertical component of the ray vector or gradient of travel time. In the upper left corner of Figure 4, \( \tilde{\gamma}_{pp} \) is shown. The interval \( \alpha_1 \in [0, 1] \) corresponds to wave propagation under an angle of 0 to 90 degrees and is called the pre-critical region. The interval \( \alpha_1 \in (1, \infty) \) corresponds to the evanescent wave constituents and is called the post-critical region.

For our numerical scheme, we consider Thiele's third-order continued fraction's approximation of the left symbol (see Serrafin and De Hoop [30] and De Hoop and De Hoop [31])

\[ \tilde{\gamma}_{pp}^{III} = \frac{1}{c} \left[ 1 - \beta_3 c^3 \alpha_1^2 \right]^{-1} \left[ -\beta_1 c \alpha_1^3 + \beta_2 c^3 \alpha_1^4 \right]. \]

(5.3)

The wave front associated with this approximation is free from artificial singularities in a cone of propagation angles about vertical, unlike the second-order approximation that is commonly used. The operator associated with this left symbol is not symmetric nor self-adjoint. To create a symmetric Thiele approximation, we extract the local wave speed \( c \) from the square root expression; the vertical slowness operator becomes

\[ \tilde{\Gamma} \approx c^{-1/2} (1 + \Xi)^{1/2} c^{-1/2}, \]

(5.4)

where

\[ \Xi \equiv -c D_1^2 (c \cdot \cdot). \]

(5.5)

This approximation is valid for small values of the commutator \([c, D_1]\), which will be the case for media varying smoothly on the scale of the dominant wavelength.

The third-order Thiele approximation is now applied to the symbol of the operator \((1 + \Xi)^{1/2}\). This leads to a symmetric operator \(\tilde{\Gamma}^{III}\), the principal symbol of which still equals the expression in Eq. (5.3),

\[ \tilde{\Gamma}^{III} = c^{-1/2} \left( 1 + [1 + \beta_3 \Xi]^{-1} [\beta_1 \Xi + \beta_2 \Xi^2] \right) c^{-1/2}. \]

(5.6)

The internal structure of \(\tilde{\Gamma}^{III}\) is such that its symbol captures some of the contributions being the principal part. From now on we will freely omit the superscript \(^{III}\).

According to Thiele's formula, we have

\[ \beta_1 = 1/2, \quad \beta_2 = 1/8, \quad \beta_3 = 1/2. \]

(5.7)

The parameters, \( \beta_1 \), \( \beta_2 \), and \( \beta_3 \), however, can be adjusted by minimizing the difference between the exact spectral-domain vertical slowness and its continued-fraction approximation, defined by Eq. (5.3), with respect to the \( L_2 \)-norm, over all the propagating modes (i.e., the real slowness surface). In this minimization, we must be aware of the fact that the pole \((c^{-2} = c^{-2}\beta_1^{-1})\) introduced by the Thiele-type approximation will lie outside the pre-critical region of the \( \alpha_1 \)-plane. The optimization procedure can be viewed as a replacement of Thiele's continued fraction by Newton interpolation.

Using an optimization routine based on the simplex method [32], the following values are obtained

\[ \beta_1 = 0.526, \quad \beta_2 = 0.364, \quad \beta_3 = 0.825. \]

(5.8)
In case the propagation angles appearing in the wave field are restricted, the optimization should be carried out accordingly. This type of optimization was also considered by Lee and Suh [33] and followed by Bunks [34].

Figure 4 shows the exact and approximated vertical slownesses with $\beta_1$, $\beta_2$ and $\beta_3$ as in Eq.(5.7) and as in Eq.(5.8), respectively. In the second row of figures the differences with the exact slowness are drawn. In the bottom row the real part of the vertical slownesses as a function of complex-valued $\alpha_1$ is shown. In these pictures we recognise the singularities in the vertical slownesses: a branch cut from $\alpha_1 = 1$ to $+\infty$ in the exact vertical slowness and poles at $\sqrt{2}$ and 1.10 for the two approximated vertical slownesses, respectively. Since no singularities are present on the imaginary axis, the approximation is accurate for a large range of imaginary values of $\alpha_1$.

Fundamental differences between the approximate and exact vertical slownesses occur near the singularities in the horizontal slowness plane. Since it is impossible to approximate a branch cut with a finite set of poles from a rational approximation, we have to restrict ourselves to approximating the pre-critical wave propagation as complete as possible. The influences of the artificial post-critical wave propagation must then be suppressed. A dissipation trick can be designed for that purpose.

The dissipation trick

To suppress artifacts, and aliasing (which may arise from the discretisation to be carried out in the horizontal direction), from the large horizontal wave number components, we may replace the real frequency $\omega$ by a complex one,

$$\omega' = \omega (1 - i\Omega),$$

just in the expression for $i\omega\hat{f}$ (and later in $M_i$ and $M'$, see equations (5.23)). In a homogeneous medium, it implies per Fourier component or plane wave, $A(x_3, \alpha_1, \omega) \exp(-i\omega\alpha_1 x_1)$, an amplification factor of the form

$$|A(x_3 + \Delta x_3, \alpha_1, \omega)/A(x_3, \alpha_1, \omega)| = \exp \left[ -\frac{\omega \Delta x_3}{2c Q(\alpha_1, \omega, i\Omega)} \right]. \quad (5.9)$$

The amplification factor is attenuative in nature. The $Q$-factor is given by

$$Q(\alpha_1, \omega, i\Omega) = \frac{1}{\Im \{2c \hat{f}(\alpha_1; \omega(1 - i\Omega))\}}, \quad (5.10)$$

and is plotted in Figure 5. Here, $\Omega$ is assumed to be very small ($\Omega \ll 1$). This particular complexification results in a local bandlimitation filter, suppressing the artifacts associated with the post-critically propagating modes.

The comoving frame of reference

To reduce discretisation artifacts in the vertical derivative, the numerical computations are done in a comoving frame of reference, traveling in the direction of preference. The accuracy is improved, since this transformation enables us to take into account the leading phase shift exactly. The change of frame yields (see Claerbout [2])

$$\hat{h}(x_k, \omega) = \exp[i\omega \tau(x_k)] \hat{h}^{(0)}_2(x_k, \omega), \quad (5.11)$$
in which

$$\tau(x_1) = \int_{\zeta=0}^{x_1} c^{-1}(x_1, \zeta) \, d\zeta$$

(5.12)

is the local vertical travel time. Substituting Eq.(5.11) into Eq.(5.1), leads to

$$\partial_3 \hat{h} = i\omega \left[ \exp(i\omega \tau) \hat{h} \cdot \exp(-i\omega \tau) - c^{-1} \right] \hat{h} = 0 .$$

(5.13)

**Discretisation in the transverse direction**

First, we introduce the discretisation in the transverse direction, of the partial differentiation $\partial_3^2$. (In three-dimensional wave propagation this would be the two-dimensional Laplace operator.) Our starting point is a low-order, implicit, finite-difference approximation. The spectrum or symbol of the resulting operator is a rational function in horizontal slowness. The medium has to be smooth on the scale of the size of the operator. A uniform grid is employed in the horizontal direction. The horizontal sampling interval is denoted as $\Delta x_1$.

The discretisation of the one-dimensional Laplace operator is formulated in terms of recursive filters based on nearest-neighbours interaction (see Mitchell and Griffiths [35]). Our recursive filter acting on a function $\hat{h}$ is defined through

$$(1 + a_2 \delta_1^2)(\partial_1^2 \hat{h}) = (\Delta x_1)^{-2} a_2 \delta_1^2 \hat{h} ,$$

(5.14)

where

$$\delta_1^2 \hat{h} = \hat{h}(x_1 + \Delta x_1) - 2\hat{h}(x_1) + \hat{h}(x_1 - \Delta x_1) .$$

(5.15)

In Eq.(5.14), 1 denotes the identity and $(\partial_1^2 \hat{h}) = (\delta_1^2 \hat{h})$ represents the approximate Laplace operator having acted on $\hat{h}$. Using a Taylor series expansion of $\hat{h}$ at $x_1 = \Delta x_1$ about $x_1$, the following values of $a_1$ and $a_2$ are found (see Mitchell and Griffiths [35])

$$a_1 = 1 , \quad a_2 = 1/12 .$$

(5.16)

Then

$$\langle \partial_1^2 \rangle = \delta_1^2 + \mathcal{O}((\Delta x_1)^4) .$$

(5.17)

For a finite-bandwidth solution generated with a given sampling interval, however, we leave this order estimate, and use $a_2$ as a parameter to improve the overall accuracy. Requiring that in the limit $\Delta x_1 \downarrow 0$ the spectrum of $\langle \partial_1^2 \rangle$ tends to the spectrum of the Laplace operator up the lowest order, leads necessarily to $a_1 = 1$. The parameter $a_2$ is then determined by minimising non-linearly the difference between the approximate and the exact Laplace operator spectra with respect to the $L_2$-norm over the Nyquist interval. Thus a more accurate approximate spectrum is obtained over the spatial bandwidth as a whole. Using an optimization routine based on the simplex method [36], the following value for $a_2$ is obtained

$$a_2 = 0.130 .$$

(5.18)

It is noted, however, that if it is known beforehand that the actual spatial bandwidth of the wave field to be extrapolated is limited by a horizontal wave number less the Nyquist wave number $k_{1,Nyq} = \pi/\Delta x_1$, the optimization should be carried out over this sub-band only.
Figure 6 illustrates the spectra of the exact and the approximate Laplace operators with \(a_1\) and \(a_2\) as in Eq. (5.18) and as in Eq. (5.16), respectively. The approximated Laplace operator has poles at

\[
\alpha_1^f = (\omega \Delta x_1)^{-1} \left\{ \pm \arccos \left[ 1 - \frac{1}{2} a_2^{-2} \right] + 2m\pi \right\},
\]

\((m = 0, \pm 1, \pm 2, \cdots)\). In practice \(a_2 < 1\), and these poles are situated in the complex \(\alpha_1\)-plane at \(\text{Re}\{\alpha_1\} = \pi/(\omega \Delta x_1)\) (determined by the Nyquist theorem), far away from the real axis.

The poles of the approximate Laplace operator carry over in the approximation of the vertical slowness operator. The vertical slowness symbol becomes periodic in accordance with the Nyquist theorem, and the Laplace poles are slightly shifted:

\[
\alpha_1^f \to (\omega \Delta x_1)^{-1} \left\{ \pm \arccos \left[ 1 - \frac{1}{2} (a_2^{-2} + \omega \Delta x_1 \beta_3) \right] + 2m\pi \right\},
\]

see Figure 7. For example, taking \(a_2 = 0.130\), \(\beta_3 = 0.825\), and a sampling rate of 8 points per wavelength, the poles are located at

\[
\alpha_1^f = \pm 1.1023 + 8m \quad \text{and} \quad \alpha_2^f = \pm (4 - 2.4010 i) + 8m.
\]

The system of ordinary differential equations

Substituting \((\delta_1^2)\), i.e., Eq. (5.14) into Eq. (5.5) and the result into the one-way wave equation (5.13) with vertical slowness (5.6), we obtain a system of ordinary differential equations of the form

\[
\dot{\mathcal{M}} \mathcal{B} \hat{h} + i \omega \mathcal{M}^\prime \hat{C}^{-1} \hat{h} = 0.
\]

The operators involved are \(\hat{\mathcal{M}} = \hat{\mathcal{M}}(x_3, \omega; x_1, \delta_1^2, \Delta x_1)\), \(\hat{\mathcal{M}}^\prime = \hat{\mathcal{M}}^\prime(x_3, \omega; x_1, \delta_1^2, \Delta x_1)\) and \(\hat{\mathcal{C}} = \hat{\mathcal{C}}(x_3, \omega; x_1, \delta_1^2, \Delta x_1)\); they are sparse, viz., 3, 5 and 3 bands respectively, and are given by

\[
\hat{\mathcal{M}} : = \left( 1 + a_2 \delta_1^2 \right) \left( c^{-1/2} \exp(-i \omega \tau) \cdot \right) + \beta_3 \beta_3 c^{-2} \delta_1^2 \left( c^{3/2} \exp(-i \omega \tau) \cdot \right),
\]

\[
\hat{\mathcal{M}}^\prime : = \beta_3 \beta_3 c^{-2} \left( 1 + a_2 \delta_1^2 \right) (\delta_1^2 \cdot) + \beta_2 \beta_3 c^{-4} \delta_1^2 \left( c^2 \delta_1^2 \cdot \right),
\]

while

\[
\hat{\mathcal{C}} : = \exp(i \omega \tau) \cdot c^{-1/2} \left( 1 + a_2 \delta_1^2 \right) \cdot,
\]

with sampling rate parameter

\[
\beta_3 = (\omega \Delta x_1 / c)^{-2}.
\]

Integration of the vertical derivative

The solution of equation (5.22) can formally be written as a product integral. This product integral will have to be evaluated, recursively, for every propagation step with size \(\Delta x_3\). The product integral is given by (see also Eq. (3.6))

\[
\hat{h}(., x_3 + \Delta x_3) = \prod_{\zeta = x_3}^{x_3 + \Delta x_3} \exp \left[ -i \omega \left( \mathcal{M}^{-1} \mathcal{M}^\prime \hat{C}^{-1} : (., \zeta) \cdot d \zeta \right) \hat{h}(., x_3) \right].
\]
It is approximated by

\[ \hat{h}(., x_3 + \Delta x_3) \approx \exp \left[ -i \omega \Delta x_3 \left( \hat{M}^{-1} \hat{M}' \hat{C}^{-1} \right)(., x_3 + \frac{1}{2} \Delta x_3) \right] \hat{h}(., x_3). \]  

(5.27)

To speed up the computations, the exponent is approximated by the solution of a unitary matrix equation. The \((n,n)\)-Padé approximants can be used for this purpose. They are given by (see Numerical Recipes [37])

\[ \exp(Z) \simeq \begin{cases} 
\frac{1 + \beta_0 Z}{1 - \beta_0 Z}, & \text{lowest order: (1,1),} \\
\left( \frac{1 + \beta_1 Z}{1 - \beta_1 Z} \right) \left( \frac{1 + \beta_0 Z}{1 - \beta_0 Z} \right), & \text{higher order: (2,2),}
\end{cases} \]  

(5.28)

where

\[ Z \equiv \Delta x_3 \left. \frac{\partial}{\partial x_3} \right| x_3 + \frac{1}{2} \Delta x_3, \]  

(5.29)

is the derivative that appears in Eq.(5.22). The higher the order of the approximation, the larger the stepsize \(\Delta x_3\) can be. For stability considerations of the associated higher-order implicit finite-difference schemes, see Weinlud [38].

According to Padé's formula, we have

\[ \beta_0 = \frac{1}{2}. \]  

(5.30)

Instead of taking this exact value, we let \(\beta_0\) be a free parameter for an optimization procedure. Minimizing the difference for the pre-critical wave constituents, gives

\[ \beta_0 = 0.766 \quad \text{for 2 points per wavelength,} \]  

(5.31)

\[ \beta_0 = 0.540 \quad \text{for 5 points per wavelength,} \]  

(5.32)

\[ \beta_0 = 0.510 \quad \text{for 10 points per wavelength.} \]  

(5.33)

The \((1,1)\)-Padé approximation of the exponent yields the well-known Crank-Nicholson scheme (see Richtmeyer and Morton [39]), in our notation given by

\[ (\hat{M} \hat{C})(x_3 + \frac{1}{2} \Delta x_3) (\delta \hat{h}'(x_3)) = 
\]

\[ (-i \beta_0 \omega \Delta x_3) \hat{M}'(x_3 + \frac{1}{2} \Delta x_3) (\hat{h}'(x_3) + \hat{h}'(x_3 + \Delta x_3)), \]  

(5.34)

with

\[ (\delta \hat{h}'(x_3)) = \hat{h}'(x_3 + \Delta x_3) - \hat{h}'(x_3), \]  

(5.35)

where

\[ \hat{h}'(\zeta) = \hat{C}^{-1}(x_3 + \Delta x_3) \hat{h}(\zeta) \quad \text{for} \; \zeta \in (x_3, x_3 + \Delta x_3]. \]  

(5.36)

The scheme Eq.(5.34) should be interpreted as a centered difference approximation of the vertical derivative, and hence is accurate up to \(O((\Delta x_3)^2)\).
Propagator matrix

To arrive at an explicit matrix representation for the one-way propagator, we introduce the array,

\[ \hat{h}_t(x_3, \omega) = \hat{h}(x_{1i}, x_3, \omega) , \]

(5.37)

where

\[ x_{1i} = i \Delta x_i , \]

(5.38)

and \( i = 1, \ldots, N_x \), labels the samples in the transverse direction. In this notation, with the aid of Eqs.(5.34) and (5.36), the approximate one-way wave propagation can finally be written in the matrix form

\[ \hat{h}_t(x_3 + \Delta x_3, \omega) = \hat{C}_{ij}(\hat{A}^{-1})_{jk} \hat{A}_{kl}(\hat{C}^{-1})_{lm} \biggl|_{x_3 + \frac{1}{2} \Delta x_3} \hat{h}_m(x_3, \omega) , \]

(5.39)

where the matrices \( \hat{A}, \hat{A}' \) and \( \hat{C} \) are functions of \( x_3 \) and \( \omega \), and are constructed from the operators \( \hat{M}, \hat{M}' \) and \( \hat{C} \):

\[ \hat{A} = \hat{M} \hat{C} + i \beta_9 \omega \Delta x_2 \hat{M}' , \]

(5.40)

\[ \hat{A}' = \hat{M} \hat{C} - i \beta_9 \omega \Delta x_2 \hat{M}' . \]

(5.41)

The matrices \( \hat{A} \) and \( \hat{A}' \) contain 5 non-vanishing bands, while \( \hat{C} \) has 3 non-vanishing bands only. Observe that

\[ \hat{A}'(\omega) = \overline{\hat{A}(-\omega)} , \]

(5.42)

thus \( \hat{A}'(\omega) = \overline{\hat{A}(\omega)} \) if \( \omega \) is real-valued (\( \Omega = 0 \)).

The numerical scheme following Eq.(5.39), involves two matrix-vector multiplications and twice solving a matrix equation per propagation step:

\[ \hat{C}_{im} \hat{h}_m'(x_3, \omega) = \hat{h}_t(x_3, \omega) , \]

(5.43)

\[ \hat{h}_m''(x_3, \omega) = \hat{A}_{kl} \hat{h}_l'(x_3, \omega) , \]

(5.44)

\[ \hat{A}_{jk} \hat{h}_k'(x_3 + \Delta x_3, \omega) = \hat{h}_m''(x_3, \omega) , \]

(5.45)

\[ \hat{h}_t(x_3 + \Delta x_3, \omega) = \hat{C}_{ij} \hat{h}_j'(x_3 + \Delta x_3, \omega) . \]

(5.46)

The first and last steps adjust for the change to the comoving frame of reference. Introduce the diagonal matrix \( \hat{T} \), representing the change of frame, with diagonal elements (cf. Eq.(5.11))

\[ \hat{T}_{mm}(x_3) = \exp[i \omega \tau(x_{1m}, x_3)] . \]

(5.47)

Then the propagator matrix (cf. Eqs.(4.25) and (5.39)) is given by

\[ \langle \hat{P}(k+1) \rangle_{im} = \hat{T}(x_3 + \Delta x_3)^{-i \omega} \hat{C}_{ij}(\hat{A}^{-1})_{jk} \hat{A}_{kl}(\hat{C}^{-1})_{lm} \biggl|_{x_3 + \frac{1}{2} \Delta x_3} \hat{T}(x_3)^{m'm} . \]

(5.48)

Appendix A contains the FORTRAN code for constructing the various matrices. Several techniques exist for carrying out the matrix inversions [40]. The inversion of the tridiagonal matrix \( \hat{C} \) in Eq.(5.43) is carried out by forward/backward substitution (see Numerical Recipes [37]).
6 Transverse transparent boundaries

Our discretisation coexists with applying periodic boundary conditions in the transverse direction. It is standard practice to make the boundaries of the computational cell absorbing, to simulate an unbounded configuration instead. Several approaches have been developed for this adjustment; we mention the work of Clayton and Engquist [41] based on the parabolic equation in the transverse directions. Competing approaches are the Transparent Boundary Condition by Hadley [42], and the one by Berenger [43] based on a perfectly matched layer for absorption at the boundary. We follow the methodology described by Arai et al. [44]), which is a generalization of Hadley’s approach. We employ its simplest form, the so-called Robin conditions (see Vassallo [45]).

The Robin boundary condition employs the one-way wave equations (as in Eq.(5.1)) in both transverse directions at the edges of the computational domain,

$$\partial_t \hat{h} + i\omega \eta(\pm) \hat{h} = 0, \quad \text{at } x_1 (\pm) 0 \quad \text{and at } x_1 (\pm) (N_{x_1} + 1)\Delta x_1. \quad (6.1)$$

By choosing $\eta(\pm)$ adaptively, we will show that these boundary conditions can be made highly transparent.

The key matrices in our numerical scheme consist of 5 bands. Thus, we require estimates of the field at two discretisation points outside the computational domain, in accordance with transparent boundary conditions. The estimate of the field at the furthest sample points will be based on the additional constraint,

$$\partial_t^2 \hat{h} + i\omega \eta(\pm) \partial_t \hat{h} = \text{invariant}, \quad \text{at } x_1 (\pm) 0 \quad \text{and at } x_1 (\pm) (N_{x_1} + 1)\Delta x_1. \quad (6.2)$$

We will illustrate the estimation procedure at the left boundary at $x_1 = \Delta x_1$. Then the field samples $\hat{h}_0$ and $\hat{h}_{-1}$ have to be determined. In terms of a lowest-order finite-difference representation, Eq.(6.1) becomes,

$$\hat{h}_1 - \hat{h}_0 + i\omega \Delta x_1 \eta(\pm) \hat{h}_0 = 0. \quad (6.3)$$

This leads to

$$\hat{h}_0 = (1 - i\omega \Delta x_1 \eta(\pm))^{-1} \hat{h}_1. \quad (6.4)$$

The constraint (6.2) is used to construct $\hat{h}_{-1}$, with the finite-difference representation

$$\frac{\hat{h}_2 - 2\hat{h}_0 + \hat{h}_{-1}}{(\Delta x_1)^2} + i\omega \eta(\pm) \frac{\hat{h}_0 - \hat{h}_{-1}}{\Delta x_1} = \frac{\hat{h}_2 - 2\hat{h}_1 + \hat{h}_0}{(\Delta x_1)^2} + i\omega \eta(\pm) \frac{\hat{h}_1 - \hat{h}_0}{\Delta x_1}, \quad (6.5)$$

with the solution

$$\hat{h}_{-1} = \frac{\hat{h}_2 + \hat{h}_1 [i\omega \Delta x_1 \eta(\pm) - 3 + (1 - i\omega \Delta x_1 \eta(\pm))^{-1}(3 - 2i\omega \Delta x_1 \eta(\pm))] - (3 - 2i\omega \Delta x_1 \eta(\pm))]}{1 - i\omega \Delta x_1 \eta(\pm)}. \quad (6.6)$$

The numerical scheme associated with the Robin boundary condition is

- neutrally stable if $\text{Im}\{\eta(\pm)\} = 0$,
- stable if $\text{Im}\{\eta(\pm)\} < 0$,
- unstable if $\text{Im}\{\eta(\pm)\} > 0$.
The boundary conditions reduce to the Neumann boundary condition for \( \eta^{(-)} = 0 \), and to the Dirichlet boundary condition for \( \eta^{(-)} \to \infty \).

We adapt the value of \( \eta^{(\pm)} \) by assuming that the field behaves like a plane wave \( \sim \exp(-i\omega \eta^{(\pm)} x_1) \) near the computational boundaries. At the previous propagation step, the horizontal slowness associated with this plane wave can be estimated; at the current step, that value of \( \eta^{(\pm)} \) can then be applied. At the left boundaries, with Eq. (6.3), we get

\[
\eta^{(-)} \simeq (i \omega \Delta x_1)^{-1} \left( 1 - \frac{\hat{h}_2}{\hat{h}_1} \right)^{-1}.
\]

(6.7)

If the plane wave were to travel inward (Re\( \{\eta^{(-)}\} > 0 \) at the left boundary), \( \eta^{(\pm)} \) is reset to its imaginary part Im\( \{\eta^{(\pm)}\} \). In this way, inward travelling waves are attenuated. The effectiveness of the transparent boundary conditions can be determined from the graph of reflection coefficient associated with a plane-wave hitting the boundary, see Clayton and Engquist [41]. Keys [46] extended the one-way absorbing boundary equation to the case where two distinguishable plane waves hit the computational boundary simultaneously.

In our numerical scheme the transparent boundary conditions are translated into an adjustment of the outer elements of the matrices \( A' \) and \( A \). For example, the left boundary elements of \( A \) become (cf. Eqs. (6.4)-(6.6))

\[
A_{1,1} \rightarrow A_{1,1} \times (1 - i \omega \Delta x_1 \eta^{(-)})^{-1} \times A_{1,0} +
\]

\[
(\omega \Delta x_1 \eta^{(-)} - 3 + (1 - i \omega \Delta x_1 \eta^{(-)})^{-1} (3 - 2 i \omega \Delta x_1 \eta^{(-)})) A_{1,-1},
\]

(6.8)

\[
A_{1,2} \rightarrow A_{1,2} \times (1 - i \omega \Delta x_1 \eta^{(-)})^{-1} A_{1,-1},
\]

(6.9)

\[
A_{2,1} \rightarrow A_{2,1} + (1 - i \omega \Delta x_1 \eta^{(-)})^{-1} A_{2,0}.
\]

(6.10)

7 The vertical phase and group slownesses

We recall the well-known concepts of vertical phase and group slownesses \( \gamma^{ph} \) and \( \gamma^g \), for a real-valued plane wave constituent:

\[
\mathbf{\int} \int_{\omega = -\infty}^{\infty} \mathbf{\hat{w}}_0(\omega) \cos[\omega(\tau - \gamma^{ph} x_3)] d\omega \]

\[
\simeq \cos[\omega_0(\tau - \gamma^{ph}(\omega_0) x_3)] \text{sinc}[\Delta \omega/2(\tau - \gamma^{g}(\omega_0) x_3)],
\]

(7.1)

where

\[
\tau = t - \alpha_1 x_1,
\]

(7.2)

if

\[
\mathbf{\hat{w}}_0(\omega) = \frac{1}{\Delta \omega} \chi_{\omega_0 - \Delta \omega/2, \omega_0 + \Delta \omega/2} \chi_{\omega}(\omega),
\]

(7.3)

with \( \Delta \omega/\omega_0 \) small. This describes the distortion of a modulated sinc-pulse traveling over a distance \( x_3 \). Now, we will derive expressions for \( \gamma^{ph} \) and \( \gamma^g \). The analysis of the accuracy of finite-difference propagation schemes with phase and group velocities can be found in Trefethen [13]. Beaumont et al. [14] and Holberg [13, 16] assessed and improved the accuracy of such schemes.
General considerations

Substitute a plane wave constituent into Eqs.(5.11) and (5.46):

\[ W_1^{(0)}(x_k, \omega) = \exp(-i\omega \delta^ph \Delta x_3) \exp[-i\omega \alpha_1 \Delta x_1] , \]  

or, in the moving frame of reference,

\[ \hat{h}(x_k, \omega) = \exp(i\omega \tau) \exp(-i\omega \delta^ph \Delta x_3) \exp[-i\omega \alpha_1 \Delta x_1] . \]  

We then obtain the amplification factor of the finite-difference scheme (see Richtmeyer and Morton [39]),

\[ a^{-1}a' = \exp[-i\omega \delta^ph \Delta x_3] \exp[i\omega \gamma^{-1} \Delta x_3] , \]  

in which \( \delta^ph \), \( a \) and \( a' \) are the left symbols of the numerical representation for \( \hat{\Gamma} \), and \( A \) and \( A' \), respectively. For a homogeneous medium (constant \( c \)), we find

\[ a = A_{i,i} + 2A_{i,i-1} \cos(\omega \alpha_1 \Delta x_1) + 2A_{i,i+2} \cos(2\omega \alpha_1 \Delta x_1) , \]  

independent of \( i \), together with a similar expression for \( a' \), replacing \( A \) by \( A' \). For real-valued \( \omega \), we have \( a'(\omega) = a(-\omega) \), thus \( a^{-1}a' = 1 \) in a homogeneous medium. The symbol of the finite difference scheme is the vertical phase slowness \( \delta^ph \) given by

\[ \delta^ph(\alpha_1, \omega) = c^{-1} - \frac{\log(a^{-1}a')}{i\omega \Delta x_3} , \]  

which reduces, if \( \omega \) is real-valued, to

\[ \delta^ph(\alpha_1, \omega) = c^{-1} + \frac{2}{\omega \Delta x_3} \arg(a) . \]  

From the vertical phase slowness the vertical group slowness \( \delta^g \) can be derived, viz.,

\[ \delta^g = \partial_\omega(\omega \delta^ph) = \delta^ph + \omega(\partial_\omega \delta^ph) . \]  

The term \( (\partial_\omega \delta^ph) \) can be identified with the delay of the envelope per wavelength. Together with Eq.(7.6), this leads to

\[ \delta^g(\alpha_1, \omega) = c^{-1} - i(\Delta x_3)^{-1} \{ a^{-1} \partial_\omega a - (a')^{-1} \partial_\omega a' \} , \]  

and reduces, if \( \omega \) is real-valued, to

\[ \delta^g(\alpha_1, \omega) = c^{-1} - 2(\Delta x_3)^{-1} \text{Im}\{ a^{-1} \partial_\omega a \} . \]  

It is observed that \( \delta^ph \) and \( \delta^g \) are functions of the horizontal and vertical sampling rates per wavelengths, \( c/(\omega \Delta x_1) \) and \( c/(\omega \Delta x_3) \). Note that \( \delta^g - \gamma \) is indicative for the numerical anisotropy, whereas \( \delta^ph - \delta^g \) is indicative for the numerical dissipation.
Optimization

The vertical phase and group slownesses are functions of our parameters \(\{\beta_1, \beta_2, \beta_3, a_2, \beta_9, \Omega\}\) arising from the various approximations made to arrive at a sparse matrix representation for the propagator. So far, we have found two distinct parameter sets: one arising from consistent rational expansions, and one arising from step-wise optimization. For a sampling rate of 3 points per wavelength in both the horizontal and vertical directions \((\Delta x_1 = \Delta x_3 = 0.2\lambda)\) the phase and group slownesses for the first parameter set are shown in Figure 8. The exact vertical slowness is shown in the same graph. Significant deviations are observed for large propagation angles. The parameter set corresponding to the step-wise optimization leads to a better result (Figure 9). In the step-wise optimization, however, it is hard to control the movement of the poles in the complex horizontal slowness plane, introduced by the approximations. One can control this movement in an optimization scheme of the phase and group slownesses for all parameters together; then one expects a more accurate result as well.

The optimization of the phase and group slownesses results in a minimization of numerical anisotropy and numerical dissipation for the pre-critical wave constituents. In the optimization procedure special attention must be paid to the contraint that the poles of a stay outside the pre-critical region in the \(\alpha_1\)-plane. The overall optimization is carried out for an objective function given by the sum of the squared differences between \(\hat{\gamma}^g\) and \(\hat{\gamma}\), and \(\hat{\gamma}^{ph}\) and \(\hat{\gamma}\). The results are illustrated in Figure 10, as before for 5 points per wavelength. From numerical tests, we find that the accuracy remains more or less the same when we vary the transverse sampling rate from 6 up to 2.5 points per wavelength, keeping the vertical sampling rate the same. The procedure is even less sensitive to the vertical sampling rate.

So far, the optimization procedure was focussed on the real slowness surface. However, as we have seen in the preceding sections, the approximations lead to artifacts in the post-critical regime. In fact, the post-critical constituents have been mapped on propagating constituents (slow waves). We have designed a dissipation trick to attenuate these constituents. Besides this trick, some of the artificial constituents are forced to leave the computational domain through the transparent boundary conditions. The dissipation trick induces a complexification of the parameters used in the optimization, and hence it seems to be natural to complexify the optimization procedure accordingly (see also Collins [47]). Pursuing this complexification, the optimization procedure is extended with the constraint that the post-critical power is attenuated strongly while the attenuation/amplification in the pre-critical region is kept minimal (less than 3%). The resulting parameter set is shown in the fourth column of Table 1. The first column of this table represents the expansion values, and the second column represents the values obtained by step-wise optimization. The accuracy of the complex parameter set is illustrated in Figure 11.

We note, however, that the complexification of parameters is not always desired. For example, in the application of our numerical scheme to long-range propagation in low-losy waveguides, the accumulative power loss is of key importance. In such configurations, we set \(\Omega = 0\) and keep the parameters real. The optimization leads to the third column of Table 1.

The poles arising in vertical phase slowness, for the different parameter sets, are given in Table 2.

A numerical example

We illustrate the final, optimized, one-way propagation by computing the wave field excited by a line source in a homogeneous medium. The wave speed of the medium is chosen to be \(c = 1900\)m/s. The numerical grid consists of 99 points along the \(x_1\)-direction and 60 points along the \(x_2\)-direction. The discretisation step is 16m in both directions. Given a source-signature with trapezoidal spectrum with corner frequencies 10, 20, 25, and 40Hz, we encounter sampling rates of 2.97 to 11.9 points per wavelength. The line source is located at \(x_1 = 800\)m, \(x_2 = 0\)m, and we will show snapshots of the pressure field at
0.3s. The source is a vertical force. For this configuration we have analytic expressions.

In Figure 12 the snapshot is shown using the first parameter set, while in Figure 13 the snapshot for the fourth parameter set is shown. The left parts of the figures \(0 < x_1 < 800\) represent the numerical solutions, while the right parts \(800 < x_1 < 1600\) represent the exact solution. The errors occurring in Figure 13 are illustrated in Figure 14, the absolute ones for \(0 < x_1 < 800\) and the relative ones for \(800 < x_1 < 1600\). From this figure we find that the relative error at the numerical wave front is less than 2.5% up to angles of approximately 65 degrees. Repeating this accuracy analysis for the other parameter sets, we conclude that the approximations are acceptable up to 20 degrees, 40 degrees, and 80 degrees for the standard parabolic approximation, parameter set 1, and parameter set 3, respectively. The third parameter set is more accurate for higher angles than the fourth parameter set; however, post-critical artifacts remain apparent in the numerical results obtained with this real-valued parameter set.
8 Sparse matrix representations of the (de)composition operators

The computation of the generalized Bremmer coupling series starts with the decomposition of the initial field into up- and downgoing constituents, see also Eq. (2.29)

\[
\begin{pmatrix}
    \tilde{W}_1 \\
    \tilde{W}_2
\end{pmatrix} = \frac{1}{2} \begin{pmatrix}
    \hat{A}^{-1}_{1,2} & \hat{\Gamma}^{-1} \\
    \hat{A}^{-1}_{1,2} & -\hat{\Gamma}^{-1}
\end{pmatrix}
\begin{pmatrix}
    \tilde{p} \\
    \tilde{v}_3
\end{pmatrix}.
\]

Upon completing the calculation of a sufficiently large number of terms of the Bremmer coupling series, the constituents are recomposed into observables, see also Eq. (2.28)

\[
\begin{pmatrix}
    \tilde{p} \\
    \tilde{v}_3
\end{pmatrix} = \begin{pmatrix}
    \hat{A}_{1,2} & \hat{A}_{1,2} \\
    \hat{\Gamma} & -\hat{\Gamma}
\end{pmatrix}
\begin{pmatrix}
    \tilde{W}_1 \\
    \tilde{W}_2
\end{pmatrix}.
\]

In these procedures, the operator \( \hat{\Gamma} \) and its inverse must be computed. Here, using the results of Section 5, we will derive sparse matrix representations for these operators.

The composition operator

The composition operator contains the vertical slowness operator. To find a sparse matrix representation for this operator, we use the same approximation as in Eq. (5.6) for one-way wave propagation. We write, however, the fraction in a slightly different way,

\[
\hat{\Gamma}^{III} = c^{-1/2} \{1 + \beta_4 \Xi + [1 + \beta_3 \Xi]^{-1} \beta_5 \Xi\} c^{-1/2},
\]

with

\[
\beta_4 = \frac{\beta_2}{\beta_3}, \quad \beta_5 = \beta_1 - \frac{\beta_2}{\beta_3}.
\]

We set

\[
\hat{\Gamma}^{III} = \hat{\Gamma}_1^{III} + \hat{\Gamma}_2^{III},
\]

with

\[
\hat{\Gamma}_1^{III} = c^{-1/2} \{1 + \beta_4 \Xi\} c^{-1/2}, \quad \hat{\Gamma}_2^{III} = c^{-1/2} \{[1 + \beta_3 \Xi]^{-1} \beta_5 \Xi\} c^{-1/2}.
\]

Substituting discretisation Eq. (5.14) into Eq. (5.5) and the result into these operators, yields

\[
(1 + a_2 \delta_i^2) \left( c^{-1/2} (\hat{\Gamma}_1^{III} \hat{h}) \right) = (1 + a_2 \delta_i^2) c^{-2} + \beta_4 \beta_8 c^{-2} \delta_i^2 \left( c^{1/2} \hat{h} \right),
\]

\[
(1 + a_2 \delta_i^2) c^{-2} + \beta_3 \beta_8 c^{-2} \delta_i^2 \left( c^{1/2} (\hat{\Gamma}_2^{III} \hat{h}) \right) = \beta_3 \beta_8 c^{-2} \delta_i^2 \left( c^{1/2} \hat{h} \right).
\]

The two equations contain tridiagonal matrices only. Hence, they can be solved rapidly with the forward/backward substitution procedure (see Numerical Recipes [37]).
The decomposition operator

The decomposition operator essentially contains the inverse of the vertical slowness operator. Using approximation (5.6) as before, we find that

\[ c^{1/2} \tilde{\Gamma}^{III} c^{1/2} = \beta_2 [1 + \beta_3 \Xi^{-1}] [\beta_2^{-1} + \beta_2^{-1} (\beta_1 + \beta_3) \Xi + \Xi^2] . \]  

(8.6)

Factoring the operator between brackets, yields

\[ \beta_2^{-1} + \beta_2^{-1} (\beta_1 + \beta_3) \Xi + \Xi^2 = [\beta_6 - \Xi] [\beta_7 - \Xi] , \]

(8.7)

with

\[ \beta_6 = (2 \beta_2)^{-1} \{- \beta_1 - \beta_3 + [(\beta_1 + \beta_3)^2 - 4 \beta_2]^{1/2}\} , \]

(8.8)

\[ \beta_7 = (2 \beta_2)^{-1} \{- \beta_1 - \beta_3 - [(\beta_1 + \beta_3)^2 - 4 \beta_2]^{1/2}\} . \]

(8.9)

Then it is straightforward to invert the vertical slowness operator, viz.,

\[ (\tilde{\Gamma}^{III})^{-1} = c^{1/2} \{ \beta_2^{-1} \beta_7 - \Xi^{-1} [\beta_6 - \Xi]^{-1} [1 + \beta_3 \Xi] \} c^{1/2} . \]

(8.10)

Observe that the symbol of this operator has two poles on both sides of the origin (the zero crossings of \( \tilde{\Gamma}^{III} \)).

Using the factorization, we decompose the inverse vertical slowness into three operators,

\[ (\tilde{\Gamma}^{III})^{-1} = \tilde{Y}_3^{-1} \tilde{Y}_2^{-1} \tilde{Y}_1 , \]

(8.11)

with

\[ \tilde{Y}_1 = [1 + \beta_3 \Xi] c^{1/2} , \quad \tilde{Y}_2 = [\beta_6 - \Xi] , \quad \tilde{Y}_3 = [\beta_7 - \Xi] \beta_2 c^{-1/2} . \]

(8.12)

The inverse vertical slowness operator is then applied in two steps,

\[ \tilde{Y}_2 \left( \tilde{Y}_3 ((\tilde{\Gamma}^{III})^{-1} \tilde{h}) \right) = \tilde{Y}_1 \tilde{h} , \]

which involves twice solving an equation. Substituting Eq.(5.5), with the aid of Eq.(5.14) the two equations are discretised

\[ [\beta_6 (1 + a_2 \delta_t^2) - \beta_6 c^{-2} \delta_t^2 c^2] \left( c^{-1} (\tilde{Y}_3 ((\tilde{\Gamma}^{III})^{-1} \tilde{h})) \right) = \left[ (1 + a_2 \delta_t^2) + \beta_6 \beta_6 c^{-2} \delta_t^2 c^2 \right] (c^{-1} \tilde{h}) , \]

\[ \beta_2 \left[ \beta_7 (1 + a_2 \delta_t^2) - \beta_6 c^{-2} \delta_t^2 c^2 \right] \left( c^{-3/2} ((\tilde{\Gamma}^{III})^{-1} \tilde{h}) \right) = (1 + a_2 \delta_t^2) \left( c^{-1} (\tilde{Y}_3 ((\tilde{\Gamma}^{III})^{-1} \tilde{h})) \right) . \]

(8.13)

The associated tridiagonal matrices are inverted using the forward/backward substitution procedure (see Numerical Recipes [37]).
A numerical example

At the computational boundaries we apply, as before, the Robin boundary conditions (Section 6). We also employ an optimization procedure for the (de)composition operators, considering the parameter subset \( \{ \beta_1, \beta_2, \beta_3, a_2, \Omega \} \). To avoid instabilities, we have to move the poles arising in the symbols of the approximate operators into the complex horizontal slowness plane. Table 3 contains the outcome of the optimization for a sampling rate of 5 points per wavelength. Figure 15 shows the symbols of the composition operator for different values of \( \Omega \).

To demonstrate the effect of wavefield decomposition, we have computed the pressure excited by a line explosion source. The configuration was otherwise the same as in Figures 12-13. Figure 16 shows a snapshot at 0.3s. Note that the artificial head wave enforces the approximate wave front to be vertical at the level of the source. The radiation pattern becomes close to isotropic.

9 Sparse matrix representations of the reflection/transmission operators

From Eq.(2.31) we obtain the reflection and transmission operators,

\[
\hat{T} = -\hat{R} = -\frac{1}{\delta} \hat{\Delta}^{-1}(\partial_3 \hat{\Delta})
\]

(9.1)

Up to principal parts, we can simplify these operators, since then

\[
\partial_3 \hat{\Delta} \approx \frac{1}{\delta} \hat{\Delta}^{-1}(\partial_3 c^{-2}) \cdot \cdot \cdot 
\]

(9.2)

Substituting this approximation into Eq.(9.1), yields

\[
\hat{R} \approx \frac{1}{\delta} \Delta^{-1}((\partial_3 c^{-2}) \cdot \cdot \cdot )
\]

(9.3)

no rational approximation is required to arrive at a sparse matrix representation. and its validity extends into the post-critical regime. To stay conceptually close to the original expression (9.1) we replace \( \partial_3 c^{-2} \) by \( 2c^{-1}\partial_3 c^{-2} \). This replacement is justified only prior to applying the finite-difference approximation. In this approximation, we have

\[
\frac{1}{2} c \partial_3 c^{-1} F G \Delta^{-1}(x_3 - \frac{1}{2} \Delta x_3) = \Delta^{-1}(x_3 + \frac{1}{2} \Delta x_3)
\]

(9.4)

which resembles the linearized reflection coefficient at normal incidence.

On principal symbol level, the operator ordering can be further interchanged. Thus, the reflection operator can be written as

\[
\hat{R} \approx \frac{1}{\delta} \Delta^{-1}(\partial_3 c^{-1})^{1/2} \Delta^{-1}(\partial_3 c^{-1})^{1/2} \cdot \cdot \cdot 
\]

(9.5)

This form guarantees that the reflection/transmission operators vanish in any region where the medium is \( x_3 \)-invariant. Upon discretising this equation, using Eqs.(3.14) and (3.14), we get

\[
2 \left[ (1 + a_2 \delta_1^2) c^{-2} + \delta_3 c^{-2} \delta_1^2 \right] \left( \partial_3 c^{-1} \right)^{1/2} \hat{\Delta} \hat{h} = (1 + a_2 \delta_1^2) \left( c^{-1} \partial_3 c^{-1} \right)^{1/2} \hat{h}
\]

(9.6)
A numerical example

The discretised reflection operator contains only two free parameters: \( a_2 \) and \( \Omega \), through \( \beta_2 \). To optimize for these parameters, we would have to consider every possible change in wave speed. It turns out, however, that the optimization is rather insensitive to the medium changes, so that we can restrict ourselves to using a single parameter set,

\[
a_2 = 1/12, \quad \Omega = -0.1. \tag{9.7}
\]

The value of \( \Omega \) is taken negative, to move the pole at 90 degrees propagation angle into the complex horizontal slowness plane.

Figure 17 shows the symbol of the discretised reflection operator for \( a_2 = 1/12 \) and \( \Omega = 0.0 \) (the plane-wave reflection coefficient). In this figure are also shown the symbols for \( \Omega = -0.05 \) and \( \Omega = -0.1 \), as well as the exact reflection coefficient. In Figure 18 we show a snapshot of the pressure field corresponding to the first two terms of the Bremmer series, excited by a vertical line force in a two-layer medium. The numerical grid, and the position of the source, are the same as in Figures 12-13. The upper medium has a wave speed \( c = 1900\text{m/s} \) and the lower medium has a wave speed \( c = 3800\text{m/s} \). The interface is located at \( z_3 = 400\text{m} \), and the snapshot time is 0.3s, as before. The reflection and transmission are accurately modelled; the head wave, however, is only mimicked by our numerical scheme (for details, see De Hoop and De Hoop [48]).

10 Numerical simulations

In this section, we show examples of wave field modelling in various configurations, using our Bremmer coupling series algorithm. First, we consider two examples that originate from integrated optics. Here, we have translated those configurations into their acoustic counterparts. Second, we consider an example from exploration seismics, viz., a salt structure embedded in a sparse system of sedimentary layers.

Waveguide

Our first example is, in fact, a benchmark test (see Nolting [49]). It considers the propagation of a single mode through a waveguide, oriented under different angles with respect to the numerical grid. The details of this test configuration can be found in Nolting [49]. The wave speed is approximately 1825m/s inside the waveguide and 1900m/s outside. The waveguide has a thickness of approximately 180m. The frequency is 190Hz. We take a numerical grid with 779 points along the \( z_1 \)-direction and 1360 points along the \( z_3 \)-direction. The discretisation step is approximately 1.5m in both directions (about 6.7 points per wavelength). The total length of the waveguide is 2.04km. In this simulation, we used the third (real-valued) parameter set. In this waveguide there exist only 11 discrete modes.

Two of the waveguiding modes (the fundamental and the tenth mode) are launched into the waveguide. In the first experiment the wall of the waveguide is parallel to the \( z_3 \)-axis. In the second experiment, the waveguide is tilted with respect to the computational coordinate system by 20 degrees. The field at the output level plane is compared with original field at the input level plane. Theoretically, these fields must be identical. Figure 19 shows the fundamental mode at the end of the vertical waveguide, while Figure 20 shows the fundamental mode at the end of the tilted waveguide. Figures 21 and 22 are the same, but then for the tenth-order mode. For the vertical waveguide, the symbols account for all the interactions; for the tilted waveguide, an interplay between the Bremmer coupling and the symbol interactions takes place. Up to the tenth-order mode the results are satisfactory. It is conjectured that some of the radiating modes can be properly modelled with our scheme as well.
Block

In the second example, we investigate the time-harmonic plane wave interaction in a homogeneous medium ($c = 1900\text{m/s}$) with a square block (sized 100m by 100m) with significantly smaller wave speed ($c = 1344\text{m/s}$). The frequency is 19Hz. The numerical grid contains 99 points along the $x_1$-direction and 60 points along the $x_2$-direction. We plot only 59 points along the $x_1$-direction. The discretisation step is 5m in both directions (14 points per wavelength inside the heterogeneity). Full, accurate, numerical responses have been presented by Martin et al. [50]. Our results, shown in Figure 23, are very similar to theirs. (The right half of the figure is the Bremmer series solution, the left half is the full solution obtained by an iterative integral-equation technique based upon minimizing an integrated square error criterion.) Figure 24 corresponds to the second term of the Bremmer coupling series, while Figure 25 shows the field corresponding to the leading term of the Bremmer coupling series. The higher-order terms in the Bremmer series become rapidly smaller. For the calculations we used parameter set 4; in the reflection operator, we set $\Omega = -0.02$.

Geological dome

As a third example, we consider the salt model of Figure 26. This model contains three sedimentary layers. The wave speed is constant in the upper and bottom layers, while the wave speed varies in the left part of the middle layer. An approximate, transient vertical line force is applied at $z_1 = 1600\text{m}$, $z_3 = 0\text{m}$. Figures 27-29 show snapshots at 0.75s. For the calculations we used parameter set 4. The numerical grid contains 199 points in the $x_1$-direction and 156 points in the $x_3$-direction. The discretisation step is 16m in both directions. In Figure 27 we have calculated the first six terms of the Bremmer coupling series. (The computation time was 50780s on a Hewlett Packard 9000/735 workstation.) Figures 28 and 29 show snapshots of the leading term (890s CPU) and the second term of the series.

This example illustrates that our approximate scheme models the transmitted and reflected body waves accurately up to large scattering angles. Also the line diffraction at the major kink of the second deepest interface is captured by the method. Some head wave energy is mimiced by the artifacts of our approximations, but is inaccurate.
11 Discussion of the results

We have developed a fast numerical implementation of the Bremmer coupling series in two dimensions, for scattering angles up to critical. The key ingredient was a rational approximation of the vertical slowness in terms of the horizontal slowness. Our scheme has added to the various existing one-way wave propagation and BPM schemes in the following ways: we considered the third-order Thiele approximation rather than the first- or second-order one, to enhance the accuracy and to remove artificial body waves around the vertical direction; we enhanced the accuracy by optimizing the phase and group slowness surfaces for any desired sampling rates; we have improved the transformation to a moving frame of reference; we introduced (de)composition operators to incorporate any desired source- or receiver-type with the appropriate radiation characteristics; we have taken care of the backscattered field with the aid of the Bremmer coupling series. The backscattered field includes phenomena like turning ray waves. Though we have shown results in two dimensions only, the algorithm can be readily extended to three dimensions by employing a hexagonal grid in the horizontal directions (for reference, see Mersereau [51], and Peterson and Middleton [52]).

We have illustrated the generalized Bremmer series with rational approximation with various examples. The examples have been taken from the different fields of application of the method. With a view to exploration geophysics, the waveguide is associated with coal beds. Figure 16 represents the wave-equation migration-operator response, the tangents to the front defining the dips. In migration, the Bremmer series approach is particularly useful if multiple arrivals play a rôle creating a proper image.

The efficiency of the method depends on the application. The method is designed in the frequency domain, and hence allows for a relatively coarse sampling of the field: the lower the frequency, the coarser the grid can be. If only a small number of frequencies has to be evaluated, the Bremmer series approach has a particular advantage over full-wave modelling schemes. Though the range of scattering angles treated accurately is limited just below critical, the medium can be relatively strongly heterogeneous. If the medium is weakly heterogeneous, however, competing approaches exist to calculate the propagator. We mention the phase-shift (Gazdag and Sguazzero [53]), the split-step Fourier (Stoffa et al. [54]), the McClellan transform (Hale [55]), and the phase-screen (Wu [56]) representations. If the configuration consists of weakly and strongly heterogeneous regions, the different approaches can be combined into a hybrid scheme, matching them through the vertical phase slownesses.

In case critical angle phenomena are important, a more precise symbol for the vertical slowness operator has to be used. A candidate is the uniform expansion developed by Fishman and Gauzen [57]. Then the thin slab propagator is directly evaluated using Fourier transforms.
References


A FORTRAN code for constructing the involved matrices

The FORTRAN code for setting up the matrices in the propagation scheme, consists of two subroutines. In one subroutine, the following global constants are calculated, once for every frequency,

\[
\beta_8 = (1D0, 0D0) / (\Omega (1D0 + (3D0, 1D0) \omega \text{Im}) \Delta x)^{2} \\
\text{const} = (1D0, 1D0) \Omega \Delta x \beta_8 \text{const} \\
\text{const} = (1D0, 2D0) \beta_2 \beta_8 \text{const} \\
\text{const} = (1D0, 3D0) \beta_1 \beta_8 (1 - 2D1 a_2) \text{const} \\
\text{const} = (1D0, 4D0) a_2 \text{const} \\
\text{const} = (1D0, 5D0) \beta_3 \beta_8 \Delta x \text{const} \\
\text{const} = (1D0, 6D0) \beta_2 \beta_8 \beta_8 \text{const} \\
\text{const} = (1D0, 7D0) \beta_1 \beta_8 (6D1 a_2 - 2D0) \text{const} \\
\text{const} = (1D0, 8D0) a_2 \text{const} \\
\text{const} = (1D0, 9D0) (1D0 - 2D0 a_2) \text{const} \\
\text{const} = (1D0, 10D0) \beta_3 \beta_8 \text{const} \\
\text{const} = (1D0, 11D0) - \beta_2 \beta_8 \Delta x (1D0 - 2D0 a_2) \text{const} \\
\text{const} = (1D0, 12D0) \beta_2 \beta_8 \beta_8 \text{const}
\]

where \( \omega \text{Im} = \Omega \). In the second subroutine, the matrix elements of \( A \) and \( A' \) are evaluated. In this subroutine, we set

\[
\text{cp} = c(i + 1) \\
\text{cc} = c(i) \\
\text{cm} = c(i - 1) \\
\text{dp} = 1D0 / \text{cp} \\
\text{dd} = 1D0 / \text{cc} \\
\text{dm} = 1D0 / \text{cm} \\
\text{gp} = \text{cp} / \text{cp} \\
\text{gg} = \text{cc} / \text{cc} \\
\text{gm} = \text{cm} / \text{cm}
\]

The function \( c(i) \) returns the value of \( c \) at position \( x_i = i \Delta x \). Since most configurations will contain large homogeneous regions, the above variables have not to be recalculated at every position.

The evaluation of the elements of matrices \( A \) and \( A' \) are very similar. We show the set up and storage of matrix \( A \),

\[
A(3, i + 2) = A(i, i - 2), \quad (A.1) \\
A(4, i + 1) = A(i, i - 1), \quad (A.2) \\
A(5, i) = A(i, i), \quad (A.3) \\
A(6, i - 1) = A(i, i - 1). \quad (A.4)
\]
\[ A(7, i-2) = A_{i, i-2}. \] (A.5)

The matrix elements follow as

\[
\begin{align*}
A(3, i+2) &= \text{const}12 + \text{const}6 \cdot \text{gp} + \text{const}8 \cdot \text{dp} + \text{const}10 \cdot \text{cp} \\
A(4, i+1) &= \text{const}1 \cdot (\text{gg} + \text{gp}) + \text{const}2 + \text{const}3 \cdot (\text{dd} + \text{dp}) + \text{const}4 \cdot \text{cc} \\
& \quad + \text{const}5 \cdot \text{cp} \\
A(5, i) &= \text{const}6 \cdot (4 \cdot \text{gg} + \text{gm} + \text{gp}) + \text{const}7 + \text{const}8 \cdot \text{dm} + \text{const}9 \cdot \text{dd} \\
& \quad + \text{const}10 \cdot \text{cm} + \text{const}11 \cdot \text{cc} + \text{const}10 \cdot \text{cp} \\
A(6, i-1) &= \text{const}1 \cdot (\text{gg} + \text{gm}) + \text{const}2 + \text{const}3 \cdot (\text{dd} + \text{dm}) + \text{const}4 \cdot \text{cc} \\
& \quad + \text{const}5 \cdot \text{cm} \\
A(7, i-2) &= \text{const}12 + \text{const}6 \cdot \text{gm} + \text{const}8 \cdot \text{dm} + \text{const}10 \cdot \text{cm}
\end{align*}
\]

The matrix \( C \) is stored in a likewise manner,

\[
\begin{align*}
K(2, i+1) &= C_{i, i+1}, \\
K(3, i) &= C_{i, i}, \\
K(4, i-1) &= C_{i, i-1},
\end{align*}
\] (A.6) \hspace{1cm} (A.7) \hspace{1cm} (A.8)

and its elements follow as

\[
\begin{align*}
K(4, i-1) &= a2 \cdot e(i)/\sqrt{cc} \\
K(3, i) &= e(i)/\sqrt{cc} \cdot (1 - 2 \cdot a2) \\
K(2, i+1) &= a2 \cdot e(i)/\sqrt{cc}
\end{align*}
\]

where the function \( e(i) \) returns the value

\[
e = \text{EXP}((0D0, 1D0) \cdot \text{DeltaX3} \cdot c(i) \cdot \text{Omega})
\]
Figure 1: A sample configuration. Between the entrance and the exit planes, the configuration's medium may vary in the vertical and horizontal directions. Outside that region the medium is constant.
Figure 2: The levels at which the field is calculated.
Incident field: $F_1(0, t) = p(0, t)$ or $F_2(0, t) = v_3(0, t)$ is given

Fourier Transform $\rightarrow \hat{F}^{\text{incident}}(x_1, 0, \omega)$

For every $\omega$

| decomposition: $\hat{W}_I = (\hat{L}^{-1})_{I,J} \hat{F}_J$ |
| For every term $j$ of Bremmer Coupling Series, $(j = 0, 1, 2, \cdots)$ |
| $\hat{I}^{(j)}_{1,J}(x_1, 0) = 0$, $\hat{I}^{(0)}_{1,J}(x_1, 0) = \hat{W}_I(x_1, 0)$ |
| For every downward propagation step $m$, $(m = 1, 2, \cdots, M)$ |
| $\hat{I}^{(j)}_{1,J}(x_1, m) = \hat{P}(x_1, m)\hat{I}^{(j)}_{1,J}(x_1, m-1) + \hat{q}^{(j)}_{2,J}(x_1, m)$ |
| $\hat{W}^{(j)}_1(x_1, m\Delta x_3) = \hat{I}^{(j)}_{1,J}(x_1, m) - \hat{I}^{(j)}_{1,2}(x_1, m)$ |
| Calculate $\hat{Q}^{(j+1)}_1(x_1, m\Delta x_3)$ |
| composition $\rightarrow \hat{F}^{(j)}_1$ and $\hat{F}^{(j)}_2$ |
| adding to previous fields + storage |

$\hat{I}^{(j)}_{2,J}(x_1, M) = 0$

For every upward propagation step $m$, $(m = M - 1, \cdots, 0)$

| $\hat{I}^{(j)}_{2,J}(x_1, m) = \hat{P}(x_1, m + 1)\hat{I}^{(j)}_{2,J}(x_1, m + 1) + \hat{q}^{(j)}_{2,J}(x_1, m)$ |
| $\hat{W}^{(j)}_2(x_1, m\Delta x_3) = -\hat{I}^{(j)}_{2,1}(x_1, m) + \hat{I}^{(j)}_{2,2}(x_1, m)$ |
| Calculate $\hat{Q}^{(j+1)}_2(x_1, m\Delta x_3)$ |
| composition $\rightarrow \hat{F}^{(j)}_1$ and $\hat{F}^{(j)}_2$ |
| adding to previous fields + storage |

Inverse Fourier Transform $\rightarrow p(x_1, x_3, t)$ and $v_3(x_1, x_3, t)$ at the positions of interest

Figure 3: Flow chart of the numerical implementation of the Bremmer coupling series.
Figure 4: The exact vertical slowness (left column) and its Thiele approximants with $\beta_1 = 1/2$, $\beta_2 = 1/8$, $\beta_3 = 1/2$ (middle) and with $\beta_1 = 0.531$, $\beta_2 = 0.364$, $\beta_3 = 0.825$ (right). In the upper left figure the dotted line corresponds to the imaginary part of the vertical slowness, while in the two other upper figures the dotted lines correspond to the real part of the exact vertical slowness. The figures in the second row show the differences between the approximants and the exact expression. In the bottom row, the real parts of the vertical slownesses as functions of complex $\alpha_1$ are shown. The contour lines correspond to the values $0.1, 0.2, \ldots, 2.0$. 
Figure 3: The Q-factor (left) and the corresponding amplification factor (right) associated with the dissipation crisis in the third-order Thiele approximation: $\beta_0 = 0.264$, $\beta_2 = 0.304$, and $\beta_3 = 0.82$. The relative imaginary circular frequency $\Omega = 0.02$. Note that the notch occurs beyond the critical angle.
Figure 6: The spectra or symbols of the approximate $\delta^2$ operators with $a_2 = 1/12$ (upper left) and $a_2 = 0.130$ (upper right). The bottom figures show the corresponding differences with the exact spectra, which are also dotted in the upper figures.
Figure 7: The third-order Thiele approximation of the vertical slowness symbol including the spectrum of the discretised Laplace operator ($\alpha_2 = 0.130$, $\beta_1 = 0.526$, $\beta_2 = 0.364$, $\beta_3 = 0.825$, and a sampling rate of 8 points per wavelength). The poles are located at $\alpha_1 \simeq 1.102, 6.898, 9.102, \cdots$. 
Figure 8: The phase slowness (upper left) and the group slowness (upper right) associated with the discretised, approximate, one-way wave equation with $\beta_1 = 1/2$, $\beta_2 = 1/8$, $\beta_3 = 1/2$, $\alpha_2 = 1/12$, $\beta_3 = 1/2$, and a sampling rate of 5 points per wavelength. The exact vertical wave-slowness symbol is dotted. In the bottom row the difference between the phase slowness and the group slowness, and the difference between the group slowness and the exact slowness are plotted.
Figure 9: The phase slowness (upper left) and the group slowness (upper right) associated with the discretised, approximate, one-way wave equation with $\beta_1 = 0.526$, $\beta_2 = 0.364$, $\beta_3 = 0.825$, $\alpha_2 = 0.0888$, $\beta_0 = 0.529$, and a sampling rate of 5 points per wavelength. The exact vertical wave-slowness symbol is dotted. In the bottom row the difference between the phase slowness and the group slowness, and the difference between the group slowness and the exact slowness are plotted.
Figure 10: The phase slowness (upper left) and the group slowness (upper right) associated with the discretised, approximate, one-way wave equation with $\beta_1 = 0.486$, $\beta_2 = 0.349$, $\beta_3 = 0.841$, $\alpha_2 = 0.114$, $\beta_0 = 0.529$, and a sampling rate of 5 points per wavelength. The exact vertical wave-slowness symbol is dotted. In the bottom row the difference between the phase slowness and the group slowness, and the difference between the group slowness and the exact slowness are plotted.
Figure 11: The difference between the real parts of the phase slowness and the group slowness (upper left), and the difference between the real parts of the group slowness and the exact slowness (upper right). The middle row shows the imaginary parts of the phase slowness (left) and the group slowness (right). The parameter set is given in column 4 of Table 1. Two sampling rates are shown: 5 points per wave length (solid line) and 10 points per wave length (dotted line). The bottom figure shows the imaginary parts of the phase slowness (solid line) and the group slowness (dotted line) on a larger scale.
Figure 12: Snapshot at 0.3s of a vertical line force source response in a homogeneous medium ($c = 1900$ m/s). The right-hand side is obtained analytically, while the left-hand side is obtained with our numerical scheme. Here, we used parameter set 1 of Table 1.
Figure 13: Snapshot at 0.3s of a vertical line force source response in a homogeneous medium ($c = 1900$ m/s). The right-hand side is obtained analytically, while the left-hand side is obtained with our numerical scheme. Here, we used parameter set 4 of Table 1.
Figure 14: The absolute error \((0 < x_1 < 800\text{m})\) and the relative error \((800 < x_1 < 1600\text{m})\) of the snapshot shown in the previous figure. The relative error is in \%.
Figure 15: The real part of the symbol of the vertical slowness (composition) operator (upper figure) and its difference with the exact symbol (lower figure). The dotted line corresponds to the real part of the exact slowness symbol; the other lines correspond to $\beta_1 = 0.526$, $\beta_2 = 0.364$, $\beta_3 = 0.825$ and $c_2 = 0.130$, while $\Omega$ equals 0 (solid line), $-0.01$ (dashed line) and $-0.1$ (dashdot line).
Figure 16: Snapshot of an explosion line source in a homogeneous medium ($c = 1900 \text{ m/s}$). The right-hand side is obtained analytically, while the left-hand side is obtained with our numerical scheme.
Figure 17: The real part of the symbols of the discretised and the exact (dotted line) reflection operators (plane-wave reflection coefficients). The differences between the discretised and exact symbols are plotted in the lower figure. The wave speed equals 1 above and 2 below the interface. We used $\alpha_2 = 1/12$, and $\Omega = 0, -0.1, -0.05$ (solid, dashed and dashdot lines, respectively).
Figure 18: Reflection and transmission at an interface.
Figure 19: The fundamental waveguiding mode at the entrance plane (upper figure) and at the exit plane (bottom figure). The waveguide is not tilted with respect to the coordinate system.
Figure 20: The fundamental waveguiding mode at the entrance plane (upper figure) and at the exit plane (bottom figure). The waveguide is tilted with an angle of 20 degrees with respect to the coordinate system. (Leading term of the Bremmer coupling series only.)
Figure 21: The tenth-order waveguide mode at the entrance plane (upper figure) and at the exit plane (bottom figure). The waveguide is not tilted with respect to the coordinate system.
Figure 22: The tenth-order waveguide mode at the entrance plane (upper figure) and at the exit plane (bottom figure). The waveguide is tilted with an angle of 20 degrees with respect to the coordinate system. (Leading term of the Bremmer coupling series only.)
Figure 23: Plane wave (from the top) response in a configuration with a block in a homogeneous embedding.
Figure 24: Plane wave (from the top) response in a configuration with a block in a homogeneous embedding. Second term of the Bremmer coupling series.
Figure 25: Plane wave (from the top) response in a configuration with a block in a homogeneous embedding. Leading term of the Bremmer coupling series.
Figure 26: Wave speeds of a two-dimensional salt model.
Figure 27: Snapshot at $t = 0.750s$ of the acoustic pressure due to a vertical line force in the model of Figure 26. Six terms of the Bremmer coupling series.
Figure 28: Snapshot at $t = 0.75s$ of the acoustic pressure due to a vertical line force in the model of Figure 26. Leading term of the Bremer coupling series.
Figure 20: Snapshot at $t = 0.730s$ of the acoustic pressure due to a vertical line force in the model of the Bremmer coupling series.
Table 1: The parameter sets for propagation.

<table>
<thead>
<tr>
<th>parameter set</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta x_1/\lambda$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>$\Delta x_2/\lambda$</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
<td>0.2</td>
</tr>
<tr>
<td>Re{$\beta_1$}</td>
<td>1/2</td>
<td>0.526</td>
<td>0.486</td>
<td>0.5104</td>
</tr>
<tr>
<td>Im{$\beta_1$}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.0340</td>
</tr>
<tr>
<td>Re{$\beta_2$}</td>
<td>1/8</td>
<td>0.364</td>
<td>0.349</td>
<td>0.2207</td>
</tr>
<tr>
<td>Im{$\beta_2$}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.0131</td>
</tr>
<tr>
<td>Re{$\beta_3$}</td>
<td>1/2</td>
<td>0.825</td>
<td>0.841</td>
<td>0.6685</td>
</tr>
<tr>
<td>Im{$\beta_3$}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.0310</td>
</tr>
<tr>
<td>Re{$\alpha_2$}</td>
<td>1/12</td>
<td>0.089</td>
<td>0.114</td>
<td>0.1207</td>
</tr>
<tr>
<td>Im{$\alpha_2$}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0.0063</td>
</tr>
<tr>
<td>Re{$\beta_9$}</td>
<td>1/2</td>
<td>0.540</td>
<td>0.529</td>
<td>0.4679</td>
</tr>
<tr>
<td>Im{$\beta_9$}</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.0066</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>-0.0406</td>
</tr>
<tr>
<td>Figures</td>
<td>8</td>
<td>9</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>13</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

Table 2: The poles of the numerical vertical phase slowness.

<table>
<thead>
<tr>
<th>parameter set</th>
<th>pole 1</th>
<th>pole 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.4588 + 0.2807i</td>
<td>1.7135 + 1.3074i</td>
</tr>
<tr>
<td>2</td>
<td>1.1054 - 0.0040i</td>
<td>2.2899 + 1.7530i</td>
</tr>
<tr>
<td>3</td>
<td>1.0762 - 0.0271i</td>
<td>1.3686 + 1.1214i</td>
</tr>
<tr>
<td>4</td>
<td>1.5974 - 1.1406i</td>
<td>1.1816 + 0.1211i</td>
</tr>
</tbody>
</table>
Table 3: The parameter set for (de)composition.

<table>
<thead>
<tr>
<th>parameter set</th>
<th>2'</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\Delta x_1/\lambda$</td>
<td>0.2</td>
</tr>
<tr>
<td>$\Delta x_2/\lambda$</td>
<td>0.2</td>
</tr>
<tr>
<td>Re{$\beta_1$}</td>
<td>0.526</td>
</tr>
<tr>
<td>Im{$\beta_1$}</td>
<td>0</td>
</tr>
<tr>
<td>Re{$\beta_2$}</td>
<td>0.364</td>
</tr>
<tr>
<td>Im{$\beta_2$}</td>
<td>0</td>
</tr>
<tr>
<td>Re{$\beta_3$}</td>
<td>0.825</td>
</tr>
<tr>
<td>Im{$\beta_3$}</td>
<td>0</td>
</tr>
<tr>
<td>Re{$\alpha_2$}</td>
<td>0.13</td>
</tr>
<tr>
<td>Im{$\alpha_2$}</td>
<td>0</td>
</tr>
<tr>
<td>$\Omega$</td>
<td>-0.01</td>
</tr>
</tbody>
</table>