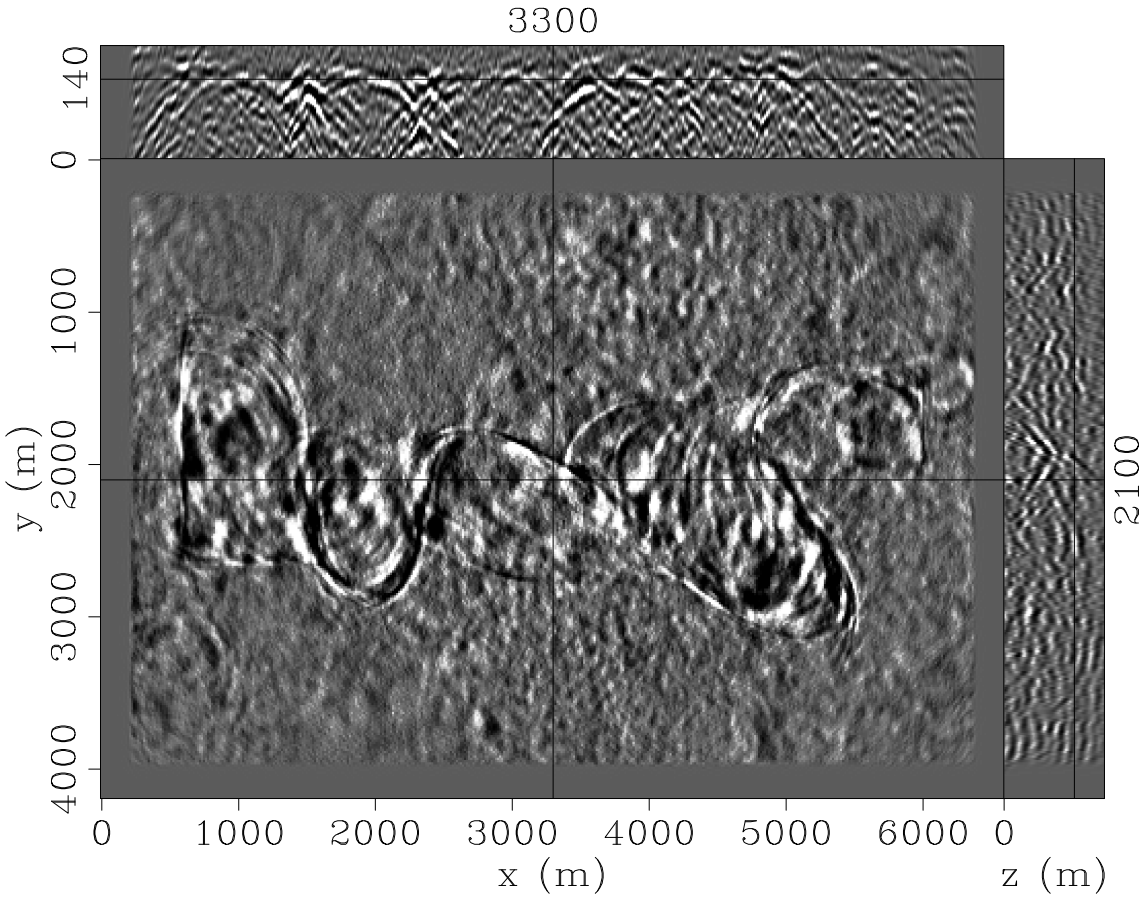
TEXAS CONSORTIUM FOR COMPUTATIONAL SEISMOLOGY

Tariq Alkhalifah, Thomas Browaeys, William Burnett, Yihua Cai, Maria Cameron, Keling Chen, Xiaohong Chen, Yangkang Chen, Zhonghuan Chen, Luke Decker, Laurent Demanet, Qizhen Du, Bj"orn Engquist, Gang Fang, Robert Ferguson, Sergey Fomel, Jingwei Hu, Xavier Janson, Long Jin, Zhaoyu Jin, Parvaneh Karimi, Roman Kazinnik, Alexander Klokov, Evgeny Landa, Jingye Li, Siwei Li, Cai Liu, Guochang Liu, Tingting Liu, Yang Liu, Wenkai Lu, Jitao Ma, Jack Poulson, James Sethian, Xiaolei Song, Alexey Stovas, M. Turhan Taner, Mirko van der Baan, Alexander Vladimirsky, Dian Wang, Erying Wang, Lexing Ying, Jiang Yuan, Hongliu Zeng, and Lele Zhang

Selected papers



Copyright © 2014-15

by The University of Texas at Austin

TCCS — TABLE OF CONTENTS

Alexander Klokov and Sergey Fomel, Selecting an optimal aperture in Kirchhoff migration using dip-angle images 1

William Burnett and Robert Ferguson, A reversible transform for seismic data processing 35

Sergey Fomel, Lexing Ying, and Xiaolei Song, Seismic wave extrapolation using lowrank symbol approximation 55

Yang Liu and Sergey Fomel, OC-seislet: seislet transform construction with differential offset continuation 75

Parvaneh Karimi and Sergey Fomel, Stratigraphic coordinates, a coordinate system tailored to seismic interpretation 103

Xiaolei Song, Sergey Fomel, and Lexing Ying, Lowrank finite-differences and lowrank Fourier finite-differences for seismic wave extrapolation in the acoustic approximation 119

Zhonghuan Chen, Sergey Fomel, and Wenkai Lu, Accelerated plane-wave destruction 135

Sergey Fomel, On anelliptic approximations for qP velocities in transversally isotropic media 151

Yangkang Chen, Sergey Fomel, and Jingwei Hu, Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization 171

Sergey Fomel, Seismic data decomposition into spectral components using regularized nonstationary autoregression 195

Sergey Fomel, Theory of 3-D angle gathers in wave-equation seismic imaging 217

Jingwei Hu, Sergey Fomel, Laurent Demanet, and Lexing Ying, A fast butterfly algorithm for generalized Radon transforms 231

Yang Liu and Sergey Fomel, Seismic data interpolation beyond aliasing using regularized nonstationary autoregression 267

Sergey Fomel, Local seismic attributes 287

Sergey Fomel, Shaping regularization in geophysical estimation problems 295

Siwei Li and Sergey Fomel, Kirchhoff migration using eikonal-based computation of traveltime source-derivatives 315

Yang Liu and Sergey Fomel, Seismic data analysis using local time-frequency decomposition 339

Yangkang Chen, Tingting Liu, Xiaohong Chen, Jingye Li, and Erying Wang, Time-frequency analysis of seismic data using synchrosqueezing wavelet transform 359

Xiaolei Song and Sergey Fomel, Fourier finite-difference wave propagation 369

Zhonghuan Chen, Sergey Fomel, and Wenkai Lu, Omnidirectional plane-wave destruction 389

Maria Cameron, Sergey Fomel, and James Sethian, Time-to-depth conversion and seismic velocity estimation using time-migration velocity 407

Yang Liu, Cai Liu, and Dian Wang, A 1-D time-varying median filter for seismic random, spike-like noise elimination 421

Sergey Fomel, Velocity-independent time-domain seismic imaging using local event slopes 437

Yangkang Chen and Sergey Fomel, Random noise attenuation using local signal-and-noise orthogonalization 461

Sergey Fomel, Evgeny Landa, and M. Turhan Taner, Post-stack velocity analysis by separation and imaging of seismic diffractions 481

Guochang Liu, Sergey Fomel, Long Jin, and Xiaohong Chen, Stacking seismic data using local correlation 499

Sergey Fomel and Mirko van der Baan, Local skewness attribute as a seismic phase detector 511

Jingwei Hu, Sergey Fomel, and Lexing Ying, A fast algorithm for 3D azimuthally anisotropic velocity scan 529

Guochang Liu, Sergey Fomel, and Xiaohong Chen, Stacking angle-domain common-image gathers for normalization of illumination 541

Yihua Cai, Sergey Fomel, and Hongliu Zeng, Automated spectral recomposition with application in stratigraphic interpretation 553

Yang Liu, Sergey Fomel, and Guochang Liu, Nonlinear structure-enhancing filtering using plane-wave prediction 569

Jack Poulson, Bj"orn Engquist, Siwei Li, and Lexing Ying, A parallel sweeping preconditioner for heterogeneous 3D Helmholtz equations 595

Xiaolei Song and Tariq Alkhalifah, Modeling of pseudo-acoustic P-waves in orthorhombic media with a lowrank approximation 621

Siwei Li and Sergey Fomel, A robust approach to time-to-depth conversion and interval velocity estimation from time migration in the presence of lateral velocity variations 639

Sergey Fomel, Predictive painting of 3-D seismic volumes 679

Sergey Fomel and Yang Liu, Seislet transform and seislet frame 693

William Burnett and Sergey Fomel, 3D velocity-independent elliptically-anisotropic moveout correction 727

Sergey Fomel and Evgeny Landa, Structural uncertainty of time-migrated seismic images 745

Sergey Fomel and Long Jin, Time-lapse image registration using the local similarity attribute 759

Thomas J. Browaeys and Sergey Fomel, Fractal heterogeneities in sonic logs and low-frequency scattering attenuation 771

Sergey Fomel, Adaptive multiple subtraction using regularized nonstationary regression 799

Siwei Li, Alexander Vladimirsky, and Sergey Fomel, First-break traveltime tomography with the double-square-root eikonal equation 817

Sergey Fomel and Alexey Stovas, Generalized nonhyperbolic moveout approximation 849

Gang Fang, Sergey Fomel, Qizhen Du, and Jingwei Hu, Lowrank seismic wave extrapolation on a staggered grid 871

Yangkang Chen, Jiang Yuan, Zhaoyu Jin, Keling Chen, and Lele Zhang, Deblending using normal moveout and median filtering in common-midpoint gathers 899

William Burnett and Sergey Fomel, Azimuthally anisotropic 3D velocity continuation 921

Guochang Liu, Sergey Fomel, and Xiaohong Chen, Time-frequency analysis of seismic data using local attributes 935

Sergey Fomel, Velocity analysis using AB semblance 959

Sergey Fomel and Roman Kazinnik, Non-hyperbolic common reflection surface 977

Luke Decker, Xavier Janson, and Sergey Fomel, Carbonate reservoir characterization using seismic diffraction imaging 991

Yangkang Chen and Jitao Ma, Random noise attenuation by f-x empirical mode decomposition predictive filtering 1007

Selecting an optimal aperture in Kirchhoff migration using dip-angle images

Alexander Klokov and Sergey Fomel

ABSTRACT

We present a method for selecting a migration aperture in Kirchhoff migration. We first split migrated data into constant-dip-angle partial images. Then, in every partial image, we estimate the consistency between each event and the constant dip of the analyzed section. We filter out events whose slope is far from the corresponding dip. Stacking of the filtered partial images corresponds to migration having an optimal aperture. Synthetic and real data examples demonstrate that the proposed approach to migration-aperture optimization is able to reduce migration noise while preserving diffraction energy, which characterizes small geological objects and brings additional resolution to the image.

INTRODUCTION

Kirchhoff migration remains a popular tool for seismic imaging (17). It handles every input trace separately, enabling the user to work with data sets of any configuration (including irregular geometry) and to apply target-oriented analysis. One of the features of Kirchhoff migration is its global aperture. The migration operator distributes energy throughout a wide area described by an impulse-response surface. However, a seismic image is constructed only by a local part of the surface that touches a reflector. The remainder does not make a constructive contribution and, even worse, may generate image artifacts (28; 11).

The artifact problem can be solved by limiting migration aperture. Schleicher et al. (24) analyzed influence of aperture parameters on imaging results and proposed a technique based on projected Fresnel zones for shortening the migration curve. After preliminary picking of target reflections and computing diffraction curves, they estimated the migration-aperture size by analytical equations. Sun (28) analyzed the structure of a seismic image and proposed rules for its optimal construction. Two main principles were identified: (1) the tangent point between a diffraction curve and a reflector must be located in the central part of the aperture and (2) the aperture must be at least as large as the first Fresnel zone. Several methods were proposed to achieve these two objectives. Tillmanns and Gebrande (31) detected instantaneous slowness in the data domain. Sun and Schuster (27) described wave-path migration smearing energy within a small zone centered on the specular reflection point and defined under a stationary-phase approximation. Lüth et al. (18) smeared migrated multicomponent

data only inside a Fresnel volume built around rays computed using polarization of the wavefield. Buske et al. (7) constructed a Fresnel volume for single-component seismic data, determining the emergent angle by local slowness analysis. Tabti et al. (29) worked with diffraction-operator panels and picked Fresnel apertures after preliminary low-pass filtering. Kabbej et al. (12) introduced an attribute that characterizes the distance between a common-midpoint position and a currently imaged depth point then migrated the attribute and used it for weighting-function construction. Spinner and Mann (25) used the common-reflection-surface operator to determine parameters for the optimal migration aperture. Alerini and Ursin (1) estimated horizon slopes in the image domain.

The migrated dip-angle domain has properties that are favorable for taking both issues — the aperture position and its width — into consideration. In this domain, a reflection event has a concave shape with an apex whose position corresponds to the reflector dip (2; 16; 15). The event summation in the dip-angle direction produces an image in accordance with the stationary-point principle (4). To limit aperture in the dip-angle domain and to restrict summation to stationary points, Bienati et al. (3) applied an automatic slope estimation followed by muting. Aperture size was defined on the basis of wavelet bandwidth. Dafni and Reshef (8) proposed analyzing migrated gathers simultaneously in dip-angle and scattering-angle directions.

A seismic wavefield may contain reflections and diffractions. There are a number of important differences between these two components (14). One of the differences is that reflections require a narrow migration aperture, whereas diffractions require an aperture as wide as possible (while allowing matching of a diffraction curve). Therefore, all methods that imply migration-aperture limiting are oriented toward optimal imaging of reflection boundaries and not diffraction objects. Diffractions characterize small but important geological objects and play a significant role in imaging of rough reflection boundaries (13). Their attenuation may cause a significant loss of resolution (19). For image resolution to be preserved, a migration-optimization method should aim to protect the diffraction component. In this paper, we demonstrate an approach that allows us to achieve an optimal aperture size for reflection boundaries while also protecting the diffraction component. The main idea is analyzing slope information in constant-dip partial images.

DIFFRACTIONS AND REFLECTIONS IN CONSTANT-DIP PARTIAL IMAGES

Common-reflection-angle migration (32) can produce a set of dip-angle gathers in which traces are partial images of a fixed lateral position for different migration dips. At the same time, the migrated volume can be considered as a set of constant-dip partial images. Integration of the partial images along the dip angle provides a seismic image:

$$I(\bar{\mathbf{x}}, z) = \int \hat{I}(\bar{\mathbf{x}}, z, \alpha) d\alpha \ , \tag{1}$$

where $\hat{I}(\bar{\mathbf{x}}, z, \alpha)$ is a true-amplitude partial image for migration dip α (5).

A dip-angle gather exposes the difference between reflections and diffractions (2; 16; 15). Reflection events have a concave shape, and an image of a reflection boundary is constructed by a limited range of dip angles near the event apex. The effective range contains migrated impulses whose shift from the event apex does not exceed half the prevailing period. This area is equivalent to the Fresnel zone in the dip-angle domain. Diffraction events appear flat, however, which means that the full range of migration-dip angles provides an effective contribution to the image (Figure 1). Thus, an optimal migration aperture should simultaneously be limited for reflections and expanded for diffractions.

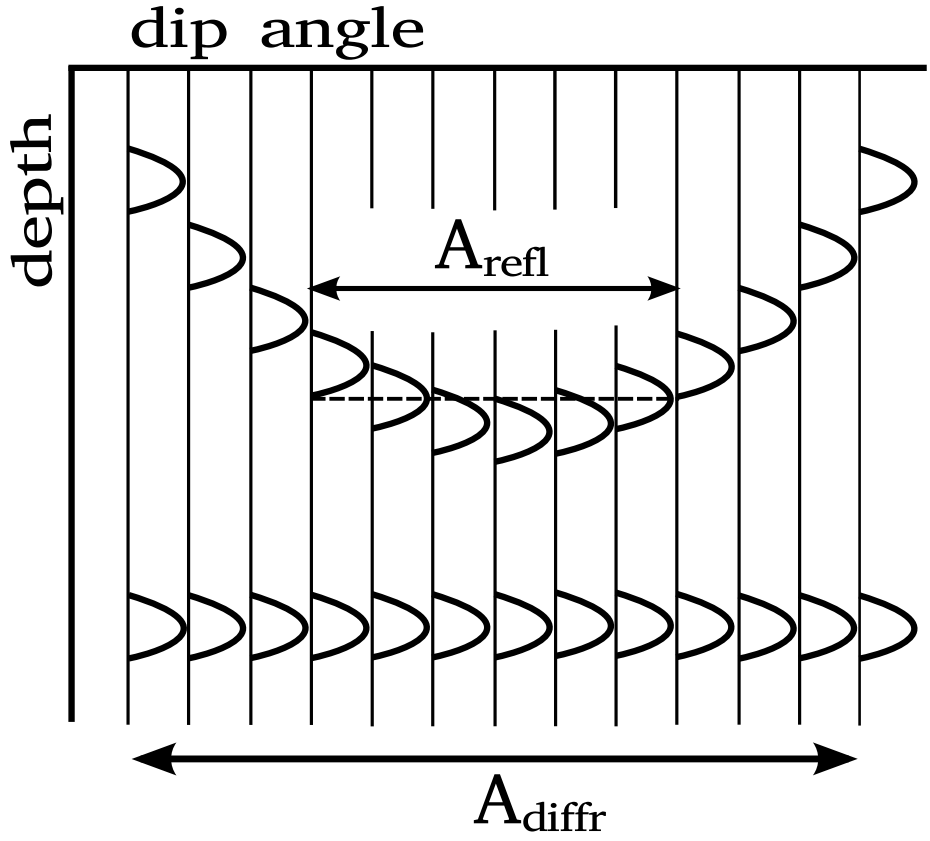


Figure 1: Reflection event and diffraction event in the 2D dip-angle gather. optapert/. smile

In the constant-dip partial-image domain, reflections and diffractions appear different as well, although they have some analogous features. To examine this, let us start with the simple case of zero-offset migration having the correct constant velocity. Similar derivations can be considered for prestack data.

A diffraction event behaves hyperbolically (Figure 2(a)). Its shape is defined by the following equation (see derivation in Appendix):

$$z(x_m) = (x_m - x_0)\tan\alpha + \sqrt{(x_m - x_0)^2/\cos^2\alpha + z_0^2}\,, \tag{2}$$

where α is the constant-dip of the partial image, x_m is an imaging position, and x_0 and z_0 are position and real depth of the diffraction point, respectively.

At the position of the diffraction point, when $x_m = x_0$, the event has the correct depth $z(x_m) = z_0$. The slope of the event at this point corresponds to the derivative:

$$\frac{\partial z}{\partial x_m}|_{x_m=x_0} = \tan\alpha \ . \tag{3}$$

Equation 3 indicates that an effective part of a diffraction event, which corresponds to the correct position and contributes to the image, has the same slope as the migration dip, or the constant-dip of the corresponding partial image.

The response of a plane reflector in the constant-dip partial image is a segment. Let us consider a small part of the reflection boundary, which has a dip α_0 and lateral length d (Figure 2(b)). In the correct-constant-velocity case, coordinates of the segment edges after constant-dip migration can be defined by rotation around escape points y_1 and y_2:

$$\bar{x_1} = z_1\frac{\sin\alpha_0 - \sin\alpha}{\cos\alpha_0} \ , \tag{4}$$

$$\bar{z_1} = z_1\frac{\cos\alpha}{\cos\alpha_0} \ , \tag{5}$$

and

$$\bar{x_2} = z_2\frac{\sin\alpha_0 - \sin\alpha}{\cos\alpha_0} + d \ , \tag{6}$$

$$\bar{z_2} = z_2\frac{\cos\alpha}{\cos\alpha_0} \ . \tag{7}$$

The migrated edge points define a segment, whose slope is

$$\tan\beta = \frac{\bar{z_2} - \bar{z_1}}{\bar{x_2} - \bar{x_1}} = \frac{\cos\alpha\sin\alpha_0}{1 - \sin\alpha\sin\alpha_0} \ . \tag{8}$$

Equation 8 illustrates some features of the migrated reflection boundary. For instance, a migrated horizontal reflector ($\alpha_0 = 0$) keeps the zero-slope independent of the migration dip.

Equation 8 also allows us to find the condition in which the migrated slope is consistent with the constant-dip partial image ($\beta = \alpha$). In this case,

$$\frac{\cos\alpha\sin\alpha_0}{1 - \sin\alpha\sin\alpha_0} = \frac{\sin\alpha}{\cos\alpha} \ . \tag{9}$$

If α is derived from equation 9, note that the equation has only an $\alpha = \alpha_0$ solution. In other words, the slope of a migrated reflection event equals that of the migration dip only after migration with the correct dip. Such is the case when the segment corresponds to the stationary point and contributes to the image. Thus, an effective

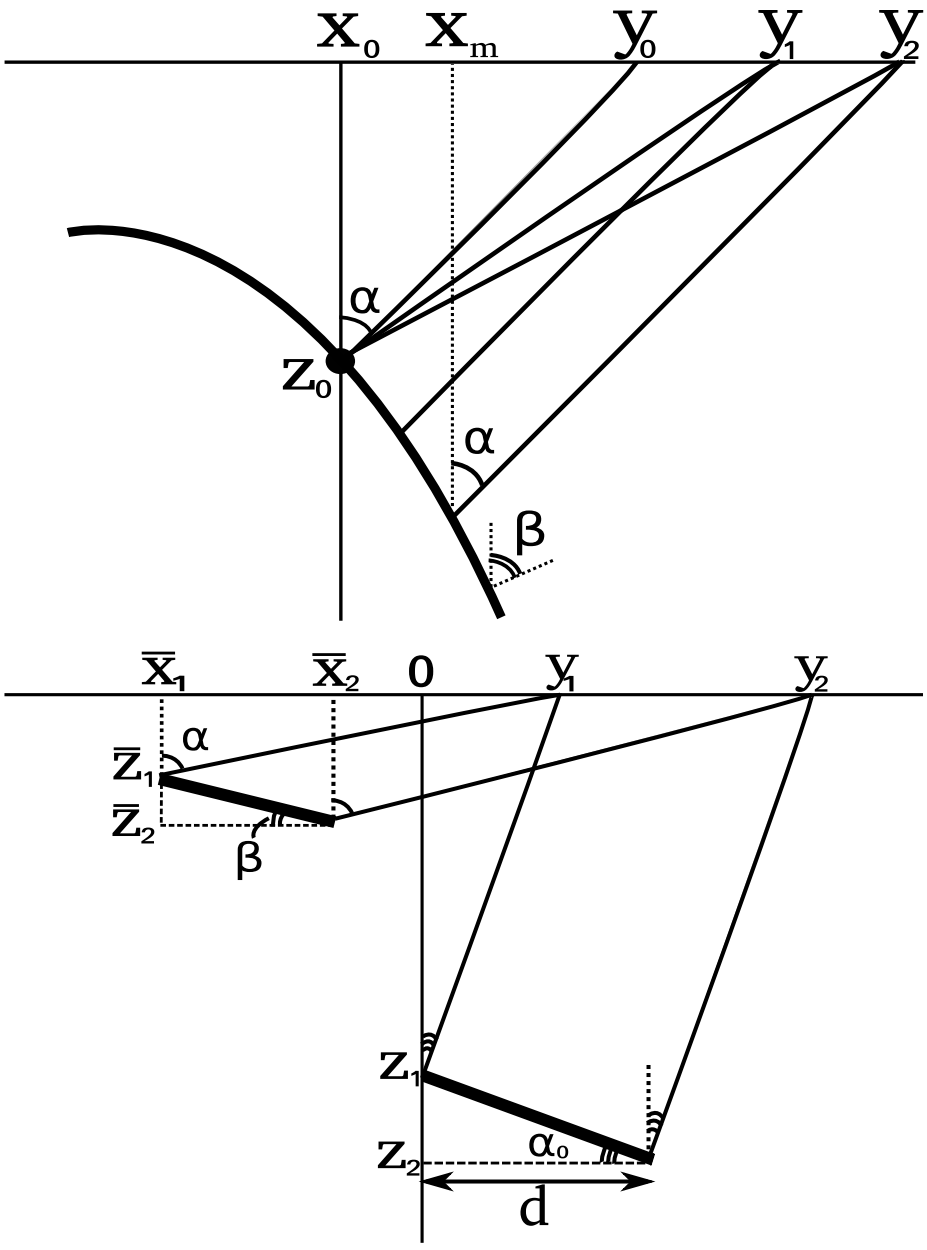


Figure 2: Constant-dip migration (scheme) of (a) a diffraction point and (b) a reflection boundary. optapert/. scheme-diffr,scheme-refl

reflection event is consistent with the constant-dip partial image, as it is for diffractions. The derivation in the Appendix shows that this fact is true for curved reflectors as well.

To verify the assertions presented, we ran the following experiment. We put two reflectors and one scattering point in the constant-velocity field (Figure 3). One reflector is flat, and the second has a dip of 15 degrees. We assume that reflectivity is generated by density variations. The modeled zero-offset section is presented in Figure 4. We migrated the data using Kirchhoff common-angle migration with the correct velocity. Figure 5 displays the image and a dip-angle gather extracted from the position where the scattering point is located. The diffraction event appears flat in the dip-angle gather, as expected, while the reflection events have concave shapes. Figure 6 shows two partial images corresponding to 0 and 15 degrees. Note that reflection boundaries and an effective part of the diffraction event are consistent with the partial image — 0 and 15 degrees, respectively (the slope is indicated by dashed lines).

To represent a realistic situation, we placed the same reflectors and scattering point within a smoothed velocity field from the Marmousi model. Both reflection and diffraction events have complicated shapes in the data domain (Figure 7). Figure 8 displays the image and the dip-angle gather after migration with the correct velocity. The traveltimes for Kirchhoff angle-domain migration were computed by Huygens wavefront-tracing (22), which allowed us to handle multiple arrivals. The diffraction event is flat, as it was in the constant-velocity example, but the shape of the reflection events appears much more complicated. The gather is contaminated by strong migration artifacts that produce noise in the image. Figure 9 shows two partial images corresponding to 0 and 15 degrees. Both images appear to be quite noisy. However, note that, even after migration with a complicated-velocity model, effective reflection events are consistent with the dips of partial images. The diffraction event at the scattering-point position has the appropriate slope as well (the slopes are indicated by dashed lines).

An event-slope-consistency analysis may provide information about separation between constructive and nonconstructive parts of migrated data. The key observation is that constructive reflection and diffraction events have slopes that are close to the constant-dip of the partial image. Slopes of nonconstructive events differ significantly from those of partial-image dips. Integration of constructive events only corresponds to migration having an optimal aperture.

SLOPE-CONSISTENCY ANALYSIS

One approach to analyzing constant-dip partial-image consistency is to measure the local slope in every partial image and to construct a taper function that is based on this measurement. The taper should protect events whose slope is close to the corresponding constant dip. Slopes that differ significantly from that of the constant

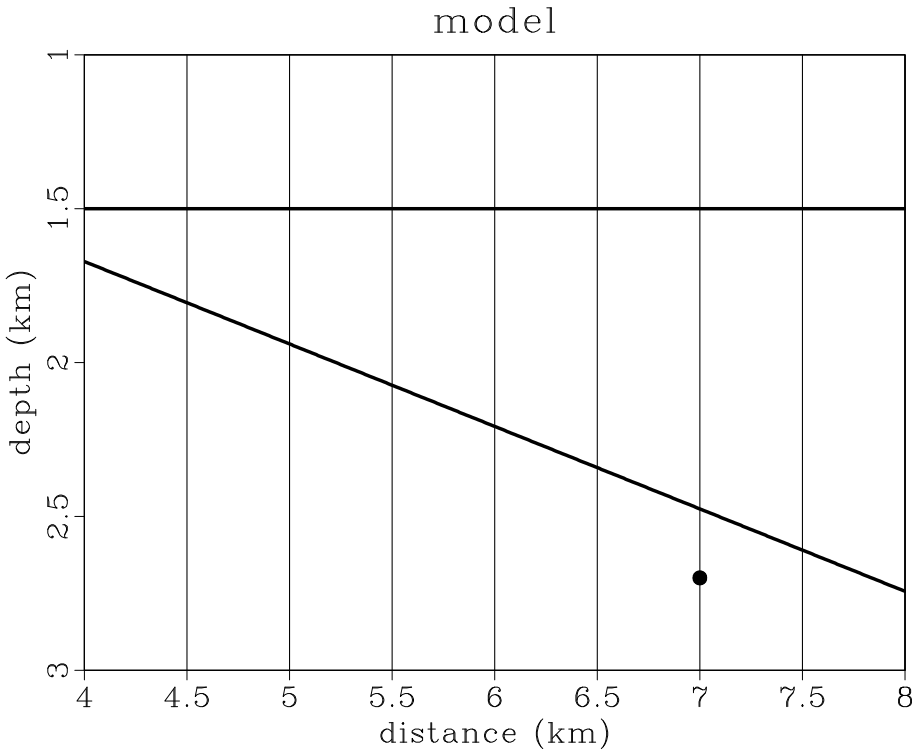


Figure 3: Theoretical model with two plane reflectors and one diffractor. optapert/theo theo-model

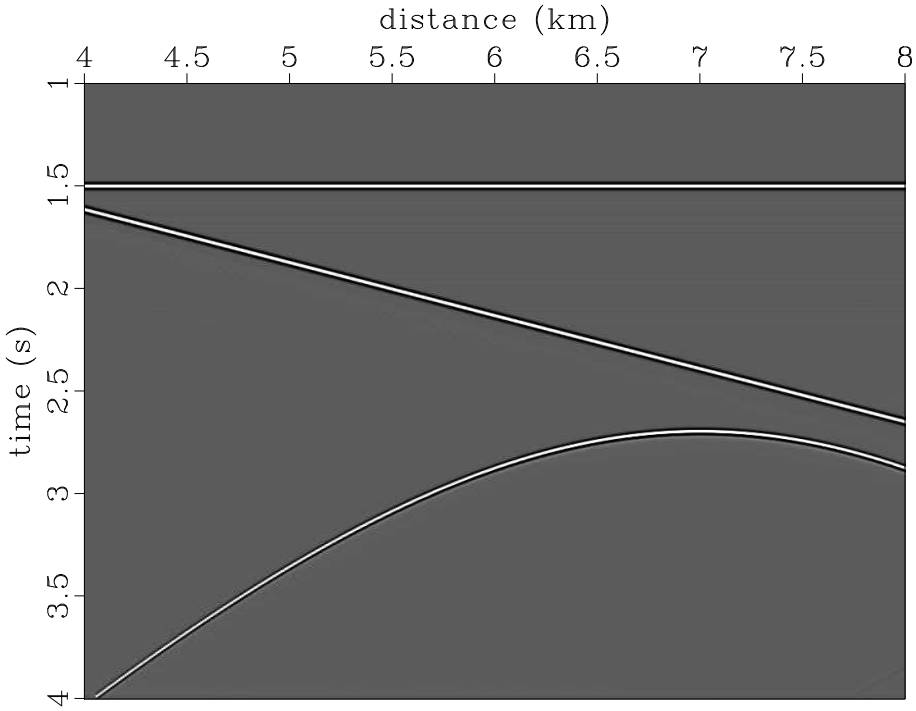


Figure 4: Zero-offset section corresponding to the constant-velocity model. optapert/theo data0

dip should be attenuated. However, if conflicting dips occur in the partial image, dips estimation becomes a non-trivial task.

To avoid this issue, we estimate segment conformity with a partial image by stacking along a local trajectory defined by partial-image constant dip. Constructive segments match the trajectory that yields high values of coherency. On the other hand, destructive segments get stacked in the wrong direction and provide low values of coherence. Classic semblance (30), as a coherence measure, can be used for defining a weight function. Weighting of migrated data corresponds to migration aperture optimization.

For illustration, we use the constant-dip partial images shown in Figure 9. They correspond to migration dips of 0 and 15 degrees, and we first apply local stacking along directions corresponding to 0 and 15 degrees, respectively. We extend stacking to the dip-angle direction as well. Effective contributions are locally plane in this direction, and local stacking along migration dips allows for signal enhancement. We measure coherency of the stacked data and get the semblance function $S(\alpha, z, x)$ presented in Figure 10. High semblance values correspond to events consistent with the dips of partial images. The next step is to transform the semblance function to a weight function $W(\alpha, z, x)$, which should optimally handle migrated amplitudes. The weight should not change a plane diffraction event or an effective apex area of reflections. At the same time, the weight should smoothly attenuate away from the

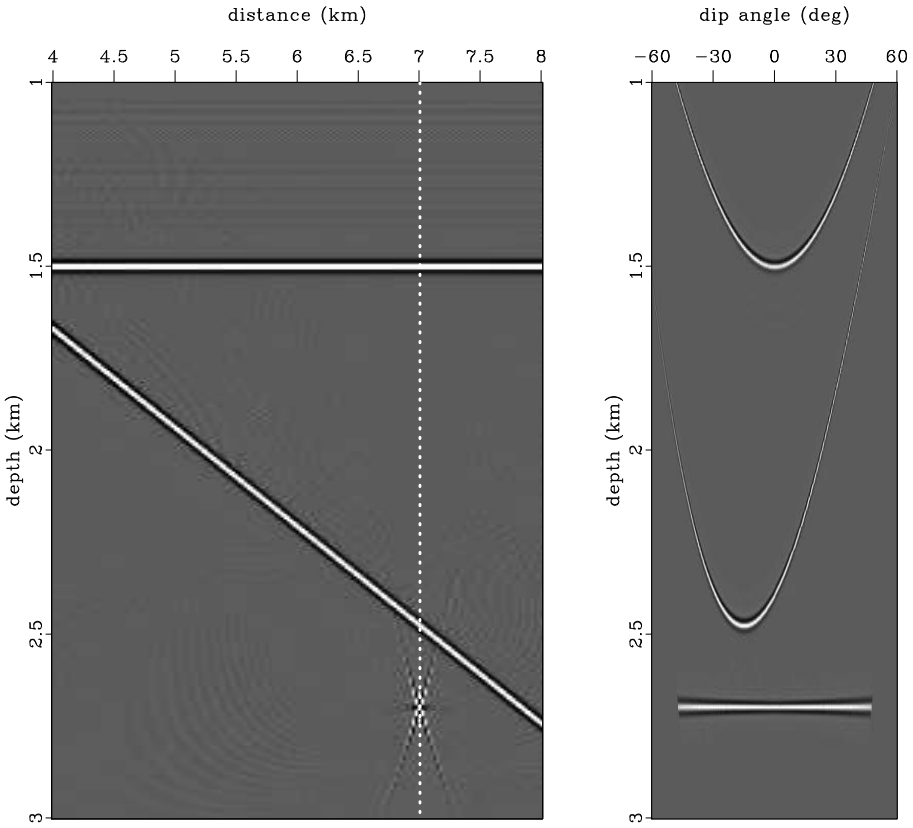


Figure 5: Depth-migrated image and common-image-gather in dip-angle domain corresponding to distance of 7 km. optapert/theo zo-res-init

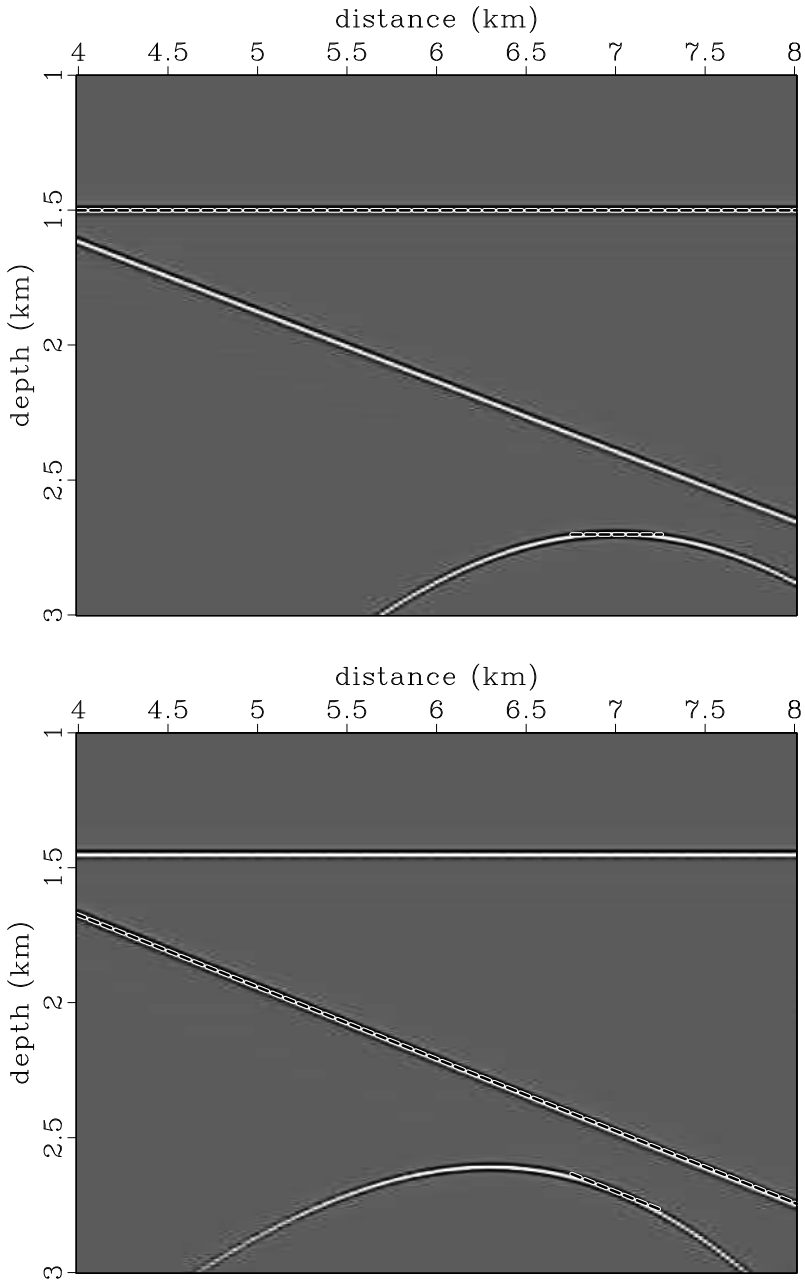


Figure 6: (a) Zero-degree partial image and (b) fifteen-degree partial image for the constant-velocity model. optapert/theo zodpis0,zodpis15

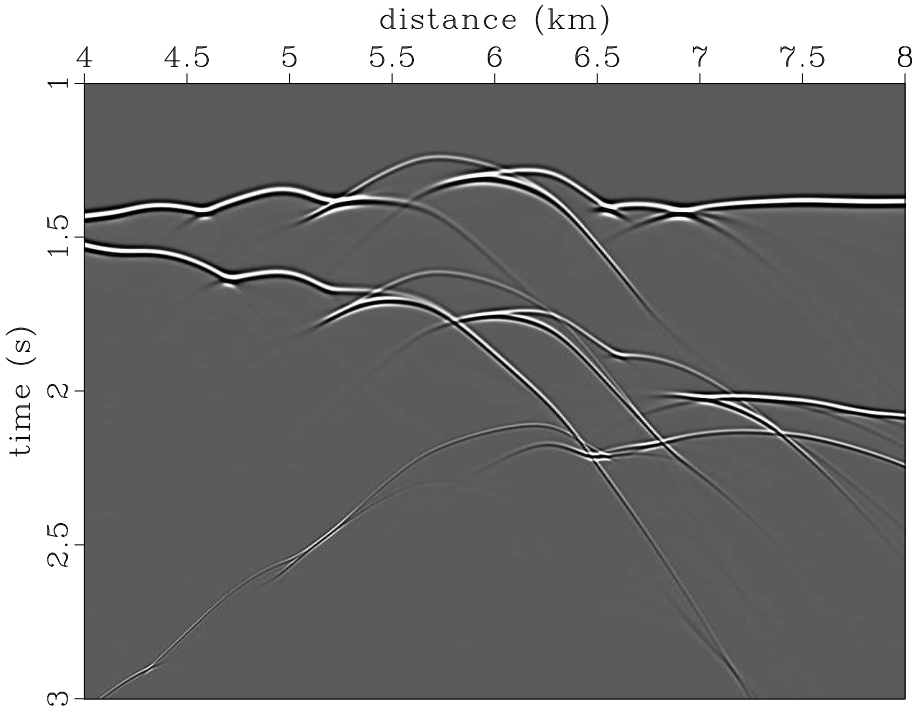


Figure 7: Zero-offset section corresponding to the Marmousi-velocity model. optapert/depthMig dm-zo

apexes so that edge effects can be eliminated (11). We define our weight function using the following thresholding rule:

$$W(\alpha,z,x)=\begin{cases} 1 & \text{for} \quad S(\alpha,z,x)>s_2\ , \\ \frac{S(\alpha,z,x)-s_1}{s_2-s_1} & \text{for} \quad s_1<S(\alpha,z,x)<s_2\ , \\ 0 & \text{for} \quad S(\alpha,z,x)<s_1\ , \end{cases} \tag{10}$$

where s_1 and s_2 define two thresholds. Thus, weight function is defined by summation bases and thresholding parameters. In complex inhomogeneous media, these parameters depend on migrated data and can be estimated by trial and error. Local summation in the dip-angle direction can be performed along constructive parts of the reflection concave event, whose size is defined after analysis of migrated dip-angle gathers. An optimal lateral window should suppress migration artifacts at a minimal computational cost. The semblance function in equation 10 appears to have smooth features that simplifies a heuristic choice of the threshold parameters.

In our synthetic example, we used a dip-angle window of 10 degrees, a lateral window of 150 m, and parameters s_1 and s_2 equal to 0.2 and 0.4, respectively. Figure 11 shows the weighted partial images. Reflectors consistent with the partial images, as well as appropriate parts of the diffraction event, are preserved. At the same time, a large amount of noneffective contribution, which may produce migration noise, has been eliminated. Summation of the weighted partial images provides the stacked image, which has the migration artifacts being suppressed significantly (Figure 12).

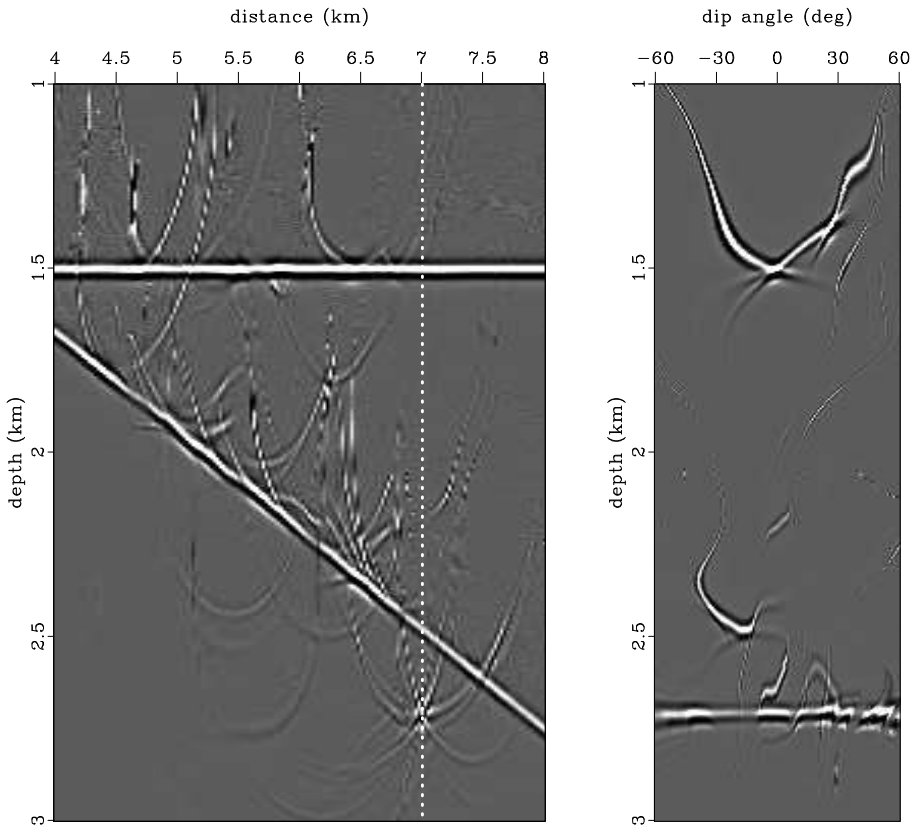


Figure 8: Depth-migrated image and common-image-gather in dip-angle domain corresponding to distance of 7 km. optapert/depthMig res-init

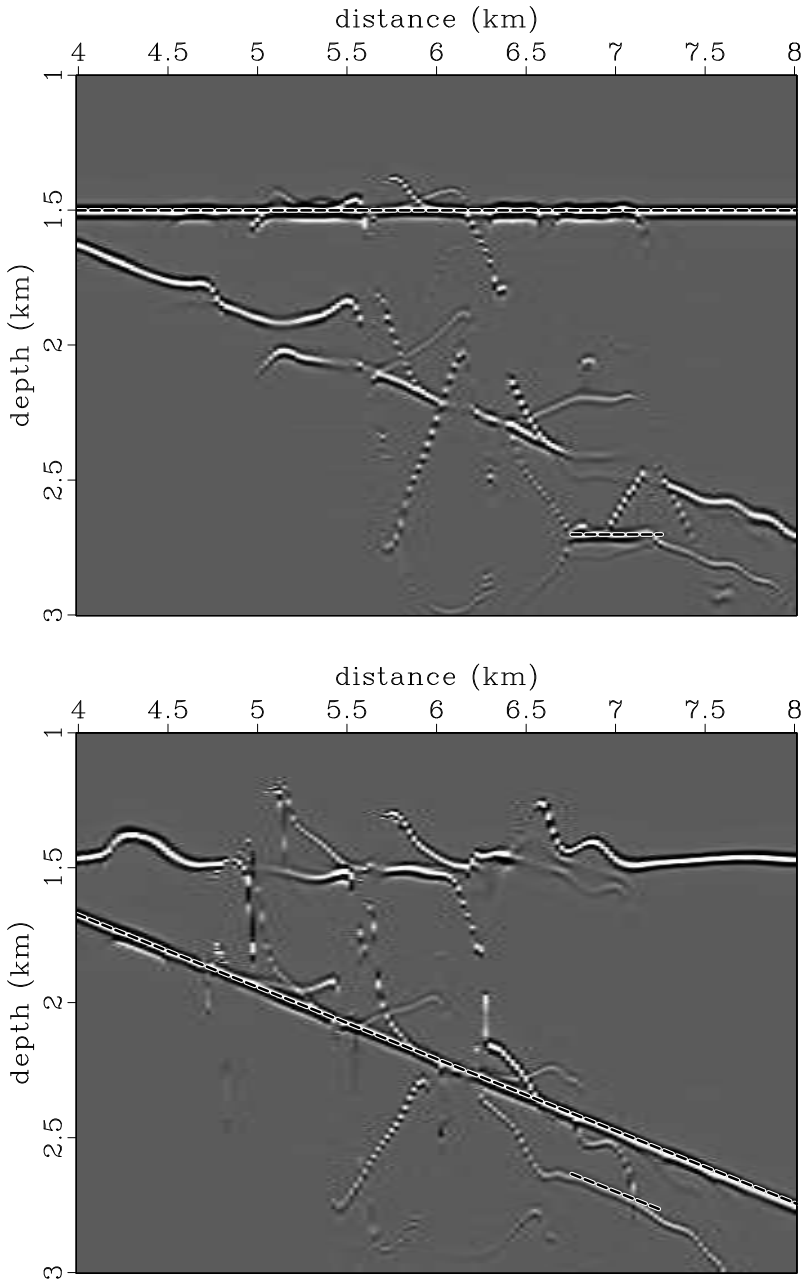


Figure 9: (a) Zero-degree partial image and (b) fifteen-degree partial image for the Marmousi-velocity model. optapert/depthMig dpis0,dpis15

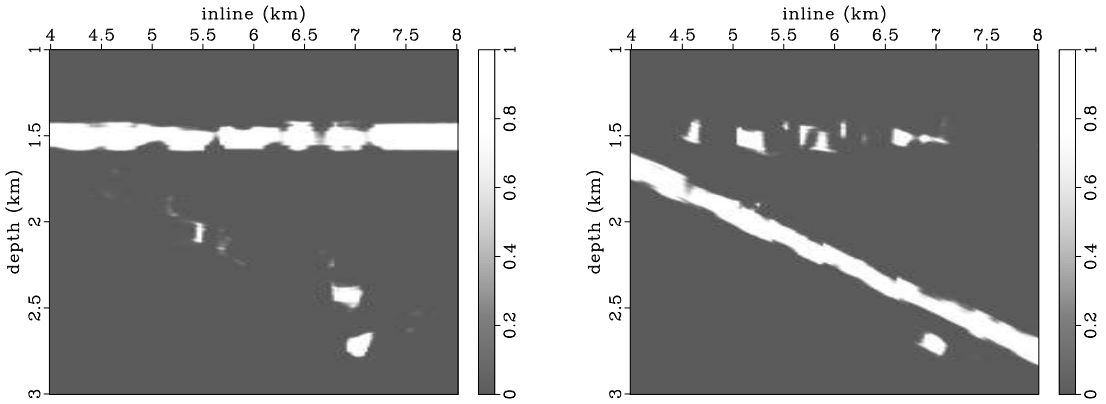


Figure 10: Semblance function for (a) zero-degree partial image and (b) fifteen-degree partial image. optapert/depthMig semb-0,semb-15

A weight function constructed by using the dip-angle direction only allows for attenuating reflection events around their apex areas while preserving diffraction events as well. Figure 13 shows the result of migrated data weighting using local summation along dip-angle window of 10 degrees. Reflection events are localized around their effective areas, and the diffraction event has its full width being preserved. However, the image remains contaminated by migration artifacts, which happen to be stationary in the dip-angle direction.

We next address the situation when migration velocity is not correct. Figure 14 shows the image and the dip-angle gather after migration with 10% lower velocity. Reflection boundaries look distorted. The diffractor is undermigrated and has a complicated shape in the dip-angle gather. The image is contaminated by strong migration artifacts. For migration-aperture optimization by slope-consistency analysis, we used the same parameters as in the previous example (Figure 15). Reflection and diffraction events both are optimized to their effective parts, eliminating migration noise in the image.

These experiments demonstrate that slope-consistency filtering can be used after migration even when using a complicated velocity model or an incorrect model. In the previous section, we considered theoretically the case of the zero-offset migration. The proposed principles may apply also to prestack migration in which stacking of the full offset range provides accumulation of effective migrated data parts. The experiments also reveal that noneffective segments may have an appropriate slope as well. For instance, the optimized fifteen-degree partial image (Figure 11(b)) has some events at a depth of 1.5 km. These events appear to have a correct slope that is close to 15 degrees, and they therefore evade nonconsistent-slope filtering and bring noise to the image of the flat reflector. This kind of noise is inherent in Kirchhoff migration and is produced by the complexity of the velocity model and incompleteness of the migrated data (26). In practice, it can be eliminated by stacking the full range of

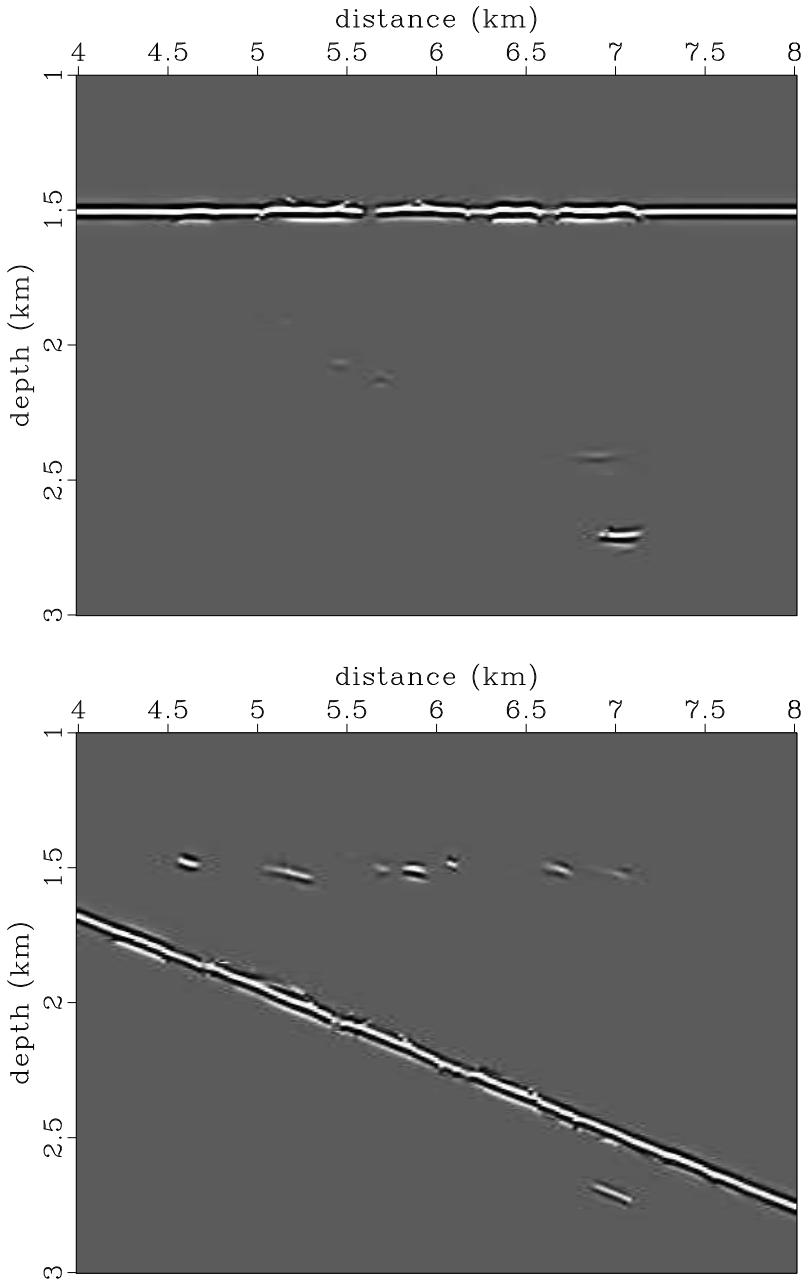


Figure 11: (a) Zero-degree partial image and (b) fifteen-degree partial image after migration-aperture optimization. optapert/depthMig dpis-cln-0,dpis-cln-15

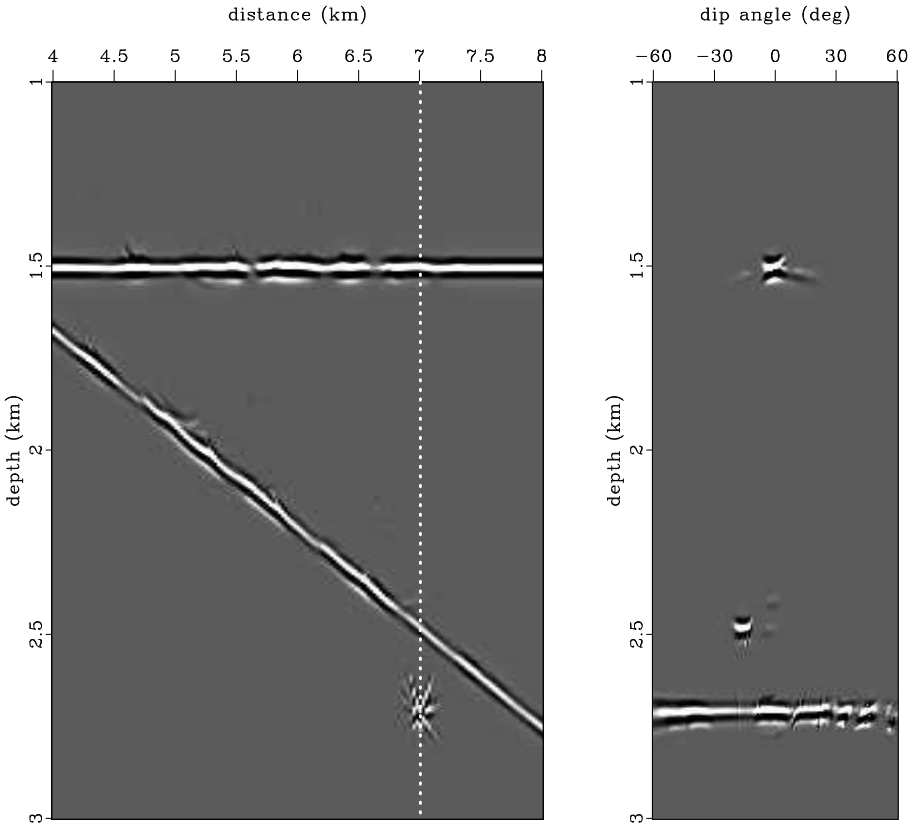


Figure 12: Image and dip-angle gather after migration-aperture optimization by slope-consistency analysis. optapert/depthMig res-clean

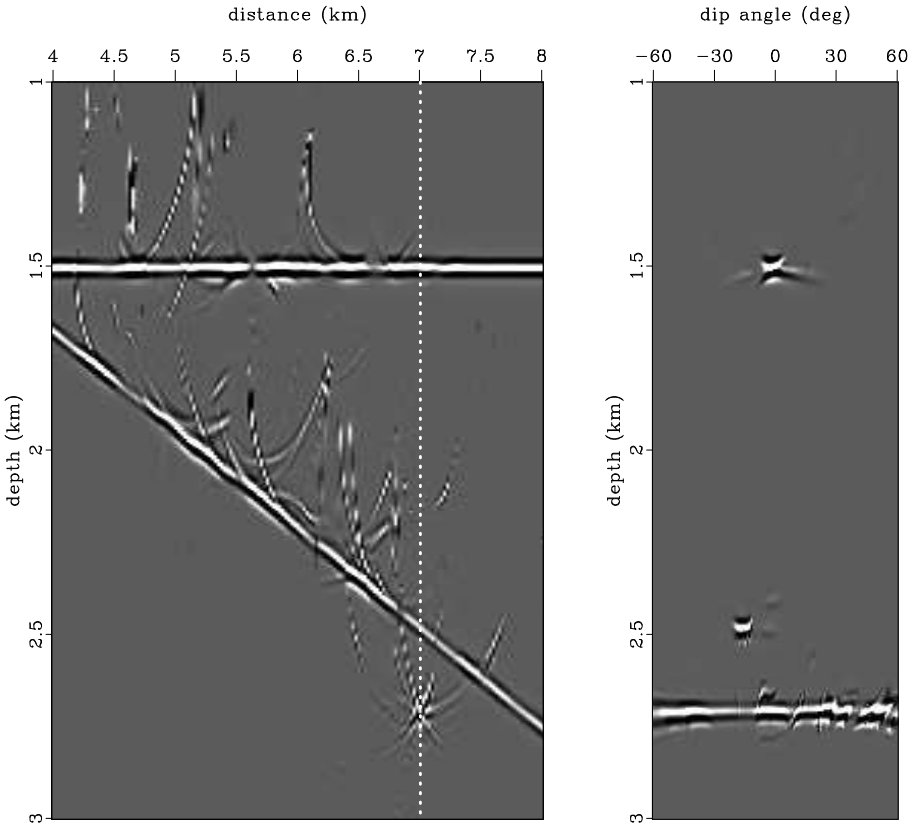


Figure 13: Image and dip-angle gather after migration-aperture optimization by local summation in the dip-angle direction. optapert/depthMig res-clean1

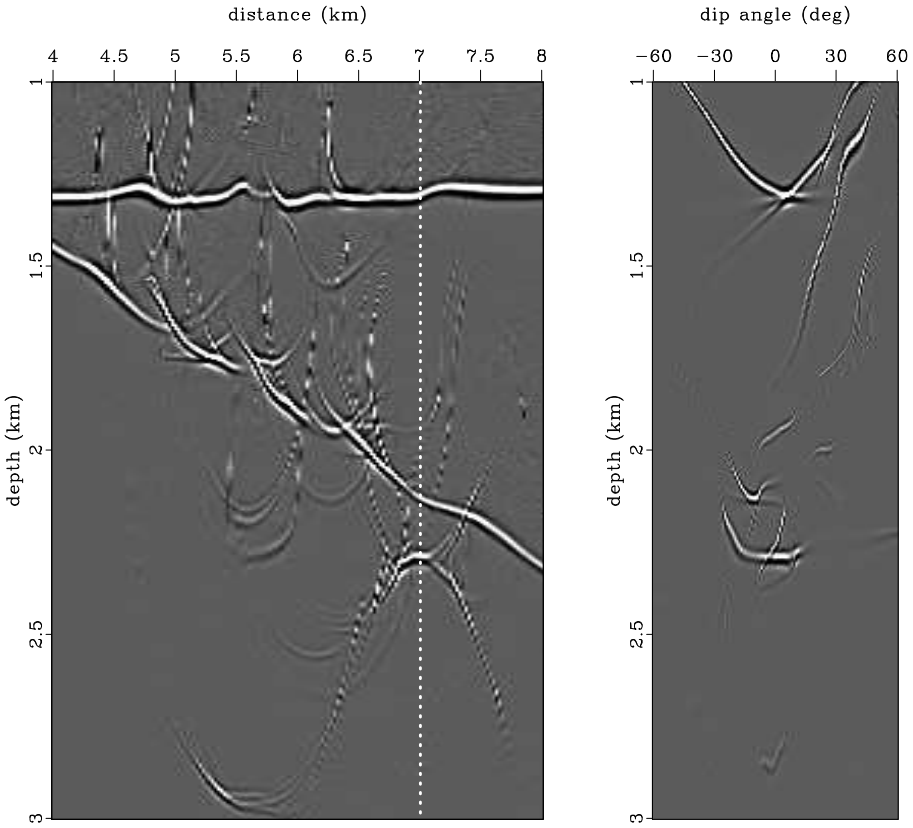


Figure 14: Initial image and common-image-gather in dip-angle domain corresponding to distance of 7 km (wrong velocity case). optapert/depthMig fres-init

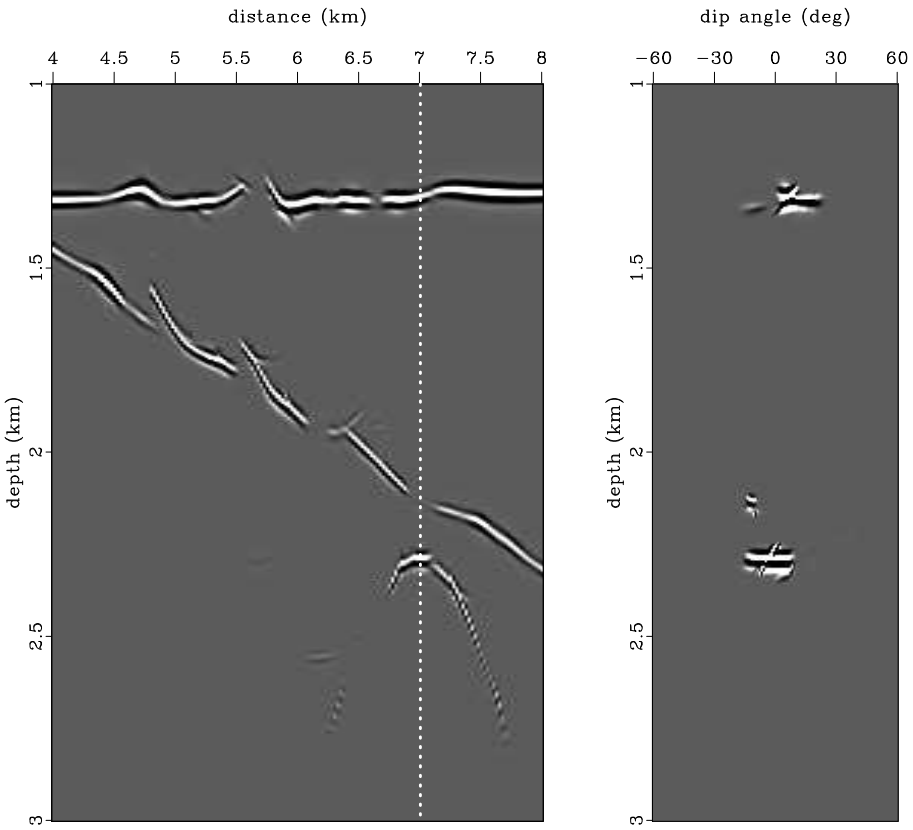


Figure 15: Image and dip-angle gather after migration-aperture optimization by slope-consistency analysis (wrong velocity case). optapert/depthMig fres-clean

offsets.

The proposed approach can be used easily in three dimensions, in which migrated events are defined by two orthogonal directions (15). Migrated constant-dip traces compose a 3D volume, and slope-consistency analysis should be performed along a plane surface whose slope is defined by the two migration dips. The method does not require slope estimation and has a relatively low computational cost.

Note that the effectiveness of the method is conditioned to that of semblance estimation. In the case of conflicting events, semblance operator effectively deals with events that are relatively strong to be detected. In practice, weak diffraction events might get suppressed by reflections and noise, which might cause difficulties in the slope-consistency estimation and, consequently, in the events protection. These weak events can be enhanced by diffraction imaging techniques (10; 15).

SYNTHETIC DATA EXAMPLES

We next tested the presented approach on prestack migration of the Sigsbee2B data set (20). The model contains a number of artificial point scatterers. The top of the salt body has a high curvature that acts as a transition between reflections and diffractions. There is a number of faults, which produce diffraction energy as well. Therefore, the data set is appropriate for testing of diffraction protection.

Figure 16 shows a seismic image of the left part of the model and one dip-angle gather extracted from the position in which two strong diffraction points are present. The gather contains two flat diffraction events at depths of 5.2 and 7.6 km, which correspond to the artificial scatterers. Some diffraction energy is also found at a depth of 4.3 km scattered by the fault.

Figure 17 shows the image and dip-angle gather after migration-aperture limiting. First, we detected reflection apex positions, estimated effective dip ranges around them, and rejected migrated amplitudes that did not correspond to effective dip intervals. This procedure enables elimination of migration noise with correct imaging of reflection boundaries (3). However, aperture limiting causes significant shortening of diffraction events and, as a result, weakening, or even disappearance, of diffraction objects in the image.

To optimize the migration aperture using the proposed slope-consistency approach, we performed semblance analysis using a dip-angle aperture of 10 degrees , a lateral base of 200 m, and parameters s_1 and s_2 equal to 0.2 and 0.4, respectively. (Figure 18). Reflection events become limited to the vicinity of their apexes, as in the previous example. At the same time, the procedure protects diffraction events in the gather and, hence, the corresponding diffraction objects in the image. Artificial point scatterers and faults are imaged correctly.

Figure 19(a) demonstrates one constant-dip partial image corresponding to 15 degrees. It resembles the seismic image of the model, with some differences. As

expected, dipping reflection boundaries are shifted from the correct positions and diffraction objects are represented by hyperbolic events. After slope-consistency analysis, we construct a filtering mask (Figure 19(b)) to protect events whose slope is close to that of the dip of the partial image (Figure 19(c)). As a result, diffraction hyperbolas became confined to the vicinities of their correct positions, where they have the appropriate slope. Faults at a depth of 4 km and other reflection boundaries were protected as well. Remaining migrated energy was eliminated significantly.

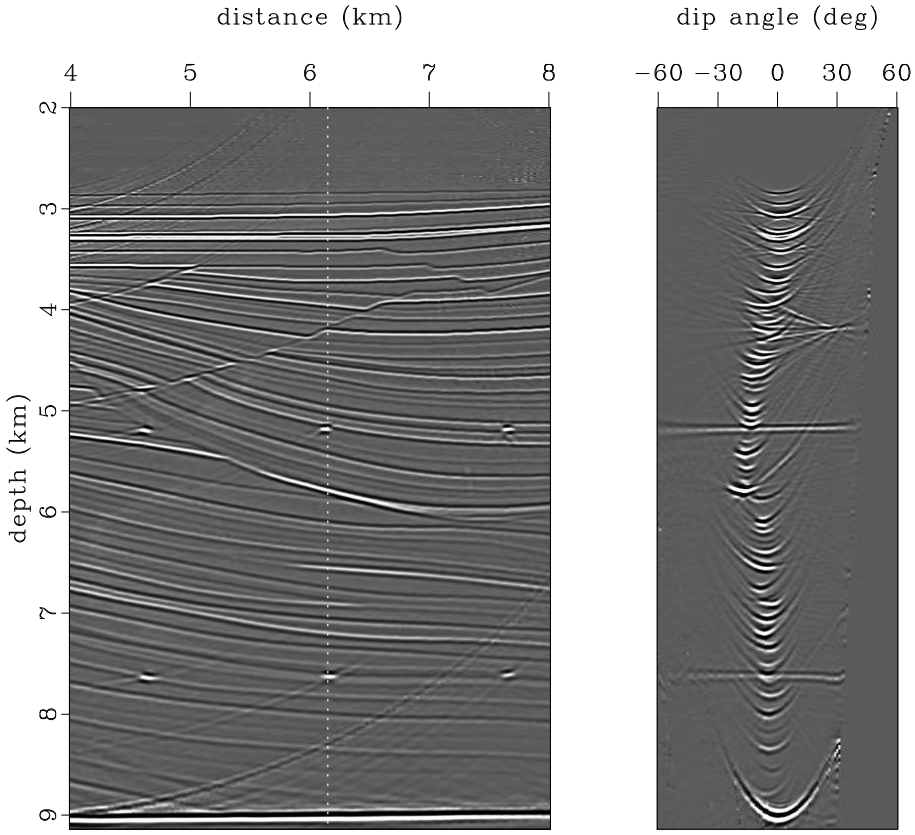


Figure 16: Initial image and common-image gather in dip-angle domain corresponding to distance of 6.1 km. optapert/sigsbee sigs-input

The salt body complicates Kirchhoff imaging in the right part of the Sigsbee model and increases the amount of migration noise. Figure 18 shows a seismic image and one dip-angle gather extracted from the position in which strong diffraction points are present. The scatterer at 7.6 km is illuminated well, and the dip-angle gather contains the corresponding plane diffraction event. A fault can be identified at the 15-km position and depth of 7 km. Some amount of migration noise is concentrated just below the salt and in the salt cavities above.

Migration-aperture optimization by slope-consistency analysis allows us to significantly decrease migration noise and, at the same time, to protect diffraction objects — artificial point scatterers, the fault, and the salt boundary are all imaged correctly (Figure 21). In comparison, simple aperture limiting leads to noise suppression as

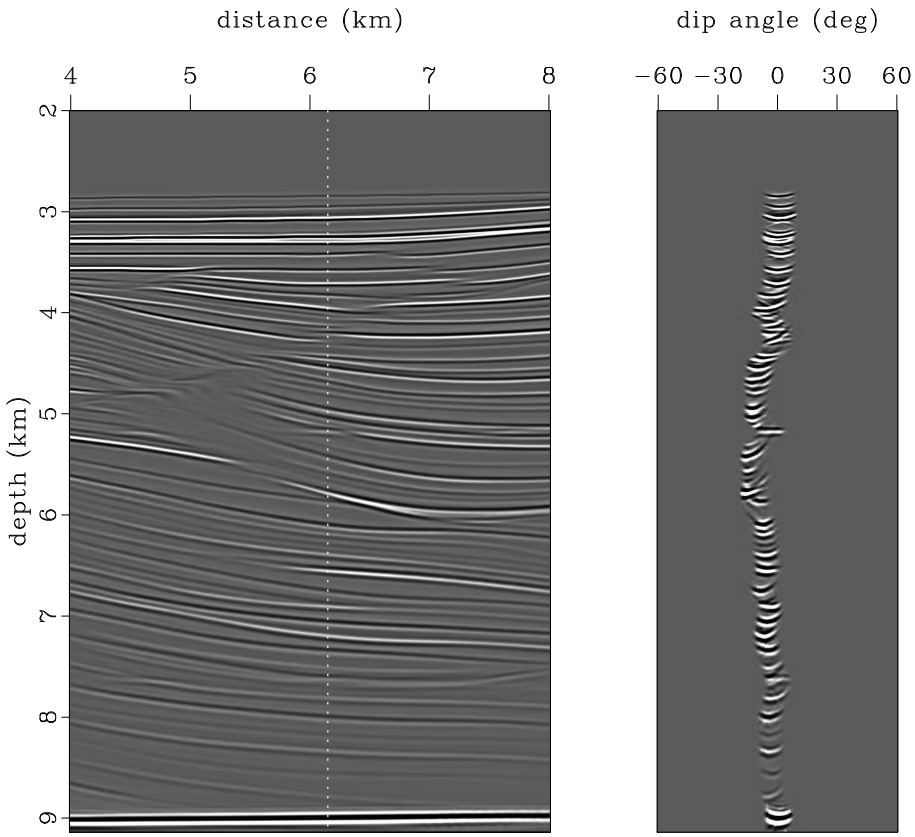


Figure 17: Image and dip-angle gather after migration-aperture limiting. optapert/sigsbee sigs-ap

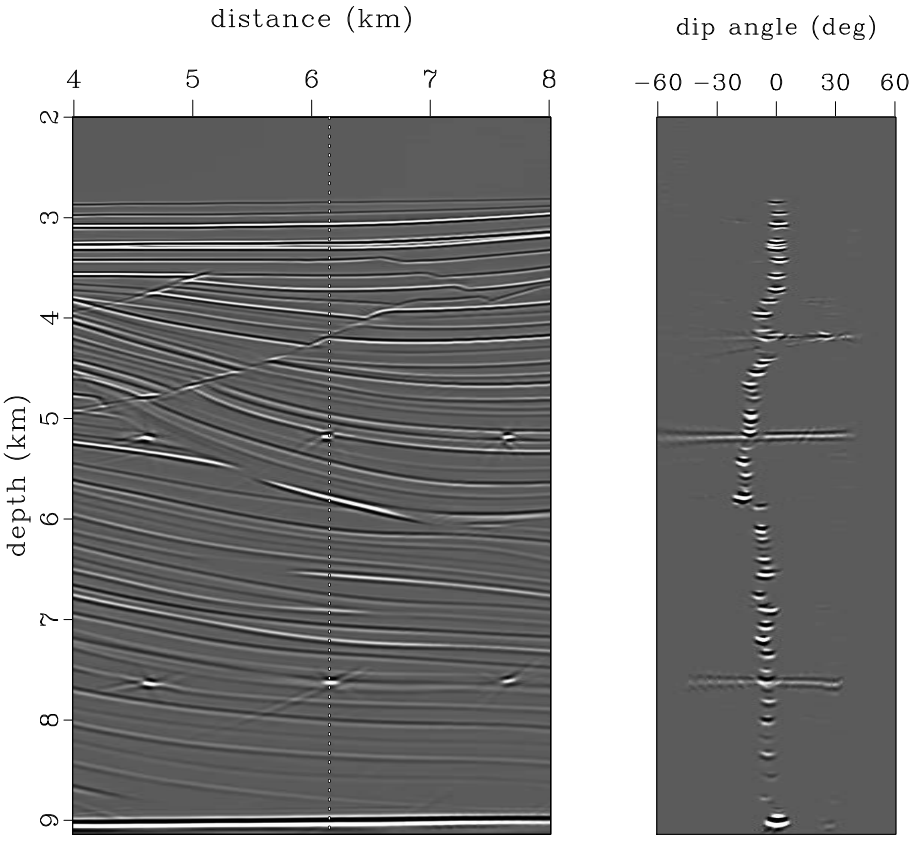


Figure 18: Image and dip-angle gather after migration-aperture optimization by slope-consistency analysis. optapert/sigsbee sigs-semb

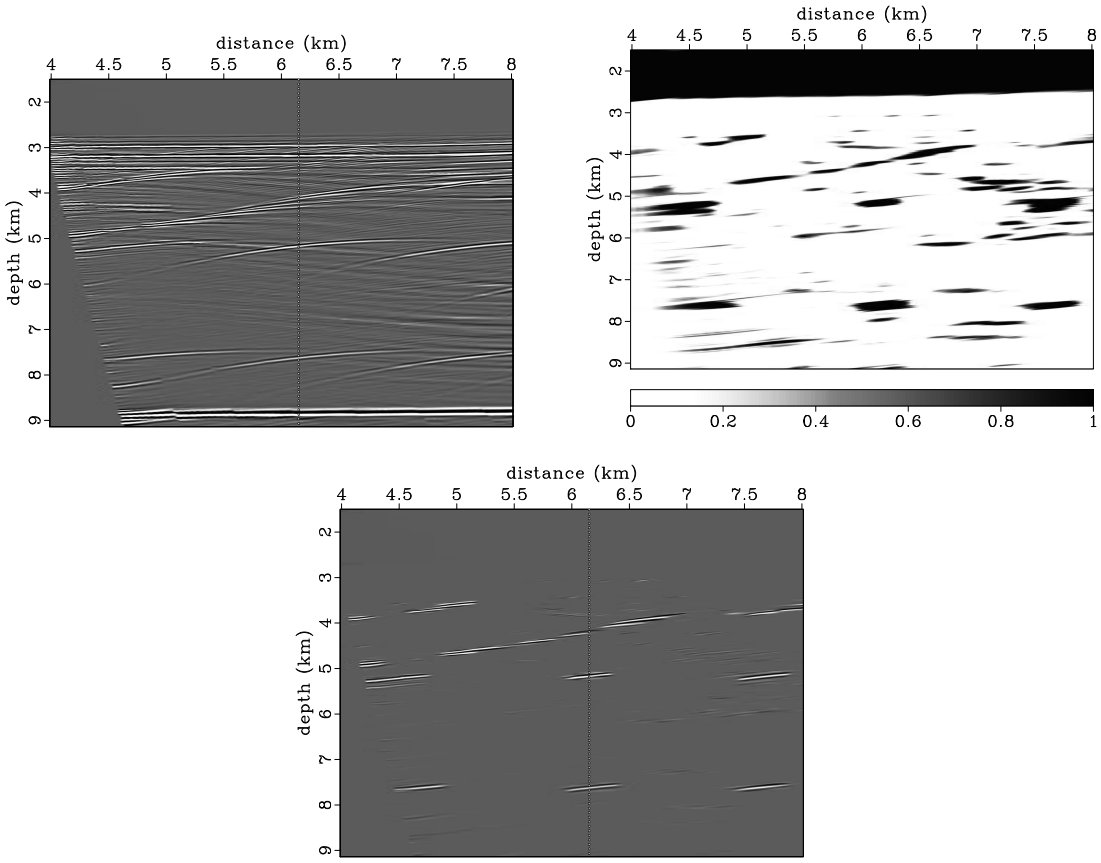


Figure 19: (a) Fifteen-degree constant-dip partial image, (b) weight function produced by slope-consistency analysis, and (c) partial image after migration-aperture optimization. Dashed line indicates position of considered artificial scattering points.
optapert/sigsbee pimage15-input-l,pimage15-mask,pimage15-semb-l

well; however, the diffraction objects are lost, and the salt body appears distorted (Figure 22).

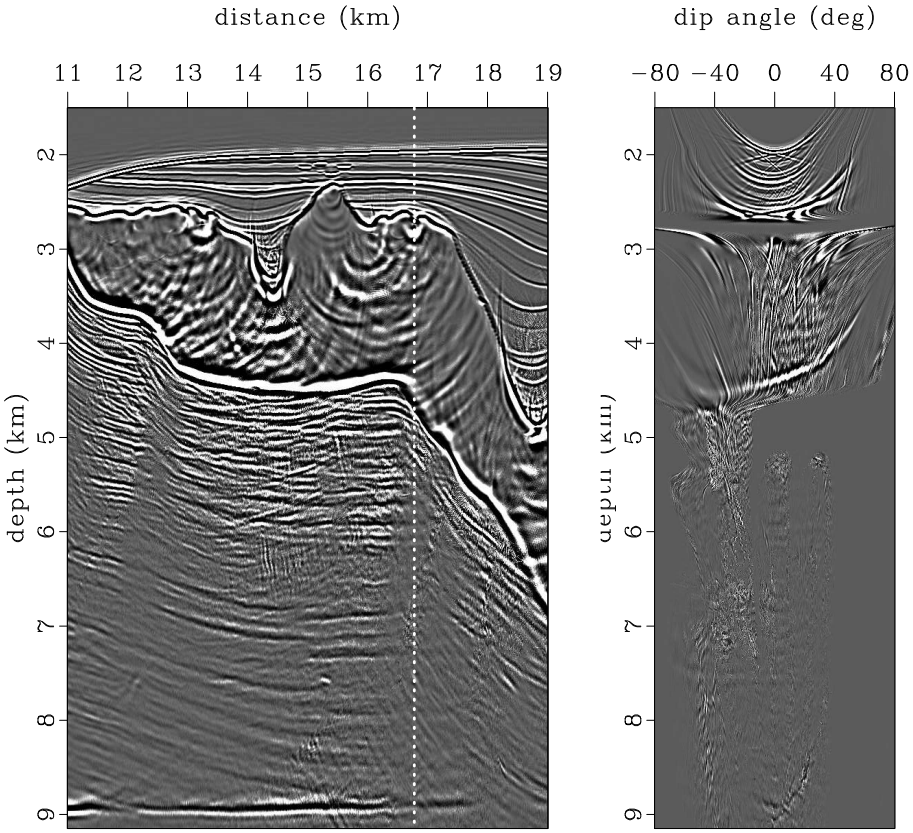


Figure 20: Initial image and common-image gather in dip-angle domain corresponding to distance of 16.8 km. optapert/rightSigsbee init

FIELD DATA APPLICATION

We next applied our aperture-optimization technique to a field 3D data set obtained from the Piceance Basin area in Colorado, USA. Figure 23 shows a part of the Williams Fork Formation, which is a well-known reservoir that contains a number of channels and point-bar sand bodies (21). (6) and (15) previously analyzed the data by diffraction imaging and showed that a number of diffraction objects were clustered in the top part of the volume. The image is corrupted by migration artifacts, and diffraction objects are buried under the noise.

We performed slope-consistency analysis using a 10-degree dip-angle windows, a 50-m-square stacking base, and parameters s_1 and s_2 equal to 0.1 and 0.3, respectively (Figure 24). Migration noise appears to be filtered out, whereas point objects in the shallow part are preserved. Diffraction objects are clearly observed in the slice.

For comparison, we migrated the data using a limited aperture of 10 degrees in

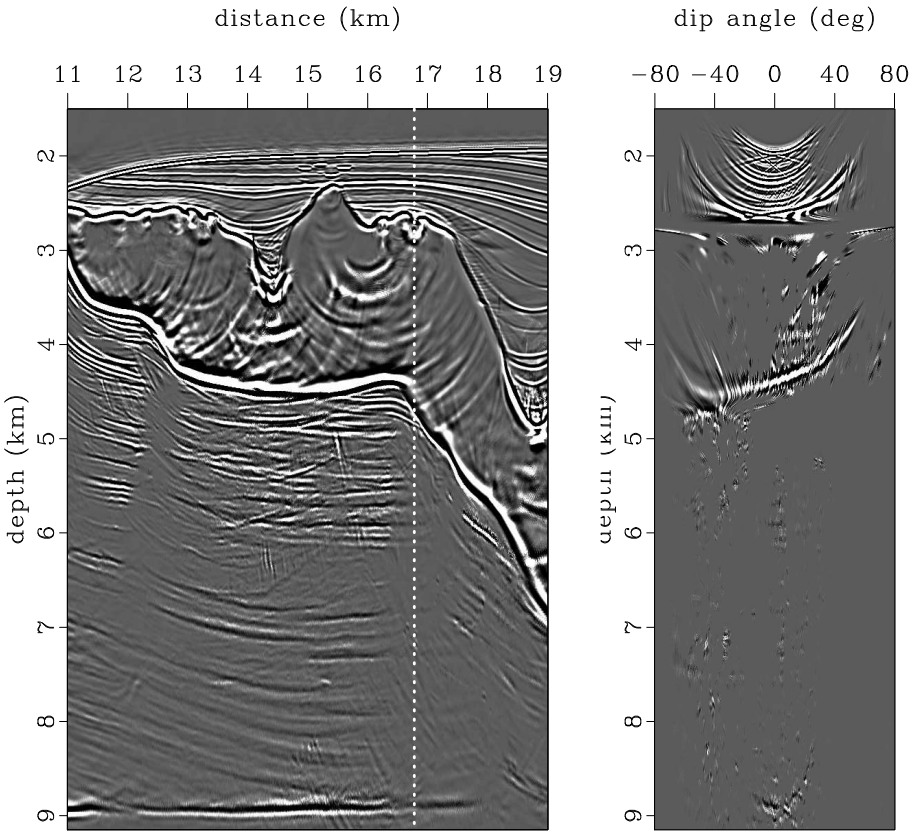


Figure 21: Image and dip-angle gather after migration-aperture optimization by slope-consistency analysis. optapert/rightSigsbee sca

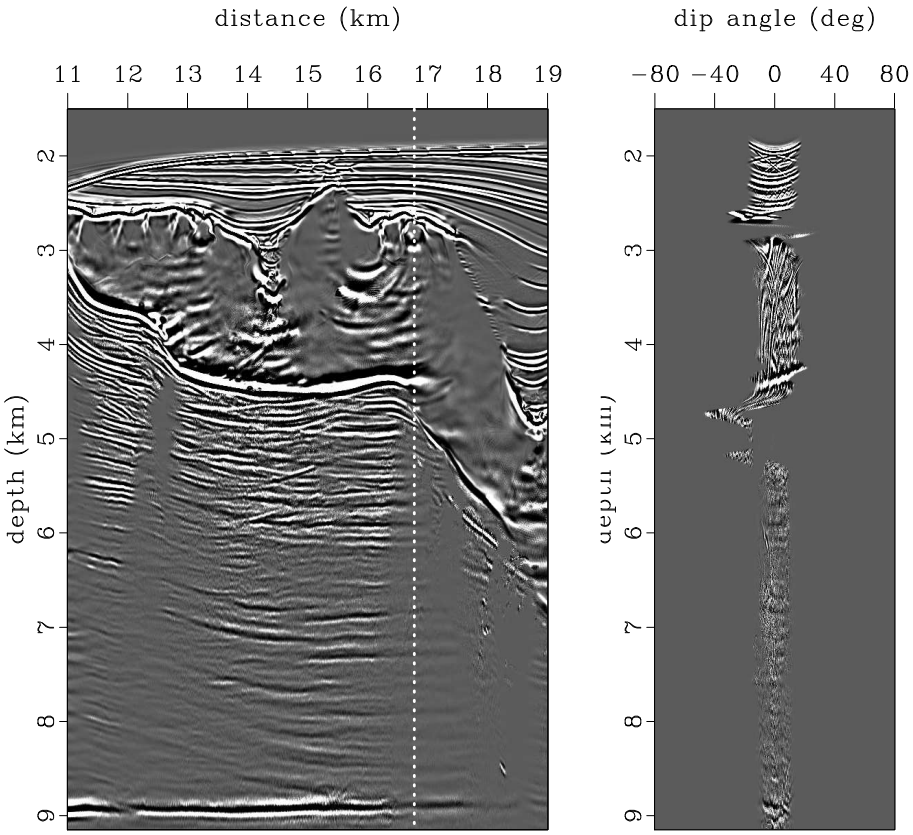


Figure 22: Image and dip-angle gather after migration-aperture limiting. optapert/rightSigsbee ap

two principal directions (Figure 25). Although most of the migration artifacts have been successfully eliminated, diffraction objects from the top part get smeared. These objects have low energy caused by scarcity of migration dips in the aperture.

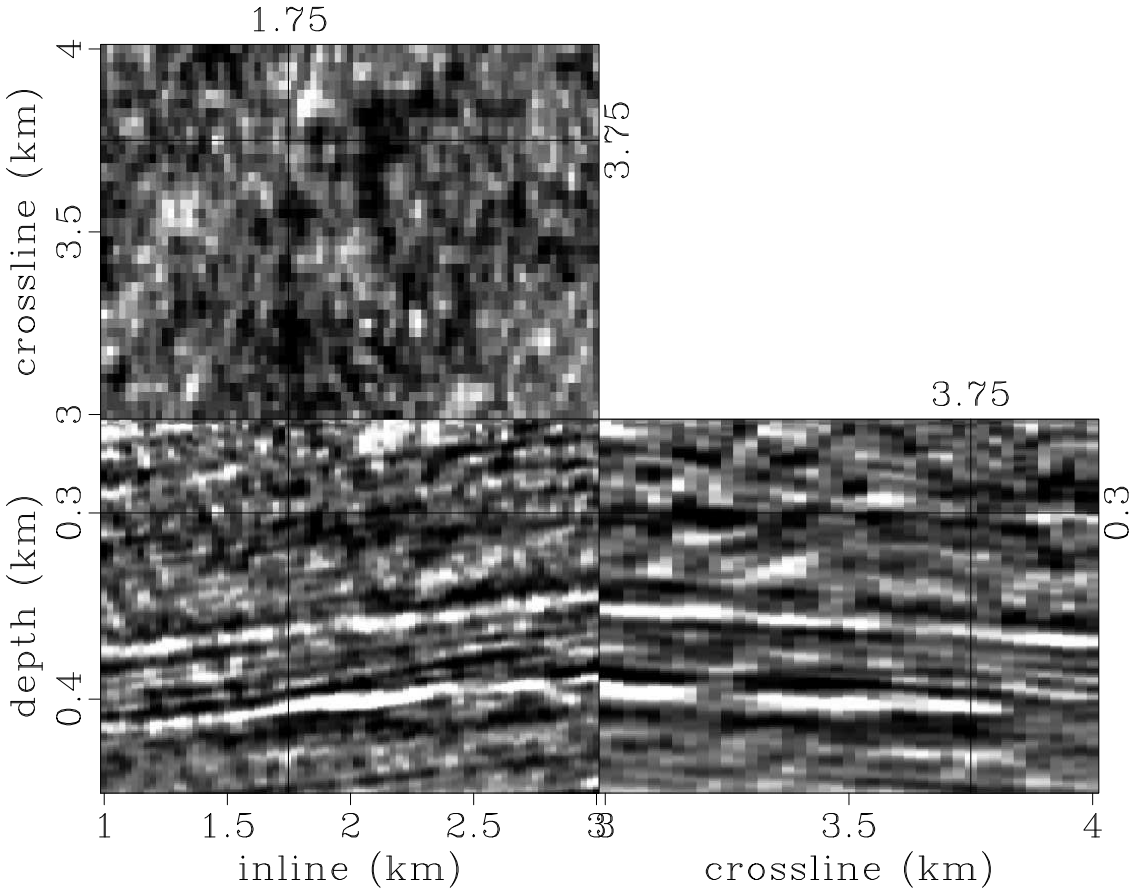


Figure 23: Initial 3D image for Piceance Basin data set. optapert/gibson image3

CONCLUSIONS

Migration-aperture optimization in Kirchhoff angle-domain migration can significantly enhance the quality of a seismic image. The aperture-optimization method needs to eliminate migration artifacts without distorting the useful signal. The diffraction component of the seismic wavefield characterizes small but important geological objects and brings extra resolution to seismic imaging. For an optimal image of diffraction objects, as wide a migration aperture as possible should be used. However, imaging of reflections requires an aperture narrowed around the tangent point.

To meet both requirements, we propose utilizing constant-dip partial images. An analysis of the consistency between local slopes and constant dips of a partial image makes evaluating the contribution of any part of the migrated data possible. This

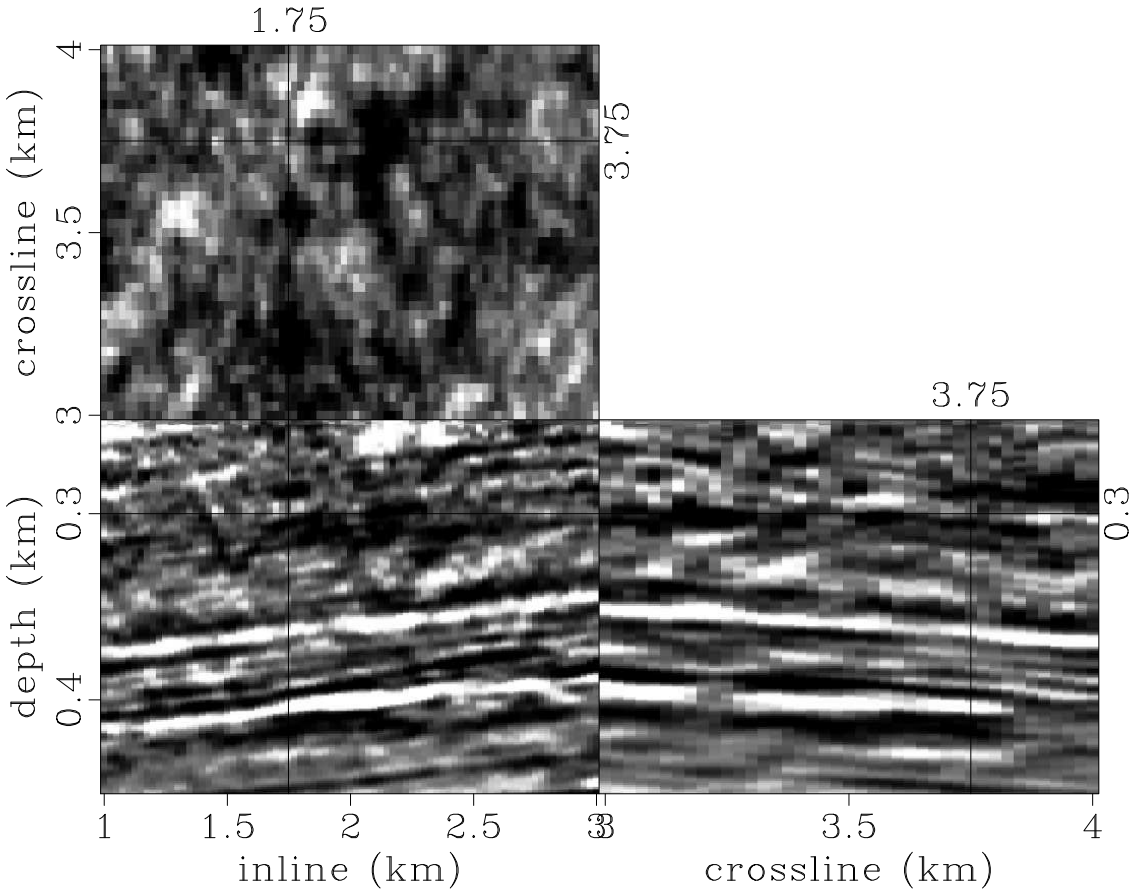


Figure 24: Image after migration-aperture optimization by slope-consistency analysis.
optapert/gibson image-clean

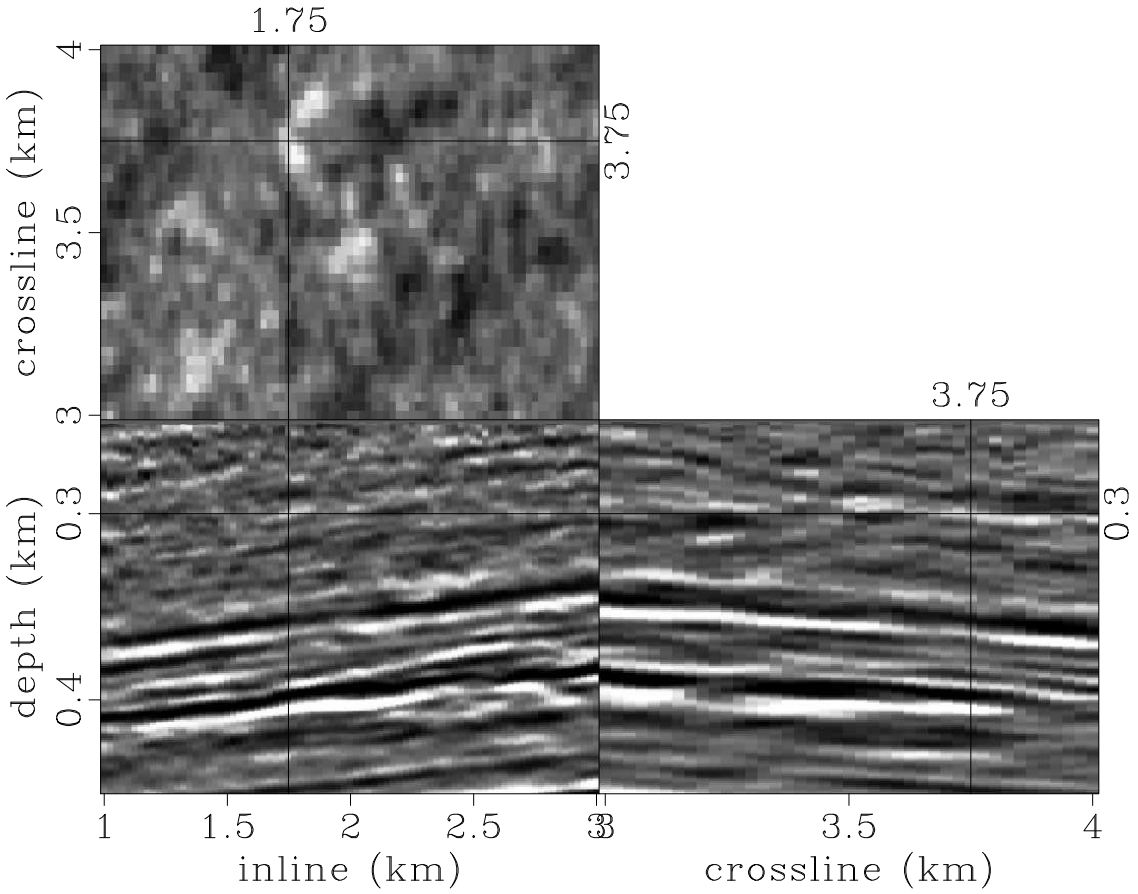


Figure 25: Image obtained after migration with limited aperture. optapert/gibson image3s

analysis allows us to extract constructive events. Stacking these events is equivalent to migration using an optimal aperture.

ACKNOWLEDGMENTS

We thank Igor Ravve, three anonymous reviewers, and the associate editor Jörg Schleicher for constructive suggestions that improved the paper.

Partial funding for this project was provided by RPSEA through the Ultra-Deepwater and Unconventional Natural Gas and Other Petroleum Resource program authorized by the US Energy Policy Act of 2005. RPSEA (`http://www.rpsea.org/`) is a nonprofit corporation whose mission is to provide a stewardship role in ensuring the focused research, development and deployment of safe and environmentally responsible technology that can effectively deliver hydrocarbons from domestic resources to the citizens of the United States. RPSEA, operating as a consortium of premier US energy research universities, industry, and independent research organizations, manages the program under a contract with the US Department of Energy's National Energy Technology Laboratory.

APPENDIX: HYPERBOLIC REFLECTOR IN THE DIP-ANGLE DOMAIN

For insight into the appearance of reflector images in the dip-angle domain, let us consider the case of a hyperbolic reflector (9). A special property of hyperbolic reflectors is that they can transform to plane dipping reflectors or point diffractors with an appropriate choice of parameters. Reflector depth is given by the function

$$z(x) = \sqrt{z_0^2 + (x - x_0)^2 \, \tan^2 \beta} \, , \tag{A-1}$$

and zero-offset reflection traveltime is given by

$$t(y) = \frac{2}{v} \sqrt{z_0^2 + (y - x_0)^2 \, \sin^2 \beta} \, . \tag{A-2}$$

When the reflector is imaged by time migration in the dip-angle domain (23) using velocity v_m, point $\{y, t\}$ in the data domain migrates to $\{x_m, t_m\}$ in the image domain according to

$$x_m = y - \frac{v_m}{2} \, t \, \sin \alpha = y - \frac{v_m \, \sin \alpha}{v} \sqrt{z_0^2 + (y - x_0)^2 \, \sin^2 \beta} \, , \tag{A-3}$$

$$t_m = t \, \cos \alpha = \frac{2 \, \cos \alpha}{v} \sqrt{z_0^2 + (y - x_0)^2 \, \sin^2 \beta} \, , \tag{A-4}$$

where α is the migration dip angle. Eliminating y from equations A-3 and A-4, we arrive at the equation

$$t_m(x_m) = \frac{2 \, \cos \alpha}{v} \, \frac{\gamma \, (x_m - x_0) \, \sin \alpha \, \sin^2 \beta + \sqrt{(x_m - x_0)^2 \, \sin^2 \beta + z_0^2 \, D}}{D} \, , \tag{A-5}$$

where $\gamma = v_m/v$ and $D = 1 - \gamma^2 \sin^2\alpha \sin^2\beta$. Equation A-5 describes the shape of the image of the hyperbolic reflector (A-1) in the dip-angle domain.

When the dip of the migrated event, imaged at a correct velocity ($\gamma = 1$),

$$\tan\alpha_m = \frac{v}{2}\, t'_m(x_m) = \frac{\cos\alpha \, \sin^2\beta}{D} \left[\sin\alpha + \frac{x_m - x_0}{\sqrt{(x_m - x_0)^2 \, \sin^2\beta + z_0^2 \, D}}\right] , \quad \text{(A-6)}$$

is equal to the dip of the image ($\alpha_m = \alpha$), it also becomes equal to the true dip of the reflector ($\alpha_m = \alpha_0$), where

$$\tan\alpha_0 = z'(x_m) = \frac{(x_m - x_0) \, \tan^2\beta}{\sqrt{z_0^2 + (x - x_0)^2 \, \tan^2\beta}} \, . \quad \text{(A-7)}$$

We can specify these conditions for two special cases described next.

Point diffractor

The hyperbolic reflector in equation A-1 creates a point diffractor at coordinates $\{x_0, z_0\}$ when $\beta = \pi/2$. In this case, equation A-5 simplifies to (15)

$$t_m(x_m) = \frac{2\,\cos\alpha}{v} \, \frac{\gamma\,(x_m - x_0)\,\sin\alpha + \sqrt{(x_m - x_0)^2 + z_0^2\,\left(1 - \gamma^2\,\sin^2\alpha\right)}}{1 - \gamma^2\,\sin^2\alpha} \, . \quad \text{(A-8)}$$

At a correct velocity ($\gamma = 1$),

$$t_m(x_m) = \frac{2}{v} \, \frac{(x_m - x_0)\,\sin\alpha + \sqrt{(x_m - x_0)^2 + z_0^2\,\cos^2\alpha}}{\cos\alpha} \, , \quad \text{(A-9)}$$

which is equivalent to equation 2 in the main text. The dip of the image is

$$\tan\alpha_m = \frac{v}{2}\, t'_m(x_m) = \tan\alpha + \frac{x_m - x_0}{\cos\alpha \, \sqrt{(x_m - x_0)^2 + z_0^2 \, \cos^2\alpha}} \, . \quad \text{(A-10)}$$

It is easy to verify that, above the diffraction point ($x_m = x_0$), $\alpha_m = \alpha$.

Plane dipping reflector

The hyperbolic reflector in equation A-1 becomes a plane dipping reflector when $z_0 = 0$. In this case, equation A-5 simplifies to (15)

$$t_m(x_m) = \frac{2}{v} \, \frac{(x_m - x_0)\,\cos\alpha\,\sin\beta}{1 - \gamma\,\sin\alpha\,\sin\beta} \, . \quad \text{(A-11)}$$

The dip of the image at a correct velocity is

$$\tan\alpha_m = \frac{v}{2}\, t'_m(x_m) = \frac{\cos\alpha\,\sin\beta}{1 - \sin\alpha\,\sin\beta} \, , \quad \text{(A-12)}$$

which is equivalent to equation 8 in the main text. The dip of the reflector in this case is simply $\alpha_0 = \beta$. It is easy to verify that when $\alpha = \beta$, $\alpha_m = \alpha$.

REFERENCES

[1] Alerini, M., and B. Ursin, 2009, Adaptive focusing window for seismic angle migration: Geophysics, **74**, S1–S10.

[2] Audebert, F., P. Froidevaux, H. Rakotoarisoa, and J. Svay-Lucas, 2002, Insights into migration in the angle domain: SEG Technical Program Expanded Abstracts, **21**, 1188–1191.

[3] Bienati, N., C. Andreoletti, F. Perrone, V. Lipari, and M. Giboli, 2009, Limited aperture migration in the angle domain: 71st EAGE Conference and Exhibition Extended Abstracts, **U005**.

[4] Bleistein, N., J. Cohen, and W. Stockwell, 2001, Mathematics of multidimensional seismic imaging, migration, and inversion: Springer.

[5] Bleistein, N., and S. Gray, 2002, A proposal for common-opening-angle migration/inversion: Center for Wave Phenomena, **CWP-420**, 293–303.

[6] Burnett, W., 2011, Multiazimuth velocity analysis using velocity-independent seismic imaging: PhD thesis, The University of Texas at Austin.

[7] Buske, S., S. Gutjahr, and C. Sick, 2009, Fresnel volume migration of single-component seismic data: Geophysics, **74**, WCA47–WCA55.

[8] Dafni, R., and M. Reshef, 2012, Image enhancement by CIG multi-parameter characterization: 74nd EAGE Conference and Exhibition.

[9] Fomel, S., and R. Kazinnik, 2013, Nonhyperbolic common reflection surface: Geophysical Prospecting, **61**, 21–27.

[10] Fomel, S., E. Landa, and M. T. Taner, 2007, Poststack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.

[11] Hertweck, T., C. Jäger, A. Goertz, and J. Schleicher, 2003, Aperture effects in 2.5D kirchhoff migration: a geometrical explanation: Geophysics, **68**, 1673–1684.

[12] Kabbej, A., R. Baina, and B. Duquet, 2007, Aperture optimized two-pass Kirchhoff migration: SEG Technical Program Expanded Abstracts, **26**, 2339–2343.

[13] Khaidukov, V., E. Landa, and T. J. Moser, 2004, Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution: Geophysics, **69**, 1478–1490.

[14] Klem-Musatov, K., 1994, Theory of seismic diffractions: Society of Exploration Geophysicists.

[15] Klokov, A., and S. Fomel, 2012, Separation and imaging of seismic diffractions using migrated dip-angle gathers: Geophysics, **77**, S131–S143.

[16] Landa, E., S. Fomel, and M. Reshef, 2008, Separation, imaging, and velocity analysis of seismic diffractions using migrated dip-angle gathers: SEG Technical Program Expanded Abstracts, **27**, 2176–2180.

[17] Leveille, J. P., I. F. Jones, Z.-Z. Zhou, B. Wang, and F. Liu, 2011, Subsalt imaging for exploration, production, and development: A review: Geophysics, **76**, WB3–WB20.

[18] Lüth, S., S. Buske, R. Giese, and A. Goertz, 2005, Fresnel volume migration of multicomponent data: Geophysics, **70**, S121–S129.

[19] Neidell, N., 1997, Perceptions in seismic imaging part 2: Reflective and diffractive contributions to seismic imaging: The Leading Edge, **16**, 1121–1123.

[20] Paffenholz, J., B. McLain, J. Zaske, and P. J. Keliher, 2002, Subsalt multiple

attenuation and imaging: Observations from the Sigsbee2B synthetic dataset: SEG Technical Program Expanded Abstracts, 2122–2125.

[21] Pranter, M. J., A. I. Ellison, R. D. Cole, and P. E. Patterson, 2007, Analysis and modeling of intermediate-scale reservoir heterogeneity based on a fluvial point-bar outcrop analog, Williams Fork Formation, Piceance Basin, Colorado: AAPG Bulletin, **91**, 1025–1051.

[22] Sava, P., and S. Fomel, 2001, 3-d traveltime computation using huygens wavefront tracing: Geophysics, **66**, 883–889.

[23] Sava, P. C., and S. Fomel, 2003, Angle-domain common-image gathers by wavefield continuation methods: Geophysics, **68**, 1065–1074.

[24] Schleicher, J., P. Hubral, M. Tygel, and M. S. Jaya, 1997, Minimum apertures and Fresnel zones in migration and demigration: Geophysics, **62**, 183–194.

[25] Spinner, M., and J. Mann, 2007, CRS-based minimum-aperture time migration – A 2D land-data case study: SEG Technical Program Expanded Abstracts, **26**, 2354–2358.

[26] Stolk, C., and W. Symes, 2004, Kinematic artifacts in prestack depth migration: Geophysics, **69**, 562–575.

[27] Sun, H., and G. T. Schuster, 2001, 2-D wavepath migration: Geophysics, **66**, 1528–1537.

[28] Sun, J., 1998, On the limited aperture migration in two dimensions: Geophysics, **63**, 984–994.

[29] Tabti, H., L. J. Gelius, and T. Hellmann, 2004, Fresnel aperture prestack depth migration: First Break, **22**, no. 3, 39–46.

[30] Taner, M. T., and F. Koehler, 1969, Velocity spectra—digital computer derivation applications of velocity functions: Geophysics, **34**, 859–881.

[31] Tillmanns, M., and H. Gebrande, 1999, Focusing in prestack isochrone migration using instantaneous slowness information: Pure and Applied Geophysics, **156**, no. 1, 187–206.

[32] Xu, S., H. Chauris, G. Lambaré, and M. Noble, 2001, Common-angle migration: A strategy for imaging complex media: Geophysics, **66**, 1877–1894.

A reversible transform for seismic data processing

William A. Burnett[1] *and Robert J. Ferguson*[2]

ABSTRACT

We use the nonstationary equivalent of the Fourier shift theorem to derive a general one-dimensional integral transform for the application and removal of certain seismic data processing steps. This transform comes from the observation that many seismic data processing steps can be viewed as nonstationary shifts. The continuous form of the transform is exactly reversible, and the discrete form provides a general framework for unitary and pseudounitary imaging operators. Any processing step which can be viewed as a nonstationary shift in any domain is a special case of this transform. Nonstationary shifts generally produce coordinate distortions between input and output domains, and those that preserve amplitudes do not conserve the energy of the input signal. The nonstationary frequency and time distortions and nonphysical energy changes inherent to such operations are predicted and quantified by this transform. Processing steps of this type are conventionally implemented using interpolation operators to map discrete data values between input and output coordinate frames. Although not explicitly derived to perform interpolation, the transform here assumes the Fourier basis to predict values of the input signal between sampling locations. We demonstrate how interpolants commonly used in seismic data processing and imaging approximate the proposed method. We find that our transform is equivalent to the conventional sinc-interpolant with no truncation. Once the transform is developed, we demonstrate its numerical implementation by matrix-vector multiplication. As an example, we use our transform to apply and remove normal moveout.

INTRODUCTION

It is well known that the Fourier transform has a variety of properties that make it useful for signal processing in geophysics (24). The Fourier transform essentially rearranges the data an input function into its complex frequency components (13). This can be done because continuous functions and their discretely-sampled equivalents can be expressed in terms of the Fourier basis. The organization of data in the Fourier domain allows many otherwise complex operations to be applied easily. Properties described by the Faltung, Weiner-Khintchine, and shift theorems (23) can exploit this organization because of another property, described by the inversion theorem (23), which states that the Fourier transform is exactly reversible.

Many individual seismic data processing steps also simply rearrange input data, and can therefore be viewed as transforms. Barring destructive processes such as muting or band-limited frequency filtering, even an entire data processing flow is itself simply transformations the input data into a corrected set. However, conventional methods used to perform processing steps of this type are usually not implemented as transforms. Instead, they are implemented with data-mapping algorithms that require interpolation (12).

Most conventional interpolation schemes contradict the primary assumption of the Fourier transform, as they do not assume the continuous equivalent of the recorded data is constructed with the Fourier basis. Instead, they assume some other basis in order to design efficient but approximate interpolants (13). These methods not only have the risk of immediately altering the data, but they are also inherently irreversible. Applying or removing processing steps in this way causes a small amount of information loss. With quantitative analysis methods as a goal of modern seismic data processing, it is important to preserve as much of the input information as possible during the flow.

In order to address the problem of data loss due to interpolation-based methods, we propose a reversible one-dimensional transform for the implementation of seismic data processing steps. The proposed transform provides a general framework for unitary and pseudounitary operators common to seismic data processing and imaging (2; 8). This framework is developed in the context of nonstationary filtering (14). Implementing imaging steps as reversible transforms does not require interpolation in the continuous case, and correctly assumes the Fourier basis in agreement with the Fourier transform for interpolation in the discrete case. Although several seismic data processing operations have been individually described by transforms of exactly this type (14; 16; 15; 8), the transform here is the general form (4; 7).

Since the Fourier basis is used, the forward transform is theoretically exactly invertible. Just as the inversion theorem quantitatively justifies the Fourier domain as a valid processing domain, a reversible transform for data processing steps justifies the use of corrected data sets as if they are in valid processing domains themselves. Any processing step that can be viewed as a nonstationary shift can be easily implemented by this transform. Seismic imaging steps are naturally useful for separating signal from noise, so they offer familiar, exploitable organizations of data (17; 25). Therefore, the proposed reversible transform for seismic data processing offers a useful set of quantitatively valid domains in which to work.

The general transform we derive is based on classical Fourier transform theory (19; 23), and the theory of nonstationary filtering (14). We first develop the general form for the forward data processing transform, and then derive the inverse of the transform. Following the transform development, we implement the normal moveout correction as an example, and then we discuss how conventional interpolation methods relate to the proposed transform.

THEORY

Many conventional seismic imaging steps are special cases of nonstationary shifts, which can be performed simultaneously with the Fourier transform via nonstationary convolution or combination (14). This section shows how the nonstationary shift operator can be combined with the Fourier transform kernel to create a general data processing transform kernel.

It is important to define a convention for the forward and inverse Fourier transform before proceeding. The shift-direction determined by the sign in a shifting exponential will reverse if the opposite sign convention for the Fourier transform is used. We use the forward Fourier transform convention,

$$F(\beta) = \int_{-\infty}^{\infty} f(p)\ e^{-i\beta p}\, dp, \tag{1}$$

where $F(\beta)$ is the complex Fourier transform (commonly a function of frequency) of the input signal, $f(p)$ (commonly a function of time). With the Fourier transform convention defined in equation 1, a positive sign in the shifting exponential causes a backward shift of the input function. The inverse Fourier transform,

$$g(q) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\beta)\ e^{i\beta q}\, d\beta, \tag{2}$$

has an output function, $g(q)$, that is strictly different from the original input function. A special property of the Fourier transform is that the input and output axes, p and q, are identical, and $f(p) = g(q)$, meaning that the input function passes unchanged when put through both the forward and inverse Fourier transforms alone.

In general, any processing step that requires a mapping from an input coordinate that is a function of the output coordinate, to that output coordinate, can be viewed as a nonstationary shift by combination. Examples of such processing steps include normal moveout (NMO), dip-moveout (DMO), and frequency-wavenumber (f-k) migration. NMO is a nonstationary shift in the time-space domain, while f-k migration performs a similar shift in the frequency-wavenumber domain. To remain general though, it is better at this point to think in terms of input and output coordinates rather than input and output times, so that frequency, wavenumber, space, or time shifts can be admitted, depending on the desired domain.

To emphasize that the input and output coordinates do not necessarily correspond to time, we use the above $p - q$ notation. A general nonstationary data processing shift in this notation is,

$$\Delta(q) = p(q) - q, \tag{3}$$

where p and q are the input and output coordinates, respectively. Using these general coordinates, the nonstationary data processing shift applied to an input function, $f(p)$, is,

$$g(q) \equiv f(q + \Delta(q)) = \frac{1}{2\pi} \int_{-\infty}^{\infty} F(\beta)\ e^{i\beta\Delta(q)}\, e^{i\beta q}\, d\beta, \tag{4}$$

where $F(\beta)$ is the Fourier transform of $f(p)$, β is the Fourier dual of p, and the output signal, $g(q)$ is the desired result of processing the input signal $f(p)$. Unlike the Fourier transform, p and q are not identical, although they represent the same dimension. The convenient form of $\Delta(q)$ allows the exponentials to be reduced by combining the Fourier transform kernel with the nonstationary shifting exponential. The result is a concise form of the forward data processing transform,

$$g\left(q\right) = \frac{1}{2\,\pi} \int_{-\infty}^{\infty} F\left(\beta\right)\, e^{i\,\beta\,p(q)}\, d\beta, \tag{5}$$

which confirms the desired result,

$$g\left(q\right) = f\left(p\right). \tag{6}$$

The forward transform given in 5 is in the mixed-domain. For example, if the input function is defined over time, 5 takes the frequency spectrum of that input function, and outputs a new time series. This should not be alarming, as the Fourier transform itself is a mixed-domain transform. It may seem questionable how $p\left(q\right)$ is handled in the exponent. However, since the nonstationary shift is formulated here as a nonstationary combination, and the integration is over β, nothing new is needed to handle the relation between p and q.

Although nonstationary combination kinematically performs nonstationary shifts correctly, amplitudes are not changed, and therefore energy is not conserved between input and output traces. Figure 1 demonstrates how the area under the input signal changes under a nonstationary combination shift. This nonphysical energy change within a trace is inherent to any processing step that applies or approximates a nonstationary shift. Rayleigh's theorem (also known as Plancherel's theorem) states that the integral of the square of an input function must be the same as the integral of its squared Fourier transform (13). Physically, this means that energy should be conserved between the input and Fourier domains, and clearly, the forward data processing transform violates this. Although nonphysical in most cases, nonstationary combination provides exactly the type of shift desired for the NMO correction and many other seismic data processing steps. The well-known issue of NMO stretch (1) is a clear example of this type of energy change, where the seismic amplitudes remain unchanged after each trace has undergone the time-variant shift of the NMO correction. Relative to the input signal, this energy change means distortions in both the input and Fourier domains. These distortions are accounted for and handled by the inverse transform developed below, which relies on physically valid nonstationary convolution instead.

To quantify how the energy has been changed by the forward transform, take the ratio of areas of a shifted and unshifted rectangular element as in Figure 1:

$$\alpha\left(q\right) = \frac{\Delta p \left(\frac{f(p_1)+f(p_2)}{2}\right)}{\Delta q \left(\frac{g(q_1)+g(q_2)}{2}\right)}. \tag{7}$$

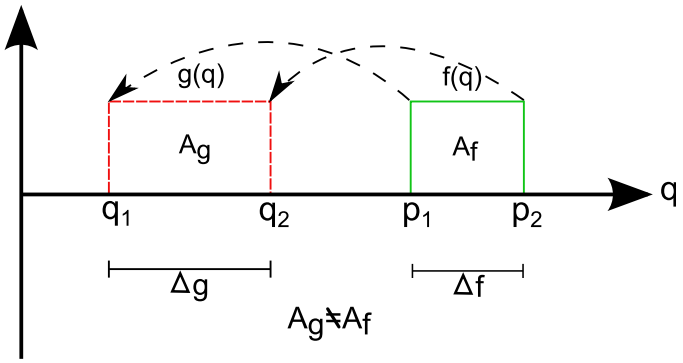


Figure 1: Element area change caused by a nonstationary combination shift. The values of f at p_1 and p_2 are mapped to g at q_1 and q_2, respectively. Since Δf does not equal Δg, the areas A_f and A_g are not equal. Therefore, by Rayleigh's Theorem, energy is not conserved. Scaling $g\,(q)$ by $\alpha(q)$ makes the areas the same. reversible/. NSArea

Equation 6 shows that the shifted values of the input function are not changed. This corresponds to the height of the rectangular elements shown in Figure 1 remaining unchanged by the forward transform ($f(p_i) = g(q_i)$), giving,

$$\alpha\,(q) = \frac{\Delta p}{\Delta q}. \tag{8}$$

In the continuous limit, 8 goes to,

$$\alpha\,(q) = \frac{dp}{dq}. \tag{9}$$

The output function, $g\,(q)$, is of course distorted by different amounts for various q since the applied shift is nonstationary. By scaling $g(q)$ by $\alpha(q)$ and integrating, we claim that the energy change can be accounted for via Rayleigh's theorem in the inverse transform.

Now that there is a mechanism for removing amplitude distortions, the rest of the inverse transform development comes from removing the kinematic shift. The output function of the forward transform, $g\,(q)$, becomes the input function of the inverse transform. Since the forward shift is applied simultaneously with the inverse Fourier transform, the reverse shift should be applied simultaneously with the forward Fourier transform. That is, the inverse data processing transform should naturally recover the original input spectrum, $F\,(\beta)$, rather than go directly back to the original input signal.

We first apply the Fourier transform to $g\,(q)$ in terms of its own coordinate, q, then reverse the nonstationary shift simultaneously,

$$G\,(\gamma) = \int_{-\infty}^{\infty} \alpha\,(q)\, g\,(q)\; e^{-i\,\gamma\,q}\, e^{-i\,\gamma\,\Delta(q)}\, dq\,, \tag{10}$$

where γ is the Fourier dual of q, and $G(\gamma)$ is the desired uncorrected spectrum. Two interesting simplifications come from 10. First, we can again exploit the convenient

form of $\Delta(q)$ to reduce the exponential giving,

$$G(\gamma) = \int_{-\infty}^{\infty} \alpha(q)\, g(q)\; e^{-i\,\gamma\, p(q)}\, dq\,. \tag{11}$$

Doing so forces integrating $p(q)$ in the exponential over q, which may at first seem like an unattractive approach, but we must do so to account for the nonphysical distortion caused by the forward transform. The second simplification helps, as the scaling function, $\alpha(q) = \frac{dp}{dq}$ conveniently acts as a change-of-integration-variable factor, giving,

$$G(\gamma) = \int_{-\infty}^{\infty} g(q)\, e^{-i\,\gamma\, p(q)} dp. \tag{12}$$

By Rayleigh's theorem, the energy change can be accounted for by scaling $g(q)$ by $\alpha(q)$ inside an integral over q. This observation, combined with 6, shows that 12 is equivalent to a regular Fourier transform of $f(p)$. Therefore, we can recover the unprocessed Fourier domain function–the goal of the mixed-domain general inverse data processing transform,

$$G(\gamma) = F(\beta)\,. \tag{13}$$

The original unprocessed input function can then be exactly recovered by a regular inverse Fourier transform:

$$h(q) = \frac{1}{2\,\pi} \int_{-\infty}^{\infty} G(\gamma)\, e^{i\,\gamma\, q} d\gamma, \tag{14}$$

where, $h(q)$ is the final unprocessed function of the inverse Fourier transform. Equation 14 is just a regular inverse Fourier transform of $G(\gamma)$ over its own coordinate, γ. Therefore, using 13, it is clear that 14 is equivalent to a regular IFT of $F(\beta)$ over β, and that the desired result of the inverse transform is confirmed:

$$h(q) = f(q)\,. \tag{15}$$

The kinematics of this general data processing transform are clear under nonstationary filtering theory. We claim that for any data processing step implemented by transform, the kinematics of the step can also be clearly understood, and that $\alpha(q)$ trivially predicts the amplitude corrections which are otherwise difficult to quantify using conventional filtering theory.

IMPLEMENTATION

Now that the general form of the transform has been developed, an appropriate method is needed to implement it. The method we review and discuss here is matrix vector multiplication.

In order to computationally implement an integral transform, recognize that the input function can be viewed as a vector, or a one-dimensional array, and that the

transform kernel can be viewed as a matrix, or a two-dimensional array. The output of a matrix vector multiplication is analogous to the output function of the integral, but it is another vector along a new coordinate axis. The important feature of matrix vector multiplication is that it can be used to simultaneously multiply and integrate two functions over an input index. This concept is certainly not new to engineering or geophysics, and many of its benefits have been explored in detail (8).

Constructing a matrix operator for the forward and inverse data processing transform follows the same procedure one would use for any transform operator. The general procedure is summarized here:

1. Identify the kernel of the transform and the input and output signals.
2. Construct the discrete input and output axes based on the sampling interval and domain of the input signal.
3. For an input signal of N-samples, initialize an empty matrix of size $N \times N$, and associate the matrix indices with the input and output axes.
4. For each (j, l) matrix (row,column) coordinate pair, calculate the kernel value and place it in the matrix.

The concise form of the forward data processing transform can be cast as a matrix vector multiplication following the above steps as:

$$g\left(q_j\right) = \sum_{l=1}^{N} \frac{e^{i\,\beta_l p(q_j)}}{2\,\pi}\, F\left(\beta_l\right). \tag{16}$$

A convenient feature of the data processing transform matrices is that they do not differ from the form of the Fourier transform matrices other than a distorted time axis. So by carefully matching the frequency axis, any standard FFT can be used together with the discrete data processing transform.

The inverse data processing transform is more similar in form to the forward Fourier transform, in that it takes the series from the input domain and outputs a Fourier spectrum. Again following the procedure to construct a transform matrix, we cast the inverse data processing transform (11) as a matrix-vector multiplication:

$$F\left(\beta_l\right) = \sum_{l=1}^{N} \frac{\left[\alpha\left(q_l\right) e^{-i\,\beta_j\, p(q_l)}\right]}{2\,\pi}\, f\left(p_l\right), \tag{17}$$

with no summation over l inside brackets. Again by carefully matching the frequency axis, a standard inverse fast Fourier transform (IFFT) can be applied to the output vector of equation 17 to recover the unprocessed input signal.

Although not necessary for implementation, it is insightful to separate the data processing transform matrices as a composition of a Fourier transform matrix with

a shifting matrix. For a set of traces that are of the same length, the only part of the data processing transform matrix that changes from trace to trace is the shifting exponential. It is only because of the unique form of the shift itself that the exponentials reduce to a single transform exponential. Before this reduction, the Fourier transform matrices can be constructed separately from the forward and reverse shifting matrices for the same data. The construction of the Fourier transform matrices is well-known, and elements of the forward shift matrix, $\mathbf{A}$, are given by:

$$a_{jl} \equiv e^{i\,\omega_l\,\Delta_j} = e^{i\,\omega_l(p(q_j)-q_j)}. \tag{18}$$

Elements of the reverse shifting matrix, $\mathbf{B}$, are then given by:

$$b_{jl} = \alpha\,(q_l)\,e^{-i\,\omega_j\,(p(q_l)-q_l)}, \tag{19}$$

with no summation over indices. This approach requires the Fourier matrices to be calculated only once for an input data set. For each trace in the given data set, the shifting matrix is computed and then combined with the Fourier matrix as a Hadamard (entry-wise) matrix product.

The inverse of the discrete data processing transform can be exactly formulated as in the continuous case. The discrete data processing transform has very similar form to the Fourier matrices, and its approximate inverse is easily calculated. From equations 18 and 19, we claim that the inverse matrix for each trace can be accurately approximated by the adjoint of the forward matrix, scaled by $\alpha(q_l)$.

NMO BY TRANSFORM

The kinematics of the shift required for the NMO correction simply shift data values backwards from input time, t_x, to output time, t_0, where t_x depends on t_0. For ideal continuous data, the input and output times are both arbitrary, but for discrete field data, both must coincide exactly with a sampling time or values must be interpolated. In the continuous case, this is exactly a nonstationary shift by combination as described by (14), where for the NMO correction, the nonstationary shift is expressed as,

$$\Delta_{NMO}\,(t_0) = t_x\,(t_0) - t_0, \tag{20}$$

(5) with,

$$t_x(t_0) = \sqrt{t_0^2 + \frac{x^2}{v^2}}, \tag{21}$$

where, x is the source-receiver offset, and v is the normal moveout velocity. It is the goal of the NMO correction to flatten each event to its associated zero-offset arrival time. The NMO correction can be clearly recognized as a nonstationary shift, where $t_x(t_0)$ is a special case of $p(q)$, and t_0 is a special case of q. Direct substitution into 5 yields the forward NMO transform:

$$g\,(t_0) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{i\omega t_x(t_0)}d\omega, \tag{22}$$

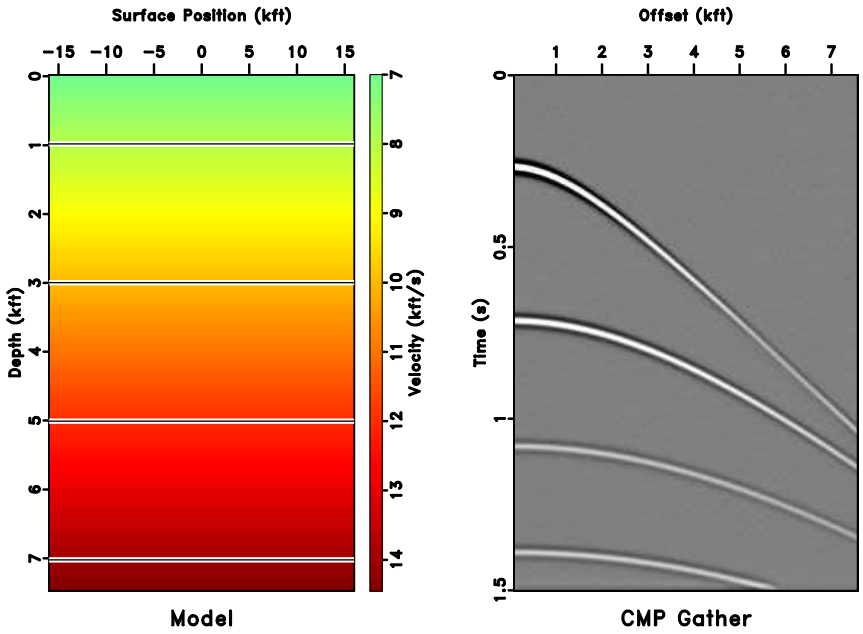


Figure 2: Synthetic CMP gather. Left: Velocity model. Four reflectors are placed in a background velocity that increases linearly with depth ($v(z) = 7.0 + z$ kft/s). Right: CMP gather generated using the velocity model. Arrival times predicted by Kirchhoff modeling are convolved with a 25 Hz Ricker wavelet. reversible/nscnmo data

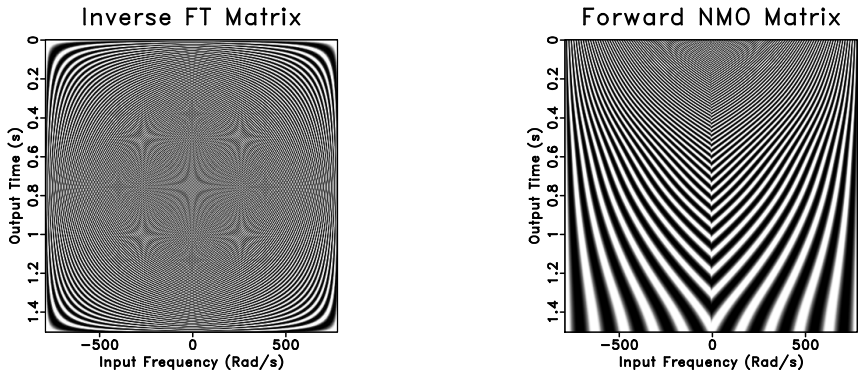


Figure 3: The forward NMO transform matrix can be composed as the Hadamard (entry-wise) product of the inverse Fourier transform matrix (left) and the NMO forward shifting matrix for an example trace at $x = 3,600$ ft (right). The real part of each matrix is shown here. reversible/nscnmo nmomat

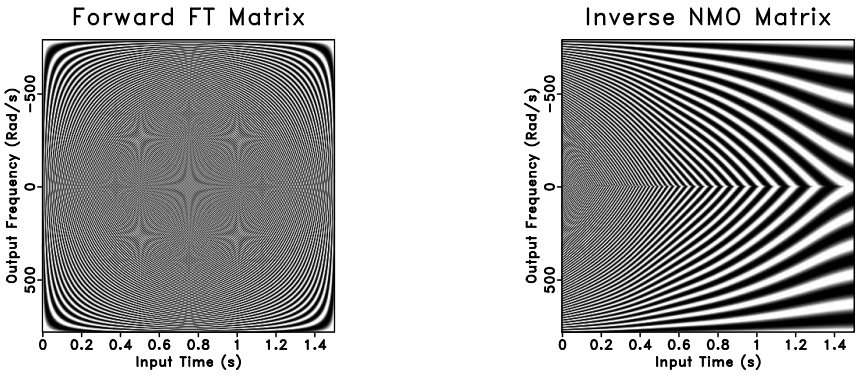


Figure 4: The inverse NMO transform matrix can be composed as the Hadamard product of the forward Fourier transform matrix (left) and the NMO reverse shifting matrix for an example trace at $x = 3,600$ ft (right). The real part of each matrix is shown here. reversible/nscnmo inmomat

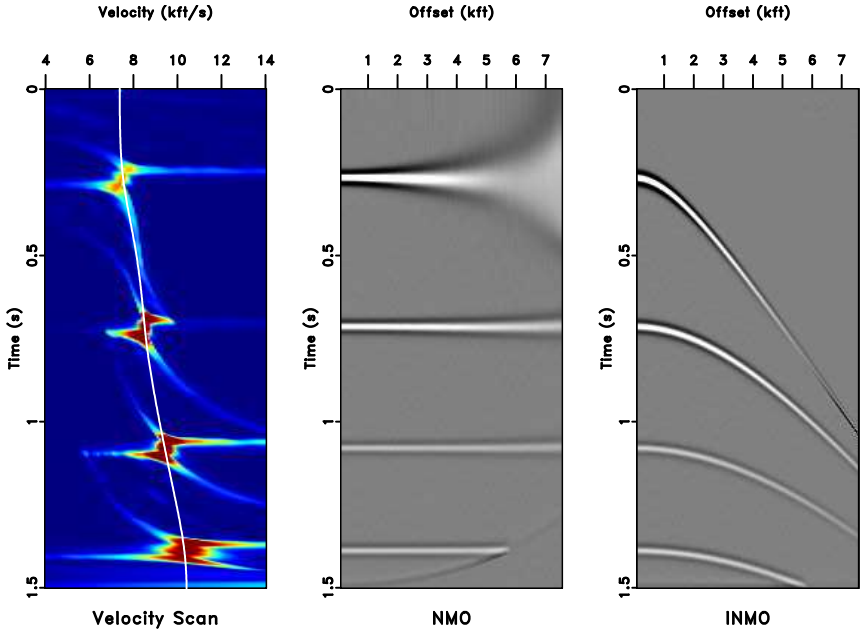


Figure 5: Reversible NMO applied to gather from Figure 9(a). Left: Semblance panel and picked velocity. Center: NMO-corrected gather. Right: Inverse NMO applied to gather. reversible/nscnmo nmo

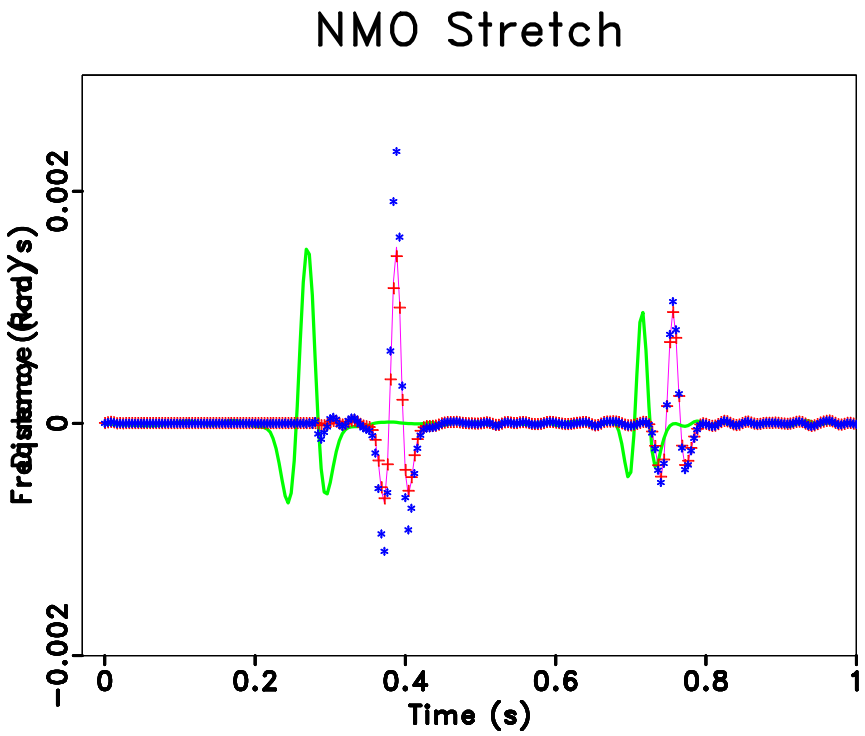


Figure 6: NMO applied to and removed from a single trace at $x = 2,100$ ft. Input trace: thin-magenta line. NMO-corrected trace: thick-green line. Inverse NMO without α: line of * symbols. Inverse NMO using α: line of + symbols. The use of α allows one to recover the input trace by accounting for the non-physical energy distortion caused by forward NMO (NMO stretch). reversible/nscnmo stretch

where ω is the Fourier dual of t_x. This transform takes the frequency spectrum of an input seismic trace, $F(\omega)$, and directly outputs the NMO-corrected trace, $g(t_0)$. Figure 9(a) shows a synthetic common midpoint (CMP) gather generated using Kirchhoff modeling. We estimate a reasonable NMO velocity from the synthetic gather in Figure 9(a) using a semblance scan, and then apply the NMO transform to yield the result in Figure 4.

In the case of the NMO correction, the time distortion caused by the shift is well-known, and referred to as *NMO stretch* (3; 9). Various authors have developed strategies for handling the effects of NMO stretch in practical seismic data processing (18; 20; 22). NMO stretch manifests the most at far offsets and early times, where the NMO correction is the most severe. The waveform of an event in this area is stretched to a longer period as the NMO correction is applied (24). The amplitude of that event is not affected, but its time distortion causes an alteration in its frequency content. Specifically, the lower frequency content is boosted, and since the amplitude of the waveform is maintained, a nonphysical addition of energy to the input signal has occurred. (24) quantifies the frequency distortion caused by NMO stretch in terms of Δ_{NMO} and t_0:

$$\frac{\Delta f}{f} = \frac{\Delta_{NMO}}{t_0}, \tag{23}$$

where, f is the dominant frequency of the input waveform. This expression is derived from geometric arguments and assumes $t_0 >> f^{-1}$ (24). Another approach to describing NMO stretch comes from analyzing the effects of the NMO correction in terms of instantaneous frequency (1). This approach provides excellent insight into the frequency distortion, and yields an exact expression for NMO stretch, S:

$$S = \frac{t_x}{t_0 - \frac{x^2 v'(t_0)}{v^3(t_0)}}, \tag{24}$$

where v' is the derivative of v with respect to time (1). This expression quantifies NMO stretch without approximations, and is not dependent only on the dominant frequency as in equation 23. It is valid for variable velocity models, and for a constant velocity model, the derivative of velocity goes to zero, and $S - 1$ agrees with the approximation in equation 23.

As equation 24 suggests, NMO stretch is itself nonstationary, and therefore cannot be removed exactly using stationary filtering theory. However, since the data processing transform takes advantage of nonstationary filtering, NMO stretch can be removed exactly. Substituting the NMO variables, Δ_{NMO} and $\alpha_{NMO}(t_0)$ into 11, gives the inverse NMO transform:

$$F(\omega) = \int_{-\infty}^{\infty} \alpha_{NMO}(t_0)\, g(t_0)\, e^{-i\omega t_x(t_0)} dt_0. \tag{25}$$

Here, insight into the role of α can finally be gained. $\alpha_{NMO}(t_0)$ is given as,

$$\alpha_{NMO}(t_0) \equiv \frac{\partial t_x}{\partial t_0} = \frac{t_0 - \frac{x^2 v'(t_0)}{v^3(t_0)}}{t_x}. \tag{26}$$

The inverse NMO transform does account for NMO stretch, and in fact, $\alpha_{NMO} = S^{-1}$ exactly. Realizing that the two are equivalent leads to a straightforward (once nonstationary filtering is allowed) perspective on predicting and quantifying NMO stretch. Further, Barnes (1992) proves, in terms of instantaneous power and energy spectra, that S, and therefore $\alpha_{NMO}(t_0)$ indeed describe the nonphysical energy change caused by the NMO correction. Figure 6 shows the application and removal of the NMO correction to a single trace. Two cases are shown for inverse NMO: one where α is not included and one where it is. In the case where α is used, the amplitude is recovered correctly, whereas when α is not used, inverse NMO preserves the energy of the NMO-corrected trace, leading to too much energy in the recovered signal. It can be shown that α plays a similar role in accounting for nonphysical energy changes introduced by other imaging steps such as so-called "Stolt stretch" associated with Stolt migration (6).

The NMO transform acts as a map between the uncorrected spectrum and the NMO-corrected trace. As such, the first step in implementing the forward NMO transform on a trace is to apply a standard forward Fourier transform to the input trace. This gives the uncorrected spectrum, which becomes the input column vector for a matrix vector multiplication. Following 16, the discrete NMO transform is given as,

$$g\left(\left[t_0\right]_j\right) = \sum_{l=1}^{N} \frac{e^{i\,\omega_l p(t_{0,j})}}{2\,\pi} F(\omega_l)\,. \tag{27}$$

One can also decompose the forward NMO transform operator as the Hadamard product of the inverse Fourier transform matrix and the NMO shifting matrix following 18 (see Figure 3). The frequency axis of the shifting operator must exactly match the frequency axis used for the forward Fourier transform, since it will simultaneously remove that Fourier transform as it NMO-corrects the data. The inverse NMO transform can also be implemented by matrix-vector multiplication by substituting t_x and α_{NMO} into 17. As seen in Figure 4, the full inverse NMO transform operator can be decomposed as the Hadamard product of the forward Fourier transform matrix and the NMO reverse shifting matrix. We conclude this example by removing the NMO correction by transform, shown in Figure 4. Most of the gather is recovered nearly perfectly–only parts which were mapped to $t_0 = 0$ are not recovered well.

RELATION TO INTERPOLATION

Casting a theoretically continuous input signal as a vector is a natural way to view recorded seismic data. The time-sampling interval of the recorded data determines the Nyquist frequency of the recorded signal (11), and it can be assumed that the data is frequency band-limited between zero and the Nyquist frequency (11). This means that the recorded signal is exactly determined by this small and discrete range of frequencies. No information is gained by using filters that operate on frequencies higher than the Nyquist frequency. This observation, combined with the Fourier sampling theorem, suggests that a continuous, but band-limited form of the recorded signal exists, and it is exactly described by the Fourier basis. This continuous form is not necessarily the same as the true continuous function that a seismometer attempts to record. Conventional processing algorithms estimate the values of this continuous function by assuming that the function behaves like some interpolant between samples. A repeatable and accurate approach to discrete data processing is to return the exact value of the band-limited continuous equivalent of the recorded signal at any desired input coordinate. The data processing transform does exactly this.

The data processing transform is a special case of the nonstationary Fourier shift theorem. It follows then, for the continuous case, that data processing steps implemented by transform are effectively performing convolution with a shifting scaled Dirac delta function. The sifting property of the delta function justifies using convolution with a delta function to extract the exact values of a continuous input function at exact target times. This is the ideal way to apply data processing corrections in either transform or mapping approaches, but the discrete nature of seismic data prevents this. By showing that interpolation operators are approximations to the Dirac delta function, we suggest that any interpolation scheme can be viewed as an approximation to a continuous and exact nonstationary shift.

The Dirac delta function can be formulated as the generalized limit of other more common functions (19). Two clear examples of this are boxcar (rectangle) and Gaussian functions, which both approach an impulse as their widths go to zero and their heights go to infinity (19). Take for example the boxcar function:

$$r_\epsilon(t) = \begin{cases} \frac{1}{\epsilon} & |t| \leq \epsilon \\ 0 & |t| > \epsilon \end{cases} . \tag{28}$$

As ϵ approaches zero, the width of the boxcar goes to zero, and the height goes to infinity (see Figure 7). As this occurs, the boxcar function approaches the Dirac delta function:

$$\lim_{\epsilon \to 0} r_\epsilon(t) = \delta(t) . \tag{29}$$

The scaled boxcar function is used as an interpolant for nearest-neighbour interpolation, where the width of the boxcar is equal to the sampling interval. This is a poor choice in most cases, as nearest-neighbour interpolation in the time domain, for example, has the unwanted effect of multiplying the frequency spectrum by a

sinc function. Conversely, nearest neighbour interpolation in the frequency domain introduces ringing into the time series.

Linear interpolation is equivalent to convolution with a scaled triangle function with a width equal to twice the sampling interval (12). The triangle function also approaches the delta function as its width goes to zero and height goes to infinity:

$$\Lambda_\epsilon(t) = \begin{cases} \frac{1}{\epsilon} - \frac{|t|}{\epsilon^2} & |t| \leq \epsilon \\ 0 & |t| > \epsilon \end{cases}, \tag{30}$$

$$\lim_{\epsilon \to 0} \Lambda_\epsilon(t) = \delta(t). \tag{31}$$

This behaviour can be seen in Figure 8. A more common interpolant used in seismic data processing is the sinc function. Data processing corrections implemented by sinc-interpolation algorithms effectively perform convolution of the input signal with a shifting scaled sinc-function. In the general continuous case, this has no clear justification, but for band-limited or discrete data, the purpose of the sinc-function becomes apparent. The Fourier transform of the sinc-function is a boxcar function over its own frequency content (12). Since the ideal interpolant should not affect the frequency content of the data while it estimates an intermediate value, a sinc function with frequency content at least up to the Nyquist frequency of the input signal will be the ideal interpolant (12). Computational efficiency is often increased by truncating the length of the sinc operator (12), and tapering of the truncated sinc function is often performed (21) to reduce the Gibbs phenomena ringing.

The sinc function is a very good choice for an interpolant as it attempts to respect the finite bandwidth of the recorded data. However, convolution with a sinc function is still an approximation to ideal continuous convolution with the delta function. For the following expression (10), as ϵ approaches infinity, the normalized sinc function approaches an impulse (See Figure 9):

$$\mathrm{sinc}_\epsilon(t) = \frac{\epsilon \sin \pi \epsilon t}{\pi \epsilon t}, \tag{32}$$

$$\lim_{\epsilon \to \infty} \mathrm{sinc}_\epsilon(t) = \delta(t). \tag{33}$$

There are more similarities between the narrowing sinc function and the Dirac delta function than just the shape of the pulse. When ϵ is less than infinity in 32, the side lobes of the sinc function mimic Gibbs phenomenon in Fourier transform theory (19). Since the Fourier transform of a sinc function is a boxcar function, one can also visualize how it approaches the delta function in the Fourier domain. As the sinc function narrows, its frequency content increases, and its boxcar Fourier transform broadens. When the sinc function finally becomes an impulse, its Fourier spectrum is an infinitely wide boxcar, which is simply a constant, just as the Fourier transform of the delta function. Applying a band-pass filter to a delta function yields a sinc function, and conversely, the sinc-function is a band-limited estimation of the Dirac delta function. Therefore, the sinc-function interpolant, *before* truncation or tapering,

respects the Fourier basis assumed by the Fourier transform, and is equivalent to processing by discrete nonstationary filtering.

The inverse to any processing step is exactly formulated here for the continuous case using the nonstationary version of the Fourier shift theorem. This is not surprising, as even the conventional procedure of mapping data directly between input and output times is exactly reversible assuming continuous data. Casting the integral kernel as a discrete matrix defined on the sampling intervals allows a precise method of approximating the exact continuous form assuming a Fourier basis.

CONCLUSIONS

Using the theory of nonstationary filtering, we have developed a general transform that applies and removes many common seismic data processing and imaging operations. Any data processing or imaging step that can be viewed as a nonstationary shift in any domain is a special case of this transform. We have provided a derivation of the general data processing transform and we have used NMO as a familiar example.

Like the Fourier transform, the data processing transform is a mixed-domain transform. In the general case, the input data should be viewed as an image, and then a distorted version of that image is the desired output. The data processing transform moves data directly between the Fourier spectrum of the input image and the corrected image. Regardless of the physical meaning of input and the Fourier coordinates, viewing the input and output data each as an image makes the general application of the transform to any processing step clear. The physical meaning of the transform coordinates and distortions can be defined only when discussing specific processing steps.

Common nonstationary shifts in seismic data processing are often nonphysical, as they do not conserve the energy of a given input signal. The general data processing transform predicts and quantifies the change in energy caused by such processing steps. Processing artifacts such as NMO stretch associated with the NMO correction and Stolt stretch associated with Stolt and frequency-wavenumber migration are examples of this type of nonphysical energy change. Under the framework of the general data processing transform, these effects are both special cases of, and are easily quantified by, the nonstationary scaling function $\alpha(q)$. We have shown that the data processing transform is reversible by accounting for these energy changes while simultaneously reversing the kinematic shift applied by the forward transform.

Rather than explicitly interpolating discrete data, the data processing transform assumes that the input signal behaves according to the Fourier basis between samples. The Fourier transform itself predicts a continuous form of given discrete data, and the transform here performs a nonstationary shift on this continuous function. Just as in the stationary case, nonstationary shifts can be viewed as convolution with a Dirac delta function. We have demonstrated how common interpolants such as boxcar,

triangle, and sinc functions approximate the delta function, but add assumptions about how the input signal behaves in between samples which may contradict the Fourier basis. By analysis in both input and Fourier domains, we have shown that the use of an untruncated sinc function as an interpolant is equivalent to the discrete form of this transform.

Having a reversible transform for seismic data processing allows the various stages of a processing flow to be viewed as valid processing domains. For example, one may use this transform to apply a correction to the input data which organizes it into a format better suited for multiple attenuation, then perform multiple attenuation, and last transform back to the input domain with no residual effects of the correction itself. Further, the discrete form of this general transform provides a straightforward framework for constructing and understanding the matrix operators associated with seismic imaging steps. Future studies may be able to exploit symmetries of these operators to improve computational efficiency while mitigating interpolation artifacts. We anticipate a variety of future applications of the reversible transform using these concepts.

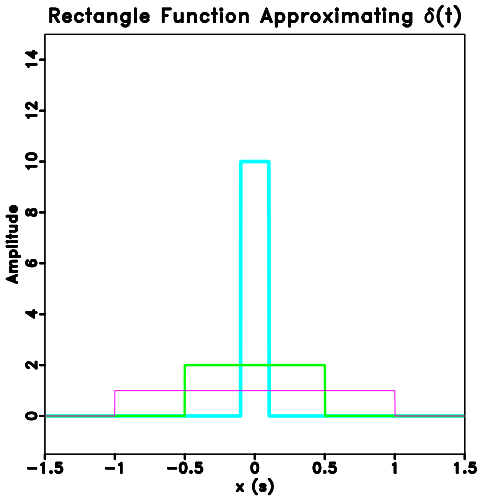


Figure 7: Nearest-neighbour interpolant approaching $\delta(t)$. As ϵ approaches zero, the rectangle function defined by equation 28 approaches the Dirac delta function. Thin magenta line: $\epsilon = 1.0$. Green line: $\epsilon = 0.5$. Thick cyan line: $\epsilon = 0.1$. reversible/delta nearest

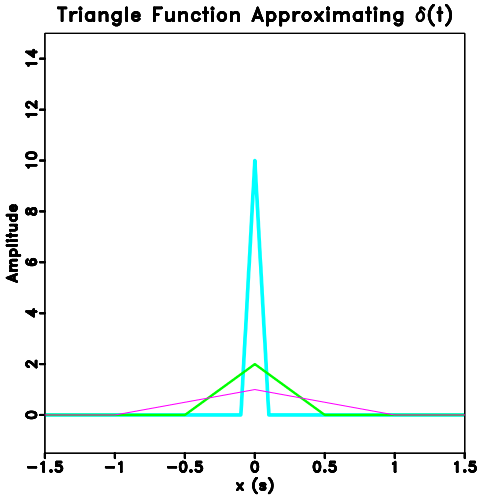


Figure 8: Linear interpolant approaching $\delta(t)$. As ϵ approaches zero, the triangle function defined by equation 30 approaches the Dirac delta function. Thin magenta line: $\epsilon = 1.0$. Green line: $\epsilon = 0.5$. Thick cyan line: $\epsilon = 0.1$. reversible/delta linear

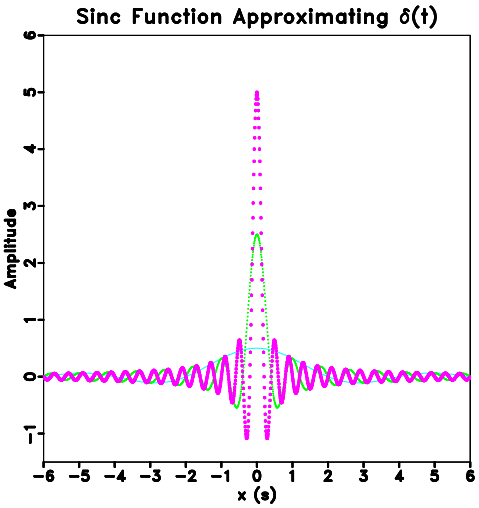


Figure 9: Sinc interpolant approaching $\delta(t)$. As ϵ approaches infinity, the sinc-function defined by equation 32 approaches the Dirac delta function. Magenta * line: $\epsilon = 5.0$, Dotted green line: $\epsilon = 2.5$. Solid cyan line: $\epsilon = 0.5$. reversible/delta sinc

ACKNOWLEDGMENTS

The authors thank the Jackson School of Geosciences and the EDGER Forum at The University of Texas at Austin, and the sponsors of CREWES at The Department of Geoscience, The University of Calgary for their support of this work.

REFERENCES

[1] Barnes, A. E., 1992, Another look at nmo stretch: Geophysics, **57**, 749–751.
[2] Biondi, B., and J. F. Claerbout, 1985, Pseudounitary NMO: SEP, **44**, 75–84.
[3] Buchholtz, H., 1972, A note on signal distortion due to dynamic (nmo) corrections: Geophysical Prospecting, **20**, 395–402.
[4] Burnett, W. A., 2007, A general transform for reversible seismic data processing by nonstationary filtering: Master's thesis, University of Texas at Austin.
[5] Burnett, W. A., and R. J. Ferguson, 2008a, Reversible NMO: CREWES Research Report, **20**, 42.
[6] ———, 2008b, Reversible Stolt migration: CREWES Research Report, **20**, 43.
[7] ———, 2008c, A reversible transform for seismic data processing: CREWES Research Report, **20**, 41.
[8] Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Science.
[9] Dunkin, J. W., and F. K. Levin, 1973, Effect of normal moveout on a seismic pulse: Geophysics, **38**, 635–642.
[10] Elliot, D. F., and K. R. Rao, 1982, Fast transforms: Algorithms, analysis, applications: Academic Press, Inc.
[11] Gubbins, D., 2004, Time series analysis and inverse theory for geophysicists: Cambridge University Press.
[12] Harlan, W., 1982, Avoiding interpolation artifacts in Stolt migration: SEP, **30**, 103–110.

[13] Karl, J. H., 1989, An introduction to digital signal processing: Academic Press.
[14] Margrave, G. F., 1998, Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering: Geophysics, **63**, 244–259.
[15] ———, 2001, Direct Fourier migration for vertical velocity variations: Geophysics, **66**, 1504–1514.
[16] Margrave, G. F., and R. J. Ferguson, 1999, Wavefield extrapolation by nonstationary phase shift: Geophysics, **64**, 1067–1078.
[17] McMechan, G. A., and R. Sun, 1991, Depth filtering of first breaks and ground roll: Geophysics, **56**, 390–396.
[18] Miller, R. D., 1992, Normal moveout stretch mute on shallow-reflection data: Geophysics, **57**, 1502–1507.
[19] Papoulis, A., 1962, The fourier integral and its applications: McGraw Hill Classic Textbook Reissue Series.
[20] Perroud, H., and M. Tygel, 2004, Nonstretch NmO: Geophysics, **69**, 599–607.
[21] Rosenbaum, J. H., and G. F. Boudeaux, 1981, Rapid convergence of some seismic processing algorithms: Geophysics, **46**, 1667–1672.
[22] Shatilo, A., and F. Aminzadeh, 2000, Constant normal-moveout(CnMo) correction: A technique and test results: Geophys. Prosp., **48**, 473–488.
[23] Sneddon, I. N., 1995, Fourier transforms: Dover Publications.
[24] Yilmaz, O., 2001, Seismic data analysis: Society of Exploration Geophysics.
[25] Yu, Z., N. Kabir, and K. Matson, 2005, Diffracted multiple attenuation using wavelet filter in the focused domain: Expanded Abstracts 75^{th} SEG International Convention, Houston, Texas, 2072–2077.

Seismic wave extrapolation using lowrank symbol approximation

Sergey Fomel[*], *Lexing Ying*[†], *and Xiaolei Song*[*]

ABSTRACT

We consider the problem of constructing a wave extrapolation operator in a variable and possibly anisotropic medium. Our construction involves Fourier transforms in space combined with the help of a lowrank approximation of the space-wavenumber wave-propagator matrix. A lowrank approximation implies selecting a small set of representative spatial locations and a small set of representative wavenumbers. We present a mathematical derivation of this method, a description of the lowrank approximation algorithm, and numerical examples which confirm the validity of the proposed approach. Wave extrapolation using lowrank approximation can be applied to seismic imaging by reverse-time migration in 3D heterogeneous isotropic or anisotropic media.

INTRODUCTION

Wave extrapolation in time plays an important role in seismic imaging (reverse-time migration), modeling, and full waveform inversion. Conventionally, extrapolation in time is performed by finite-difference methods (10). Spectral methods (26; 22) have started to gain attention recently and to become feasible in large-scale 3-D applications thanks to the increase in computing power. The attraction of spectral methods is in their superb accuracy and, in particular, in their ability to suppress dispersion artifacts (7; 11).

Theoretically, the problem of wave extrapolation in time can be reduced to analyzing numerical approximations to the mixed-domain space-wavenumber operator (27). In this paper, we propose a systematic approach to designing wave extrapolation operators by approximating the space-wavenumber matrix symbol with a lowrank decomposition. A lowrank approximation implies selecting a small set of representative spatial locations and a small set of representative wavenumbers. The optimized separable approximation or OSA (23) was previously employed for wave extrapolation (28; 8) and can be considered as another form of lowrank decomposition. However, the decomposition algorithm in OSA is significantly more expensive, especially for anisotropic wave propagation, because it involves eigenfunctions rather than rows and columns of the original extrapolation matrix. Our algorithm can also be regarded as an extension of the wavefield interpolation algorithm of (11), with optimally selected reference velocities and weights. Another related method is the

Fourier finite-difference (FFD) method proposed recently by (24). FFD may have an advantage in efficiency, because it uses only one pair of multidimensional forward and inverse FFTs (fast Fourier transforms) per time step. However, it does not offer flexible controls on the approximation accuracy.

Our approach to wave extrapolation is general and can apply to different types of waves, including both acoustic and elastic seismic waves, as well as velocity continuation (13), offset continuation (12), prestack exploding reflector extrapolation (3), etc.

The paper is organized as follows. We first present the theory behind the proposed algorithm, then describe the algorithm and test its accuracy on a number of synthetic benchmark examples of increasing complexity.

WAVE EXTRAPOLATION

Let $P(\mathbf{x},t)$ be the seismic wavefield at location $\mathbf{x}$ and time t. The wavefield at the next time step $t+\Delta t$ can be approximated by the following mixed-domain operator (27)

$$P(\mathbf{x},t+\Delta t) = \int \widehat{P}(\mathbf{k},t)\, e^{i\,\phi(\mathbf{x},\mathbf{k},\Delta t)}\, d\mathbf{k}\ , \tag{1}$$

where $\widehat{P}(\mathbf{k},t)$ is the spatial Fourier transform of $P(\mathbf{x},t)$

$$\widehat{P}(\mathbf{k},t) = \frac{1}{(2\,\pi)^3} \int P(\mathbf{x},t) e^{-i\,\mathbf{k}\cdot\mathbf{x}}\, d\mathbf{x}\ , \tag{2}$$

where $\mathbf{k}$ is the spatial wavenumber. To define the phase function $\phi(\mathbf{x},\mathbf{k},t)$, which appears in equation (1), one can substitute approximation (1) into the wave equation and extract the geometrical (high-frequency) asymptotic of it. In case of seismic wave propagation, this leads to the eikonal-like equation

$$\frac{\partial \phi}{\partial t} = \pm V(\mathbf{x},\mathbf{k})\ |\nabla \phi|\ , \tag{3}$$

where $V(\mathbf{x},\mathbf{k})$ is the phase velocity, and the choice of the sign corresponds, in the case of a point source, to expanding or contracting waves. In the isotropic case, V does not depend on $\mathbf{k}$. The initial condition for equation (1) is

$$\phi(\mathbf{x},\mathbf{k},0) = \mathbf{k}\cdot\mathbf{x}\ , \tag{4}$$

which turns equation (1) into the simple inverse Fourier transform operation.

Assuming small steps Δt in equation (1), one can build successive approximations for the phase function ϕ by expanding it into a Taylor series. In particular, let us represent the phase function as

$$\phi(\mathbf{x},\mathbf{k},t) \approx \mathbf{k}\cdot\mathbf{x} + \phi_1(\mathbf{x},\mathbf{k})\, t + \phi_2(\mathbf{x},\mathbf{k})\, \frac{t^2}{2} + \cdots \tag{5}$$

Correspondingly,

$$|\nabla\phi| \approx |\mathbf{k}| + \frac{\nabla\phi_1 \cdot \mathbf{k}}{|\mathbf{k}|}\, t + O(t^2)\ . \tag{6}$$

Substituting expansions (6) and (6) into equation (1) and separating terms with different powers of t, we find that

$$\phi_1(\mathbf{x},\mathbf{k}) = V(\mathbf{x},\mathbf{k})\,|\mathbf{k}|\ , \tag{7}$$
$$\phi_2(\mathbf{x},\mathbf{k}) = V(\mathbf{x},\mathbf{k})\,\nabla V\cdot\mathbf{k}\ . \tag{8}$$

When either the velocity gradient ∇V or the time step Δt are small, the Taylor expansion (6) can be reduced to only two terms, which in turn reduces equation (1) to the familiar expression (11)

$$P(\mathbf{x},t+\Delta t) \approx \int \widehat{P}(\mathbf{k},t)\, e^{i\,[\mathbf{k}\cdot\mathbf{x}+V(\mathbf{x},\mathbf{k})\,|\mathbf{k}|\,\Delta t]}\, d\mathbf{k}\ , \tag{9}$$

or

$$P(\mathbf{x},t+\Delta t)+P(\mathbf{x},t-\Delta t) \approx 2\int \widehat{P}(\mathbf{k},t)\, e^{i\,\mathbf{k}\cdot\mathbf{x}} \cos\left[V(\mathbf{x},\mathbf{k})\,|\mathbf{k}|\,\Delta t\right]\, d\mathbf{k}\ . \tag{10}$$

In rough velocity models, where the gradient ∇V does not exist, one can attempt to solve the eikonal equation 1 numerically or to apply approximations other than the Taylor expansion (6). In the examples of this paper, we used only the ϕ_1 term.

Note that the approximations that we use, starting from equation (1), are focused primarily on the phase of wave propagation. As such, they are appropriate for seismic migration but not necessarily for accurate seismic modeling, which may require taking account of amplitude effects caused by variable density and other elastic phenomena.

The computational cost for a straightforward application of equation (1) is $O(N_x^2)$, where N_x is the total size of the three-dimensional $\mathbf{x}$ grid. Even for modest-size problems, this cost is prohibitively expensive. In the next section, we describe an algorithm that reduces the cost to $O(M\,N_x\,\log N_x)$, where M is a small number.

LOWRANK APPROXIMATION

The key idea of the lowrank decomposition is decomposing the wave extrapolation matrix

$$W(\mathbf{x},\mathbf{k}) = e^{i\,[\phi(\mathbf{x},\mathbf{k},\Delta t)-\mathbf{k}\cdot\mathbf{x}]} \tag{11}$$

for a fixed Δt into a separated representation

$$W(\mathbf{x},\mathbf{k}) \approx \sum_{m=1}^{M}\sum_{n=1}^{N} W(\mathbf{x},\mathbf{k}_m)\,a_{mn}\,W(\mathbf{x}_n,\mathbf{k}). \tag{12}$$

Representation (15) speeds up the computation of $P(\mathbf{x}, t+\Delta t)$ since

$$\begin{aligned} P(\mathbf{x}, t+\Delta t) &= \int e^{i\mathbf{x}\mathbf{k}} W(\mathbf{x},\mathbf{k}) \widehat{P}(\mathbf{k},t) d\mathbf{k} \\ &\approx \sum_{m=1}^{M} W(\mathbf{x},\mathbf{k}_m) \left(\sum_{n=1}^{N} a_{mn} \left(\int e^{i\mathbf{x}\mathbf{k}} W(\mathbf{x}_n,\mathbf{k}) \widehat{P}(\mathbf{k},t) d\mathbf{k} \right) \right) . \quad (13) \end{aligned}$$

The evaluation of the last formula is effectively equivalent to applying N inverse Fast Fourier Transforms. Physically, a separable lowrank approximation amounts to selecting a set of N representative spatial locations and M representative wavenumbers.

In order to discuss the construction of approximation (15), let us view it as a matrix decomposition problem

$$\mathbf{W} \approx \mathbf{W}_1 \, \mathbf{A} \, \mathbf{W}_2 \quad (14)$$

where $\mathbf{W}$ is the $N_x \times N_x$ matrix with entries $W(\mathbf{x},\mathbf{k})$, $\mathbf{W}_1$ is the submatrix of $\mathbf{W}$ that consists of the columns associated with $\{\mathbf{k}_m\}$, $\mathbf{W}_2$ is the submatrix that consists of the rows associated with $\{\mathbf{x}_n\}$, and $\mathbf{A} = \{a_{mn}\}$. In practice, we find that the matrix $\mathbf{W}$ has a low rank separated representation provided that Δt is sufficiently small, which, in the case of smooth models, can be partially explained by the separation of terms in the Taylor series 6. Let ε be a prescribed accuracy of this separated representation, and r_ε be the numerical rank of $\mathbf{W}$. The construction of the separated representation in equation (9) follows the method of Engquist and Ying (2007, 2009) and is detailed in the appendix. The main observation is that the columns of $\mathbf{W}_1$ and the rows of $\mathbf{W}_2$ should span the column space and row space of $\mathbf{W}$, respectively, as well as possible. The algorithm for computing (9) takes the following steps:

1. Pick a *uniformly random* set S of $\beta \cdot r_\varepsilon$ columns of $\mathbf{W}$ where β is chosen to be 3 or 4 in practice. Perform the pivoted QR factorization to $(\mathbf{W}(:,S))^*$ (16). The first r_ε pivoted columns correspond to r_ε rows of the matrix $\mathbf{W}(:,S)$. Define $\mathbf{W}_1$ to be the submatrix of $\mathbf{W}$ that consists of these rows and set $\mathbf{x}_1, \ldots, \mathbf{x}_N$ with $n = r_\varepsilon$ to be the corresponding x values of these rows.

2. Pick a *uniformly random* set T of $\beta \cdot r_\varepsilon$ rows of $\mathbf{W}$ and perform the pivoted QR factorization to $\mathbf{W}(T,:)$. Define $\mathbf{W}_2$ to be the submatrix of $\mathbf{W}$ that consists of these columns and set $\mathbf{k}_1, \ldots, \mathbf{k}_M$ with $m = r_\varepsilon$ to be the corresponding k values of these columns.

3. Set the middle matrix $\mathbf{A} = \mathbf{W}^\dagger(\mathbf{x}_n, \mathbf{k}_m)_{1 \le n \le N, 1 \le m \le M}$ where $\dagger$ stands for the pseudoinverse.

4. Combining the result of the previous three steps gives the required separated representation $\mathbf{W} \approx \mathbf{W}_1 \, \mathbf{A} \, \mathbf{W}_2$.

The algorithm does not require, at any step, access to the full matrix $\mathbf{W}$, only to its selected rows and columns. Once the decomposition is complete, it can be used at

every time step during the wave extrapolation process. In multiple-core implementations, the matrix operations in equation (15) are easy to parallelize. The algorithm details are outlined in the appendix.

The cost of the algorithm is $O(M\,N_x\,\log N_x)$ operations per time step, where $N_x\,\log N_x$ refers to the cost of the Fourier transform. In comparison, the cost of finite-difference wave extrapolation is $O(L\,N_x)$, where L is the size of the finite-difference stencil. (25) present an application of the proposed lowrank approximation algorithm for devising accurate finite-different schemes. There is a natural trade-off in the selection of M: larger values lead to a more accurate wave representation but require a longer computational time. In the examples of the next section, we select these parameters based on an estimate of the approximation accuracy and generally aiming for the relative accuracy of 10^{-4}. The resulting M is typically smaller than the number of Fourier transforms required for pseudo-spectral algorithms such as pseudo-spectral implementations of the rapid expansion method (21).

EXAMPLES

We start with a simple 1-D example. The 1-D velocity model contains a linear increase in velocity, from 1 km/s to 2.275 km/s. The extrapolation matrix, $2\,(\cos\,[V(x)\,|k|\,\Delta t]-1)$, or pseudo-Laplacian in the terminology of (11), for the time step $\Delta t = 0.001\,\mathrm{s}$ is plotted in Figure 1(a). Its lowrank approximation is shown in Figure 1(b) and corresponds to $N = M = 2$. The x locations selected by the algorithm correspond to velocities of 1.59 and 2.275 km/s. The wavenumbers selected by the algorithm correspond to the Nyquist frequency and 0.7 of the Nyquist frequency. The approximation error is shown in Figure 1(c). The relative error does not exceed 0.34%. Such a small approximation error results in accurate wave extrapolation, which is illustrated in Figure 2. The extrapolated wavefield shows a negligible error in wave amplitudes, as demonstrated in Figure 2(c).

Our next example (Figures 3 and 4) corresponds to wave extrapolation in a 2-D smoothly variable isotropic velocity field. As shown by (24), the classic finite-difference method (second-order in time, fourth-order in space) tends to exhibit dispersion artifacts with the chosen model size and extrapolation step, while spectral methods exhibit high accuracy. As yet another spectral method, the lowrank approximation is highly accurate. The wavefield snapshot, shown in Figures 3(b) and 10(a), is free from dispersion artifacts and demonstrates high accuracy. The approximation rank decomposition in this case is $N = M = 2$, with the expected error of less than 10^{-4}. In our implementation, the CPU time for finding the lowrank approximation was 2.45 s, the single-processor CPU time for extrapolation for 2500 time steps was 101.88 s or 2.2 times slower than the corresponding time for the finite-difference extrapolation (46.11 s).

To show that the same effect takes place in case of rough velocity model, we use first a simple two-layer velocity model, similar to the one used by (15). The difference

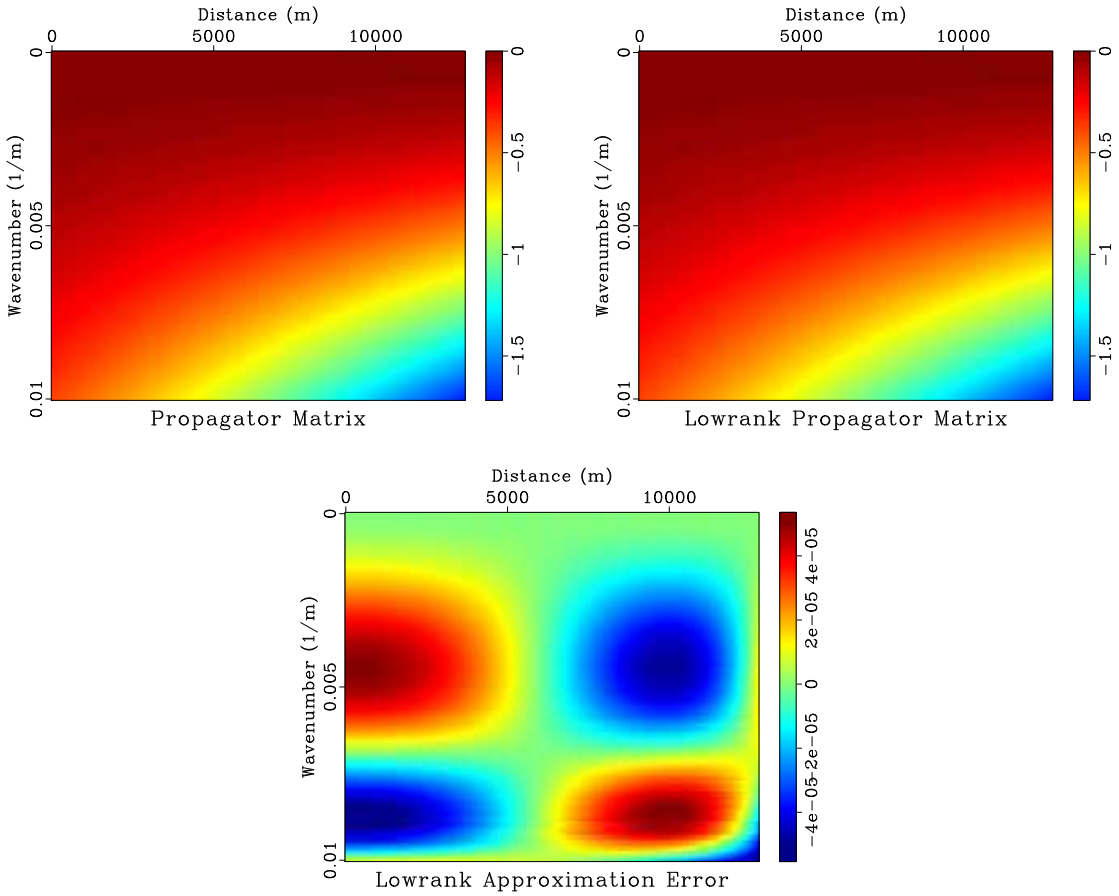


Figure 1: Wave extrapolation matrix for 1-D wave propagation with linearly increasing velocity (a), its lowrank approximation (b), and Approximation error (c). lowrank/fio prop,prod,proderr

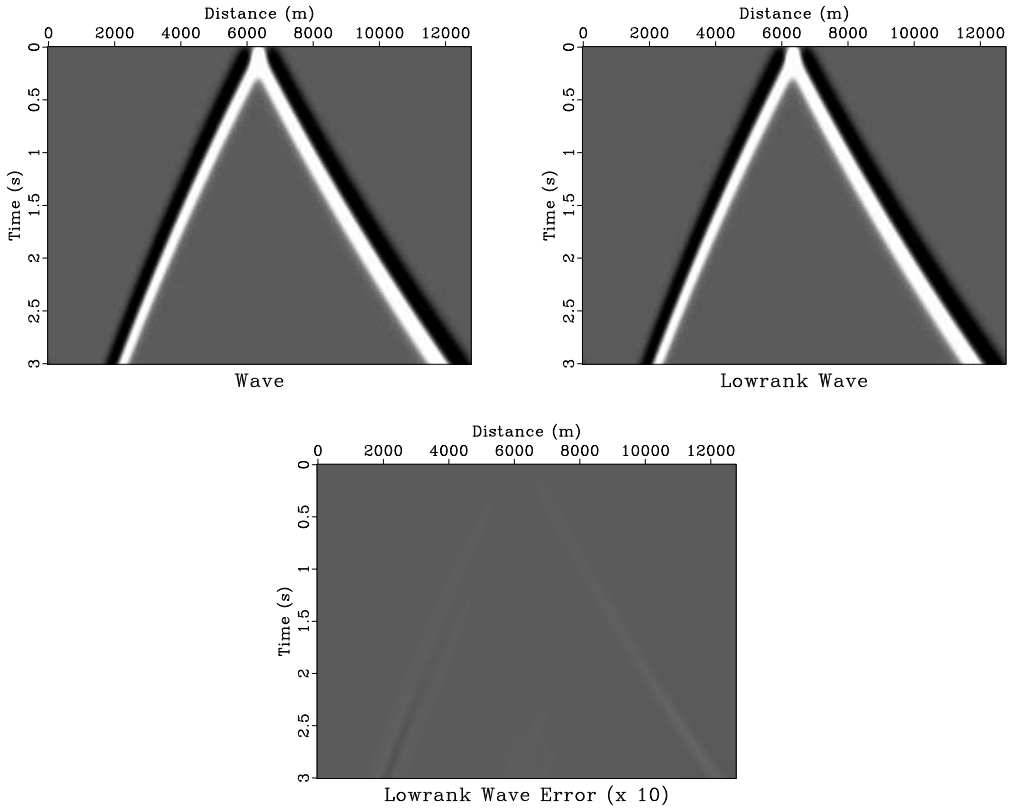


Figure 2: (a) 1-D wave extrapolation using the exact extrapolation symbol. (b) 1-D wave extrapolation using lowrank approximation. (c) Difference between (a) and (b), with the scale amplified 10 times compared to (a) and (b). lowrank/fio wave2,awave2,waverr

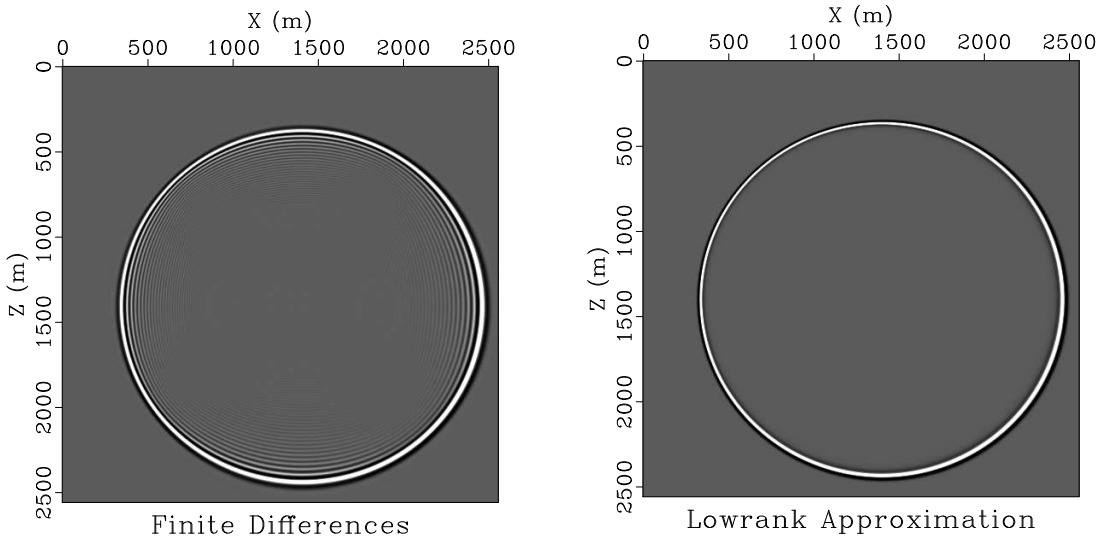


Figure 3: Wavefield snapshot in a smooth velocity model computed using (a) fourth-order finite-difference method and (b) lowrank approximation. The velocity model is $v(x,z) = 550 + 0.00015\,(x-800)^2 + 0.001\,(z-500)^2$. The wave source is a point-source Ricker wavelet, located in the middle of the model. The finite-difference result exhibits dispersion artifacts while the result of the lowrank approximation, similarly to that of the FFD method, is dispersion-free. lowrank/impres wavefd,wave

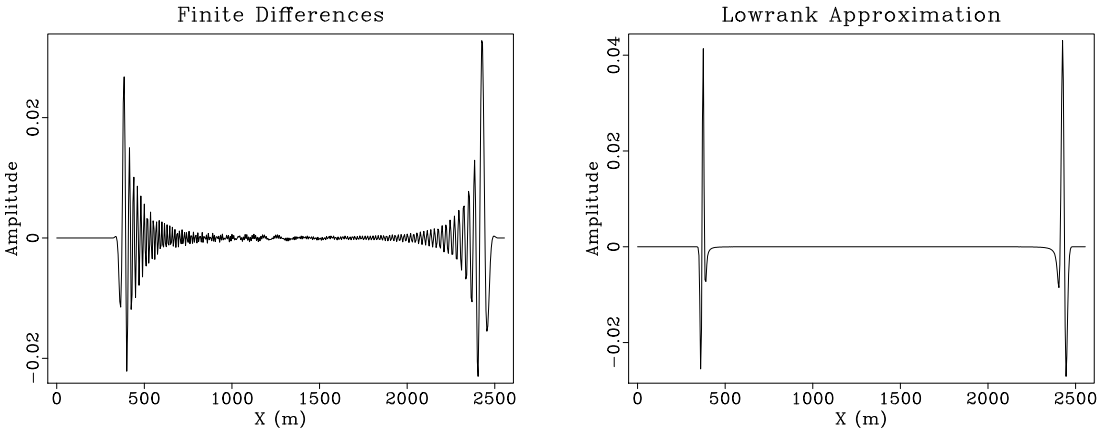


Figure 4: Horizontal slices through wavefield snapshots in Figure 3 lowrank/impres slicefd,slice

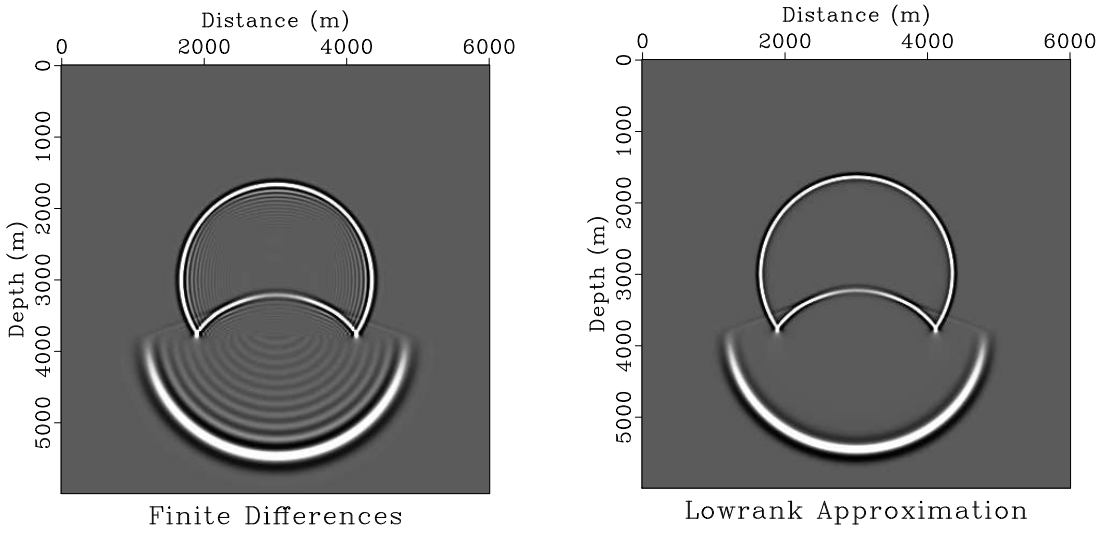


Figure 5: Wavefield snapshot in a simple two-layer velocity model using (a) fourth-order finite-difference method and (b) lowrank approximation. The upper-layer velocity is 1500 m/s, and the bottom-layer velocity is 4500 m/s. The finite-difference result exhibits clearly visible dispersion artifacts while the result of the lowrank approximation is dispersion-free. lowrank/fowler fwavefd,fwave

between a dispersion-infested result of the classic finite-difference method (second-order in time, fourth-order in space) and a dispersion-free result of the lowrank approximation is clearly visible in Figure 5. The time step was 2 ms, which corresponded to the approximation rank of 3. In our implementation, the CPU time for finding the lowrank approximation was 2.69 s, the single-processor CPU time for extrapolation for 601 time steps was 19.76 s or 2.48 times slower than the corresponding time for the finite-difference extrapolation (7.97 s). At larger time steps, the finite-difference method in this model becomes unstable, while the lowrank method remains stable but requires a higher rank.

Next, we move to isotropic wave extrapolation in a complex 2-D velocity field. Figure 6 shows a portion of the BP velocity model (6), containing a salt body. The wavefield snapshot (shown in Figure 7) confirms the ability of our method to handle complex models and sharp velocity variations. The lowrank decomposition in this case corresponds to $N = M = 3$, with the expected error of less than 10^{-7}. Increasing the time step size Δt does not break the algorithm but increases the rank of the approximation and correspondingly the number of the required Fourier transforms. For example, increasing Δt from 1 ms to 5 ms leads to $N = M = 5$.

Our next example is isotropic wave extrapolation in a 3-D complex velocity field: the SEG/EAGE salt model (5) shown in Figure 8. A dispersion-free wavefield snapshot is shown in Figure 9. The lowrank decomposition used $N = M = 2$, with the expected error of 10^{-5}.

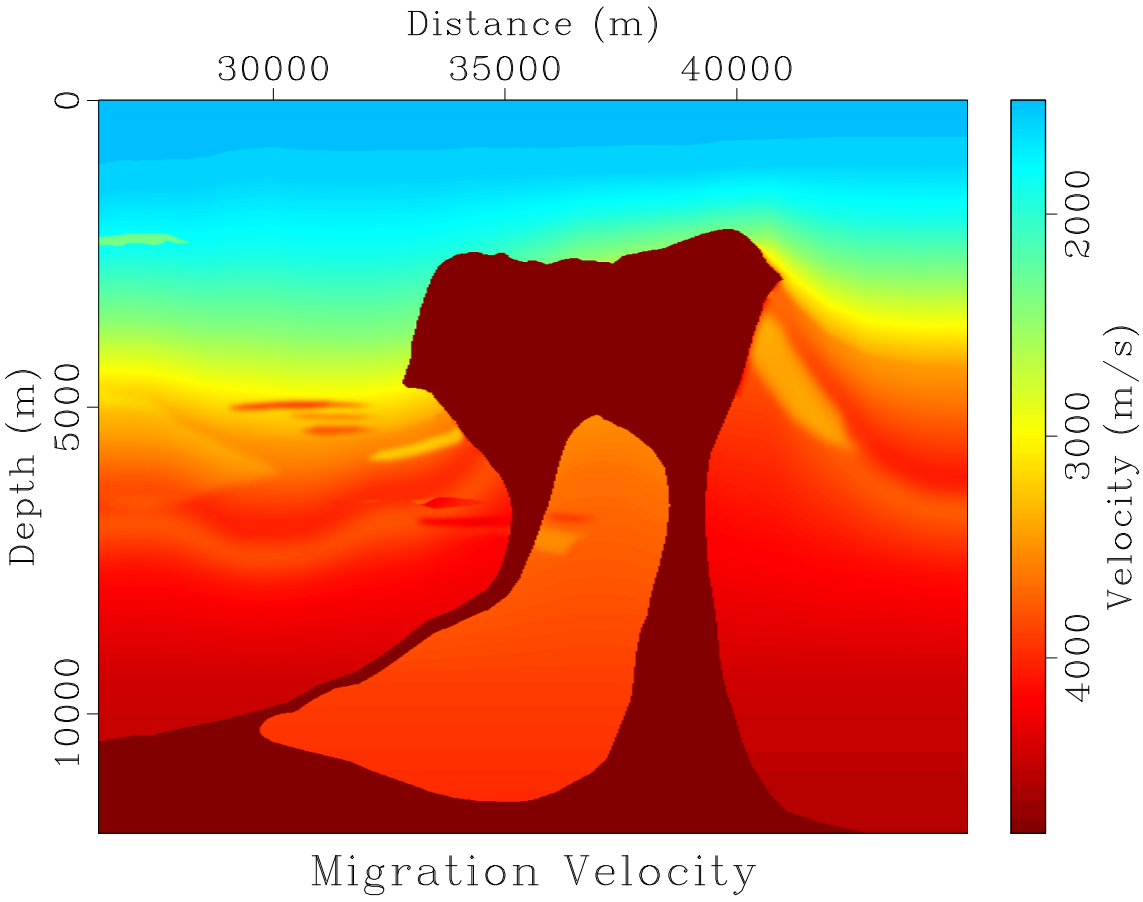


Figure 6: Portion of BP-2004 synthetic isotropic velocity model. lowrank/bp sub

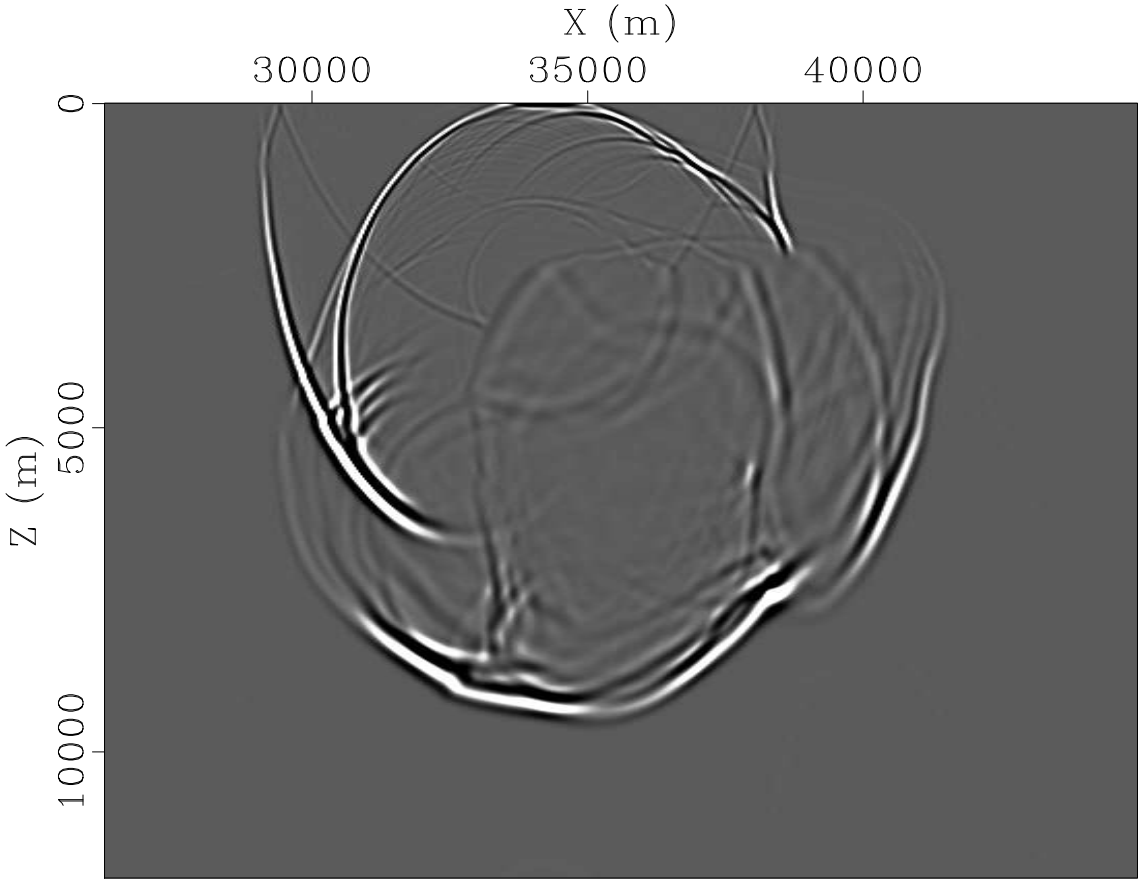


Figure 7: Wavefield snapshot for the velocity model shown in Figure 6. lowrank/bp snap

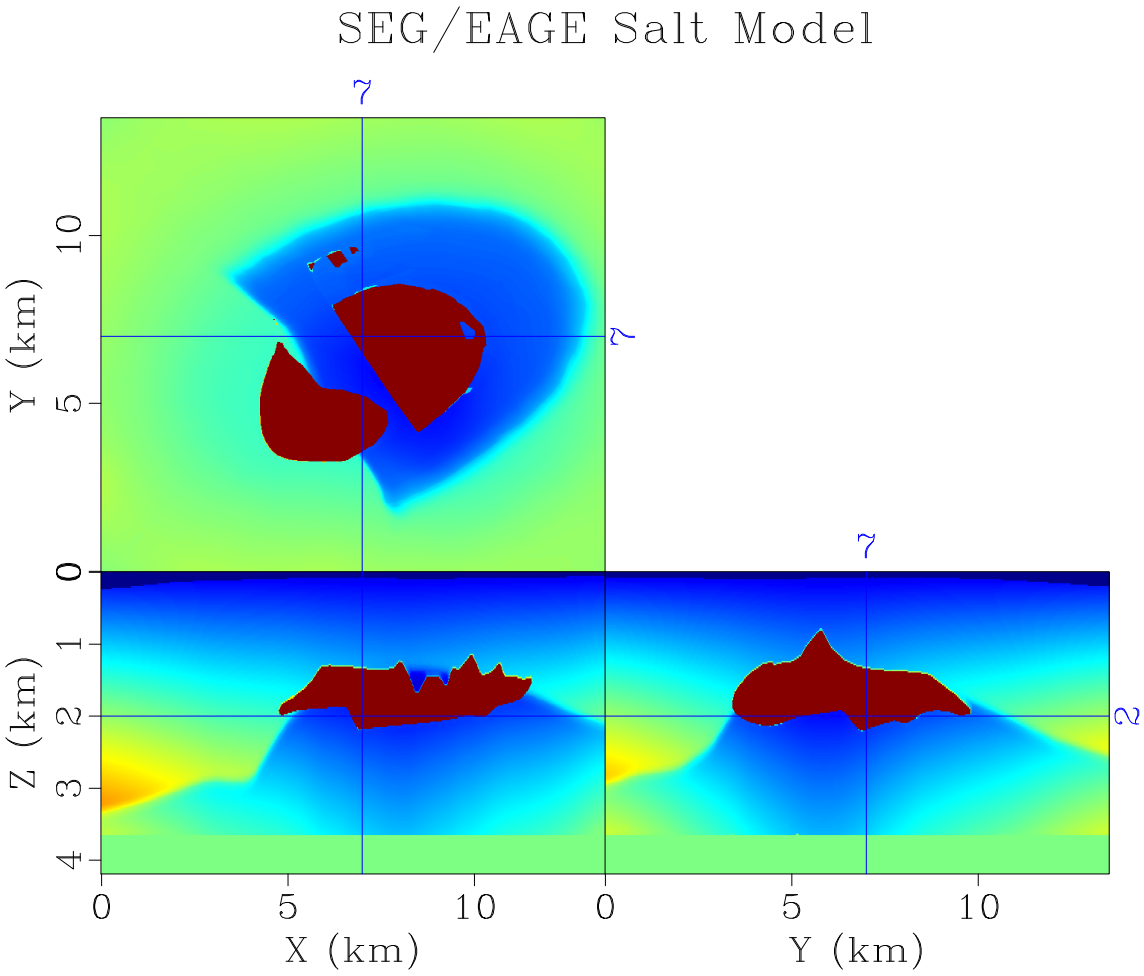


Figure 8: SEG/EAGE 3-D salt model. lowrank/threed salt

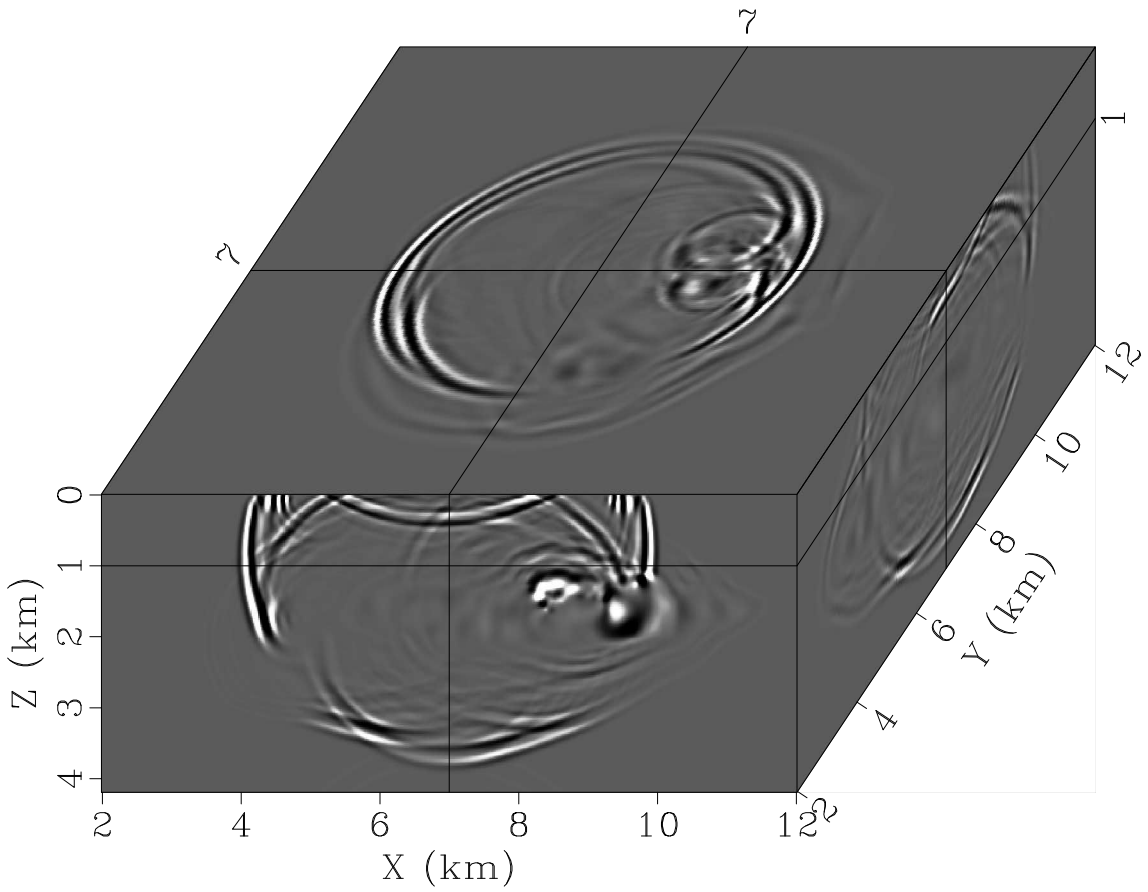


Figure 9: Snapshot of a point-source wavefield propagating in the SEG/EAGE 3-D salt model. lowrank/threed wave3

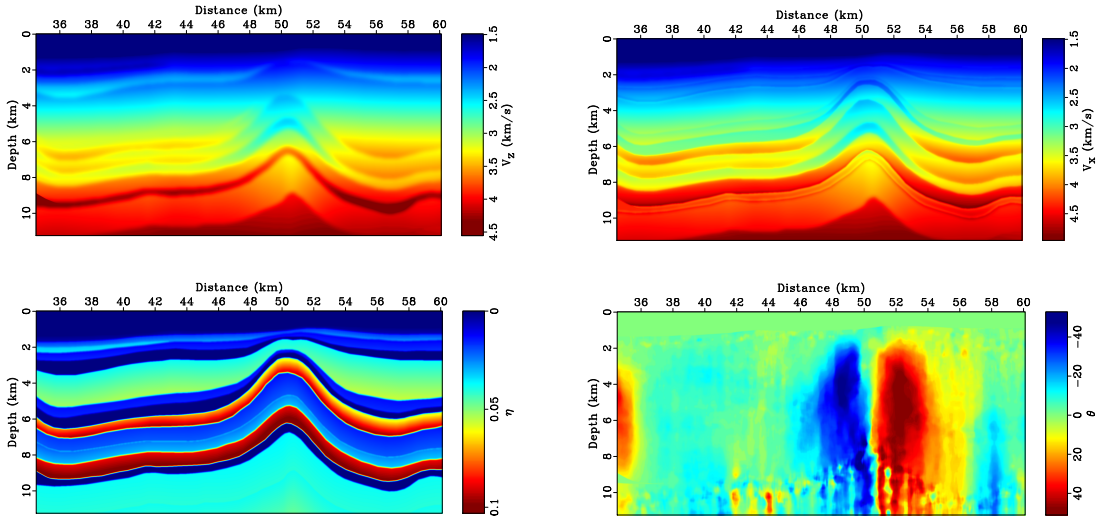


Figure 10: Portion of BP-2007 anisotropic benchmark model. (a) Velocity along the axis of symmetry. (b) Velocity perpendicular to the axis of symmetry. (c) Anellipticity parameter η. (d) Tilt of the symmetry axis. lowrank/bptti vpend2,vxend2,etaend2,thetaend2

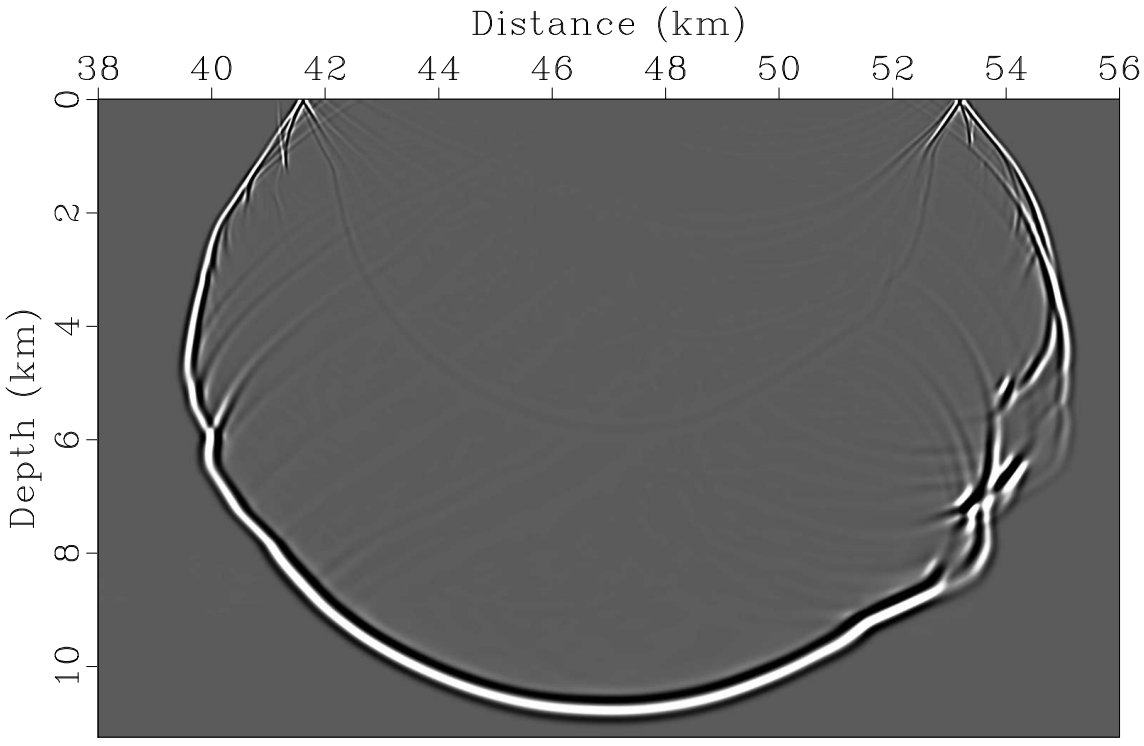


Figure 11: Wavefield snapshot for the velocity model shown in Figure 10. lowrank/bptti snap4299

Finally, we illustrate wave propagation in a complex anisotropic model. The model is a 2007 anisotropic benchmark dataset from BP*. It exhibits a strong TTI (tilted transverse isotropy) with a variable tilt of the symmetry axis (Figure 10). A wavefield snapshot is shown in Figure 9. Because of the complexity of the wave propagation patterns, the lowrank decomposition took $N = M = 10$ in this case and required 10 FFTs per time step. In a TTI medium, the phase velocity $V(\mathbf{x}, \mathbf{k})$ from equation (10) can be expressed with the help of the acoustic approximation (1; 2; 14)

$$V(\mathbf{x},\mathbf{k}) = \sqrt{\frac{1}{2}(v_x^2\,\hat{k}_x^2 + v_z^2\,\hat{k}_z^2) + \frac{1}{2}\sqrt{(v_x^2\,\hat{k}_x^2 + v_z^2\,\hat{k}_z^2)^2 - \frac{8\eta}{1+2\eta}v_x^2 v_z^2\,\hat{k}_x^2\,\hat{k}_z^2}}\;, \tag{15}$$

where v_x is the P-wave phase velocity in the symmetry plane, v_z is the P-wave phase velocity in the direction normal to the symmetry plane, η is the anellipticity parameter (4), and $\hat{k}_x$ and $\hat{k}_z$ stand for the wavenumbers evaluated in a rotated coordinate system aligned with the symmetry axis:

$$\begin{aligned} \hat{k}_x &= k_x \cos\theta + k_z \sin\theta \\ \hat{k}_z &= k_z \cos\theta - k_x \sin\theta\;, \end{aligned} \tag{16}$$

where θ is the tilt angle measured with respect to horizontal.

CONCLUSIONS

We have presented a novel algorithm for wave extrapolation in heterogeneous and anisotropic media. The algorithm is based on a lowrank approximation of the extrapolation symbol. It reduces the cost of extrapolation to that of a small number of FFT operations per time step, which correspond to the approximation rank. The algorithm has a high, spectral accuracy. In that sense, it is comparable with a number of other recently proposed FFT-based methods. Its advantage is a direct control on the accuracy-efficiency trade-off by controlling the rank of the approximation and the corresponding approximation error. We propose to incorporate the lowrank extrapolation algorithm in seismic imaging by reverse-time migration.

ACKNOWLEDGMENTS

We thank KAUST for partial financial support; Tariq Alkhalifah, Björn Engquist, Laurent Demanet, and Paul Fowler for useful discussions; and BP for releasing benchmark synthetic models.

The reproducible computational examples in this paper use the Madagascar open-source software package `http://www.ahay.org/`.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

*The dataset was created by Hemang Shah and is provided at `http://software.seg.org/` courtesy of BP Exploration Operation Company Limited.

LINEAR-TIME ALGORITHM FOR A LOWRANK MATRIX APPROXIMATION

In this appendix, we outline the lowrank matrix approximation algorithm in more details.

Let N_x be the number of samples both in space and wavenumber. Let us denote the samples in the spatial domain by $\mathbf{x} = \{x_1, \ldots, x_{N_x}\}$ and the ones in the Fourier domain by $\mathbf{k} = \{k_1, \ldots, k_{N_x}\}$. The elements of the interaction matrix $\mathbf{W}$ from equation (6) are then defined as

$$W_{ij} = e^{\imath[\phi(x_i,k_j,\Delta t]-x_i\cdot k_j]}, \quad 1 \le i,j \le N_x. \tag{A-1}$$

Here we describe an algorithm by (9) that generates, in a time linear with respect to N_x, an approximate factorization of $\mathbf{W}$ of rank r in the following form

$$\mathbf{W} \approx \mathbf{U}\mathbf{M}\mathbf{V}^* \,, \tag{A-2}$$

where $\mathbf{U}$ consists of r selected columns from $\mathbf{W}$, $\mathbf{M}$ is a matrix of size $r \times r$ and $\mathbf{V}^*$ consists of r selected rows from $\mathbf{W}$.

The first question is: which columns of $\mathbf{W}$ shall one pick for the matrix $\mathbf{U}$? It has been shown by (17) and (18) that the r-dimensional volume spanned by these columns should be the maximum or close to the maximum among all possible choices of r columns from $\mathbf{W}$. More precisely, suppose $\mathbf{W} = [w_1, \ldots, w_{N_x}]$ is a column partitioning of $\mathbf{W}$. Then one aims to find $\{j_1, \ldots, j_r\}$ such that

$$\{j_1, \ldots, j_r\} = \mathrm{argmin}_{\{j_1,\ldots,j_r\}} \mathrm{vol}_r(w_{j_1}, \ldots, w_{j_r}). \tag{A-3}$$

However, finding a set of r columns with almost the maximum r-dimensional volume is a computationally difficult problem due to the following two reasons. First, the length of the vectors N is typically very large for three dimensional problems, hence manipulating these vectors can be costly. Second, the number of the vectors N_x is also large. A exhaustive search over all possible choices of r vectors to find the one with the maximum volume is prohibitive expensive, so one needs to find a more practical approach.

In order to overcome the problem associated with long vectors, the first idea is to project to a lower dimensional space and search for the set of vectors with maximum volume among the projected vectors. However, one needs to ensure that the volume is roughly preserved after the projection so that the set of vectors with the maximum projected volume also has a near-maximum volume in the original space. One of the most celebrated theorems in high dimensional geometry and probability is the following Johnson-Lindenstrauss lemma (19).

Theorem 1 *Let $v_1, \ldots, v_N$ be a set of N vectors in R^d. Let T be a randomly generated subspace of dimension $t = O(\log N/\varepsilon^2)$ and use P_T to denote the orthogonal projection*

onto T. Then with high probability,

$$(1-\varepsilon)\|v_i - v_j\| \leq \sqrt{\frac{d}{t}}\|P_T v_i - P_T v_j\| \leq (1+\varepsilon)\|v_i - v_j\|$$

for $1 \leq i, j \leq N$.

This theorem essentially says that projecting to a subspace of dimension $O(\log N)$ preserves the pairwise distance between N arbitrary vectors. There is an immediate generalization of this theorem due to (20), formulated slightly differently for our purpose.

Theorem 2 *Let $v_1, \ldots, v_N$ be a set of N vectors in R^d. Let T be a randomly generated subspace of dimension $t = O(r^3 \log N/\varepsilon^2)$ and use P_T to denote the orthogonal projection onto T. Then with high probability,*

$$(1-\varepsilon)\cdot \mathrm{vol}_r(v_{i_1}, \ldots, v_{i_r}) \leq \left(\frac{d}{t}\right)^{r/2} \mathrm{vol}_r(P_T v_{i_1}, \ldots, P_T v_{i_r}) \leq (1+\varepsilon)\cdot \mathrm{vol}_r(v_{i_1}, \ldots, v_{i_r})$$

for any $\{i_1, \ldots, i_r\} \subset \{1, \ldots, N\}$.

The main step of the proof is to bound the singular values of a random matrix between $(1-\varepsilon)^{1/r}$ and $(1+\varepsilon)^{1/r}$ (after a uniform scaling) and this ensures that the r-dimensional volume is preserved within a factor of $(1-\varepsilon)$ and $(1+\varepsilon)$. In order to obtain this bound on the singular values, we need t to be $O(r^3 \log N)$. However, bounding the singular values is only one way to bound the volume, hence it is possible to improve the dependence of t on r. In fact, in practice, we observe that t only needs to scale like $O(r \log N)$.

Given a generic subspace T of dimension t, computing the projections $P_T w_1, \ldots, P_T w_N$ takes $O(tN^2)$ steps. Recall that our goal is to find an algorithm with linear complexity, hence this is still too costly. In order to reduce the cost of the random projection, the second idea of our approach is to randomly choose t coordinates and then project (or restrict) each vector only to these coordinates. This is a projection with much less randomness but one that is much more efficient to apply. Computationally, this is equivalent to restricting $\mathbf{W}$ to t randomly selected rows. We do not yet have a theorem regarding the volume for this projection. However, it preserves the r-dimensional volume very well for the matrix $\mathbf{W}$ and this is in fact due to the oscillatory nature of the columns of $\mathbf{W}$. We denote the resulting vectors by $\{\widetilde{\mathbf{w}}_1, \ldots, \widetilde{\mathbf{w}}_{N_x}\}$.

The next task is to find a set of columns $\{j_1, \ldots, j_r\}$ so that the volume $\mathrm{vol}_r(\widetilde{\mathbf{w}}_{j_1}, \ldots, \widetilde{\mathbf{w}}_{j_r})$ is nearly maximum. As we mentioned earlier, exhaustive search is too costly. To overcome this, the third idea is to use the following pivoted QR algorithm (or pivoted Gram-Schmidt process) to find the r columns.

1: **for** $s = 1, \ldots, r$ **do**
2: Find j_s among $\{1, \ldots, N\} \setminus j_1, \ldots, j_{s-1}$ such that $\widetilde{\mathbf{w}}_{j_s}$ has the largest norm

3: Orthogonalize the vectors $\widetilde{\mathbf{w}}_j$ for $j \in \{1,\ldots,N\}\setminus j_1,\ldots,j_s$ with $\widetilde{\mathbf{w}}_{j_s}$ and update them
4: **end for**
5: $\{j_1,\ldots,j_r\}$ is the column set required

Once the column set is found, we set $\mathbf{U} = [\mathbf{w}_{j_1},\ldots,\mathbf{w}_{j_r}]$.

In order to identify $\mathbf{V}^*$, one needs to find a set of r rows of $\mathbf{W}$ that has an almost maximum volume. To do that, we repeat the same steps now to $\mathbf{W}^*$. More precisely, let

$$\mathbf{W} = \begin{bmatrix} \mathbf{m}_1 \\ \vdots \\ \mathbf{m}_{N_x} \end{bmatrix} \tag{A-4}$$

be the row partitioning of the matrix $\mathbf{W}$. The algorithm takes the following steps:

1: Select uniform randomly a set of t columns and obtain an $N_x \times t$ tall matrix
2: Perform pivoted QR algorithm on the rows of this tall matrix and denote the first r rows selected by $\{i_1,\ldots,i_r\}$
3: The matrix $\mathbf{V}^*$ is

$$\mathbf{V}^* = \begin{bmatrix} \mathbf{m}_{i_1} \\ \vdots \\ \mathbf{m}_{i_r} \end{bmatrix} . \tag{A-5}$$

Once both $\mathbf{U}$ and $\mathbf{V}^*$ are identified, the last task is to compute the $r \times r$ matrix $\mathbf{M}$ for $\mathbf{W} \approx \mathbf{U}\mathbf{M}\mathbf{V}^*$. Minimizing

$$\min_{\mathbf{M}} \|\mathbf{W} - \mathbf{U}\mathbf{M}\mathbf{V}^*\|_F \tag{A-6}$$

yeilds $\mathbf{M} = (\mathbf{U})^\dagger \mathbf{W} (\mathbf{V}^*)^\dagger$ where $\dagger$ stands for the pseudo-inverse. However, this formula requires taking matrix product with $\mathbf{W}$, which takes $O(t\,N_x^2)$ steps. In order to achieve linear scaling, the fourth idea of our approach is to select randomly a set of t rows A and a set of t columns B and minimize

$$\min_{\mathbf{M}} \|\mathbf{W}(A,B) - \mathbf{U}(A,:)\,\mathbf{M}\,\mathbf{V}(B,:)^*\|_F . \tag{A-7}$$

The solution for this problem is

$$\mathbf{M} = (\mathbf{U}(A,:))^\dagger\, \mathbf{W}(A,B)\, (\mathbf{V}(B,:)^*)^\dagger . \tag{A-8}$$

Let us now discuss the overall cost of this algorithm. Random sampling of t rows and t columns of the matrix $\mathbf{W}$ clearly takes $O(t\,N_x)$ steps. Pivoted QR factorization on the projected columns $\{\widetilde{\mathbf{w}}_1,\ldots,\widetilde{\mathbf{w}}_{N_x}\}$ takes $O(t^2\,N_x)$ steps and the cost for for the pivoted QR factorization on the projected rows. Finally, performing pseudo-inverses takes $O(t^3)$ steps. Therefore, the overall cost of the algorithm is $O(t\,N_x) + O(t^2\,N_x) + O(t^3) = O(t^2\,N_x)$. As we mentioned earlier, in practice $t = O(r \log N_x)$. Hence, the overall cost is linear in N_x.

REFERENCES

[1] Alkhalifah, T., 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.
[2] ———, 2000, An acoustic wave equation for anisotropic media: Geophysics, **65**, 1239–1250.
[3] Alkhalifah, T., and S. Fomel, 2010, Source-receiver two-way wave extrapolation for prestack exploding-reflector modeling and migration, *in* 80th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2950–2955.
[4] Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.
[5] Aminzadeh, F., J. Brac, and T. Kunz, 1997, 3-D salt and overthrust models: SEG.
[6] Billette, F. J., and S. Brandsberg-Dahl, 2005, The 2004 BP velocity benchmark: 67th Annual EAGE Meeting, EAGE, Expanded Abstracts, B305.
[7] Chu, C., and P. Stoffa, 2008, A pseudospectral-finite difference hybrid approach for large-scale seismic modeling and RTM on parallel computers, *in* 78th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2087–2091.
[8] Du, X., R. P. Fletcher, and P. J. Fowler, 2010, Pure P-wave propagators versus pseudo-acoustic propagators for RTM in VTI meida: 72nd Annual EAGE Meeting, EAGE, Expanded Abstracts, C013.
[9] Engquist, B., and L. Ying, 2009, A fast directional algorithm for high frequency acoustic scattering in two dimensions: Commun. Math. Sci., **7**, 327–345.
[10] Etgen, J., 1986, High order finite-difference reverse time migration with the two way nonreflecting wave equation, *in* SEP-48: Stanford Exploration Project, 133–146.
[11] Etgen, J., and S. Brandsberg-Dahl, 2009, The pseudo-analytical method: application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation, *in* 79th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2552–2556.
[12] Fomel, S., 2003a, Theory of differential offset continuation: Geophysics, **68**, 718–732.
[13] ———, 2003b, Velocity continuation and the anatomy of residual prestack time migration: Geophysics, **68**, 1650–1661.
[14] ———, 2004, On anelliptic approximations for qP velocities in VTI media: Geophys. Prosp., **52**, 247–259.
[15] Fowler, P., X. Du, and R. P. Fletcher, 2010, Recursive integral time extrapolation methods for scalar waves, *in* 80th Ann. Internat. Mtg: Soc. of Expl. Geophys., 3210–3215.
[16] Golub, G. H., and C. F. Van Loan, 1996, Matrix computations: John Hopkins.
[17] Goreinov, S., E. Tyrtyshnikov, and N. Zamarashkin, 1997, A theory of pseudoskeleton approximations: Linear Algebra Appl., **261**, 1–21.
[18] Gu, M., and S. C. Eisenstat, 1996, Efficient algorithms for computing a strong rank-revealing QR factorization: SIAM J. Sci. Comput., **17**, 848–869.
[19] Johnson, W. B., and J. Lindenstrauss, 1984, Extensions of Lipschitz mappings into a Hilbert space, *in* Conference in modern analysis and probability (New Haven, Conn., 1982): Amer. Math. Soc., volume **26** *of* Contemp. Math., 189–206.

[20] Magen, A., 2002, Dimensionality reductions that preserve volumes and distance to affine spaces, and their algorithmic applications, *in* Randomization and approximation techniques in computer science: Springer, volume **2483** *of* Lecture Notes in Comput. Sci., 239–253.

[21] Pestana, R. C., and P. L. Stoffa, 2011, Using the rapid expansion method for accurate time-stepping in modeling and reverse-time migration: Geophysics, **76**, S177–S185.

[22] Reshef, M., D. Kosloff, M. Edwards, and C. Hsiung, 1988, Three-dimensional acoustic modeling by the Fourier method: Geophysics, **53**, 1175–1183.

[23] Song, J., 2001, The optimized expression of a high dimensional function/manifold in a lower dimensional space: Chinese Scientific Bulletin, **46**, 977–984.

[24] Song, X., and S. Fomel, 2011, Fourier finite-difference wave propagation: Geophysics, **76**, T123–T129.

[25] Song, X., S. Fomel, L. Ying, and T. Ding, 2011, Lowrank finite-differences for wave extrapolation, *in* 81st Ann. Internat. Mtg: Soc. of Expl. Geophys., 3372–3376.

[26] Tal-Ezer, H., D. Kosloff, and Z. Koren, 1987, An accurate scheme for seismic forward modelling: Geophys. Prosp., **35**, 479–490.

[27] Wards, B. D., G. F. Margrave, and M. P. Lamoureux, 2008, Phase-shift time-stepping for reverse-time migration, *in* 78th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2262–2266.

[28] Zhang, Y., and G. Zhang, 2009, One-step extrapolation method for reverse time migration: Geophysics, **74**, A29–A33.

OC-seislet: seislet transform construction with differential offset continuation

Yang Liu[*†] *and Sergey Fomel*[†]

ABSTRACT

Many of the geophysical data analysis problems, such as signal-noise separation and data regularization, are conveniently formulated in a transform domain, where the signal appears sparse. Classic transforms such as the Fourier transform or the digital wavelet transform, fail occasionally in processing complex seismic wavefields, because of the nonstationarity of seismic data in both time and space dimensions. We present a sparse multiscale transform domain specifically tailored to seismic reflection data. The new wavelet-like transform – the *OC-seislet transform* – uses a differential offset-continuation (OC) operator that predicts prestack reflection data in offset, midpoint, and time coordinates. It provides high compression of reflection events. In the transform domain, reflection events get concentrated at small scales. Its compression properties indicate the potential of OC-seislets for applications such as seismic data regularization or noise attenuation. Results of applying the method to both synthetic and field data examples demonstrate that the OC-seislet transform can reconstruct missing seismic data and eliminate random noise even in structurally complex areas.

INTRODUCTION

Digital wavelet transforms are excellent tools for multiscale data analysis. The wavelet transform is more powerful when compared with the classic Fourier transform, because it is better fitted for representing non-stationary signals. Wavelets provide a sparse representation of piecewise regular signals, which may include transients and singularities (28). In recent years, many wavelet-like transforms that explore directional characteristics of images (33; 12; 29; 36) were proposed. The curvelet transform in particular has found important applications in seismic imaging and data analysis (13; 7; 23). (19) and (20) investigated the possibility of designing a wavelet-like transform tailored specifically to seismic data and introduced it under the name of the *seislet transform*. Based on the digital wavelet transform (DWT), the seislet transform follows patterns of seismic events (such as local slopes in 2-D and frequencies in 1-D) when analyzing those events at different scales. The seislet transform's compression ability finds applications in common data processing tasks such as data regularization and noise attenuation. However, the problem of pattern detection limits its further applications. In 2-D, conflicting slopes at a single data point are difficult

to detect reliably even using advanced methods (15). It is also difficult to estimate local slopes in the presence of strong noise. A similar situation occurs in the 1-D case, in which it is difficult to exactly represent a known seismic signal using a limited set of frequencies.

Offset continuation is a process of seismic data transformation between different offsets (11; 6; 31; 18). Different types of dip moveout (DMO) operators (22) can be regarded as continuation to zero offset and derived as solutions to initial-value problems with the offset-continuation differential equation. In the shot-record domain, offset continuation transforms to shot continuation, which describes the process of transforming reflection seismic data along shot location (2; 32; 17). The 3-D analog is known as azimuth moveout or AMO (3; 16). (5) developed a general platform for Kirchhoff data mapping, which includes offset continuation as a special case.

In this paper, we propose to incorporate offset continuation as the prediction operator into the seislet transform. We design the transform in the log-stretch-frequency domain, where each frequency slice can be processed independently and in parallel. We expect the new seislet transform to perform better than the previously proposed seislet transform by plane-wave destruction, PWD-seislet transform (20), in cases of moderate velocity variations and complex structures that generate conflicting dips in the data.

THEORETICAL BASIS

OC-seislet transform follows the same construction strategy, the lifting scheme, as the PWD-seislet transform but employs the offset-continuation operator for prediction at each transform scale.

The lifting scheme for the seislet transform

The lifting scheme (34) provides a convenient approach for designing digital wavelet transforms. The general recipe is as follows:

1. Organize the input data as a sequence of records. For OC-seislet transform of 2-D seismic reflection data, the input is in the 'frequency'-'midpoint wavenumber'-'offset' domain after the log-stretched NMO correction (6), and the transform direction is offset.

2. Divide the data records (along the offset axis in the case of the OC-seislet transform) into even and odd components $\mathbf{e}$ and $\mathbf{o}$. This step works at one scale level.

3. Find the residual difference $\mathbf{r}$ between the odd component and its prediction from the even component:
$$\mathbf{r} = \mathbf{o} - \mathbf{P}[\mathbf{e}] \ , \tag{1}$$

where $\mathbf{P}$ is a *prediction* operator. For example, one can obtain Cohen-Daubechies-Feauveau (CDF) 5/3 biorthogonal wavelets (9) by defining the prediction operator as a linear interpolation between two neighboring samples,

$$\mathbf{P}[\mathbf{e}]_k = (\mathbf{e}_{k-1} + \mathbf{e}_k)/2 \ , \tag{2}$$

where k is an index number at the current scale level.

4. Find an approximation $\mathbf{c}$ of the data by updating the even component:

$$\mathbf{c} = \mathbf{e} + \mathbf{U}[\mathbf{r}] \ , \tag{3}$$

where $\mathbf{U}$ is an *update* operator. Constructing the update operator for CDF 5/3 biorthogonal wavelets aims at preserving the running average of the signal (35):

$$\mathbf{U}[\mathbf{r}]_k = (\mathbf{r}_{k-1} + \mathbf{r}_k)/4 \ . \tag{4}$$

5. The coarse approximation $\mathbf{c}$ becomes the new data, and the sequence of steps is repeated on the new data to calculate the transform coefficients at a coarser scale level.

Next, we define new prediction and update operators using offset-continuation operators.

OC-seislet structure

We define the OC-seislet transform by specifying prediction and update operators with the help of the offset-continuation operator. Prediction and update operators for the OC-seislet transform are specified by modifying the biorthogonal wavelet construction in equations 2 and 5 as follows (19; 20):

$$\mathbf{P}[\mathbf{e}]_k = \left(\mathbf{S}_k^{(+)}[\mathbf{e}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{e}_k]\right)/2 \tag{5}$$

$$\mathbf{U}[\mathbf{r}]_k = \left(\mathbf{S}_k^{(+)}[\mathbf{r}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{r}_k]\right)/4 \ , \tag{6}$$

where $\mathbf{S}_k^{(+)}$ and $\mathbf{S}_k^{(-)}$ are operators that predict the data record (a common-offset section) by differential offset continuation from its left and right neighboring common-offset sections with different offsets. Offset continuation operators provide the physical connection between data records. The theory of offset continuation is reviewed in Appendix A.

One can also employ a higher-order transform, for example, by using the template of the CDF 9/7 biorthogonal wavelet transform, which is used in JPEG-2000 compression (25). There is only one stage (one prediction and one update) for the CDF 5/3 wavelet transform, but there are two cascaded stages and one scaling operation

for CDF 9/7 wavelet transform. Prediction and update operators for a high-order OC-seislet transform are defined as follows:

$$\mathbf{P}_1[\mathbf{e}]_k = (\mathbf{S}_k^{(+)}[\mathbf{e}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{e}_k]) \cdot \alpha \ , \tag{7}$$

$$\mathbf{U}_1[\mathbf{r}]_k = (\mathbf{S}_k^{(+)}[\mathbf{r}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{r}_k]) \cdot \beta \ , \tag{8}$$

$$\mathbf{P}_2[\mathbf{e}]_k = (\mathbf{S}_k^{(+)}[\mathbf{e}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{e}_k]) \cdot \gamma \ , \tag{9}$$

$$\mathbf{U}_2[\mathbf{r}]_k = (\mathbf{S}_k^{(+)}[\mathbf{r}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{r}_k]) \cdot \delta \ , \tag{10}$$

where the subscripts 1 and 2 represent the first and the second stage. α, β, γ, and δ are defined numerically as follows:

$$\begin{aligned} \alpha &= -1.586134342, \\ \beta &= -0.052980118, \\ \gamma &= 0.882911076, \\ \delta &= 0.443506852. \end{aligned}$$

One can combine equations 10, 2, 7, and 8 to finish the first stage, and repeatedly process the result by using equations 10, 2, 9, and 10. The scale normalization factors correspond to the CDF 9/7 biorthogonal wavelet transform (10). Scaling and coefficients are as follows:

$$\mathbf{e} = \mathbf{e} \cdot K \ , \tag{11}$$

$$\mathbf{o} = \mathbf{o} \cdot (1/K) \ , \tag{12}$$

where K= 1.230174105.

We used the high-order version of OC-seislet transform to process the synthetic and field data examples used in this paper.

Simple example

Figure 10(a) shows a 2-D slice out of the benchmark French model (21). We created a 2-D prestack dataset (Figure 2a) by Kirchhoff modeling. Three sections in Figure 2a show the time slice at time position 0.6 s (top section), common-offset section at offset position of 0.2 km (bottom-left section), and common-midpoint gather at midpoint position of 1.0 km (bottom-right section). The reflector with a round dome and corners creates complicated reflection events along both midpoint and offset axes. The inflection points of the reflector leads to traveltime triplications at some offsets. Figure 2b shows a preprocessed data cube in the F-K-offset domain after the log-stretched NMO correction and a double Fourier transform along the stretched time and midpoint axes. We apply the OC-seislet transform described above along the offset axis in Figure 2b. Thus, the offset axis becomes the scale axis. The cube of the transform coefficients is shown in Figure 3b and should be compared with the corresponding Fourier transform along the offset direction in Figure 3a. The

OC-seislet transform coefficients get concentrated at small scales, which enables an effective compression. In contrast, the Fourier transform develops large coefficients at coarser scales but has small residual coefficients at fine scales. Figure 2 shows a comparison between the decay of coefficients (sorted from large to small) between the Fourier transform and the OC-seislet transform. A significantly faster decay of the OC-seislet coefficients is evident.

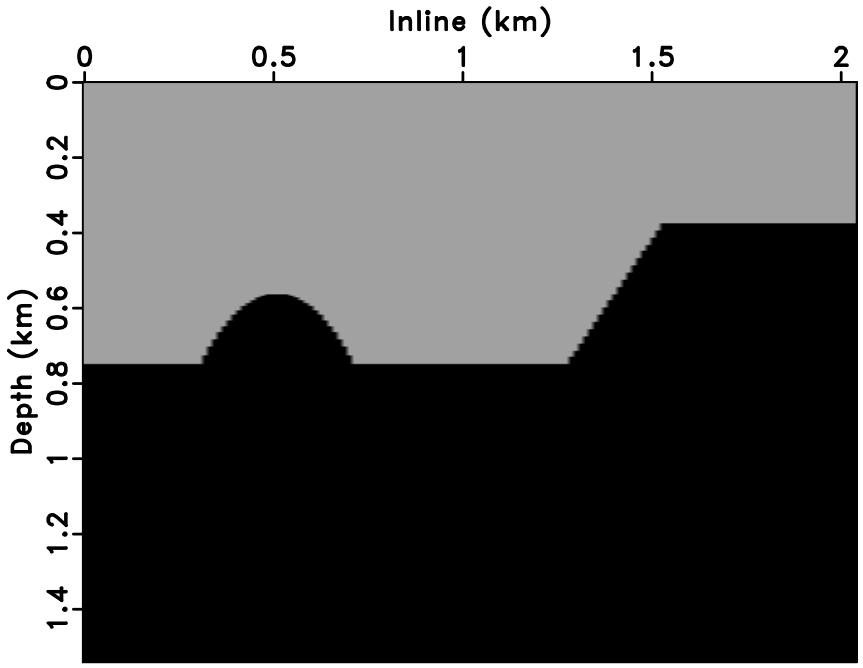


Figure 1: 2-D slice out of the benchmark French model (21). oc3/french slice

Iterative soft-thresholding

The proposed OC-seislet transform uses physical offset continuation to compress the reflection data after NMO and log-stretch transform of the time coordinate, followed by double Fourier transforms of the stretched time axis and midpoint axis. If seismic traces in the midpoint direction are missing, the Fourier transform may produce artifacts (spatial leakage) along the midpoint-wavenumber axis (39). Additionally, the missing seismic traces in the offset direction affect the continuity of predicted data. Once the OC-seislet transform is applied for predicting and compressing, the artifacts, discontinuity information, and random noise spread over different scales while the predictable reflection information gets compressed to large coefficients at small scales. A simple thresholding method can easily remove the small coefficients of artifacts. Finally, applying inverse OC-seislet transform, inverse FFTs both in time and midpoint axes, inverse log-stretch, and inverse NMO reconstructs the data while attenuating random noise, reducing artifacts along the midpoint axis, reconstructing continuity in the offset axis, and recovering main structural features without using

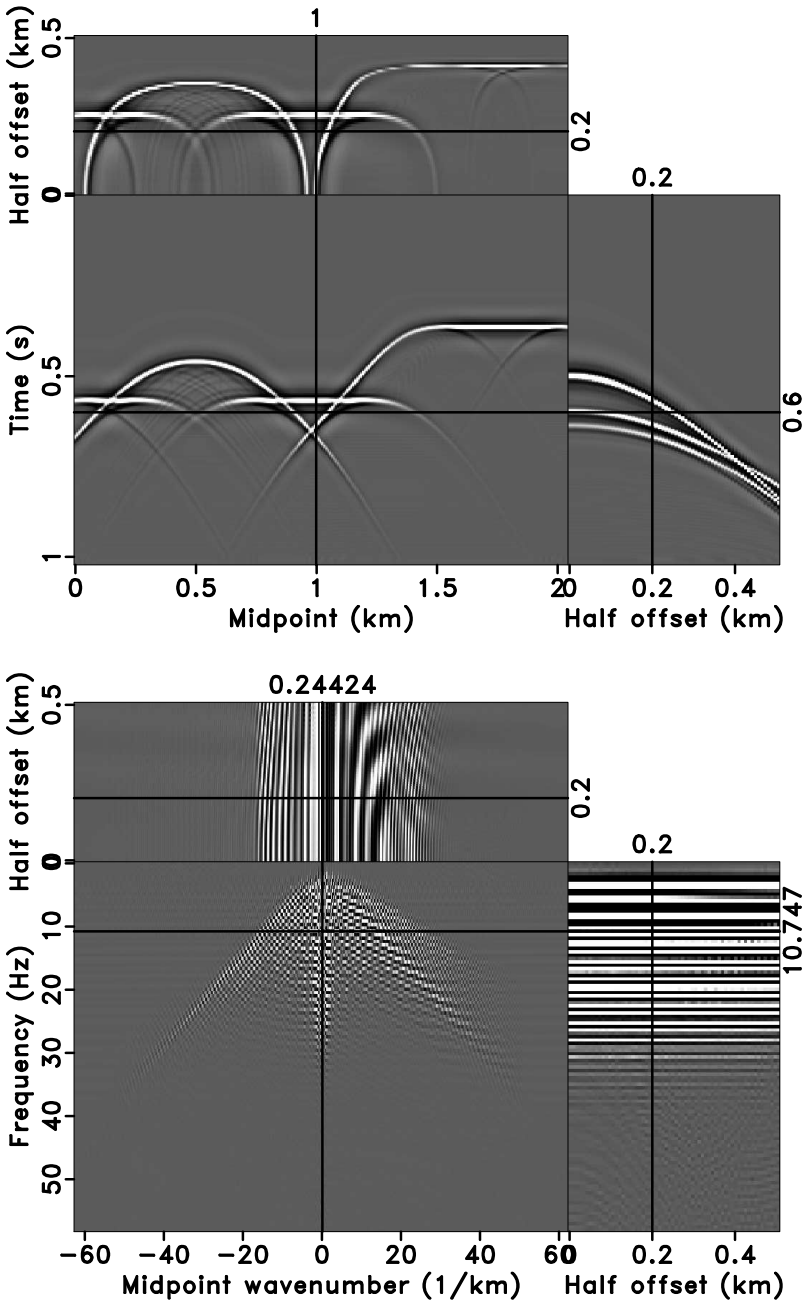


Figure 2: 2-D synthetic prestack data in t-x-offset domain (a) and in F-K-offset domain (b). oc3/french data,dinput

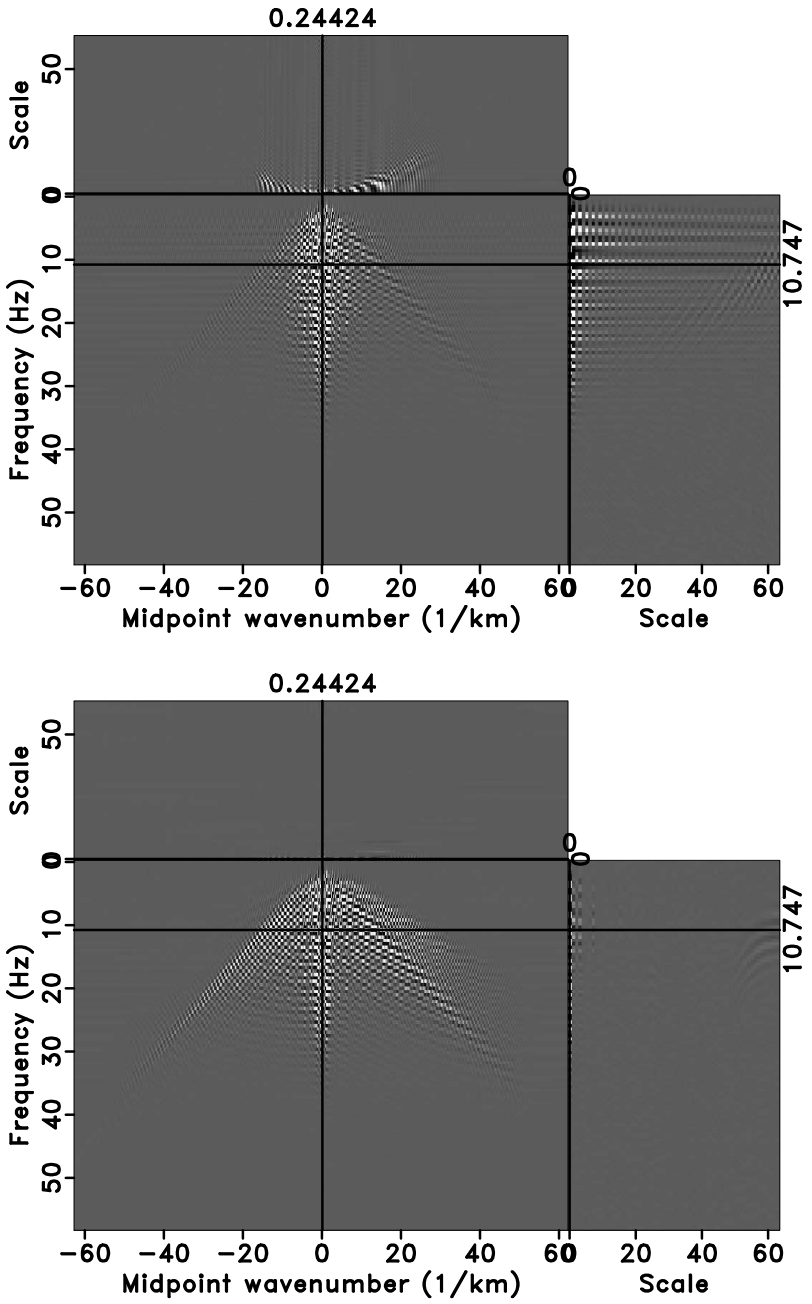


Figure 3: Fourier transform (a) and OC-seislet transform (b) of the input data from Figure 2b along the offset axis. oc3/french dfourier,dtran

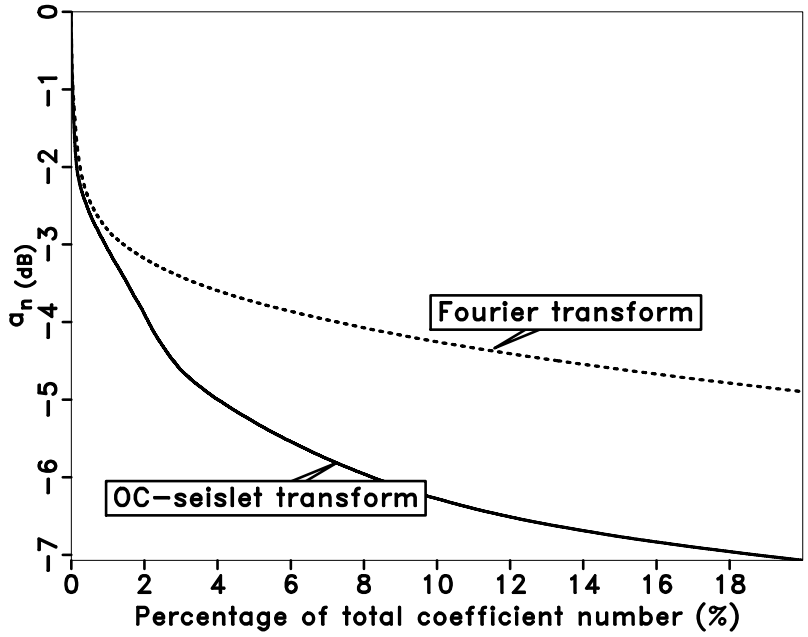


Figure 4: Transform coefficients sorted from large to small, normalized, and plotted on a decibel scale. Solid line: OC-seislet transform. Dashed line: Fourier transform. oc3/french compare

any assumptions about structural continuity in the midpoint-offset-time domain. The key steps are shown in Figure 5a.

The idea of sparse transforms has been thoroughly explored in the literature, with application to Fourier and curvelet transforms. (27) proposed a Fourier-based minimum weighted norm interpolation (MWNI) algorithm with iterative inversion to perform multidimensional reconstruction of seismic wavefields. (37) suggested an iterative Fourier-based matching pursuit (antileakage Fourier transform) for seismic data regularization. (1) used a Fourier-based method with iterative thresholding to solve seismic data interpolation problems. (24) presented a recovery method that exploits the curvelet frame. Missing data interpolation is a particular case of data regularization, where the input data are already given on a regular grid, and one needs to reconstruct only the missing values in empty bins (14). For input data with nonuniform spatial sampling, one can bin the data to a regular grid first and then use the proposed method for filling empty bins. It is also possible to generalize the method for combining irregular data interpolation with a sparse transform (39). The thresholding iteration helps a sparse transform to recover the missing information. We adopt a similar thresholding strategy (fixed thresholding value for each iteration) with the OC-seislet transform. The key steps of the algorithm are illustrated schematically in Figure 5b.

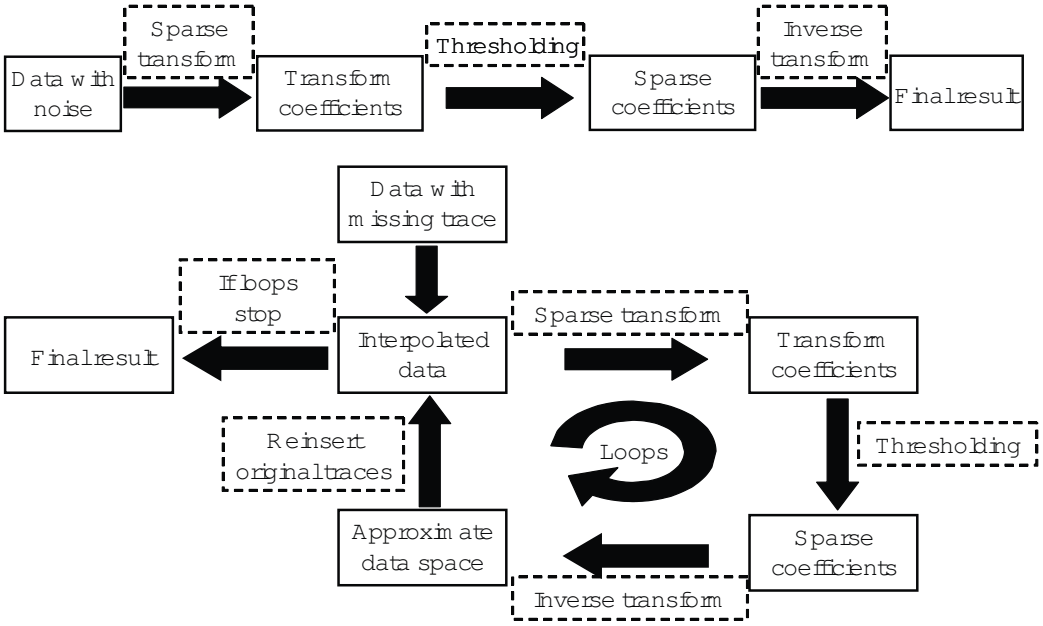


Figure 5: Schematic illustration of the thresholding workflow. Denoising with a simple soft-thresholding (a) and missing data interpolation with iterative soft-thresholding (b). oc3/XFig denoise,jpocs

SYNTHETIC DATA TESTS

We continue to use the synthetic model from Figure 10(a) to test the proposed method. Figure 6 is the data with normally-distributed random noise added. Figure 7a shows the datacube in the F-K-offset domain after the log-stretched NMO correction and a double Fourier transform along the stretched time axis and midpoint axis. We run the OC-seislet transform in parallel on individual frequency slices. Figure 7b shows the OC-seislet coefficients. All reflection information is concentrated in a small scale range. However, since random noise cannot be predicted well by the offset-continuation operator, it spreads throughout the whole transform domain. The inverse Fourier transform both in time and midpoint directions and the inverse log-stretch return the OC-seislet coefficients to the time-midpoint-scale domain. The coefficients at the zero scale represent stacking along the offset direction, which is equivalent to the DMO stack (Figure 8b).

We use a soft-thresholding method to separate reflection and random noise in the OC-seislet domain. Figure 8a displays the result after the inverse OC-seislet transform. Compared with Figure 7a, data in the F-K-offset domain contain only useful information. Figure 9b shows the denoising result in the t-x-offset domain after the double inverse Fourier transform, the inverse log-stretch and the inverse NMO. All characteristics of reflection and diffraction events are preserved well. For comparison, we used PWD-seislet transform and soft-thresholding method with the same threshold values to process the noisy data (Figure 6). The result is shown in Figure 9a.

Because PWD-seislet transform is based on the local dip information, a mixture of different dips from the triplications makes it difficult to process the data in individual common-midpoint gathers. The PWD-seislet transform compresses all information along the local events slopes in each common-midpoint gather. It separates reflection signal and noise but smears the crossing events, especially at the far offset. The corresponding signal-to-noise ratios for denoised results with PWD-seislet transform and OC-seislet transform are 24.95 dB and 41.45 dB, respectively. The differences (Figure 10) between noisy data (Figure 6) and denoised results with PWD-seislet transform (Figure 9a) and OC-seislet transform (Figure 9b) further illustrate the effectiveness of the OC-seislet transform.

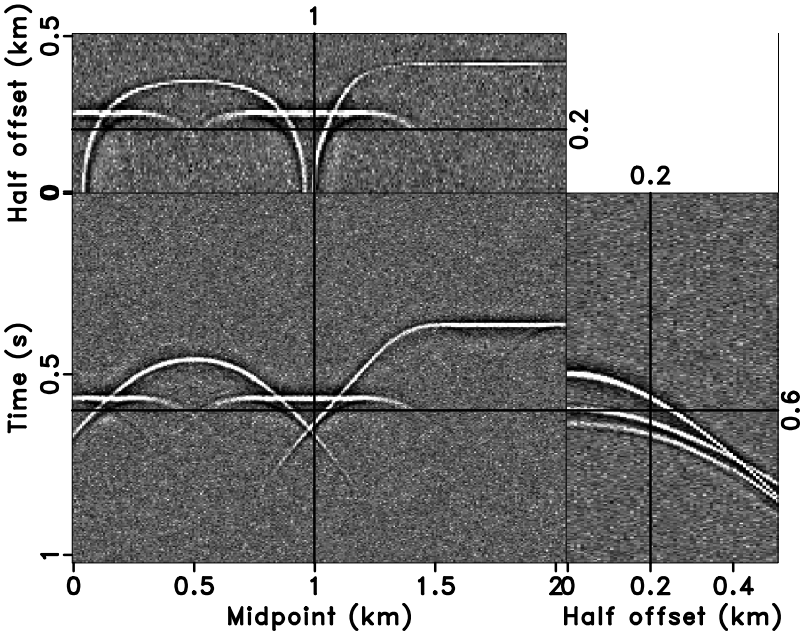


Figure 6: 2-D noisy data in t-x-offset domain. oc3/french noise

For a data regularization test, we remove 80% of randomly selected traces (Figure 11a) from the ideal data (Figures 2a). The complex dip information makes it extremely difficult to interpolate the data in individual common-offset gathers. The dataset is also non-stationary in the offset direction. Therefore, a simple offset interpolation scheme would also fail. Figure 11b shows the data after NMO correction, log-stretch transform, and double Fourier transforms. The missing traces introduce spatial artifacts in midpoint-wavenumber axis and discontinuities along the offset direction. After the OC-seislet transform, the reflection information can be predicted and compressed. Meanwhile, the artifacts spread over the whole transform domain (Figure 12a). The simple soft-thresholding algorithm (i.e., the iterative strategy with only one iteration) removes most of the artifacts, and the inverse OC-seislet transform reconstructs the major reflection information according to the offset-continuation prediction (Figure 12b).

One can also employ an iterative soft-thresholding strategy to implement miss-

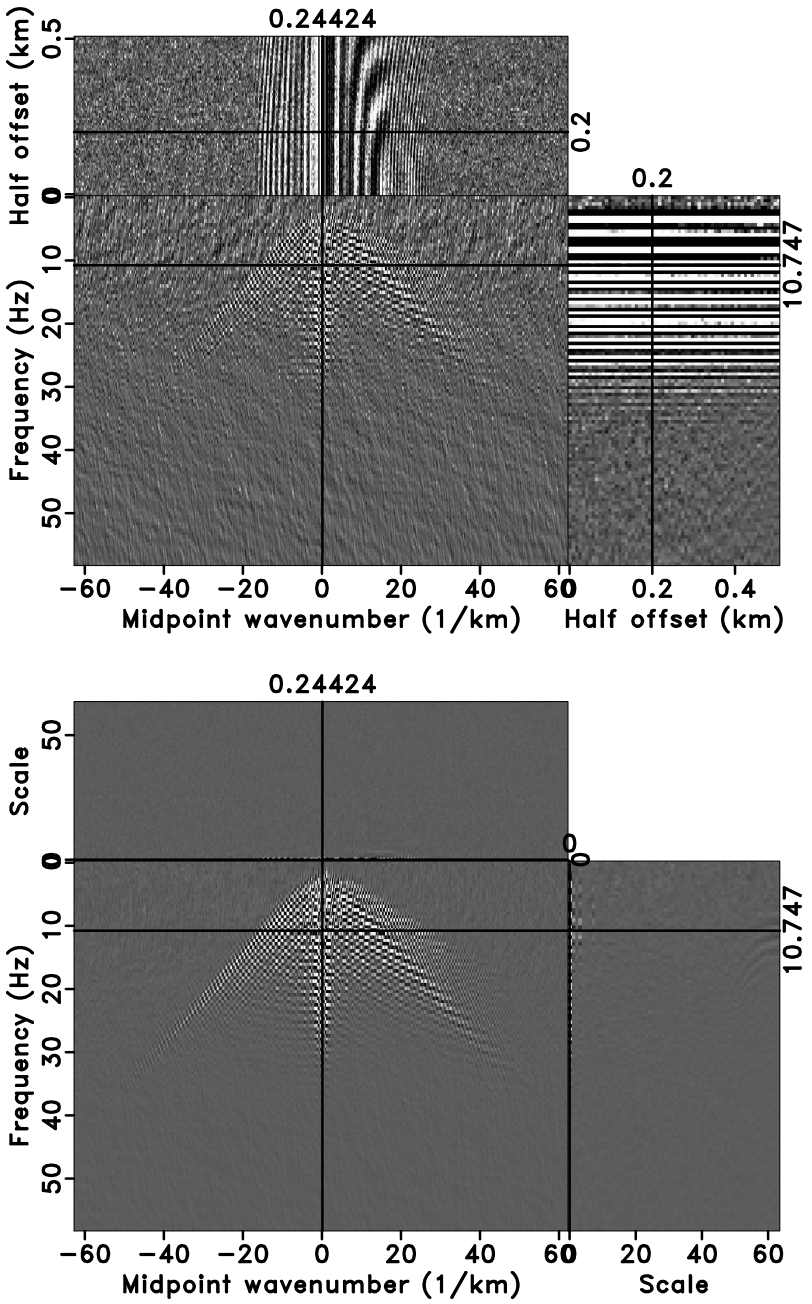


Figure 7: Noisy data in F-K-offset domain (a) and OC-seislet coefficients in F-K-scale domain (b). oc3/french input,tran

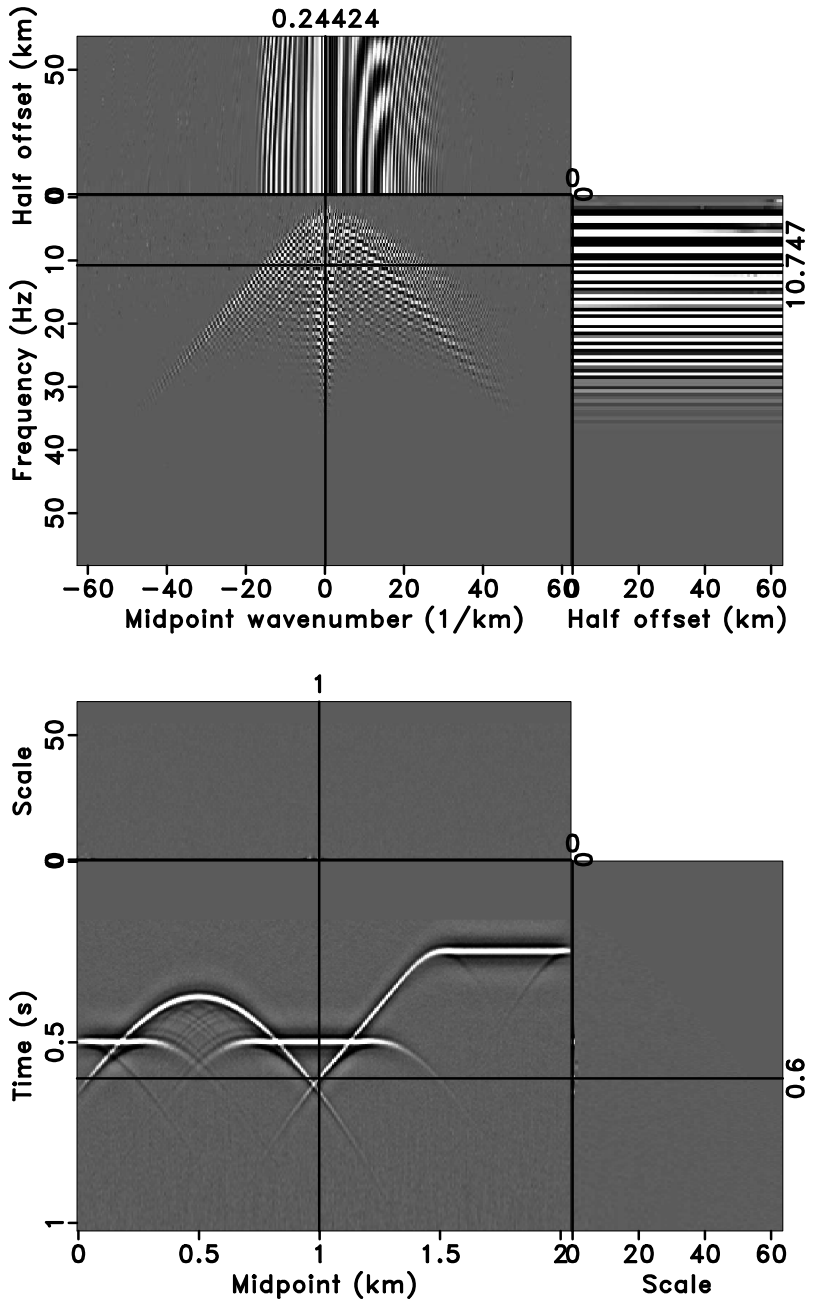


Figure 8: Thresholded data in F-K-offset domain (a) and OC-seislet coefficients in t-x-scale domain (b). oc3/french ithr,inv-tran

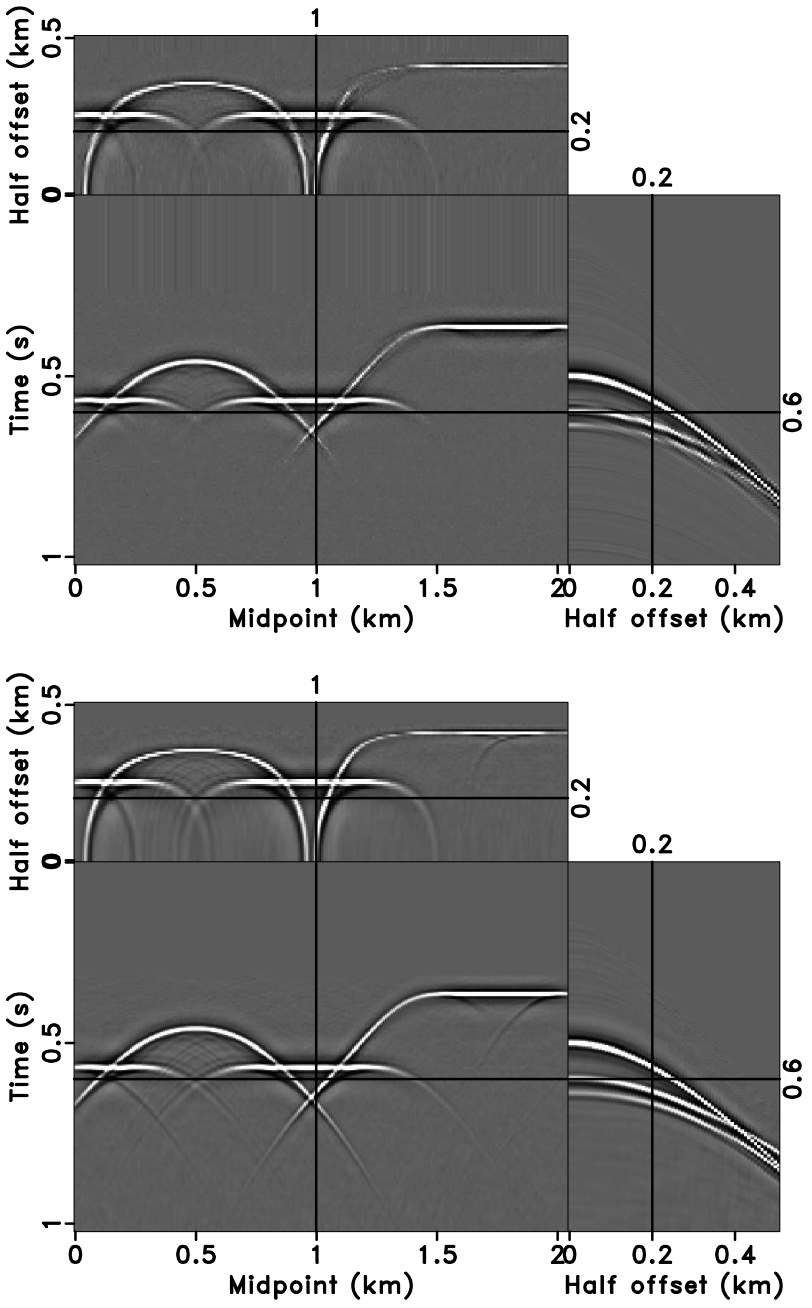


Figure 9: Denoised result by different methods. PWD-seislet transform (a) and OC-seislet transform (b). (Compare with Figure 2.) oc3/french nidwt,inver

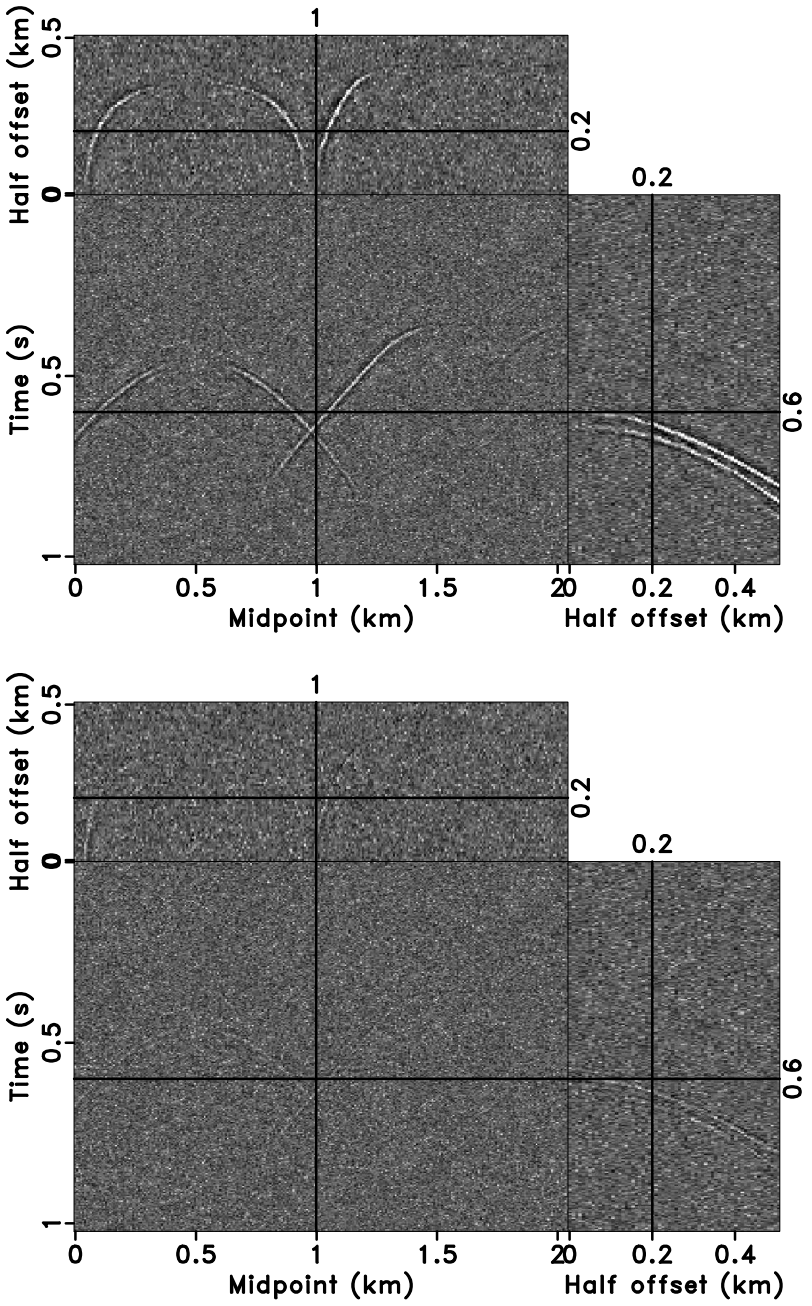


Figure 10: Difference between noisy data (Figure 6) and denoised results with different methods (Figure 9). PWD-seislet transform (a) and OC-seislet transform (b). oc3/french diff1,diff2

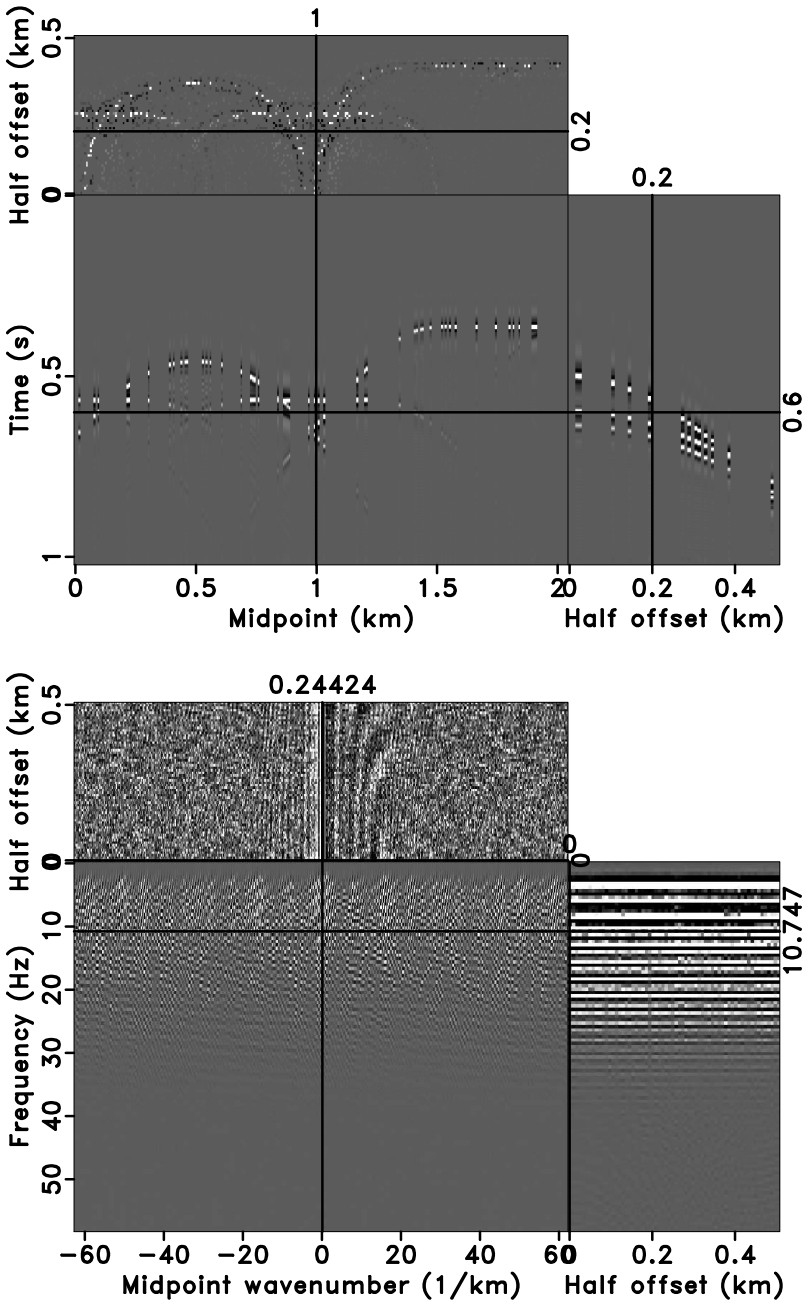


Figure 11: Synthetic data with 80% traces removed (a) and missing data in F-K-offset domain (b). oc3/frenchint czero,cinput

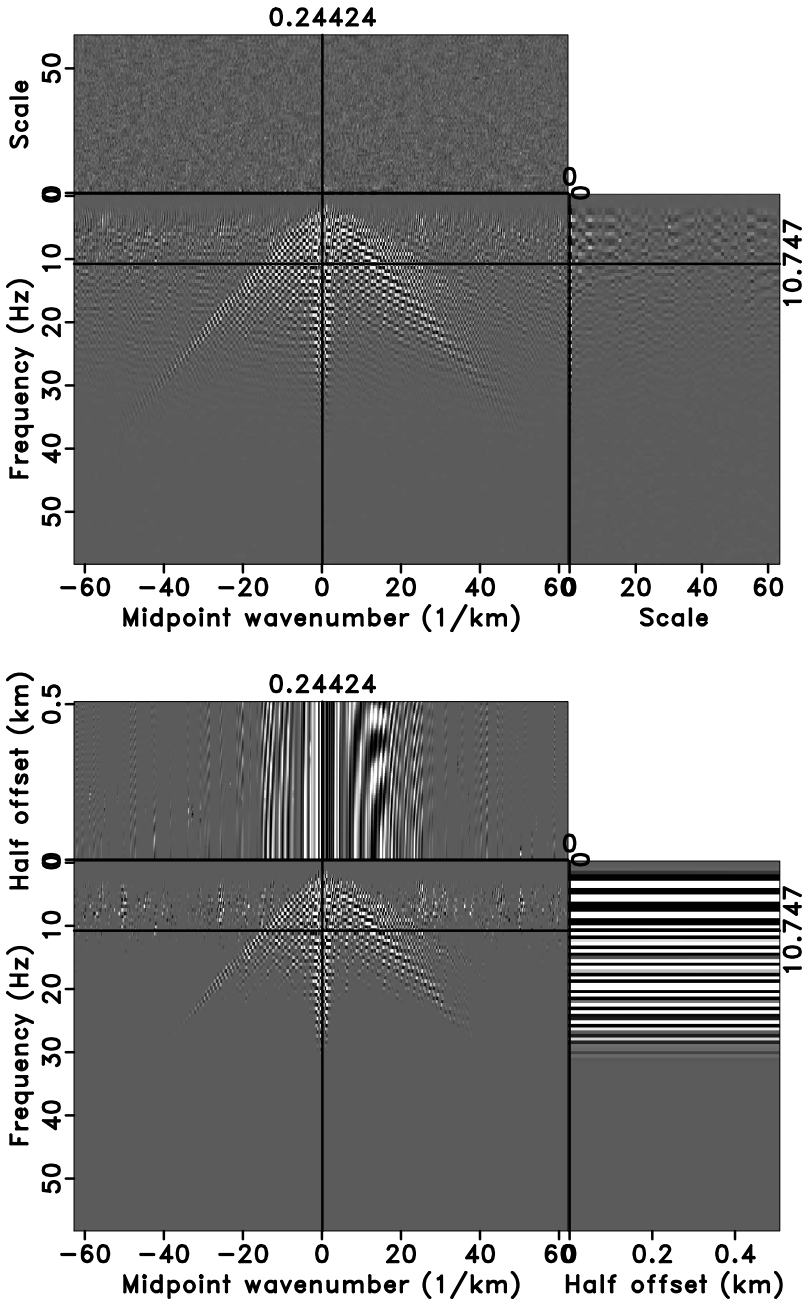


Figure 12: OC-seislet coefficients in F-K-scale domain (a) and data of iteration 1 after soft-thresholding in F-K-offset domain (b). oc3/frenchint ctran,cithr

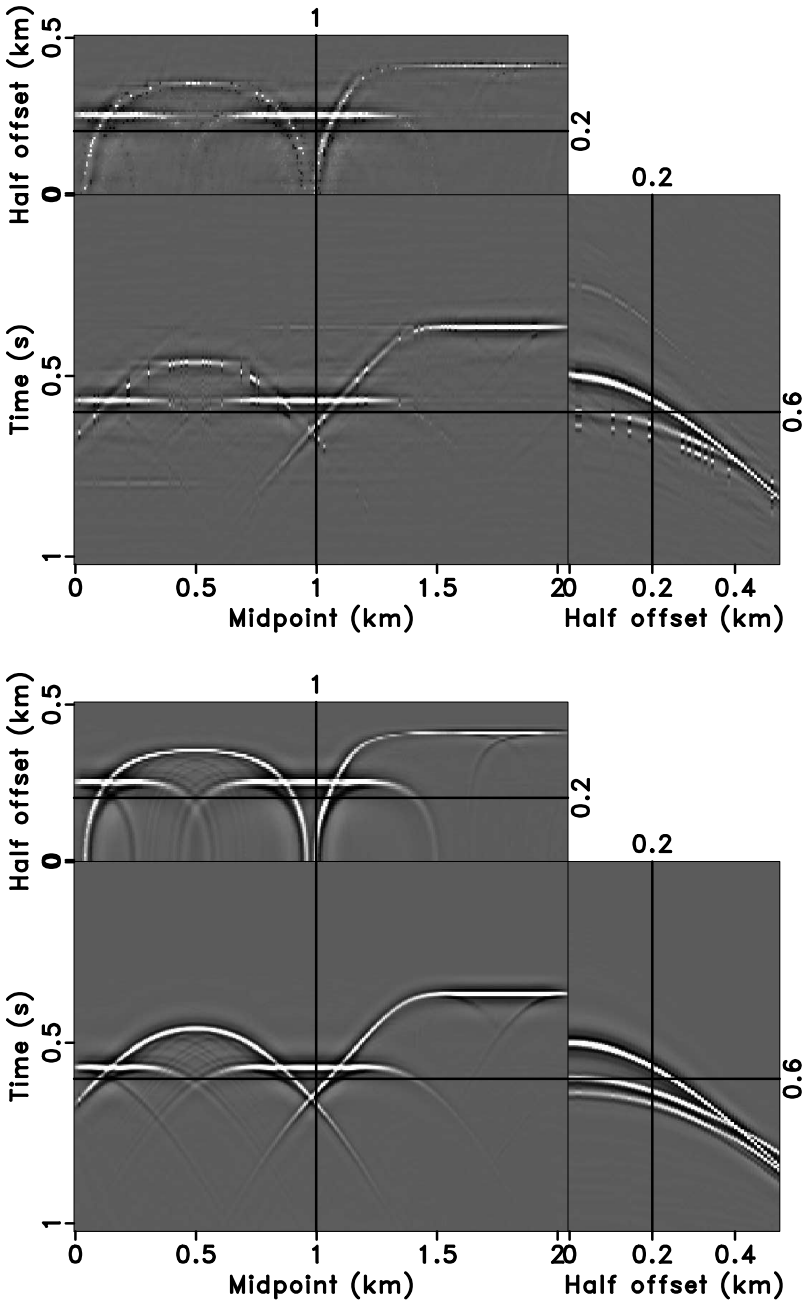


Figure 13: Interpolated results using iterative thresholding with different sparse transforms. 3-D Fourier transform (a) and OC-seislet transform (b). (Compare with Figure 2) oc3/frenchint four3pocs,cpocs

ing data interpolation. This method recovers missing traces as long as seismic data are sparse enough in the transform domain. To demonstrate the superior sparseness of the OC-seislet coefficients, we compare the proposed method with the 3-D Fourier transform. Figure 13a displays the interpolated result after a 3-D Fourier interpolation using iterative thresholding (1). The Fourier transform cannot provide enough sparseness of coefficients for complex reflections and, therefore, fails in recovering all missing traces. The OC-seislet transform is based on a physical prediction, and provides a much sparser domain for both reflections and diffractions. Iterative thresholding succeeds in interpolating the missing traces (Figure 13b).

FIELD DATA TESTS

We use a historic marine dataset from the Gulf of Mexico (8) to evaluate the proposed method for both noise attenuation and missing data interpolation problems. The input data are shown in Figure 14a. Near- and far- offset information is completely missing. After log-stretched NMO and Fourier transforms along time and midpoint axes convert the original data from time-midpoint-offset domain to frequency-wavenumber-offset domain, the OC-seislet transform compresses the predictable information according to the physical connection between different offsets. Unpredictable artifacts and reflections display different tendencies: artifacts disperse throughout the whole transform domain while reflections are compressed to a small scale range (Figure 14b). If we choose the significant coefficients using soft-thresholding, the inverse processing steps (including the inverse OC-seislet transform, the inverse FFTs both in time and midpoint axes, and the inverse log-stretch NMO) effectively remove incoherent artifacts from the data (Figure 15a). The difference between input data (Figure 14a) and denoised result (Figure 15a) is shown in Figure 15b. Most coherent events including reflections and diffractions are well protected by the proposed method thanks to the ability of the offset-continuation operator to predict structurally complex wavefields. However, some energy of the events is partly missing. To achieve a better result, one may need to use a more accurate NMO velocity. Thus, denoising is a naturally defined operation in the OC-seislet domain. The soft-thresholding techniques eliminate only part of spatial aliasing because of sparse offset sampling. Additional offset interpolation can further remove spatial aliasing.

For a data regularization test, we removed 70% of randomly selected traces from the input data (Figure 16). After applying the OC-seislet transform and iterative soft-thresholding, spatial artifacts and offset discontinuity in the transform domain coming from missing traces get attenuated. The interpolated result is shown in Figure 17b. Missing traces have been interpolated well even where diffractions are present. For comparison, we applied 3-D Fourier transform with iterative soft-thresholding, as proposed by (1). The same parameters of soft-thresholding for those in the OC-seislet transform were used. The final result is shown in Figure 17a. The Fourier transform fails to recover the missing information as accurately as the OC-seislet transform. Fundamentally, the Fourier coefficients are simply not sparse enough for

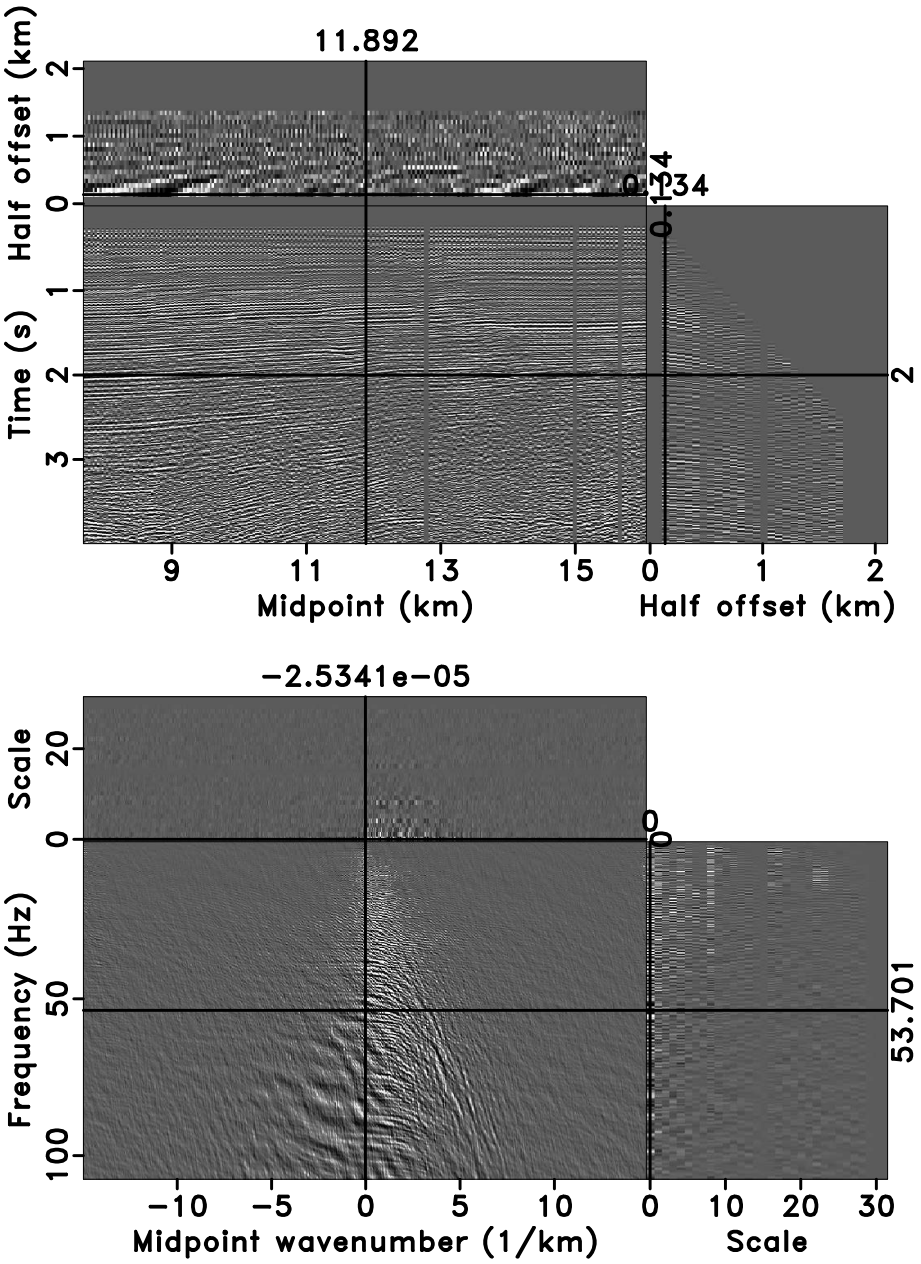


Figure 14: Marine data in t-x-offset domain (a) and OC-seislet coefficients in F-K-scale domain (b). oc3/bei bei,bei-tran

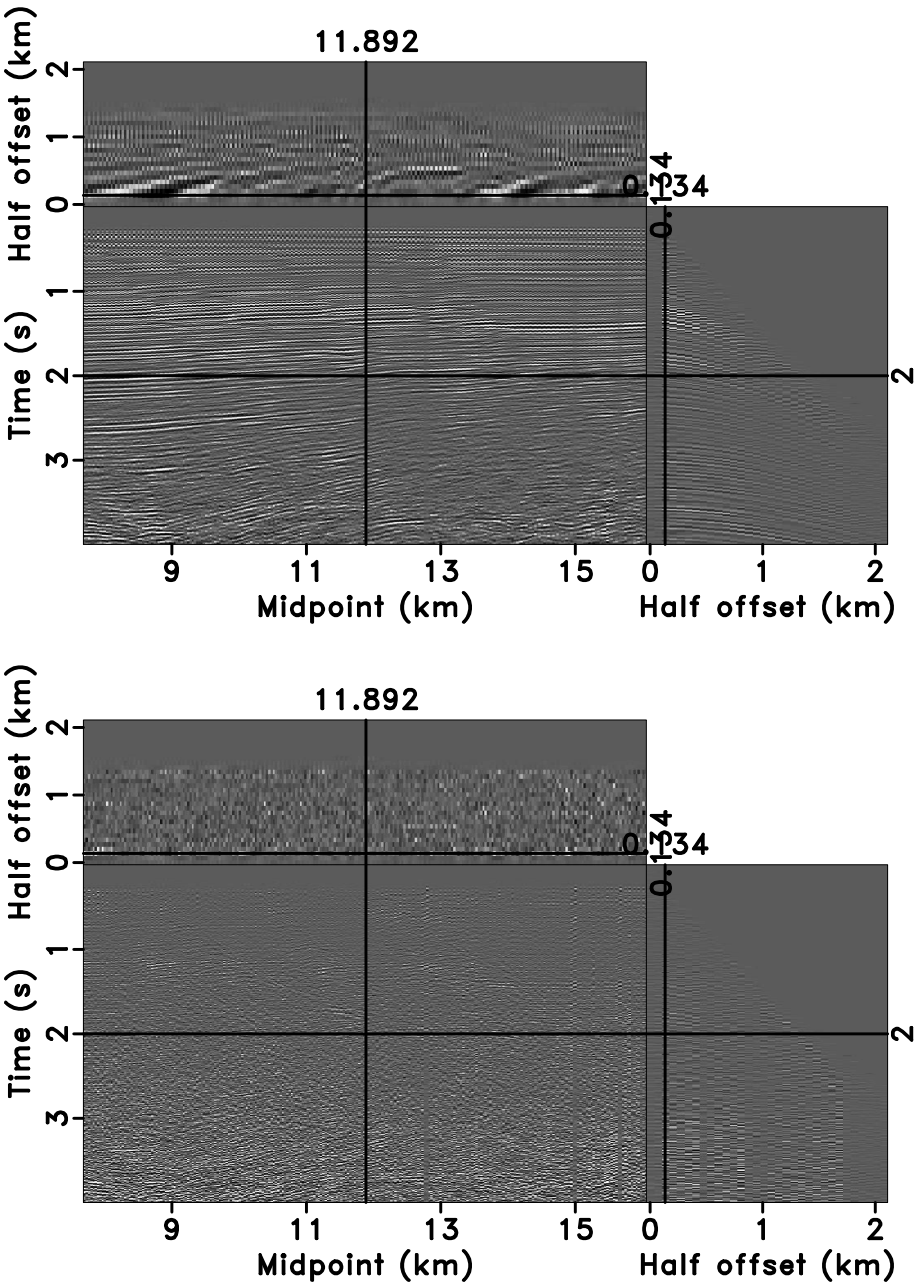


Figure 15: Denoised result in t-x-offset domain (a) and difference between input data (Figure 14a) and denoised result (Figure 15a) (b). oc3/bei bei-denoise,bei-diff

this dataset. Figure 18 shows the comparison after DMO stacking. The result from 3-D Fourier transform displays obvious gaps which are caused by inperfect interpolation (Figure 18a), while the OC-seislet transform and iterative soft-thresholding strategy provide a reasonably accurate result (Figure 18b).

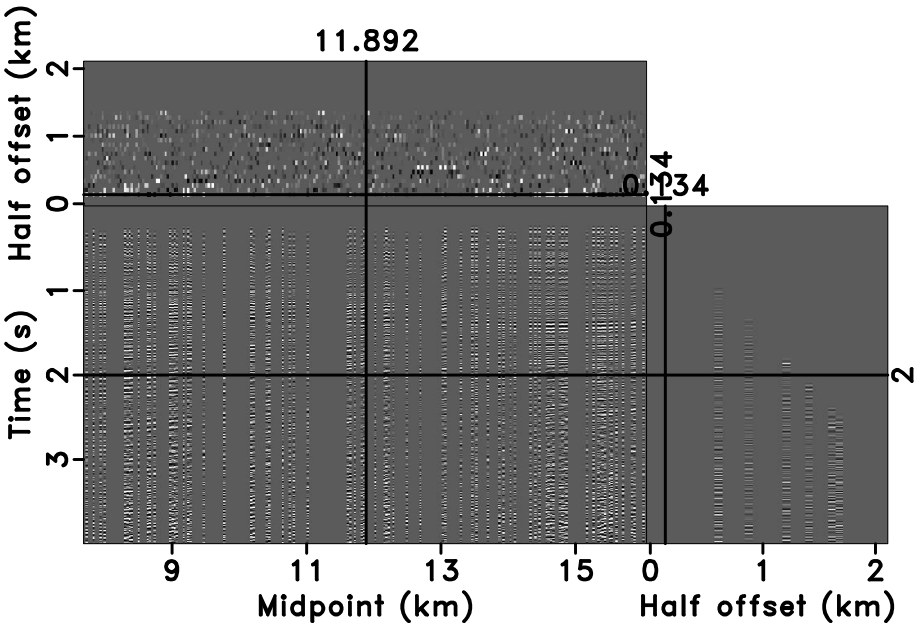


Figure 16: Marine data with 70% traces removed. oc3/fault bei-zero

CONCLUSIONS

We have introduced the OC-seislet transform, a new domain for analyzing prestack reflection data in offset, midpoint, and time coordinates. Thanks to its physical basis, the new transform is able to characterize and compress structurally complex seismic data better than either the 3-D Fourier transform or the PWD-seislet transform. The offset-continuation operator serves as a transformational bridge between different offsets, which allows the OC-seislet transform to compress all predictable reflections and diffractions into small scales. In the new transform domain, random noise spreads over different scales while the predictable reflection seismic events get compressed to large coefficients at small scales. Therefore, a thresholding approach is successful in separating seismic signal and noise. Reconstructing missing seismic data is a more difficult task, where iterative thresholding is required to help the proposed OC-seislet transform recover the missing seismic traces. Tests using both synthetic and field data demonstrate that the proposed transform can succeed in handling structurally complex data even in the presence of strong random noise.

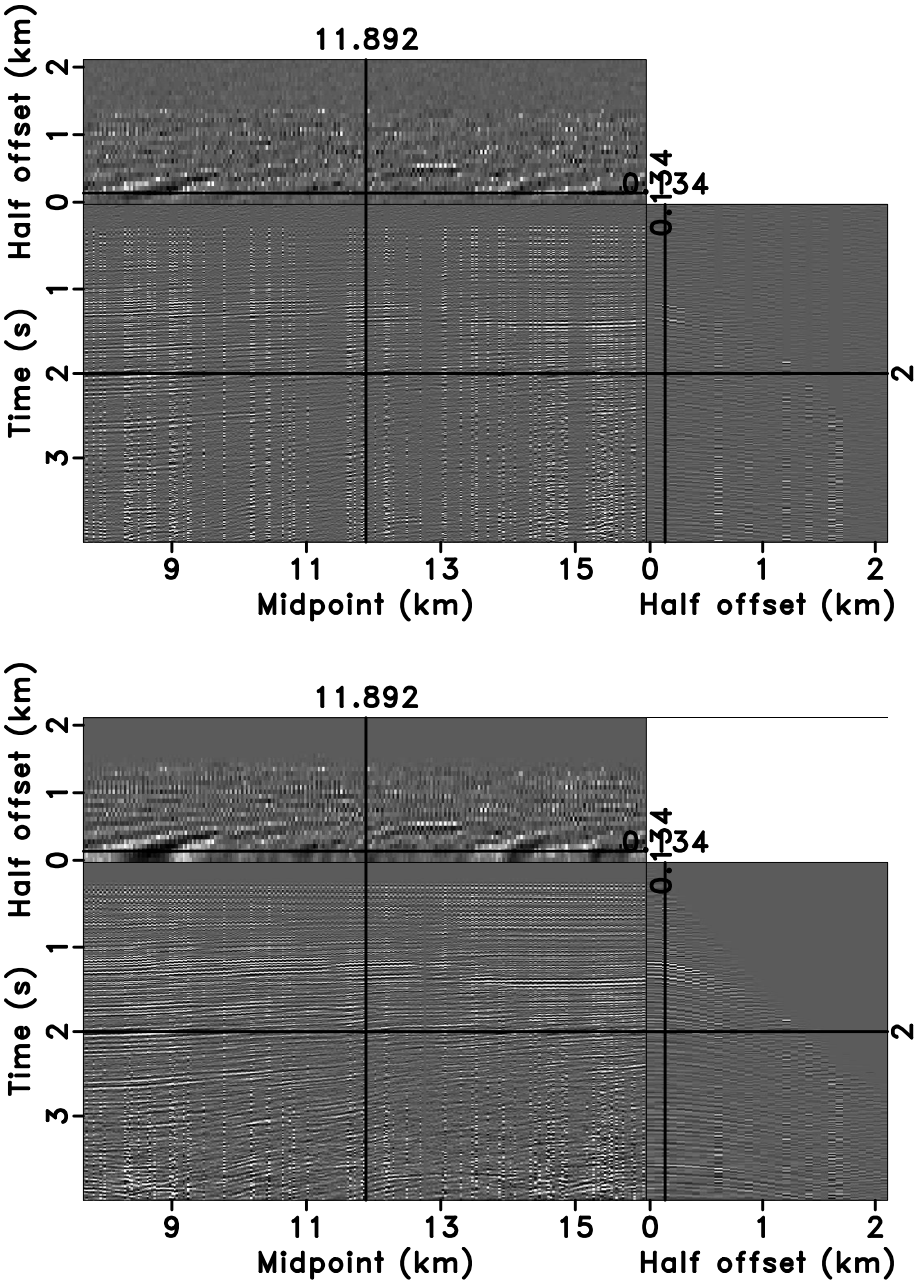


Figure 17: Interpolated results using iterative thresholding with different transforms. 3-D Fourier transform (a) and OC-seislet transform (b). oc3/fault bfour3pocs,bei-pocs

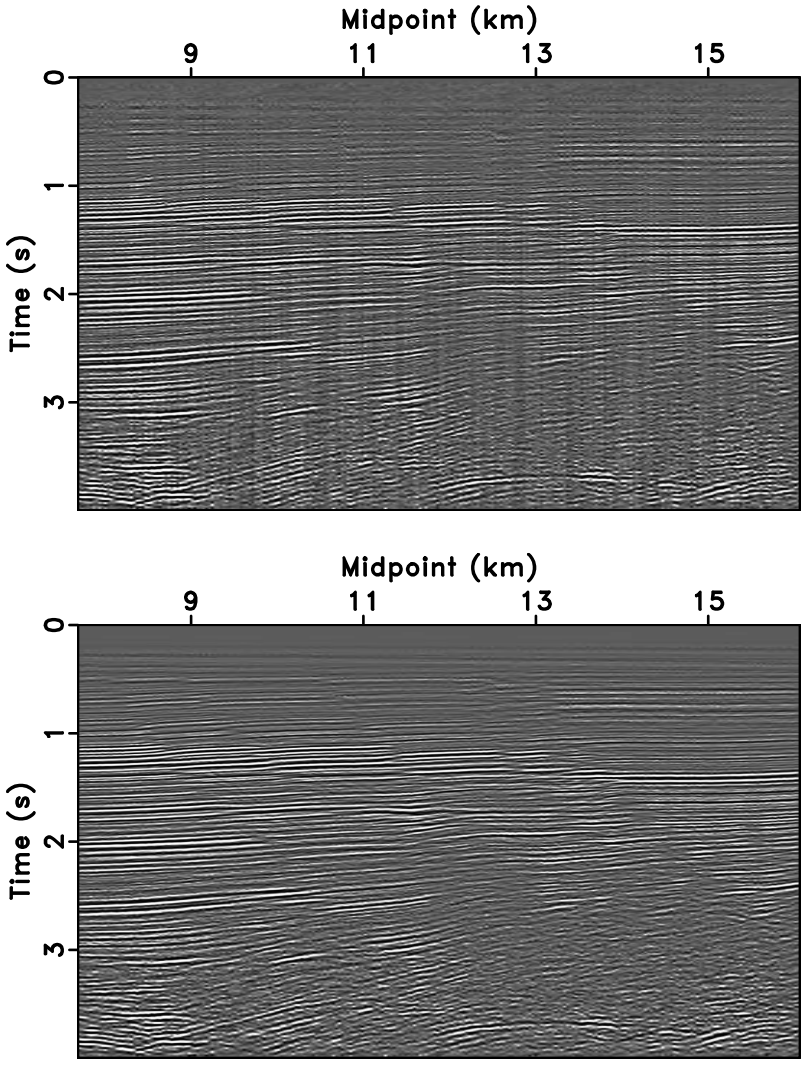


Figure 18: Stacking of interpolated results corresponding to Figure 17. 3-D Fourier transform (a) and OC-seislet transform (b). oc3/fault fdmos,dmos

ACKNOWLEDGMENTS

We thank BGP Americas for a partial financial support of this work. We thank Tamas Nemeth, Mauricio Sacchi, Sandra Tegtmeier-Last, and two anonymous reviewers for their constructive comments and suggestions. This publication was authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

REVIEW OF DIFFERENTIAL OFFSET CONTINUATION

In this appendix, we review the theory of differential offset continuation from (16; 18). The partial differential equation for offset continuation (differential azimuth moveout) takes the form

$$\left(\mathbf{h}^T \mathbf{P}_{xx}\, \mathbf{h} - h^2\, \frac{\partial^2 P}{\partial h^2}\right) \;=\; h\, t_n\, \frac{\partial^2 P}{\partial t_n\, \partial h}\;, \tag{A-1}$$

where $P(\mathbf{x}, \mathbf{h}, t_n)$ is the seismic data in the midpoint-offset-time domain, t_n is the time coordinate after the normal moveout (NMO) correction, $\mathbf{h}^T$ denotes the transpose of $\mathbf{h}$, and $\mathbf{P}_{xx}$ is the tensor of the second-order midpoint derivatives.

A particularly efficient implementation of offset continuation results from a log-stretch transform of the time coordinate (6), followed by a Fourier transform of the stretched time axis. After these transforms, the offset-continuation equation takes the form

$$\left(\mathbf{h}^T \tilde{\mathbf{P}}_{xx}\, \mathbf{h} - h^2\, \frac{\partial^2 \tilde{P}}{\partial h^2}\right) \;=\; i\,\Omega\, h\, \frac{\partial \tilde{P}}{\partial h}\;, \tag{A-2}$$

where Ω is the dimensionless frequency corresponding to the stretched time coordinate and $\tilde{P}(\mathbf{x}, \mathbf{h}, \Omega)$ is the transformed data. As in other frequency-space methods, equation A-2 can be applied independently and in parallel on different frequency slices.

In the frequency-wavenumber domain, the extrapolation operator is defined by solving an initial-value problem for equation A-2. The analytical solution takes the form

$$\tilde{\tilde{P}}(\mathbf{k}, \mathbf{h}_2, \Omega) = \tilde{\tilde{P}}(\mathbf{k}, \mathbf{h}_1, \Omega)\, \frac{Z_\lambda(\mathbf{k}\cdot\mathbf{h}_2)}{Z_\lambda(\mathbf{k}\cdot\mathbf{h}_1)}\;, \tag{A-3}$$

where $\tilde{\tilde{P}}(\mathbf{k}, \mathbf{h}, \Omega)$ is the double-Fourier-transformed data, $\lambda = (1 + i\Omega)/2$, Z_λ is the special function defined as

$$Z_\lambda(x) = \Gamma(1-\lambda)\, \left(\frac{x}{2}\right)^\lambda\, J_{-\lambda}(x) = {}_0F_1\left(;1-\lambda;-\frac{x^2}{4}\right)\;, \tag{A-4}$$

Γ is the gamma function, $J_{-\lambda}$ is the Bessel function, and ${}_0F_1$ is the confluent hypergeometric limit function (30). The wavenumber $\mathbf{k}$ in equation A-3 corresponds

to the midpoint $\mathbf{x}$ in the original data domain. In high-frequency asymptotics, the offset-continuation operator takes the form

$$\tilde{\tilde{P}}(\mathbf{k},\mathbf{h}_2,\Omega) \approx \tilde{\tilde{P}}(\mathbf{k},\mathbf{h}_1,\Omega)\, \frac{F(\frac{2\mathbf{k}\cdot\mathbf{h}_2}{\Omega})\, \exp\left[i\Omega\, \psi\left(\frac{2\mathbf{k}\cdot\mathbf{h}_2}{\Omega}\right)\right]}{F(\frac{2\mathbf{k}\cdot\mathbf{h}_1}{\Omega})\, \exp\left[i\Omega\, \psi\left(\frac{2\mathbf{k}\cdot\mathbf{h}_1}{\Omega}\right)\right]}\ , \tag{A-5}$$

where

$$F(\epsilon) = \sqrt{\frac{1+\sqrt{1+\epsilon^2}}{2\sqrt{1+\epsilon^2}}}\, \exp\left(\frac{1-\sqrt{1+\epsilon^2}}{2}\right)\ , \tag{A-6}$$

and

$$\psi(\epsilon) = \frac{1}{2}\left(1-\sqrt{1+\epsilon^2}+\ln\left(\frac{1+\sqrt{1+\epsilon^2}}{2}\right)\right)\ . \tag{A-7}$$

The phase function ψ defined in equation A-7 corresponds to the analogous term in the exact-log DMO and AMO (26; 38; 4).

REFERENCES

[1] Abma, R., and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, E91–E97.

[2] Bagaini, C., and U. Spagnolini, 1996, 2-D continuation operators and their applications: Geophysics, **61**, 1846–1858.

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

[4] Biondi, B., and I. Vlad, 2002, Amplitude preserving prestack imaging of irregularly sampled 3-D data: 72nd Annual International Meeting, SEG, Expanded Abstracts, 2170–2173.

[5] Bleistein, N., and H. Jaramillo, 2000, A platform for kirchhoff data mapping in scalar models of data acquisition: Geophysical Prospecting, **48**, 135–161.

[6] Bolondi, G., E. Loinger, and F. Rocca, 1982, Offset continuation of seismic sections: Geophysical Prospecting, **30**, 813–828.

[7] Chauris, H., and T. Nguyen, 2008, Seismic demigration/migration in the curvelet domain: Geophysics, **73**, S35–S46.

[8] Claerbout, J. F., 2000, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

[9] Cohen, A., I. Daubechies, and J. Feauveau, 1992, Biorthogonal bases of compactly supported wavelets: Communications on Pure and Applied Mathematics, **45**, 485–560.

[10] Daubechies, I., and W. Sweldens, 1998, Factoring wavelet transforms into lifting steps: The Journal of Fourier Analysis and Application, **4**, 247–269.

[11] Deregowski, S. M., and F. Rocca, 1981, Geometrical optics and wave theory of constant offset sections in layered median: Geophysical Prospecting, **29**, 374–406.

[12] Do, M. N., and M. Vetterli, 2005, The contourlet transform: an efficient directional multiresolution image representation: IEEE Trans. Image Process., **14**, 2091–2106.

[13] Douma, H., and M. V. de Hoop, 2007, Leading-order seismic imaging using curvelets: Geophysics, **72**, S231–S248.
[14] Fomel, S., 2001, Three-dimensinal seismic data regularization: PhD thesis, Ph.D. thesis, Stanford University.
[15] ———, 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.
[16] ———, 2003a, Differential azimuth moveout: 73th Annual International Meeting, SEG, Expanded Abstracts, 2068–2071a.
[17] ———, 2003b, Seismic reflection data interpolation with differential offset and shot continuation: Geophysics, **68**, 733–744.
[18] ———, 2003c, Theory of differential offset continuation: Geophysics, **68**, 718–732.
[19] ———, 2006, Towards the seislet transform: 76th Annual International Meeting, SEG, Expanded Abstracts, 2847–2851a.
[20] Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, V25–V38.
[21] French, W. S., 1974, Two-dimensional and three-dimensional migration of model-experiment reflection profiles: Geophysics, **39**, 265–277.
[22] Hale, D., 1991, Dip moveout processing: SEG, Course Notes Series 4.
[23] Herrmann, F., D. Wang, G. Hennenfent, and P. Moghaddam, 2008, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, A1–A5.
[24] Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.
[25] Lian, C., K. Chen, H. Chen, and L. Chen, 2001, Lifting based discrete wavelet transform architecture for JPEG2000: The 2001 IEEE International Symposium on Circuits and Systems, IEEE, II445–II448.
[26] Liner, C., 1990, General theory and comparative anatomy of dip moveout: Geophysics, **55**, 595–607.
[27] Liu, B., and M. D. Sacchi, 2004, Minimum weighted norm interpolation of seismic records: Geophysics, **69**, 1560–1568.
[28] Mallat, S., 2009, A wavelet tour of signal processing, the sparse way: Academic Press, Third Edition.
[29] Pennec, E. L., and S. Mallat, 2005, Sparse geometrical image representation with bandelets: IEEE Trans. Image Process., **14**, 423–438.
[30] Petkovsek, M., H. S. Wilf, and D. Zeilberger, 1996, $A = B$: A K Peters Ltd.
[31] Salvador, L., and S. Savelli, 1982, Offset continuation for seismic stacking: Geophysical Prospecting, **30**, 829–849.
[32] Spagnolini, U., and S. Opreni, 1996, 3-D shot continuation operator: 66th Annual International Meeting, SEG, Expanded Abstracts, 439–442.
[33] Starck, J. L., E. J. Candés, and D. L. Donoho, 2000, The curvelet transform for image denoising: IEEE Trans. Image Process., **11**, 670–684.
[34] Sweldens, W., 1995, The lifting scheme: A new philosophy in biorthogonal wavelet constructions: Wavelet Applications in Signal and Image Processing III, Proc. SPIE 2569, 68–79.
[35] Sweldens, W., and P. Schröder, 1996, Building your own wavelets at home, *in*

Wavelets in Computer Graphics: ACM SIGGRAPH Course Notes, 15–87.
[36] Velisavljevic, V., 2005, Directionlets: anisotropic multi-directional representation with separable filtering: PhD thesis, Ecole Polytechnique Fédérale de Lausanne.
[37] Xu, S., Y. Zhang, D. Pham, and G. Lambaré, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, V87–V95.
[38] Zhou, B., I. M. Mason, and S. A. Greenhalgh, 1996, An accurate formulation of log-stretch dip moveout in the frequency-wavenumber domain: Geophysics, **61**, 815–820.
[39] Zwartjes, P., and A. Gisolf, 2007, Fourier reconstruction with sparse inversion: Geophysical Prospecting, **55**, 199–221.

Stratigraphic coordinates, a coordinate system tailored to seismic interpretation

Parvaneh Karimi and Sergey Fomel

ABSTRACT

In certain seismic data processing and interpretation tasks, such as spiking deconvolution, tuning analysis, impedance inversion, spectral decomposition, etc., it is commonly assumed that the vertical direction is normal to reflectors. This assumption is false in the case of dipping layers and may therefore lead to inaccurate results. To overcome this limitation, we propose a coordinate system in which geometry follows the shape of each reflector and the vertical direction corresponds to normal reflectivity. We call this coordinate system *stratigraphic coordinates*. We develop a constructive algorithm that transfers seismic images into the stratigraphic coordinate system. The algorithm consists of two steps. First, local slopes of seismic events are estimated by plane-wave destruction; then structural information is spread along the estimated local slopes, and horizons are picked everywhere in the seismic volume by the predictive-painting algorithm. These picked horizons represent level sets of the first axis of the stratigraphic coordinate system. Next, an upwind finite-difference scheme is used to find the two other axes, which are perpendicular to the first axis, by solving the appropriate gradient equations. After seismic data are transformed into stratigraphic coordinates, seismic horizons should appear flat, and seismic traces should represent the direction normal to the reflectors. Immediate applications of the stratigraphic coordinate system are in seismic image flattening and spectral decomposition. Synthetic and real data examples demonstrate the effectiveness of stratigraphic coordinates.

INTRODUCTION

In certain seismic data processing and interpretation tasks, such as spiking deconvolution, tuning analysis, impedance inversion, spectral decomposition, etc., it is commonly assumed that the vertical direction is normal to reflectors. This assumption does not hold true in the case of dipping layers and may therefore lead to inaccurate results (7). (13) defined a mathematical framework, called GeoChron, for transforming the geologic space into a new space in which all horizons appear flat, and faults, if any, disappear. In this paper, we propose the *stratigraphic coordinate* system, in which geometry follows the shape of each reflector and the vertical direction corresponds to normal reflectivity .

Flattening post-stack seismic data is an immediate use of the proposed coordinate system. Flattened seismic images facilitate the interpreter's ability to extract detailed stratigraphic information from the seismic data. In interpretational applications, several different algorithms for image flattening have been developed by different authors. The idea of seismic image flattening by extracting stratal slices was introduced by (18). Automatic picking of horizons using local shifts was studied by (1) and (16). (11) and (14) presented inversion methods in which horizons are calculated on the basis of local slopes and are then used to flatten seismic events. (5) proposed the method of predictive painting that uses the prediction operators extracted by plane-wave destruction to spread information inside the seismic volume recursively.

Conventionally, seismic image flattening is performed by shifting samples in the original image up or down - in other words, differentially stretching and squeezing the original image in order to flatten the reflection events. (12) proposed a method for image flattening that uses the vector shift field instead of the scalar field of vertical shifts to define deformations in the image. Flattening by vector shift uses either vertical shear or rotation or a combination of the two, depending on the type of geologic deformation.

The stratigraphic coordinate system, introduced in this paper, represents a new framework for seismic interpretation and processing. To construct stratigraphic coordinates, we combine predictive painting with an upwind finite-difference scheme (6) for solving relevant gradient equations. The stratigraphic coordinate system is semi-orthogonal; i.e., picked horizons that are level sets of the first axis are orthogonal to the other two axes. In other words, stratigraphic coordinates are aligned with horizons, and the vertical direction in stratigraphic coordinates corresponds to the direction normal to the major reflection boundaries. Application of the stratigraphic coordinate system is not limited to seismic image flattening and may be extended to many data processing and interpretation tasks in which the vertical direction is commonly assumed to be normal to reflection boundaries: a crude assumption in all structures but flat geology. In the following sections, we start by describing a constructive algorithm for generating stratigraphic coordinates. We then illustrate applications of the stratigraphic coordinate system to seismic image flattening and spectral decomposition using synthetic and field data examples.

THEORY

In order to define the first step for transformation to stratigraphic coordinates, we follow the predictive-painting algorithm (5), which is reviewed in appendix A. Predictive painting spreads the time values along a reference trace into the seismic volume to output the *relative geologic age* attribute ($Z_0(x, y, z)$). The painted horizons output by predictive painting are used as the first axis of our stratigraphic coordinate system. Several alternative methods exist to track horizons automatically in a seismic volume and produce horizon cubes (8; 17). We choose predictive painting because of its simplicity and efficiency.

In the next step, we find the two other axes, $X_0(x,y,z)$ and $Y_0(x,y,z)$, orthogonal to the first axis, $Z_0(x,y,z)$, by numerically solving the following gradient equations:

$$\nabla Z_0 \cdot \nabla X_0 = 0 \tag{1}$$

and

$$\nabla Z_0 \cdot \nabla Y_0 = 0. \tag{2}$$

Equations 1 and 2 simply state that the X_0 and Y_0 axes should be perpendicular to Z_0. We can define the boundary condition for the first gradient equation (equation 1) as

$$X_0(x,y,0) = x \tag{3}$$

and the boundary condition for equation 2 as

$$Y_0(x,y,0) = y. \tag{4}$$

These two boundary conditions mean that the stratigraphic coordinate system and the regular coordinate system (x,y,z) become equivalent at the surface $(z=0)$.

The stratigraphic coordinates are originally designed for depth images. When applied to time-domain images, the definition of the gradient operator becomes

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\frac{\partial z}{\partial t}\right), \tag{5}$$

so a scaling factor with dimensions of velocity-squared is needed in equations 1 and 2.

Algorithm

In summary, our algorithm for transferring seismic images from the regular Cartesian coordinate system into the stratigraphic coordinate system consists of the following steps:

1. Extract the first axis of the stratigraphic coordinate system $Z_0(x,y,z)$ from seismic image $P(x,y,z)$ by predictive painting;

2. Start with X_0 at $(z=0)$ as an initial value and solve equation 1 with boundary condition 3 numerically for X_0;

3. Start with Y_0 at $(z=0)$ as an initial value and solve equation 2 with boundary condition 4 numerically for Y_0.

We solve equations 1 and 2 numerically with an explicit upwind finite-difference scheme (6; 9).

EXAMPLES OF STRATIGRAPHIC COORDINATES

For a simple illustration of stratigraphic coordinates, we use a 2D synthetic seismic image from (3), which contains layers with sinusoidal dip variations, faulted and truncated by an unconformity, and dipping beds with constant slope above the unconformity (Figure 2(a)). Figure 1(b) shows local slopes measured by plane-wave destruction and delineates the slope field variation. Figure 2(a) shows automatically picked horizons obtained by the predictive-painting algorithm. After solving the gradient equation 1, we acquire the other axis of the stratigraphic coordinate system (Figure 2(b)). Figure 3(a) shows the stratigraphic coordinates grid overlain on the regular Cartesian coordinates of the seismic image. Complex tectonic deformations expressed in faults and folds cannot in general be undone by a simple vertical stretch and squeeze operator. In contrast, the transformation to stratigraphic coordinates allows for complex displacements, which can better capture and thus undo non-trivial tectonic deformations. Arrows in Figure 3(b) show the amount and direction of shift that it takes for different samples to be transformed from their original position in the seismic image to their corresponding positions in the flattened image through the stratigraphic coordinates algorithm. For comparison, arrows in Figure 3(c) represent how different samples shift under conventional flattening methods. The seismic image gets flattened when the data from the regular coordinate system are transferred to stratigraphic coordinates. The result is shown in Figure 4(a). Apart from the structural (fault) and stratigraphic (erosional truncation) discontinuities, the input image is successfully flattened. Figure 1(a) shows that by returning from stratigraphic coordinates to regular coordinates, one can recostruct the features of the original image effectively.

Figure 5(a) shows the input image for a field-data test reproduced from (11) and (5). The input is a depth-migrated 3-D image with structural folding and angular unconformities. The three axes of the stratigraphic coordinates are shown in Figure 6. Figure 7(a) displays the image in the regular Cartesian coordinates overlain by its stratigraphic coordinates grid. The flattened image in the stratigraphic coordinate system, shown in Figure 7(b), can be transferred back to the regular Cartesian coordinates to reconstruct the original image (Figure 7(b)).

APPLICATION OF STRATIGRAPHIC COORDINATES TO SPECTRAL DECOMPOSITION

Improving the accuracy of spectral decomposition is one of the possible applications of the stratigraphic coordinate system. Spectral decomposition is a window-based analysis to characterize the reflecting wavelet of an interpretation target and refers to any method that produces a continuous time-frequency analysis of a seismic trace (15). According to the convolutional model, seismic traces are considered as normal-incidence 1D seismograms, which is true in the case of horizontal layers and allows for capturing the signal wavelet while performing spectral decomposition along the seis-

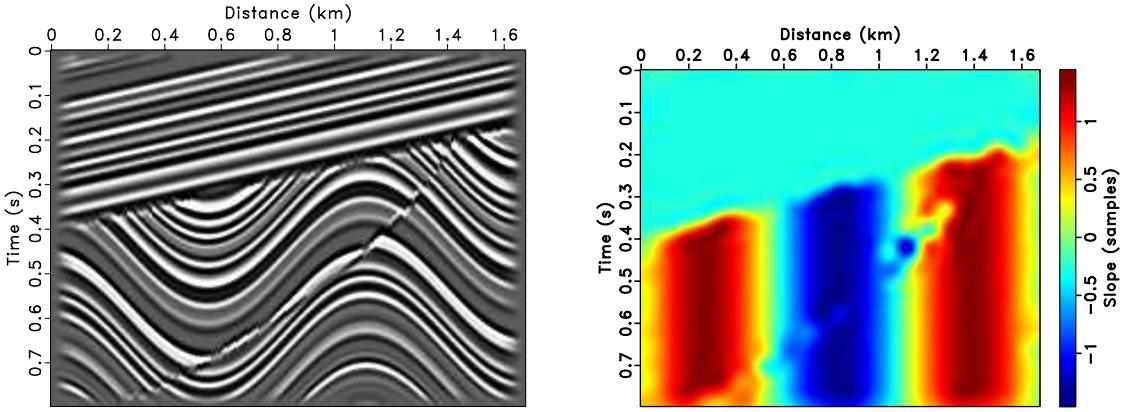


Figure 1: (a) Synthetic seismic image from (3). (b) Local slopes measured by plane-wave destruction. strat/sigmoid1 sigmoid,sdip

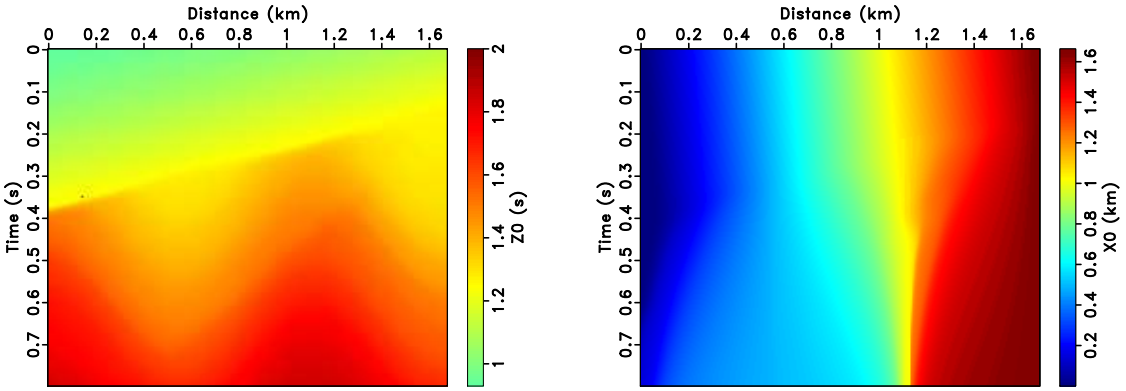


Figure 2: (a) First axis (Z_0) of stratigraphic coordinates in the regular coordinates obtained by predictive painting. (b) Second axis (X_0) of the stratigraphic coordinate system acquired by solving the gradient equation (equation 1). strat/sigmoid1 t0sig,x0sig

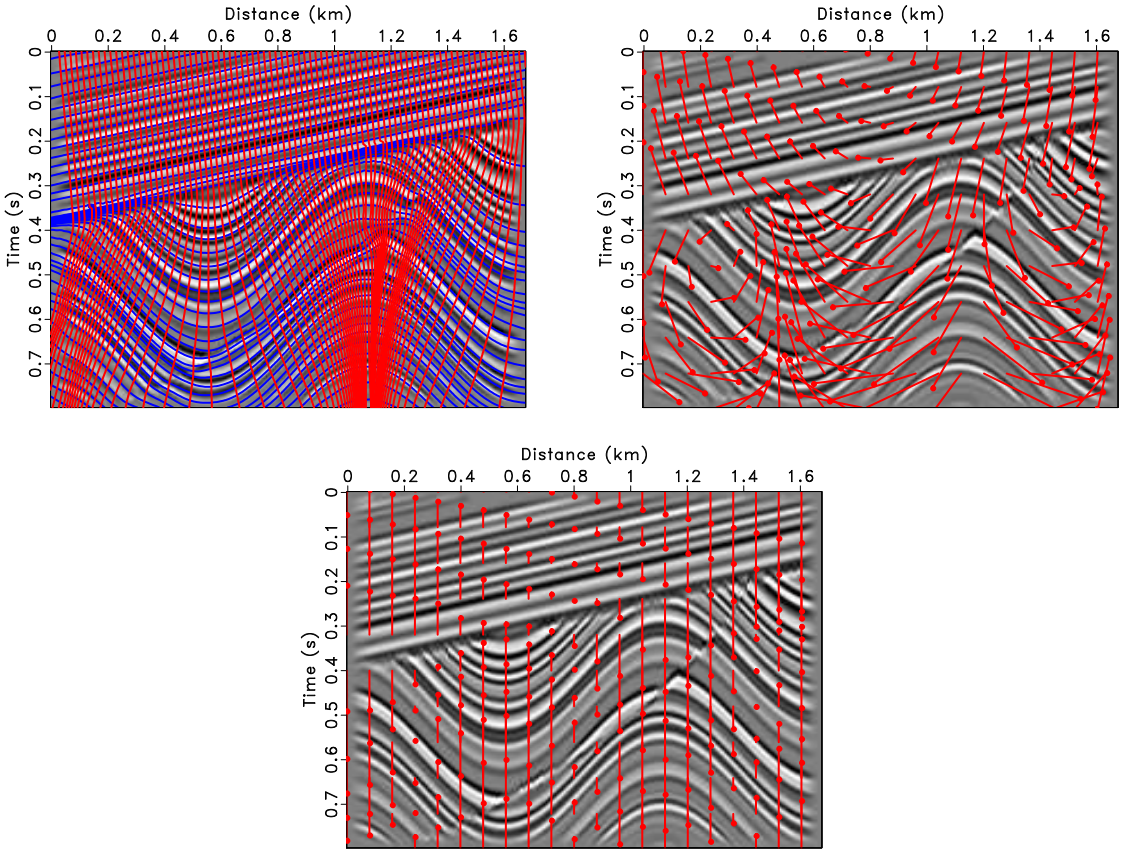


Figure 3: (a) Two axes of the stratigraphic coordinate system relative to the regular coordinate system. The synthetic image with corresponding shift represented by red arrows through using stratigraphic coordinates (b) and conventional flattening methods (c). strat/sigmoid1 sigmoid-t0sigx0sig,shift,shift0

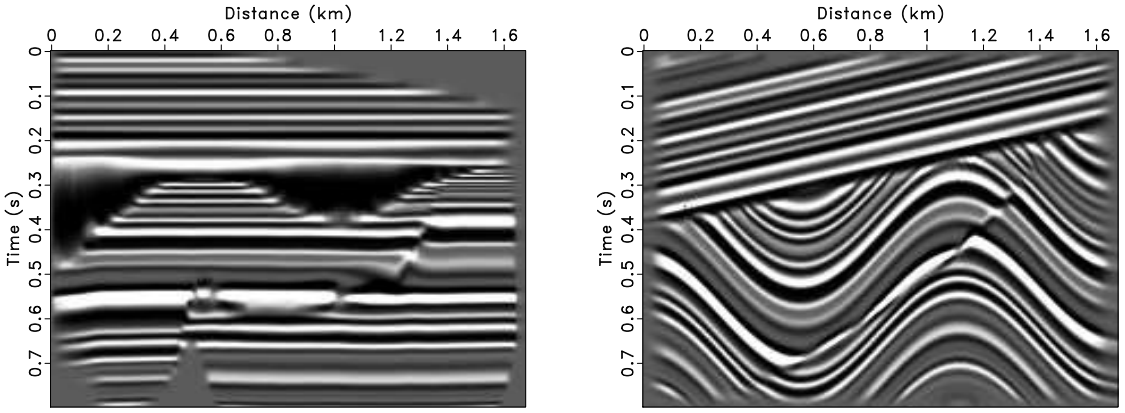


Figure 4: (a) Synthetic image from Figure 2(a) flattened by transferring the image to the stratigraphic coordinate system. (b) Unflattened synthetic image reconstruction by returning from stratigraphic coordinates to regular coordinates. strat/sigmoid1 sigmoid2,sigmoid1

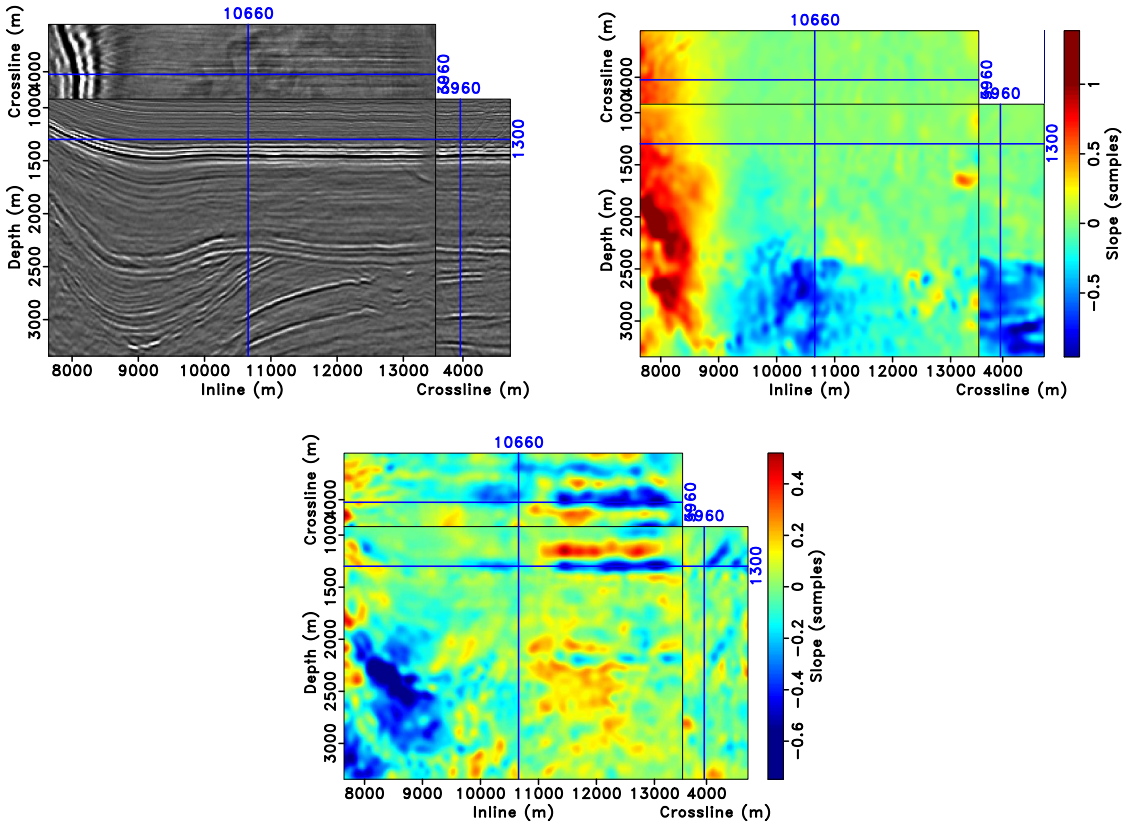


Figure 5: (a) North Sea image from (11) and its inline (b) and cross-line (c) slopes estimated by plane-wave destruction. strat/three-d win,wdip1,wdip2

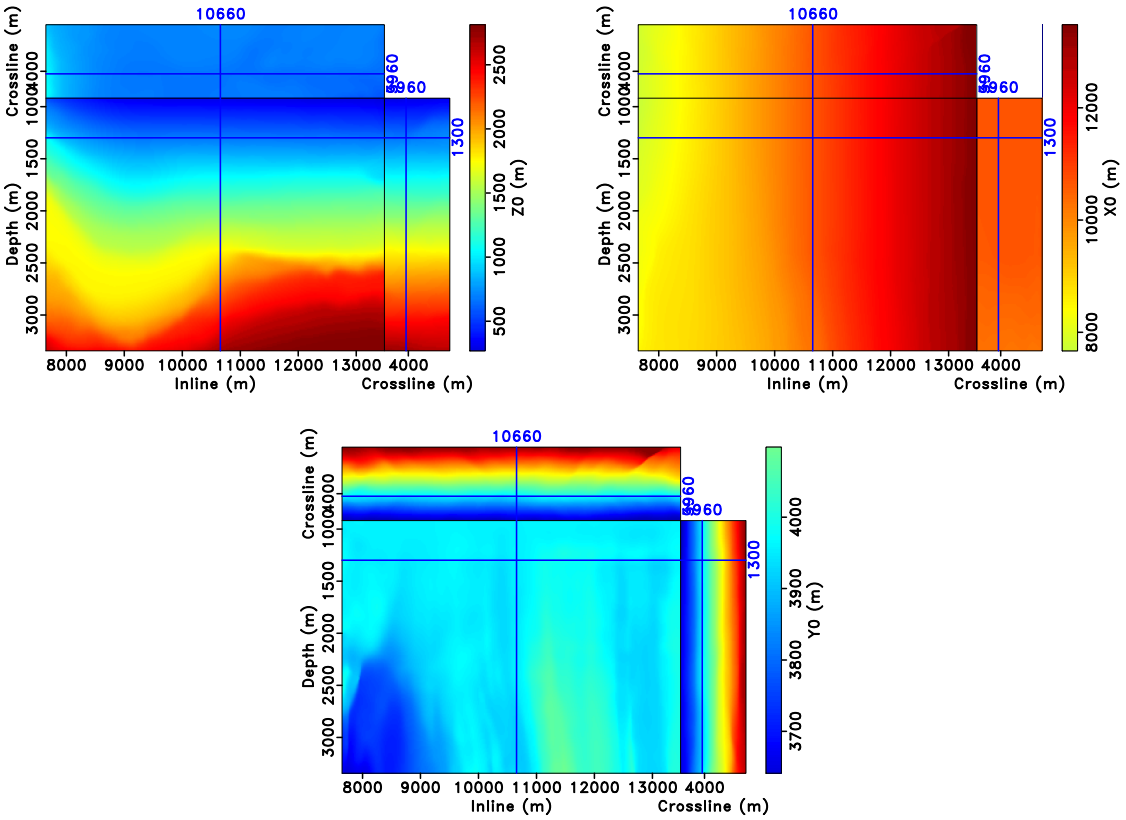


Figure 6: (a) First axis (Z_0), (b) second axis (X_0), and (c) third axis (Y_0) of stratigraphic coordinates in the North Sea image. strat/three-d t0real1,x0real1,y0real2

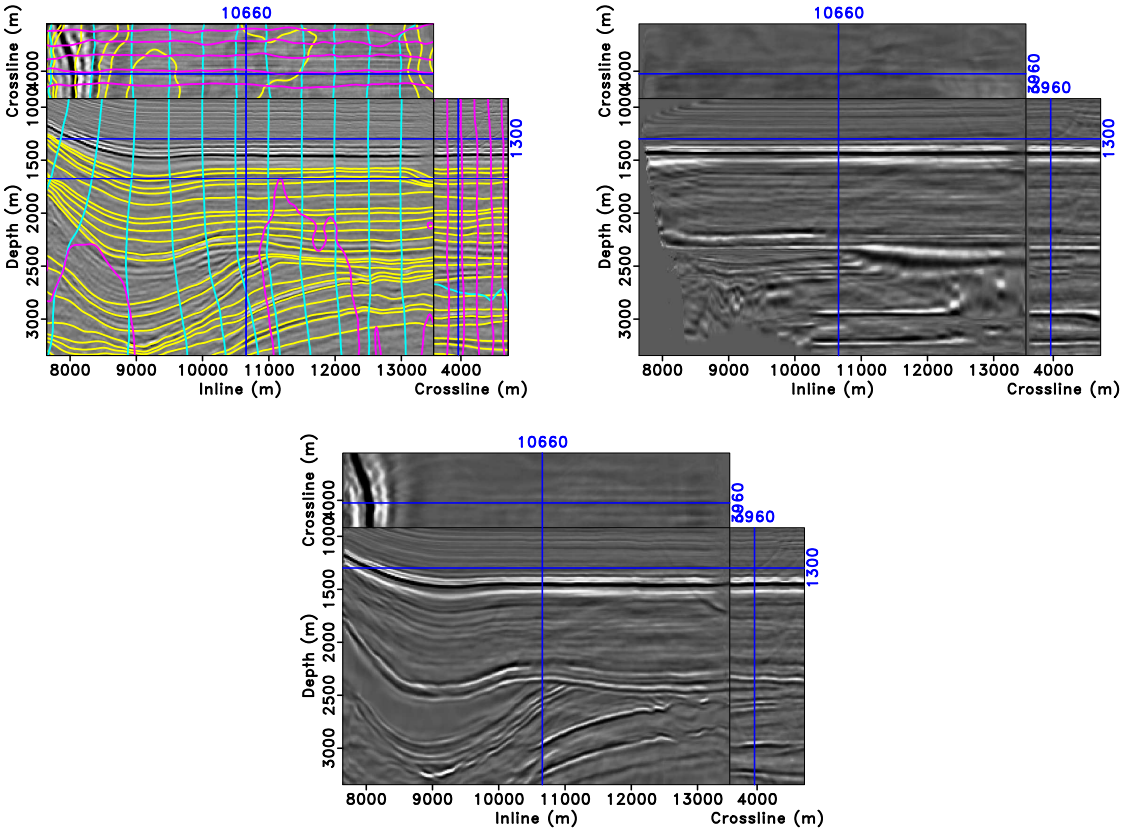


Figure 7: (a) Three axes of stratigraphic coordinates of Figure 5(a) plotted as a grid in their Cartesian coordinates. (b) North Sea image after flattening (transferring image to stratigraphic coordinates). (c) North Sea image reconstruction by returning from stratigraphic coordinates to regular coordinates. strat/three-d coord,win1,win2

mic trace. However, when the subsurface exhibits dipping layers, the convolutional model no longer holds true, and sampling the seismic waveform vertically instead of perpendicularly to reflectors introduces a dip-dependent stretch that will carry over to any frequency estimation or spectral decomposition. (7) proposed to solve this problem by sampling the signal wavelet along the ray-path on which the wavelet travels. Because this path is normal to reflectors, we implement the same idea by employing the stratigraphic coordinate system, which honors the convolutional model and can capture and analyze seismic waveforms perpendicularly to seismic reflectors. Figure 7 shows Gulf of Mexico data reproduced from (11) and (10) that contain a salt dome and horizons that dip steeply on the flank of the dome because of the salt piercement. Following (10) we calculated the spectral decomposition of the data in the Cartesian coordinate system. Figure 9 shows horizon slices from spectral decomposition calculated in the Cartesian coordinate system at different frequencies because depositional elements of different thicknesses tune it at different frequencies. Figure 10 also shows the same horizon slices as Figure 9, but this time from spectral decomposition calculated in the stratigraphic coordinate system. Compared with horizon slices in Figure 9, those from spectral decomposition calculated in the stratigraphic coordinate system better highlight detailed geologic features such as sand channels. That is because in the stratigraphic coordinates seismic horizons get flattened and the vertical direction corresponds to the normal direction to reflectors. We can therefore analyze the unbiased seismic wavefrom and achieve a more accurate spectral decomposition result. Indeed, since seismic reflectors appear flat in the stratigraphic coordinate system, (vertical) trace analysis methods such as spectral decomposition probe the unbiased seismic waveform and thus yield more accurate measurements and attributes. Conversely, methods such as post-stack seismic inversion, spiking deconvolution, tuning analysis, etc., usually assume that layers are flat and might therefore lead interpreters to incur errors in the presence of dips. The same methods, just like spectral decomposition, may benefit from being applied in the stratigraphic coordinate system and thereby produce results unbiased by structural dip.

DISCUSSION AND CONCLUSIONS

We have introduced the stratigraphic coordinate system, a novel framework for seismic interpretation. Our algorithm for constructing the stratigraphic coordinate system consists of two steps. In the first step, we use predictive painting to produce an implicit horizon volume that defines the first axis of the stratigraphic coordinates, aligned with reflection boundaries (seismic horizons). We obtain the remaining two axes of the stratigraphic coordinate system by solving the relevant gradient equations using an upwind finite-difference scheme. Seismic image flattening is an immediate application of stratigraphic coordinates. Other possible applications include post-stack impedance inversion, tuning analysis, spiking deconvolution, or any other process that implicitly assumes that reflectors are flat or that the seismic waveforms should be sampled vertically. In all structures but layercake geology, trace-based attributes can be biased in the presence of dipping layers. In contrast, the stratigraphic coordi-

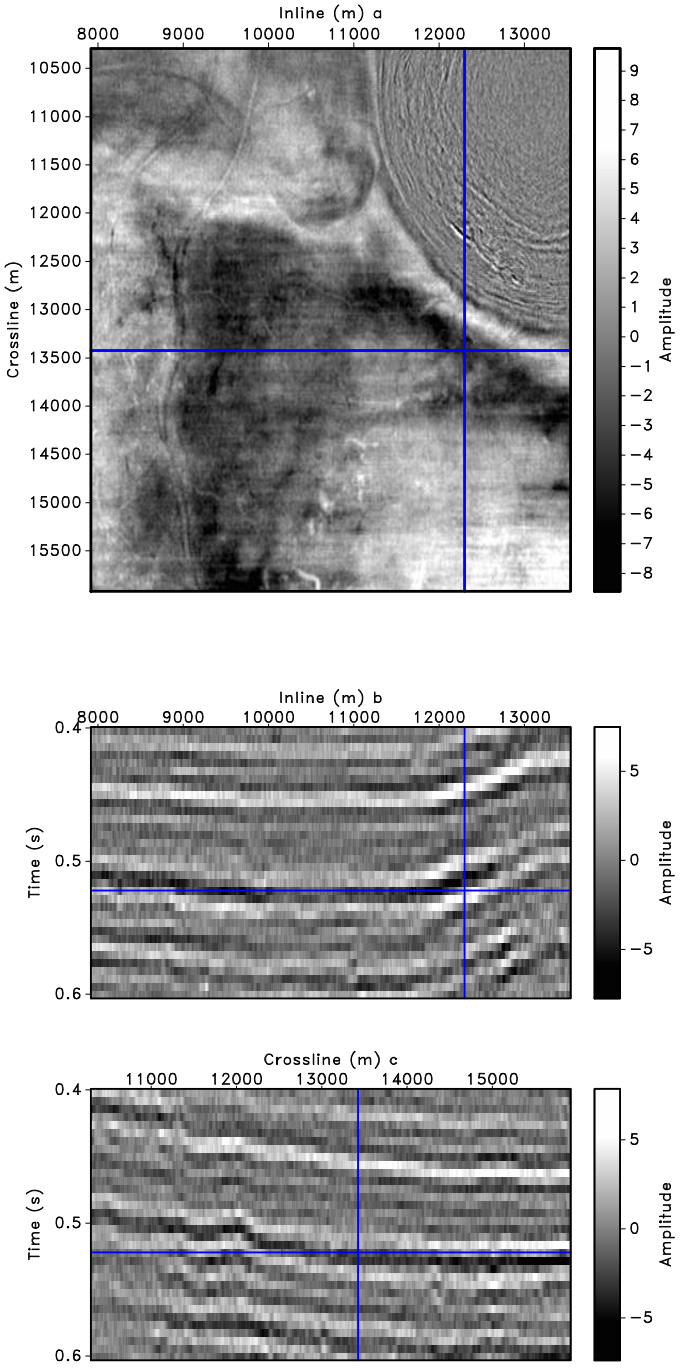


Figure 8: Seismic image from Gulf of Mexico. (a) Time slice. (b) Inline section. (c) Cross-line section. strat/spec-decom chev

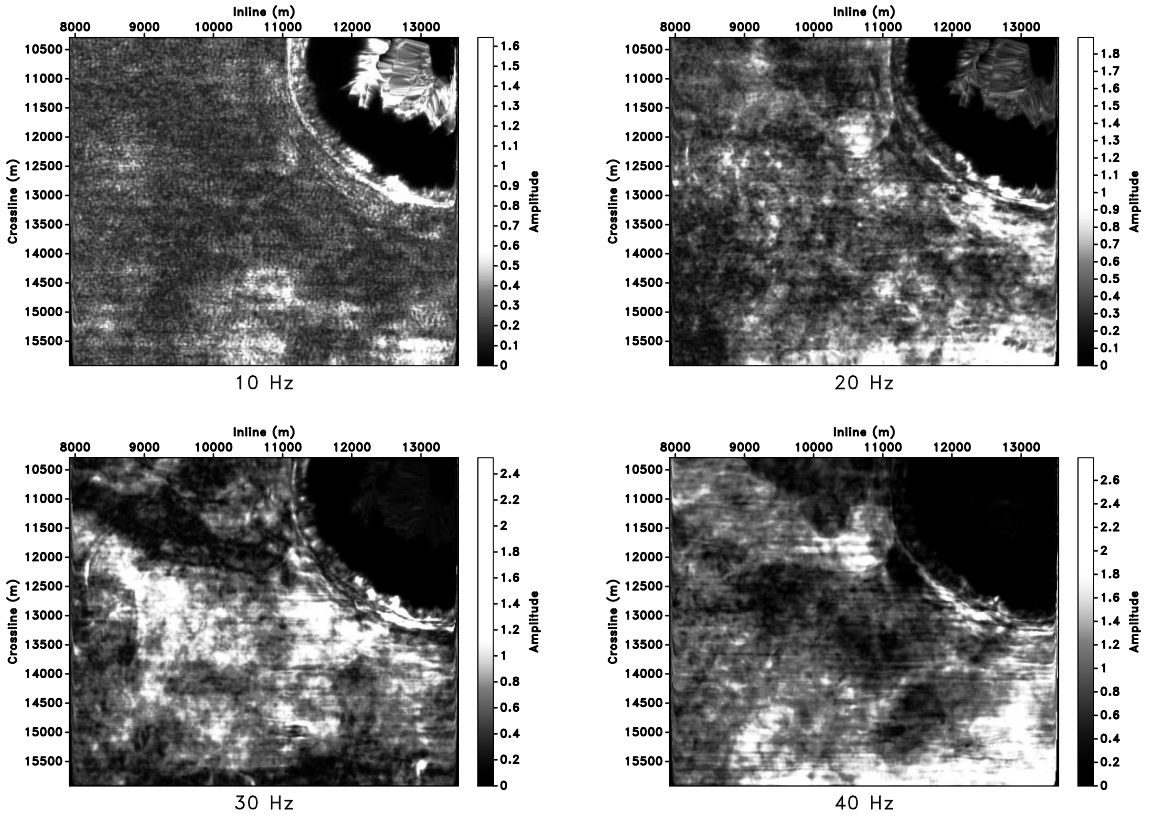


Figure 9: Horizon slices from spectral decomposition at (a) 10 Hz, (b) 20 Hz, (c) 30 Hz, and (d) 40 Hz in Cartesian coordinate system. strat/spec-decom slice-310,slice-320,slice-330,slice-340

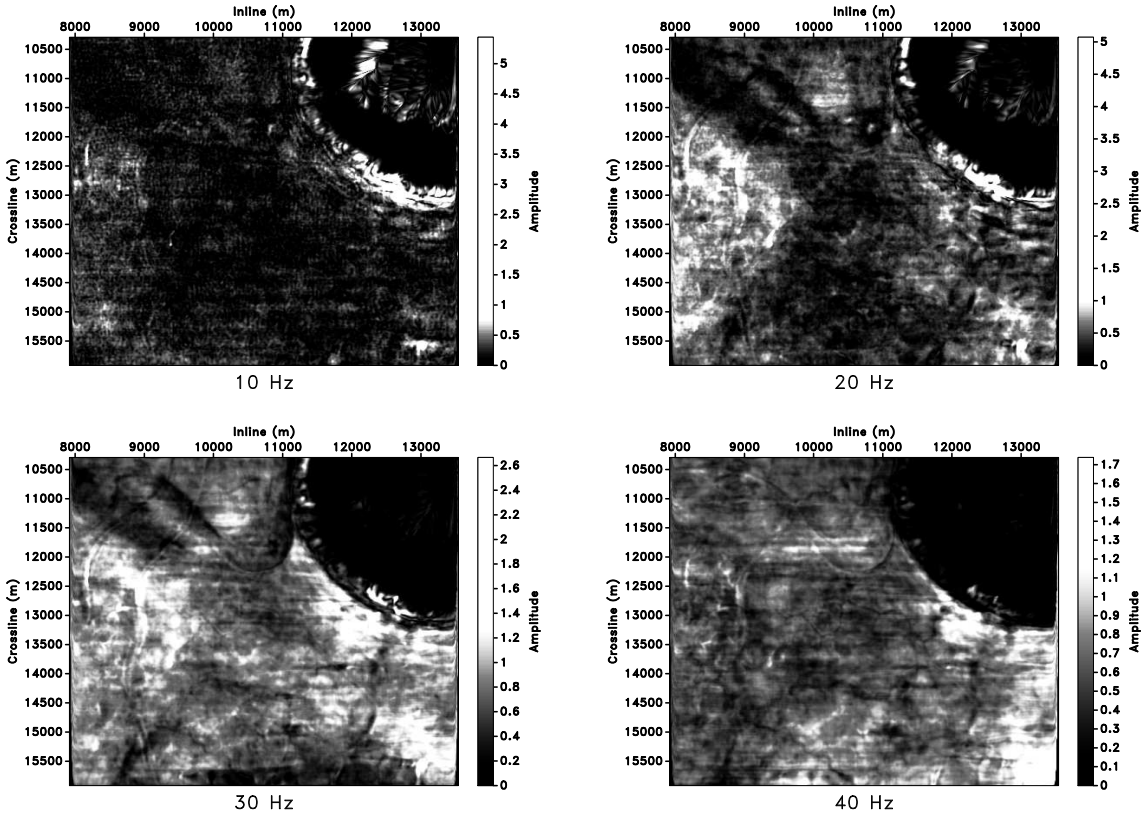


Figure 10: Horizon slices from spectral decomposition at (a) 10 Hz, (b) 20 Hz, (c) 30 Hz, and (d) 40 Hz in stratigraphic coordinate system. The 30 Hz slice most clearly displays visible channel features (red arrows). strat/spec-decom slice-110,slice-120,slice-130,slice-140

nate system offers a local reference frame naturally oriented to sample the unbiased seismic waveform and, hence, promises to yield more accurate waveform analysis and trace attributes.

Our implementation of stratigraphic coordinates is based on the predictive-painting algorithm, which produces the best results when traces can be predicted by their neighbors. Note that the predictive-painting algorithm may fail to capture some of the events across structural or stratigraphic discontinuities. Also, in the presence of crossing dips or incoherent events, predictive painting is susceptible to errors and warrants further improvement.

ACKNOWLEDGMENTS

We thank Stephane Gesbert and other reviewers for their valuable comments and suggestions.

APPENDIX A: PREDICTIVE PAINTING

In the most general case, the predictive-painting method (5) can be described as follows:

Local spatially-variable inline and cross-line slopes of seismic events are estimated by the plane-wave destruction method (4). Plane-wave destruction originates from a local plane-wave model for characterizing seismic data, which is based on the plane-wave differential equation (2):

$$\frac{\partial P}{\partial x} + \sigma \frac{\partial P}{\partial t} = 0. \tag{A-1}$$

Here $P(t, x)$ is the seismic wave-field at time t and location x, and σ is the local slope, which can be either constant or variable in both time and space. The local plane differential equation can easily be solved where the slope is constant, and it has a simple general solution

$$P(t, x) = f(t - \sigma x), \tag{A-2}$$

where $f(t)$ is an arbitrary waveform. Equation A-2 is just a mathematical description of a plane wave. In the case of variable slopes, a local operator is designed to propagate each trace to its neighbors by shifting seismic events along their local slopes.

By writing the plane-wave destruction operation in the linear operator notation, we have

$$\mathbf{r} = \mathbf{D}\mathbf{s}, \tag{A-3}$$

where $\mathbf{s}$ is a seismic section as a collection of traces $\left(\mathbf{s} = [s_1 s_2 \dots s_N]^T\right)$, $\mathbf{r}$ is the destruction residual, and $\mathbf{D}$ is the destruction operator, defined as

$$\mathbf{D} = \begin{bmatrix} \mathbf{I} & 0 & 0 & \dots & 0 \\ \mathbf{P}_{1,2} & \mathbf{I} & 0 & \dots & 0 \\ 0 & -\mathbf{P}_{2,3} & \mathbf{I} & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -\mathbf{P}_{N-1,N} & \mathbf{I} \end{bmatrix}. \tag{A-4}$$

$\mathbf{I}$ is the identity operator, and $\mathbf{P}_{i,j}$ is an operator that predicts trace j from trace i. By minimizing the prediction residual $\mathbf{r}$ using least-squares optimization and smooth regularization, the dominant slopes will be obtained. For $3D$ structure characterization, a pair of inline and crossline slopes, $\sigma_x(t,x,y)$ and $\sigma_y(t,x,y)$, and a pair of destruction operators, $\mathbf{D}_x$ and $\mathbf{D}_y$, are required. The prediction of trace $\mathbf{s}_k$ from reference trace $\mathbf{s}_r$ can be defined as $\mathbf{P}_{r,k}\mathbf{s}_r$, where

$$\mathbf{P}_{r,k} = \mathbf{P}_{k-1,k} \dots \mathbf{P}_{r+1,r+2}\mathbf{P}_{r,r+1}. \tag{A-5}$$

This is a simple recursion, and $\mathbf{P}_{r,k}$ is called the predictive-painting operator. After obtaining elementary prediction operators in equation A-4 by plane-wave destruction, predictive painting spreads the information contained in a seed trace to its neighbors by following the local slope of seismic events. In order to be able to paint all events in the seismic volume, one can use multiple references and average painting values extrapolated from different reference traces.

REFERENCES

[1] Bienati, N., and U. Spagnolini, 1999, Surface slice generation and interpretation: A review: IEEE Transactions on Geoscience and Remote Sensing, **39**, 1287–1290.

[2] Claerbout, J. F., 1992, Earth sounding analysis: Processing versus inversion: Blackwell Science.

[3] ———, 2006, Basic earth imaging: Stanford Exploration Project.

[4] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, no. 6, 1946–1960.

[5] ———, 2010, Predictive painting of 3-D seismic volumes: Geophysics, **75**, no. 4, A25–A30.

[6] Franklin, J. B., and J. M. Harris, 2001, A fast marching level set method for monotonically advancing fronts: Computational Acoustics, **9**, 1095–1109.

[7] Guo, H., and K. J. Marfurt, 2010, Spectral decomposition along the travel path of signal wavelets: Expanded Abstracts 80^{th} SEG International Convention, 1478–1482.

[8] Hoyes, J., and T. Cheret, 2011, A review of "global" interpretation methods for automated 3D horizon picking: The Leading Edge, **30**, 38–47.
[9] Li, S., and S. Fomel, 2013, Kirchhoff migration using eikonal-based computation of traveltime source-derivatives: Geophysics, **78**, no. 4, S211–S219.
[10] Liu, G., S. Fomel, and X. Chen, 2011, Time-frequency analysis of seismic data using local attributes: Geophysics, **76**, no. 6, P23–P34.
[11] Lomask, J., A. Guitton, S. Fomel, J. Claerbout, and A. Valenciano, 2006, Flattening without picking: Geophysics, **71**, no. 4, P12–P20.
[12] Luo, S., and D. Hale, 2013, Unfaulting and unfolding 3d seismic images: Geophysics, **78**, no. 4, O45–O56.
[13] Mallet, J. L., 2004, Space-time mathematical framework for sedimentary geology: Mathematical Geology, **36**, A25–A30.
[14] Parks, D., 2010, Seismic image flattening as a linear inversion problem: Master thesis, Colorado School of Mines.
[15] Partyka, G. A., and J. Lopez, 1999, Interpretational applications of spectral decomposition in reservoir characterization: The Leading Edge, **18**, 353–360.
[16] Stark, T. J., 2005, Generation of a 3D seismic "wheeler diagram" from a high resolution age volume: Expanded Abstracts 75^{th} SEG International Convention, **9**, 782–785.
[17] Wolak, J., N. Hemstra, J. Ochoa, and M. Pelissier, 2013, Reconstruction of depocenter evolution through time using relative stratigraphic thickness: The Leading Edge, **32**, 172–177.
[18] Zeng, H., S. C. Henry, and J. P. Riola, 1998, Stratal slicing, part ii: Real 3-D seismic data: Geophysics, **63**, no. 2, 514–522.

Lowrank finite-differences and lowrank Fourier finite-differences for seismic wave extrapolation in the acoustic approximation

Xiaolei Song, Sergey Fomel*, and Lexing Ying†*

ABSTRACT

We introduce a novel finite-difference (FD) approach for seismic wave extrapolation in time. We derive the coefficients of the finite-difference operator from a lowrank approximation of the space-wavenumber, wave-propagator matrix. Applying the technique of lowrank finite-differences, we also improve the finite difference scheme of the two-way Fourier finite differences (FFD). We call the new operator lowrank Fourier finite differences (LFFD). Both the lowrank FD and lowrank FFD methods can be applied to enhance accuracy in seismic imaging by reverse-time migration. Numerical examples confirm the validity of the proposed technique.

INTRODUCTION

Wave extrapolation in time is crucial in seismic modeling, imaging (reverse-time migration), and full-waveform inversion. The most popular and straightforward way to implement wave extrapolation in time is the method of explicit finite differences (FDs), which is only conditionally stable and suffers from numerical dispersion (39; 13). In practice, a second-order FD for temporal derivatives and a high-order FD for spatial derivatives are often employed to reduce dispersion and improve accuracy. FD coefficients are conventionally determined using a Taylor-series expansion around zero wavenumber (7; 26). Therefore, traditional FD methods are accurate primarily for long-wavelength components.

More advanced methods have been applied previously to FD schemes in the case of one-way wave extrapolation (downward continuation). (24; 25) designed the derivative operator by matching the spectral response in the wavenumber domain. (34) adopted the Remez exchange algorithm to obtain the L_∞-norm-optimized coefficients for second-derivative filters. (29) designed stable explicit depth extrapolators using projections onto convex sets (POCS). These approaches have advantages over conventional FD methods in their ability to propagate shorter-wavelength seismic waves correctly. To satisfy the general criterion for optimal accuracy (21), (22) derived an optimally accurate time-domain finite difference method for computing synthetic seismograms for 1-D problems extended later to 2-D and 3-D (37). (27) proposed

FD schemes for two-way scalar waves on the basis of time-space dispersion relations and plane-wave theory. Later on, they suggested adaptive variable-length spatial operators in order to decrease computing costs significantly without reducing accuracy (28). The Liu-Sen scheme satifies the exact dispersion relation and has greater accuracy and better stability than a conventional one. However, it still uses an expansion around the zero wavenumber.

In sedimentary rocks, anisotropic phenomena are often observed as a result of layering lithification, which is described as transversely isotropic (TI). Tectonic movement of the crust may rotate the rocks and tilt the natural vertical orientation of the symmetry axis (VTI), causing a tilted TI (TTI) anisotropy. Wavefields in anisotropic media are well described by the anisotropic elastic-wave equation. However, in practice, seismologists often have little information about shear waves and prefer to deal with scalar wavefields. Conventional P-wave modeling may contain shear-wave numerical artifacts in the simulated wavefield (23; 42; 8). Those artifacts as well as sharp changes in symmetry-axis tilting may introduce severe numerical dispersion and instability in modeling. (40) proposed to reduce the instability by introducing elliptical anisotropy in regions with rapid tilt changes. (14) suggested to include a finite S-wave velocity in order to enhance stability when solving coupled equations. These methods can alleviate the instability problem; however, they may alter the wave propagation kinematics or leave residual S-wave components in the P-wave simulation. A number of spectral methods are proposed to provide solutions which can completely avoid the shear-wave artifacts (11; 6; 32; 17; 19; 41; 5) at the cost of several Fourier transforms per time step. These methods differ from conventional pseudo-spectral methods (20; 18), because they approximate the space-wavenumber mixed-domain propagation matrix instead of a Laplacian operator.

Our goal is to design an FD scheme that matches the spectral response in the mixed space-wavenumber domain for a wide range of spatial wavenumbers. The scheme is derived from the lowrank approximation of the mixed-domain operator (16; 17) and its representation by FD with adapted coefficients. We derive this kind of FD schemes which we call lowrank FD or LFD for both isotropic and TTI media. Using this approach, we only need to compute the FD coefficients once and save them for the whole process of wave extrapolation or reverse-time migration. The method is flexible enough to control accuracy by the rank of approximation and by FD order selection.

The paper is organized as follows. We first give a brief review of the lowrank approximation method. As a spectral method, it provides an accurate wave extrapolation, but it is not optimally efficient. Next, we present the derivation of LFD. LFD as an FD method can reduce the cost and is also more adaptable for parallel computing on distributed computer systems. We also propose lowrank Fourier FD (LFFD), by replacing the original FD operator in the two-way Fourier FD (FFD) (32) with the corresponding LFD. LFFD improves the accuracy of FFD, in particular in tilted transversely isotropic (TTI) media. A number of synthetic examples of increasing complexity validate the proposed methods. In this paper, we solve the acoustic

wave equation in constant-density media, aiming at incorporating wave extrapolation with LFD and LFFD into seismic imaging by reverse-time migration. It is possible to extend LFD and LFFD to variable-density media by factoring the second-order k-space operator into first-order parts (36; 33).

THEORY

Lowrank Approximation

The following acoustic-wave equation is widely used in seismic modeling and reverse-time migration (12):

$$\frac{\partial^2 p}{\partial t^2} = v(\mathbf{x})^2 \, \nabla^2 p \; , \tag{1}$$

where $\mathbf{x} = (x_1, x_2, x_3)$, $p(\mathbf{x}, t)$ is the seismic pressure wavefield and $v(\mathbf{x})$ is the propagation velocity. Assuming a constant velocity, v, after Fourier transform in space, we could obtain the following explicit expression,

$$\frac{d^2 \hat{p}}{dt^2} = -v^2 |\mathbf{k}|^2 \hat{p} \; , \tag{2}$$

where

$$\hat{p}(\mathbf{k}, t) = \int_{-\infty}^{+\infty} p(\mathbf{x}, t) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \; , \tag{3}$$

and $\mathbf{k} = (k_1, k_2, k_3)$.

Equation 2 has an explicit solution:

$$\hat{p}(\mathbf{k}, t + \Delta t) = e^{\pm i |\mathbf{k}| v \Delta t} \hat{p}(\mathbf{k}, t) \; . \tag{4}$$

A second-order time-marching scheme and the inverse Fourier transform lead to the well-known expression (10; 35):

$$p(\mathbf{x}, t + \Delta t) + p(\mathbf{x}, t - \Delta t) = 2 \int_{-\infty}^{+\infty} \hat{p}(\mathbf{k}, t) \cos(|\mathbf{k}| v \Delta t) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \; . \tag{5}$$

Equation 5 provides an efficient solution in the case of a constant-velocity medium with the aid of the fast Fourier transform (FFT). When velocity varies in space, equation 5 can provide an approximation by replacing v with $v(\mathbf{x})$. In such a case, a mixed-domain term, $\cos(|\mathbf{k}| v(x) \Delta t)$, appears in the expression. As a result, the computational cost of a straightforward application of equation 5 is $O(N_x^2)$, where N_x is the total size of the three-dimensional space grid.

(16; 17) showed that the mixed-domain matrix,

$$W(\mathbf{x}, \mathbf{k}) = \cos(|\mathbf{k}| v \Delta t), \tag{6}$$

can be efficiently decomposed into a separate representation of the following form:

$$W(\mathbf{x},\mathbf{k}) \approx \sum_{m=1}^{M}\sum_{n=1}^{N} W(\mathbf{x},\mathbf{k}_m)a_{mn}W(\mathbf{x}_n,\mathbf{k}), \tag{7}$$

where $W(\mathbf{x},\mathbf{k}_m)$ is a submatrix of $W(\mathbf{x},\mathbf{k})$ that consists of selected columns associated with $\mathbf{k}_m$, $W(\mathbf{x}_n,\mathbf{k})$ is another submatrix that contains selected rows associated with $\mathbf{x}_n$, and a_{mn} stands for the middle matrix coefficients. The construction of the separated form 15 follows the method of (9). The main observation is that the columns of $W(\mathbf{x},\mathbf{k}_m)$ need to span the column space of the original matrix and that the rows of $W(\mathbf{x}_n,\mathbf{k})$ need to span the row space as well as possible.

Representation (15) speeds up the computation of $p(\mathbf{x},t+\Delta t)$ because

$$\begin{aligned} p(\mathbf{x},t+\Delta t)+p(\mathbf{x},t-\Delta t) &= 2\int e^{-i\mathbf{x}\mathbf{k}}W(\mathbf{x},\mathbf{k})\hat{p}(\mathbf{k},t)d\mathbf{k} \\ &\approx 2\sum_{m=1}^{M} W(\mathbf{x},\mathbf{k}_m)\left(\sum_{n=1}^{N} a_{mn}\left(\int e^{-i\mathbf{x}\mathbf{k}}W(\mathbf{x}_n,\mathbf{k})\hat{p}(\mathbf{k},t)d\mathbf{k}\right)\right). \end{aligned} \tag{8}$$

Evaluation of equation 8 requires N inverse FFTs. Correspondingly, the lowrank approximation reduces the cost to $O(NN_x\log N_x)$, where N is a small integer, which is related to the rank of the above decomposition and can be automatically calculated at some given error level with a pre-determined Δt. Increasing the time step size Δt may increase the rank of the approximation (M and N) and correspondingly the number of the required Fourier transforms.

As a spectral method, the lowrank approxmation is highly accurate. However, its cost is several FFTs per time step. Our goal is to reduce the cost further by deriving an FD scheme that matches the spectral response of the output from the lowrank decomposition.

Lowrank Finite Differences

In a matrix notation, the lowrank decomposition problem takes the following form:

$$\mathbf{W} \approx \mathbf{W}_1 \cdot \mathbf{A} \cdot \mathbf{W}_2, \tag{9}$$

where $\mathbf{W}$ is the $N_x \times N_x$ matrix with entries $W(\mathbf{x},\mathbf{k})$, $\mathbf{W}_1$ is the submatrix of $\mathbf{W}$ that consists of the columns associated with $\{\mathbf{k}_m\}$, $\mathbf{W}_2$ is the submatrix that consists of the rows associated with $\{\mathbf{x}_n\}$, and $\mathbf{A}=\{a_{mn}\}$.

Note that $\mathbf{W}_2$ is a matrix related only to wavenumber $\mathbf{k}$. We propose to further decompose it as follows:

$$\mathbf{W_2} \approx \mathbf{C} \cdot \mathbf{B}, \tag{10}$$

where we determine $\mathbf{B}$ to be an $L \times N_x$ matrix, and the entry, $\mathbf{B}(\xi,\mathbf{k})$, has the form of $\cos(\sum_{j=1}^{3}\xi^j k_j \Delta x_j)$, in which ξ is a 3-D integer vector, $\xi=(\xi^1,\xi^2,\xi^3)$, k_j is the jth

component of wavenumber $\mathbf{k}$, Δx_j is the space grid size in the jth direction, $j = 1, 2, 3$, and $\mathbf{C}$ is the matrix product of $\mathbf{W}_2$ and the pseudo-inverse of $\mathbf{B}$. In practice, we apply a weighted-inversion to achieve the pseudo-inverse: putting a larger weight on the low-wavenumber part and a smaller weight on the high-wavenumber part to enhance the stability. Now we have a new decompostion for the mixed-domain matrix:

$$\mathbf{W} \approx \mathbf{G} \cdot \mathbf{B}, \tag{11}$$

where $\mathbf{G}$ is an $N_x \times L$ matrix,

$$\mathbf{G} = \mathbf{W}_1 \cdot \mathbf{A} \cdot \mathbf{C}, \tag{12}$$

and

$$\begin{aligned} & p(\mathbf{x}, t+\Delta t) + p(\mathbf{x}, t-\Delta t) = \\ 2\int e^{-i\mathbf{x}\mathbf{k}} W(\mathbf{x},\mathbf{k})\hat{p}(\mathbf{k},t)d\mathbf{k} &\approx 2\sum_{m=1}^{L} G(\mathbf{x},m)\left(\int e^{-i\mathbf{x}\mathbf{k}} B(\xi_m,\mathbf{k})\hat{p}(\mathbf{k},t)d\mathbf{k}\right) \\ &\approx \sum_{m=1}^{L} G(\mathbf{x},m)\left(\int e^{-i\mathbf{x}\mathbf{k}} 2\cos(\sum_{j=1}^{3}\xi_m^j k_j \Delta x_j)\hat{p}(\mathbf{k},t)d\mathbf{k}\right) \\ &\approx \sum_{m=1}^{L} G(\mathbf{x},m)\left[\int e^{-i\mathbf{x}\cdot\mathbf{k}}(e^{i\sum\limits_{j=1}^{3}\xi_m^j k_j\Delta x_j} + e^{-i\sum\limits_{j=1}^{3}\xi_m^j k_j\Delta x_j})\hat{p}(\mathbf{k},t)d\mathbf{k}\right]. \end{aligned} \tag{13}$$

According to the shift property of FFTs, we finally obtain an expression in the space-domain

$$p(\mathbf{x}, t+\Delta t) + p(\mathbf{x}, t-\Delta t) = \sum_{m=1}^{L} G(\mathbf{x},m)[p(\mathbf{x}_L, t) + p(\mathbf{x}_R, t)], \tag{14}$$

where $\mathbf{x}_L = (x_1 - \xi_m^1\Delta x_1, x_2 - \xi_m^2\Delta x_2, x_3 - \xi_m^3\Delta x_3)$, and $\mathbf{x}_R = (x_1 + \xi_m^1\Delta x_1, x_2 + \xi_m^2\Delta x_2, x_3 + \xi_m^3\Delta x_3)$.

Equation 14 indicates a procedure of finite differences for wave extrapolation: the integer vector, $\xi_{\mathbf{m}} = (\xi_m^1, \xi_m^2, \xi_m^3)$ provides the stencil information, and $G(\mathbf{x}, m)$ stores the corresponding coefficients. We call this method *lowrank finite differences* (LFD) because the finite-difference coefficients are derived from a lowrank approximation of the mixed-domain propagator matrix. We expect the derived LFD scheme to accurately propagate seismic-wave components within a wide range of wavenumbers, which has advantages over conventional finite differences that focus mainly on small wavenumbers. In comparison with the Fourier-domain approach, the cost is reduced to $O(L\,N_x)$, where L, as the row size of matrix $\mathbf{B}$, is related to the order of the scheme. L can be used to characterize the number of FD coefficients in the LFD scheme, shown in equation 14. Take the 1-D 10th order LFD as an example, there are 1 center point, 5 left points ($\mathbf{x}_L$) and 5 right ones ($\mathbf{x}_R$). So $\xi_m^1 = 0, 1, 2, 3, 4, 5,$

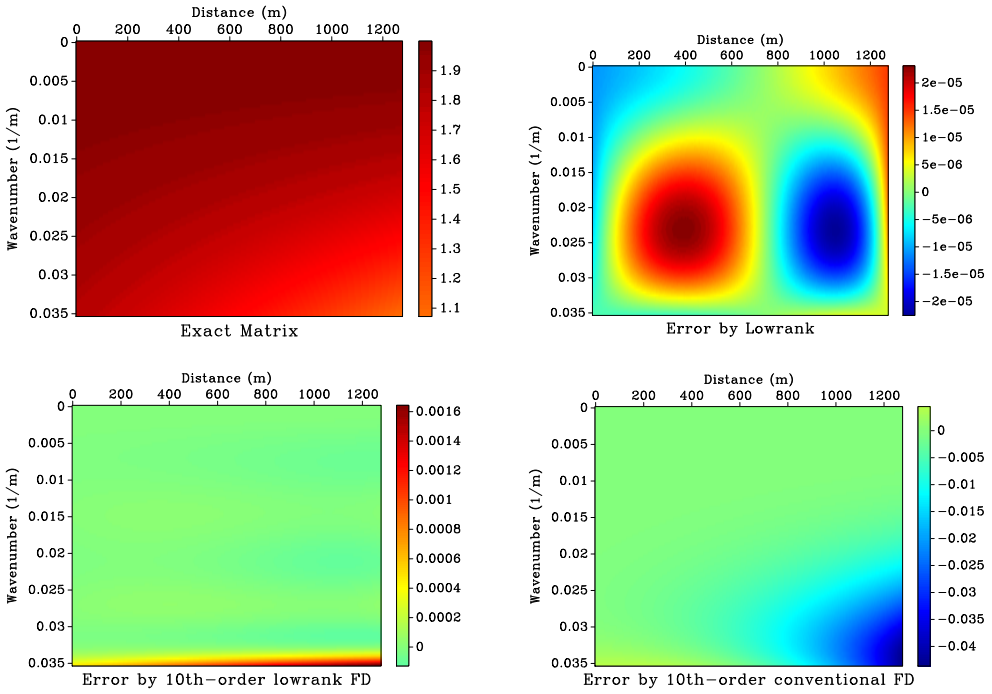


Figure 1: (a) Wavefield extrapolation matrix for 1-D linearly increasing velocity model. Error of wavefield extrapolation matrix by:(b) lowrank approximation, (c) the 10th-order lowrank FD (d) the 10th-order FD. lfd/oned Mexact,Mlrerr,Mapperr,Mfd10err

and $\xi_m = (\xi_m^1, 0, 0)$. Thanks to the symmetry of the scheme, coefficients of $\mathbf{x}_L$ and $\mathbf{x}_R$ are the same, as indicated by equation 14. As a result, one only needs 6 coefficients: $L = 6$.

We use a one-dimentional example shown in Figure 1 to demonstrate the accuracy of the proposed LFD method. The velocity linearly increases from 1000 to 2275 m/s. The rank is 3 ($M = N = 3$) for lowrank decomposition for this model with 1 ms time step. The propagator matrix is shown in Figure 4(a). Figure 4(b)-Figure 4(d) display the errors corresponding to different approximations. The error by the 10th-order lowrank finite differences (Figure 4(c)) appears significantly smaller than that of the 10th-order finite difference (Figure 4(d)). Figure 5 displays the middle column of the error matrix. Note that the error of the LFD is significantly closer to zero than that of the FD method.

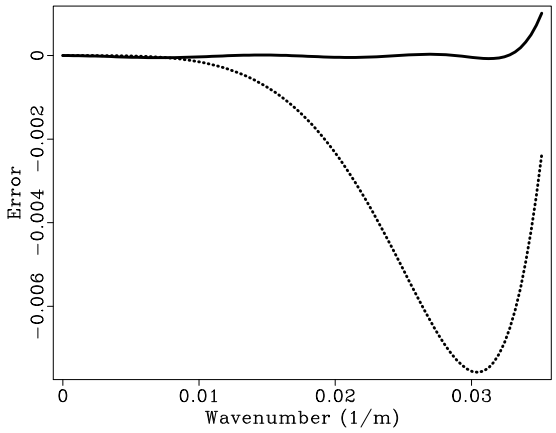


Figure 2: Middle column of the error matrix. Solid line: the 10th-order LFD. Dash line: the 10th-order FD. lfd/oned slicel

To analyze the accuracy, we let

$$p(\mathbf{x}, t) = e^{i(\mathbf{k}\cdot\mathbf{x} - \omega t)}, \tag{15}$$

by using the plane wave theory. Inserting 1 into equation 14 and also adopting the dispersion relation $\omega = |k|\, v$, defines the phase velocity of LFD (v_{LFD}) as follows:

$$v_{LFD} = \frac{1}{|k|\Delta t} \arccos\Big(\sum_{m=1}^{L} G(\mathbf{x}, m)(\cos(\xi_m^1\, k_1 \Delta x_1) + \cos(\xi_m^2\, k_2 \Delta x_2) + \cos(\xi_m^3\, k_3 \Delta x_3))\Big), \tag{16}$$

For 1-D 10th order LFD, $L = 6$, $\xi_m = (\xi_m^1, 0, 0)$ and $\xi_m^1 = 0, 1, 2, 3, 4, 5$. With equation 16, we can calculate phase-velocities (v_{LFD}) by 1-D 10th order LFD with different velocities ($v = 2500, 3000, 3500, 4000$), and we use the ratio $\delta = v_{LFD}/v$ to describe the dispersion of FD methods. Figure 3(a) displays 1D dispersion curves by 1-D 10th order LFD, and Figure 3(b) shows those by conventional FD method. Note that compared with the conventional FD method, LFD is accurate in a wider range of wavenumbers (up to 70% of the Nyquist frequency).

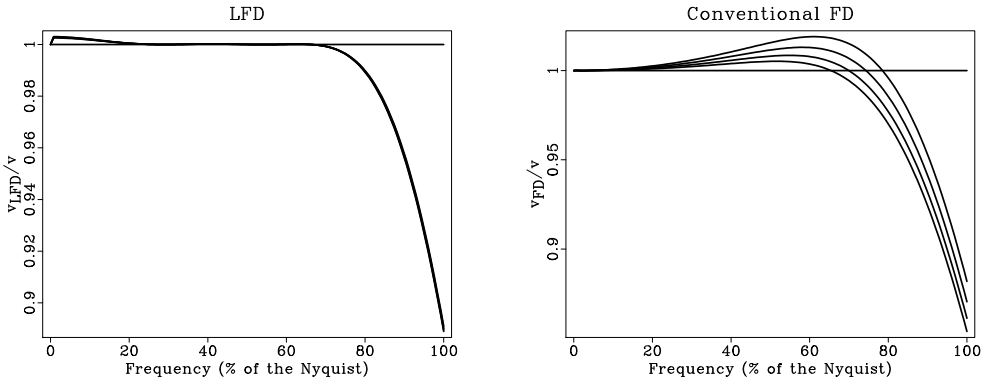


Figure 3: Plot of 1-D dispersion curves for different velocities, $v = 2500$ (red), 3000 (pink), 3500 (green), 4000 (blue) m/s, $\Delta t = 1$ ms, $\Delta x = 10$ m by: (a) the 10th-order LFD (b) the 10th-order conventional FD. lfd/dispersion1 app,fd10

TTI Lowrank Finite Differences

The LFD approach is not limited to the isotropic case. In the case of transversely isotropic (TTI) media, the term $v(\mathbf{x})\,|\mathbf{k}|$ on the right-hand side of equation 6, can be replaced with the acoustic approximation (1; 2; 15),

$$f(\mathbf{v},\hat{\mathbf{k}},\eta) = \sqrt{\frac{1}{2}(v_1^2\,\hat{k}_1^2 + v_2^2\,\hat{k}_2^2) + \frac{1}{2}\sqrt{(v_1^2\,\hat{k}_1^2 + v_2^2\,\hat{k}_2^2)^2 - \frac{8\eta}{1+2\eta}v_1^2 v_2^2\,\hat{k}_1^2\,\hat{k}_2^2}}\,, \tag{17}$$

where v_1 is the P-wave phase velocity in the symmetry plane, v_2 is the P-wave phase velocity in the direction normal to the symmetry plane, η is the anisotropic elastic parameter (3) related to Thomsen's elastic parameters ϵ and δ (38) by

$$\frac{1+2\delta}{1+2\epsilon} = \frac{1}{1+2\eta}\,; \tag{18}$$

and $\hat{k}_1$ and $\hat{k}_2$ stand for the wavenumbers evaluated in a rotated coordinate system aligned with the symmetry axis:

$$\begin{aligned} \hat{k}_1 &= k_1\cos\theta + k_2\sin\theta \\ \hat{k}_2 &= -k_1\sin\theta + k_2\cos\theta \end{aligned} \tag{19}$$

where θ is the tilt angle measured with respect to vertical. Using these definitions, we develop a version of the lowrank finite-difference scheme for 2D TTI media.

Lowrank Fourier Finite Differences

(32) proposed FFD approach to solve the two-way wave equation. The FFD operator is a chain operator that combines FFT and FD, analogous to the concept introduced

previously for one-way wave extrapolation by (30). The FFD method adopts the pseudo-analytical solution of the acoustic wave equation, shown in equation 5. It first extrapolates the current wavefield with some constant reference velocity and then applies FD to correct the wavefield according to local model parameter variations. In the TTI case, the FD scheme in FFD is typically a 4th-order operator, derived from Taylor's expansion around $k = 0$. However, it may exhibit some dispersion caused by the inaccuracy of the FD part. We propose to replace the original FD operator with lowrank FD in order to increase the accuracy of FFD in isotropic and anisotropic media. We call the new operator lowrank Fourier Finite Differences (LFFD).

NUMERICAL EXAMPLES

Our first example is wave extrapolation in a 2-D, smoothly variable velocity model. The velocity ranges between 500 and 1300 m/s, and is formulated as

$$v(x,z) \;=\; 500 + 1.2 \times 10^{-4}(x-800)^2 + 10^{-4}(z-500)^2; \tag{20}$$

$0 \le x \le 2560,\ 0 \le z \le 2560$. A Ricker-wavelet source with a 20 Hz dominant frequency (f_d) is located at the center of the model. The maximum frequency (f_{max}) is around 60 Hz. The amplitude corresponding to f_{max} is about 10^{-5} of that of f_d. For numerical simulations based on this model, we use the same grid size: $\Delta x = 5$ m and $\Delta t = 2$ ms. We use $\alpha = v_{max}\,\Delta t/\Delta x$ to specify the stability condition and $\beta = v_{min}/(f_{max}\Delta x)$ as the dispersion factor, where v_{max} and v_{min} are the maximum and minimum velocities of the model. The dispersion factor β indicates the number of sampling points for the minimum wavelength. For simulations with the above parameters, $\alpha \approx 0.52$ and $\beta \approx 1.67$.

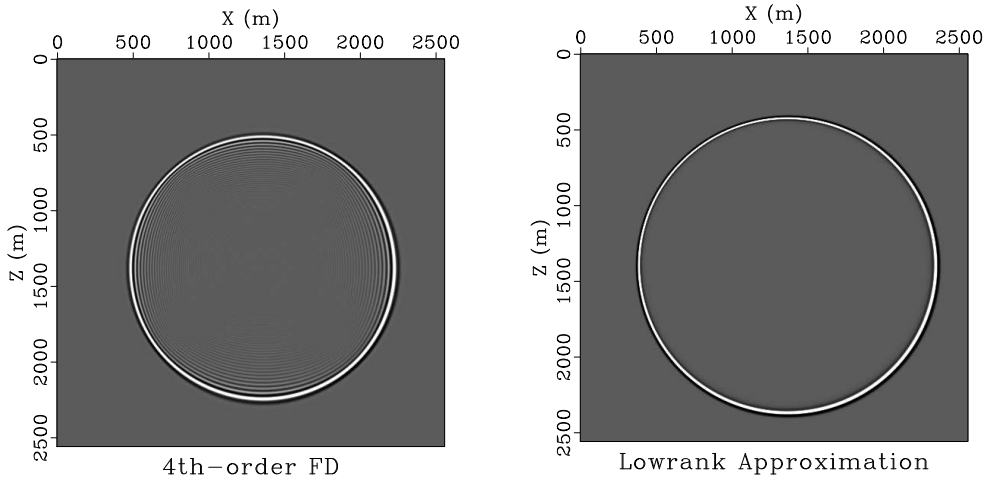


Figure 4: Wavefield snapshot in a varable velocity field by: (a) conventional 4th-order FD method (b) Lowrank method. lfd/twod wavfd7sn,wavel

It is easy to observe obvious numerical dispersions on the snapshot computed by the 4th-order FD method (Figure 4(a)). The lowrank FD method with the same

order exhibits higher accuracy and fewer dispersion artifacts (Figure 5(a)). The approximation rank decomposition in this case is $N = 3\,M = 4$, with the expected error of less than 10^{-4}. Figure 5(b) displays the snapshot by the 10th-order LFD method with a larger time step: $\Delta t = 2.5$ ms, $\alpha \approx 0.65$. Note that the result is still accurate. However, the regular FD method becomes unstable in this case. For comparison, Figure 4(b) displays the snapshot by the lowrank method with the same time step. The approximation rank decomposition in this case is $N = M = 3$, with the expected error of less than 10^{-4}. Thanks to the spectral nature of the algorithm, the result appears accurate and free of dispersion artifacts.

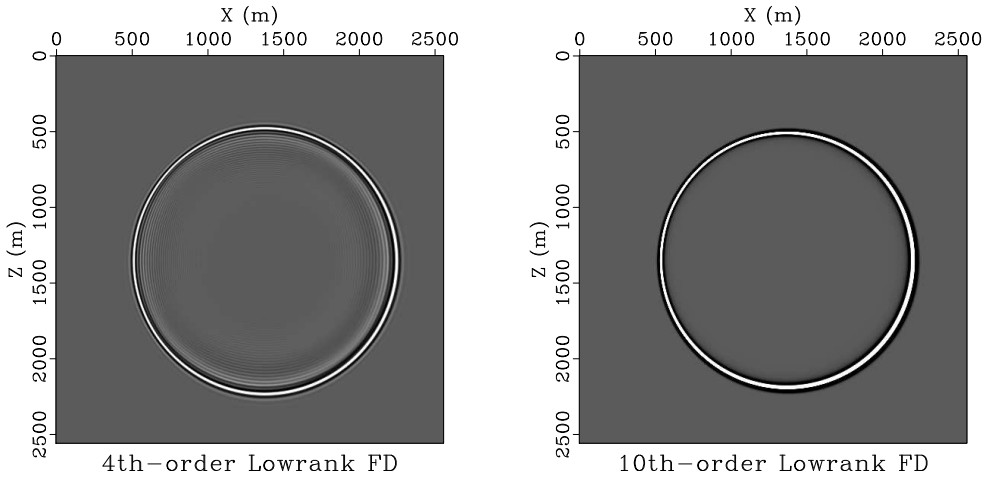


Figure 5: Wavefield snapshot in a varable velocity field by: (a) the 4th-order lowrank FD method (b) the 10th-order lowrank FD method. Note that the time step is 2.5 ms and the LFD result is still accurate. However, the FD method is unstable in this case. lfd/twod wavapp7sn,wavapp10sn

Next, we test the lowrank FD method in a complex velocity model. Figure 6 shows a part of the BP velocity model (4), which is a complicated model containing a salt body and sharp velocity contrasts on the flanks of the salt body. We use a Ricker-wavelet at a point source. The dominant frequency is 17 Hz ($f_{max} \approx 54$). The horizontal grid size Δx is 12.5 m, the vertical grid size Δz is 12.5 m, and the time step is 1.5 ms. Thus $\alpha \approx 0.57$ and $\beta \approx 2.2$. The approximation rank decomposition in this case is $N = 4\,M = 5$, with the expected error of less than 10^{-4}. In this case, we adopt a disk-shaped compact scheme (8th-order) for LFD with a 4-point radius ($|\xi| \leq 4$, $L = 25$). Figure 7 displays a wavefield snapshot in the above velocity model. The snapshot is almost free of dispersions. This experiment confirms that the lowrank FD method is able to handle sharp velocity variations.

Our next example is wave propagation in a TTI model with a tilt of 45° and smooth velocity variation (v_x: 800-1225.41 m/s, v_z: 700-883.6 m/s). Figure 8(a) shows wavefield snapshots at different time steps by a 16th-order LFD operator in the TTI model. The space grid size is 5 m and the time step size is 2 ms. So $\alpha \approx 0.49$ and $\beta \approx 2.3$. The approximation rank decomposition in this case is $N = 6\,M = 6$,

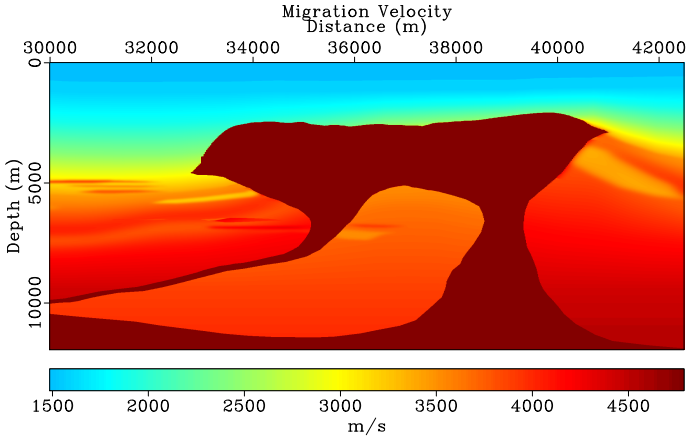


Figure 6: Portion of BP 2004 synthetic velocity model. lfd/bp sub1

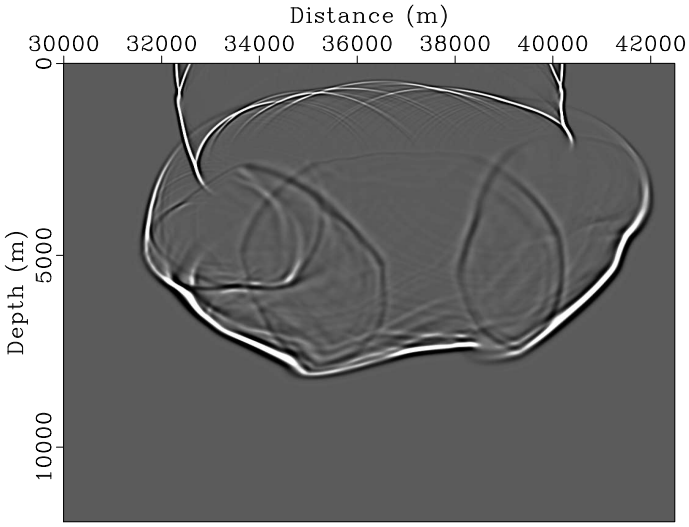


Figure 7: Wavefield snapshot by the 8th-order lowrank FD (compact scheme) in the BP Model shown in Figure 6. lfd/bp wavsnapabc

with the expected error of less than 10^{-4}. For TTI model, we adopt a high-order (16th order) LFD operator in order to reduce dispersions. The scheme is compact and shaped as a disk with a radius of 8 points ($L = 99$).

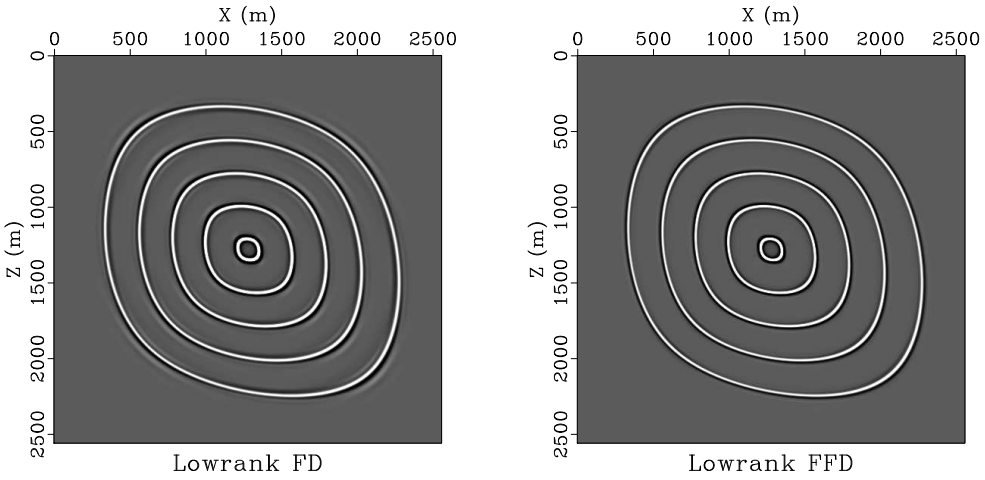


Figure 8: Wavefield snapshots in a TTI medium with a tilt of 45 ° by: (a) Lowrank FD method; (b) Lowrank FFD method.$v_x(x,z) = 800+10^{-4}(x-1000)^2+10^{-4}(z-1200)^2$; $v_z(x,z) = 700+10^{-4}(z-1200)^2$; $\eta = 0.3$; $\theta = 45\,^\circ$. lfd/aniso snapshotlfd,snapshotlffd

(32) showed an application of FFD method for TTI media. However, the example wavefield snapshot by FFD method still had some dispersion caused by the fact that the FD scheme in the FFD operator is derived from Taylor's expansion around zero wavenumber. It was apparent that 4th-order FD scheme is not accurate enough for TTI case and requires denser sampling per wavelength ($\beta \approx 4.6$). We first apply lowrank approximation to the mixed-domain velocity correction term in FFD. The rank is $N = 9\,M = 9$, with the expected error of less than 10^{-6}. Then we propose to replace that 4th-order FD operator with an 8th-order LFD compact scheme. The scheme has the shape of a disk with a radius of 4 points ($L = 25$), the same as the one for LFD in the above BP model. Figure 8(b) shows wavefield snapshots by the proposed LFFD operator. The time step size is 1.5 ms ($\alpha \approx 0.37$). Note that the wavefront is clean and almost free of dispersion with $\beta \approx 2.3$. Because we use the exact dispersion relation, Equation 10 for TTI computation, there is no coupling of q-SV wave and q-P wave (23; 42; 8) in our snapshots by either LFD or LFFD methods.

Next we test the LFD and LFFD methods in a complex TTI model. Figure 9(a)-9(d) shows parameters for part of the BP 2D TTI model (31). The dominant frequency is 15 Hz ($f_{max} \approx 50$). The space grid size is 12.5 m and the time step is 1 ms. Thus $\alpha \approx 0.42$ and $\beta \approx 2.4$. The approximation rank decomposition for LFD method is $N = 22\,M = 22$, with the expected error of less than 10^{-6}. For FFD, $N = 24\,M = 30$, with the expected error of less than 10^{-6}. Both methods are able to simulate an accurate qP-wave field in this model as shown in Figure 10(a) and 10(b).

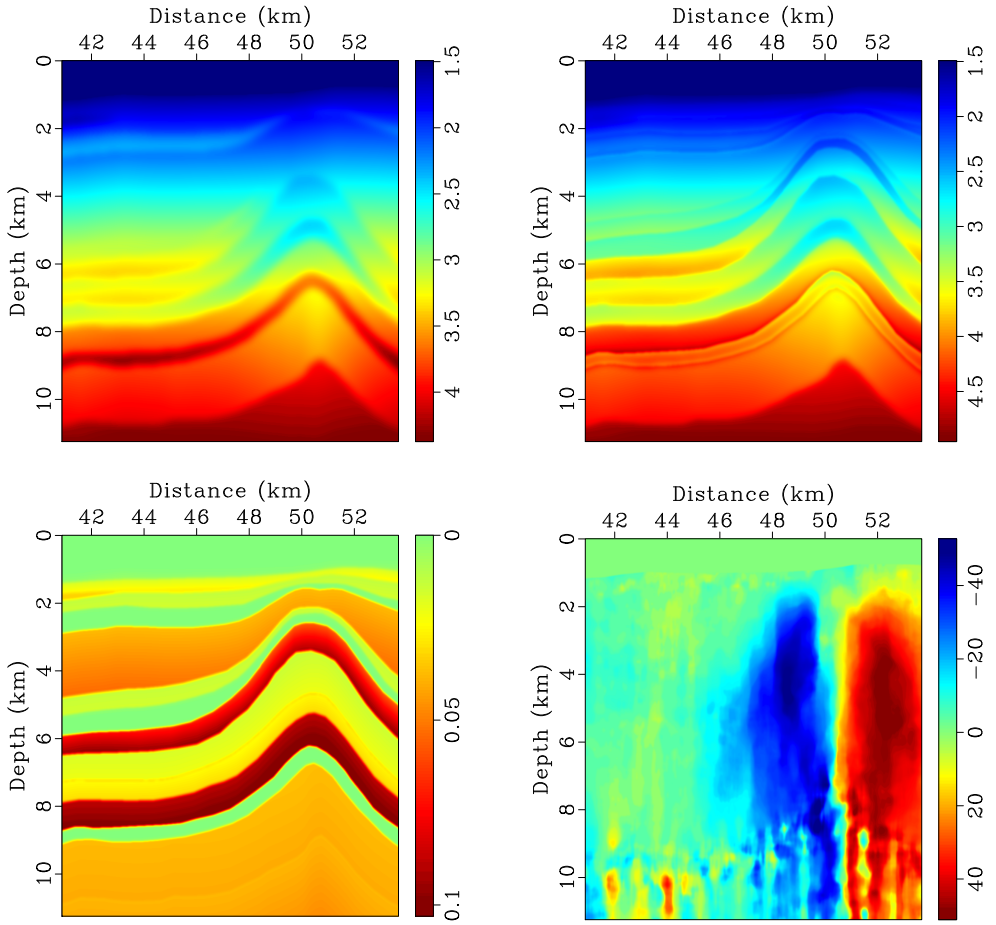


Figure 9: Partial region of the 2D BP TTI model. a: v_z. b: v_x. c: η. d:θ. lfd/bptti vp0,vx0,yita0,theta0

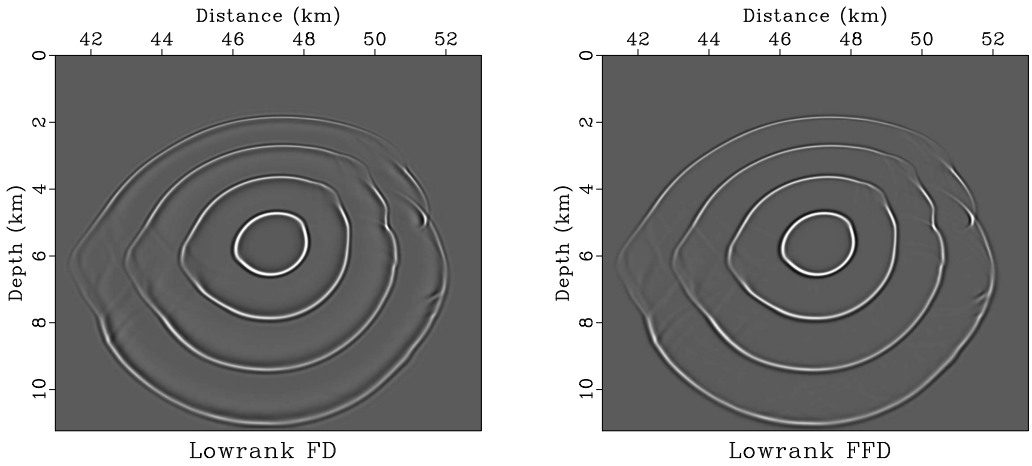


Figure 10: Scalar wavefield snapshots by LFD and LFFD methods in the 2D BP TTI model. lfd/bptti snapshotslfd,snapshotslffd

It is difficult to provide analytical stability analysis for LFD and LFFD operators. In our experience, the values of α are around 0.5 for 2D LFD and LFD methods appear to allow for a larger time step size than that of the LFFD method. In TTI case, the conventional FD method for acoustic TTI has known issues of instability caused by shear-wave numerical artifacts or sharp changes in the symmetry-axis tilting (23; 42; 8). Conventional methods may place limits on anisotropic parameters, smooth parameter models or include a finite shear-wave velocity to alleviate the instability problem(40; 43; 14). Both LFD and LFFD methods are free of shear-wave artifacts. They require no particular bounds for anisotropic parameters and can also handle sharp tilt changes.

CONCLUSIONS

Explicit finite difference (FD) methods are the most popular and straightforward methods for seismic modeling and seismic imaging, particularly for reverse-time migration. Traditionally the coefficients of FD schemes are derived from a Taylor series expansion around the zero wavenumber. We present a novel FD scheme: Lowrank Finite Differences (LFD), which is based on the lowrank approximation of the mixed-domain space-wavenumber propagator. LFD uses compact FD schemes, which are more suitable for parallelization on multi-core computers than spectral methods that require FFT operations. This technique promises higher accuracy and better stability than those of the conventional, explicit FD method. We also propose to replace the 4th-order FD operator based on Taylor's expansion in Fourier Finite Differences (FFD) with an 8th-order LFD operator to reduce dispersion, particularly in the TTI case. Results from synthetic experiments illustrate the stability of the proposed methods in complicated velocity models. In TTI media, there is no coupling of qP-waves and qSv-waves by either method. Both methods can be incorporated in seismic imaging by reverse-time migration to enhance its accuracy and stability.

REFERENCES

[1] Alkhalifah, T., 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.

[2] ———, 2000, An acoustic wave equation for anisotropic media: Geophysics, **65**, 1239–1250.

[3] Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

[4] Billette, F. J., and S. Brandsberg-Dahl, 2004, The 2004 BP velocity benchmark: 67th Annual EAGE Meeting, EAGE, Expanded Abstracts, B305.

[5] Cheng, J., and W. Kang, 2012, Propagating pure wave modes in 3D general anisotropic media, Part I: P-wave propagator: 82nd Ann. Internat. Mtg., Soc. Expl. Geophys., Accepted.

[6] Chu, C., and P. L. Stoffa, 2011, Application of normalized pseudo-Laplacian to elastic wave modeling on staggered grids: Geophysics, **76**, T113–T121.
[7] Dablain, M. A., 1986, The application of high-order differencing to the scalar wave equation: Geophysics, **51**, 54–66.
[8] Duveneck, E., P. Milcik, P. M. Bakker, and C. Perkins, 2008, Acoustic VTI wave equations and their application for anisotropic reverse-time migration: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2186–2190.
[9] Engquist, B., and L. Ying, 2009, A fast directional algorithm for high frequency acoustic scattering in two dimensions: Commun. Math. Sci., **7**, 327–345.
[10] Etgen, J., 1989, Accurate wave equation modeling: SEP-60, 131–148.
[11] Etgen, J., and S. Brandsberg-Dahl, 2009, The pseudo-analytical method: application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation: 79nd Annual International Meeting, SEG, Expanded Abstracts, 2552–2556.
[12] Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: Geophysics, **74**, WCA5–WCA17.
[13] Finkelstein, B., and R. Kastner, 2007, Finite difference time domain dispersion reduction schemes: Journal of Computational Physics, **221**, 422–438.
[14] Fletcher, R. P., X. Du, and P. J. Fowler, 2009, Reverse time migration in tilted transversely isotropic (TTI) media: Geophysics, **74**, WCA179–WCA187.
[15] Fomel, S., 2004, On anelliptic approximations for qP velocities in VTI media: Geophysical Prospecting, **52**, 247–259.
[16] Fomel, S., L. Ying, and X. Song, 2010, Seismic wave extrapolation using a lowrank symbol approximation: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3092–3096.
[17] ———, 2012, Seismic wave extrapolation using lowrank symbol approximation: Geophysical Prospecting.
[18] Fornberg, B., 2002, A practical guide to pseudo-spectral methods (cambridge monographs on applied and computational mathematics): Cambridge University Press.
[19] Fowler, P. J., and C. Lapilli, 2012, Generalized pseudospectral methods for modeling and reverse-time migration in orthorhombic media: 74th Annual EAGE Meeting, EAGE, Expanded Abstracts, AO22.
[20] Gazdag, J., 1981, Modeling of the acoustic wave equation with transform methods: Geophysics, **46**, 854–859.
[21] Geller, R. J., and N. Takeuchi, 1995, A new method for computing highly accurate DSM synthetic seismograms: Geophys. J. Int., **123**, 449–470.
[22] ———, 1998, Optimally accurate second-order time domain finite difference scheme for the elastic equation of motion: 1-D case: Geophys. J. Int., **135**, 48–62.
[23] Grechka, V., L. Zhang, and J. W. Rector, 2004, Shear waves in acoustic anisotropic media: Geophysics, **69**, 576–582.
[24] Holberg, O., 1987, Computational aspect of the choice of operator and sampling interval for numerical differentiation in large scale simulation of wave phenomena: Geophysical Prospecting, **35**, 629–655.
[25] ———, 1988, Towards optimum one-way wave propagation: Geophysics, **36**, 99–114.

[26] Kindelan, M., A. Kamel, and P. Sguazzero, 1990, On the construction and efficiency of staggered numerical differentiators for the wave equation: Geophysics, **55**, 107–110.

[27] Liu, Y., and M. K. Sen, 2009, A new time-space domain high-order finite-difference method for the acoustic wave equation: Journal of Computational Physics, **228**, 8779–8806.

[28] ———, 2011, Finite-difference modeling with adaptive variable-length spatial operators: Geophysics, **76**, T79–T89.

[29] Mousa, W. A., M. van der Baan, S. Boussakta, and D. C. McLernon, 2009, Designing stable extrapolators for explicit depth extrapolation of 2D and 3D wavefields using projections onto convex sets: Geophysics, **74**, S33–S45.

[30] Ristow, D., and T. Ruhl, 1994, Fourier finite-difference migration: Geophysics, **59**, 1882–1893.

[31] Shah, H., 2007, The 2007 BP anisotropic velocity-analysis benchmark: 70th Annual EAGE Meeting, EAGE, workshop.

[32] Song, X., and S. Fomel, 2011, Fourier finite-difference wave propagation: Geophysics, **76**, T123–T129.

[33] Song, X., K. Nihei, and J. Stefani, 2012, Seismic modeling in acoustic variable-density media by fourier finite differences: 82th Ann. Internat. Mtg., Accepted.

[34] Soubaras, R., 1996, Explicit 3-D migration using equiripple polynomial expansion and laplacian synthesis: Geophysics, **61**, 1386–1393.

[35] Soubaras, R., and Y. Zhang, 2008, Two-step explicit marching method for reverse time migration: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2272–2276.

[36] Tabei, M., T. D. Mast, and R. C. Waag, 2002, A k-space method for coupled first-order acoustic propagation equations: J. Acoust. Soc. Am., **111**, 53–63.

[37] Takeuchi, N., and R. J. Geller, 2000, Optimally accurate second order time-domain finite difference scheme for computing synthetic seismograms in 2-D and 3-D media: Phys. Earth Planet. Int., **119**, 99–131.

[38] Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.

[39] Wu, W., L. R. Lines, and H. Lu, 1996, Analysis of high-order, finite-difference schemes in 3-D reverse-time migration: Geophysics, **61**, 845–856.

[40] Yoon, K., S. Suh, J. Ji, J. Cai, and B. Wang, 2010, Stability and speedup issues in TTI RTM implementation: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3221–3225.

[41] Zhan, G., R. C. Pestana, and P. L. Stoffa, 2012, Decoupled equations for reverse time migration in tilted transversely isotropic media: Geophysics, **77**, T37–T45.

[42] Zhang, H., G. Zhang, and Y. Zhang, 2009, Removing s-wave noise in TTI reverse time migration: 79th Ann. Internat. Mtg., Soc. Expl. Geophys., 2849–2853.

[43] Zhang, H., and Y. Zhang, 2008, Reverse time migration in 3D heterogeneous TTI media: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2196–2200.

Accelerated plane-wave destruction

Zhonghuan Chen, Sergey Fomel†, and Wenkai Lu**

ABSTRACT

When plane-wave destruction (PWD) is implemented by implicit finite differences, the local slope is estimated by an iterative algorithm. We propose an analytical estimator of the local slope that is based on convergence analysis of the iterative algorithm. Using the analytical estimator, we design a noniterative method to estimate slopes by a three-point PWD filter. Compared with the iterative estimation, the proposed method needs only one regularization step, which reduces computation time significantly. With directional decoupling of the plane-wave filter, the proposed algorithm is also applicable to 3D slope estimation. We present both synthetic and field experiments to demonstrate that the proposed algorithm can yield a correct estimation result with shorter computational time.

INTRODUCTION

Local slope fields have been widely used in geophysical applications, such as wavefield separation and denoising (14; 11), antialiased seismic interpolation (1), seislet transform (12), velocity-independent NMO correction and imaging (9; 5), predictive painting (10), seismic attribute analysis (16), etc.

Several tools exist for local slope estimation: local slant stack (17; 14), complex trace analysis (2), multiwindow dip search (15), local structure tensor (6; 13), and plane-wave destruction (3; 7). Plane-wave destruction (PWD) approximates the local wave-field by a local plane wave, and models it using a linear differential equation (3).

When plane-wave destruction is applied on discrete sampled seismic signals, the corresponding differential equation needs to be discretized by finite differences. (3) used explicit finite differences. In this method, plane-wave approximation of the wavefield can be seen as applying a linear finite impulse reponse (FIR) filter to the wavefield. Slope estimation is equivalent to estimating a parameter of the FIR filter. A least-squares estimator of the local slope can be obtained by minimizing the prediction error of the filter. To improve estimation performance of the explicit finite difference scheme, (18) proposed total least-squares estimation.

The implicit finite difference scheme was applied to the differential equation by (7). Using an infinite impulse response (IIR) filter, known as the *Thiran allpass filter*

(19), to approximate the phase-shift operator, the plane-wave destruction equation becomes a nonlinear equation of the slope. An iterative algorithm was designed to estimate the slope. In order to improve stability in the iterative algorithm, a smoothing regularization (8) of the increment can be applied at each iteration. Iterations of regularization can be time consuming, however, particularly in the 3D case.

In this paper, we prove the fact that the plane-wave destruction equation is a polynomial equation of an unkown slope. In the case of a three-point approximation of *Thiran's filter*, the convergence results of the iterative algorithm can be analytically analyzed. In this case, we obtain an analytical estimator of the local slope and show that the smoothing regularization can be applied on the final estimator only once. This approach reduces the computational time significantly. We present both 2D and 3D examples, which demonstrate that the proposed algorithm can obtain a slope-estimation result faster than the iterative algorithm, with a similar or even better accuracy.

THEORY

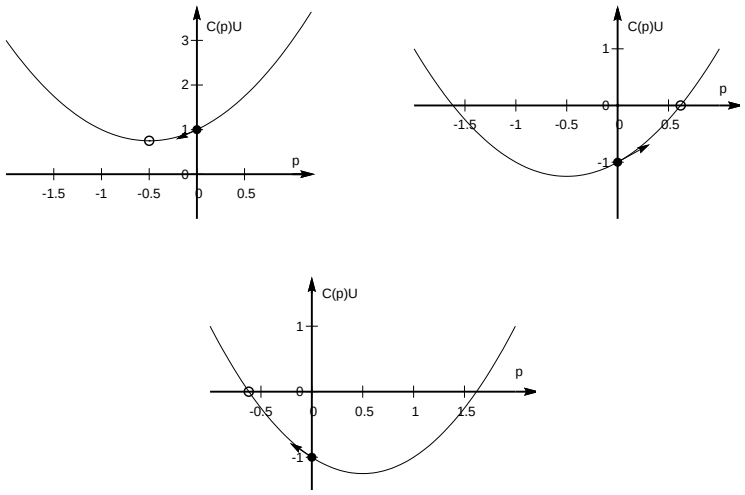


Figure 1: The iterative results of Newton's algorithm. The initial point is $p_0 = 0$ (solid circle), and the convergence result is marked by a blank circle: (a) when $D \leq 0$, (b) when $D > 0$ and $a_1 a_2 > 0$, (c) when $D > 0$ and $a_1 a_2 < 0$. fpwd/Gnuplot iter0,iter1,iter2

Review of PWD

The local plane wave can be represented by the following differential equation (3):

$$\frac{\partial u}{\partial x} + \sigma \frac{\partial u}{\partial t} = 0 \,, \tag{1}$$

where σ is the local slope in continuous space, with dimension time/length. The wavefields observed at the two positions x_1, x_2 have a time delay which is proportional to their distance, $\sigma|x_1 - x_2|$. In the sampled system with space and time intervals Δx and Δt, we define the discrete space slope in the unit of $\Delta t/\Delta x$, as $p = \sigma \Delta x/\Delta t$. As p is independent of the sampling interval, it can be directly used in irregular dataset (in this case, the unit of the slopes becomes space variant). The time delay between two adjacent positions is then the slope $p\Delta t$:

$$u(x,t) = u(x + \Delta x, t + p\Delta t) \; . \tag{2}$$

With the Z transform applied along both time and space directions, the above equation becomes

$$(1 - Z_x Z_t^p) U(Z_x, Z_t) = 0 \; , \tag{3}$$

where Z_t is the unit time-shift operator, Z_x denotes the unit space-shift operator and $U(Z_x, Z_t)$ is the Z transform of $u(x,t)$. The operator $1 - Z_x Z_t^p$ is the plane-wave destructor. Using *Thiran's fractional delay filter* $H(Z_t) = \dfrac{B(1/Z_t)}{B(Z_t)}$ (19) to approximate the time-shift operator $Z_t^p = e^{j\omega p}$, where ω is the circular frequency, the plane-wave destructor can be expressed as (7),

$$C(p) = B(Z_t) - Z_x B(\frac{1}{Z_t}), \tag{4}$$

where

$$B(Z_t) = \sum_{k=-N}^{N} b_k(p) Z_t^{-k}, \tag{5}$$

N is the order of the noncausal temporal filter and $b_k(p)$ are functions of the local slope p.

Equation 3 is a 2D filter. Applying the filter at an arbitrary point in the wavefield, the plane-wave destruction equation 3 becomes a nonlinear equation for the local slope p:

$$C(p, Z_x, Z_t) U(Z_x, Z_t) \approx 0 \; . \tag{6}$$

An iterative method, such as Newton's method, can be applied to find the slope. In practice, wavefields are polluted by noise and the plane wave assumption may not hold true where faults and conflicting boundaries exist. To obtain a stable slope estimation, an additional smoothing regularization process (8) is needed at each step. The total computational cost of slope estimation by plane-wave destruction becomes $O(N_d N_f N_l N_n)$, where N_d is the size of the data, $N_f = 2N + 1$ is the size of the filter, N_l is the number of linear iterations for regularization, and N_n is the number of nonlinear iterations for solving equation 6. Typical values are $N_f = 3, 5$, $N_l =$ 10-50, and $N_n =$ 5-10.

Accelerated PWD

Gauss-Newton's iteration searches the solutions for nonlinear equation 6 as follows: Let p_k be the estimated slope at step k, with estimating error (or destructive error) $e_k = C(p_k)U(Z_x, Z_t)$. In order to find the correct solution p_{k+1} that minimizes e_{k+1}, we need to find the increment Δp_k from the local linearization:

$$e_{k+1} = C(p_k)U(Z_x, Z_t) + C'(p_k)U(Z_x, Z_t)\Delta p_k \approx 0 \ , \tag{7}$$

where $C'(p_k)$ is the derivative of $C(p)$ at p_k with respect to p.

The iterative algorithm stops when a stationary point or a root of $C(p)U$ is reached. They are:

1. Points where $C'(p_k)U = 0$: When p_k satisfies $C'(p)U = 0$, then $e_{k+1} = e_k$, and the Δp_k dependency in equation 7 is removed, stopping further iterations on p_k.
2. Points where $C(p_k)U = 0$ and $C'(p)U \neq 0$: In this case, $\Delta p_k = 0$, thus $p_{k+1} = p_k$, eliminating the need for further improvements on p_k.

The iterative algorithm for equation 6 may converge at different points, depending on the initial point that we chose; $p_0 = 0$ is a common practical choice for the initial solution. In this case, the iterative algorithm may converge to the least absolute root, which denotes the event with smallest dip angle.

In order to analyze the convergence results, the maximally flat fractional delay filter (19; 20) is designed with polynomial coefficients:

$$b_k(p) = \frac{(2N)!(2N)!}{(4N)!(N+k)!(N-k)!} \prod_{m=0}^{N-1-k} (m - 2N + p) \prod_{m=0}^{N-1+k} (m - 2N - p). \tag{8}$$

Details on how to design the filter can be found in the Appendix.

Since $b_k(p)$ is a polynomial of p, expanding it, we get $b_k(p) = \sum_{i=0}^{2N} c_{ki} p^i$ and

$$B(Z_t, p) = \sum_{k=-N}^{N} \sum_{i=0}^{2N} c_{ki} Z_t^{-k} p^i \ . \tag{9}$$

From equation 5, it is obvious that $b_k(p) = b_{-k}(-p)$, therefore $c_{ki} = (-1)^i c_{-k,i}$ and

$$B(Z_t, p) = B(\frac{1}{Z_t}, -p). \tag{10}$$

Substituting the above two equations, the nonlinear equation 6 becomes a $2N$-th degree polynomial equation for p:

$$\sum_{i=0}^{2N} a_i p^i = 0, \tag{11}$$

and the coefficients of the polynomial plane-wave destruction can be expressed as

$$a_i = [1 - (-1)^i Z_x] \sum_{k=-N}^{N} c_{ki} Z_t^{-k} U, \tag{12}$$

which says that the coefficients of the polynomial PWD can be obtained by applying a 2D filter on the wavefield u. Moreover, the 2D filter can be decoupled into the cascade of two 1D directional filters: the temporal filter $\sum_{k=-N}^{N} c_{ki} Z_t^{-k}$ and the spatial filter $1 - (-1)^i Z_x$.

In the special case of $N = 1$, we get a three-point approximation of $B(Z_t)$. It takes the following form (7):

$$B(Z_t) = \frac{(1+p)(2+p)}{12} Z_t^{-1} + \frac{(2+p)(2-p)}{6} + \frac{(1-p)(2-p)}{12} Z_t \ . \tag{13}$$

The plane-wave destruction equation 11 is a quadratic equation. The coefficients $a_i (i = 0, 1, 2)$ can be solved for and expressed as

$$a_i = \frac{1}{12}[1 - (-1)^i Z_x] v_i \ , \tag{14}$$

where v_i are outputs of the following three-point temporal filters:

$$v_0 = 2(Z_t^{-1} + 4 + Z_t)U \ , \tag{15}$$
$$v_1 = 3(Z_t - Z_t^{-1})U \ , \tag{16}$$
$$v_2 = (Z_t^{-1} - 2 + Z_t)U \ . \tag{17}$$

In this case, the quadratic plane-wave destruction equation 11 has one stationary point and two roots, which can be analytically expressed as: $\left\{ \frac{-a_1}{2a_2}, \quad \frac{-a_1 \pm \sqrt{D}}{2a_2} \right\}$, where $D = a_1^2 - 4a_0 a_2$.

The plots in Figure 1 show the convergence process of the iterative algorithm when we choose $p_0 = 0$ as the starting value. Geometrically when $D \leq 0$, the iteration converges to the stationary point $\frac{-a_1}{2a_2}$, as shown in Figure 1a. When $D > 0$, it converges to the least absolute solution of equation 11. Figure 1b and 1c shows the convergence process to the least absolute solution in different cases. We can summarize the convergence result of the iterative algorithm as follows:

$$p = \begin{cases} \dfrac{-a_1}{2a_2} & D \leq 0 \\ \dfrac{-2a_0}{a_1 - \sqrt{D}} & D > 0, \ a_1 < 0 \\ \dfrac{-2a_0}{a_1 + \sqrt{D}} & D > 0, \ a_1 > 0 \end{cases} . \tag{18}$$

As $\frac{-2a_0}{a_1 \pm \sqrt{D}} = \frac{-a_1 \pm \sqrt{D}}{2a_2}$ in the above equation, we use $\frac{-2a_0}{a_1 \pm \sqrt{D}}$ instead of $\frac{-a_1 \pm \sqrt{D}}{2a_2}$ to obtain better numerical stability.

When the data is polluted by noise, in order to obtain a robust slope estimation, we can combine the equations in a local window into the following equation set:

$$\mathbf{F}\mathbf{p} \approx \mathbf{g}, \tag{19}$$

where $\mathbf{F}$ is a normalized diagonal matrix and $\mathbf{g}$ is a vector. Their elements are denuminators and numerators of equation 18 respectively. When we are solving the above equation set by least squares, we can use Tikhonov's regularization (7) or the shaping regularization (8, equation 13) to obtain a smooth solution as follows

$$\mathbf{p} = \mathbf{H}[\mathbf{I} + \mathbf{H}^T(\mathbf{F}^T\mathbf{F} - \mathbf{I})\mathbf{H}]^{-1}\mathbf{H}^T\mathbf{F}^T\mathbf{g}\ , \tag{20}$$

where $\mathbf{H}$ is an appropriate smoothing operator. In this case, the regularization runs only once, therefore the computational cost is reduced to $O(N_d N_f N_l)$.

In 3D applications, there are two polynomial PWD equations for inline and crossline slopes separately. Note that, using the decoupling, inline and crossline slope estimations can share the temporal filtering results in equations 15−17. We can obtain the coefficients of the crossline plane-wave destruction equation as

$$a_i = \frac{1}{12}[1 - (-1)^i Z_y]v_i \tag{21}$$

The five-point or longer approximations of $B(Z_t)$ can achieve higher accuracy. Equation 6 in this case becomes a higher-order polynomial equation (see details in the Appendix), which can be solved numerically. However, there are multiple stationary points, and it is difficult to determine the right one analytically. For applications that need five-point or higher accuracy, we suggest obtaining an initial slope estimation by the proposed three-point method and using it to make the iterative algorithm converge faster (to decrease N_n).

EXAMPLES

Synthetic examples

To test the performance of the proposed slope-estimation method, we generated a harmonic wave field with constant slope $p_0 = 0.3$ shown in Figure 5. We added different scales of additive white Gaussian noise (AWGN) to the wave field and estimate the slope by the proposed method. To compare with the iterative algorithm by (7), the mean square error (MSE) is used as the criterion:

$$\mathrm{MSE}(p) = \mathrm{E}\{(p - p_0)^2\}. \tag{22}$$

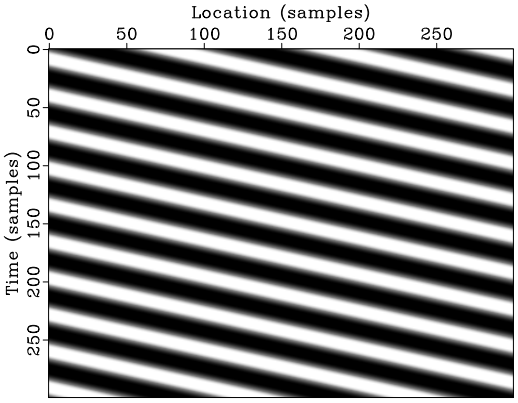


Figure 2: Harmonic waves with constant slope $p_0 = 0.3$. fpwd/const modl

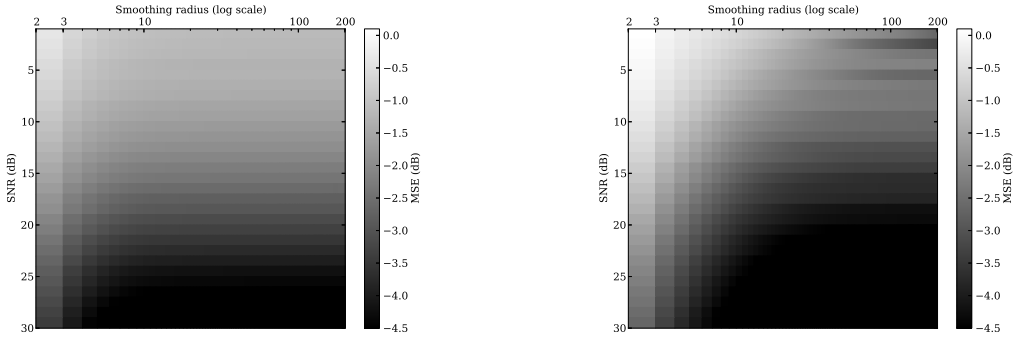


Figure 3: Mean square errors of of the slope estimations by the proposed method (a) and the three-point ($N = 1$) iterative method (b). fpwd/const mse-fdip,mse-dip

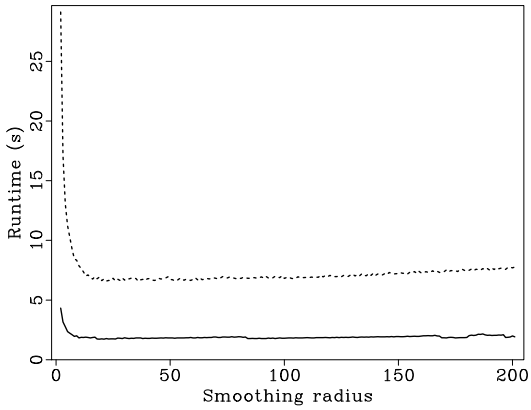


Figure 4: Run time of the proposed method (solid line) and the three-point ($N = 1$) iterative method (dash line). fpwd/const runtime

We use five iterations in the iterative method, $N_n = 5$. For constant slope model, using a large smoothing window, the smoothing regularization can converge faster (with less N_l) and we can obtain a better estimation accuracy. For each methods, we try the smoothing windows from 2 to 200 and show the mean square errors in Figure 3. Compared with the iterative method, the proposed method has better accuracy at the left upper (low SNR and small smoothing window) and worse accuracy at right bottom corner (high SNR and large smoothing window).

We show the total runtime of all the noise scale data in Figure 4. For all smoothing windows in the regularization, the proposed method (solid line) only uses about one fifth run time of the three-point ($N = 1$) iterative method (dash line).

A more complicated model from (4) and (7) is shown in Figure 5a. It has variable slopes in synclines, anticlines, and faults. The slope estimated by the proposed method is shown in Figure 5b. The estimation takes about 20 ms. The three-point ($N = 1$) iterative method can obtain a similar estimation (shown in Figure 5c) by five iterations, which takes about 130ms. Both methods use a 4-point smoothing window in the regularization, but the proposed method obtains a slightly smoother estimation. In Figure 5d we show the faults detected by the residuals of the proposed plane-wave destruction.

Application

The 2D seislet transform (12) uses local slopes to predict and update even and odd traces in the wavelet lifting scheme. The seislet transform itself is fast, but the slope estimation step is comparatively slow. The 3D seislet transform can be constructed in the same way by using 3D slopes and cascading 2D transforms in inline and crossline directions. So that an efficient transform can be built, the proposed accelerated plane-wave destruction is applied to slope estimation.

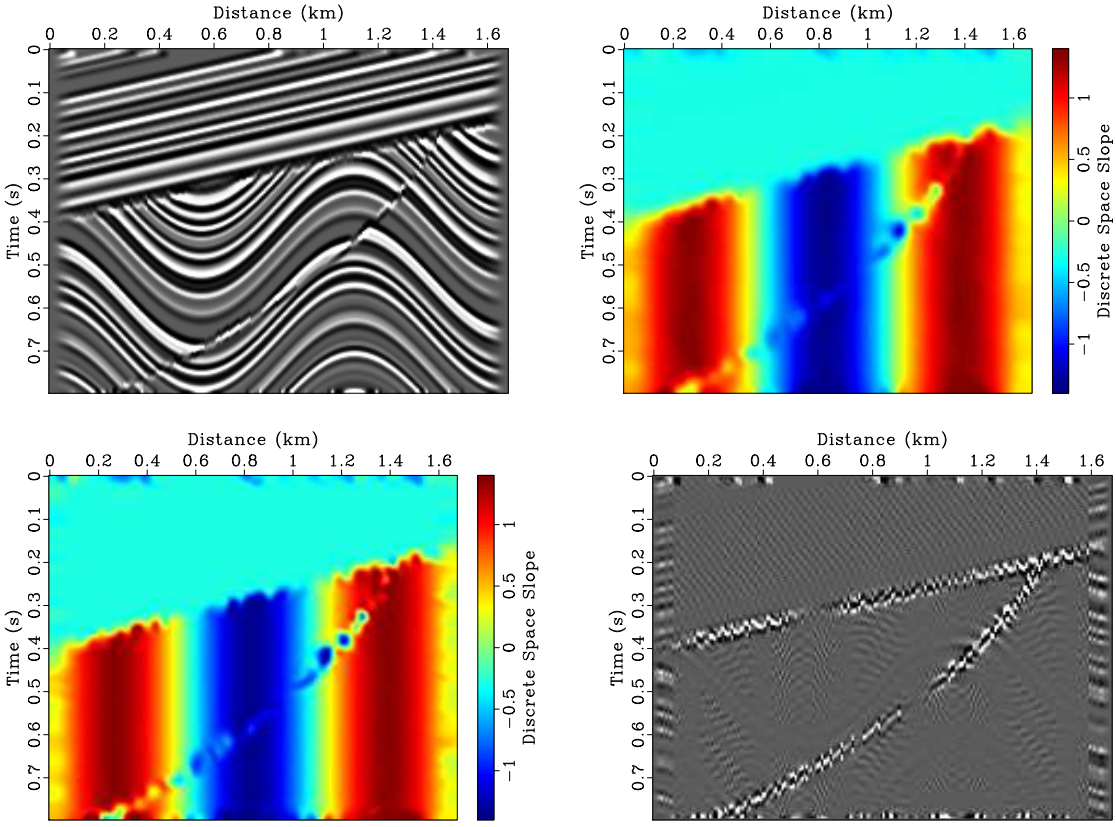


Figure 5: 2D slope estimation example: (a) synthetic data, (b) slope field estimated by the proposed algorithm, (c) slope field estimated by the iterative algorithm after five iterations, (d) faults detection by plane-wave destruction with the estimated slope. fpwd/sigmoid sigmoid,fdip,dip,pwd-fdip

Figure 6 shows a part of the Teapot Dome image from Wyoming. The 3D slope estimation yields two slope fields along two space directions. Figure 7 shows the local inline and crossline slope fields estimated by the proposed algorithm.

The two slopes are used by the lifting scheme in the seislet transform to obtain the seislet transform coefficients. The coefficients of the 2D seislet transform are concentrated at the planes near the zero inline plane, as shown in Figure 8a, whereas in Figure 8b, the 3D transform coefficients are concentrated in the corner region near the origin.

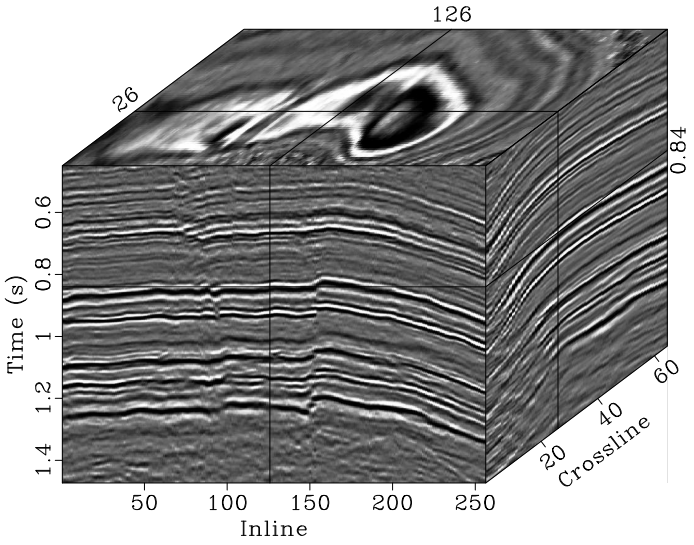


Figure 6: A portion of the 3D data from the Teapot dataset. fpwd/teapot cuber

To illustrate the compressive performance of the 3D seislet transform, Figure 9 shows the reconstruction results at different percentages of the compression. Most of the data can be reconstructed using only 1% of the seislet coefficients (1:100 compression ratio). In order to compare with the iterative methods quantitatively, we define the following normalized cross-correlation (NCC)

$$NCC(\mathbf{x},\mathbf{y}) = \frac{\mathbf{x}^T\mathbf{y}}{\|\mathbf{x}\|_2\|\mathbf{y}\|_2}. \tag{23}$$

In the seislet transform, the more accurate dip we use, the better compression we can obtain. The NCC between the reconstructed data and the original data can be used to quantify the compressive performance. In Table 1, the proposed method is compared with both three-point ($N = 1$) and five-point ($N = 2$) iterative methods.

In all the three methods, we use ten-point smoothing windows in both inline and crossline dimensions, and use five-point smoothing window in time direction. In order to obtain a similar slope estimation as the proposed method, the three-point ($N = 1$) and five-point ($N = 2$) iterative mathods need about six and five iterations respectively. The five-point method can obtain a better compression ratio than the

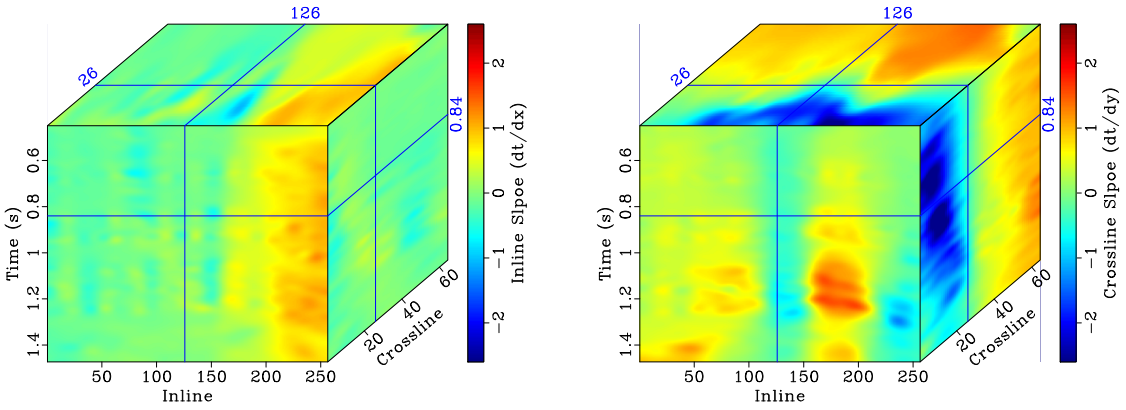


Figure 7: 3D slopes estimated by the proposed algorithm: (a) inline slope, (b) crossline slope. fpwd/teapot fdip1,fdip2

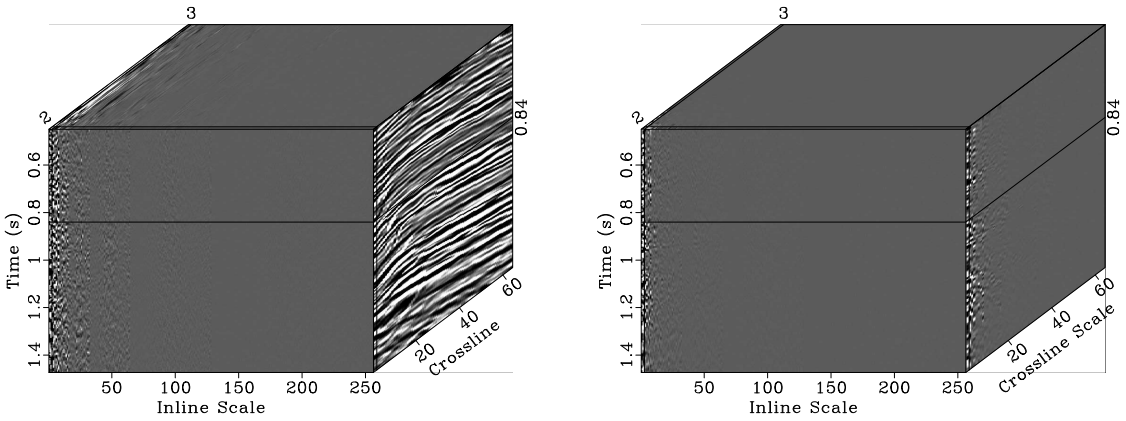


Figure 8: Coefficients of: (a) the 2D seislet transform along inline direction only, (b) the 3D seislet transform. fpwd/teapot fseis1,fseis2

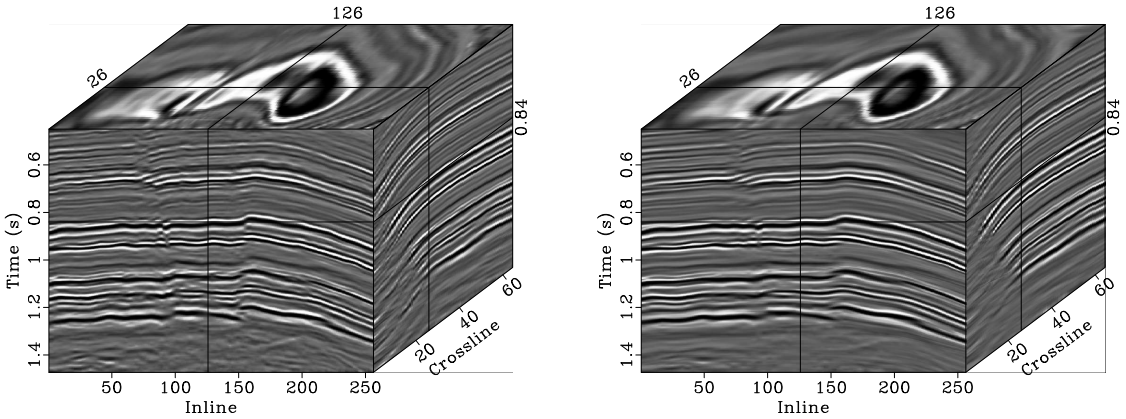


Figure 9: Reconstruction of 3D-seislet-compressed data using: (a) 5% of the seislet coefficients, (b) 1% of the seislet coefficients. fpwd/teapot fapprox2,fapprox1

three-point method, because of its better accuracy in slope estimation. However, although the iterative methods have smaller PWD residuals, neighter of them achieves a better NCC than the proposed method. In this case, using the noniterative estimation as the initial in five-point estimation can save at least 250s. That is to say, the computational time cost is reduced by a factor of about 6.

Method	N	Iterations	Runtime (s)	RES-inline	RES-xline	NCC-1	NCC-5
noniterative	1	0	54.63	0.2314	0.2242	0.8858	0.9618
iterative	1	6	334.5	0.2267	0.2172	0.8814	0.961
iterative	2	5	301.6	0.2194	0.2075	0.8838	0.9615

Table 1: Performance of different dip estimation methods in 3D seislet transform: Runtime is the run time of the slope estimation process; RES-inline is the inline residual; RES-xline is the crossline residual; NCC-1 is the normalized cross-correlation between the original data and the reconstruction using one percent of the seislet coefficients; NCC-5 uses five percents of the coefficients. The run time for 2D and 3D seislet transform are about 22.45 s and 45.0 s respectively.

CONCLUSIONS

In this paper, we derived an analytical estimator of the local slope in three-point implicit plane-wave destruction. On the basis of this result, we built an accelerated slope estimation algorithm. Examples show that the proposed method can produce a result that is similar to that of the iterative algorithm, at a reduced computation time.

Two or more conflicting slopes can be estimated simultaneously by the iterative algorithm. In this case, the PWD equation is a multi-variable polynomial, and the convergence analysis becomes complicated. The proposed noniterative method is not yet suitable for multi-slope estimation. However, we believe that it can find many applications in situations where one dominant slope is sufficient.

ACKNOWLEDGMENTS

We thank Alexander Klokov for his helpful technical suggestions. We also thank the reviewers and the editors for their careful review of the paper.

The Teapot Dome data is provided by the Rocky Mountain Oilfield Testing Center, sponsored by the U.S. Department of Energy. Zhonghuan Chen's joint research is sponsored by China Scholarship Council and China State Key Science and Technology Projects (2011ZX05023-005-007).

APPENDIX A: POLYNOMIAL FORM OF PWD

If all the coefficients of $B(Z_t)$ are polynomials of p, equation 3 is also a polynomial of p, and the plane-wave destruction equation becomes in turn a polynomial equation of p. The problem is to design a $2N+1$ points filter $B(Z_t)$ with polynomial coefficients such that the allpass system $H(Z_t) = \frac{B(1/Z_t)}{B(Z_t)}$ can approximate the phase-shift operator $Z_t^p = e^{j\omega p}$. Denoting the phase response of the system as $\theta(\omega)$, that is $H(e^{j\omega}) = e^{j\theta(\omega)}$, the group delay of the system is

$$\tau(\omega) = \frac{\partial \theta(\omega)}{\partial \omega}. \tag{A-1}$$

The maximally flat criteria designs a filter with a smoothest phase response. There are $2N$ unknown coefficients in $H(Z_t)$, so we can add $2N$ flat constraints for the first $2N-$th order deviratives of the phase response. It becomes (20, equation 7)

$$\begin{cases} \tau(\omega) = p \\ \dfrac{\partial^n \tau(\omega)}{\partial \omega^n} = 0 \qquad n = 1, 2, \ldots, 2N \end{cases}, \tag{A-2}$$

which is equivalent to the following linear maximally flat conditions (19):

$$\sum_{k=-N}^{N} (d-k)^{2n+1} b_k = 0, \tag{A-3}$$

where $n = 0, 1, \ldots, 2N-1$ and $d = p/2$ is the fractional delay of $B(1/Z_t)$ or $1/B(Z_t)$.

In order to solve b_k from the above equations, (19) used an additional condition $b_0 = 1$, which leads $b_k (k \neq 0)$ to be a fractional function of p. Differently from that, we use the following condition,

$$\sum_{k=-N}^{N} b_k = 1, \tag{A-4}$$

where b_k can be proved to be polynomials of p.

Let vector $\mathbf{b} = [b_0, b_N, \ldots, b_1, b_{-1}, \ldots, b_{-N}]^{\mathrm{T}}$. Combining equations A-3 and A-4, we rewrite them into the following matrix form:

$$\left[\begin{array}{c|ccccccc} 1 & 1 & \ldots & 1 & 1 & \ldots & 1 \\ \hline d & d-N & \ldots & d-1 & d+1 & \ldots & d+N \\ d^3 & (d-N)^3 & \ldots & (d-1)^3 & (d+1)^3 & \ldots & (d+N)^3 \\ \vdots & \vdots & \ldots & \ldots & \ldots & \ldots & \vdots \\ d^{4N-1} & (d-N)^{4N-1} & \ldots & (d-1)^{4N-1} & (d+1)^{4N-1} & \ldots & (d+N)^{4N-1} \end{array}\right] \mathbf{b} = \left[\begin{array}{c} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{array}\right].$$

The matrix on the left side, denoted as $\mathbf{V}$, can be split into four blocks $\left[\begin{array}{c|c} \mathbf{A} & \mathbf{B} \\ \hline \mathbf{C} & \mathbf{D} \end{array}\right]$ as shown above. Following the *lemma of matrix inversion*,

$$\mathbf{V}^{-1} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & -(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} & \mathbf{D}^{-1} + \mathbf{D}^{-1}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1}\mathbf{B}\mathbf{D}^{-1} \end{bmatrix}, \tag{A-5}$$

therefore the coefficients

$$\mathbf{b} = \mathbf{V}^{-1}[1, 0, \ldots, 0]^{\mathrm{T}} = \begin{bmatrix} (\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \\ -\mathbf{D}^{-1}\mathbf{C}(\mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C})^{-1} \end{bmatrix}. \tag{A-6}$$

Let subindex $i = -N, -N+1, \ldots, -1, 1, 2, \ldots, N$ and $x_i = d + i$. Submatrix $\mathbf{D}$ can be expressed as

$$\begin{aligned} \mathbf{D} &= \mathbf{E}\mathbf{X} \\ &= \begin{bmatrix} 1 & \ldots & 1 & 1 & \ldots & 1 \\ (d-N)^2 & \ldots & (d-1)^2 & (d+1)^2 & \ldots & (d+N)^2 \\ (d-N)^4 & \ldots & (d-1)^4 & (d+1)^4 & \ldots & (d+N)^4 \\ \vdots & \ldots & \vdots & \vdots & \ldots & \vdots \\ (d-N)^{4N-2} & \ldots & (d-1)^{4N-2} & (d+1)^{4N-2} & \ldots & (d+N)^{4N-2} \end{bmatrix} \mathrm{diag} \begin{bmatrix} x_{-N} \\ \vdots \\ x_{-1} \\ x_1 \\ \vdots \\ x_N \end{bmatrix}, \end{aligned}$$

so $\mathbf{D}^{-1} = \mathbf{X}^{-1}\mathbf{E}^{-1}$. Denoting $\mathbf{U} = \mathbf{E}^{-1}$ with elements $u_{ij},\ j = 1, 2, \ldots, 2N$, as $\mathbf{E}$ is a Vandermonde matrix, u_{ij} and Lagrange intepolating polynomials have the following relationship:

$$\sum_{j=1}^{2N} u_{ij} x^{2j-2} = \ell_i(x), \tag{A-7}$$

where $i = -N, \ldots, -1, 1, \ldots, N$, and $\ell_i(x)$ is the Lagrange polynomial related to the basis $d + i$,

$$\ell_i(x) = \prod_{-N \le m \le N}^{m \ne i, m \ne 0} \frac{x^2 - (d+m)^2}{(d+i)^2 - (d+m)^2}. \tag{A-8}$$

Substituting the above equation, u_{ij} and x into equation A-7, we can prove equation A-7. It follows that

$$[\mathbf{E}^{-1}\mathbf{C}]_i = d\ell_i(d), \tag{A-9}$$

$$[\mathbf{D}^{-1}\mathbf{C}]_i = [\mathbf{X}^{-1}\mathbf{E}^{-1}\mathbf{C}]_i = \frac{d}{d+i}\ell_i(d), \tag{A-10}$$

with

$$\ell_i(d) = \frac{(-1)^{i+1} N! N!}{(N+i)!(N-i)!} \frac{d+i}{d} \prod_{m=-N}^{N} \frac{2d+m}{2d+m+i}. \tag{A-11}$$

Thus hence

$$\begin{aligned} \mathbf{A} - \mathbf{B}\mathbf{D}^{-1}\mathbf{C} &= \sum_{i=-N}^{N} (-1)^i \frac{N!N!}{(N+i)!(N-i)!} \prod_{m=-N}^{N} \frac{p+m}{p+m+i} \\ &= \frac{(4N)!N!N!}{(2N)!(2N)!} \frac{1}{\prod_{m=N+1}^{2N} (m^2 - p^2)} \end{aligned} \tag{A-12}$$

and

$$[\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}]^{-1}=\frac{(2N)!(2N)!}{(4N)!N!N!}\prod_{m=N+1}^{2N}(m^2-p^2). \tag{A-13}$$

It is the coefficient b_0, a $2N$-th degree polynomial of p. Substituting it into equation A-6, the coefficients at $k=\pm1,\pm2,\cdots\pm N$ are expressed as

$$\begin{aligned} b_k &= -[\mathbf{D}^{-1}\mathbf{C}]_k[\mathbf{A}-\mathbf{B}\mathbf{D}^{-1}\mathbf{C}]^{-1} \\ &= \frac{(2N)!(2N)!}{(4N)!(N+k)!(N-k)!}\prod_{m=0}^{N-1-k}(m-2N+p)\prod_{m=0}^{N-1+k}(m-2N-p). \end{aligned} \tag{A-14}$$

With the additional condition A-4 in $2N+1$ points approximation, all the coefficients are polynomials of p of $2N$-th degree. Thus the plane-wave destruction equation 6 therefore is proved to be a polynomial equation of $2N$-th degree.

REFERENCES

[1] Bardan, V., 1987, Trace interpolation in seismic data processing: Geophysical Prospecting, **35**, 343–358.

[2] Barnes, A. E., 1996, Theory of 2-D complex seismic trace analysis: Geophysics, **61**, 264–272.

[3] Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications.

[4] ———, 1999, Geophysical estimation by example: Geophysical soundings image construction: Stanford Exploration Project.

[5] Cooke, D., A. Bóna, and B. Hansen, 2009, Simultaneous time imaging, velocity estimation, and multiple suppression using local event slopes: Geophysics, **74**, WCA65–WCA73.

[6] Fehmers, G. C., and C. F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, 1286–1293.

[7] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

[8] ———, 2007a, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[9] ———, 2007b, Velocity-independent time-domain seismic imaging using local event slopes: Geophysics, **72**, S139–S147.

[10] ———, 2010, Predictive painting of 3D seismic volumes: Geophysics, **75**, A25–A30.

[11] Fomel, S., E. Landa, and M. T. Taner, 2007, Poststack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.

[12] Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, V25–V38.

[13] Hale, D., 2007, Local dip filtering with directional laplacians: CWP Report, **567**.

[14] Harlan, W. S., J. F. Claerbout, and F. Rocca, 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, 1869–1880.
[15] Marfurt, K. J., 2006, Robust estimates of 3d reflector dip and azimuth: Geophysics, **71**, P29–P40.
[16] Marfurt, K. J., V. Sudhaker, A. Gersztenkorn, K. D. Crawford, and S. E. Nissen, 1999, Coherency calculations in the presence of structural dip: Geophysics, **64**, 104–111.
[17] Ottolini, R., 1983, Signal/noise separation in dip space: SEP Report, **37**.
[18] Schleicher, J., J. C. Costa, L. T. Santos, A. Novais, and M. Tygel, 2009, On the estimation of local slopes: Geophysics, **74**, P25–P33.
[19] Thiran, J., 1971, Recursive digital filters with maximally flat group delay: IEEE Transactions on Circuit Theory, **18**, 659–664.
[20] Zhang, X., 2009, Maxflat fractional delay iir filter design: IEEE Transactions on Signal Processing, **57**, 2950 –2956.

On anelliptic approximations for qP velocities in transversally isotropic media

*Sergey Fomel**

ABSTRACT

I develop a unified approach for approximating phase and group velocities of qP seismic waves in a transversally isotropic medium with the vertical axis of symmetry (VTI). While the exact phase velocity expressions involve four independent parameters to characterize the elastic medium, the proposed approximate expressions use only three parameters. This makes them more convenient for use in surface seismic experiments, where estimation of all the four parameters is problematic. The three-parameter phase-velocity approximation coincides with the previously published "acoustic" approximation of Alkhalifah. The group velocity approximation is 'new and noticeably more accurate than some of the previously published approximations. I demonstrate an application of the group velocity approximation for finite-difference computation of traveltimes.

INTRODUCTION

Anellipticity (deviation from ellipse) is an important characteristic of elastic wave propagation. One of the simplest and yet practically important cases of anellipticity occurs in transversally isotropic media with the vertical axis of symmetry (VTI). In this type of media, the phase velocities of qSH waves and the corresponding wavefronts are elliptic, while the phase and group velocities of qP and qSV waves may exhibit strong anellipticity (53).

The exact expressions for the phase velocities of qP and qSV waves in VTI media involve four independent parameters. However, it has been observed that only three parameters influence wave propagation and are of interest to surface seismic methods (10). Moreover, the exact expressions for the group velocities in terms of the group angle are difficult to obtain and too cumbersome for practical use. This explains the need for developing practical three-parameter approximations for both group and phase velocities in VTI media.

Numerous different successful approximations have been previously developed (16; 21; 10; 4; 7; 40; 47; 57). In this paper, I attempt to construct a unified approach for deriving anelliptic approximations.

*e-mail: sergey.fomel@beg.utexas.edu

The starting point is the anelliptic approximation of Muir (31; 21). Although not the most accurate for immediate practical use, this approximation possesses remarkable theoretical properties. The Muir approximation correctly captures the linear part of anelliptic behavior. It can be applied to find more accurate approximations with nonlinear dependence on the anelliptic parameter. A particular way of "unlinearizing" the linear approximation is the shifted hyperbola approach, familiar from the isotropic approximations in vertically inhomogeneous media (30; 49; 20; 18) and from the theory of Stolt stretch (46; 24). I show that applying this idea to approximate the phase velocity of qP waves leads to the known "acoustic" approximation of (4; 6), derived in a different way. Applying the same approach to approximate the group velocity of qP waves leads to a new remarkably accurate three-parameter approximation.

One practical use for the group velocity approximation is traveltime computations, required for Kirchhoff imaging and tomography. In the last part of the paper, I show examples of finite-difference traveltime computations utilizing the new approximation.

EXACT EXPRESSIONS

Wavefront propagation in the general anisotropic media can be described with the anisotropic eikonal equation

$$v^2\left(\frac{\nabla T}{|\nabla T|},\mathbf{x}\right)\,|\nabla T|^2 = 1\;, \tag{1}$$

where $\mathbf{x}$ is a point in space, $T(\mathbf{x})$ is the traveltime at that point for a given source, and $v(\mathbf{n},\mathbf{x})$ is the *phase velocity* in the phase direction $\mathbf{n}=\frac{\nabla T}{|\nabla T|}$.

In the case of VTI media, the three modes of elastic wave propagation (qSH, qSV, and qP) have the following well-known explicit expressions for the phase velocities (25):

$$v^2_{SH}(\mathbf{n},\mathbf{x}) = m\,\sin^2\theta + l\,\cos^2\theta\;; \tag{2}$$

$$v^2_{SV}(\mathbf{n},\mathbf{x}) = \frac{1}{2}\left[(a+l)\,\sin^2\theta + (c+l)\,\cos^2\theta\right] - \frac{1}{2}\sqrt{\left[(a-l)\,\sin^2\theta - (c-l)\,\cos^2\theta\right]^2 + 4\,(f+l)^2\,\sin^2\theta\,\cos^2\theta}\;; \tag{3}$$

$$v^2_{P}(\mathbf{n},\mathbf{x}) = \frac{1}{2}\left[(a+l)\,\sin^2\theta + (c+l)\,\cos^2\theta\right] + \frac{1}{2}\sqrt{\left[(a-l)\,\sin^2\theta - (c-l)\,\cos^2\theta\right]^2 + 4\,(f+l)^2\,\sin^2\theta\,\cos^2\theta}\;, \tag{4}$$

where, in the notation of (11) and (12), $a=c_{11}$, $c=c_{33}$, $f=c_{13}$, $l=c_{55}$, $m=c_{66}$, $c_{ij}(\mathbf{x})$ are the density-normalized components of the elastic tensor, and θ is the phase angle between the phase direction $\mathbf{n}$ and the axis of symmetry.

The group velocity describes the propagation of individual ray trajectories $\mathbf{x}(\tau)$. It can be determined from the phase velocity using the general expression

$$\mathbf{V} = \frac{d\mathbf{x}}{d\tau} = v\mathbf{n} + \left(\mathbf{I} - \mathbf{n}\,\mathbf{n}^T\right)\nabla_{\mathbf{n}} v \;, \tag{5}$$

where $\mathbf{I}$ denotes the identity matrix, $\mathbf{n}^T$ stands for the transpose of $\mathbf{n}$, and $\nabla_{\mathbf{n}} v$ is the gradient of v with respect to $\mathbf{n}$. The two terms in equation (11) are clearly orthogonal to each other. Therefore, the group velocity magnitude is (35; 12; 15)

$$V = |\mathbf{V}| = \sqrt{v^2 + v_\theta^2}\;, \tag{6}$$

where

$$v_\theta^2 = \left|\left(\mathbf{I} - \mathbf{n}\,\mathbf{n}^T\right)\nabla_{\mathbf{n}} v\right|^2 = |\nabla_{\mathbf{n}} v|^2 - |\mathbf{n}\cdot\nabla_{\mathbf{n}} v|^2 \;\; . \tag{7}$$

The group velocity has a particularly simple form in the case of elliptic anisotropy. Specifically, the phase velocity squared has the quadratic form

$$v_{\rm ell}^2(\mathbf{n},\mathbf{x}) = \mathbf{n}^T\,\mathbf{A}(\mathbf{x})\,\mathbf{n} \tag{8}$$

with a symmetric positive-definite matrix $\mathbf{A}$, and the group velocity is

$$\mathbf{V}_{\rm ell} = \mathbf{A}\,\mathbf{p}\;, \tag{9}$$

where $\mathbf{p} = \nabla T = \mathbf{n}/v(\mathbf{n},\mathbf{x})$. The corresponding group slowness squared has the explicit expression

$$\frac{1}{V_{\rm ell}^2(\mathbf{N},\mathbf{x})} = \mathbf{N}^T\,\mathbf{A}^{-1}(\mathbf{x})\,\mathbf{N}\;, \tag{10}$$

where $\mathbf{N}$ is the group direction, and $\mathbf{A}^{-1}$ is the matrix inverse of $\mathbf{A}$. For example, the elliptic expression (2) for the phase velocity of qSH waves in VTI media transforms into a completely analogous expression for the group slowness

$$\frac{1}{V_{SH}^2(\mathbf{N},\mathbf{x})} = M\,\sin^2\Theta + L\,\cos^2\Theta \tag{11}$$

where $M = 1/m$, $L = 1/l$, and Θ is the angle between the group direction $\mathbf{N}$ and the axis of symmetry.

The situation is more complicated in the anelliptic case. Figure 1 shows the qP and qSV phase velocity profiles in a transversely isotropic material – Greenhorn shale (27), which has the parameters $a = 14.47\,\mathrm{km}^2/\mathrm{s}^2$, $l = 2.28\,\mathrm{km}^2/\mathrm{s}^2$, $c = 9.57\,\mathrm{km}^2/\mathrm{s}^2$, and $f = 4.51\,\mathrm{km}^2/\mathrm{s}^2$. Figure 2 shows the corresponding group velocity profiles. The non-convexity of the qSV phase velocity causes a multi-valued (triplicated) group velocity profile. The shapes of all the surfaces are clearly anelliptic.

A simple model of anellipticity is suggested by the Muir approximation (31; 21), reviewed in the next section.

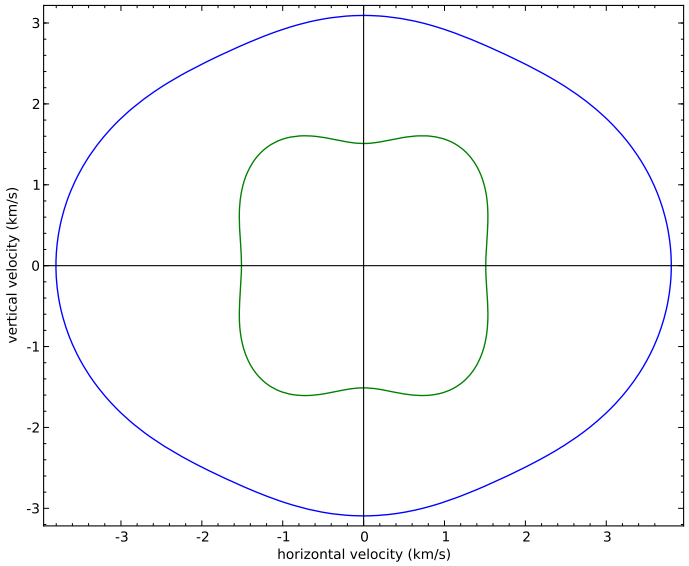


Figure 1: Phase velocity profiles for qP (outer curve) and qSV (inner curve) waves in a transversely isotropic material (Greenhorn shale). group/Sage exph

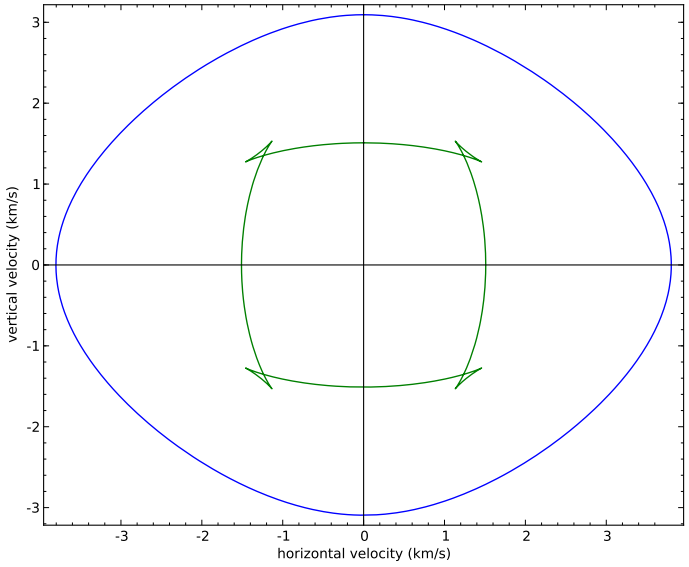


Figure 2: Group velocity profiles for qP (outer curve) and qSV (inner curve) waves in a transversely isotropic material (Greenhorn shale). group/Sage exgr

MUIR APPROXIMATION

(31) suggested representing anelliptic qP phase velocities with the following approximation:

$$v_P^2(\theta) \approx e(\theta) + \frac{(q-1)\,a\,c\,\sin^2\theta\,\cos^2\theta}{e(\theta)}\ , \tag{12}$$

where $e(\theta)$ is the elliptical part of the velocity, defined by

$$e(\theta) = a\,\sin^2\theta + c\,\cos^2\theta\ , \tag{13}$$

and q is the anellipticity coefficient ($q = 1$ in case of elliptic velocities). Approximation (12) uses only three parameters to characterize the medium (a, c, and q) as opposed to the four parameters (a, c, l, and f) in the exact expression.

There is some freedom in choosing an appropriate value for the coefficient q. Assuming near-vertical wave propagation and the vertical axis of symmetry (a VTI medium) and fitting the curvature ($d^2v_P/d\theta^2$) of the exact phase velocity (4) near the vertical phase angle ($\theta = 0$), leads to the definition (21)

$$q = \frac{l\,(c-l) + (l+f)^2}{a\,(c-l)}\ . \tag{14}$$

In terms of Thomsen's elastic parameters ϵ and δ (51) and the elastic parameter η of (10),

$$q = \frac{1+2\,\delta}{1+2\epsilon} = \frac{1}{1+2\,\eta}\ . \tag{15}$$

This confirms the direct relationship between η and anellipticity. If we were to fit the phase velocity curvature near the horizontal axis $\theta = \pi/2$ (perpendicular to the axis of symmetry), the appropriate value for q would be

$$\hat{q} = \frac{l\,(a-l) + (l+f)^2}{c\,(a-l)}\ . \tag{16}$$

(31) also suggested approximating the VTI group velocity with an analogous expression

$$\frac{1}{V_P^2(\Theta)} \approx E(\Theta) + \frac{(Q-1)\,A\,C\,\sin^2\Theta\,\cos^2\Theta}{E(\Theta)} \tag{17}$$

where $A = 1/a$, $C = 1/c$, $Q = 1/q$, Θ is the group angle, and $E(\Theta)$ is the elliptical part:

$$E(\Theta) = A\,\sin^2\Theta + C\,\cos^2\Theta\ . \tag{18}$$

Equations (12) and (17) are consistent in the sense that both of them are exact for elliptic anisotropy ($q = Q = 1$) and accurate to the first order in $(q-1)$ or $(Q-1)$ in the general case of transversally isotropic media.

To the same approximation order, the connection between the phase and group directions is

$$\tan\Theta = \tan\theta\,\frac{a}{c}\,\left(1 - (q-1)\,\frac{a\,\sin^2\theta - c\,\cos^2\theta}{a\,\sin^2\theta + c\,\cos^2\theta}\right)\ . \tag{19}$$

SHIFTED HYPERBOLA APPROXIMATION FOR THE PHASE VELOCITY

Despite the beautiful symmetry of Muir's approximations (12) and (17), they are less accurate in practice than some other approximations, most notably the weak anisotropy approximation of (51), which can be written as (52)

$$v_P^2(\theta) \approx c \left(1 + 2\,\epsilon \sin^4\theta + 2\,\delta \sin^2\theta \cos^2\theta\right) , \tag{20}$$

where

$$\epsilon = \frac{a-c}{2\,c} \quad \text{and} \quad \delta = \frac{(l+f)^2 - (c-l)^2}{2\,c\,(c-l)} . \tag{21}$$

Note that both approximations involve the anellipticity factor ($q-1$ or $\epsilon - \delta$) in a linear fashion. If the anellipticity effect is significant, the accuracy of Muir's equations can be improved by replacing the linear approximation with a nonlinear one. There are, of course, infinitely many nonlinear expressions that share the same linearization. In this study, I focus on the shifted hyperbola approximation, which follows from the fact that an expression of the form

$$x + \frac{\alpha}{x} \tag{22}$$

is the linearization (Taylor series expansion) of the form

$$x\,(1-s) + s\,\sqrt{x^2 + \frac{2\,\alpha}{s}} \tag{23}$$

for small α. Linearization does not depend on the parameter s, which affects only higher-order terms in the Taylor expansion. Expression (23) is reminiscent of the shifted hyperbola approximation for normal moveout in vertically heterogeneous media (30; 49; 20; 18) and the Stolt stretch correction in the frequency-wavenumber migration (46; 24). It is evident that Muir's approximation (12) has exactly the right form (22) to be converted to the shifted hyperbola approximation (23).

Thus, we seek an approximation of the form

$$v_P^2(\theta) \approx e(\theta)\,(1-s) + s\,\sqrt{e^2(\theta) + \frac{2\,(q-1)\,a\,c\,\sin^2\theta\,\cos^2\theta}{s}} \tag{24}$$

with $e(\theta)$ defined by equation (13). The plan is to select a value of the additional parameter s to fit the exact phase velocity expression (4) and then to constrain s so that it depends only on the three parameters already present in the original approximation (12).

One can verify that the velocity curvature $d^2v_P/d\theta^2$ around the vertical axis $\theta = 0$ for approximation (24) depends on the chosen value of q but does not depend on the value of the shift parameter s. This means that the velocity profile $v_P(\theta)$

becomes sensitive to s only further away from the vertical direction. This separation of influence between the approximation parameters is an important and attractive property of the shifted hyperbola approximation. I find an appropriate value for s by fitting additionally the fourth-order derivative $d^4v_P/d\theta^4$ at $\theta = 0$ to the corresponding derivative of the exact expression. The fit is achieved when s has the value

$$s = \frac{c-l}{2}\,\frac{(a-l)\,(c-l)-(l+f)^2}{a\,(c-l)^2-c\,(l+f)^2}\ . \tag{25}$$

It is more instructive to express it in the form

$$s = \frac{1}{2}\,\frac{(a-c)\,(q-1)\,(\hat{q}-1)}{a\,\left(1-\hat{q}-q\,(1-q)\right)-c\,\left((\hat{q}-1)^2+\hat{q}\,(q-\hat{q})\right)}\ , \tag{26}$$

where q and $\hat{q}$ are defined by equations (14) and (16). In this form of the expression, $\hat{q}$ appears as the extra parameter that we need to eliminate. This parameter was defined by fitting the velocity profile curvature around the horizontal axis, which would correspond to infinitely large offsets in a surface seismic experiment. One possible way to constrain it is to set $\hat{q}$ equal to q, which implies that the velocity profile has similar behavior near the vertical and the horizontal axes. Setting $\hat{q} \approx q$ in equation (26) yields

$$s \approx \lim_{\hat{q}\to q} s = \frac{1}{2}\ . \tag{27}$$

Substituting (27) in equation (24) produces the final approximation

$$v_P^2(\theta) \approx \frac{1}{2}\,e(\theta) + \frac{1}{2}\,\sqrt{e^2(\theta)+4\,(q-1)\,a\,c\,\sin^2\theta\,\cos^2\theta}\ . \tag{28}$$

Approximation (28) is exactly equivalent to the *acoustic approximation* of (4; 6), derived with a different set of parameters by formally setting the S-wave velocity ($l = v_S^2$) in equation (4) to zero. A similar approximation is analyzed by (47). Approximation (28) was proved to possess a remarkable accuracy even for large phase angles and significant amounts of anisotropy. Figure 3 compares the accuracy of different approximations using the parameters of the Greenhorn shale. The acoustic approximation appears especially accurate for phase angles up to about 25 degrees and does not exceed the relative error of 0.3% even for larger angles.

SHIFTED HYPERBOLA APPROXIMATION FOR THE GROUP VELOCITY

Similar strategy is applicable for approximating the group velocity. Applying the shifted hyperbola approach to "unlinearize" Muir's approximation (17), we seek an approximation of the form

$$\frac{1}{V_P^2(\Theta)} \approx E(\Theta)\,(1-S) + S\,\sqrt{E^2(\Theta)+\frac{2\,(Q-1)\,A\,C\,\sin^2\Theta\,\cos^2\Theta}{S}} \tag{29}$$

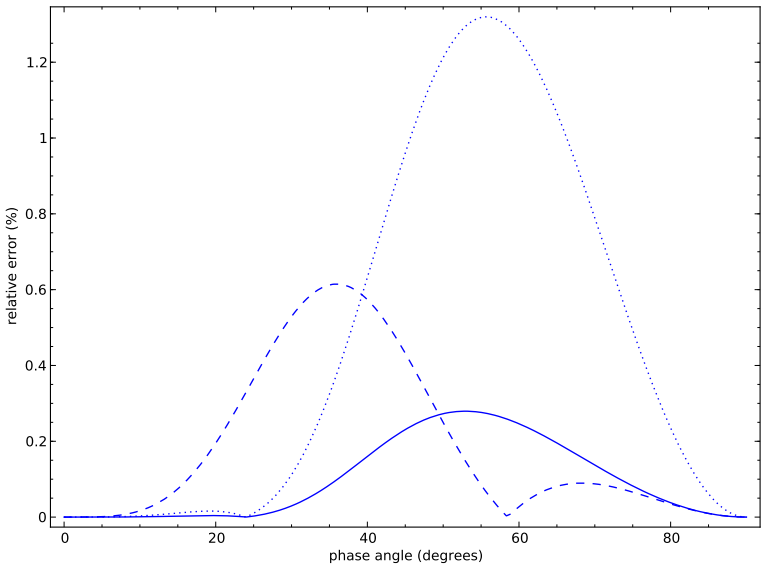


Figure 3: Relative error of different phase velocity approximations for the Greenhorn shale anisotropy. Short dash: Thomsen's weak anisotropy approximation. Long dash: Muir's approximation. Solid line: suggested approximation (similar to Alkhalifah's acoustic approximation.) group/Sage errphp

An approximation of this form with S set to $1/2$ was proposed earlier by (57). Similarly to the case of the phase velocity approximation, I constrain the value of S by Taylor fitting of the velocity profiles near the vertical angle.

Although there is no simple explicit expression for the transversally isotropic group velocity, we can differentiate the parametric representations of V_P and Θ in terms of the phase angle θ that follow from equation (11). The group velocity is an even function of the angle Θ because of the VTI symmetry. Therefore, the odd-order derivatives are zero at the axis of symmetry ($\Theta = \theta = 0$). Fitting the second-order derivative $d^2V_P/d\Theta^2$ at $\theta = 0$ produces $Q = 1/q = 1 + 2\,\eta$, consistent with Muir's approximation (17). Fitting additionally the fourth-order derivative $d^4V_P/d\Theta^4$ at $\theta = 0$ produces

$$S = \frac{1}{2}\,\frac{\left[(l+f)^2 + l\,(c-l)\right]^2\,\left[(c-l)\,(a-l) - (l+f)^2\right]}{a^2\,c\,(c-l)\,(l+f)^2 - \left[l\,(c-l) + (l+f)^2\right]^3} \tag{30}$$

or, equivalently,

$$S = \frac{1}{2}\,\frac{(C-A)\,(Q-1)\,(\hat{Q}-1)}{C\,\left(\hat{Q}\,(Q^2-Q-1)+1\right) + A\,\left(\hat{Q} - Q^3 + Q^2 - 1\right)}\;, \tag{31}$$

where $\hat{Q} = 1/\hat{q}$. As in the previous section, I approximate the optimal value of S by

setting $\hat{Q}$ equal to Q, as follows:

$$S \approx \lim_{\hat{Q}\to Q} S = \frac{1}{2\,(1+Q)} = \frac{1}{4\,(1+\eta)}\ . \tag{32}$$

Selected in this way, the value of S depends on the anelliptic parameter Q (or η) and, for small anellipticity, is close to $1/4$, which is different from the value of $1/2$ in the approximation of (57).

The final group velocity approximation takes the form

$$\frac{1}{V_P^2(\Theta)} \approx \frac{1+2\,Q}{2\,(1+Q)}\,E(\Theta) + \frac{1}{2\,(1+Q)}\,\sqrt{E^2(\Theta) + 4\,(Q^2-1)\,A\,C\,\sin^2\Theta\,\cos^2\Theta}\ . \tag{33}$$

In Figure 4, the accuracy of approximation (33) is compared with the accuracy of Muir's approximation (17) and the accuracy of the weak anisotropy approximation (51) for the elastic parameters of the Greenhorn shale. The weak anisotropy approximation, used in this comparison, is

$$V_P^2(\Theta) \approx c\,\left(1 + 2\,\epsilon\,\sin^4\Theta + 2\,\delta\,\sin^2\Theta\,\cos^2\Theta\right)\ , \tag{34}$$

where ϵ and δ are Thomsen's parameters, defined in equations (21). A similar form (in a different parameterization) was introduced by (16).

Approximation (33) turns out to be remarkably accurate for this example. It appears nearly exact for group angles up to 45 degrees from vertical and does not exceed 0.3% relative error even at larger angles. It is compared with two other approximations in Figure 5. These are the Zhang-Uren approximation (57) and the Alkhalifah-Tsvankin approximation, which follows directly from the normal moveout equation suggested by (10):

$$t^2(x) \approx t_0^2 + \frac{x^2}{V_n^2} - \frac{2\,\eta\,x^4}{V_n^2\,\left[t_0^2\,V_n^2 + (1+2\,\eta)\,x^2\right]}\ , \tag{35}$$

where $t(x)$ is the moveout curve, t_0 is the vertical traveltime, and $V_n = \sqrt{a/(1+2\,\eta)}$ is the NMO velocity. In a homogeneous medium, equation (35) corresponds to the group velocity approximation

$$\frac{1}{V_P^2(\Theta)} \approx \frac{\cos^2\Theta}{V_z^2} + \frac{\sin^2\Theta}{V_n^2} - \frac{2\,\eta\,\sin^4\Theta}{V_n^2\,\left[\cos^2\Theta\,V_n^2/V_z^2 + (1+2\,\eta)\,\sin^2\Theta\right]}\ , \tag{36}$$

where $V_z = \sqrt{c}$. In the notation of this paper, the Alkhalifah-Tsvankin equation (36) takes the form

$$\frac{1}{V_P^2(\Theta)} \approx E(\Theta) + \frac{(Q-1)\,A\,C\,\sin^2\Theta\,\cos^2\Theta}{E(\Theta) + (Q^2-1)\,A\,\sin^2\Theta} \tag{37}$$

and differs from approximation (17) by the correction term in the denominator. Approximation (33) is noticeably more accurate for this example than any of the other approximations considered here.

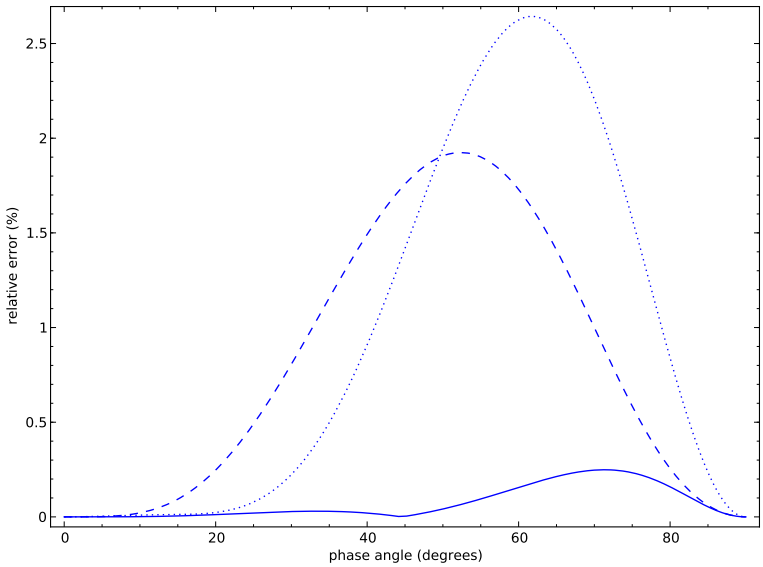


Figure 4: Relative error of different group velocity approximations for the Greenhorn shale anisotropy. Short dash: Thomsen's weak anisotropy approximation. Long dash: Muir's approximation. Solid line: suggested approximation. group/Sage errgrp

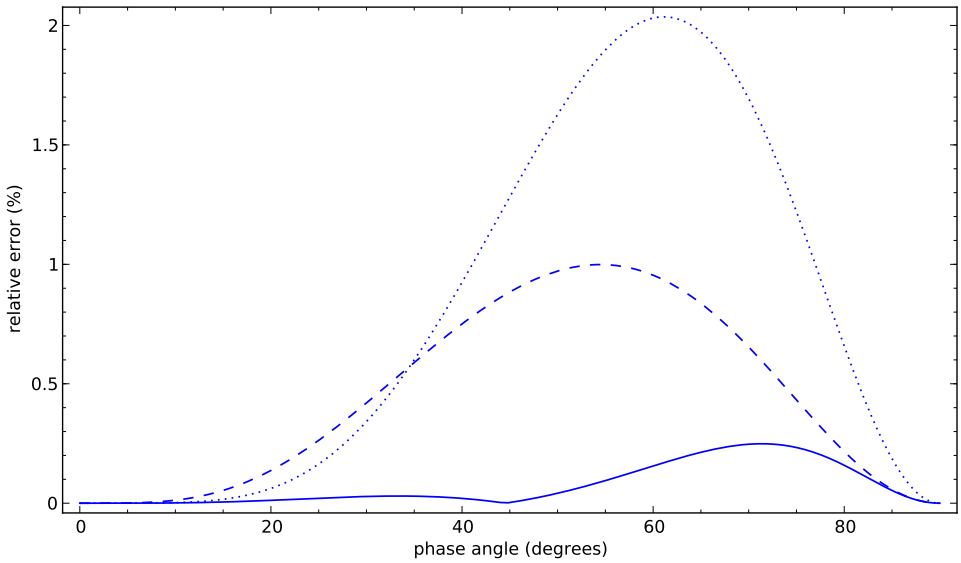


Figure 5: Relative error of different group velocity approximations for the Greenhorn shale anisotropy. Short dash: Alkhalifah-Tsvankin approximation. Long dash: Zhang-Uren approximation. Solid line: suggested approximation. group/Sage errgrp2

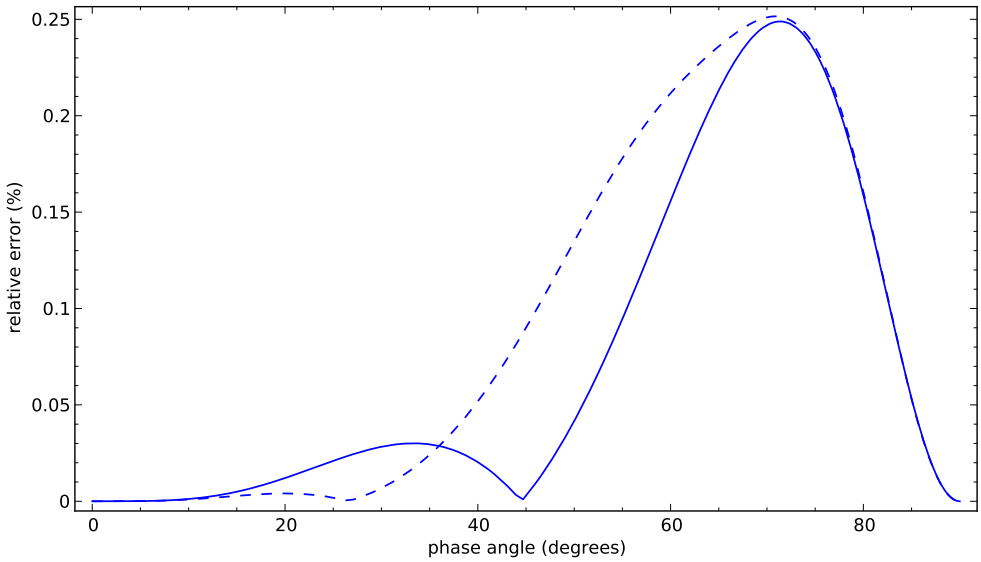


Figure 6: Relative error of different group velocity approximations for the Greenhorn shale anisotropy. Dashed line: Alkhalifah approximation. Solid line: suggested approximation. group/Sage errgrp4

Another accurate group velocity approximation was suggested by (7). However, the analytical expression is complicated and inconvenient for practical use. The accuracy of Alkhalifah's approximation for the Greenhorn shale example is depicted in Figure 6.

It is similarly possible to convert a group velocity approximation into the corresponding moveout equation. In a homogeneous anisotropic medium, the reflection traveltime t as a function of offset x is

$$t(x) = \frac{2\sqrt{(x/2)^2 + z^2}}{V_P\left(\arctan\left(\frac{x}{2\,z}\right)\right)} \;, \tag{38}$$

where $z = t_0\,V_P(0)/2$ is the depth of the reflector. The moveout equation corresponding to approximation (33) is

$$\begin{aligned} t^2(x) &\approx \frac{1+2\,Q}{2\,(1+Q)}\,H(x) + \frac{1}{2\,(1+Q)}\sqrt{H^2(x) + 4\,(Q^2-1)\,\frac{t_0^2\,x^2}{Q\,V_n^2}} \\ &= \frac{3+4\,\eta}{4\,(1+\eta)}\,H(x) + \frac{1}{4\,(1+\eta)}\sqrt{H^2(x) + 16\,\eta\,(1+\eta)\,\frac{t_0^2\,x^2}{(1+2\,\eta)\,V_n^2}} \;, \end{aligned} \tag{39}$$

where $H(x)$ represents the hyperbolic part:

$$H(x) = t_0^2 + \frac{x^2}{Q\,V_n^2} = t_0^2 + \frac{x^2}{(1+2\,\eta)\,V_n^2} \;. \tag{40}$$

For small offsets, the Taylor series expansion of equation (39) is

$$\begin{aligned} t^2(x) &\approx t_0^2 + \frac{x^2}{V_n^2} - (Q-1)\,\frac{x^4}{t_0^2\,V_n^4} + (Q-1)\,(2\,Q^2-1)\,\frac{x^6}{Q\,t_0^4\,V_n^6} + O(x^8) \\ &= t_0^2 + \frac{x^2}{V_n^2} - 2\,\eta\,\frac{x^4}{t_0^2\,V_n^4} + 2\,\eta\,(1+8\,\eta+8\eta^2)\,\frac{x^6}{(1+2\,\eta)\,t_0^4\,V_n^6} + O(x^8)\;. \end{aligned} \tag{41}$$

Figure 7 compares the accuracy of different moveout approximations assuming reflection from the bottom of a homogeneous anisotropic layer of 1 km thickness with the elastic parameters of Greenhorn shale. Approximation (39) appears extremely accurate for half-offsets up to 1 km and does not develop errors greater than 5 ms even at much larger offsets.

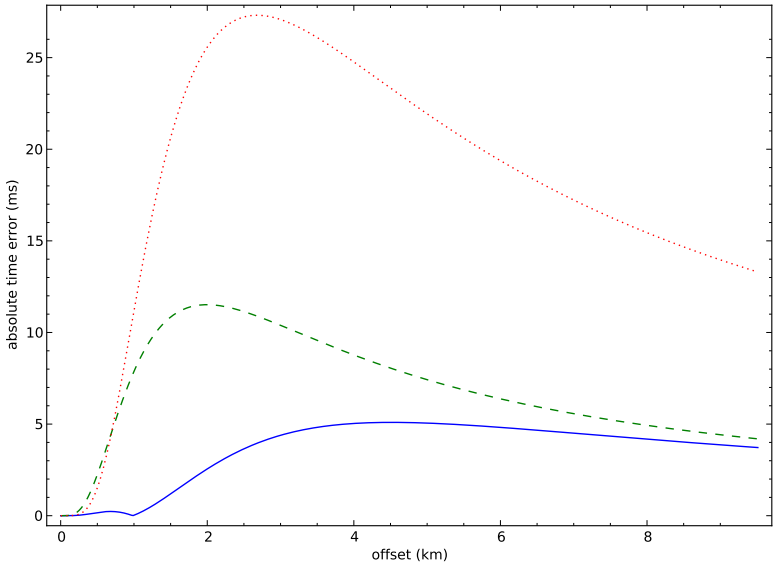


Figure 7: Traveltime moveout error of different group velocity approximations for Greenhorn shale anisotropy. The reflector depth is 1 km. Short dash: Alkhalifah-Tsvankin approximation. Long dash: Zhang-Uren approximation. Solid line: suggested approximation. group/Sage timepp

It remains to be seen if the suggested approximation proves to be useful for describing normal moveout in layered media. The next section discusses its application for traveltime computation in heterogenous velocity models.

APPLICATION: FINITE-DIFFERENCE TRAVELTIME COMPUTATION

As an essential part of seismic imaging with the Kirchhoff method, traveltime computation has received a lot of attention in the geophysical literature. Finite-difference eikonal solvers (56; 54; 34) provide an efficient and convenient way of computing first

arrival traveltimes on regular grids. Although they have a limited capacity for imaging complex structures (26), eikonal solvers can be extended in several different ways to accommodate multiple arrivals (13; 50; 2). A particularly attractive approach to finite-difference traveltime computation is the fast marching method, developed by (42) in the general context of level set methods for propagating interfaces (32; 41). (43) adopt the fast marching method for computing seismic isotropic traveltimes. Alternative implementations are discussed by (48), (9), and (29). The fast marching method possesses a remarkable numerical stability, which results from a cleverly chosen order of finite-difference evaluation. The order selection scheme resembles expanding wavefronts of (36) and wavefront tracking of (17).

While the anisotropic eikonal equation (4) operates with phase velocities, the kernel of the fast marching eikonal solver can be interpreted in terms of local ray tracing in a constant-velocity background (23) and is more conveniently formulated with the help of the group velocity. (44) present a thorough extention of the fast marching method to anisotropic wavefront propagation in the form of ordered upwind methods. In this paper, I adopt a simplified approach. Anisotropic traveltimes are computed in relation to an isotropic background. At each step of the isotropic fast marching method, the local propagation direction is identified, and the anisotropic traveltimes are computed by local ray tracing with the group velocity corresponding to the same direction. This is analogous to the tomographic linearization approach in ray tracing, where anisotropic traveltimes are computed along ray trajectories, traced in the isotropic background (19). (8) and (39) present different approaches for linearizing the anisotropic eikonal equation.

Many alternative forms of finite-difference traveltime computation in anisotropic media are presented in the literature (37; 22; 28; 14; 33; 38; 58). Although the method of this paper has limited accuracy because of the linearization assumption, it is simple and efficient in practice and serves as an illustration for the advantages of the explicit group velocity approximation (33). For a more accurate and robust extension of the fast marching method for anisotropic traveltime calculation, I recommend the ordered upwind methods of (44; 45).

Figure 8 shows finite-difference wavefronts for an isotropic and an anisotropic homogeneous media, compared with the exact solutions. The anisotropic media has the parameters of the Greenhorn shale. The finite-difference error decreases with finer sampling.

Figure 12 shows the first arrival wavefronts (traveltime contours) computed in the anisotropic Marmousi model created by (3) in comparison with wavefronts for the isotropic Marmousi model (55; 1). The model parameters are shown in Figure 10. The observed significant difference in the wavefront position suggests a difference in the positioning of seismic images when anisotropy is not properly taken into account.

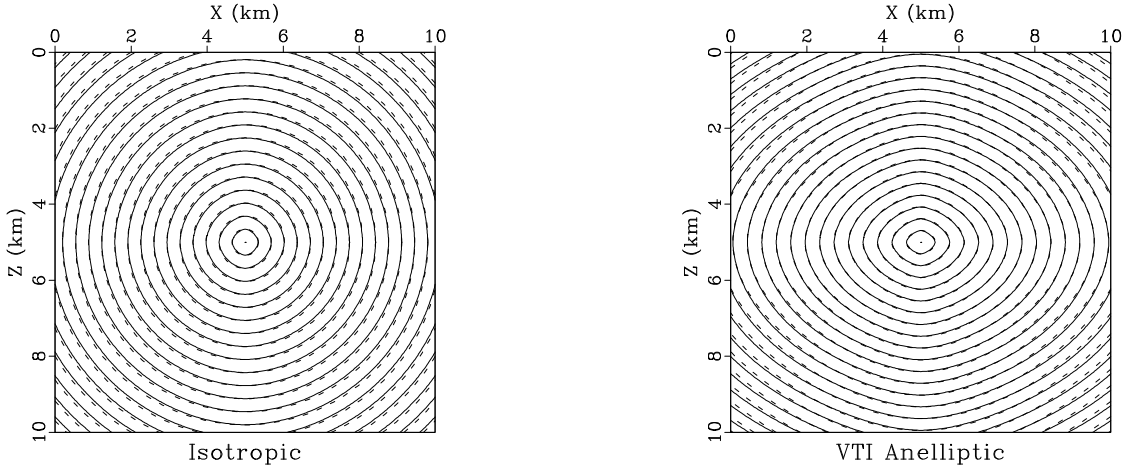


Figure 8: Finite-difference wavefronts in an isotropic (left) and an anisotropic (right) homogeneous media. The anisotropic media has the parameters of the Greenhorn shale. The finite-difference sampling is 100 m. The contour sampling is 0.1 s. Dashed curves indicate the exact solution. The finite-difference error will be reduced at finer sampling. group/ell const

CONCLUSIONS

I have developed a general approach for approximating both phase and group velocities in a VTI medium. Suggested approximations use three elastic parameters as opposed to the four parameters in the exact phase velocity expression. The phase velocity approximation coincides with the acoustic approximation of (4; 6) but is derived differently. The group velocity approximation has an analogous form and similar superior approximation properties. It is important to stress that the two approximations do not correspond exactly to each other. The exact group velocity corresponding to the acoustic approximation is different from the approximation derived in this paper and can be too complicated for practical use (5). The suggested phase and group approximations match each other in the sense that they have analogous approximation accuracy in the dual domains.

The group velocity approximation is useful for approximating normal moveout and diffraction traveltimes in applications to non-hyperbolic velocity analysis and prestack time migration. It is also useful for traveltime computations that require ray tracing in locally homogeneous cells. I have shown examples of such computations utilizing an anisotropic extension of the fast marching finite-difference eikonal solver.

ACKNOWLEDGMENTS

The paper was improved by helpful suggestions from Tariq Alkhalifah and Paul Fowler.

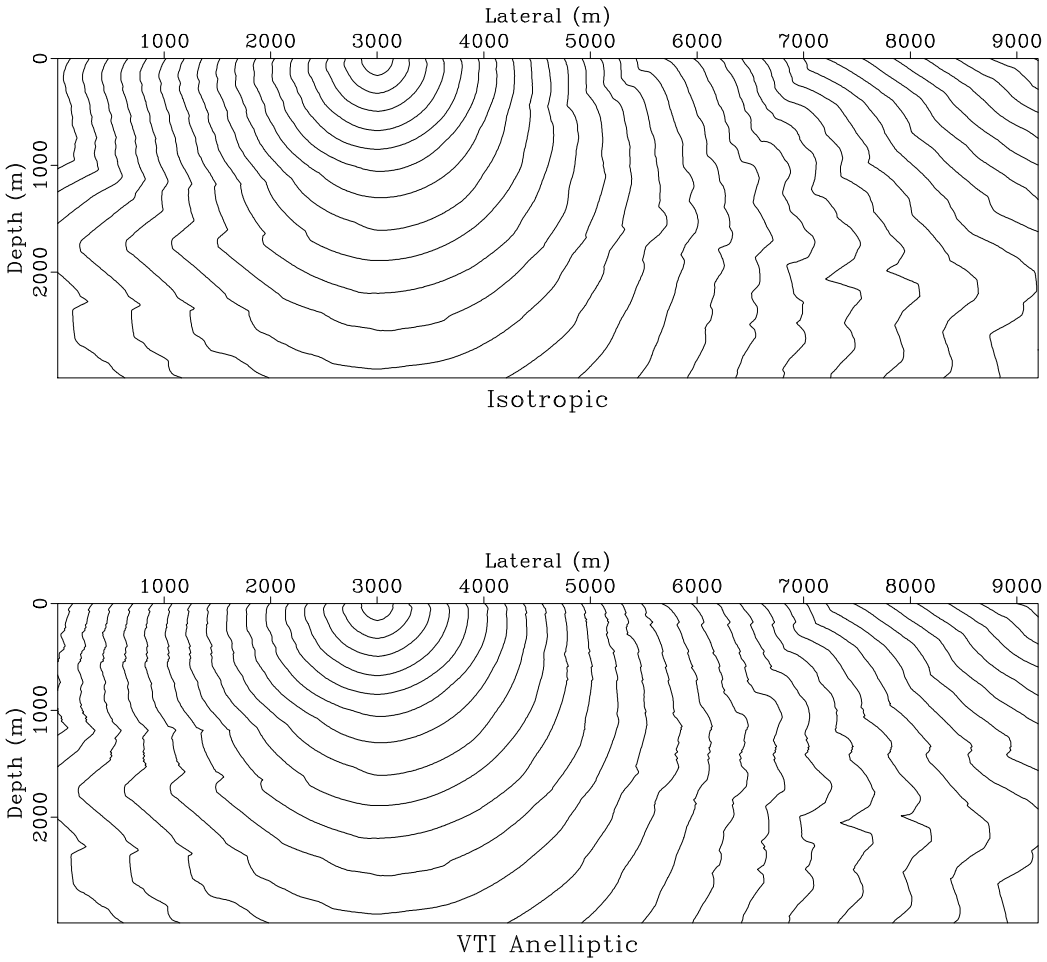


Figure 9: Finite-difference wavefronts in the isotropic (top) and anisotropic (bottom) Marmousi models. A significant shift in the wavefront position suggest possible positioning error when seismic imaging does not take anisotropy into account. group/emarm marm

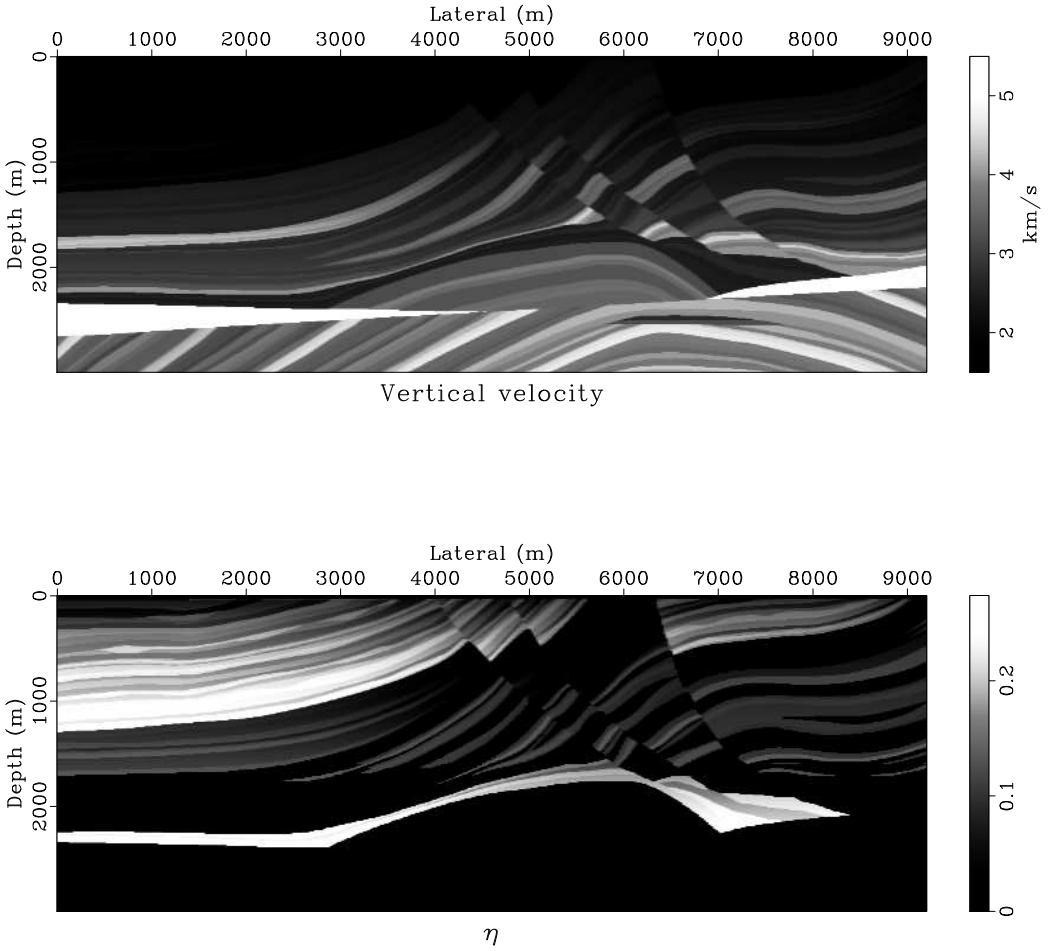


Figure 10: Alkhalifah's anisotropic Marmousi model. Top: vertical velocity. Bottom: anelliptic η parameter. The vertical velocity is taken equal to the NMO velocity V_n. group/emarm marmmod

REFERENCES

[1] 1990, Practical aspects of seismic data inversion, The Marmousi experience: 52nd Mtg., Eur. Assn. of Expl. Geophys., 1–194. (Eight comparisons of inversion techniques, Workshop Proceedings).

[2] Abgrall, R., and J. D. Benamou, 1999, Big ray-tracing and eikonal solver on unstructured grids: Application to the computation of a multivalued traveltime field in the Marmousi model: Geophysics, **64**, 230–239.

[3] Alkhalifah, T., 1997, An anisotropic Marmousi model, *in* SEP–**95**: Stanford Exploration Project, 265–282.

[4] ———, 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.

[5] ———, 1999, The offset-midpoint traveltime pyramid for transversely isotropic media: 61st Mtg., Eur. Assn. Geosci. Eng., Session:1049.

[6] ———, 2000a, An acoustic wave equation for anisotropic media: Geophysics, **65**, 1239–1250.

[7] ———, 2000b, The offset-midpoint traveltime pyramid in transversely isotropic media: Geophysics, **65**, 1316–1325.

[8] ———, 2002, Traveltime computation with the linearized eikonal equation for anisotropic media: Geoph. Prosp., **50**, 373–382.

[9] Alkhalifah, T., and S. Fomel, 2001, Implementing the fast marching eikonal solver: spherical versus Cartesian coordinates: Geophys. Prosp., **49**, 165–178.

[10] Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

[11] Backus, G. E., 1962, Long-wave elastic anisotropy produced by horizontal layering: J. Geoph. Res., **67**, 4427–4440.

[12] Berryman, J. G., 1979, Long-wave elastic anisotropy in transversely isotropic media: Geophysics, **44**, 896–917. (Discussion in GEO-45-05-0977-0980; Reply in GEO-45-05-0980-0982; Discussion in GEO-46-03-0336-0338; Reply in GEO-46-03-0338-0339).

[13] Bevc, D., 1997, Imaging complex structures with semirecursive Kirchhoff migration: Geophysics, **62**, 577–588.

[14] Bousquie, N., and R. Siliqi, 2001, 3D VTI eikonal solver for efficient acoustic travel-time computation, *in* Anisotropy 2000: Fractures, converted waves, and case studies: Soc. of Expl. Geophys., 333–338.

[15] Byun, B. S., 1984, Seismic parameters for transversely isotropic media: Geophysics, **49**, 1908–1914.

[16] Byun, B. S., D. Corrigan, and J. E. Gaiser, 1989, Anisotropic velocity analysis for lithology discrimination: Geophysics, **54**, 1564–1574.

[17] Cao, S., and S. A. Greenhalgh, 1994, Finite-difference solution of the eikonal equation using an efficient, first-arrival wavefront tracking scheme: Geophysics, **59**, 632–643. (Errata in GEO-64-03-992).

[18] Castle, R. J., 1994, Theory of normal moveout: Geophysics, **59**, 983–999.

[19] Chapman, C. H., and R. G. Pratt, 1992, Traveltime tomography in anisotropic media–I: Theory: Geophys. J. Internat., **109**, 1–19.

[20] de Bazelaire, E., 1988, Normal moveout revisited - Inhomogeneous media and curved interfaces: Geophysics, **53**, 143–157.
[21] Dellinger, J., F. Muir, and M. Karrenbach, 1993, Anelliptic approximations for TI media: Jour. Seis. Expl., **2**, 23–40.
[22] Dellinger, J., and W. Symes, 1997, Anisotropic finite-difference traveltimes using a Hamilton-Jacobi solver: 67th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1786–1789.
[23] Fomel, S., 1997, A variational formulation of the fast marching eikonal solver, *in* SEP–**95**: Stanford Exploration Project, 127–147.
[24] Fomel, S., and L. Vaillant, 2001, Evaluating the Stolt-stretch parameter: Journal of Seismic Exploration, **9**, 319–335.
[25] Gassmann, F., 1964, Introduction to seismic travel time methods in anisotropic media: Pageoph., **58**, 63–112.
[26] Geoltrain, S., and J. Brac, 1993, Can we image complex structures with first-arrival traveltime?: Geophysics, **58**, 564–575.
[27] Jones, L. E. A., and H. F. Wang, 1981, Ultrasonic velocities in Cretaceous shales from the Williston Basin: Geophysics, **46**, 288–297.
[28] Kim, S., 1999, On eikonal solvers for anisotropic traveltime: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1875–1878.
[29] ———, 2002, 3-D eikonal solvers: First arrival traveltimes: Geophysics, **67**, 1225–1231.
[30] Malovichko, A. A., 1978, A new representation of the traveltime curve of reflected waves in horizontally layered media: Applied Geophysics (in Russian), **91**, 47–53, English translation in (49).
[31] Muir, F., and J. Dellinger, 1985, A practical anisotropic system, *in* SEP–**44**: Stanford Exploration Project, 55–58.
[32] Osher, S., and J. A. Sethian, 1988, Fronts propagating with curvature-dependent speed: Algorithms based on Hamilton-Jacobi formulation: Jour. of Comp. Phys., **79**, 12–49.
[33] Perez, M., and J. Bancroft, 2001, Finite-difference methods for estimating traveltimes and raypaths in anisotropic media: 71st Ann. Internat. Mtg, Soc. of Expl. Geophys., 1225–1228.
[34] Podvin, P., and I. Lecomte, 1991, Finite difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools: Geophysical Journal International, **105**, 271–284.
[35] Postma, G. W., 1955, Wave propagation in a stratified medium: Geophysics, **20**, 780–806.
[36] Qin, F., Y. Luo, K. B. Olsen, W. Cai, and G. T. Schuster, 1992, Finite-difference solution of the eikonal equation along expanding wavefronts: Geophysics, **57**, 478–487.
[37] Qin, F., and G. T. Schuster, 1993, First-arrival traveltime calculation for anisotropic media: Geophysics, **58**, 1349–1358.
[38] Qin, J., and W. Symes, 2002, Finite-difference quasi-p traveltimes for anisotropic media: Geophysics, **67**, 147–155.
[39] Schneider, W. A., 2003, Linearization of the P-wave eikonal equation for weak

vertical transverse isotropy: Geophysics, in press.
[40] Schoenberg, M. A., and M. V. de Hoop, 2000, Approximate dispersion relations for qP-qSv-waves in transversely isotropic media: Geophysics, **65**, 919–933.
[41] Sethian, J., 1999, Level set methods and fast marching methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision and materials sciences: Cambridge University Press.
[42] Sethian, J. A., 1996, A fast marching level set method for monotonically advancing fronts: Proc. Nat. Acad. Sci., **93**, 1591–1595.
[43] Sethian, J. A., and A. M. Popovici, 1999, 3-D traveltime computation using the fast marching method: Geophysics, **64**, 516–523.
[44] Sethian, J. A., and A. Vladimirsky, 2001, Ordered upwind methods for static Hamilton-Jacobi equations: Proc. Nat. Acad. Sci., **98**, 11069–11074.
[45] ———, 2003, Ordered upwind methods for static Hamilton-Jacobi equations: Theory and applications: SIAM Journal on Numerical Analysis, **41**, 325–363.
[46] Stolt, R. H., 1978, Migration by Fourier transform: Geophysics, **43**, 23–48. (Discussion and reply in GEO-60-5-1583).
[47] Stopin, A., 2001, Comparison of (symbol) equations in Ti medium, *in* Anisotropy 2000: Fractures, converted waves, and case studies: Soc. of Expl. Geophys., 349–361.
[48] Sun, Y., and S. B. Fomel, 1998, Fast-marching eikonal solver in the tetragonal coordinates: 68th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1949–1952.
[49] Sword, C. H., 1987, A Soviet look at datum shift, *in* SEP–**51**: Stanford Exploration Project, 313–316.
[50] Symes, W. W., 1998, A slowness matching finite difference method for traveltimes beyond transmission caustics: 68th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1945–1948.
[51] Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966. (Discussion in GEO-53-04-0558-0560 with reply by author).
[52] Tsvankin, I., 1996, P-wave signatures and notation for transversely isotropic media: An overview: Geophysics, **61**, 467–483.
[53] Tsvankin, I. D., 2001, Seismic signatures and analysis of reflection data in anisotropic media: Pergamon Press.
[54] van Trier, J., and W. W. Symes, 1991, Upwind finite-difference calculation of traveltimes: Geophysics, **56**, 812–821.
[55] Versteeg, R., 1994, The Marmousi experience: Velocity model determination on a synthetic complex data set: The Leading Edge, **13**, 927–936.
[56] Vidale, J. E., 1990, Finite-difference calculation of traveltimes in three dimensions: Geophysics, **55**, 521–526.
[57] Zhang, F., and N. Uren, 2001, Approximate explicit ray velocity functions and travel times for P-waves in Ti media: 71st Ann. Internat. Mtg, Soc. of Expl. Geophys., 106–109.
[58] Zhang, L., J. W. Rector, and G. M. Hoversten, 2002, An eikonal solver in tilted TI media, *in* 72nd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 1955–1958.

Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization

*Yangkang Chen, Sergey Fomel, and Jingwei Hu**

ABSTRACT

We introduce a novel iterative estimation scheme for separation of blended seismic data from simultaneous sources. The scheme is based on an augmented estimation problem, which can be solved by iteratively constraining the deblended data using shaping regularization in the seislet domain. We formulate the forward modeling operator in the common receiver domain, where two sources are assumed to be blended using a random time-shift dithering approach. The nonlinear shaping-regularization framework offers some freedom in designing a shaping operator to constrain the model in an underdetermined inverse problem. We design the backward operator and the shaping operator for the shaping regularization framework. The backward operator can be optimally chosen as a half of the identity operator in the two-source case, and the shaping operator can be chosen as coherency-promoting operator. Three numerically blended synthetic datasets and one numerically blended field dataset demonstrate the high-performance deblending effect of the proposed iterative framework. Compared with alternative $f-k$ domain thresholding and $f-x$ predictive filtering, seislet-domain soft thresholding exhibits the most robust behavior.

INTRODUCTION

The simultaneous-source technique aims at removing the limitation of no interference between adjacent shots by allowing more than one source to be shot simultaneously. Thus it can reduce the acquisition period and increase spatial sampling (4). Because of its economic benefits and technical challenges, this technique has attracted significant attention of researchers in both industry and academia (37; 33; 1; 27; 44). The biggest issue involved in simultaneous-source processing is an intense crosstalk noise between adjacent shots, which poses a challenge for conventional processing. One way to deal with this issue is to use a first-separate and second-process strategy, which is also known as *deblending* (16). The other way is by direct imaging and waveform inversion (5; 9; 39; 24; 45). In this paper, we focus on the deblending approach.

Different filtering and inversion methods have been applied previously to deblend seismic data. Filtering methods utilize the property that the coherency of

***e-mail:** sergey.fomel@beg.utexas.edu

the simultaneous-source data varies in different domains. Therefore, one can get unblended data by filtering out randomly distributed blending noise in some transform domains, where one source record is coherent and the others are not (25; 34; 27). Inversion methods treat the separation problem as an estimation problem, which aims at estimating the desired unknown unblended data. Because of the ill-posed nature of such estimation problems, a regularization term is usually required (15). The regularization term can be chosen as a sparsity promotion in a sparsifying transformed domain (2). (42) proposed to distribute all energy in the shot records by reconstructing individual shot records at their respective locations. (33) introduced an iterative estimation and subtraction scheme that combines the properties of filtering and inversion methods and exploits the fact that the characteristics of the blending noise differ in different domains. One choice is to transform seismic data from the common-shot domain to common-receiver, common-offset or common-midpoint domain. (3) proposed a separation technique called the alternating projection method, and demonstrated it to be robust in the presence of aliasing.

In this paper, we propose a novel iterative estimation scheme for the separation of blended seismic data. We construct an augmented estimation problem, then use shaping regularization (19; 20) to constrain the characteristics of the model during the iteration process for obtaining a suitable estimation result. We adopt seislet-domain (21) soft thresholding as a robust shaping operator to remove crosstalk noise and at the same time to preserve useful components of seismic data. Shaping regularization maps each model into a more admissible model at each iteration and allows us to solve the deblending problem with a small number of iterations.

This paper is organized as follows: we first present a formulation of numerical blending using an augmented block-matrix equation. We then incorporate shaping regularization to solve the forward equation iteratively and discuss the selection of the backward operator and shaping operator involved in the shaping-regularization iterative framework. Finally, we test the proposed iterative framework on three numerically blended synthetic datasets and one numerically blended field dataset and compare three different choices of the shaping operator: $f-k$ domain thresholding, $f-x$ predictive filtering, and seislet domain thresholding.

DEBLENDING USING SHAPING REGULARIZATION

Numerical blending

We assume that the seismic record is blended using two independent sources, which correspond to two shooting vessels in the ocean bottom nodes (OBN) acquisition (36). Here, one source means a collection of shots from one shooting vessel. The two sources shoot pseudosynchronously, which means that each shot in one source has a random time dithering compared with the corresponding shot in another source. By precisely picking the shooting time of each shot of one source at the long seismic

record on the node, we can get one common receiver gather. Using the second source, we get another gather. Thus, the forward problem can be formulated as follows:

$$\mathbf{d} = \mathbf{d}_1 + \mathbf{T}\mathbf{d}_2. \tag{1}$$

Here, $\mathbf{d}_1$ and $\mathbf{d}_2$ denote the seismic records to be separated out, $\mathbf{d}$ is the blended data, and $\mathbf{T}$ denotes the dithering operator. Note that we perform the processing in the common-receiver domain, because each blended record in this domain is coherent for one source and incoherent for another (25). In this case, $\mathbf{d}$ is usually referred to as the pseudodeblended data (33), which means picking traces from the original blended field record according to the shooting time of one source to form a common receiver gather with one source coherent and the other one incoherent. We propose to augment equation 4 with another equation through applying the inverse dithering operator $\mathbf{T}^{-1}$:

$$\mathbf{T}^{-1}\mathbf{d} = \mathbf{T}^{-1}\mathbf{d}_1 + \mathbf{d}_2. \tag{2}$$

By combining equations 4 and 2, we formulate an augmented estimation problem:

$$\mathbf{F}\mathbf{m} = \tilde{\mathbf{d}}, \tag{3}$$

where

$$\tilde{\mathbf{d}} = \begin{bmatrix} \mathbf{d} \\ \mathbf{T}^{-1}\mathbf{d} \end{bmatrix}, \quad \mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix}, \quad \mathbf{m} = \begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}, \tag{4}$$

and $\mathbf{I}$ is the identity operator.

We choose to solve equation 3 rather than equation 4 because equation 3 is a more convenient form to use in the following derivations.

Shaping regularization

The unknown $\mathbf{m}$ in equation 3 can be recovered iteratively using shaping regularization:

$$\mathbf{m}_{n+1} = \mathbf{S}[\mathbf{m}_n + \mathbf{B}(\tilde{\mathbf{d}} - \mathbf{F}\mathbf{m}_n)], \tag{5}$$

where operator $\mathbf{S}$ shapes the estimated model into the space of admissible models at each iteration (19; 20) and $\mathbf{B}$ is the backward operator that provides an inverse mapping from data space to model space. (11) prove that, if $\mathbf{S}$ is a nonlinear thresholding operator (14), $\mathbf{B} = \mathbf{F}^T$ where $\mathbf{F}^T$ is the adjoint operator of $\mathbf{F}$, iteration 5 converges to the solution of equation 6 with L_1 regularization term:

$$\min_{\mathbf{m}} \parallel \mathbf{F}\mathbf{m} - \tilde{\mathbf{d}} \parallel_2^2 + \mu \parallel \mathbf{A}^{-1}\mathbf{m} \parallel_1, \tag{6}$$

where μ is the threshold and $\mathbf{A}^{-1}$ denotes a sparsity-promoting transform. A better choice for $\mathbf{B}$ is the pseudoinverse of $\mathbf{F}$: $\mathbf{B} = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T$ (12).

Performance evaluation and convergence rate

Combining equations 3, 4 and 5, we get:

$$\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}_{n+1} = \mathbf{S}\left[\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}_n + \mathbf{B}\left[\begin{bmatrix} \mathbf{d} \\ \mathbf{T}^{-1}\mathbf{d} \end{bmatrix} - \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix}\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \end{bmatrix}_n\right]\right], \tag{7}$$

where $\mathbf{S}$ is a block operator that takes the form $[\mathbf{S}_1 \quad \mathbf{0}; \mathbf{0} \quad \mathbf{S}_2]$, and $\mathbf{S}_1$ and $\mathbf{S}_2$ correspond to the shaping operators for $\mathbf{d}_1$ and $\mathbf{d}_2$, respectively. Using equation 7 and performing appropriate iterations, we aim to get an estimate of unblended seismic data.

To test the convergence performance and accuracy of iteration 7, we use the following measure (26):

$$SNR_{i,n} = 10\log_{10}\frac{\|\mathbf{d}_i\|_2^2}{\|\mathbf{d}_i - \mathbf{d}_{i,n}\|_2^2}, \tag{8}$$

where $SNR_{i,n}$ stands for signal-to-noise ratio of ith source after nth iteration, $\mathbf{d}_{i,n}$ denotes the estimated model of ith source after n iterations, $\mathbf{d}_i$ denotes the true model for ith source and $\|\cdot\|_2^2$ denotes the squared L_2 norm of a function.

BACKWARD OPERATOR

In equation 5, $\mathbf{B}$ is an approximate inverse of $\mathbf{F}$. Recall that the dithering operator applies a time-domain random time shift to the input seismic data. Taken in the frequency domain, the dithering operator takes the following form:

$$\mathbf{T} = \mathcal{F}^{-1}\mathbf{P}\mathcal{F}, \tag{9}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ are forward and inverse Fourier transforms, respectively, and $\mathbf{P}$ is a $N \times N$ diagonal block phase-shift operator given by

$$\mathbf{P} = diag(\mathbf{P}_1, \mathbf{P}_2, \mathbf{P}_3, \cdots, \mathbf{P}_N), \tag{10}$$

where $\mathbf{P}_n$ denotes the individual phase shift operator for nth trace and can be expressed as a $M \times M$ diagonal matrix:

$$\mathbf{P}_n = diag(\overbrace{1, 1, 1, \cdots, 1}^{M}) * \exp(-i\omega\delta t_n). \tag{11}$$

Here, w denotes the angular frequency, δt_n denotes the random dithering time of nth trace, and M and N in equations 10 and 11 denote the number of temporal samples and number of traces, respectively. $diag(\cdot)$ denotes a diagonal matrix.

Considering that the Fourier operator (with symmetric normalization) and the phase shift operator are both unitary, which means $\mathcal{F}^{-1} = \mathcal{F}^T$ and $\mathbf{P}^{-1} = \mathbf{P}^T$, it is easy to see that

$$\mathbf{T}^T = (\mathcal{F}^{-1}\mathbf{P}\mathcal{F})^T = \mathcal{F}^T\mathbf{P}^T\mathcal{F} = (\mathcal{F}^{-1}\mathbf{P}\mathcal{F})^{-1} = \mathbf{T}^{-1}. \tag{12}$$

Thus, we conclude that the dithering operator is also a unitary operator. Furthermore we notice that

$$\mathbf{F}^T = \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix}^T = \begin{bmatrix} \mathbf{I} & \mathbf{T}^{-T} \\ \mathbf{T}^T & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix} = \mathbf{F},$$

and:

$$\mathbf{F}^T\mathbf{F} = \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix} = 2 \begin{bmatrix} \mathbf{I} & \mathbf{T} \\ \mathbf{T}^{-1} & \mathbf{I} \end{bmatrix} = 2\mathbf{F}. \tag{13}$$

The least-squares solution of equation 3 is therefore:

$$\hat{\mathbf{m}} = (\mathbf{F}^T\mathbf{F})^{-1}\mathbf{F}^T\tilde{\mathbf{d}} = \frac{1}{2}\mathbf{F}^{-1}\mathbf{F}^T\tilde{\mathbf{d}} = \frac{1}{2}\tilde{\mathbf{d}}. \tag{14}$$

According to equation 14, an appropriate choice for $\mathbf{B}$ is simply $\mathbf{B} = \frac{1}{2}[\mathbf{I} \quad \mathbf{0}; \mathbf{0} \quad \mathbf{I}]$. This form of the derived backward operator is also referred to as scaled pseudodeblending (32).

The dithering operator is unitary only if the time-shift range is small and so the constructive summation of the useful components (events) between different traces can be ignored. Even if this condition isn't fully met, we can use the concept of interference to generalize the meaning of the dithering operator $\mathbf{T}$. Although, in this case, the backward operator might not be most appropriately chosen as half of the identity operator, we can still use it as an approximation.

SHAPING OPERATOR

The shaping operator $\mathbf{S}$ in equation 5 can be chosen as a coherency-promoting operator, which serves as a flexible constraint on the estimated model. Any coherency-promoting tool such as $f - x$ predictive filtering (6; 40; 22; 8) or median filter (31; 27; 28) can be utilized to preserve the coherent useful subsurface reflection and at the same time to remove incoherent blending noise.

Another approach to promote coherency is to promote sparseness in an appropriate transform domain. To enforce sparsity, a transform domain thresholding operation can also be chosen as the shaping operator as long as the data in the transformed domain is sparse (35; 7). Instead of a thresholding operator, we can also use a mask operator to compress the transformed domain (38; 29). A sparsity-promoting shaping operator can be defined as:

$$\mathbf{S}_i = \mathbf{A}\mathcal{T}_{\gamma_i}[\mathbf{A}^{-1}], \tag{15}$$

where i denotes the ith source, $\mathbf{A}^{-1}$ and $\mathbf{A}$ are forward and inverse sparsity-promoting transforms and $\mathcal{T}_{\gamma_i}$ corresponds to a mask operator or thresholding operator with an input parameter γ_i in the transformed domain.

Mask operator can be chosen to preserve small-scale components and remove large-scale components. It takes the following form:

$$\mathcal{M}_{\gamma_i}(v(\mathbf{x})) = \begin{cases} v(\mathbf{x}) & \text{for} \quad s(\mathbf{x}) < \gamma_i \\ 0 & \text{for} \quad s(\mathbf{x}) \geq \gamma_i \end{cases}, \tag{16}$$

where $\mathbf{x}$ is a position vector in the transformed domain, $v(\mathbf{x})$ denotes the amplitude value of point $\mathbf{x}$, $s(\mathbf{x})$ denotes the scale of point $\mathbf{x}$ and $\mathcal{M}_{\gamma_i}$ corresponds to the mask operator with a scale limitation γ_i. When the sparsity-promoting transform is the Fourier transform, the scale defined in the mask operator shown in equation 16 corresponds to the frequency and wavenumber band.

Thresholding operators can be divided into two types: soft and hard. Soft thresholding or shrinkage aims to remove data whose value is smaller than a certain level and subtract the other data values by this level (13). Hard thresholding simply removes data with small values. A soft thresholding operator takes the following form:

$$\mathcal{S}_{\gamma_i}(v(\mathbf{x})) = \begin{cases} v(\mathbf{x}) - \gamma_i \frac{v(\mathbf{x})}{|v(\mathbf{x})|} & \text{for} \quad |v(\mathbf{x})| > \gamma_i \\ 0 & \text{for} \quad |v(\mathbf{x})| \leq \gamma_i \end{cases}, \tag{17}$$

where $\mathcal{S}_{\gamma_i}$ corresponds to the soft thresholding operator with a threshold value γ_i.

Similarly, a hard thresholding operator takes the form:

$$\mathcal{H}_{\gamma_i}(v(\mathbf{x})) = \begin{cases} v(\mathbf{x}) & \text{for} \quad |v(\mathbf{x})| > \gamma_i \\ 0 & \text{for} \quad |v(\mathbf{x})| \leq \gamma_i \end{cases}, \tag{18}$$

where $\mathcal{H}_{\gamma_i}$ corresponds to the hard thresholding operator with a threshold value γ_i.

The mask operator can be more efficient because it is linear and requires only the scale coefficient below which data values are preserved. The thresholding operator needs to compare the amplitude value of each transformed domain point with a predefined coefficient. However, the mask operator is more difficult to design when the signal and noise coefficients are both spread across the transformed domain. Therefore, in the case of deblending, we prefer to use the thresholding operator rather than the mask operator. The selection criteria of γ_i in the above definitions of thresholding operators (both soft and hard) leads to different kinds of thresholding strategies, which may result in different thresholding performances (convergence rate and quality) (47). Some of the common iterative thresholding strategies are constant-value thresholding, linear-decreasing thresholding, exponential-decreasing thresholding, etc. (23). In our paper, we use percentile thresholding (43; 46), which is convenient to implement and helps accelerate the convergence (46).

Comparison of sparsity-promoting transforms

The best-known sparsity-promoting transforms are the Fourier and wavelet transform. (21) proposed a sparsity-promoting transform called *seislet* transform. The specific

adaptation for seismic data makes the seislet transform appealing to seismic data processing. In the seislet domain, useful events and blending noise reside at different scales and the useful signal tends to form a more compact regionẆe provide a brief review of the seislet transform in Appendix A.

In order to illustrate the comparative sparseness of different transforms, we first create a simple synthetic data (Figure 1). Figure 2 shows different transformed domains for the data. Figure 2(d) shows a comparison between the decay of sorted coefficients in the 2-D Fourier transform, 2-D wavelet transform, and 2-D seislet transform. The 2-D Fourier transform in this case means the $f-k$ transform. The 2-D wavelet transform means implementing the 1-D wavelet transform along the temporal direction first and along the spatial direction second. The 2-D seislet transform means implementing the seislet transform along the spatial direction first and 1-D seislet transform along the temporal direction second. Our experiments show that the seislet coefficients decay significantly faster than coefficients of the other two transforms, which indicates a more compact structure of the seislet domain. The shaping operation is thus preferably chosen as the thresholding or mask operation in the seislet domain.

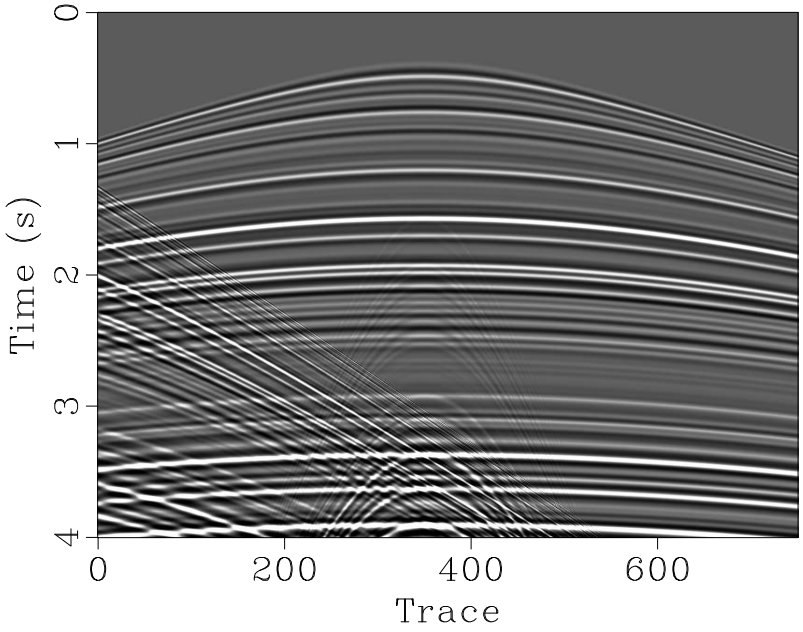


Figure 1: A synthetic seismic profile. deblend/hyper test1

EXAMPLES

In this section, in order to test the effectiveness of the proposed iterative framework defined in equation 5, we create three numerically blended examples with different kinds of shaping operations to demonstrate its applicability. The iterative framework in equation 5 is fundamentally a noise attenuation algorithm. The better an operator's

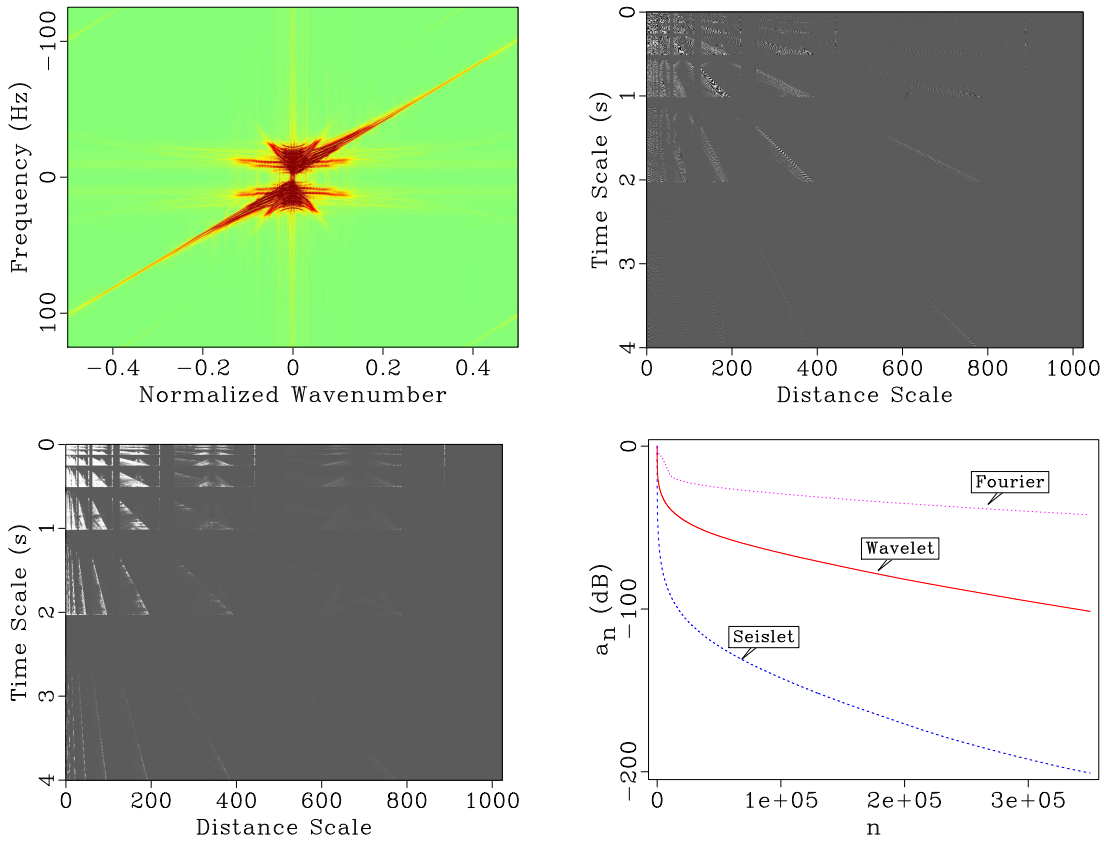


Figure 2: Comparison among different sparsity-promoting transforms. (a) 2-D Fourier transform domain. (b) 2-D Wavelet transform domain. (c) 2-D Seislet transform domain. (d) Coefficients decreasing diagram. deblend/hyper ft,wlet,slet0,sigcoef

ability to remove blending noise and preserve useful signal, the more appropriate it is for shaping. From this point of view, we first created two synthetic sections in different cases. One is a linear event case, with conflicting dip components, and the other is a hyperbolic event case, without crossing events. In order to deal with a more realistic and difficult situation, we created a more complex synthetic model (shown in Figure 1 and used previously for comparing the sparseness of different transforms) to test the proposed deblending algorithm. For the field data example, we use two common receiver gathers acquired by the OBN technique to generate a numerically blended data test.

The first three blended synthetic datasets (shown in Figures 3 to 11) show nearly perfect deblending results, particularly when the shaping operator is chosen as soft thresholding in the seislet domain. The third blended synthetic datasets shows an acceptable experiments are based on an OBN acquisition, as noted in the beginning of the paper. Thus, each seismic section in the examples corresponds to a common receiver gather. The first example is used simply for testing the denoising ability of each of the shaping operators in the case of crossing events. This example can also demonstrate the deblending performance for common offset gathers, where most seismic events are linear. For conciseness, we show the deblending performance for only one source.

Numerically blended synthetic data - linear events

Our first synthetic example contains three plane-wave events, with one high-dip-angle event crossing two other events. The original unblended and numerically blended sections are shown in Figure 4 shows the deblended results using soft thresholding in the $f-k$ domain, soft thresholding in the seislet domain, and $f-x$ predictive filtering. All of these deblending results are generally acceptable despite some artifacts left in Figure 4(a) and some weak-level noise left in Figure 4(c). By computing the difference between the deblended section and the blended section, we can obtain the blending noise section. Comparing the blending noise sections (see Figures 4(d)-4(f)), we find that there is some leakage of useful energy for $f-x$ predictive filtering. Computing the differences between the deblended sections and the unblended sections, we get the estimation error sections, shown in Figures 4(g)-4(i). Compared with the other two shaping approaches, seislet-domain soft thresholding causes a nearly zero estimation error, which indicates a nearly perfect deblending result. The estimation error for $f-x$ predictive filtering is comparatively large, because of the large predictive error problem when several dipping components are taken into consideration (8). The small estimation errors in Figure 4(g) are caused by spectrum mixture in the $f-k$ domain because some useful energy reside in the low amplitude part that is easy to be removed during soft thresholding. The small estimation errors shown in Figure 4(h) are caused by dip estimation error because of the conflicting dips when using plane wave destruction (PWD) algorithm with a single slope (17), which is a limitation of seislet-domain denoising approaches, unless a seislet frame is used instead of the seislet

transform (21). The diagrams for convergence rates are shown in Figure 5, which demonstrates a superior behavior of seislet-domain soft thresholding in comparison with the other two approaches. In order to make the comparisons fair, we try to find the best deblending performance through the parameter-selection process for corresponding shaping operators. In this case, the percentages we use for $f-k$ domain and seislet domain thresholding are both 8 %, the filter length we use for $f-x$ predictive filtering is 4 samples.

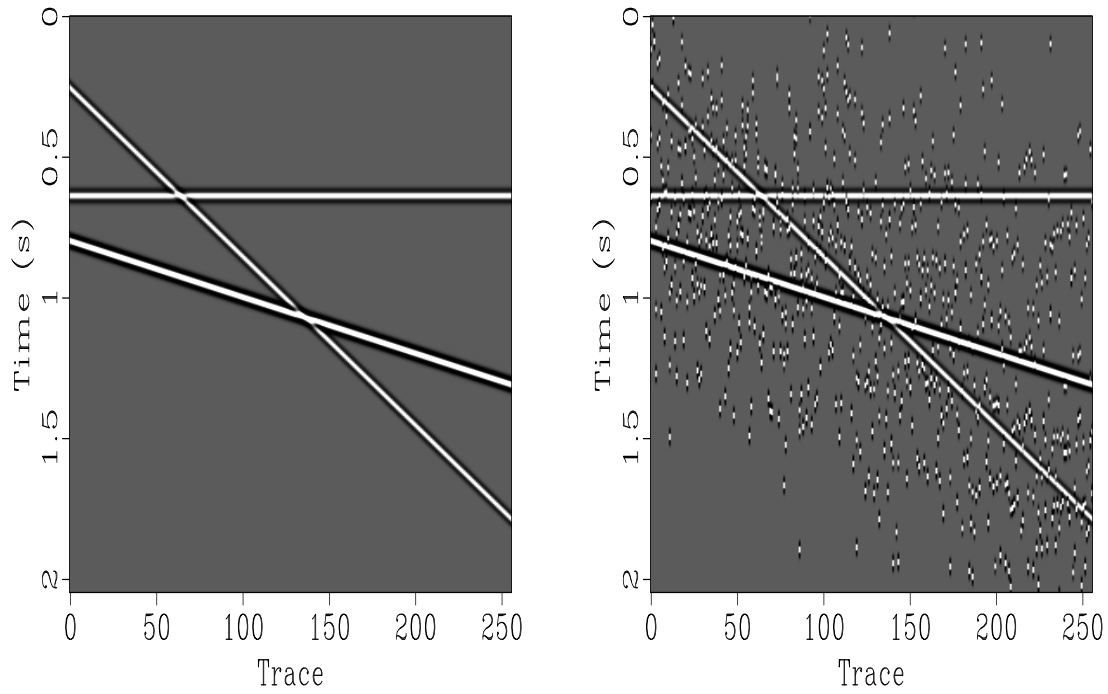


Figure 3: Numerically blended synthetic data (linear case). (a) Unblended data. (b) Blended data. deblend/linear data1,datas

Numerically blended synthetic data - hyperbolic events

Next, we create another synthetic example which contains hyperbolic events without dip conflicts. The unblended and blended data are shown in Figures 6(a) and 6(b), respectively. Figure 7 shows the deblending results for the hyperbolic case. Figure 8 shows the diagrams of changing SNR for one source. In this case, the estimation error for the seislet-domain soft thresholding becomes negligible, mainly owing to the successful dip estimation when constructing the seislet transform. The strength of the $f-x$ predictive filtering also increases and its SNR diagram shows a similar behavior as the $f-k$ domain thresholding. The seislet-based shaping still outperforms the other two approaches. In this case, the percentages we use for $f-k$ domain and seislet domain thresholding are both 9 %, the filter length we use for $f-x$ predictive filtering is 4 samples.

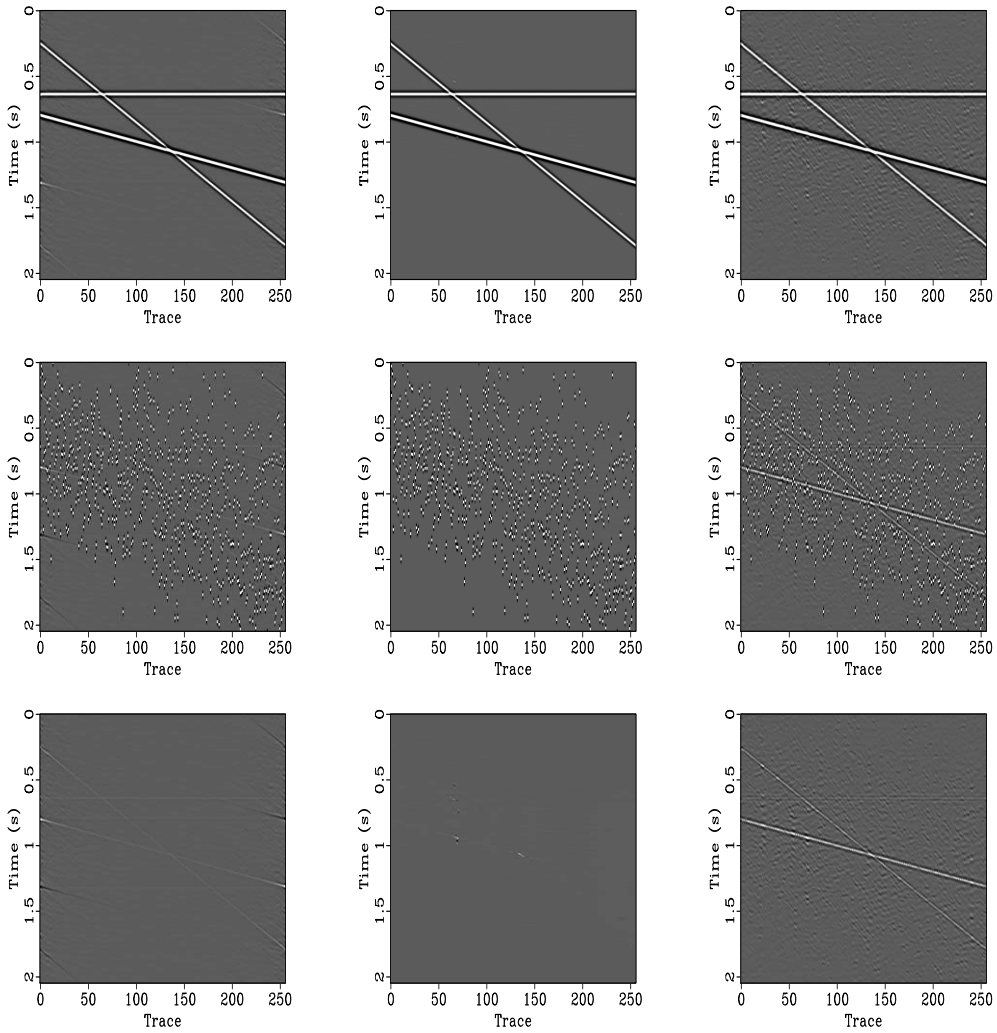


Figure 4: Deblending comparison for numerically blended synthetic data (linear case). (a) Deblended result using $f-k$ domain thresholding. (b) Deblended result using seislet-domain thresholding. (c) Deblended result using $f-x$ predictive filtering. (d) Blending noise corresponding to (a). (e) Blending noise corresponding to (b). (f) Blending noise corresponding to (c). (g) Estimation error corresponding to (a). (h) Estimation error corresponding to (b). (i) Estimation error corresponding to (c).

deblend/linear deblendedfft1,deblendedslet1,deblendedfxdecon1,diffft1,diffslet1,difffxdecon1,errorfft1,

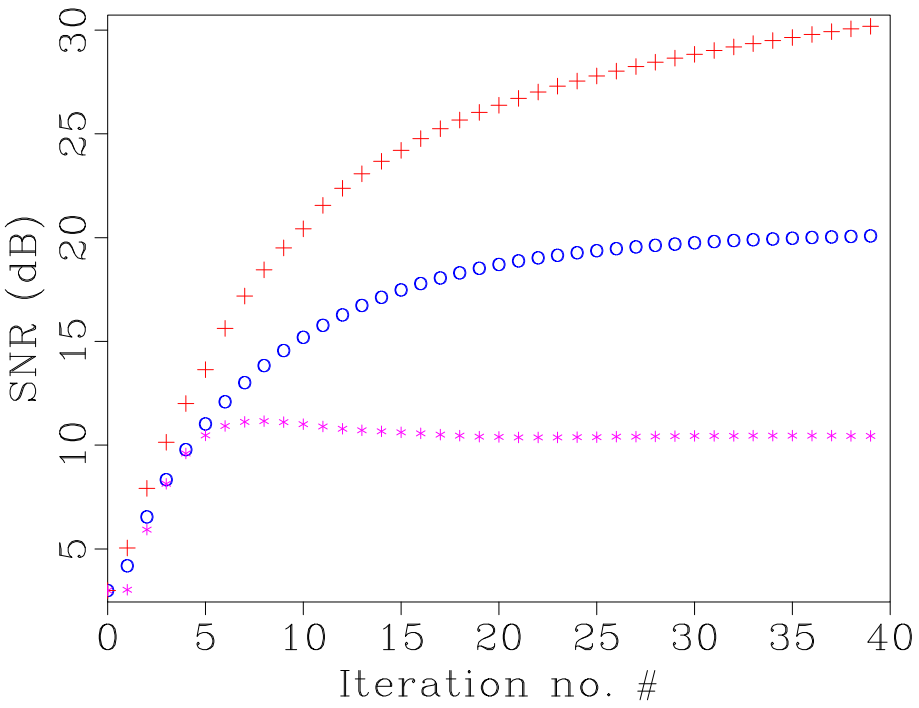


Figure 5: Diagrams of SNR for synthetic example (linear case). The "+" line corresponds to seislet-domain thresholding. The "o" line corresponds to $f-k$ domain soft thresholding. The "*" line corresponds to $f-x$ predictive filtering. deblend/linear snrsa

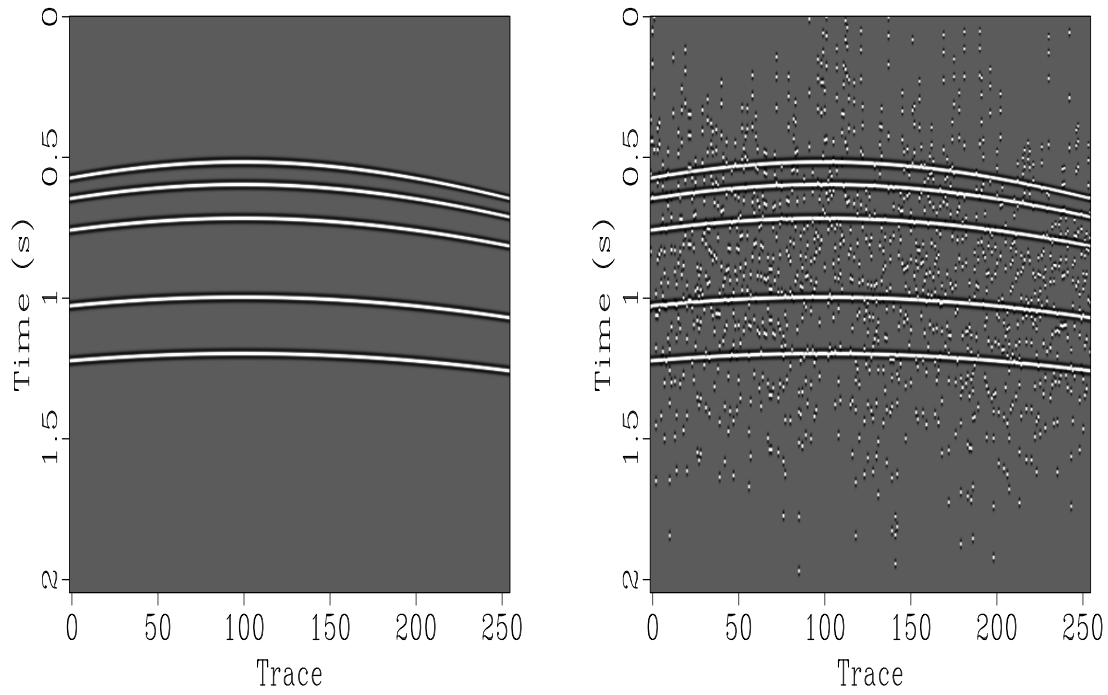


Figure 6: Numerically blended synthetic data (hyperbolic case). (a) Unblended data. (b) Blended data. deblend/synthhyper hyper1,hypers

Numerically blended synthetic data - complex gather

The third synthetic dataset consists of two complex gathers, which contain both useful seismic reflections and additional components such as ground roll and coherent dipping events. We simulate this example in order to best match the field data. The unblended and blended data are shown in Figures 9(a) and 9(b), respectively. Figure 10 shows the deblending results for this case. Figure 11 shows the diagrams of changing SNR for both sources. The SNR for the converged deblended data becomes noticeably smaller compared with the previous examples for three different shaping operators respectively, which results from the fact that the seismic gather is no longer ideally sparse as the previous examples. Notice that in the noise sections for $f-k$ domain thresholding (Figure 10(a)) and the noise sections for $f-x$ predictive filtering (Figure 10(c)), there exist a certain amount of coherent events. However, for the seislet-domain soft thresholding, there are barely any coherent events. If we look carefully at the error sections corresponding to these three shaping operators, we can find that for the $f-k$ domain thresholding, the estimation error comes from the useful hyperbolic reflections, the coherent dipping events, and the ground roll, for the $f-x$ predictive filtering, the estimation error mainly comes from the useful hyperbolic reflections and the coherent dipping events, but for the seislet-domain soft thresholding, only a small amount of the ground roll and negligible amount of coherent dipping events and deep-water horizontal reflections are left on the error

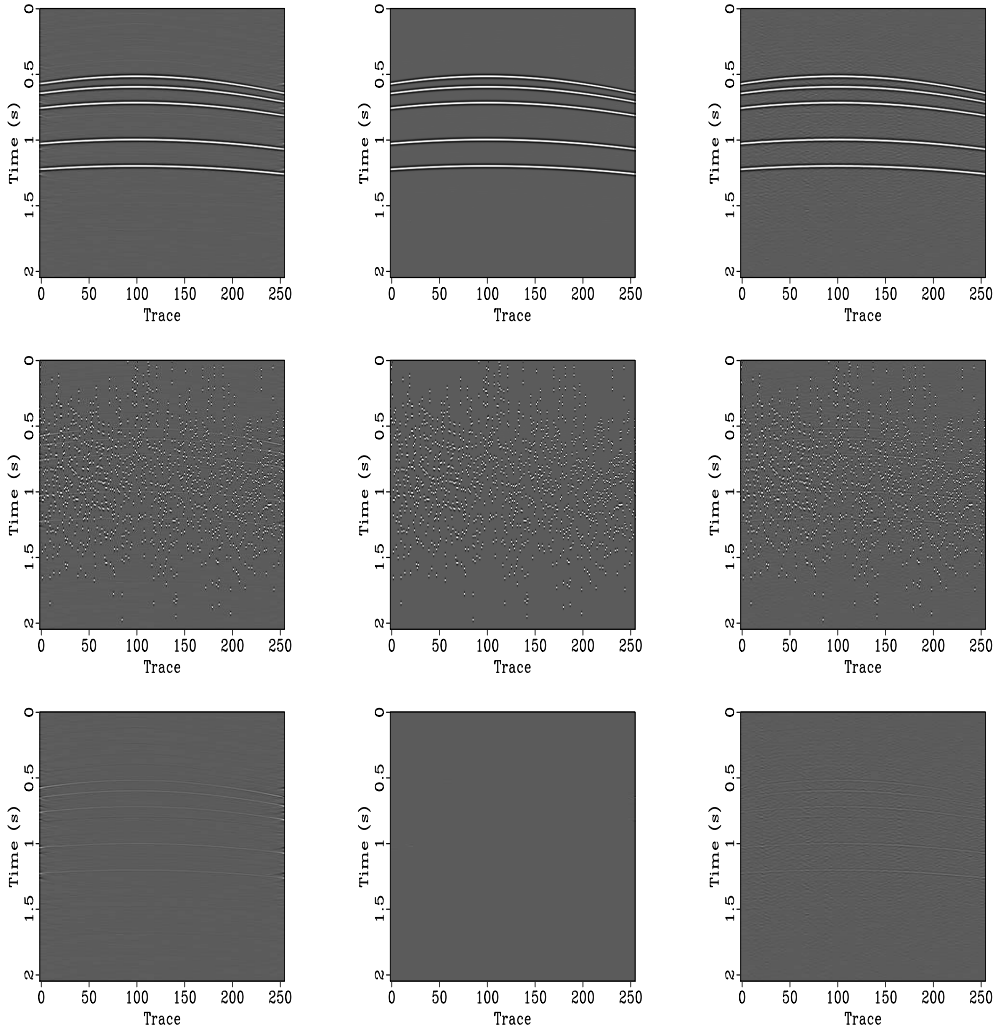


Figure 7: Deblending comparison for numerically blended synthetic data (hyperbolic case). (a) Deblended result using $f-k$ domain thresholding. (b) Deblended result using seislet-domain thresholding. (c) Deblended result using $f-x$ predictive filtering. (d) Blending noise corresponding to (a). (e) Blending noise corresponding to (b). (f) Blending noise corresponding to (c). (g) Estimation error corresponding to (a). (h) Estimation error corresponding to (b). (i) Estimation error corresponding to (c).

deblend/synthhyper hyperdeblendedfft1,hyperdeblendedslet1,hyperdeblendedfxdecon1,hyperdifffft1,hy

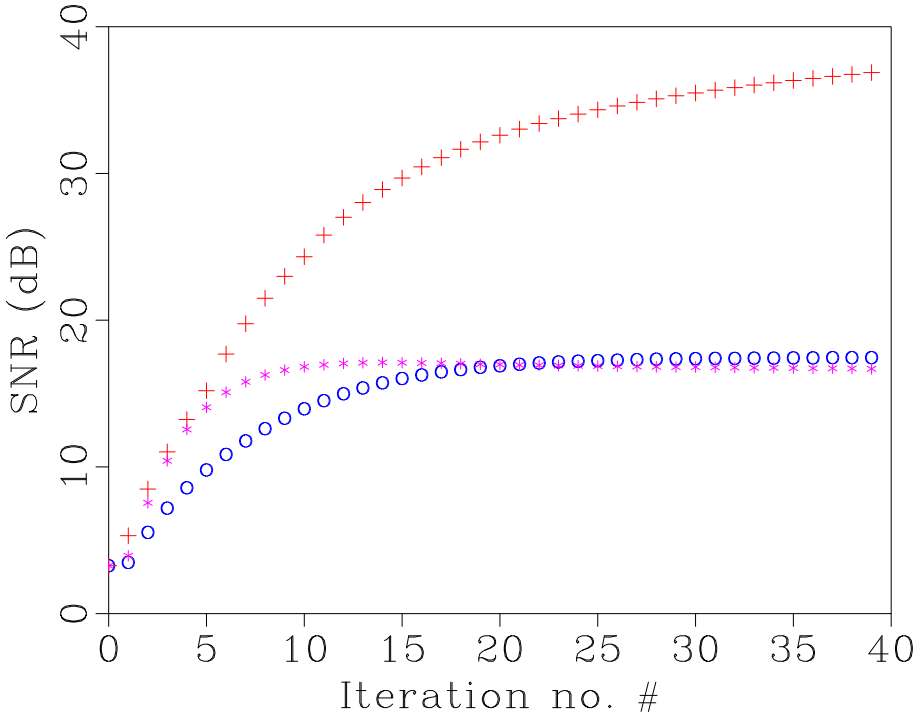


Figure 8: Diagrams of SNR for synthetic example (hyperbolic case). The "+" line corresponds to seislet-domain thresholding. The "o" line corresponds to $f-k$ domain thresholding. The "*" line corresponds to $f-x$ predictive filtering. deblend/synthhyper hypersnrsa

sections, which is not important for the whole processing tasks. We can conclude that for complex datasets like this synthetic example, seislet-domain soft thresholding is the preferable approach for shaping regularization. In this case, the percentages we use for $f-k$ domain and seislet domain thresholding are both 20 %, the filter length we use for $f-x$ predictive filtering is 4 samples.

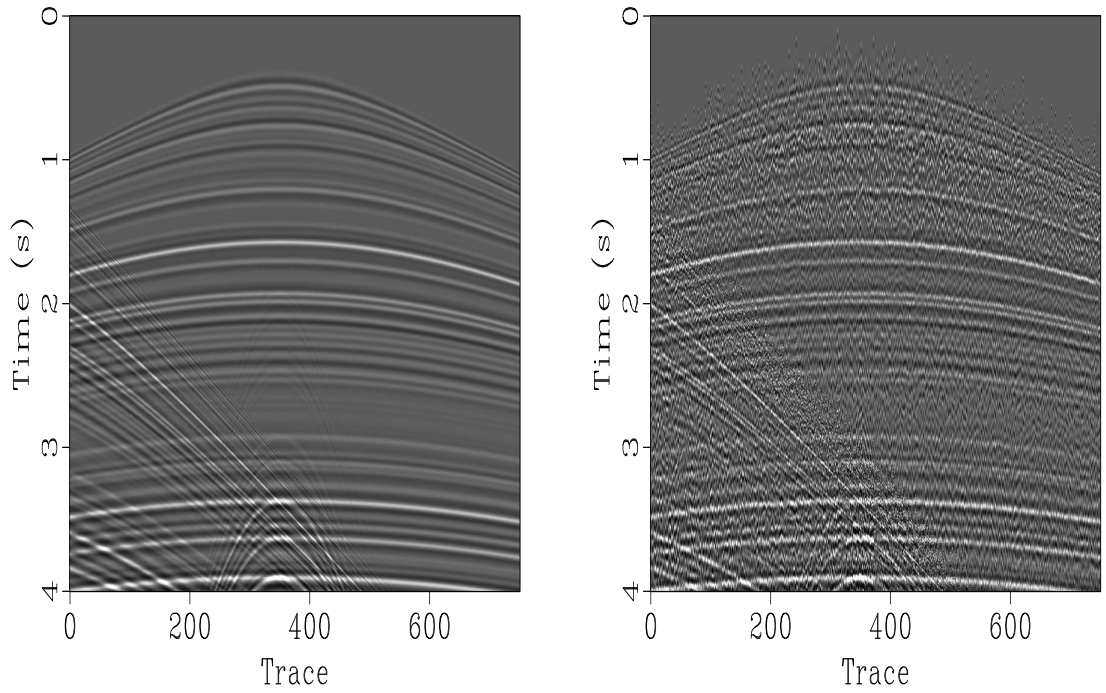


Figure 9: Numerically blended synthetic data (complex gather). (a) Unblended data. (b) Blended data. deblend/synthcomplex complex1,complexs

Numerically blended field data

Finally, we use a field data example to further test the effectiveness for our proposed algorithm. The field data is similar in appearance to the previous synthetic example. Because of the previous test, we only used seislet-domain soft thresholding as the shaping operator. The two blended data are shown in Figures 12(a) and 12(b) and appear noisy. After 30 iterations of seislet-domain soft thresholding, the deblending results are shown in Figures 12(d) and 12(c). The intense interference in the shallow part has been removed to a large extent. The deep-water reflection becomes much clearer because of a significant drop in the noise energy for the whole profile. From the noise sections, we can observe only a small amount of ground roll, which is analogous to the results of the previous synthetic test, and a even less amount of direct waves and shallow part horizontal reflections, which is not significant. The SNR defined in equation 11 obtained an increase from 1.00 dB to 5.50 dB for the first source and from 0.83 dB to 5.50 dB for the second source. The error sections for this field

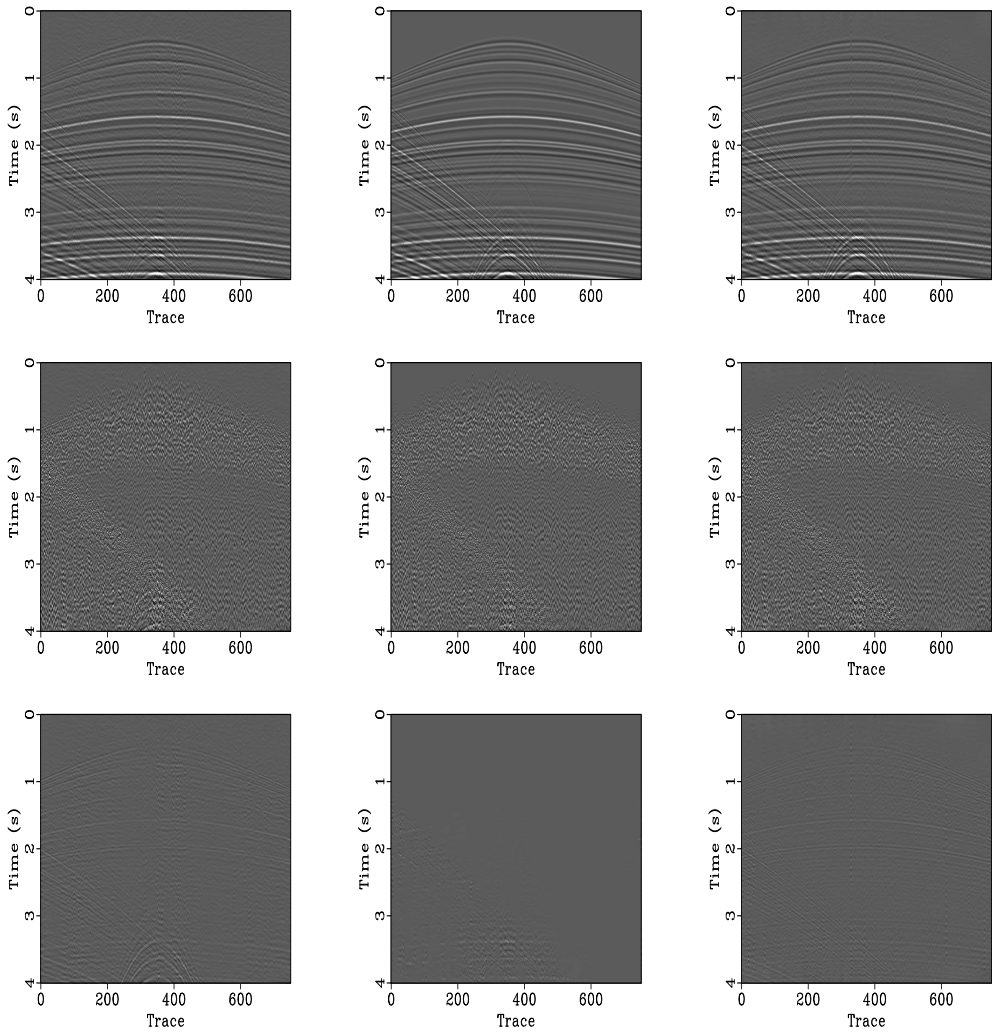


Figure 10: Deblending comparison for numerically blended synthetic data (complex gather). (a) Deblended result using $f-k$ domain thresholding. (b) Deblended result using seislet-domain thresholding. (c) Deblended result using $f-x$ predictive filtering. (d) Blending noise corresponding to (a). (e) Blending noise corresponding to (b). (f) Blending noise corresponding to (c). (g) Estimation error corresponding to (a). (h) Estimation error corresponding to (b). (i) Estimation error corresponding to (c).

deblend/synthcomplex complexdeblendedfft1,complexdeblendedslet1,complexdeblendedfxdecon1,comp

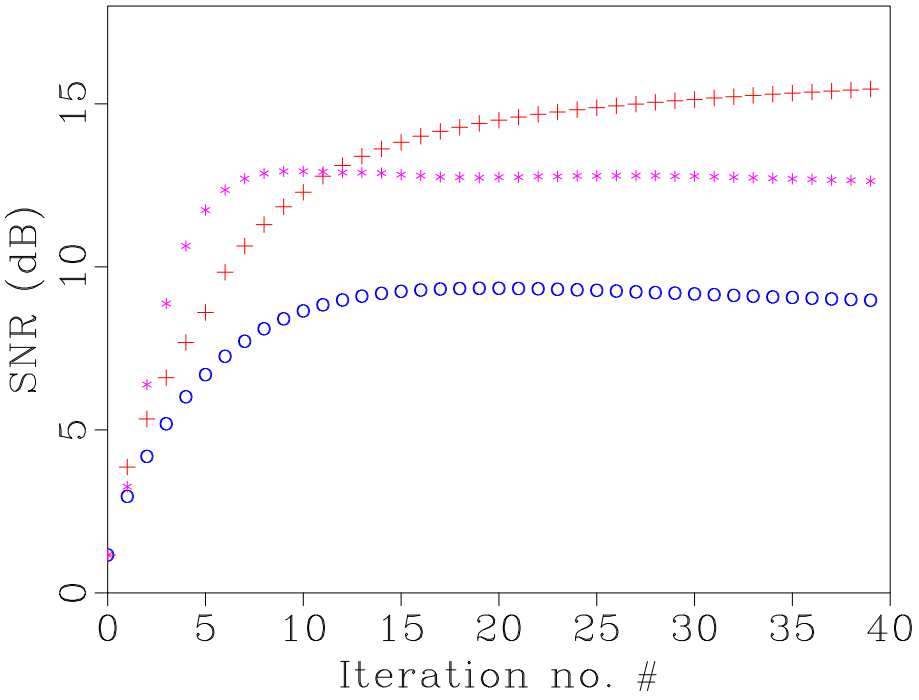


Figure 11: Diagrams of SNR for synthetic example (complex gather). The "+" line corresponds to seislet-domain thresholding. The "o" line corresponds to $f-k$ domain thresholding. The "*" line corresponds to $f-x$ predictive filtering. deblend/synthcomplex complexsnrsa

data test are not given since the unblended data for this test contain much random noise, which can not be estimated. However, from the point of view of random noise attenuation, which is concerned with the signal content of the denoised section and the randomness of the noise section, the deblending result of the field data test is acceptable. In this case, the percentage we use for seislet domain thresholding is 18%.

The main computational cost of the proposed approach lays on the forward and inverse seislet transforms. As analyzed by (21), the seislet transform has linear $O(N)$ cost. It can be more expensive than fast Fourier transform and digital wavelet transform, but is still comfortably efficient in practice. Considering that for the 2D seislet transform, the main cost may not be in the transform itself but in the iterative estimation of the slope fields, we update slope estimation every several iterations (e.g., every 5 iterations in this field data example) to reduce the computational cost.

CONCLUSIONS

We have proposed a novel iterative framework, which offers a flexible way to control deblending using sparsity or coherency constraints. When the seismic data are relatively clean and do not contain much coherent noise, $f-k$ domain sparsity-promoting operator or $f-x$ predictive filtering might be adequate to handle the blending noise. However, when the blended data become more complicated, a more robust sparsity or coherency-promoting tool should be utilized. The seislet transform has an inherent ability to compress coherent seismic data, because it is generated based on predictability between neighboring traces using the local slope information. As long as we can get an acceptable slope estimation, the seislet-transformed data appears sparse. Our experiments indicate that it is possible to get accurate results within a small number of iterations when an appropriate shaping operator is taken.

ACKNOWLEDGMENTS

We thank FairfieldNodal for providing the field data used in this study. We thank Alexander Klokov, Karl Schleicher, Parvaneh Karimi, Araz Mahdad, Josef Paffenholz, David Hays, Paul Docherty and Le-wei Mo for helpful discussions. We also thank Sam Kaplan, Chengbo Li, Jinkun Cheng, Jeff Shragge, Evert Slob, and one anonymous reviewer for their constructive suggestions, which helped improve this paper.

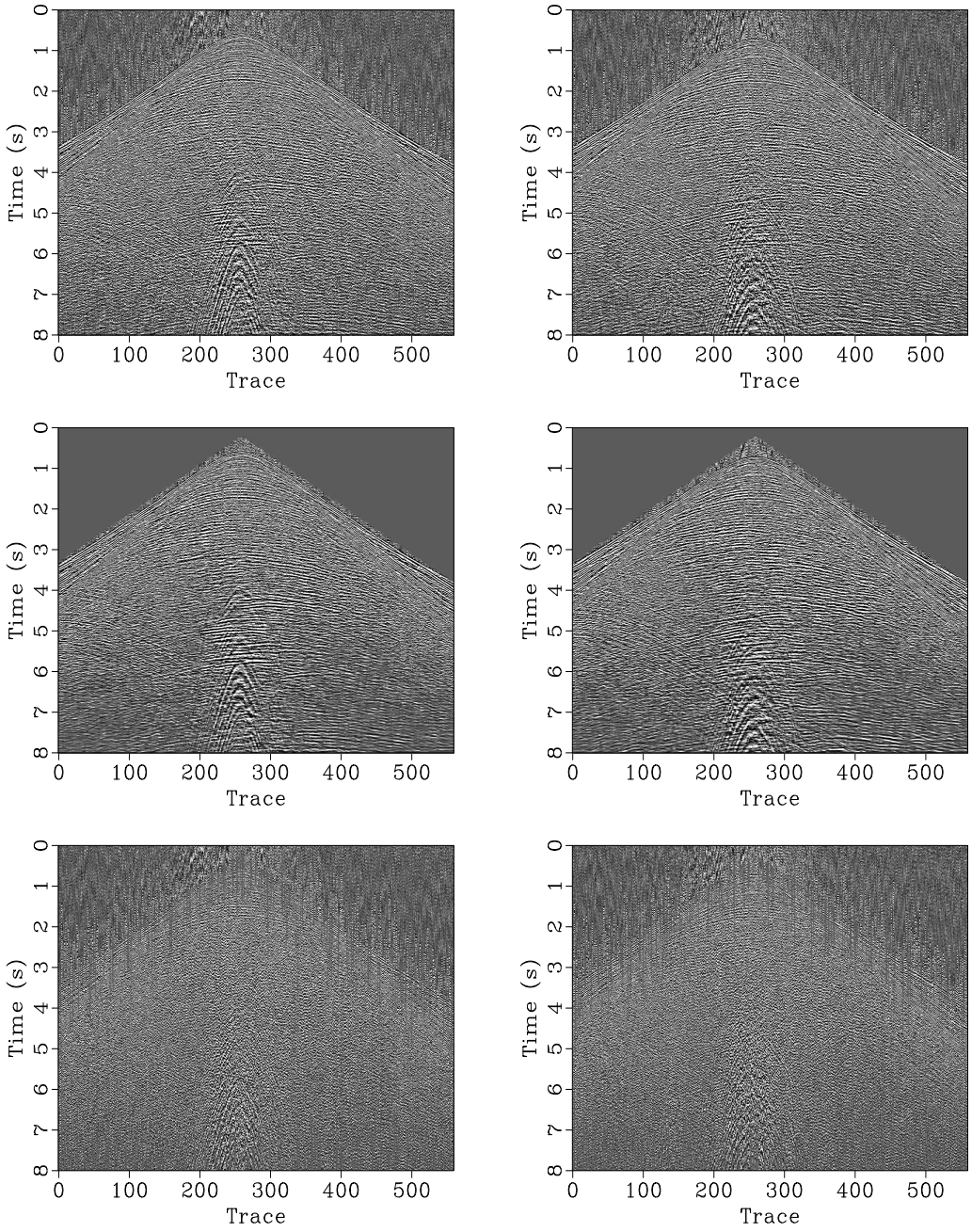


Figure 12: Deblending comparison for the real data. (a) Blended source 1. (b) Blended source 2. (c) Deblended source 1. (d) Deblended source 2. (e) Blending noise for source 1. (f) Blending noise for source 2.

deblend/fairfield-initmfnew blended763,blended754,shotdeblendedslet2,shotdeblendedslet1,shotdiffslet

REFERENCES

[1] Abma, R., Q. Zhang, A. Arogunmati, and G. Beaudoin, 2012, An overview of BP's marine independent simultaneous source field trials: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1404.1.

[2] Abma, R. L., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: 72nd Annual International Conference and Exhibition, EAGE, Extended Abstracts.

[3] Beasley, C. J., B. Dragoset, and A. Salama, 2012, A 3d simultaneous source field test processed using alternating projections: a new active separation method: Geophysical Prospecting, **60**, 591–601.

[4] Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.

[5] Berkhout, A. J., D. J. Verschuur, and G. Blacquière, 2012, illumination properties and imaging promises of blended, multiple-scattering seismic data: a tutorial: Geophysical Prospecting, **60**, 713–732.

[6] Canales, L., 1984, Random noise reduction: SEG expanded abstracts: 54th Annual international meeting, 525–527.

[7] Candès, E. J., and Y. Plan, 2010, A probabilistic and ripless theory of compressed sensing: IEEE Transactions on Information Theory, **57**, 7235–7254.

[8] Chen, Y., and J. Ma, 2014, Random noise attenuation by *f-x* empirical mode decomposition predictive filtering: Geophysics, **79**, V81–V91.

[9] Choi, Y., and T. Alkhalifah, 2012, Application of multi-source waveform inversion to marine streamer data using the global correlation norm: Geophysical Prospecting, **60**, 748–758.

[10] Cohen, A., I. Daubechies, and J. C. Feauveau, 1992, Biorthogonal bases of compactly supported wavelets: Communications on Pure and Applied Mathematics, **45**, 485–560.

[11] Daubechies, I., M. Defrise, and C. D. Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: Communications on Pure and Applied Mathematics, **57**, 1413–1457.

[12] Daubechies, I., M. Fornasier, and I. Loris, 2008, Accelerated projected gradient method for linear inverse problems with sparsity constraints: Journal of Fourier Analysis and Applications, **14**, 764–792.

[13] Donoho, D. L., 1995, De-noising by soft-thresholding: IEEE transactions on information theory, **41**, 613–627.

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

[15] Doulgeris, P., and K. Bube, 2012, Analysis of a coherency-constrained inversion for the separation of blended data: discovering the leakage subspace: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts.

[16] Doulgeris, P., K. Bube, G. Hampson, and G. Blacquiere, 2012, Covergence analysis of a coherency-constrained inversion for the separation of blended data: Geophysical Prospecting, **60**, 769–781.

[17] Fomel, S., 2002, Application of plane-wave destruction filters: Geophysics, **67**,

1946–1960.
[18] ———, 2006, Towards the seislet transform: 76th Annual International Meeting, SEG, Expanded Abstracts, 2847–2850.
[19] ———, 2007, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.
[20] ———, 2008, Nonlinear shapping regularization in geophysical inverse problems: 78th Annual International Meeting, SEG, Expanded Abstracts, 2046–2051.
[21] Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, V25–V38.
[22] Galbraith, M., and Z. Yao, 2012, Fx plus deconvolution for seismic data noise reduction: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–4.
[23] Gao, J., X. Chen, G. Liu, and J. Ma, 2010, Irregular seismic data reconstruction based on exponentional threshold model of pocs method: Applied Geophysics, **7**, 229–238.
[24] Guitton, A., and E. Diaz, 2012, Attenuating crosstalk noise with simultaneous source full waveform inversion: Geophysical Prospecting, **60**, 759–768.
[25] Hampson, G., J. Stefani, and F. Herkenhoff, 2008, Acquisition using simultaneous sources: The Leading Edge, **27**, 918–923.
[26] Hennenfent, G., and F. Herrmann, 2006, Seismic denoising with nonunformly sampled curvelets: Computing in Science & Engineering, **8**, 16–25.
[27] Huo, S., Y. Luo, and P. G. Kelamis, 2012, Simultaneous sources separation via multidirectional vector-median filtering: Geophysics, **77**, V123–V131.
[28] Liu, Y., 2013, Noise reduction by vector median filtering: Geophysics, **78**, V79–V87.
[29] Liu, Y., and S. Fomel, 2012, Seismic data analysis using local time-frequency decomposition: Geophyscial Prospecting, **60**, 1–10.
[30] Liu, Y., S. Fomel, C. Liu, D. Wang, G. Liu, and X. Feng, 2009a, High-order seislet transform and its application of random noise attenuation: Chinese Journal of Geophysics, **52**, 2142–2151.
[31] Liu, Y., C. Liu, and D. Wang, 2009b, A 1d time-varying median filter for seismic random, spike-like noise elimination: Geophysics, **74**, V17–V24.
[32] Mahdad, A., 2012, Deblending of seismic data: TU Delft, PhD thesis.
[33] Mahdad, A., P. Doulgeris, and G. Blacquiere, 2011, Separation of blended data by iterative estimation and subtraction of blending interference noise: Geophysics, **76**, Q9–Q17.
[34] ———, 2012, Iterative method for the separation of blended seismic data: discussion on the algorithmic aspects: Geophysical Prospecting, **60**, 782–801.
[35] Mallat, S. G., 2009, A wavelet tour of signal processing: The sparse way: Academic Press.
[36] Mitchell, S., C. Ray, E. Marc, D. Hays, and K. Craft, 2010, Fairfieldnodal's excellent nodal adventure: 80th Annual International Meeting, SEG, Expanded Abstracts, 3740–3745.
[37] Moore, I., B. Dragoset, T. Ommundsen, D. Wilson, C. Ward, and D. Eke, 2008, Simultaneous source separation using dithered sources: 78th Annual International Meeting, SEG, Expanded Abstracts, 2806–2809.

[38] Naghizadeh, M., and M. D. Sacchi, 2010, Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data: Geophysics, **75**, WB189–WB202.

[39] Plessix, R.-E., G. Baeten, J. W. de Maag, F. ten Kroode, and Z. Rujie, 2012, Full waveform inversion and distance separated simultaneous sweeping: a study with a land seismic data set: Geophysical Prospecting, **60**, 733–747.

[40] Sacchi, M. D., and H. Kuehl, 2000, Fx arma filters: 70th Annual International Meeting, SEG, Expanded Abstracts, 2092–2095.

[41] Sweldens, W., 1995, Lifting scheme: A new philosophy in biorthogonal wavelet constructions: Wavelet applications in signal and image processing iii: Proceedings of SPIE 2569, 68–79.

[42] van Borselen, R., R. Baardman, T. Martin, B. Goswami, and E. Fromyr, 2012, An inversion approach to separating sources in marine simultaneous shooting acquision-application to a gulf of mexico data set: Geophysical Prospecting, **60**, 640–647.

[43] Wang, D., C. Liu, Y. Liu, and G. Liu, 2008, Application of wavelet transform based on lifting scheme and percentiles soft-threshold to elimination of seismic random noise: Progress in Geophysics (in Chinese), **23**, 1124–1130.

[44] Wapenaar, K., J. van der Neut, and J. Thorbecke, 2012, Deblending by direct inversion: Geophysics, **77**, A9–A12.

[45] Xue, Z., Y. Chen, S. Fomel, and J. Sun, 2014, Imaging incomplete data and simultaneous-source data using least-squares reverse-time migration with shaping regularization: 84th Annual International Meeting, SEG, Expanded Abstracts, submitted.

[46] Yang, P., J. Gao, and W. Chen, 2012, Improved pocs interpolation using a percentage thresholding strategy and excessively zero-padded fft: 82nd Annual International Meeting, SEG, Expanded Abstracts, 1–6.

[47] ———, 2013, On analysis-based two-step interpolation methods for randomly sampled seismic data: Computers & Geosciences, **51**, 449–461.

APPENDIX A

REVIEW OF SEISLET TRANSFORM

(18) and (21) proposed a digital wavelet-like transform, which is defined with the help of the wavelet-lifting scheme (41) combined with local plane-wave destruction. The wavelet-lifting utilizes predictability of even traces from odd traces of 2-D seismic data and finds a difference $\mathbf{r}$ between them, which can be expressed as:

$$\mathbf{r} = \mathbf{o} - \mathbf{P}\left[\mathbf{e}\right], \tag{A-1}$$

where $\mathbf{P}$ is the prediction operator. A coarse approximation $\mathbf{c}$ of the data can be achieved by updating the even component:

$$\mathbf{c} = \mathbf{e} + \mathbf{U}\left[\mathbf{r}\right], \tag{A-2}$$

where $\mathbf{U}$ is the updating operator.

The digital wavelet transform can be inverted by reversing the lifting-scheme operations as follows:

$$\mathbf{e} = \mathbf{c} - \mathbf{U}\left[\mathbf{r}\right], \tag{A-3}$$

$$\mathbf{o} = \mathbf{r} + \mathbf{P}\left[\mathbf{e}\right]. \tag{A-4}$$

The foward transform starts with the finest scale (the original sampling) and goes to the coarsest scale. The inverse transfrom starts with the coarsest scale and goes back to the finest scale. At the start of forward transform, $\mathbf{e}$ and $\mathbf{o}$ corresponds to the even and odd traces of the data domain. At the start of the inverse transform, $\mathbf{c}$ and $\mathbf{r}$ will have just one trace of the coarsest scale of the seislet domain.

The above prediction and update operators can be defined, for example, as follows:

$$\mathbf{P}\left[\mathbf{e}\right]_k = \left(\mathbf{P}_k^{(+)}\left[\mathbf{e}_{k-1}\right] + \mathbf{P}_k^{(-)}\left[\mathbf{e}_k\right]\right)/2, \tag{A-5}$$

and

$$\mathbf{U}\left[\mathbf{r}\right]_k = \left(\mathbf{P}_k^{(+)}\left[\mathbf{r}_{k-1}\right] + \mathbf{P}_k^{(-)}\left[\mathbf{r}_k\right]\right)/4, \tag{A-6}$$

where $\mathbf{P}_k^{(+)}$ and $\mathbf{P}_k^{(-)}$ are operators that predict a trace from its left and right neighbors, correspondingly, by shifting seismic events according to their local slopes. This scheme is analogous to CDF biorthogonal wavelets (10). The predictions need to operate at different scales, which means different separation distances between traces. Taken through different scales, equations A-1-A-6 provide a simple definition for the 2D seislet transform. More accurate versions are based on other schemes for the digital wavelet transform (30).

Seismic data decomposition into spectral components using regularized nonstationary autoregression

*Sergey Fomel**

ABSTRACT

Seismic data can be decomposed into nonstationary spectral components with smoothly variable frequencies and smoothly variable amplitudes. To estimate local frequencies, I use a nonstationary version of Prony's spectral analysis method defined with the help of regularized nonstationary autoregression (RNAR). To estimate local amplitudes of different components, I fit their sum to the data using regularized nonstationary regression (RNR). Shaping regularization ensures stability of the estimation process and provides controls on smoothness of the estimated parameters. Potential applications of the proposed technique include noise attenuation, seismic data compression, and seismic data regularization.

INTRODUCTION

Decomposing data into components has an immediate application in noise-attenuation problems in cases where signal and noise correspond to different components. The classic Fourier transform, Radon transform (13), wavelet transform (27), curvelet frame (17), and seislet transform and frame (12) are some examples of possible decompositions applicable to seismic data. A fundamental characteristic of seismic data is non-stationarity. In 1D (time dimension), seismic data are nonstationary because of wave-attenuation effects. In 2D and 3D (time and space dimensions), non-stationarity is manifested by variable slopes of seismic events. The nonstationary character of seismic data can be captured by EMD (empirical mode decomposition) proposed by (20). EMD has found a number of important applications in seismic data analysis (26; 2; 3; 15). However, it remains "empirical" because its properties are not fully understood. (7) recently proposed an EMD-like decomposition using the continuous wavelet transform and *synchrosqueezing* (8). Synchrosqueezing improves the analysis but remains an indirect method when it comes to extracting spectral attributes (16).

In this paper, I develop an efficient decomposition algorithm, which explicitly fits seismic data to a sum of oscillatory signals with smoothly varying frequencies and smoothly varying amplitudes. Such a decomposition is close in properties to the one generated by EMD but with explicit controls on the frequencies and amplitudes of

*e-mail: sergey.fomel@beg.utexas.edu

different components and on their smoothness. Recently, (18; 19) developed an explicit data-adaptive decomposition based on matching-pursuit sparse optimization, an accurate but computationally expensive method. To implement a faster approach, I adopt regularized nonstationary regression, or RNR (11), a general method for fitting data to a set of basis functions with nonstationary coefficients. RNR was previously applied to time-frequency decomposition over a set of regularly sampled frequencies (25). When the input signal is fitted to shifted versions of itself, RNR turns into regularized nonstationary autoregression, or RNAR, and is related to adaptive prediction-error filtering. RNAR was previously applied to data regularization (24) and noise removal (22; 21). In this paper, I use it for spectral analysis and estimating different frequencies present in the data using a nonstationary extension of Prony's method of autoregressive spectral analysis (28; 1). After the frequencies have been identified, I use RNR to determine local, smoothly-varying amplitudes of different components.

The paper opens with a brief review of RNR and RNAR and explains an extension of Prony's method to the nonstationary case. Next, I use simple synthetic and field-data examples to illustrate performance of the proposed technique.

REGULARIZED NONSTATIONARY REGRESSION

Regularized nonstationary regression (11) is based on the following simple model. Let $d(\mathbf{x})$ represent the data as a function of data coordinates $\mathbf{x}$, and $b_n(\mathbf{x})$, $n = 1, 2, \ldots, N$, represent a collection of basis functions. The goal of *stationary* regression is to estimate coefficients a_n, $n = 1, 2, \ldots, N$ such that the prediction error

$$e(\mathbf{x}) = d(\mathbf{x}) - \sum_{n=1}^{N} a_n \, b_n(\mathbf{x}) \tag{1}$$

is minimized in the least-squares sense. In the case of regularized *nonstationary* regression (RNR), the coefficients become variable,

$$\hat{e}(\mathbf{x}) = d(\mathbf{x}) - \sum_{n=1}^{N} \hat{a}_n(\mathbf{x}) \, b_n(\mathbf{x}) \ . \tag{2}$$

The problem in this case is underdetermined but can be constrained by regularization (9). I use shaping regularization (10) to implement an explicit control on the resolution and variability of regression coefficients. Shaping regularization applied to RNR amounts to linear inversion,

$$\mathbf{a} = \mathbf{M}^{-1} \, \mathbf{c} \ , \tag{3}$$

where $\mathbf{a}$ is a vector composed of $\hat{a}_n(\mathbf{x})$, the elements of vector $\mathbf{c}$ are

$$c_i(\mathbf{x}) = \mathbf{S} \left[b_i^*(\mathbf{x}) \, d(\mathbf{x}) \right] \ , \tag{4}$$

the elements of matrix $\mathbf{M}$ are

$$M_{ij}(\mathbf{x}) = \lambda^2 \, \delta_{ij} + \mathbf{S} \left[b_i^*(\mathbf{x}) \, b_j(\mathbf{x}) - \lambda^2 \, \delta_{ij} \right] \ , \tag{5}$$

λ is a scaling coefficient, and $\mathbf{S}$ represents a shaping (typically smoothing) operator. When inversion in equation 6 is implemented by an iterative method, such as conjugate gradients, strong smoothing makes $\mathbf{M}$ close to identity and easier (taking less iterations) to invert, whereas weaker smoothing slows down the inversion but allows for more details in the solution. This intuitively logical behavior distinguishes shaping regularization from alternative methods (11).

Regularized nonstationary autoregression (RNAR) corresponds to the case of basis functions being causal translations of the input data itself. In 1D, with $\mathbf{x} = t$, this condition implies $b_n(t) = d(t - n \, \Delta t)$.

AUTOREGRESSIVE SPECTRAL ANALYSIS

Prony's method of data analysis was developed originally for representing a noiseless signal as a sum of exponential components (32). It was extended later to noisy signals, complex exponentials, and spectral analysis (31; 28; 1; 4). The basic idea follows from the fundamental property of exponential functions: $e^{\alpha \, (t + \Delta t)} = e^{\alpha \, t} \, e^{\alpha \, \Delta t}$. In signal-processing terms, it implies that a time sequence $d(t) = A \, e^{\alpha \, t}$ (with real or complex α) is predictable by a two-point prediction-error filter $\left(1, -e^{\alpha \, \Delta t}\right)$, or, in the Z-transform notation,

$$F_0(Z) = 1 - Z/Z_0 \ , \tag{6}$$

where $Z_0 = e^{-\alpha \, \Delta t}$. If the signal is composed of multiple exponentials,

$$d(t) \approx \sum_{n=1}^{N} A_n \, e^{\alpha_n \, t} \ , \tag{7}$$

they can be predicted simultaneously by using a convolution of several two-point prediction-error filters:

$$\begin{aligned} F(Z) &= (1 - Z/Z_1) \, (1 - Z/Z_2) \cdots (1 - Z/Z_N) \\ &= 1 + a_1 \, Z + a_2 \, Z^2 + \cdots + a_n \, Z^N \ , \end{aligned} \tag{8}$$

where $Z_n = e^{-\alpha_n \, \Delta t}$. This observation suggests the following three-step algorithm:

1. Estimate a prediction-error filter from the data by determining filter coefficients $a_1, a_2, \ldots, a_N$ from the least-squares minimization of

$$e(t) = d(t) - \sum_{n=1}^{N} a_n \, d(t - n \, \Delta t) \ . \tag{9}$$

2. Writing the filter as a Z polynomial (equation 8), find its complex roots $Z_1, Z_2, \ldots, Z_N$. The exponential factors $\alpha_1, \alpha_2, \ldots, \alpha_N$ are determined then as

$$\alpha_n = -\ln(Z_n)/\Delta t \ . \tag{10}$$

3. Estimate amplitudes $A_1, A_2, \ldots, A_N$ of different components in equation 7 by linear least-squares fitting.

Prony's method can be applied in sliding windows, which was a technique developed by Russian geophysicists (14; 30) for identifying low-frequency seismic anomalies (29). I propose to extend it to smoothly nonstationary analysis by applying the following modifications:

1. Using RNAR, the filter coefficients a_n become smoothly-varying functions of time $\hat{a}_n(t)$, which allows the filter to adapt to nonstationary changes in the input data.

2. At each instance of time, roots of the corresponding Z polynomial also become functions of time $\hat{Z}_n(t)$. I apply a robust, eigenvalue-based algorithm for root finding (36). The instantaneous frequency of different components $f_n(t)$ is determined directly from the phase of different roots:

$$f_n(t) = -Re\left[\arg\left(\frac{\hat{Z}_n(t)}{2\pi\,\Delta t}\right)\right] \ . \tag{11}$$

3. Finally, using RNR, I estimate smoothly-varying amplitudes of different components $\hat{A}_n(t)$.

The nonstationary decomposition model for a complex signal $d(t)$ is thus

$$d(t) \approx \sum_{n=1}^{N} d_n(t) \ , \quad \text{where} \quad d_n(t) = \hat{A}_n(t)\, e^{i\,\phi_n(t)} \tag{12}$$

and the local phase $\phi_n(t)$ corresponds to time integration of the instantaneous frequency determined in Step 2:

$$\phi_n(t) = 2\pi \int_0^t f_n(\tau) d\tau \ . \tag{13}$$

For ease of analysis, real signals can be transformed to the complex domain by using analytical traces (35).

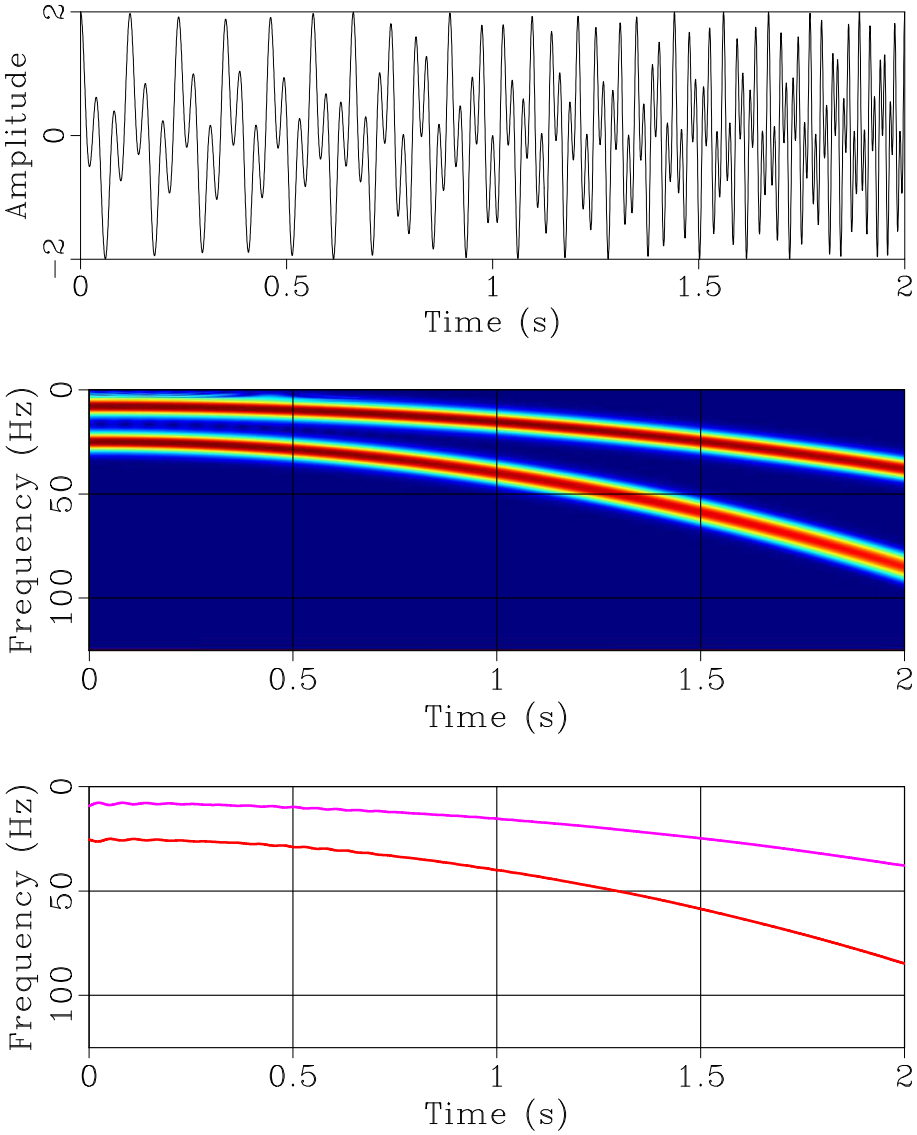


Figure 1: (a) Test signal from (23) composed of two variable-frequency components. (b) Time-frequency decomposition. (c) Instantaneous frequencies estimated by RNAR. nar/chirp sig,tf,group

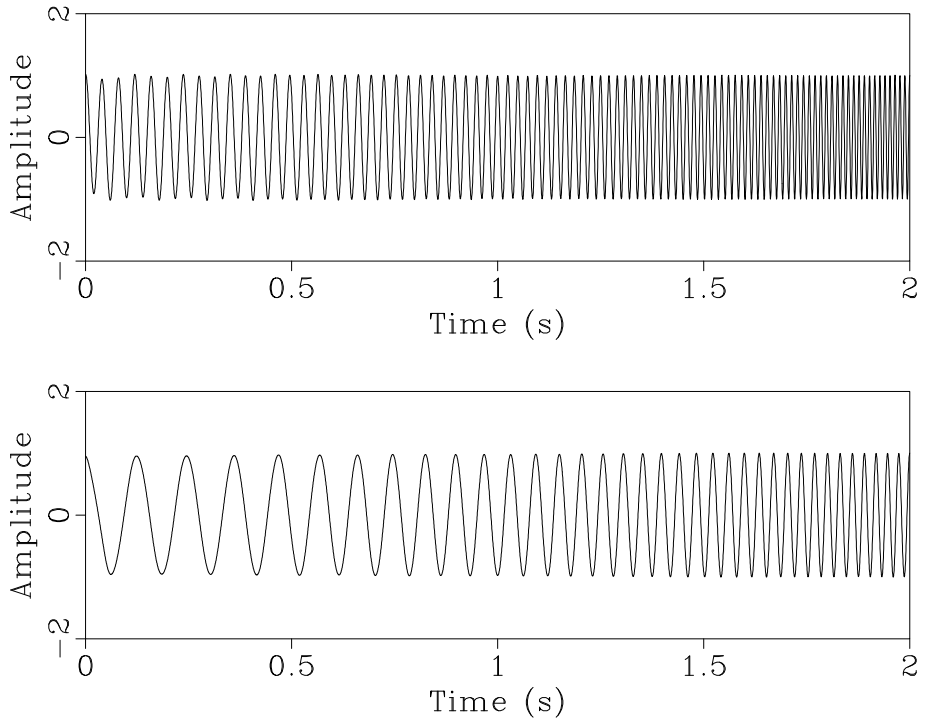


Figure 2: Decomposition of the signal from Figure 15(b) into two spectral components using RNR. nar/chirp sign1,sign2

Benchmark tests

Figure 1a shows a benchmark signal from (23), which consists of two nonstationary components with smoothly varying (parabolic) frequencies. The corresponding time-frequency analysis over a range of regularly sampled frequencies is shown in Figure 1b. Two instantaneous frequencies were extracted at Step 2 of the algorithm from a time-varying, three-point prediction-error filter (Figure 1c). They correspond precisely to the two components present in the synthetic signal. Finally, Step 3 separates these components (Figure 2.)

The next benchmark example is taken from (16). Figure 3(a) shows the input signal, which is composed of several components with variable frequencies. The components can be identified in the time-frequency decomposition analysis (Figure 3(b)) generated with the method of (23). More directly, they are extracted using RNAR (the method of this paper), with the output shown in Figure 3(c). Fitting a sum of different components to the data by RNR, we can estimate their respective amplitudes. The output is shown in Figure 4. For comparison, I show the output of EMD (empirical mode decomposition) in Figure 5. For robustness, I used EEMD (ensemble EMD) suggested by (37). Although EEMD succeeds in separating the signal into components with variable frequencies, the individual components are not as meaningful as those identified by RNAR.

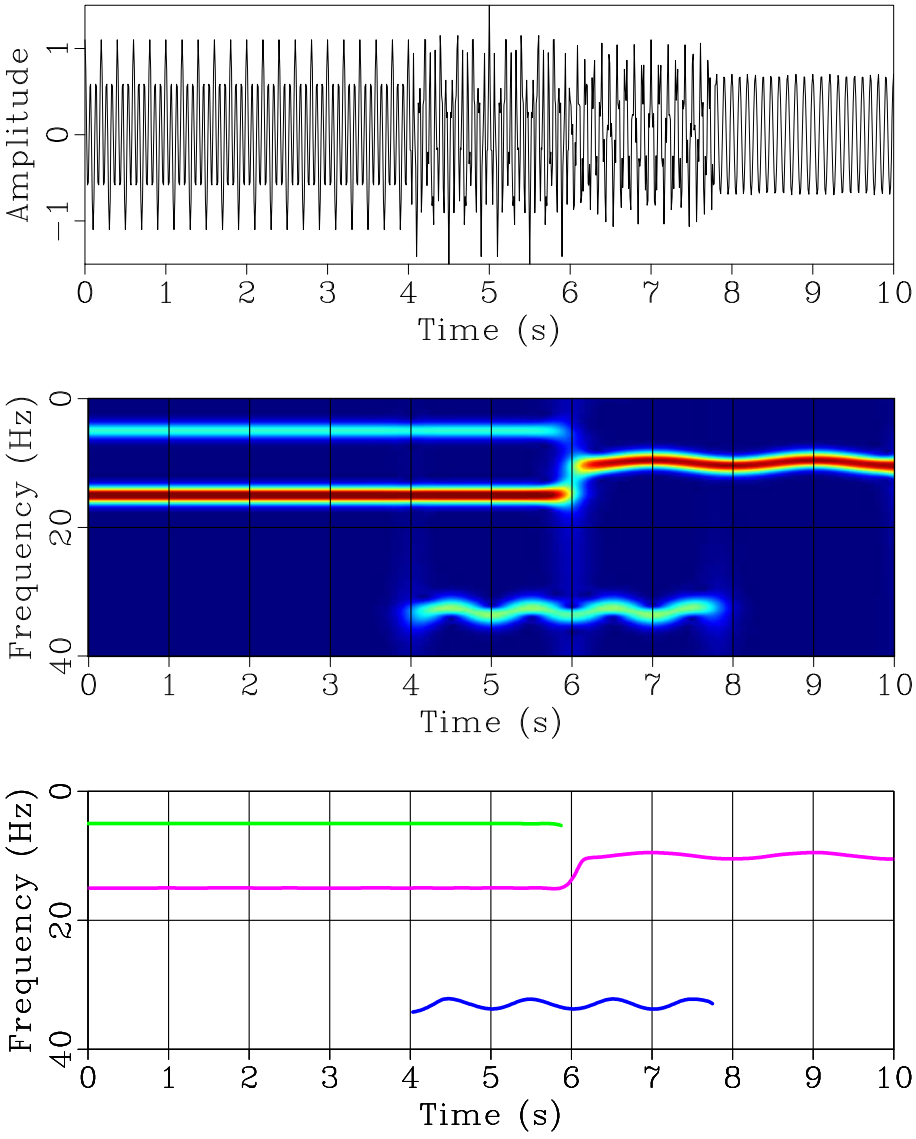


Figure 3: (a) Test signal from (16) composed of several variable-frequency components. (b) Time-frequency decomposition. (c) Instantaneous frequencies estimated by RNAR. nar/mirko msig,mtf,mgroup

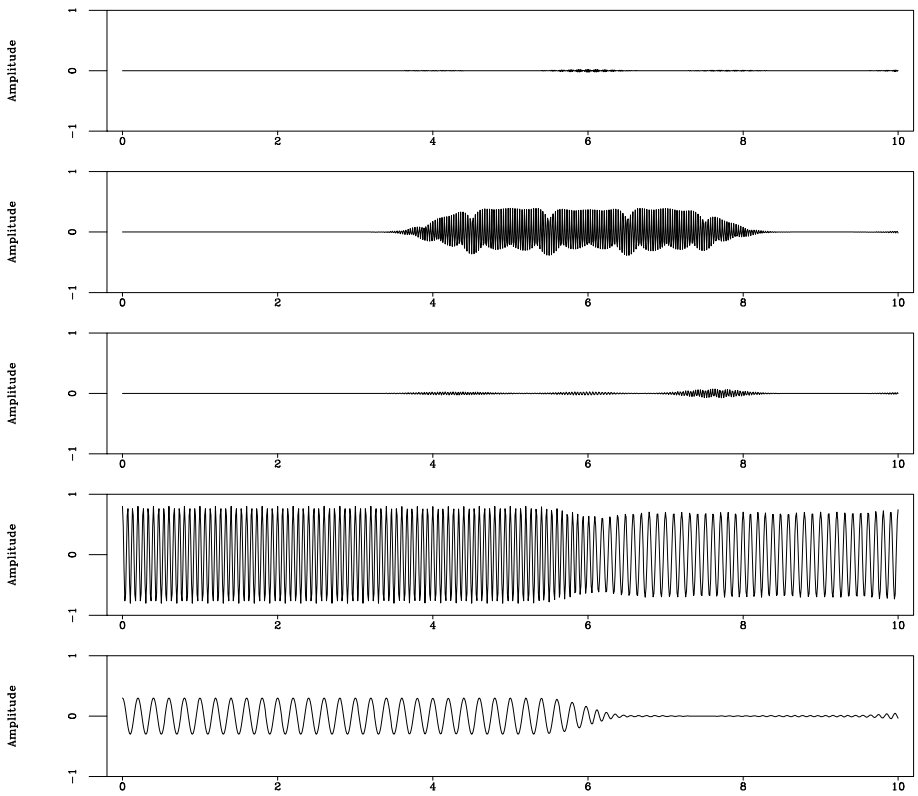


Figure 4: Decomposition of the signal from Figure 3(a) into spectral components using RNR. The components are sorted in the order of decreasing frequency. nar/mirko narall

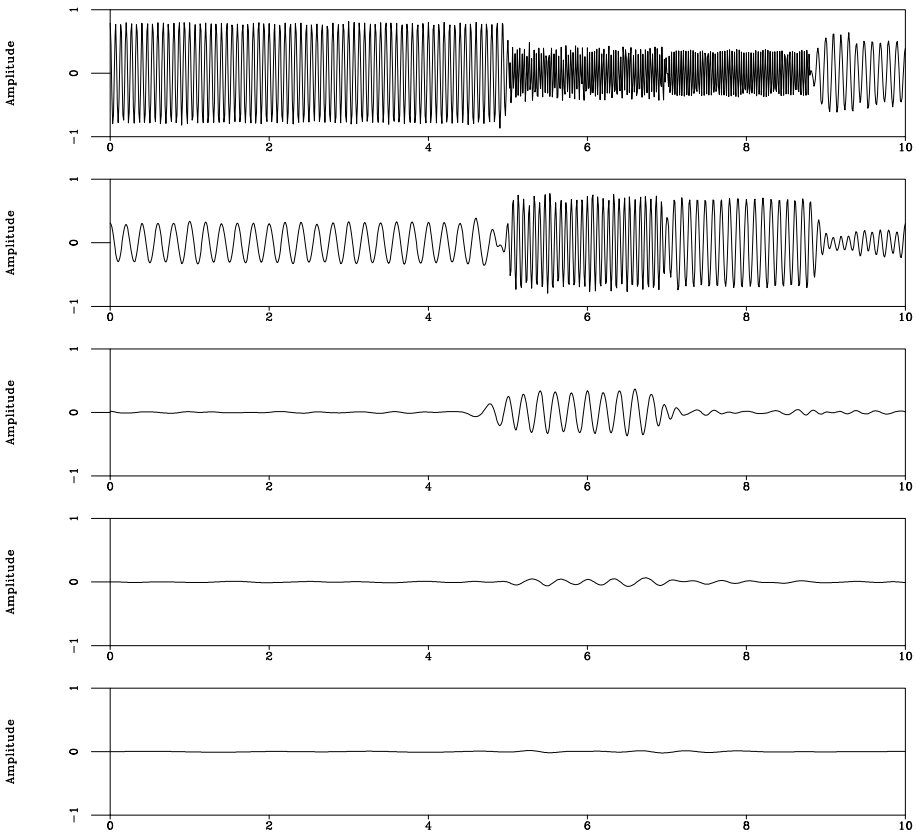


Figure 5: Decomposition of the signal from Figure 3(a) into intrinsic mode functions using EMD. Compare with Figure 4. nar/mirko emd

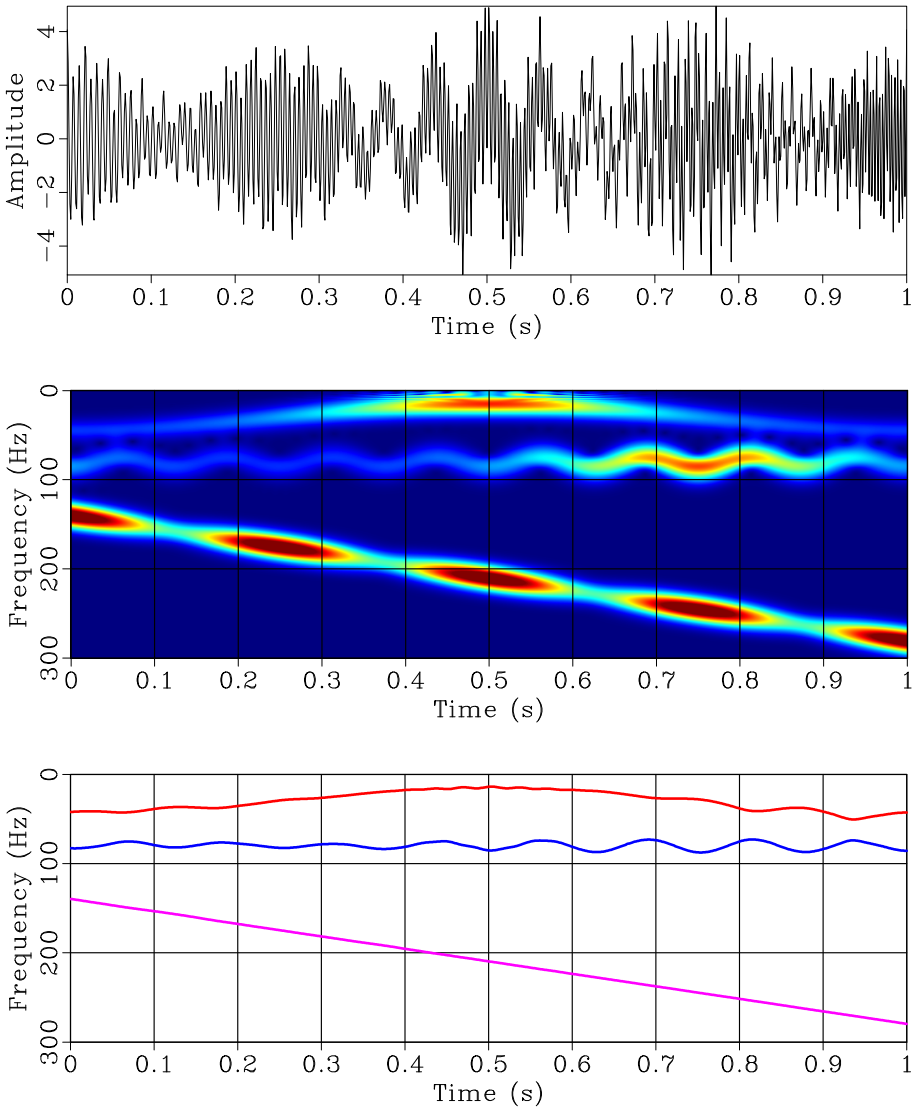


Figure 6: (a) Test signal from (19) composed of several variable-frequency and variable-amplitude components. (b) Time-frequency decomposition. (c) Instantaneous frequencies estimated by RNAR. nar/hou hsig,htf,hgroup

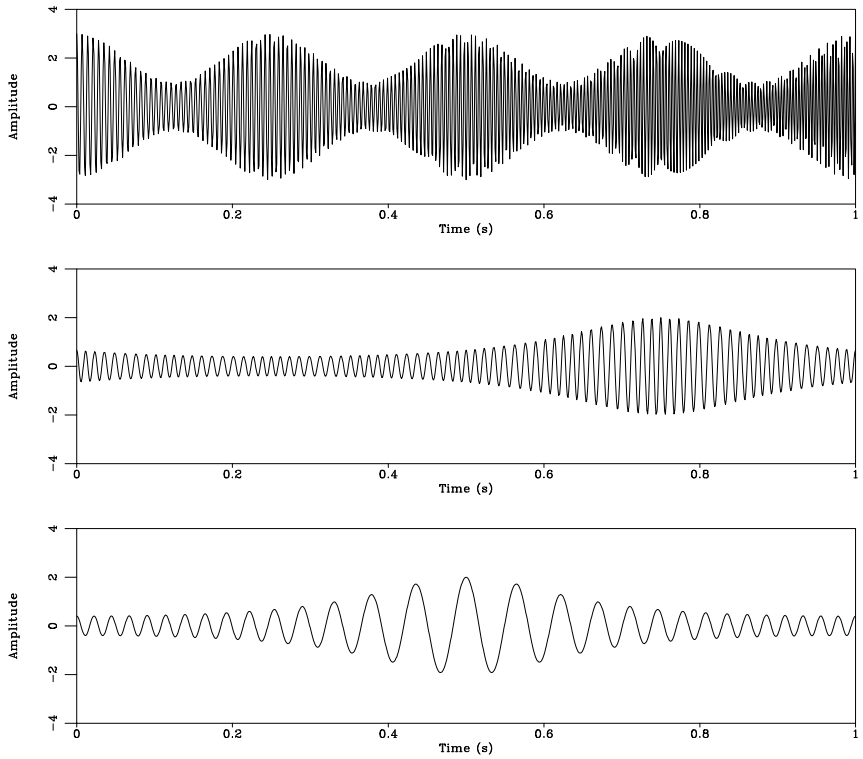


Figure 7: Signal components making the signal in Figure 6(a). nar/hou sigs

The third benchmark test is taken from (19). The signal in this case (Figure 6(a)) consists of three components with variable frequencies and amplitudes (Figure 7). RNAR correctly identifies the three components (Figure 6(c)) that are also visible on the time-frequency plot (Figure 6(b)). Next, RNR extracts the three components (Figure 8) by fitting their sum to the data (Figure 9) and adjusting the amplitudes. The sparse inversion algorithm proposed by (19) achieves a more accurate result in this case but at a much higher computational cost.

Discussion

The cost of the proposed decomposition is $O\left(N\, N_t\, N_{iter}\right)$, where N is the number of components, N_t is the number of time samples, and N_{iter} is the number of conjugate-gradient iterations for shaping regularization (typically between 10 and 100). This is significantly faster than the $O\left(N_t^2\, N_{iter}\right)$ cost of time-frequency decomposition for a regularly sampled range of frequencies.

Although the examples of this paper use only 1D analysis, the proposed technique is also directly applicable to analyzing variable slopes of 2D and 3D seismic events, where the analysis applies to different frequency slices in the f-x domain (5; 33; 34; 22; 21).

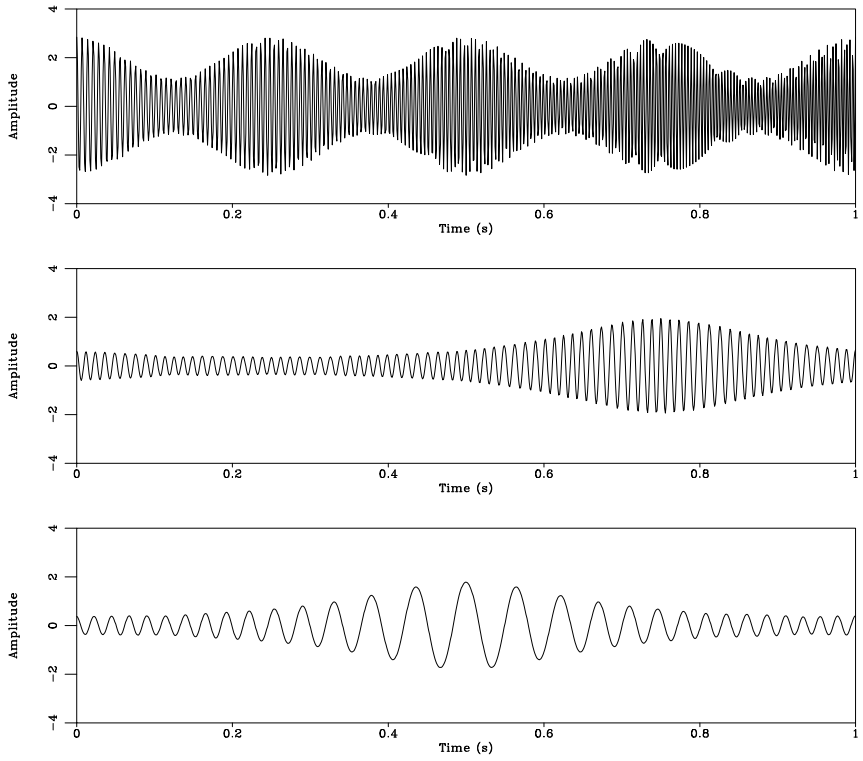


Figure 8: Decomposition of the signal from Figure 6(a) into spectral components using RNR. nar/hou nar

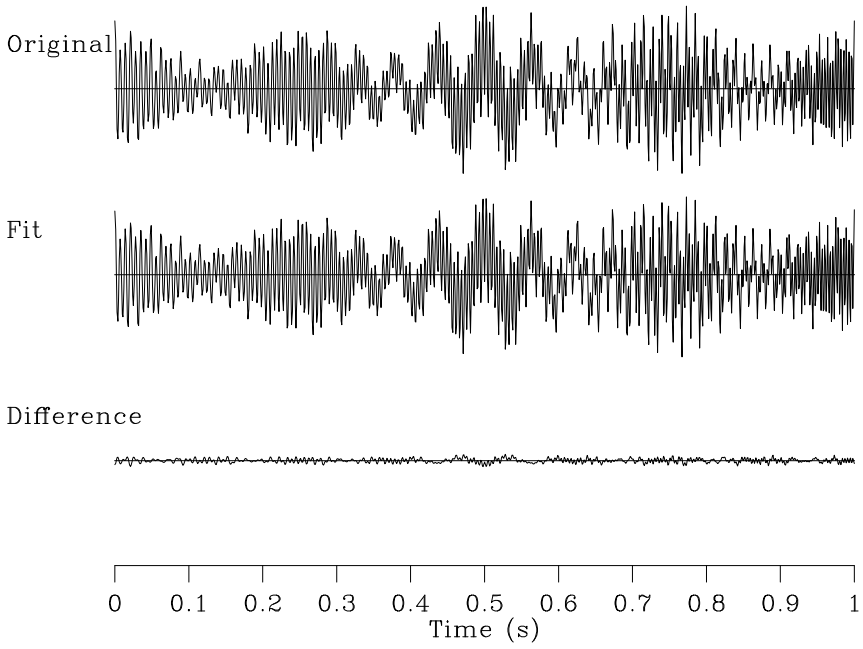


Figure 9: Fitting the signal from Figure 6(a) with the sum of three components shown in Figure 8. nar/hou hfit

EXAMPLES

To illustrate performance of the proposed approach in field-data applications, I first use a simple 1D example: a single seismic trace from a marine survey. Figure 10 shows the input trace and the output of RNAR, with a five-point adaptive prediction-error filter. The four variable instantaneous frequencies extracted from the roots of the filter are shown in Figure 11. They correspond to four different spectral components extracted from the data in Step 3 (Figure 12.) Surprisingly, only four components with smoothly varying frequencies and amplitudes are sufficient to describe a significant portion of the signal, including the effect of attenuating frequencies at later times (Figure 13.)

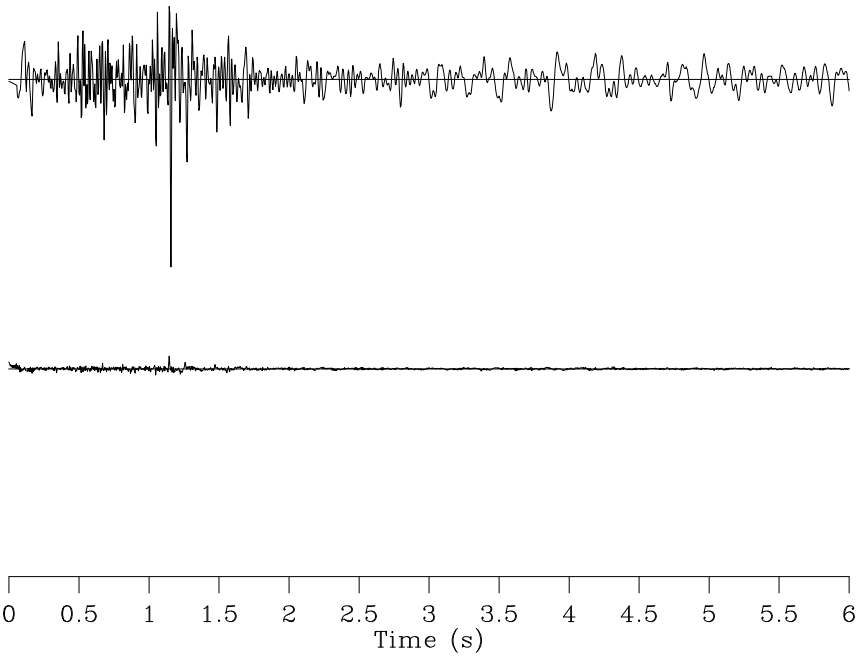


Figure 10: Seismic trace and residual after adaptive prediction-error filtering with RNAR. nar/trace cerr

The second example is a 2D section from a land seismic survey (Figure 14a), analyzed previously by (10) and (25). I choose a three-point prediction-error filter to highlight the two most significant data components. The fitting error is shown in Figure 15 and contains mostly random noise. The two estimated spectral components are shown in Figure 16, with the corresponding instantaneous frequencies $f_n)t)$ shown in Figure 17. The corresponding amplitudes $|\hat{A}_n(t)|$ are shown in Figure 18. Comparing frequency and amplitude attributes from different components, a low-frequency anomaly (a zone of attenuated high frequencies) in the top-left part of the section becomes apparent. This anomaly might indicate presence of gas (6).

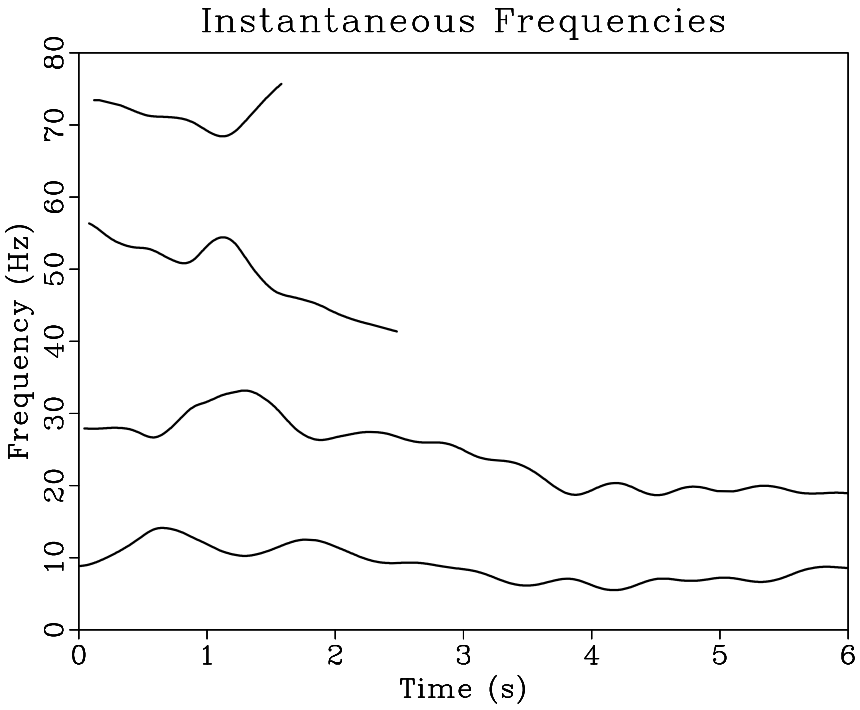


Figure 11: Instantaneous frequencies of four components extracted from seismic trace in Figure 10 using RNAR. nar/trace tgroup

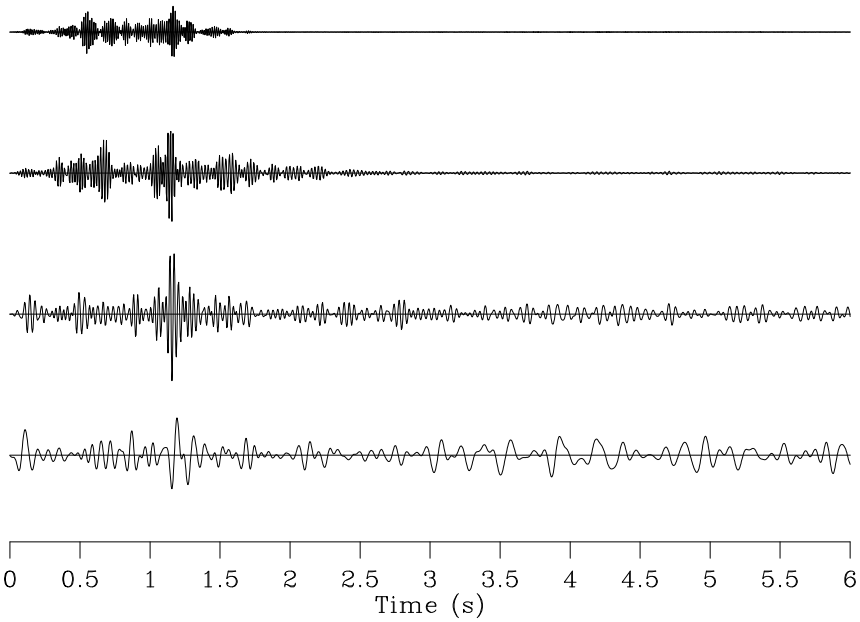


Figure 12: Four nonstationary spectral components corresponding to frequencies in Figure 11. nar/trace csign

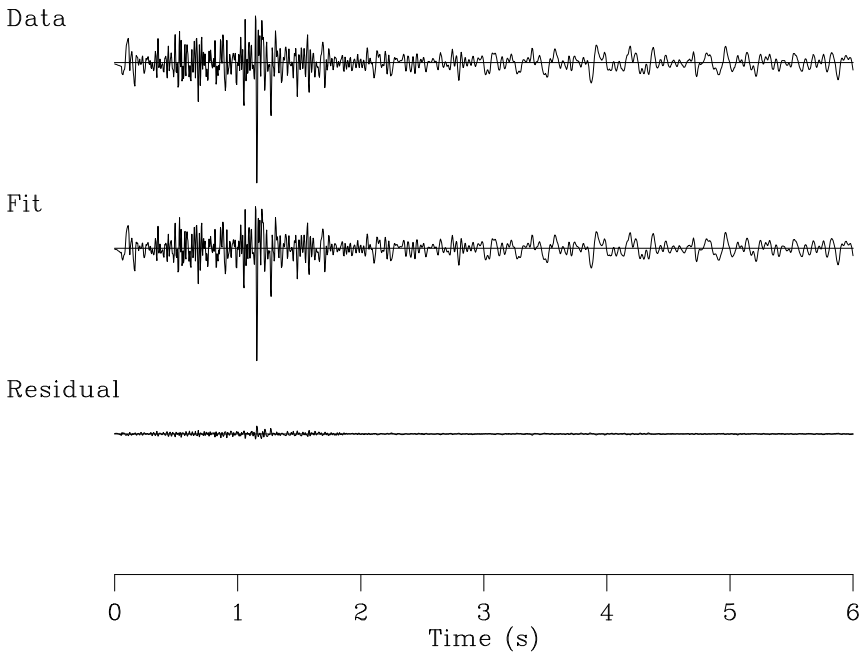


Figure 13: Fitting input seismic trace with sum of four spectral components shown in Figure 12. nar/trace cfit

CONCLUSIONS

I have presented a constructive approach to decomposing seismic data into spectral components with smoothly variable frequencies and smoothly variable amplitudes. The output of the proposed algorithm is close to that of empirical model decomposition (EMD) and related techniques, such as the synchrosqueezing transform (SST), but with a more explicit control on parameters and more direct access to instantaneous-frequency and amplitude attributes. The main tool for the task is regularized nonstationary regression (RNR), which is applied twice: first to estimate local frequencies by autoregression (RNAR) and then to estimate local amplitudes. Although all examples shown in this paper use only 1D analysis, the proposed technique is also applicable to analyzing 2D or 3D variable-slope seismic events in the f-x domain. Potential applications may include noise attenuation, data compression, and data regularization.

ACKNOWLEDGMENTS

I would like to thank Sergey Gritsenko for teaching me Prony's method more than 20 years ago. I am also grateful to Yihua Cai, Yangkang Chen, Henry Herrera, Guochang Liu, Yang Liu, Karl Schleicher, Mirko van der Baan, and Lexing Ying for inspiring discussions.

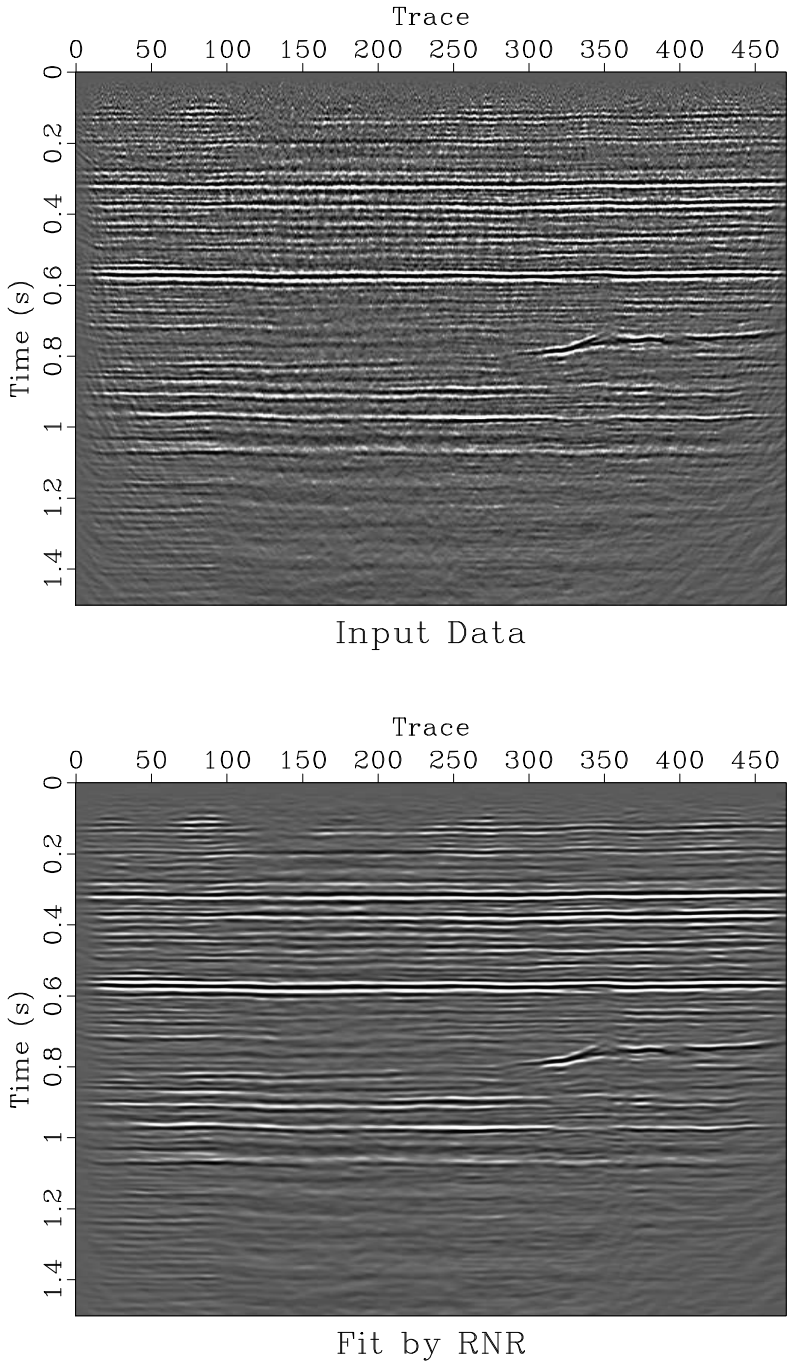


Figure 14: (a) 2D seismic data section. (b) Result of fitting data with two components shown in Figure 16. nar/vecta vdata

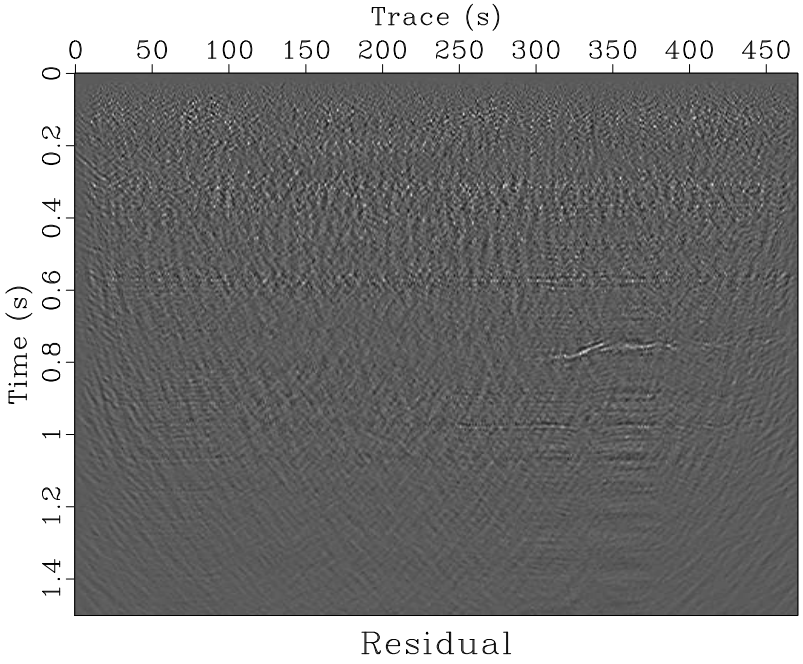


Figure 15: Residual error after fitting seismic data from Figure 14 with two components shown in Figure 16. nar/vecta vdif

REFERENCES

[1] Bath, M., 1995, Modern Spectral Analysis with Geophysical Applications: Soc. of Expl. Geophys. (Edited by Robert E. Wiley).

[2] Battista, B. M., C. Knapp, T. McGee, and V. Goebel, 2007, Application of the empirical mode decomposition and Hilbert-Huang transform to seismic reflection data: Geophysics, **72**, H29–H37.

[3] Bekara, M., and M. van der Baan, 2009, Random and coherent noise attenuation by empirical mode decomposition: Geophysics, **74**, V89–V98.

[4] Beylkin, G., and L. Monzón, 2005, On approximation of functions by exponential sums: Applied and Computational Harmonic Analysis, **19**, 17–48.

[5] Canales, L. L., 1984, Random noise reduction: 54th Ann. Internat. Mtg, Soc. of Expl. Geophys., Session:S10.1.

[6] Castagna, J. P., S. Sun, and R. W. Siegfried, 2003, Instantaneous spectral analysis: Detection of low-frequency shadows associated with hydrocarbons: The Leading Edge, **22**, 120–127.

[7] Daubechies, I., J. Lu, and H.-T. Wu, 2011, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool: Applied and Computational Harmonic Analysis, **30**, 243–261.

[8] Daubechies, I., and S. Maes, 1996, A nonlinear squeezing of the continuous wavelet transform based on auditory nerve models, *in* Wavelets in medicine and biology: CRC Press, 527–546.

[9] Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems:

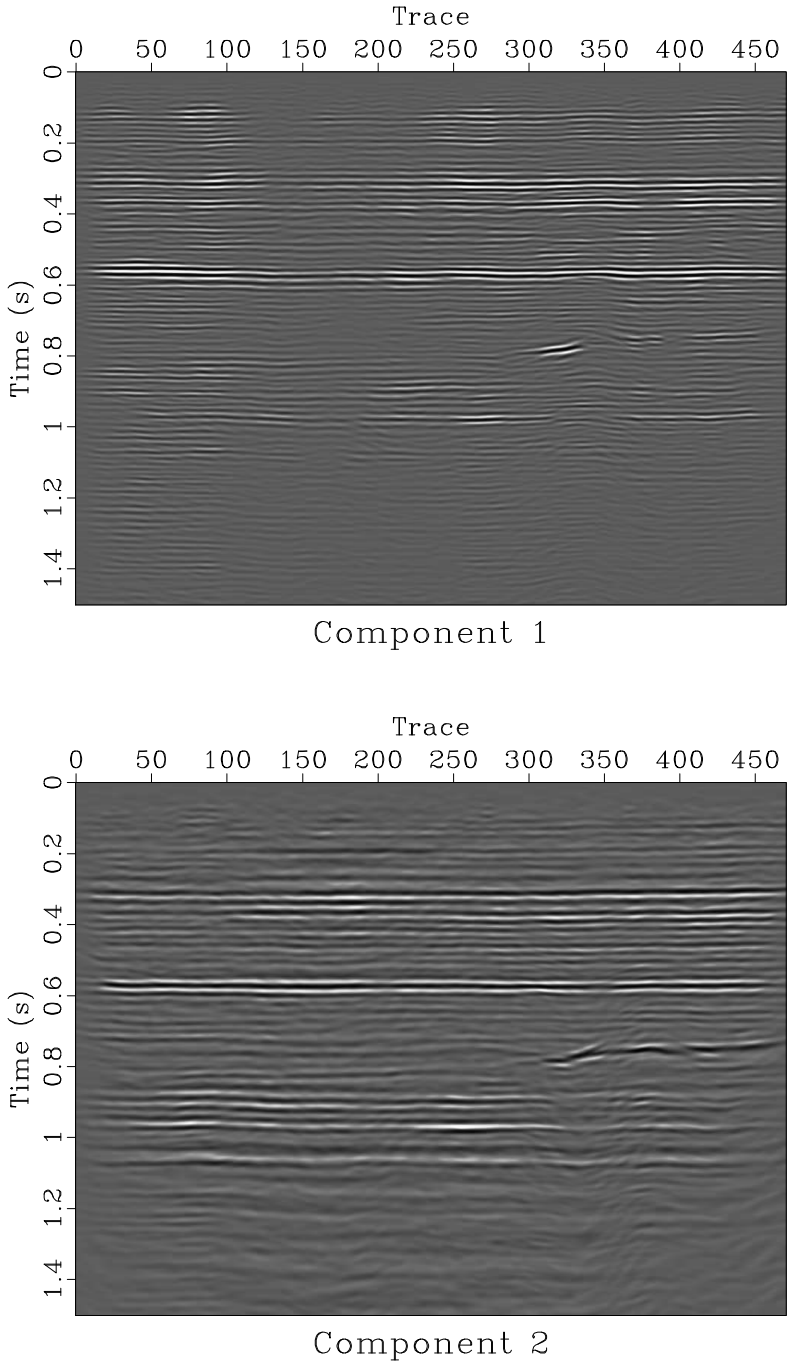


Figure 16: Two nonstationary spectral components: high-frequency (Component 1) and low-frequency (Component 2) estimated from the data shown in Figure 14a. nar/vecta vsign

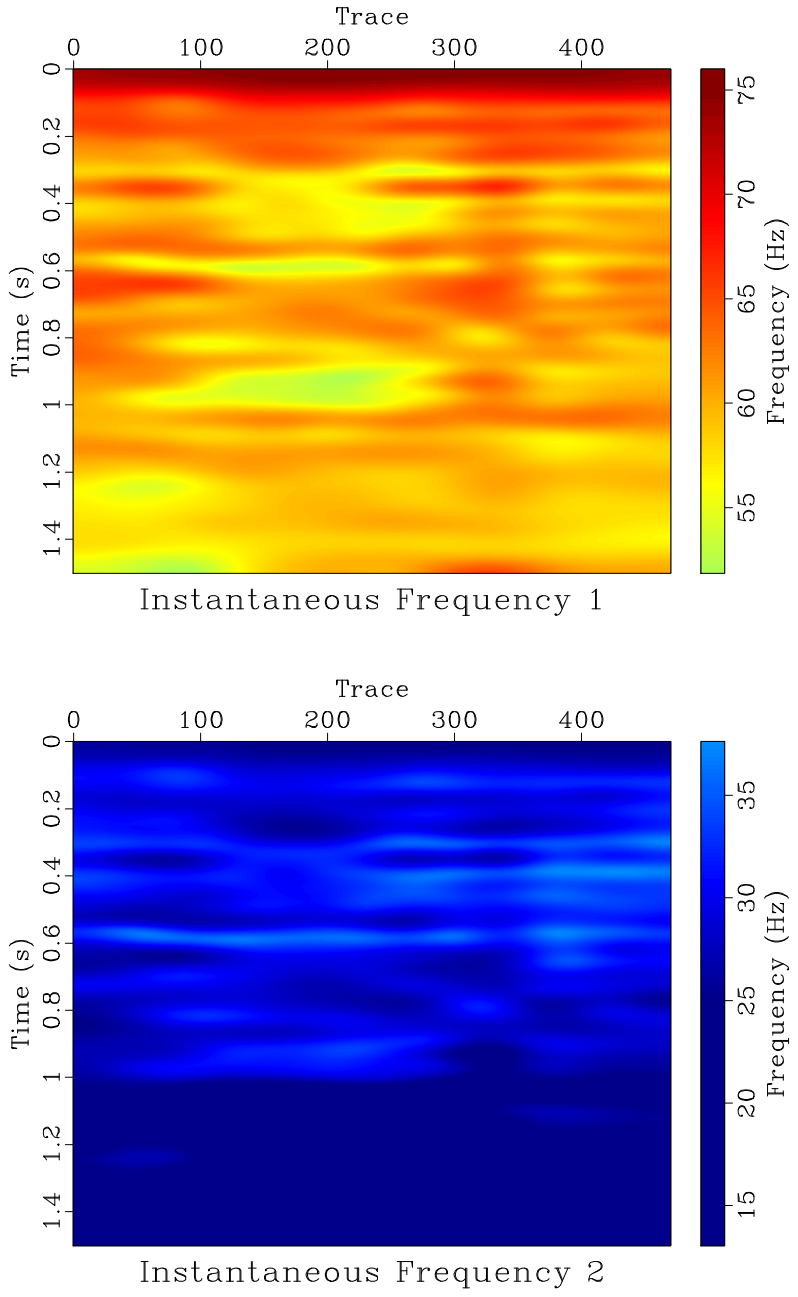


Figure 17: Instantaneous frequencies of high-frequency and low-frequency components from decomposition shown in Figure 16. nar/vecta vgroup

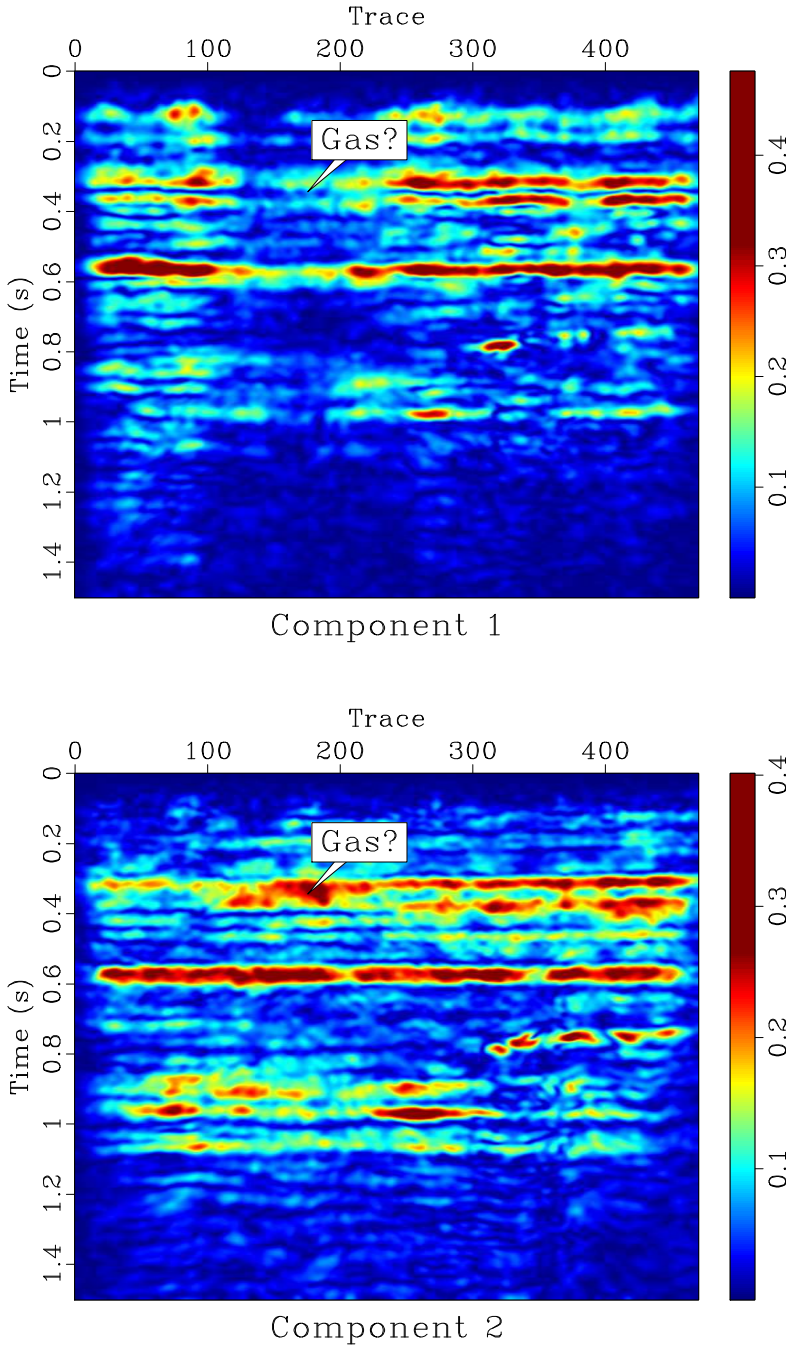


Figure 18: Amplitudes of high-frequency and low-frequency components from decomposition shown in Figure 16. The apparent attenuation of high frequencies in the top left part of the section may indicate presence of gas. nar/vecta vcwht

Kluwer Academic Publishers.
[10] Fomel, S., 2007, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.
[11] ———, 2009, Adaptive multiple subtraction using regularized nonstationary regression: Geophysics, **74**, V25–V33.
[12] Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, V25–V38.
[13] Gardner, G. H. F., and L. Lu, eds., 1991, Slant-stack processing: Society of Exploration Geophysicists. Issue 14 of Geophysics reprint series.
[14] Gritsenko, S. A., S. Fomel, and V. S. Chernyak, 2001, Filtering using Prony's method: Geofizika, 24–26 (in Russian).
[15] Han, J., and M. van der Baan, 2013, Empirical mode decomposition for seismic time-frequency analysis: Geophysics, **78**, O9–O19.
[16] Herrera, R. H., J.-B. Tary, and M. van der Baan, 2013, Time-frequency representation of microseismic signals using the synchrosqueezing transform, *in* Geoconvention: Can. Soc. Expl. Geophys.
[17] Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.
[18] Hou, T. Y., and Z. Shi, 2011, Adaptive data analysis via sparse time-frequency representation: Advances in Adaptive Data Analysis, **3**, 1–28.
[19] ———, 2013, Data-driven time-frequency analysis: Applied and Computational Harmonic Analysis, **35**, 284308.
[20] Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: Proceeding of the Royal Society of London Series A, **454**, 903–995.
[21] Liu, G., and X. Chen, 2013, Noncausal f-x-y regularized nonstationary prediction filtering for random noise attenuation on 3D seismic data: Journal of Applied Geophysics, **93**, 60–66.
[22] Liu, G., X. Chen, J. Du, and K. Wu, 2012, Random noise attenuation using f-x regularized nonstationary autoregression: Geophysics, **77**, V61–V69.
[23] Liu, G., S. Fomel, and X. Chen, 2011, Time-frequency analysis of seismic data using local attributes: Geophysics, **76**, P23–P34.
[24] Liu, Y., and S. Fomel, 2011, Seismic data interpolation beyond aliasing using regularized nonstationary autoregression: Geophysics, **76**, V69–V77.
[25] ———, 2013, Seismic data analysis using local time-frequency decomposition: Geophysical Prospecting, **61**, 516–525.
[26] Magrin-Chagnolleau, I., and R. Baraniuk, 1999, Empirical mode decomposition based frequency attributes: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1949–1952.
[27] Mallat, S., 2009, A wavelet tour of signal processing: The sparse way: Academic Press.
[28] Marple, S. L., 1987, Digital spectral analysis with applications: Prentice-Hall.
[29] Mitrofanov, G., Z. Zhan, and J. Cai, 1998, Using of the Proni transform of Chinese seismic data: 68th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1157–1159.

[30] Mitrofanov, G. M., and V. I. Priimenko, 2011, Prony filtration of seismic data: theoretical background: Revista Brasileira de Geofísica, **29**.
[31] Pisarenko, V. F., 1973, The retrieval of harmonics from a covariance function: Journal of the Royal Astronomical Society, **33**, 347–366.
[32] Prony, R., 1795, Essai expérimental et analytique: Annuaire de l'École Polytechnique, **1**, 24.
[33] Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58.
[34] ———, 2000, Model-based subtraction of multiple events in the frequency-space domain: 70th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1969–1972.
[35] Taner, M. T., F. Koehler, and R. E. Sheriff, 1979, Complex seismic trace analysis: Geophysics, **44**, 1041–1063. (Errata in GEO-44-11-1896; Discussion in GEO-45-12-1877-1878; Reply in GEO-45-12-1878-1878).
[36] Toh, K., and L. Trefethen, 1994, Pseudozeros of polynomials and pseudospectra of companion matrices: Mathematik, **68**, 403–425.
[37] Wu, Z., and N. E. Huang, 2009, Ensemble empirical mode decomposition: a noise-assisted data analysis method: Advances in Adaptive Data Analysis, **1**, 1–41.

Theory of 3-D angle gathers in wave-equation seismic imaging

*Sergey Fomel**

ABSTRACT

I present two methods for constructing angle gathers in 3-D seismic imaging by downward extrapolation. Angles in angle gathers refer to the scattering angle at the reflector and provide a natural access to analyzing migration velocity and amplitudes. In the first method, angle gathers are extracted at each downward-continuation step by mapping transformations in constant-depth frequency slices. In the second method, one extracts angle gathers after applying the imaging condition by transforming local offset gathers in the depth domain. The second approach generalizes previously published algorithms for angle-gather construction in 2-D and common-azimuth imaging.

INTRODUCTION

Wave extrapolation provides an accurate method for seismic imaging in structurally complex areas (9; 13). Wave extrapolation methods have several known advantages in comparison with direct methods such as Kirchhoff migration thanks to their ability to handle multi-pathing, strong velocity heterogeneities, and finite-bandwidth wave-propagation effects (16). However, velocity and amplitude analysis in the prestack domain are not immediately available for wave extrapolation methods. To overcome this limitation, several authors (12; 21; 20; 22; 36; 28; 24; 25; 26) suggested methods for constructing angle gathers from downward-continued wavefields. Angles in angle gathers are generally understood as the reflection (scattering) angles at reflecting interfaces (37; 10). Angle gathers facilitate velocity analysis (18; 33) and can be used in principle for extracting angle-dependent reflectivity information directly at the target reflectors (23). (29) assert that angle gathers generated with wavefield extrapolation are genuinely free of artifacts documented for Kirchhoff-generated angle gathers (30; 31).

There are two possible approaches to angle-gather construction with wavefield continuation. In the first approach, one generates gathers at each depth level converting offset-space-frequency planes into angle-space planes simultaneously with applying the imaging condition. The offset in this case refers to the local offset between source and receiver parts of the downward continued prestack data. Such a construction

***e-mail:** sergey.fomel@beg.utexas.edu

was suggested, for example, by (21). This approach is attractive because of its localization in depth. However, the method of (21) produces gathers in the offset ray parameter as opposed to angle. As a result, the angle-domain information becomes structure-dependent: the output depends not only on the scattering angle but also on the structural dip.

In the second approach, one converts migrated images in offset-depth domain to angle-depth gathers after imaging of all depth levels is completed. (24) suggested a simple Radon-transform procedure for extracting angle gathers from migrated images. The transformation is independent of velocity and structure. (22) adopted it for constructing angle gathers in the shot-gather migration. (7) demonstrate that the method of (24) is strictly valid in the 3-D case only in the absence of cross-line structural dips. They present an extension of this method for the common-azimuth approximation (5).

In this paper, I present a more complete analysis of the angle-gather construction in 3-D imaging by wavefield continuation. First, I show how to remove the structural dependence in the depth-slice approach. The improved mapping retains the velocity dependence but removes the effect of the structure. Additionally, I extend the second, post-migration approach to a complete 3-D wide-azimuth situation. Under the common-azimuth approximation, this formulation reduces to the result of (8) and, in the absence of cross-line structure, it is equivalent to the Radon construction of (24).

TRAVELTIME DERIVATIVES AND DISPERSION RELATIONSHIPS FOR A 3-D DIPPING REFLECTOR

Theoretical analysis of angle gathers in downward continuation methods can be reduced to analyzing the geometry of reflection in the simple case of a dipping reflector in a locally homogeneous medium. Considering the reflection geometry in the case of a plane reflector is sufficient for deriving relationships for local reflection traveltime derivatives in the vicinity of a reflection point (15). Let the local reflection plane be described in $\{x, y, z\}$ coordinates by the general equation

$$x \cos\alpha + y \cos\beta + z \cos\gamma = d \;, \tag{1}$$

where the normal angles α, β, and γ satisfy

$$\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1 \;, \tag{2}$$

The geometry of the reflection ray paths is depicted in Figure 1. The reflection traveltime measured on a horizontal surface above the reflector is given by the known expression (27; 17)

$$t(h_x, h_y) = \frac{2}{v}\sqrt{D^2 + h_x^2 + h_y^2 - (h_x \cos\alpha + h_y \cos\beta)^2} \;, \tag{3}$$

where D is the length of the normal to the reflector from the midpoint (distance MM' in Figure 2)

$$D = d - m_x \cos\alpha - m_y \cos\beta \ , \tag{4}$$

m_x and m_y are the midpoint coordinates, h_x and h_y are the half-offset coordinates, and v is the local propagation velocity.

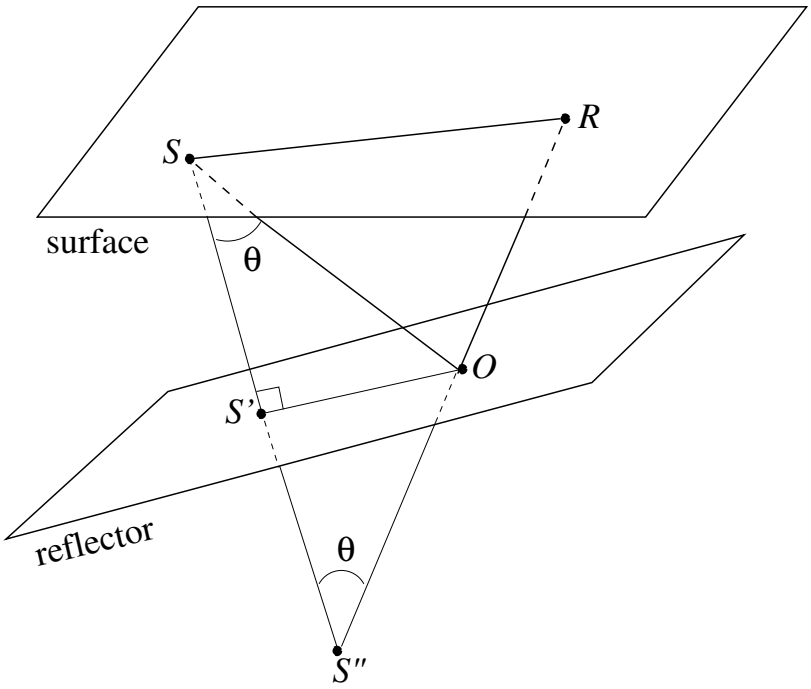


Figure 1: Reflection geometry in 3-D (a scheme). S and R and the source and the receiver positions at the surface. O is the reflection point. S' is the normal projection of the source to the reflector. S'' is the "mirror" source. The cumulative length of the incident and reflected rays is equal to the distance from S'' to R. agath/XFig plane3b

According to elementary geometrical considerations (Figures 1 and 2), the reflection angle θ is related to the previously introduced quantities by the equation

$$\cos\theta = \frac{D}{\sqrt{D^2 + h_x^2 + h_y^2 - (h_x \cos\alpha + h_y \cos\beta)^2}} \ . \tag{5}$$

Explicitly differentiating equation (3) with respect to the midpoint and offset

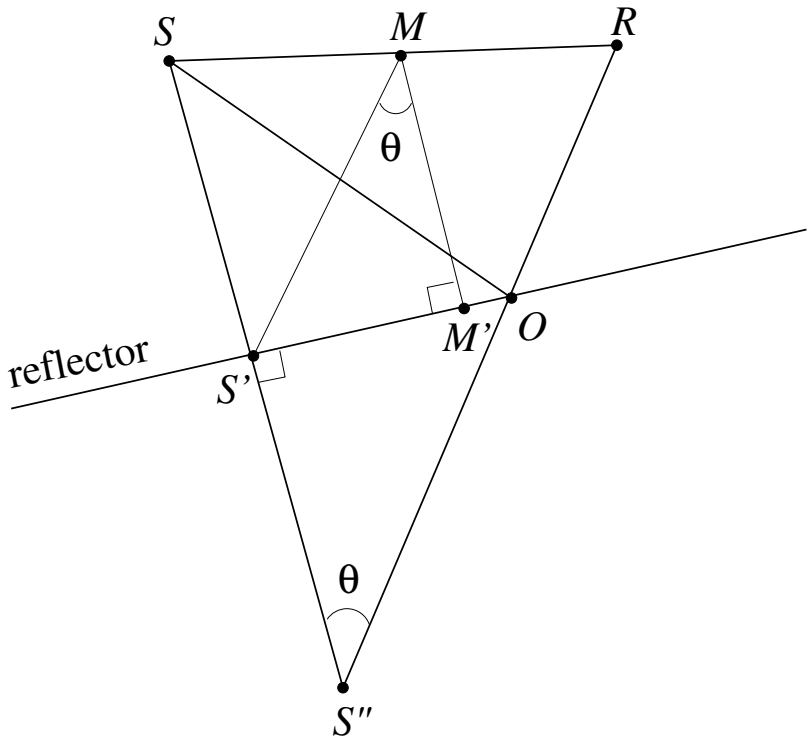


Figure 2: Reflection geometry in the reflection plane (a scheme). M is the midpoint. As follows from the similarity of triangles $S''SR$ and $S'SM$, the distance from M to S' is twice smaller than the distance from S'' to R. agath/XFig plane2b

coordinates and utilizing equation (B-5) leads to the equations

$$t_{m_x} \equiv \frac{\partial t}{\partial m_x} = -\frac{2}{v}\,\cos\theta\,\cos\alpha\;, \tag{6}$$

$$t_{m_y} \equiv \frac{\partial t}{\partial m_y} = -\frac{2}{v}\,\cos\theta\,\cos\beta\;, \tag{7}$$

$$t_{h_x} \equiv \frac{\partial t}{\partial h_x} = \frac{4}{v^2\,t}\,\left(h_x\,\sin^2\alpha - h_y\,\cos\alpha\,\cos\beta\right)\;, \tag{8}$$

$$t_{h_y} \equiv \frac{\partial t}{\partial h_y} = \frac{4}{v^2\,t}\,\left(h_y\,\sin^2\beta - h_x\,\cos\alpha\,\cos\beta\right)\;. \tag{9}$$

Additionally, the traveltime derivative with respect to the depth of the observation surface is given by

$$t_z \equiv \frac{\partial t}{\partial z} = -\frac{2}{v}\,\cos\theta\,\cos\gamma \tag{10}$$

and is related to the previously defined derivatives by the double-square-root equation

$$\begin{aligned} -\,v\,t_z &= \sqrt{1-\frac{v^2}{4}\,\left(t_{m_x}-t_{h_x}\right)^2-\frac{v^2}{4}\,\left(t_{m_y}-t_{h_y}\right)^2} \\ &+ \sqrt{1-\frac{v^2}{4}\,\left(t_{m_x}+t_{h_x}\right)^2-\frac{v^2}{4}\,\left(t_{m_y}+t_{h_y}\right)^2}\;. \end{aligned} \tag{11}$$

In the frequency-wavenumber domain, equation (13) serves as the basis for 3-D shot-geophone downward-continuation imaging. In the Fourier domain, each t_x derivative translates into $-k_x/\omega$ ratio, where k_x is the wavenumber corresponding to x and ω is the temporal frequency.

Equations (6), (7), and (10) immediately produce the first important 3-D relationship for angle gathers

$$\cos\theta = \frac{v}{2\,\omega}\,\sqrt{k_{m_x}^2+k_{m_y}^2+k_z^2}\;. \tag{12}$$

Expressing the depth derivative with the help of the double-square-root equation (13) and applying a number of algebraic transformations, one can turn equation (12) into the dispersion relationship

$$\boxed{\begin{aligned} &\left(k_{m_x}^2+k_{m_y}^2\right)\,\frac{\sin^2\theta}{v^2}+\left(k_{h_x}^2+k_{h_y}^2\right)\,\frac{\cos^2\theta}{v^2} = \\ &\frac{1}{4\,\omega^2}\,\left(k_{m_x}\,k_{h_y}-k_{m_y}\,k_{h_x}\right)^2+4\,\omega^2\,\frac{\cos^2\theta}{v^2}\,\frac{\sin^2\theta}{v^2}\;. \end{aligned}} \tag{13}$$

For each reflection angle θ and each frequency ω, equation (13) specifies the locations on the four-dimensional (k_{m_x}, k_{m_y}, k_{h_x}, k_{h_y}) wavenumber hyperplane that contribute to the common-angle gather. In the 2-D case, equation (13) simplifies by setting k_{h_y}

and k_y to zero. Using the notation $k_{m_x} = k_m$ and $k_{h_x} = k_h$, the 2-D equation takes the form

$$k_m^2 \, \sin^2\theta + k_h^2 \, \cos^2\theta = \frac{4\,\omega^2}{v^2} \, \cos^2\theta \, \sin^2\theta \tag{14}$$

and can be explicitly solved for k_h resulting in the convenient 2-D dispersion relationship

$$k_h = \frac{2\,\omega \, \sin\theta}{v} \sqrt{1 - \frac{4\,k_m^2\,v^2}{\omega^2 \, \cos^2\theta}} \; . \tag{15}$$

In the next section, I show that a similar simplification is also valid under the common-azimuth approximation. Equations (13) and (15) describe an effective migration of the downward-continued data to the appropriate positions on midpoint-offset planes to remove the structural dependence from the local image gathers.

Another important relationship follows from eliminating the local velocity v from equations (13) and (12). Expressing v^2 from equation (12) and substituting the result in equation (13), we arrive (after a number of algebraical transformations) to the frequency-independent equation

$$\boxed{\tan^2\theta = \frac{k_z^2 \left(k_{h_x}^2 + k_{h_y}^2\right) + \left(k_{h_x}\,k_{m_x} + k_{h_y}\,k_{m_y}\right)^2}{k_z^2 \left(k_{m_x}^2 + k_{m_y}^2 + k_z^2\right)} \; .} \tag{16}$$

Equation (16) can be expressed in terms of ratios k_{m_x}/k_z and k_{m_y}/k_z, which correspond at the zero local offset to local structural dips (z_{m_x} and z_{m_y} partial derivatives), and ratios k_{h_x}/k_z and k_{h_y}/k_z, which correspond to local offset slopes. As shown by (25), it can be also expressed as

$$\tan^2\theta = \frac{k_{h_x}^2 + k_{h_y}^2 + k_{h_z}^2}{k_{m_x}^2 + k_{m_y}^2 + k_z^2} \; , \tag{17}$$

where k_{h_z} refers to the vertical offset between source and receiver wavefields (6).

In the 2-D case, equation (16) simplifies to the form, independent of the structural dip:

$$\tan\theta = \frac{k_h}{k_z} \; , \tag{18}$$

which is the equation suggested by (24). Equation (18) appeared previously in the theory of migration-inversion (32).

COMMON-AZIMUTH APPROXIMATION

Common-azimuth migration (5) is a downward continuation imaging method tailored for narrow-azimuth streamer surveys that can be transformed to a single common azimuth with the help of azimuth moveout (4) Employing the common-azimuth approximation, one assumes the reflection plane stays confined in the acquisition azimuth.

Although this assumption is strictly valid only in the case of constant velocity (34), the modest azimuth variation in realistic situations justifies the use of the method (2).

To restrict equations of the previous section to the common-azimuth approximation, it is sufficient to set the cross-line offset h_y to zero assuming the x coordinate is oriented along the acquisition azimuth. In particular, from equations (8-9), we obtain

$$h_x \, \sin\alpha = \frac{v\,t}{2} \, \sin\theta \tag{19}$$

$$t_{h_x} = \frac{4\,h_x}{v^2\,t} \, \sin^2\alpha = \frac{2}{v} \, \sin\theta \, \sin\alpha \; , \tag{20}$$

$$t_{h_y} = -\frac{4\,h_x}{v^2\,t} \, \cos\alpha \, \cos\beta = -\frac{2}{v} \, \sin\theta \, \cot\alpha \, \cos\beta \; . \tag{21}$$

With the help of equations (6), (7), and (10), equation (21) transforms to the form

$$\begin{aligned} t_{h_y} &= t_{m_y} \, \frac{\tan\theta}{\tan\alpha} \\ &= t_{m_y} \, \frac{\sqrt{1-\frac{v^2}{4}\,\left(t_{m_x}+t_{h_x}\right)^2} - \sqrt{1-\frac{v^2}{4}\,\left(t_{m_x}-t_{h_x}\right)^2}}{\sqrt{1-\frac{v^2}{4}\,\left(t_{m_x}+t_{h_x}\right)^2} + \sqrt{1-\frac{v^2}{4}\,\left(t_{m_x}-t_{h_x}\right)^2}} \; , \end{aligned} \tag{22}$$

suggested by (5). Combining equations (6), (7), (10), and (20) and transforming to the frequency-wavenumber domain, we obtain the common-azimuth dispersion relationship

$$\left(k_{h_x}^2 + k_{m_y}^2 + k_z^2\right) \, \left(k_{m_x}^2 + k_{m_y}^2 + k_z^2\right) = \frac{4\,\omega^2}{v^2} \, \left(k_{m_y}^2 + k_z^2\right) \; , \tag{23}$$

which shows that, under the common-azimuth approximation and in a laterally homogeneous medium, 3-D seismic migration amounts to a cascade of a 2-D prestack migrations in the in-line direction and a 2-D zero-offset migration in the cross-line direction (11).

Under the common-azimuth approximation, the angle-dependent relationship (13) takes the form

$$k_{m_x}^2 \, \sin^2\theta + k_{h_x}^2 \, \cos^2\theta = \frac{4\,\omega^2}{v^2} \, \cos^2\theta \, \sin^2\theta \; , \tag{24}$$

which is identical to the 2-D equation (14). This proves that under this approximation, one can perform the structural correction independently for each cross-line wavenumber.

The post-imaging equation (16) transforms to the equation

$$\tan^2\theta = \frac{k_{h_x}^2}{k_{m_y}^2 + k_z^2} \; , \tag{25}$$

obtained previously by (8). In the absence of cross-line structural dips ($k_{m_y} = 0$), it is equivalent to the 2-D equation (18).

ALGORITHM I: ANGLE GATHERS DURING DOWNWARD CONTINUATION

This algorithm follows from equation (13). It consists of the following steps, applied at each propagation depth z:

1. Generate local offset gathers and transform them to the wavenumber domain. In the double-square-root migration, the local offset wavenumbers are immediately available. In the shot gather migration, local offsets are generated by cross-correlation of the source and receiver wavefields (22).

2. For each frequency ω, transform the local offset wavenumbers $\{k_{h_x}, k_{h_y}\}$ into the angle coordinates $\sin\theta/v$ according to equation (13). The angle coordinates depend on velocity but do not depend on the local structural dip. In the 2-D case, each frequency slice is simply the $\{k_m, k_h\}$ plane, and each angle coordinate corresponds to a circle in that plane centered at the origin and described by equation (14). Figure 3 shows an example of a 2-D frequency slice transformed to angles.

3. Accumulate contributions from all frequencies to apply the imaging condition in time.

This algorithm is applicable for targets localized in depth. The local offset gathers need to be computed for all lateral locations, but there is no need to store them in memory, because conversion to angles happens on the fly. The algorithm outputs not angles directly, but velocity-dependent parameters $\sin\theta/v$. (1) extend this algorithm to transversally-isotropic media.

ALGORITHM II: POST-MIGRATION ANGLE GATHERS

The second algorithm follows from equation (16). It applies after the imaging has completed and consists of the following steps applied at each common-image location:

1. Generate and store local offset gathers. In the double-square-root migration, the local offsets are immediately available. In the shot gather migration, local offsets are generated by cross-correlation of the source and receiver wavefields.

2. Estimate the dominant local structural dips at the common image point by using one of the available dip estimation methods: local slant stack, plane-wave destruction, etc.

3. After the imaging has completed, transform local-offset gathers into the slant-stack domain either by slant-stacking in the $\{z, h_x, h_y\}$ physical domain or by radial-trace construction in the $\{k_z, k_{h_x}, k_{h_y}\}$ Fourier domain (24).

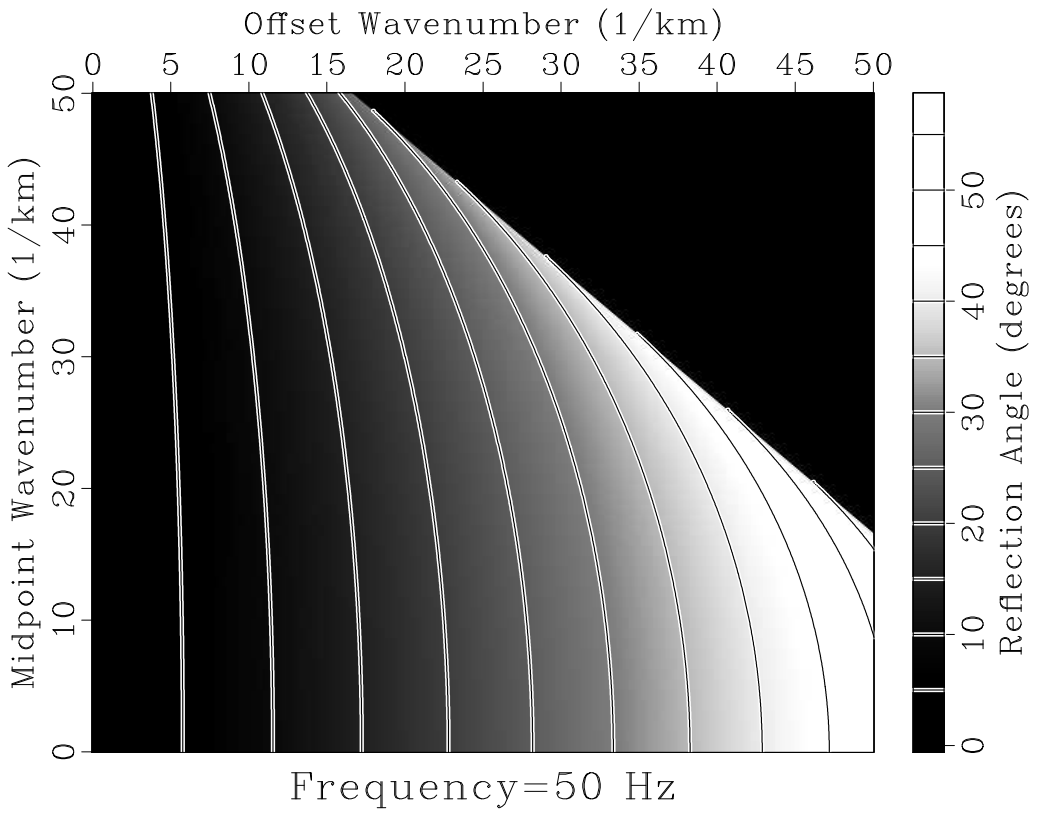


Figure 3: Constant-depth constant-frequency slice mapped to reflection angles according to the 2-D version of Algorithm I. Zero offset wavenumber maps to zero (normal incidence) angle. The top right corner is the evanescent region. agath/zslice angle1

4. Using estimated dips, convert slant stacks into angles by applying equation (16). The mapping from offset-depth slopes to angles is illustrated in Figure 4.

The last two steps can be combined into one. It is sufficient to compute the effective offset $\hat{h} = \sqrt{h_x^2 + h_y^2 + (h_x z_y - h_y z_x)^2}$ and apply the basic 2-D angle extraction algorithm to the effective offset gather.

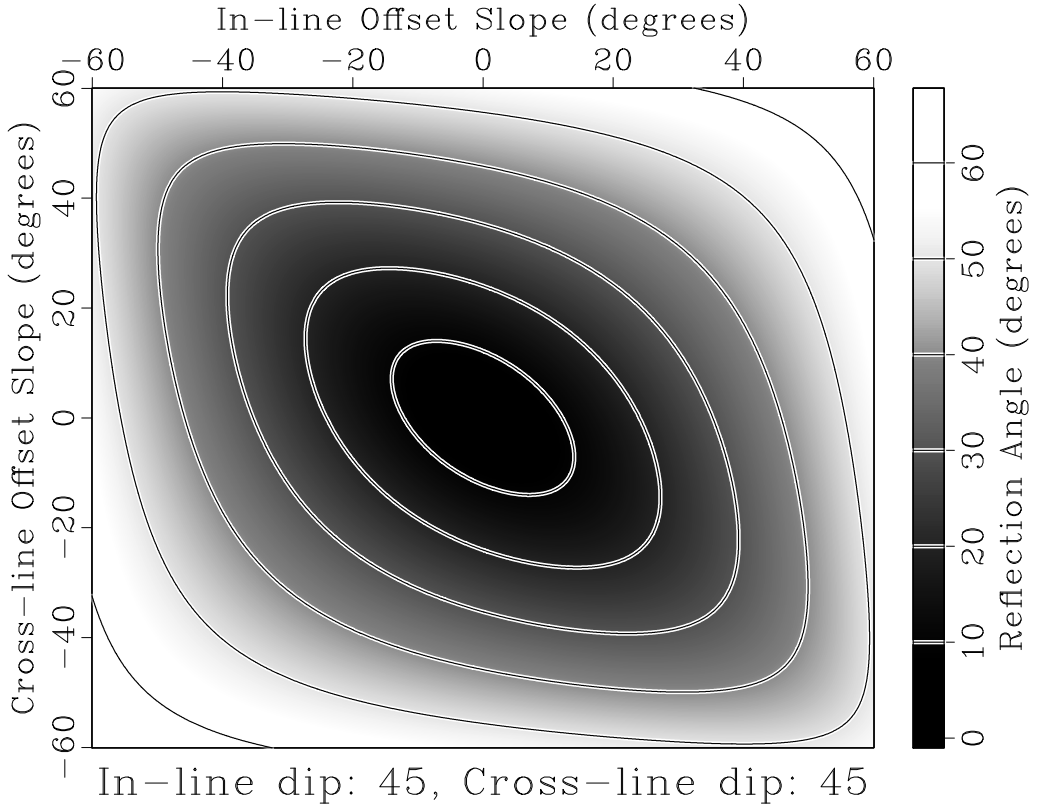


Figure 4: Mapping from the offset slope plane to angles according to Algorithm II. Zero slopes map to zero (normal-incidence) angle. agath/radon angl45-45

The second method is applicable to selected common-image gathers, which can be spread on a sparse grid. The local offset gathers need to be computed and stored at all depths. The method works independent of the velocity. The main disadvantage is the need to estimate local structural dips. In the common-azimuth approximation, only the cross-line dip is required (8). In the 2-D case (zero cross-line dip), the method is dip-independent (24).

DISCUSSION

Since the first presentation of the 3-D angle-gather theory (14), many new research results have appeared in the literature. By the end of 2000s, prestack 3-D reverse-time

migration has become a standard tool for depth imaging in structurally-complex areas, and it is becoming feasible to generate 3-D angle gathers as part of routine processing (19; 35; 38). The most important new theoretical developments are the ability to extract angle information from time-shift angle gathers (26; 35), the ability to extract not only reflection-angle but also azimuth information (38), and the extension of the angle-gather theory to anisotropy (3; 1).

CONCLUSIONS

Angle gathers present a natural tool for analyzing velocities and amplitudes in wave-equation imaging. I have discussed two approaches for angle-gather construction. In the first approach, angle gather are constructed on the fly at different depth steps of the wave extrapolation process. In the second approach, angle gathers are extracted from the local-offset gathers after imaging has completed. The second method was previously presented for the 2-D case and for the case of a common-azimuth approximation. Both approaches have advantages and disadvantages. The preference depends on the application and the input data configuration.

ACKNOWLEDGMENTS

I am grateful to Nanxun Dai, John Etgen, Sergey Goldin, and Paul Sava for enlightening discussions.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Alkhalifah, T., and S. Fomel, 2009, Angle gathers in wave-equation imaging for VTI media, *in* 79th Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 2899–2903.

[2] Biondi, B., 2003, Narrow-azimuth migration of marine streamer data, *in* 73rd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys.

[3] ———, 2007, Angle-domain common-image gathers from anisotropic migration: Geophysics, **72**, 581–591.

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

[5] Biondi, B., and G. Palacharla, 1996, 3-D prestack migration of common-azimuth data: Geophysics, **61**, 1822–1832.

[6] Biondi, B., and G. Shan, 2002, Prestack imaging of overturned reflections by reverse time migration: 72nd Ann. Internat. Mtg, Soc. of Expl. Geophys., 1284–1287.

[7] Biondi, B., and W. Symes, 2004, Angle-domain common-image gathers for migration velocity analysis by wavefield-continuation methods: Geophysics, **69**, 1283–1298.
[8] Biondi, B., T. Tisserant, and W. Symes, 2003, Wavefield-continuation angle-domain common-image gathers for migration velocity analysis, *in* 73rd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys.
[9] Biondi, B. L., 2006, 3-D seismic imaging: Soc. Expl. Geophys.
[10] Brandsberg-Dahl, S., M. V. de Hoop, and B. Ursin, 2003, Focusing in dip and AVA compensation on scattering-angle/azimuth common image gathers: Geophysics, **68**, 232–254.
[11] Canning, A., and G. H. F. Gardner, 1996, A two-pass approximation to 3-D prestack migration: Geophysics, **61**, 409–421.
[12] de Bruin, C. G. M., C. P. A. Wapenaar, and A. J. Berkhout, 1990, Angle-dependent reflectivity by means of prestack migration: Geophysics, **55**, 1223–1234.
[13] Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: Geophysics, **74**, WCA5–WCA17.
[14] Fomel, S., 2004, Theory of 3D angle gathers in wave-equation imaging: 74th Ann. Internat. Mtg., Soc. of Expl. Geophys., 1053–1056.
[15] Goldin, S. V., 2002, Theoretical aspects of 3-D DMO, *in* 72nd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 2333–2336.
[16] Gray, S. H., J. Etgen, J. Dellinger, and D. Whitmore, 2001, Seismic migration problems and solutions: Geophysics, **66**, 1622–1640.
[17] Levin, F. K., 1971, Apparent velocity from dipping interface reflections: Geophysics, **36**, 510–516. (Errata in GEO-50-11-2279).
[18] Liu, W., A. Popovici, D. Bevc, and B. Biondi, 2001, 3D migration velocity analysis for common image gathers in the reflection angle domain: 71st Ann. Internat. Mtg, Soc. of Expl. Geophys., 885–888.
[19] Luo, M., R. Lu, G. Winbow, and L. Bear, 2010, A comparison of methods for obtaining local image gathers in depth migration, *in* 80th Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 3247–3251.
[20] Mosher, C., and D. Foster, 2000, Common angle imaging conditions for prestack depth migration: 70th Ann. Internat. Mtg, Soc. of Expl. Geophys., 830–833.
[21] Prucha, M., B. Biondi, and W. Symes, 1999, Angle-domain common image gathers by wave-equation migration: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 824–827.
[22] Rickett, J. E., and P. C. Sava, 2002, Offset and angle-domain common image-point gathers for shot-profile migration: Geophysics, **67**, 883–889.
[23] Sava, P., B. Biondi, and S. Fomel, 2001, Amplitude-preserved common image gathers by wave-equation migration: 71st Ann. Internat. Mtg, Soc. of Expl. Geophys., 296–299.
[24] Sava, P. C., and S. Fomel, 2003, Angle-domain common-image gathers by wavefield continuation methods: Geophysics, **68**, 1065–1074.
[25] ———, 2005, Coordinate-independent angle-gathers for wave equation migration, *in* 75th Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of

Expl. Geophys., 2052–2055.
[26] ———, 2006, Time-shift imaging condition in seismic migration: Geophysics, **71**, S209–S217.
[27] Slotnick, M. M., 1959, Lessons in Seismic Computing: Soc. of Expl. Geophys. (Edited by R. A. Geyer).
[28] Soubaras, R., 2003, Angle gathers for shot-record migration by local harmonic decomposition: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 889–892.
[29] Stolk, C., and M. V. de Hoop, 2002, Seismic inverse scattering in the "wave-equation" approach, *in* CWP-417: Colorado School of Mines.
[30] Stolk, C. C., and W. W. Symes, 2002, Artifacts in Kirchhoff common image gathers, *in* 72nd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 1129–1132.
[31] ———, 2004, Kinematic artifacts in prestack depth migration: Geophysics, **69**, 562–575.
[32] Stolt, R. H., and A. B. Weglein, 1985, Migration and inversion of seismic data: Geophysics, **50**, 2458–2472.
[33] Stork, C., P. Kitchenside, D. Yingst, U. Albertin, C. Kostov, B. Wilson, D. Watts, J. Kapoor, and G. Brown, 2002, Comparison between angle and offset gathers from wave equation migration and Kirchhoff migration, *in* 72nd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 1200–1203.
[34] Vaillant, L., and B. Biondi, 2000, Accuracy of common-azimuth migration approximations, *in* SEP-103: Stanford Exploration Project, 157–168.
[35] Vyas, M., E. Mobley, D. Nichols, and J. Perdomo, 2010, Angle gathers for RTM using extended imaging conditions, *in* 80th Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 3252–3256.
[36] Xie, X. B., and R. S. Wu, 2002, Extracting angle domain information from migrated wavefield, *in* 72nd Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 1360–1363.
[37] Xu, S., H. Chauris, G. Lambaré, and M. Noble, 2001, Common-angle migration: A strategy for imaging complex media: Geophysics, **66**, 1877–1894.
[38] Xu, S., Y. Zhang, and B. Tang, 2010, 3D common image gathers from reverse time migration, *in* 80th Ann. Internat. Mtg, Soc. Expl. Geophys., Expanded Abstracts: Soc. of Expl. Geophys., 3257–3262.

A fast butterfly algorithm for generalized Radon transforms

Jingwei Hu[*], *Sergey Fomel*[†], *Laurent Demanet*[‡], *and Lexing Ying*[§*]

ABSTRACT

Generalized Radon transforms such as the hyperbolic Radon transform cannot be implemented as efficiently in the frequency domain as convolutions, thus limiting their use in seismic data processing. We introduce a fast butterfly algorithm for the hyperbolic Radon transform. The basic idea is to reformulate the transform as an oscillatory integral operator and to construct a blockwise low-rank approximation of the kernel function. The overall structure follows the Fourier integral operator (FIO) butterfly algorithm. For two-dimensional data, the algorithm runs in complexity $O(N^2 \log N)$, where N depends on the maximum frequency and offset in the dataset and the range of parameters (intercept time and slowness) in the model space. Using a series of examples, we show that the proposed algorithm can be significantly more efficient than the conventional time-domain integration.

INTRODUCTION

In seismic data processing, the Radon transform (RT) (14), or slant stack, is a set of line integrals that map mixed and overlapping events in seismic gathers to a new transformed domain where they can be separated (6). The integrals can also be taken along curves: parabolas (parabolic RT), or hyperbolas (hyperbolic RT or velocity stack) are most commonly used. A major difference between these transforms is that the former two are time-invariant (i.e., involve a convolution in time) whereas the latter is time-variant. When the curves are time-invariant, the transform can be performed efficiently in the frequency domain using the convolution theorem. However, this approach does not work for time-variant transforms. As a result, the hyperbolic Radon transform is usually thought of as requiring a computation in the time domain, which is computationally expensive due to the large size of seismic data. Nevertheless, the hyperbolic transform is often preferred as it better matches the true seismic events in common midpoint (CMP) gathers (16).

In this work, we construct a fast butterfly algorithm to effectively evaluate time-variant transforms such as the hyperbolic Radon transform. As opposed to the conventional, relatively costly *velocity scan* (i.e., direct integration + interpolation in the

[*]**e-mail:** sergey.fomel@beg.utexas.edu

time domain), our method provides an accurate approximation in only $O(N^2 \log N)$ (all the log in this paper refer to logarithm to base 2) operations for 2D data. Here N depends on the maximum frequency and offset in the dataset and the range of parameters (intercept time and slowness) in the model space, and can often be chosen small compared to the grid size. The adjoint of the transform can be evaluated similarly without extra difficulty. Note that the algorithm introduced in this paper only deals with the fast implementation of a single integral operator (forward Radon transform or its adjoint), not an iteration process for its inversion which is the main objective of many previous works on fast Radon transforms (15; 18; 10; 19).

Radon transforms have been widely used to separate and attenuate multiple reflections (7; 20; 5; 9; 12; 8; 17). As having fast implementations of both forward and adjoint transforms is an essential component of least-squares minimization, our hope is that the current fast algorithm will help to make the hyperbolic Radon transform an accessible tool for improving the inversion process.

The term "generalized Radon transform" connotes a broader context where integrals are taken along arbitrary parametrized sets of smooth curves. The term was introduced by (1; 2), who showed that an asymptotically-correct inverse follows from an amplitude correction to the adjoint. Kirchhoff migration and its (regularized) inverse can be expressed as generalized Radon transforms. The algorithm presented in this paper can in principle be applied in the context of Kirchhoff migration, although we do not attempt to do so here.

The rest of the paper is organized as follows. We first introduce the low-rank approximations and the butterfly structure of the hyperbolic Radon operator; then use these building elements to construct our fast algorithm. A brief description of the algorithm is given in the main text, and a complete derivation can be found in the appendix. We present numerical examples using both synthetic and field data to illustrate the accuracy and efficiency of the proposed algorithm.

ALGORITHM

Assume $d(t,h)$ is a function in the data space. The hyperbolic Radon transform R maps d to a function $(Rd)(\tau,p)$ in the model space (16) through

$$(Rd)(\tau,p) = \int d(\sqrt{\tau^2+p^2h^2},h)\,dh, \tag{1}$$

where t is time, h is offset, τ is intercept time, and p is slowness. Fixing (τ,p), hyperbola $t=\sqrt{\tau^2+p^2h^2}$ describes the traveltime for the event; hence integration along these curves can be used to identify different reflections.

Instead of approximating the integral in equation 1 directly, we reformulate it as a double integral,

$$(Rd)(\tau,p) = \iint \hat{d}(f,h)e^{2\pi i f\sqrt{\tau^2+p^2h^2}}\,df\,dh. \tag{2}$$

Here f is the frequency, $\hat{d}(f,h)$ is the Fourier transform of $d(t,h)$ in t. A simple discretization of equation 2 yields

$$(Rd)(\tau,p) = \sum_{f,h} e^{2\pi i f\sqrt{\tau^2+p^2h^2}}\hat{d}(f,h) \tag{3}$$

(the area element is omitted; the same symbols f, h, τ, and p are used for both continuous and discrete variables). The reason that hyperbolic RT is harder to compute than linear RT ($t = \tau + ph$) or parabolic RT ($t = \tau + ph^2$) should be clear from equation 3: product $f\tau$ in the phase cannot be decoupled from other terms.

To construct the fast algorithm, we first perform a linear transformation to map all discrete points in (f,h) and (τ,p) domains to points in the unit square $[0,1]\times[0,1]$ ($[a,b]\times[c,d]$ represents a 2D rectangular domain in the xy-plane with $x\in[a,b]$ and $y\in[c,d]$), i.e., a point $(f,h)\in[f_{\min},f_{\max}]\times[h_{\min},h_{\max}]$ is mapped to $\mathbf{k}=(k_1,k_2)\in[0,1]\times[0,1]=K$ via

$$f = (f_{\max}-f_{\min})k_1 + f_{\min}, \quad h = (h_{\max}-h_{\min})k_2 + h_{\min}; \tag{4}$$

a point $(\tau,p)\in[\tau_{\min},\tau_{\max}]\times[p_{\min},p_{\max}]$ is mapped to $\mathbf{x}=(x_1,x_2)\in[0,1]\times[0,1]=X$ via

$$\tau = (\tau_{\max}-\tau_{\min})x_1 + \tau_{\min}, \quad p = (p_{\max}-p_{\min})x_2 + p_{\min}. \tag{5}$$

If we define input $g(\mathbf{k}) = \hat{d}(f(k_1),h(k_2))$, output $u(\mathbf{x}) = (Rd)(\tau(x_1),p(x_2))$, and phase function $\Phi(\mathbf{x},\mathbf{k}) = f(k_1)\sqrt{\tau(x_1)^2+p(x_2)^2h(k_2)^2}$, then equation 3 can be written as

$$u(\mathbf{x}) = \sum_{\mathbf{k}\in K} e^{2\pi i\Phi(\mathbf{x},\mathbf{k})}g(\mathbf{k}), \quad \mathbf{x}\in X \tag{6}$$

(throughout the paper, K and X will either be used for sets of discrete points or square domains containing them; the meaning should be clear from the context). This form is the discrete version of an oscillatory integral of the type

$$u(\mathbf{x}) = \int_K e^{2\pi i\Phi(\mathbf{x},\mathbf{k})}g(\mathbf{k})\,d\mathbf{k}, \quad \mathbf{x}\in X, \tag{7}$$

whose fast evaluation has been considered in (3). Our method for computing the summation in equation 6 follows the FIO butterfly algorithm introduced there.

Low-rank approximations

Clearly the range and possibly other factors such as gradient of phase $\Phi(\mathbf{x},\mathbf{k})$ determine the degree of oscillations in the kernel $e^{2\pi i\Phi(\mathbf{x},\mathbf{k})}$. Let N be an integer power of two, which is on the order of the maximum value of $|\Phi(\mathbf{x},\mathbf{k})|$ for $\mathbf{x}\in X$ and $\mathbf{k}\in K$ (the exact choice of N depends on the desired efficiency and accuracy of the algorithm, which will be made specific in numerical examples). The design of the fast algorithm relies on the key observation that the kernel $e^{2\pi i\Phi(\mathbf{x},\mathbf{k})}$, when properly restricted to

subsets of X and K, admits accurate and low-rank separated approximations. More precisely, if A and B are two square boxes in X and K, with sidelengths $w(A)$, $w(B)$ obeying $w(A)w(B) \leq 1/N$ — in which case the pair (A, B) is called admissible — then

$$\left| e^{2\pi i \Phi(\mathbf{x},\mathbf{k})} - \sum_{t=1}^{r_\epsilon} \alpha_t^{AB}(\mathbf{x}) \beta_t^{AB}(\mathbf{k}) \right| \leq \epsilon, \quad \text{for } \mathbf{x} \in A, \quad \mathbf{k} \in B, \tag{8}$$

where r_ϵ is independent of N for a fixed error ϵ. Here and below the subscript t is slightly abused: t should be understood as multi-indices (t_1, t_2), and accordingly r_ϵ is the total number of terms in a double sum. Furthermore, (3) showed that this low-rank approximation can be constructed via a tensor-product Chebyshev interpolation of $e^{2\pi i \Phi(\mathbf{x},\mathbf{k})}$ in the $\mathbf{x}$ variable when $w(A) \leq 1/\sqrt{N}$, and in the $\mathbf{k}$ variable when $w(B) \leq 1/\sqrt{N}$.

Specifically, when $w(B) \leq 1/\sqrt{N}$, α_t^{AB} and β_t^{AB} are given by

$$\alpha_t^{AB}(\mathbf{x}) = e^{2\pi i \Phi(\mathbf{x},\mathbf{k}_t^B)}, \tag{9}$$

$$\beta_t^{AB}(\mathbf{k}) = e^{-2\pi i \Phi(\mathbf{x_0}(A),\mathbf{k}_t^B)} L_t^B(\mathbf{k}) e^{2\pi i \Phi(\mathbf{x_0}(A),\mathbf{k})}; \tag{10}$$

and when $w(A) \leq 1/\sqrt{N}$, α_t^{AB} and β_t^{AB} are given by

$$\alpha_t^{AB}(\mathbf{x}) = e^{2\pi i \Phi(\mathbf{x},\mathbf{k}_0(B))} L_t^A(\mathbf{x}) e^{-2\pi i \Phi(\mathbf{x}_t^A,\mathbf{k}_0(B))}, \tag{11}$$

$$\beta_t^{AB}(\mathbf{k}) = e^{2\pi i \Phi(\mathbf{x}_t^A,\mathbf{k})}. \tag{12}$$

Boldface letters $\mathbf{k}_t^B$, $\mathbf{x}_t^A, \mathbf{k}_0(B), \mathbf{x}_0(A)$ denote 2D vectors. $\mathbf{k}_t^B$ is a point on the 2D, $q_{k_1} \times q_{k_2}$ Chebyshev grid in box B centered at $\mathbf{k}_0(B)$, i.e., let $\mathbf{k}_t^B = (k_{t_1}^B, k_{t_2}^B)$, $\mathbf{k}_0(B) = (k_{0_1}^B, k_{0_2}^B)$, then

$$k_{t_1}^B = k_{0_1}^B + w(B) z_{t_1}, \quad 0 \leq t_1 \leq q_{k_1} - 1, \tag{13}$$

$$k_{t_2}^B = k_{0_2}^B + w(B) z_{t_2}, \quad 0 \leq t_2 \leq q_{k_2} - 1, \tag{14}$$

where

$$\left\{ z_{t_i} = \frac{1}{2} \cos\left(\frac{\pi t_i}{q_{k_i} - 1} \right) \right\}_{0 \leq t_i \leq q_{k_i} - 1,\ i=1,2} \tag{15}$$

is the 1D Chebyshev grid of order q_{k_i} on $[-1/2, 1/2]$. See Figure 1 for an illustration. $L_t^B(\mathbf{k})$ is the 2D Lagrange interpolation defined on the Chebyshev grid:

$$L_t^B(\mathbf{k}) = \left(\prod_{s_1=0, s_1 \neq t_1}^{q_{k_1}-1} \frac{k_1 - k_{s_1}^B}{k_{t_1}^B - k_{s_1}^B} \right) \left(\prod_{s_2=0, s_2 \neq t_2}^{q_{k_2}-1} \frac{k_2 - k_{s_2}^B}{k_{t_2}^B - k_{s_2}^B} \right). \tag{16}$$

Analogously, $\mathbf{x}_t^A$ is a point on the 2D, $q_{x_1} \times q_{x_2}$ Chebyshev grid in box A centered at $\mathbf{x}_0(A)$, and $L_t^A(\mathbf{x})$ is the 2D Lagrange interpolation defined on this grid. Based on the discussion above, the number r_ϵ in low-rank approximation 8 is equal to $q_{k_1} q_{k_2}$ when $w(B) \leq 1/\sqrt{N}$, and $q_{x_1} q_{x_2}$ when $w(A) \leq 1/\sqrt{N}$.

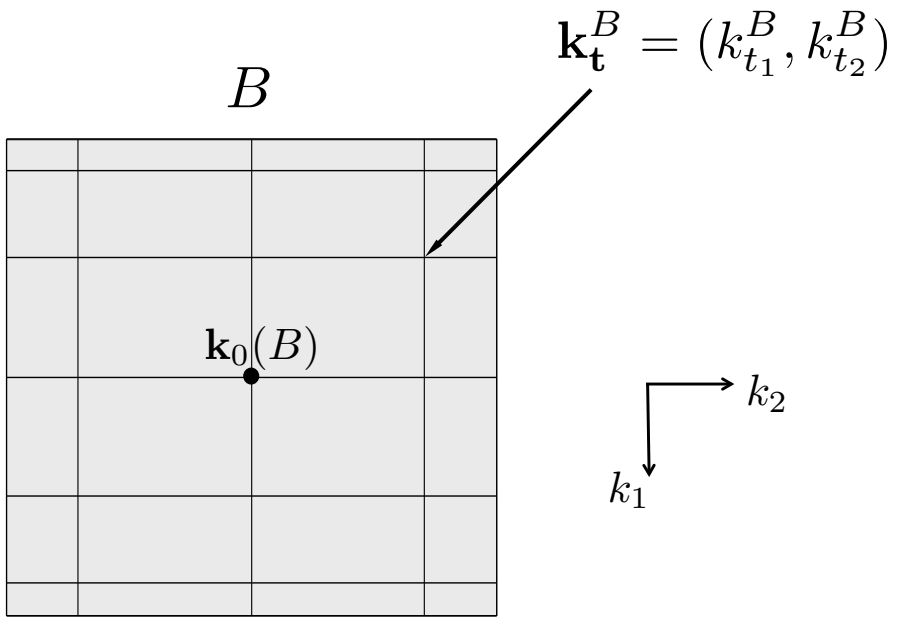


Figure 1: A 2D, $q_{k_1} \times q_{k_2}$ ($q_{k_1} = 7$, $q_{k_2} = 5$) Chebyshev grid in box B. $\mathbf{k}_0(B)$ is the center of the box. $\mathbf{k}_t^B = (k_{t_1}^B, k_{t_2}^B),\ 0 \le t_1 \le q_{k_1} - 1,\ 0 \le t_2 \le q_{k_2} - 1$ is a point on the grid. butterfly/. grid

A simple way of viewing expressions 9 – 12 is: when $w(B) \le 1/\sqrt{N}$, plugging expression 9 into approximation 8 (leaving $\beta_t^{AB}(\mathbf{k})$ as it is) yields

$$e^{2\pi i \Phi(\mathbf{x},\mathbf{k})} \approx \sum_t e^{2\pi i \Phi(\mathbf{x},\mathbf{k}_t^B)} \beta_t^{AB}(\mathbf{k}), \quad \text{for } \mathbf{x} \in A, \ \ \mathbf{k} \in B. \tag{17}$$

For fixed $\mathbf{x}$, the right hand side of equation 17 is just a special interpolation of function $e^{2\pi i \Phi(\mathbf{x},\mathbf{k})}$ in variable $\mathbf{k}$, where $\mathbf{k}_t^B$ are the interpolation points, $\beta_t^{AB}(\mathbf{k})$ are the basis functions. Likewise, when $w(A) \le 1/\sqrt{N}$, plugging expression 12 into approximation 8, we get

$$e^{2\pi i \Phi(\mathbf{x},\mathbf{k})} \approx \sum_t e^{2\pi i \Phi(\mathbf{x}_t^A,\mathbf{k})} \alpha_t^{AB}(\mathbf{x}), \quad \text{for } \mathbf{x} \in A, \ \ \mathbf{k} \in B. \tag{18}$$

For fixed $\mathbf{k}$, the right hand side of equation 18 is a special interpolation of $e^{2\pi i \Phi(\mathbf{x},\mathbf{k})}$ in variable $\mathbf{x}$: $\mathbf{x}_t^A$ are the interpolation points, $\alpha_t^{AB}(\mathbf{x})$ are the basis functions.

Once the low-rank approximation 8 is known, computing the partial sum

$$u^B(\mathbf{x}) := \sum_{\mathbf{k} \in B} e^{2\pi i \Phi(\mathbf{x},\mathbf{k})} g(\mathbf{k}) \tag{19}$$

generated by points $\mathbf{k}$ inside a box B becomes

$$u^B(\mathbf{x}) \approx \sum_{\mathbf{k} \in B} \sum_t \alpha_t^{AB}(\mathbf{x}) \beta_t^{AB}(\mathbf{k}) g(\mathbf{k}) = \sum_t \alpha_t^{AB}(\mathbf{x}) \delta_t^{AB}, \quad \text{for } \mathbf{x} \in A, \tag{20}$$

where

$$\delta_t^{AB} := \sum_{\mathbf{k}\in B} \beta_t^{AB}(\mathbf{k}) g(\mathbf{k}). \tag{21}$$

The case that the box B represents the whole domain K is of particular interest, since it corresponds to the original problem. Therefore, if we can find the set of interaction coefficients δ_t^{AB} relative to all admissible couples of boxes (A, B) with $B = K$, our problem will be solved.

Butterfly structure

The coefficients δ_t^{AB} for $B = K$ are however not readily available. The so-called butterfly algorithm turns out to be an appropriate tool. The butterfly algorithm was introduced by (11), and generalized by (13) and (3). Different applications include sparse Fourier transform (21), and radar imaging (4). (4) also provide a complete error analysis of the method introduced by (3).

The idea of the butterfly algorithm is to obtain δ_t^{AB} for $B = K$ at the last step of a hierarchical construction of *all* the coefficients δ_t^{AB} for *all* pairs of admissible boxes (A, B) belonging to a quad tree structure. The algorithm starts with very small boxes B, where δ_t^{AB} are easily computed by direct summation, and gradually increases the sizes of the boxes B in a multiscale fashion. In tandem, the sizes of the boxes A where u^B is evaluated must decrease to respect the admissibility of each couple (A, B). The computation then mostly consists in updating coefficients δ_t^{AB} from one scale to the next — from finer to coarser B boxes, and from coarser to finer A boxes.

The main data structure underlying the algorithm is a pair of quad trees T_X and T_K. The tree T_X has $[0, 1] \times [0, 1]$ as its root box (level 0) and is built by recursive, dyadic partitioning until level $L = \log N$, where the finest boxes are of sidelength $1/N$. The tree T_K is built similarly but in the opposite direction. Figure 2 shows such a partition for $N = 4$. A crucial property of this structure is that at arbitrary level l, the sidelengths of a box A in T_X and a box B in T_K always satisfy

$$w(A)w(B) = \frac{1}{2^l}\frac{1}{2^{L-l}} = \frac{1}{N}; \tag{22}$$

thus a low-rank approximation of the kernel $e^{2\pi i \Phi(\mathbf{x},\mathbf{k})}$ is available at every level of the tree, for every couple of admissible boxes (A, B).

Fast butterfly algorithm

With the previously introduced low-rank approximations and the butterfly structure, we are ready to describe the fast algorithm. Our goal is to approximate δ_t^{AB}, definition 21, so as to get $u^B(\mathbf{x})$, definition 19, by traversing the tree structure (Figure 2) from top to bottom on the X side, and from bottom to top on the K side. This can be

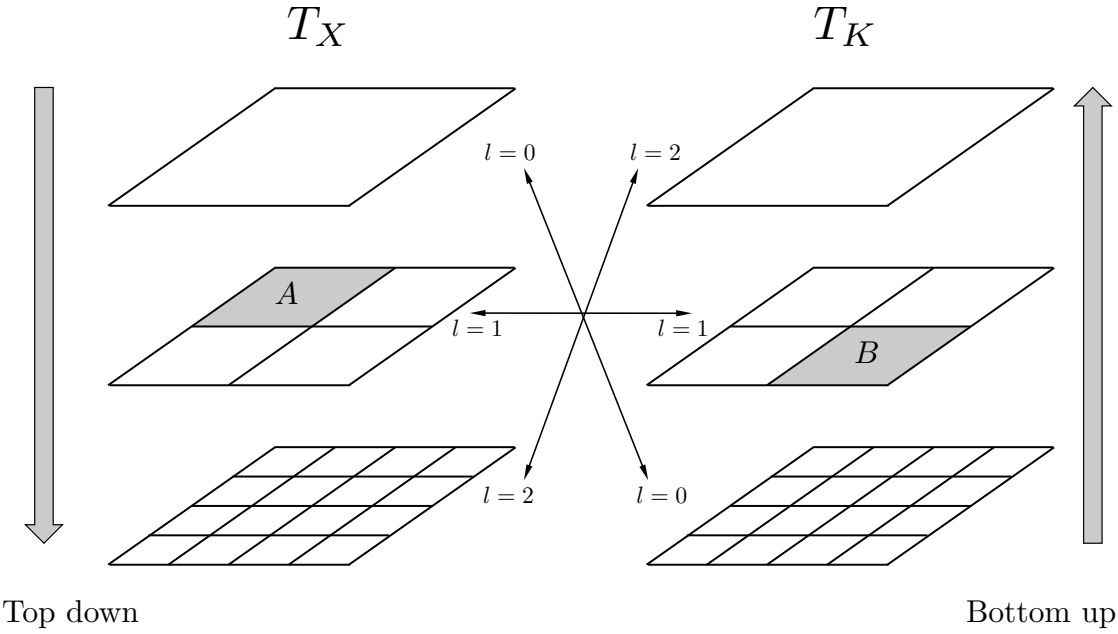


Figure 2: The butterfly quad tree structure for the special case of $N = 4$. butterfly/. tree

done in five major steps. To avoid too much technical detail, we deliberately defer the complete derivation of the algorithm until the appendix, and only summarize here the final updating formulas for each step.

1. *Initialization.* At level $l = 0$, let A be the root box of T_X. For each leaf box $B \in T_K$, construct the coefficients $\{\delta_t^{AB}\}$ by

$$\delta_t^{AB} = e^{-2\pi i \Phi(\mathbf{x}_0(A), \mathbf{k}_t^B)} \sum_{\mathbf{k} \in B} \left(L_t^B(\mathbf{k}) e^{2\pi i \Phi(\mathbf{x}_0(A), \mathbf{k})} g(\mathbf{k}) \right). \tag{23}$$

2. *Recursion.* At $l = 1, 2, ..., L/2$, for each pair (A, B), let A_p be A's parent and $\{B_c, c = 1, 2, 3, 4\}$ be B's children from the previous level. Update $\{\delta_t^{AB}\}$ from $\{\delta_{t'}^{A_p B_c}\}$ by

$$\delta_t^{AB} = e^{-2\pi i \Phi(\mathbf{x}_0(A), \mathbf{k}_t^B)} \sum_c \sum_{t'} \left(L_t^B(\mathbf{k}_{t'}^{B_c}) e^{2\pi i \Phi(\mathbf{x}_0(A), \mathbf{k}_{t'}^{B_c})} \delta_{t'}^{A_p B_c} \right). \tag{24}$$

3. *Switch.* At middle level $l = L/2$, for each (A, B) compute the new set of coefficients $\{\delta_t^{AB}\}$ from the old set $\{\delta_s^{AB}\}$ by

$$\delta_t^{AB} = \sum_s e^{2\pi i \Phi(\mathbf{x}_t^A, \mathbf{k}_s^B)} \delta_s^{AB}. \tag{25}$$

4. *Recursion.* At $l = L/2+1, ..., L$, for each pair (A, B), update $\{\delta_t^{AB}\}$ from $\{\delta_{t'}^{A_pB_c}\}$ of the previous level by

$$\delta_t^{AB} = \sum_c e^{2\pi i \Phi(\mathbf{x}_t^A, \mathbf{k}_0(B_c))} \sum_{t'} \left(L_{t'}^{A_p}(\mathbf{x}_t^A) e^{-2\pi i \Phi(\mathbf{x}_{t'}^{A_p}, \mathbf{k}_0(B_c))} \delta_{t'}^{A_pB_c} \right). \tag{26}$$

5. *Termination.* Finally at level $l = L$, B is the entire domain K. For every box A in X and every $\mathbf{x} \in A$, compute $u(\mathbf{x})$ by

$$u(\mathbf{x}) = e^{2\pi i \Phi(\mathbf{x}, \mathbf{k}_0(B))} \sum_t \left(L_t^A(\mathbf{x}) e^{-2\pi i \Phi(\mathbf{x}_t^A, \mathbf{k}_0(B))} \delta_t^{AB} \right). \tag{27}$$

Numerical complexity and accuracy

To analyze the algorithm's numerical complexity, let us assume the numbers of Chebyshev points in every box and every dimension of K and X are all equal to a small constant q, i.e., $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = q$ and $r_\epsilon \equiv q^2$. The main workload of the fast butterfly algorithm is in Steps 2 and 4. For each level, there are N^2 pairs of boxes (A, B), and the operations between each A and B is $O(r_\epsilon^2)$, which can be further reduced to $O(r_\epsilon^{3/2})$ by performing Chebyshev interpolation one dimension at a time. Since there are $\log N$ levels, the total cost is $O(r_\epsilon^{3/2} N^2 \log N)$. It is not difficult to see that Step 3 takes $O(r_\epsilon^2 N^2)$, and Steps 1 and 5 take $O(r_\epsilon N_f N_h)$ and $O(r_\epsilon N_\tau N_p)$ operations. Considering the initial Fourier transform of preparing data in the (f, h) domain, we conclude that the overall complexity of the algorithm is $O(N_h N_t \log N_t + r_\epsilon^{3/2} N^2 \log N + r_\epsilon^2 N^2 + r_\epsilon(N_f N_h + N_\tau N_p))$. The analysis in (3) shows that the relation between r_ϵ and error ϵ is $r_\epsilon = O(\log^4(1/\epsilon))$. We would like to mention that this is only the worst case estimate. Numerical results in the same paper demonstrate that the dependence of r_ϵ on $\log(1/\epsilon)$ is rather moderate in practice.

In comparison, the conventional *velocity scan* requires at least $O(N_\tau N_p N_h)$ computations, which quickly becomes a burden as the problem size increases. Yet the efficiency of our algorithm is mainly controlled by $O(N^2 \log N)$ with a constant polylogarithmic in ϵ, where N depends neither on data size nor on data content (here we mean the data after the Fourier transform). Since the Chebyshev interpolation is only performed on the kernel, our choice of parameters (N and number of Chebyshev points) relies on the preknowledge about the range of f, h, τ, and p. In other words, we need a general idea about how oscillate the kernel is. Recall that everything is mapped to a unit square, so the larger the range of $\Phi(\mathbf{x}, \mathbf{k})$ is, the more oscillations occur in the unit square. If the original data (data before the Fourier transform) contain high frequency information, the accuracy will be affected as the frequency bandwidth is now larger. A possible way to get around it is to divide the Fourier domain into two or three smaller subdomains (so the range of f in each subdomain is smaller than the original problem), and apply the fast algorithm to each part separately, finally add the results back together. This only increases the cost by a small factor, but presumably offers better accuracy.

NUMERICAL EXAMPLES

In this section we provide several numerical examples to illustrate the empirical properties of the fast butterfly algorithm. To check the results *qualitatively*, we compare with the *velocity scan* method (the nearest neighbor interpolation is used to minimize the interpolation cost); to test the results *quantitatively*, however, it makes more sense to compare with the direct evaluation of equation 3, since the fast algorithm is to speed up this summation in the frequency domain, whereas the *velocity scan* computes a slightly different sum in the time domain, which may contain interpolation artifacts.

There is no general rule for selecting parameters N, q_{k_1}, q_{k_2}, ... The larger N is, the fewer Chebyshev points are needed, and vice versa. In practice, parameters can be tuned to achieve the best efficiency and accuracy trade-off. For simplicity, in the following examples N and q_{k_1}, q_{k_2}, q_{x_1}, q_{x_2} are chosen such that the relative error between the fast algorithm and the direct computation of equation 3 is about $O(10^{-2})$. These combinations are not necessarily optimal in terms of efficiency.

Synthetic data — square sampling

We start with a simple 2D example of square sampling. Figure 9(a) is a synthetic CMP gather sampled on $N_t = N_h = 1000$. Figure 4 shows the absolute value of its Fourier transform on time axis. These band-limited data allow us to shorten the computational range for f, which can be crucial as N depends on this range. In model space, the sampling sizes are chosen as $N_\tau = N_p = 1000$. Figure 5 is the output of the fast butterfly algorithm for $N = 32$, $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = 9$ (here the range of $\Phi = f\sqrt{\tau^2 + p^2h^2}$ is about 125). Figure 6 is the output of the *velocity scan*. The two methods yield nearly the same results. The fast algorithm runs in only 1.75 s of CPU time, while the *velocity scan* takes about 37 s. In Figure 6(e), we plot the difference between the results of the fast algorithm and the direct evaluation of equation 3, where the relative error is 0.0178. For reference, if we let $N = 64$ and run the same test, the error decreases to $O(10^{-3})$ and the running time is 3.63 s.

Synthetic data — rectangular sampling

We now make two synthetic datasets using rectangular sampling $N_t = 4000$, $N_h = 400$. The first one (Figure 8) has the same range as the previous example (Figure 9(a)), while the second one (Figure 9) doubles the range of time and offset. Results of the fast algorithm are shown in Figures 10 and 11. The purpose of showing these two examples is to demonstrate that the choice of N does not depend on the problem size, but rather on the range of parameters — for the data in Figure 9, one has to increase N to preserve the same accuracy (the range of $\Phi = f\sqrt{\tau^2 + p^2h^2}$ is about 125 for the first dataset, and 250 for the second one).

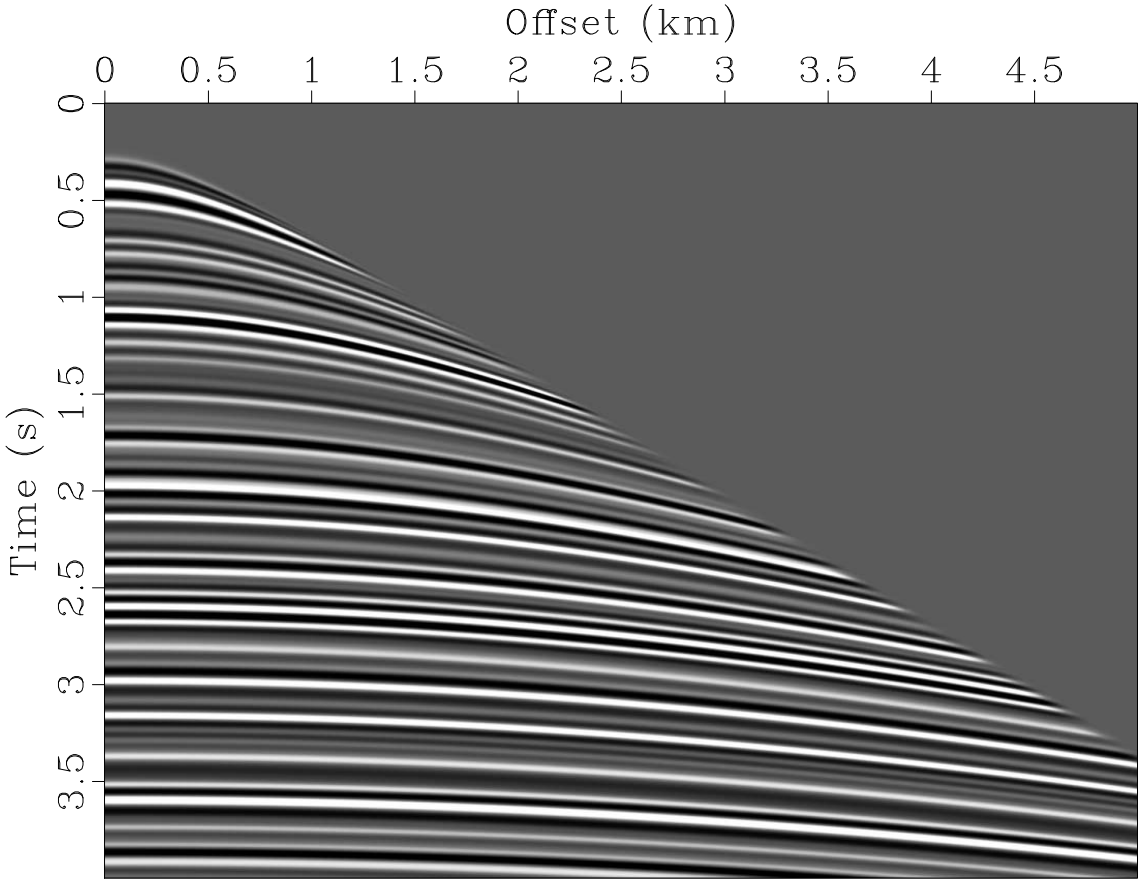


Figure 3: 2D synthetic CMP gather. $N_t = N_h = 1000$. $\Delta t = 0.004$ s, $\Delta h = 0.005$ km. butterfly/synth2D-1-timer data

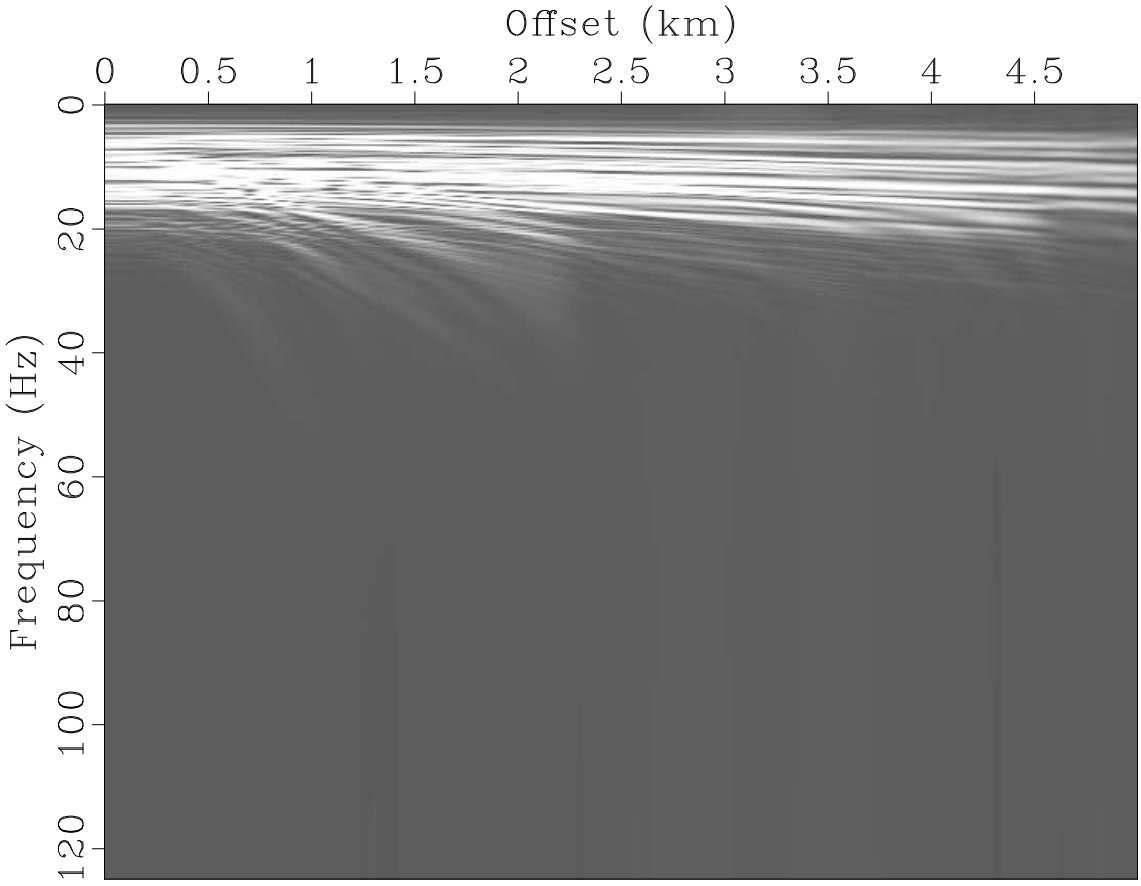


Figure 4: The Fourier transform (absolute value) on time axis of the synthetic data in Figure 9(a). butterfly/synth2D-1-timer fftabs

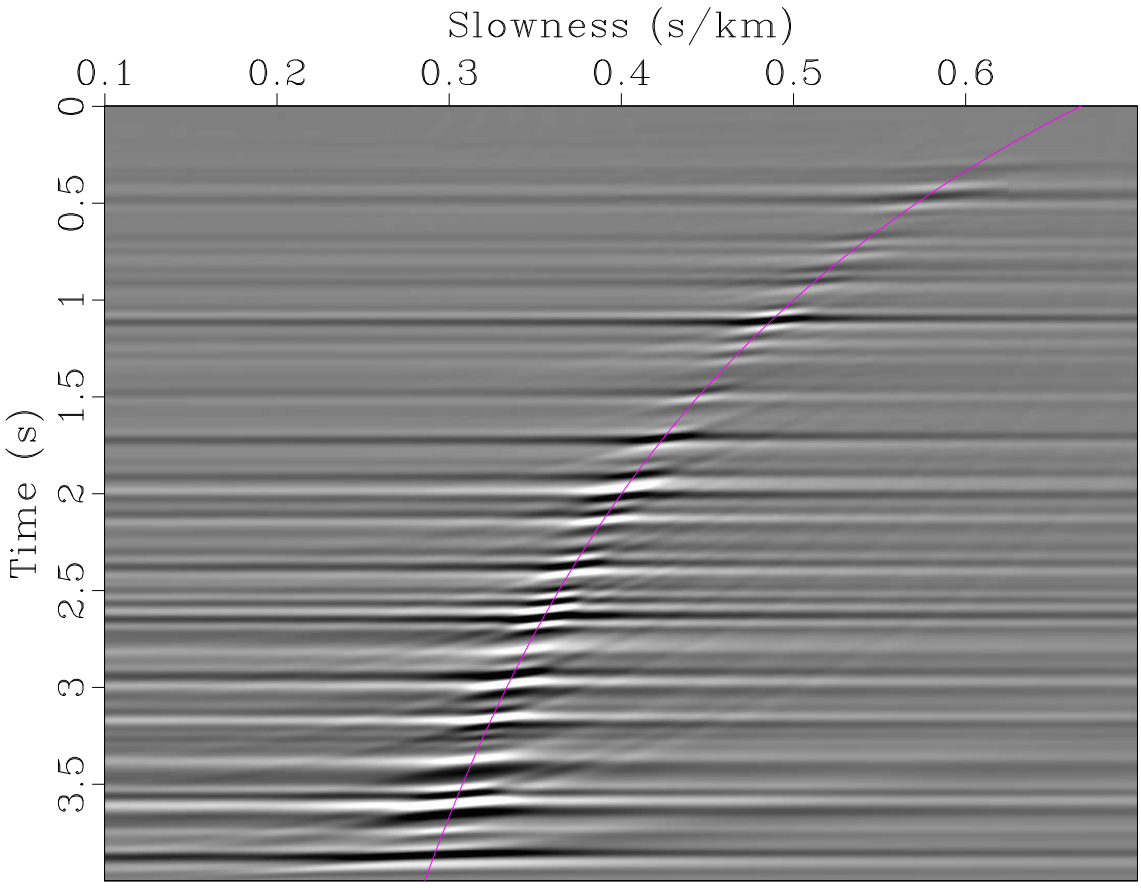


Figure 5: $N_\tau = N_p = 1000$. Output of the fast butterfly algorithm applied to the synthetic data in Figure 9(a). $N = 32$, $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = 9$. CPU time: **1.75 s**. Purple curve overlaid is the true slowness. butterfly/synth2D-1-timer fmod

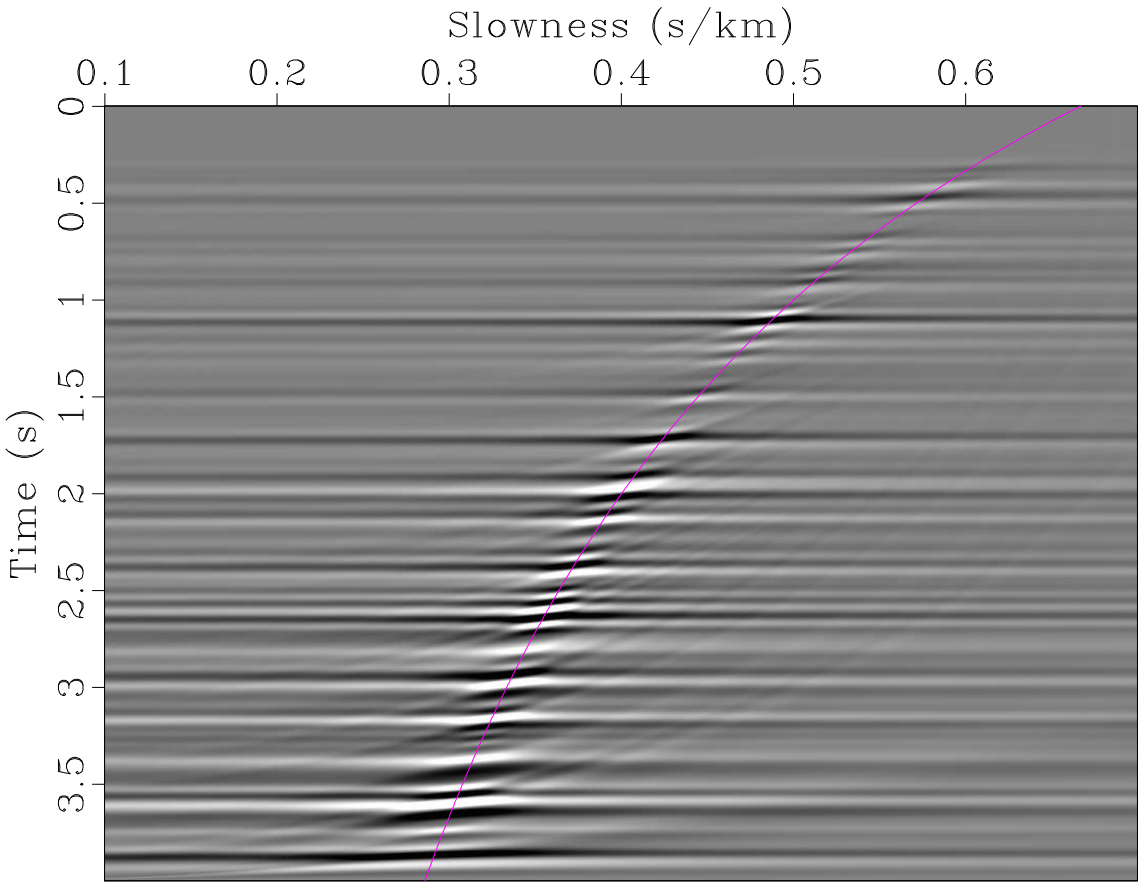


Figure 6: $N_\tau = N_p = 1000$. Output of the *velocity scan* applied to the synthetic data in Figure 9(a). CPU time: **37.23 s**. Purple curve overlaid is the true slowness. butterfly/synth2D-1-timer dimod

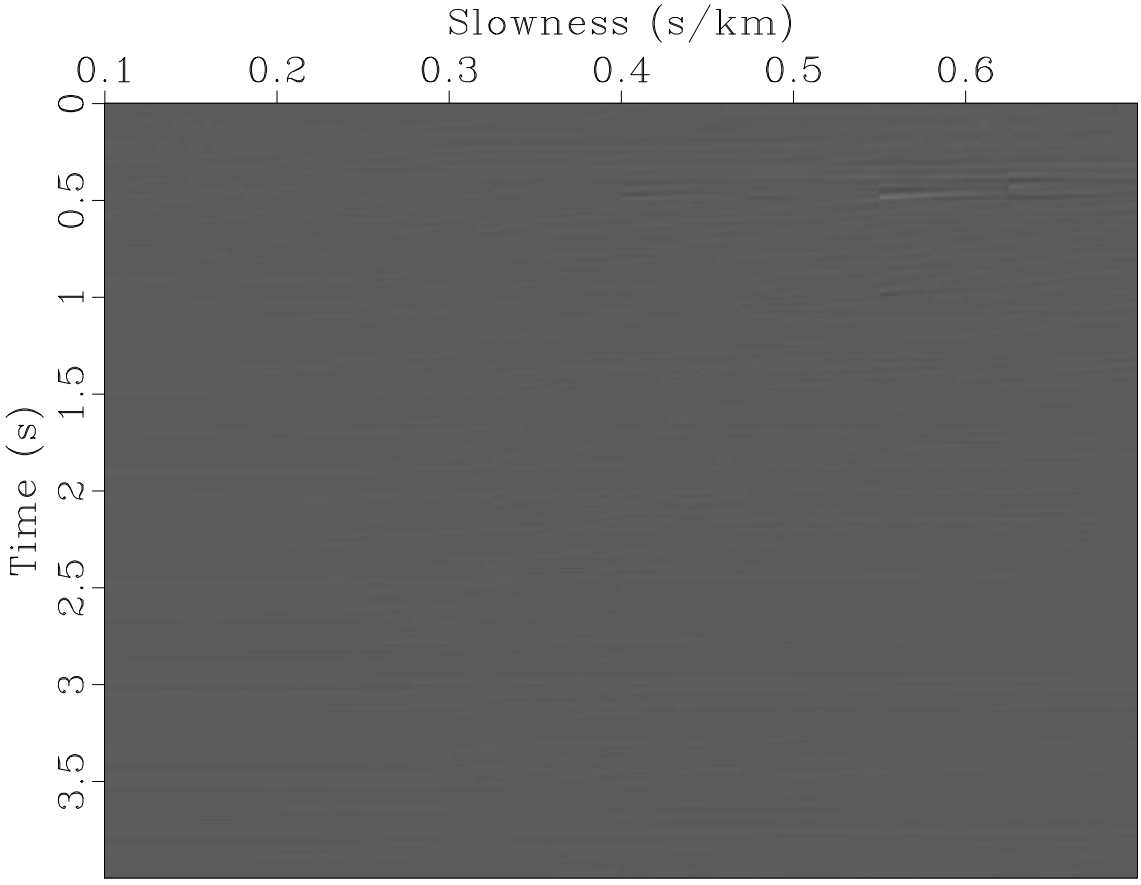


Figure 7: Difference between the results of the fast algorithm and the direct evaluation of equation 3 plotted at the same scale as in Figure 5. butterfly/synth2D-1-timer diff

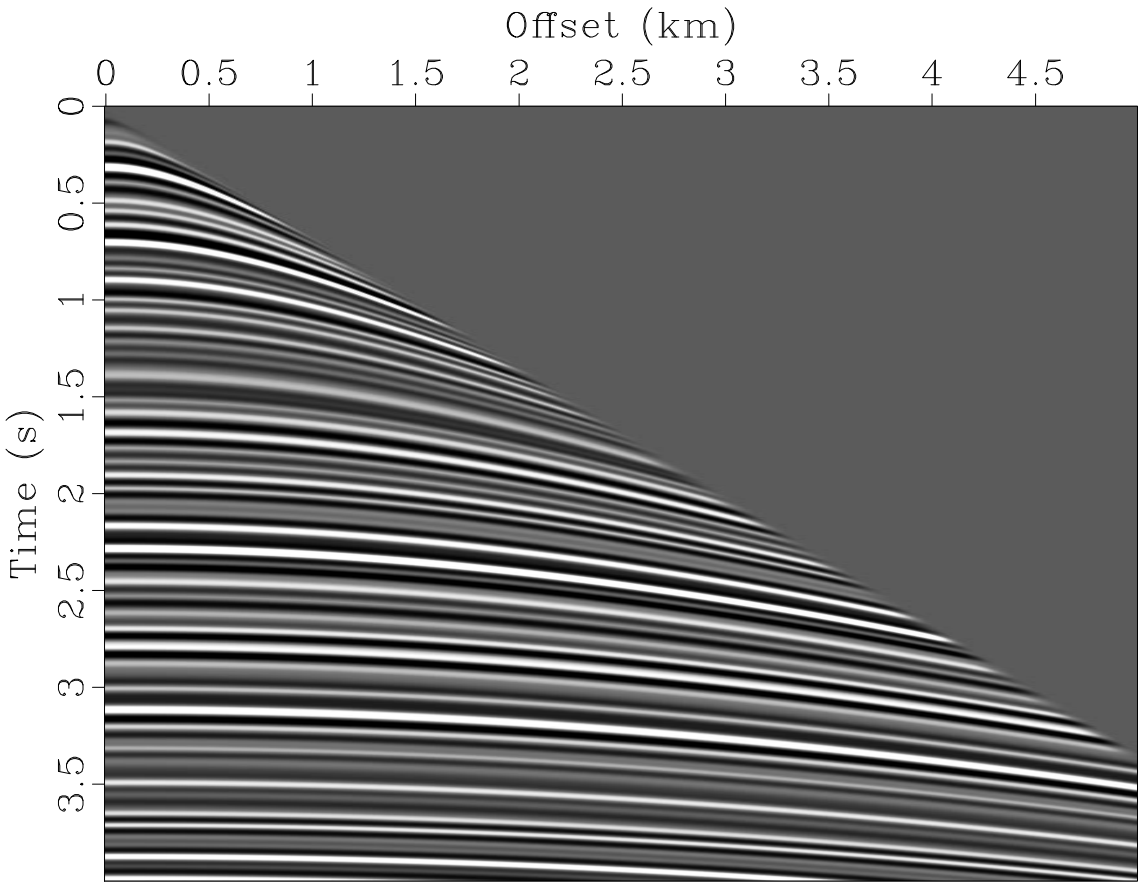


Figure 8: 2D synthetic CMP gather. $N_t = 4000$, $N_h = 400$. $\Delta t = 0.001$ s, $\Delta h = 0.0125$ km. butterfly/synth2D-2 data-2

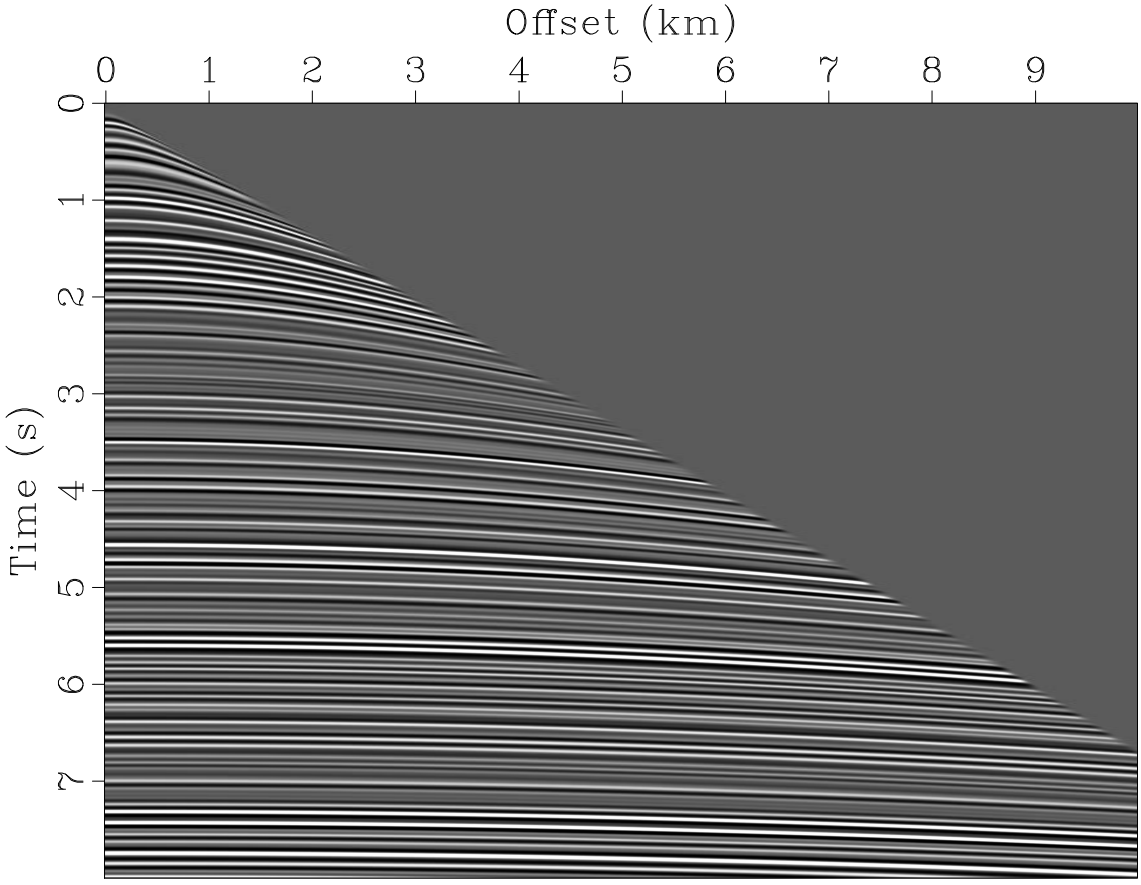


Figure 9: 2D synthetic CMP gather. $N_t = 4000$, $N_h = 400$. $\Delta t = 0.002$ s, $\Delta h = 0.025$ km. butterfly/synth2D-3 data-3

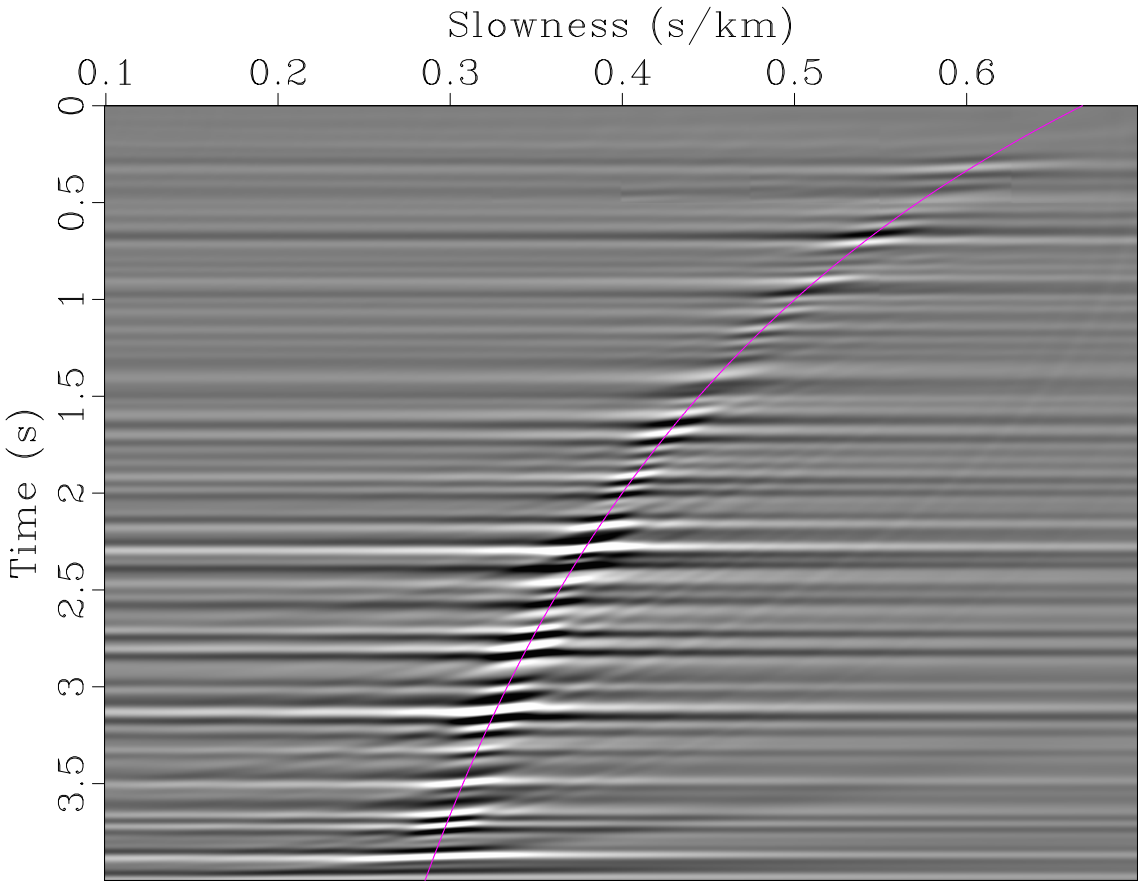


Figure 10: $N_\tau = 4000$, $N_p = 400$. Output of the fast butterfly algorithm applied to the synthetic data in Figure 8. $N = 32$, $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = 9$. CPU time: **2.46 s**. Ref: CPU time of *velocity scan*: **21.84 s**. Purple curve overlaid is the true slowness. butterfly/synth2D-2 fmod-2

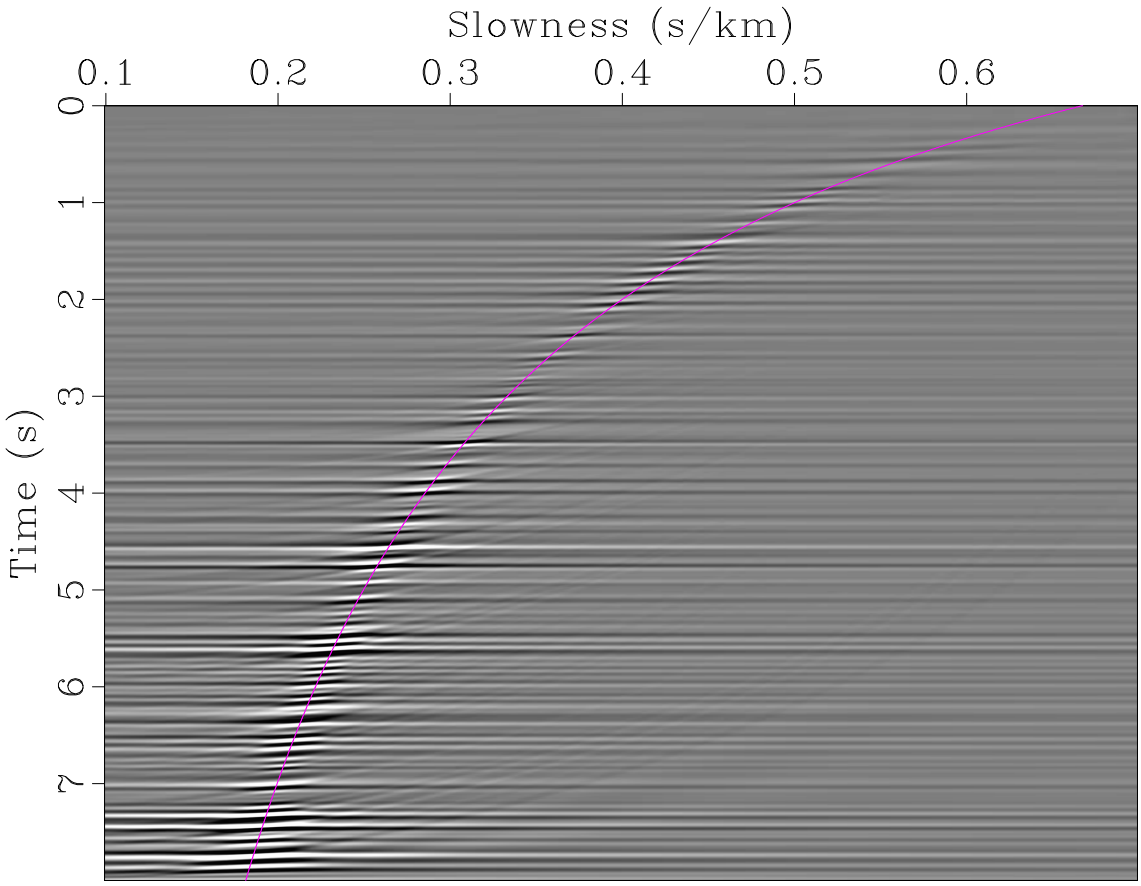


Figure 11: $N_\tau = 4000$, $N_p = 400$. Output of the fast butterfly algorithm applied to the synthetic data in Figure 9. $N = 64$, $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = 9$. CPU time: **4.35 s**. Ref: CPU time of *velocity scan*: **21.93 s**. Purple curve overlaid is the true slowness. butterfly/synth2D-3 fmod-3

Synthetic data — irregular sampling

Going back to the five steps of the butterfly algorithm, it is clear that the input data $g(\mathbf{k})$ is only involved at the very first step. Besides, for every (A, B) the operation connecting $g(\mathbf{k})$ and δ_t^{AB} amounts to a matrix-vector multiplication (see equation 23), which does not at all require the input data to be uniformly distributed (the same argument applies to the output data $u(\mathbf{x})$). Therefore, our algorithm can be easily extended to handle the following problem:

$$(Rd)(\tau, p) = \iint d(\sqrt{\tau^2 + p^2(h_1^2 + h_2^2)}, h_1, h_2)\, dh_1\, dh_2, \tag{28}$$

where $d(t, h_1, h_2)$ is a 3D function. All we need is to introduce a new variable for the absolute offset $h = \sqrt{h_1^2 + h_2^2}$, and reorder the values $d(t, h_1, h_2)$ according to h. Figure 12 shows such synthetic data sampled on $N_t = 1000$, $N_{h_1} = N_{h_2} = 128$. The output is obtained on $N_\tau = 1000$, $N_p = 128$. The fast algorithm (Figure 13) runs in only 1.67 s for $N = 64$, $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = 5$ (here the range of $\Phi = f\sqrt{\tau^2 + p^2(h_1^2 + h_2^2)}$ is about 162), while the *velocity scan* (Figure 14) takes more than 125 s.

Field data

We now consider a 2D field seismic gather shown in Figure 6(a). Its Fourier transform is shown in Figure 16. Due to the comparatively wide frequency bandwidth, N cannot be chosen too small (here the range of $\Phi = f\sqrt{\tau^2 + p^2h^2}$ is about 306). The input sampling sizes are $N_t = 1500$, $N_h = 240$, while the output sizes are chosen as $N_\tau = 1500$, $N_p = 800$. Although this small dataset is not very suitable for showcasing the fast algorithm, our method runs in 6.62 s for $N = 128$, $q_{k_1} = q_{x_1} = 7$, $q_{k_2} = q_{x_2} = 5$ (Figure 17), still outperforming the *velocity scan* which takes about 10 s (Figure 18). Note that the simplest interpolation is used in the *velocity scan*, any other higher order interpolation should take longer computation time.

Computing the adjoint operator

The last example is concerned with the computation of the adjoint of the hyperbolic Radon transform. Assuming $m(\tau, p)$ and $d(t, h)$ are two arbitrary functions (in the discrete sense) in the model domain and data domain, if we require

$$\langle m(\tau, p), (Rd)(\tau, p)\rangle = \langle (R^* m)(t, h), d(t, h)\rangle, \tag{29}$$

where $(Rd)(\tau, p)$ is given by equation 3, the inner product $\langle \cdot, \cdot \rangle$ is defined as

$$\langle g_1(x, y), g_2(x, y)\rangle = \sum_{x,y} g_1(x, y)\, \overline{g_2(x, y)}, \quad \forall g_1(x, y),\ g_2(x, y); \tag{30}$$

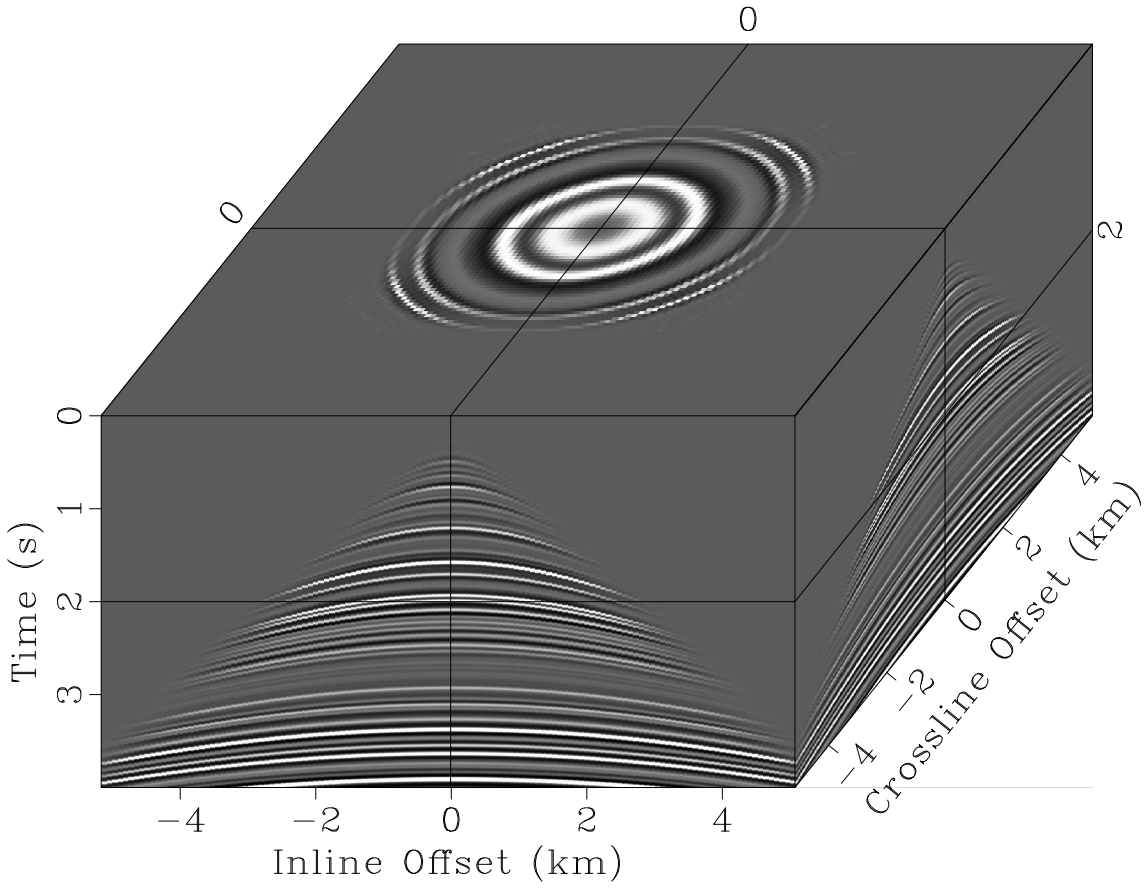


Figure 12: 3D synthetic CMP gather. $N_t = 1000$, $N_{h_1} = N_{h_2} = 128$. $\Delta t = 0.004$ s, $\Delta h_1 = \Delta h_2 = 0.08$ km. butterfly/synth2D-4 data-4

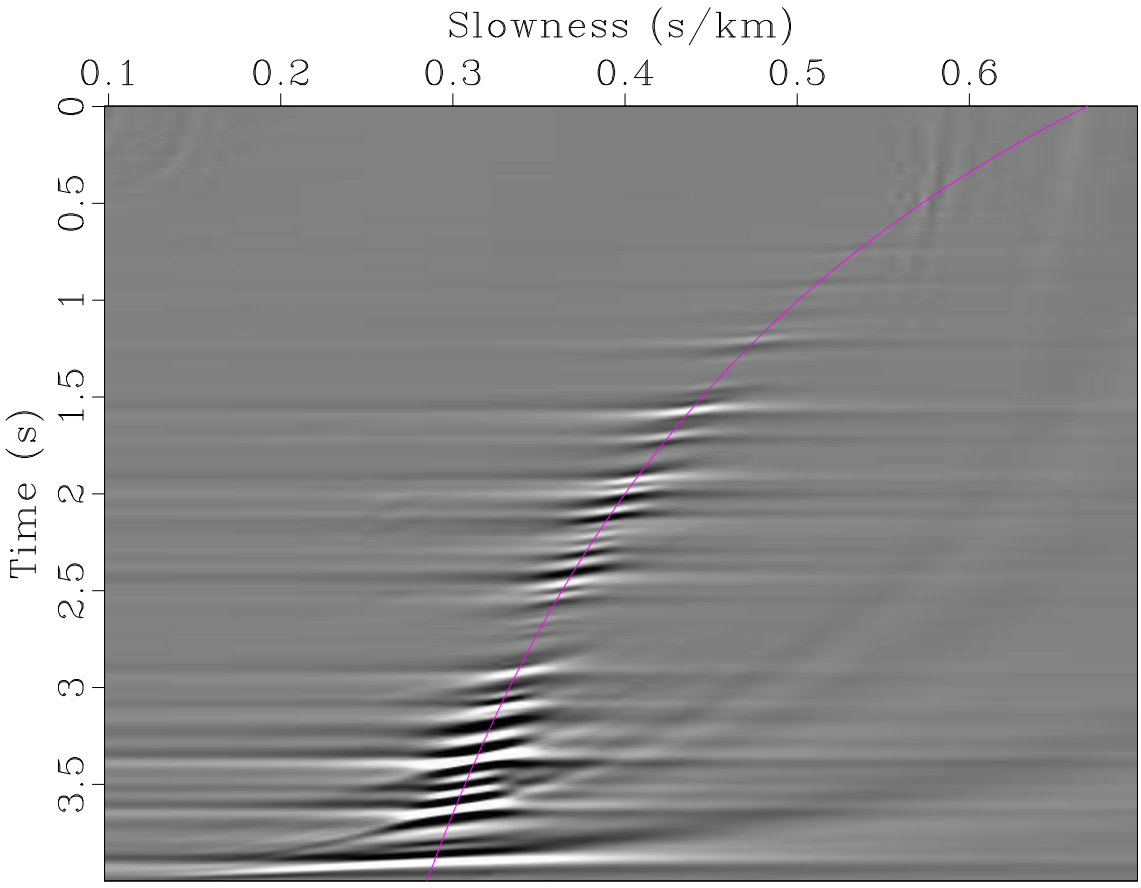


Figure 13: $N_\tau = 1000$, $N_p = 128$. Output of the fast butterfly algorithm applied to the synthetic data in Figure 12. $N = 64$, $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = 5$. CPU time: **1.67 s**. Purple curve overlaid is the true slowness. butterfly/synth2D-4 fmod-4

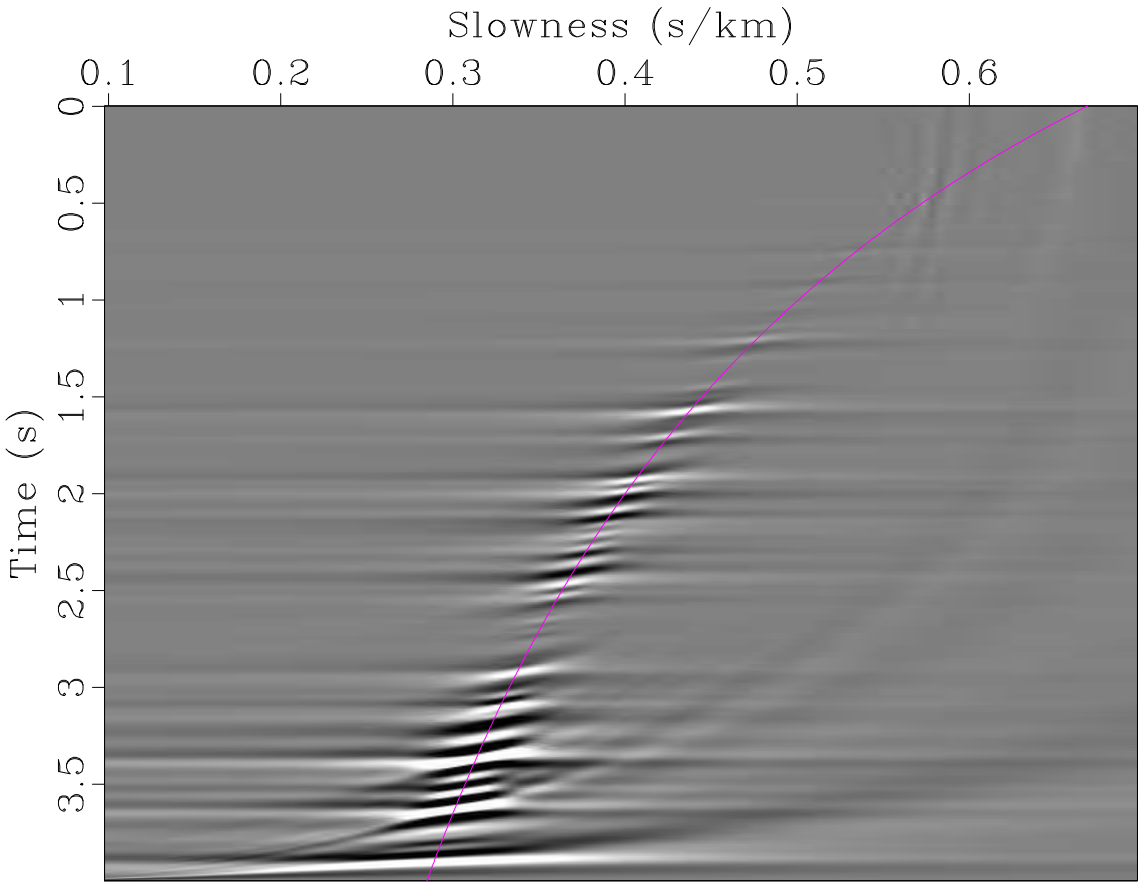


Figure 14: $N_\tau = 1000$, $N_p = 128$. Output of the *velocity scan* applied to the synthetic data in Figure 12. CPU time: **125.54 s**. Purple curve overlaid is the true slowness. butterfly/synth2D-4 dimod-4

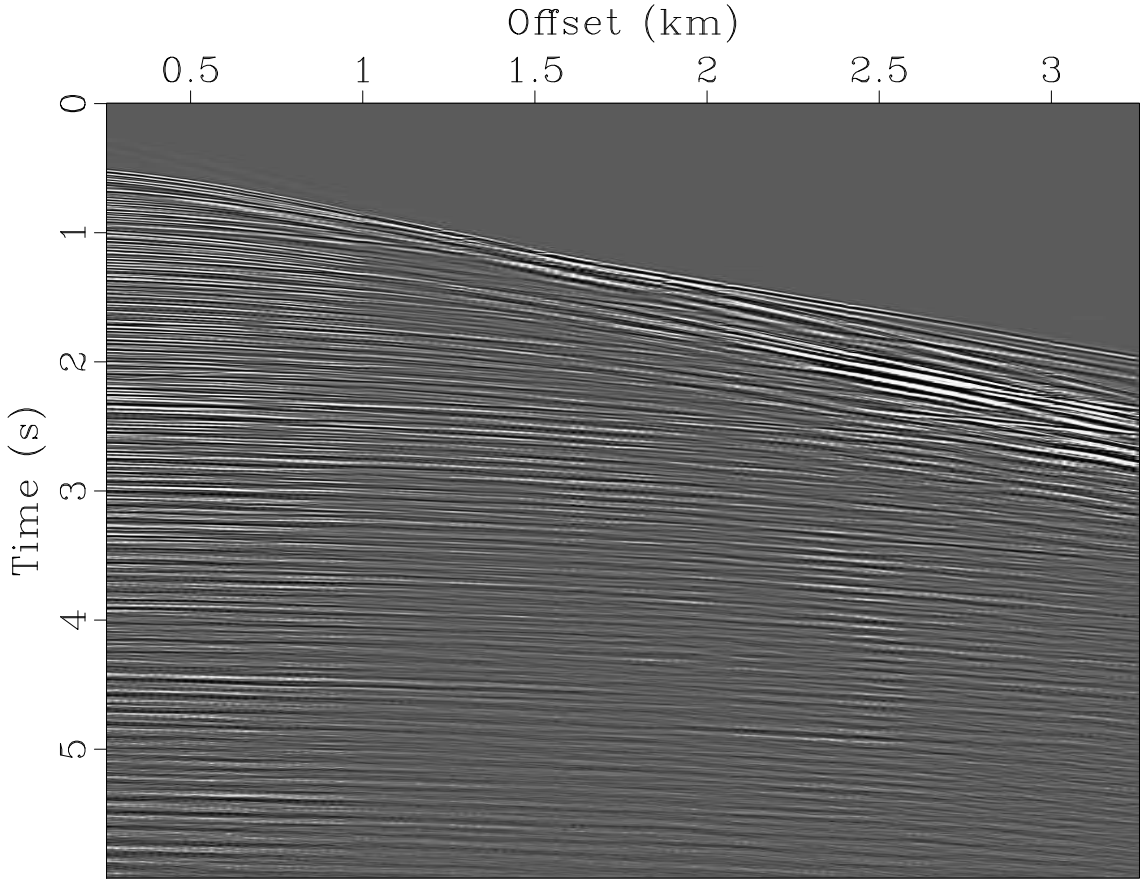


Figure 15: 2D field CMP gather. $N_t = 1500$, $N_h = 240$. $\Delta t = 0.004$ s, $\Delta h = 0.0125$ km. butterfly/field2 samiss

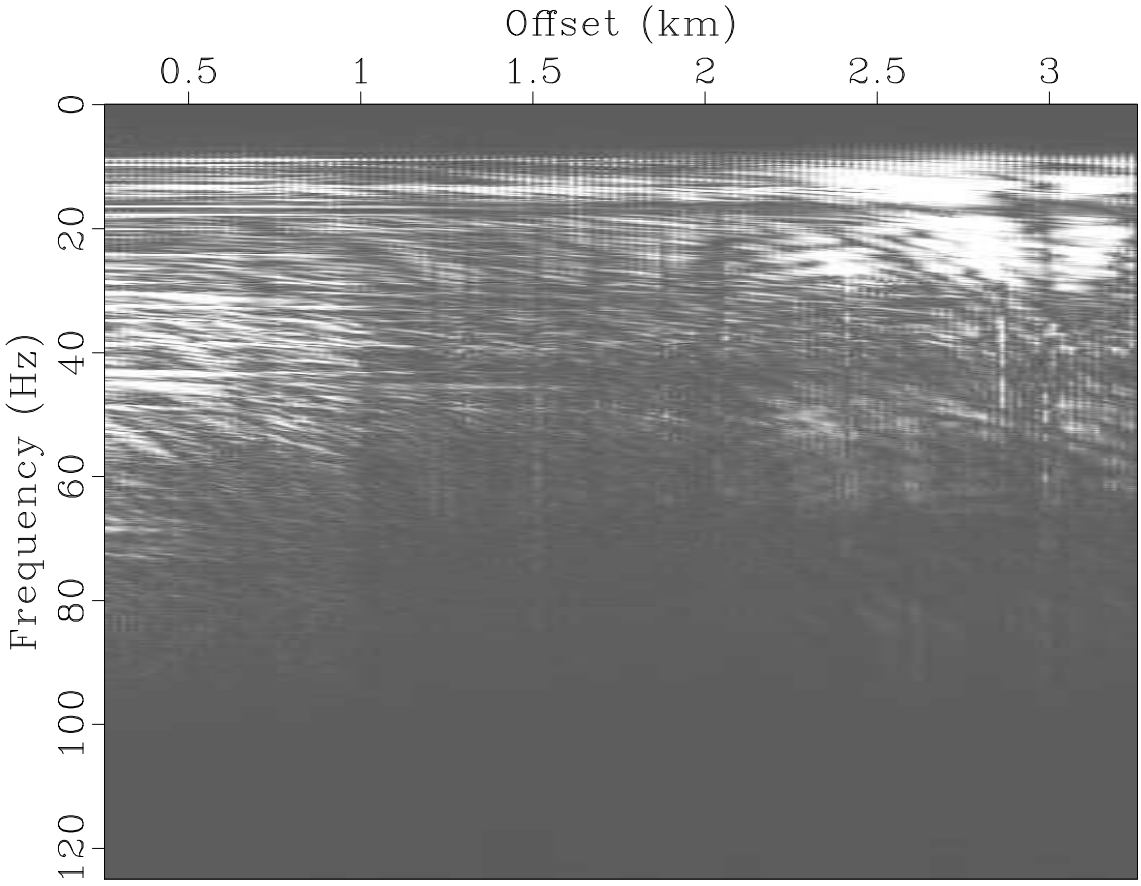


Figure 16: The Fourier transform (absolute value) on time axis of the field data in Figure 6(a). butterfly/field2 fftabs-f

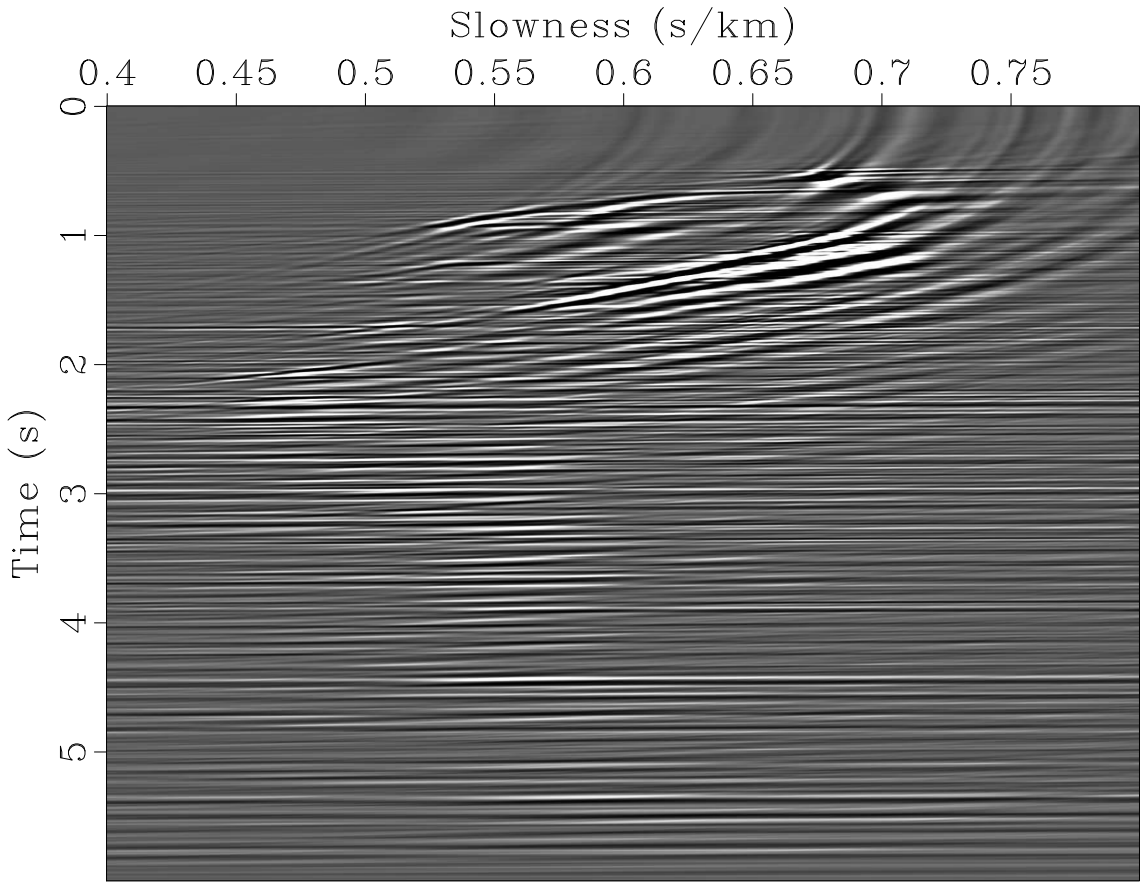


Figure 17: $N_\tau = 1500$, $N_p = 800$. Output of the fast butterfly algorithm applied to the field data in Figure 6(a). $N = 128$, $q_{k_1} = q_{x_1} = 7$, $q_{k_2} = q_{x_2} = 5$. CPU time: **6.62 s**. butterfly/field2 fmod-f

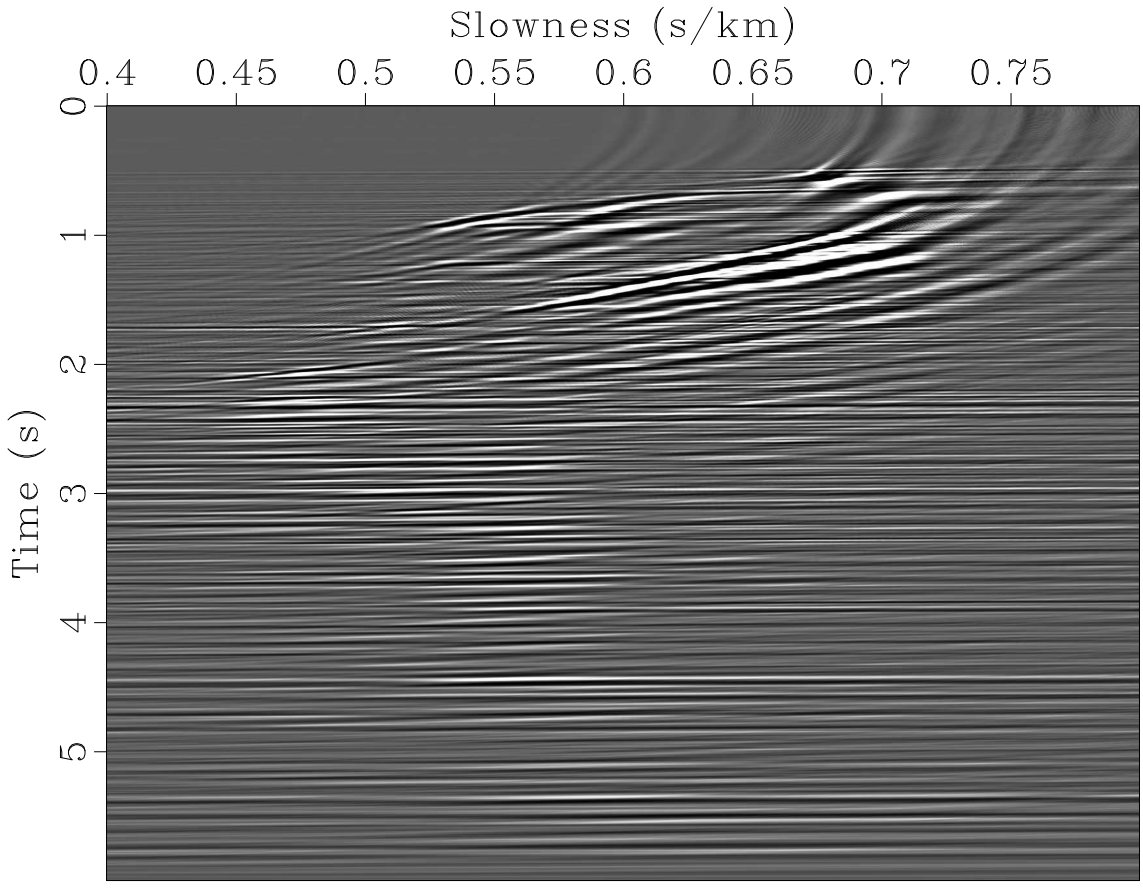


Figure 18: $N_\tau = 1500$, $N_p = 800$. Output of the *velocity scan* applied to the field data in Figure 6(a). CPU time: **9.91 s**. butterfly/field2 dimod-f

then it is easy to verify that the adjoint operator R^* is given by

$$(R^*m)(t,h) = \mathcal{F}^{-1}_{f\to t}\left(\sum_{\tau,p} e^{-2\pi i f\sqrt{\tau^2+p^2h^2}} m(\tau,p)\right), \tag{31}$$

where $\mathcal{F}^{-1}_{f\to t}$ is the inverse Fourier transform from variable f to t. The summation in equation 31 again resembles an oscillatory integral operator, therefore the fast algorithm for computing R applies with minor modifications. The computational cost remains the same.

We consider still the first example and apply the (discrete) adjoint operators of the fast butterfly algorithm and the *velocity scan* respectively to the data in Figures 5 and 6. The two methods produce similar results (see Figures 19, 20). It is also clear that the adjoint is far from the inverse, at least for this geometry, hence some kind of least-squares implementation is needed for inversion process.

To further verify that the numerically computed R^* is the adjoint operator of R, one can compare the values of $\langle Rd, Rd\rangle$ and $\langle R^*Rd, d\rangle$ for arbitrary d. Indeed, the proposed algorithm passed this dot-product test with a relative error of $O(10^{-7})$ in single precision.

CONCLUSIONS

We constructed a fast butterfly algorithm for the hyperbolic Radon transform, a type of generalized Radon transforms. Compared with expensive integration in the time domain, the new method runs in only $O(N^2 \log N)$ operations, where N depends on the range of frequency and offset in the dataset and the range of intercept time and slowness in the model space, and can often be chosen smaller than the grid size. In practice, this may lead to speedup of several orders of magnitude. Our ongoing work is studying the performance of this fast solver on the sparse iterative inversion of the hyperbolic Radon transform applied to multiple attenuation.

Due to the generality of the butterfly algorithm, its application is not limited to the hyperbolic transform considered here. Using a different phase function, one can easily extend the algorithm to higher-order transforms. If the slowness or velocity range is not constant but a corridor around a central function, then a sparse butterfly algorithm can be designed to save the cost by building the quad tree adaptively. Furthermore, many of the Radon-like integral operators, such as Kirchhoff migration, the apex-shifted Radon transform, the anisotropic multiparameter velocity scan, etc., can be reformulated in a similar fashion as we did in this paper. To address these extensions, a 3D version of the butterfly algorithm might be more appropriate.

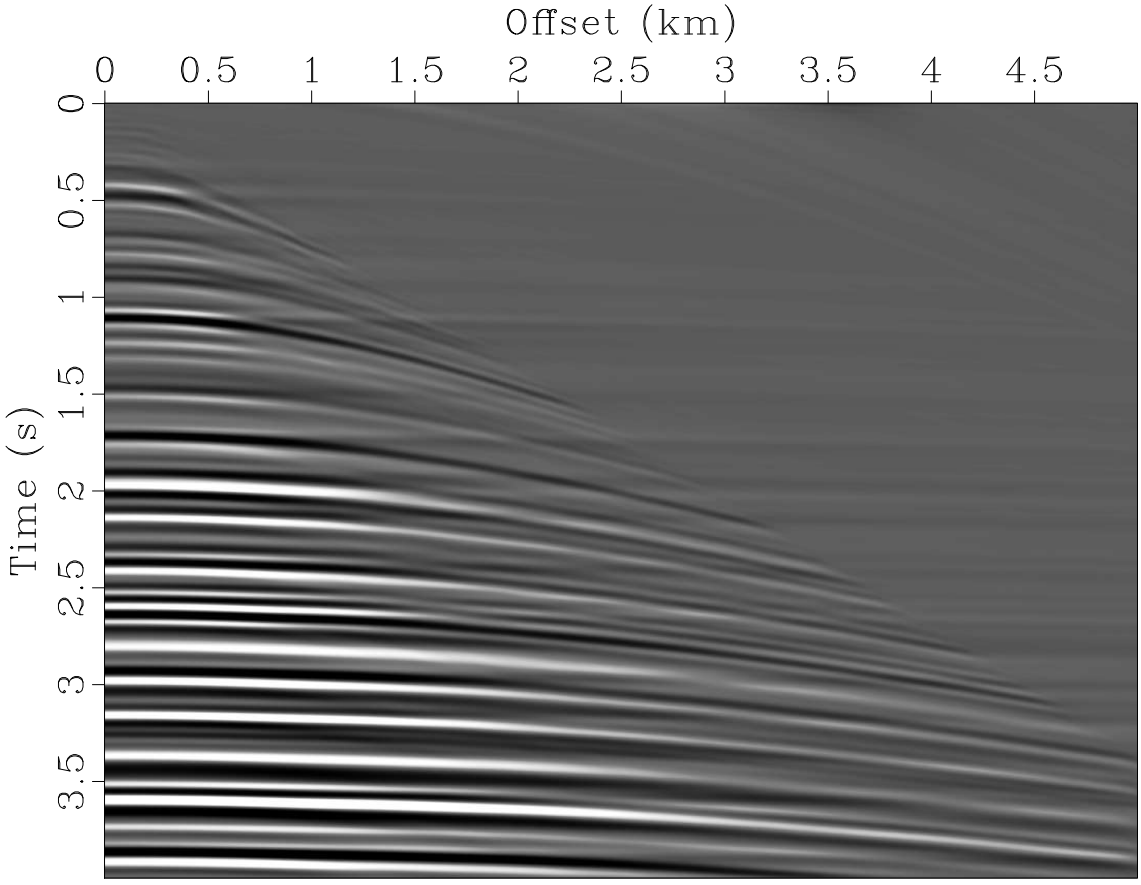


Figure 19: Output of the adjoint fast butterfly algorithm applied to the data in Figure 5. $N = 32$, $q_{k_1} = q_{k_2} = q_{x_1} = q_{x_2} = 9$. butterfly/synth2D-1-timer fdat

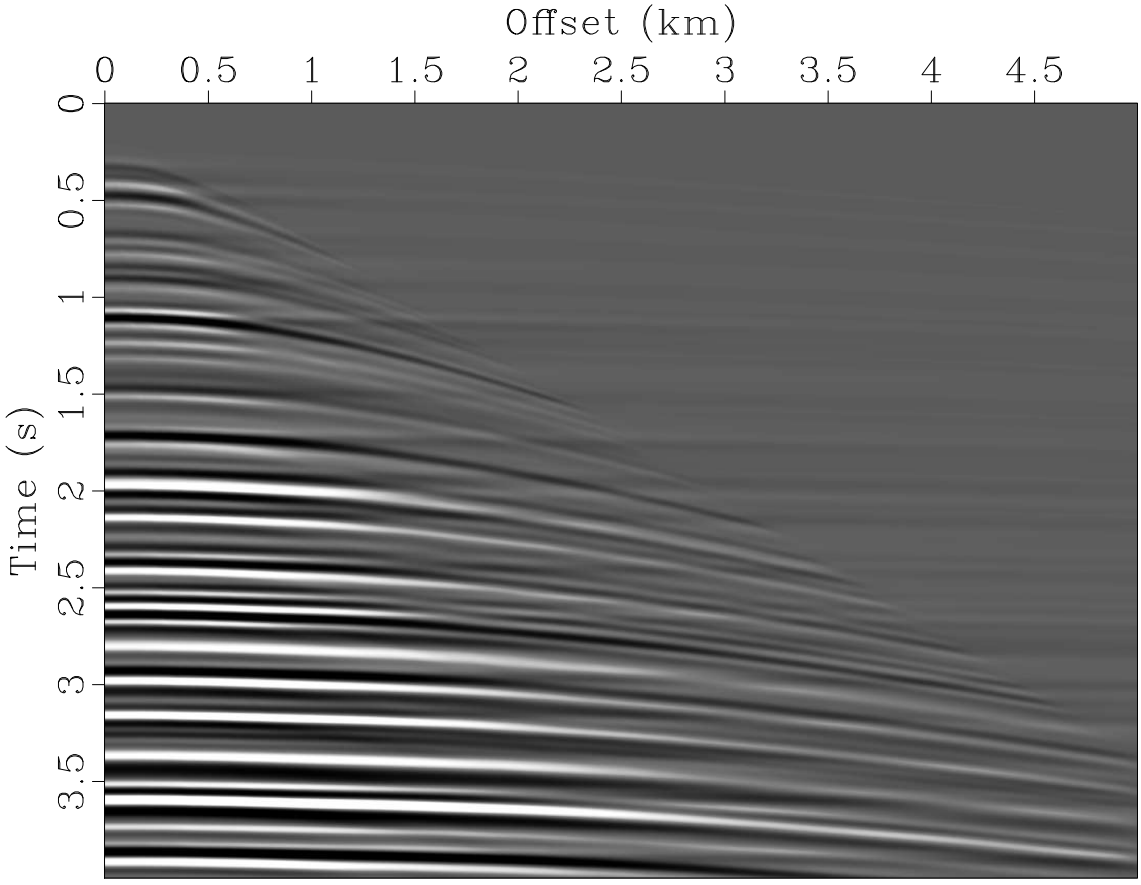


Figure 20: Output of the adjoint *velocity scan* applied to the data in Figure 6. butterfly/synth2D-1-timer didat

ACKNOWLEDGMENTS

We are grateful to Tariq Alkhalifah, Anatoly Baumstein, Ian Moore, Daniel Trad, and the anonymous reviewer for their valuable comments and suggestions. We thank Dr. Alexander Klokov for preprocessing the field data. We thank KAUST and sponsors of the Texas Consortium for Computational Seismology (TCCS) for financial support.

APPENDIX A

THE MATHEMATICAL DERIVATION OF THE FAST BUTTERFLY ALGORITHM

This appendix gives a complete derivation and description of the butterfly algorithm, which combines the low-rank approximations and the butterfly structure introduced in the main text. For more mathematical exposition, the reader is referred to (3).

To facilitate the presentation, we add a new figure (Figure A-1) to illustrate the notations.

1. *Initialization.* At level $l = 0$, let A be the root box of T_X. For each leaf box $B \in T_K$, expressions 9 and 10 are valid as $w(B) \leq 1/\sqrt{N}$. Substituting β_t^{AB} (in equation 10) into the definition of δ_t^{AB}, equation 21, we get

$$\delta_t^{AB} = e^{-2\pi i \Phi(\mathbf{x}_0(A), \mathbf{k}_t^B)} \sum_{\mathbf{k} \in B} \left(L_t^B(\mathbf{k}) e^{2\pi i \Phi(\mathbf{x}_0(A), \mathbf{k})} g(\mathbf{k}) \right), \tag{A-1}$$

i.e. the equation 23 in the main text. In addition, for $\mathbf{x} \in A$, the partial sum $u^B(\mathbf{x})$ in equation 20 is given by (with α_t^{AB} (in equation 9) plugged in)

$$u^B(\mathbf{x}) \approx \sum_t e^{2\pi i \Phi(\mathbf{x}, \mathbf{k}_t^B)} \delta_t^{AB}. \tag{A-2}$$

Comparing the right hand sides of equations 19 and A-2, if we call $g(\mathbf{k})$ the sources at $\mathbf{k}$, then coefficients δ_t^{AB} are just like the *equivalent sources* at $\mathbf{k}_t^B$. This initial step is to redistribute the original sources $g(\mathbf{k})$ located at $\mathbf{k}$ (denoted by blue dots in Figure A-1) to equivalent sources δ_t^{AB} located at Chebyshev grid $\mathbf{k}_t^B$ (not shown in the figure). We next aim at updating δ_t^{AB} until the end level L.

2. *Recursion.* At $l = 1, 2, ..., L/2$, for each pair (A, B), let A_p be A's parent and $\{B_c, c = 1, 2, 3, 4\}$ be B's children from the previous level (see Figure A-1). For each child B_c, we have available from the previous level an approximation of the form

$$u^{B_c}(\mathbf{x}) \approx \sum_{t'} e^{2\pi i \Phi(\mathbf{x}, \mathbf{k}_{t'}^{B_c})} \delta_{t'}^{A_p B_c}, \quad \text{for } \mathbf{x} \in A_p. \tag{A-3}$$

Summing over all children gives

$$u^B(\mathbf{x}) \approx \sum_c \sum_{t'} e^{2\pi i \Phi(\mathbf{x}, \mathbf{k}_{t'}^{B_c})} \delta_{t'}^{A_p B_c}, \quad \text{for } \mathbf{x} \in A_p. \tag{A-4}$$

Since $A \subset A_p$, this is of course true for any $\mathbf{x} \in A$. Also we know that equation 17 holds for $\mathbf{k}_{t'}^{B_c} \in B$, i.e.,

$$e^{2\pi\Phi(\mathbf{x},\mathbf{k}_{t'}^{B_c})} \approx \sum_t e^{2\pi i\Phi(\mathbf{x},\mathbf{k}_t^B)}\beta_t^{AB}(\mathbf{k}_{t'}^{B_c}), \quad \text{for } \mathbf{x} \in A. \tag{A-5}$$

Inserting it into expression A-4 yields

$$u^B(\mathbf{x}) \approx \sum_c \sum_{t'} \sum_t e^{2\pi i\Phi(\mathbf{x},\mathbf{k}_t^B)}\beta_t^{AB}(\mathbf{k}_{t'}^{B_c})\delta_{t'}^{A_pB_c}, \quad \text{for } \mathbf{x} \in A. \tag{A-6}$$

On the other hand, $u^B(\mathbf{x})$ admits a low-rank approximation of equivalent sources at the current level,

$$u^B(\mathbf{x}) \approx \sum_t e^{2\pi i\Phi(\mathbf{x},\mathbf{k}_t^B)}\delta_t^{AB}, \quad \text{for } \mathbf{x} \in A. \tag{A-7}$$

Equating expressions A-6 and A-7 suggests that we can take

$$\delta_t^{AB} = \sum_c \sum_{t'} \beta_t^{AB}(\mathbf{k}_{t'}^{B_c})\delta_{t'}^{A_pB_c}. \tag{A-8}$$

Substituting β_t^{AB} (in equation 10), we get

$$\delta_t^{AB} = e^{-2\pi i\Phi(\mathbf{x}_0(A),\mathbf{k}_t^B)} \sum_c \sum_{t'} \left(L_t^B(\mathbf{k}_{t'}^{B_c}) e^{2\pi i\Phi(\mathbf{x}_0(A),\mathbf{k}_{t'}^{B_c})} \delta_{t'}^{A_pB_c} \right), \tag{A-9}$$

i.e. the equation 24 in the main text.

3. *Switch.* A switch of the representation to expressions 11 and 12 is needed at the middle level $l = L/2$ since expressions 9 and 10 are no longer valid as soon as $l > L/2$ (boxes B are getting bigger and bigger so that $w(B) \leq 1/\sqrt{N}$ is no longer satisfied). Plugging β_t^{AB} (in equation 12) into the definition of δ_t^{AB}, equation 21, one has

$$\delta_t^{AB} = \sum_{\mathbf{k}\in B} e^{2\pi\Phi(\mathbf{x}_t^A,\mathbf{k})} g(\mathbf{k}) = u^B(\mathbf{x}_t^A); \tag{A-10}$$

from expression A-7,

$$u^B(\mathbf{x}_t^A) \approx \sum_s e^{2\pi i\Phi(\mathbf{x}_t^A,\mathbf{k}_s^B)}\delta_s^{AB}, \tag{A-11}$$

where we use $\{\delta_t^{AB}\}$ to denote the new set of coefficients and $\{\delta_s^{AB}\}$ the old set. Equating expressions A-10 and A-11, we can set δ_t^{AB} as

$$\delta_t^{AB} = \sum_s e^{2\pi i\Phi(\mathbf{x}_t^A,\mathbf{k}_s^B)}\delta_s^{AB}, \tag{A-12}$$

i.e. the equation 25 in the main text. This middle step is to switch from equivalent sources δ_s^{AB} located at Chebyshev grid $\mathbf{k}_s^B$ on the K side to equivalent sources δ_t^{AB} located at Chebyshev grid $\mathbf{x}_t^A$ on the X side.

4. *Recursion.* The rest of the recursion is analogous to Step 2. For $l = L/2 + 1, ..., L$, we have

$$u^B(\mathbf{x}) \approx \sum_c \sum_{t'} \alpha_{t'}^{A_p B_c}(\mathbf{x}) \delta_{t'}^{A_p B_c}, \quad \text{for } \mathbf{x} \in A_p, \tag{A-13}$$

thus

$$u^B(\mathbf{x}_t^A) \approx \sum_c \sum_{t'} \alpha_{t'}^{A_p B_c}(\mathbf{x}_t^A) \delta_{t'}^{A_p B_c}; \tag{A-14}$$

recalling expression A-10, one can simply set

$$\delta_t^{AB} = \sum_c \sum_{t'} \alpha_{t'}^{A_p B_c}(\mathbf{x}_t^A) \delta_{t'}^{A_p B_c}. \tag{A-15}$$

Inserting α_t^{AB} (in equation 11) gives the update

$$\delta_t^{AB} = \sum_c e^{2\pi i \Phi(\mathbf{x}_t^A, \mathbf{k}_0(B_c))} \sum_{t'} \left(L_{t'}^{A_p}(\mathbf{x}_t^A) e^{-2\pi i \Phi(\mathbf{x}_{t'}^{A_p}, \mathbf{k}_0(B_c))} \delta_{t'}^{A_p B_c} \right), \tag{A-16}$$

i.e. the equation 26 in the main text.

5. *Termination.* Finally we reach the level $l = L$, and B is the entire domain K. For every box A in X and every $\mathbf{x} \in A$,

$$u(\mathbf{x}) = u^B(\mathbf{x}) \approx \sum_t \alpha_t^{AB}(\mathbf{x}) \delta_t^{AB}. \tag{A-17}$$

Plugging in α_t^{AB} (in equation 11), we get

$$u(\mathbf{x}) = e^{2\pi i \Phi(\mathbf{x}, \mathbf{k}_0(B))} \sum_t \left(L_t^A(\mathbf{x}) e^{-2\pi i \Phi(\mathbf{x}_t^A, \mathbf{k}_0(B))} \delta_t^{AB} \right), \tag{A-18}$$

i.e. the equation 27 in the main text. This final step is to transform the equivalent sources δ_t^{AB} located at Chebyshev grid $\mathbf{x}_t^A$ back to the targets $u(\mathbf{x})$ located at $\mathbf{x}$ (denoted by red dots in Figure A-1).

In the above algorithm, $L = \log N$ is assumed to be an even number. If L is odd, one can either switch at level $(L-1)/2$ or $(L+1)/2$. Everything else remains unchanged.

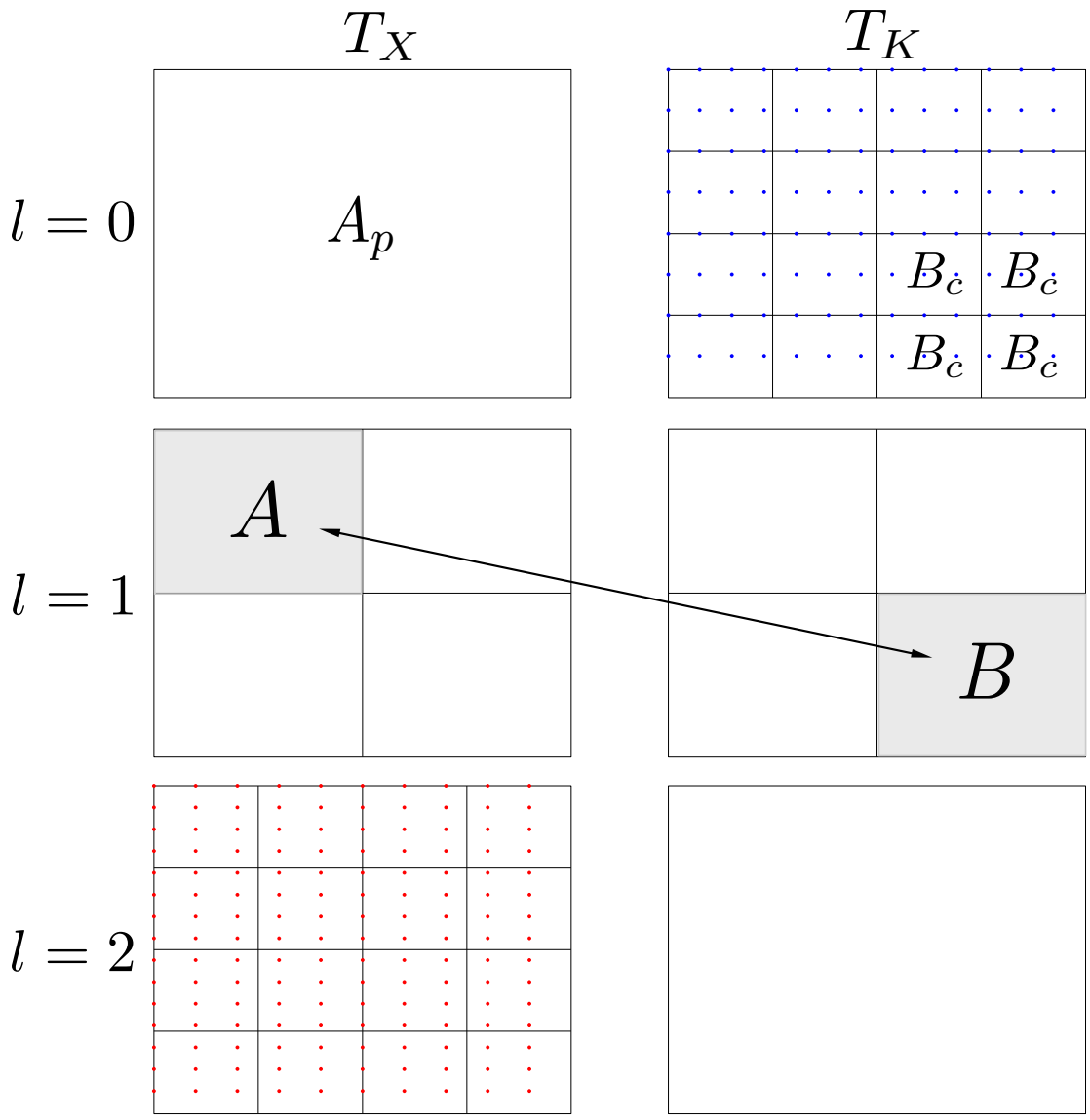


Figure A-1: The butterfly structure for the special case of $N = 4$. The top right panel represents the input domain K with sources $g(\mathbf{k})$ located at $\mathbf{k}$ (blue dots). The bottom left panel represents the output domain X with targets $u(\mathbf{x})$ located at $\mathbf{x}$ (red dots). For the pair of boxes (A, B) at level $l = 1$, box A_p is called A's parent at the previous level; four small boxes B_c are called B's children at the previous level. butterfly/. illus

REFERENCES

[1] Beylkin, G., 1984, The inversion problem and applications of the generalized Radon transform: Communications on Pure and Applied Mathematics, **37**, 579–599.

[2] ———, 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: Journal of Mathematical Physics, **26**, 99–108.

[3] Candès, E., L. Demanet, and L. Ying, 2009, A fast butterfly algorithm for the computation of Fourier integral operators: Multiscale Modeling and Simulation, **7**, 1727–1750.

[4] Demanet, L., M. Ferrara, N. Maxwell, J. Poulson, and L. Ying, 2012, A butterfly algorithm for synthetic aperture radar imaging: SIAM Journal on Imaging Sciences, **5**, 203–243.

[5] Foster, D. J., and C. C. Mosher, 1992, Suppression of multiple reflections using the Radon transform: Geophysics, **57**, 386–395.

[6] Gardner, G. H. F., and L. Lu, eds., 1991, Slant-stack processing: Society of Exploration Geophysicists. Issue 14 of Geophysics reprint series.

[7] Hampson, D., 1986, Inverse velocity stacking for multiple elimination: 56th Annual International Meeting, SEG, Expanded Abstracts, 422–424.

[8] Hargreaves, N., B. verWest, R. Wombell, and D. Trad, 2003, Multiple attenuation using an apex-shifted Radon transform: 65th Conference and Exhibition, EAGE, Extended Abstracts.

[9] Herrmann, P., T. Mojesky, M. Magesan, and P. Hugonnet, 2000, De-aliased, high-resolution Radon transforms: 70th Annual International Meeting, SEG, Expanded Abstracts.

[10] Liu, Y., and M. Sacchi, 2002, De-multiple via a fast least squares hyperbolic Radon transform: 72nd Annual International Meeting, SEG, Expanded Abstracts.

[11] Michielssen, E., and A. Boag, 1996, A multilevel matrix decomposition algorithm for analyzing scattering from large structures: IEEE Transactions on Antennas and Propagation, **44**, 1086–1093.

[12] Moore, I., and C. Kostov, 2002, Stable, efficient, high-resolution Radon transforms: 64th Conference and Exhibition, EAGE, Extended Abstracts.

[13] O'Neil, M., F. Woolfe, and V. Rokhlin, 2010, An algorithm for the rapid evaluation of special function transforms: Applied and Computational Harmonic Analysis, **28**, 203–226.

[14] Radon, J., 1917, Über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten: Berichte über die Verhandlungen der Sächsische Akademie der Wissenschaften (Reports on the proceedings of the Saxony Academy of Science), **69**, 262–277.

[15] Sacchi, M., 1996, A bidiagonalization procedure for the inversion of time-variant velocity stack operator: CDSST report, 73–92.

[16] Thorson, J. R., and J. F. Claerbout, 1985, Velocity-stack and slant-stack stochastic inversion: Geophysics, **50**, 2727–2741.

[17] Trad, D., 2003, Interpolation and multiple attenuation with migration operators:

Geophysics, **68**, 2043–2054.

[18] Trad, D., T. Ulrych, and M. Sacchi, 2002, Accurate interpolation with high resolution time-variant Radon transforms: Geophysics, **67**, 644–656.

[19] Wang, J., and M. Ng, 2009, Greedy least-squares and its application in Radon transforms: 2009 CSPG CSEG CWLS Convention.

[20] Yilmaz, O., 1989, Velocity-stack processing: Geophysical Prospecting, **37**, 357–382.

[21] Ying, L., 2009, Sparse Fourier transform via butterfly algorithm: SIAM Journal on Scientific Computing, **31**, 1678–1694.

Seismic data interpolation beyond aliasing using regularized nonstationary autoregression

Yang Liu[*], *Sergey Fomel*[†*]

ABSTRACT

Seismic data are often inadequately or irregularly sampled along spatial axes. Irregular sampling can produce artifacts in seismic imaging results. We present a new approach to interpolate aliased seismic data based on adaptive prediction-error filtering (PEF) and regularized nonstationary autoregression. Instead of cutting data into overlapping windows (patching), a popular method for handling nonstationarity, we obtain smoothly nonstationary PEF coefficients by solving a global regularized least-squares problem. We employ shaping regularization to control the smoothness of adaptive PEFs. Finding the interpolated traces can be treated as another linear least-squares problem, which solves for data values rather than filter coefficients. Compared with existing methods, the advantages of the proposed method include an intuitive selection of regularization parameters and fast iteration convergence. Benchmark synthetic and field data examples show that the proposed technique can successfully reconstruct data with decimated or missing traces.

INTRODUCTION

The regular and fine sampling along the time axis is common, whereas good spatial sampling is often more expensive or prohibitive and therefore is the main bottleneck for seismic resolution. Too large a spatial sampling interval may lead to aliasing problems that adversely affect the resolution of subsurface images. An alternative to expensive dense spatial sampling is interpolation of seismic traces. One important approach to trace interpolation is prediction interpolating methods (23), which use low-frequency non-aliased data to extract antialiasing prediction-error filters (PEFs) and then interpolates high frequencies beyond aliasing. (3) extends Spitz's method using PEFs in the t-x domain. (22) proposes a half-step PEF scheme that makes the interpolation process more efficient. (18) and (25) extend f-x trace interpolation to higher spatial dimensions. (16) introduces an algorithm similar to f-x prediction filtering, which has an elegant representation in the f-k domain. (10) uses multi-dimensional nonstationary PEFs to interpolate diffracted multiples. (19) propose an adaptive f-x interpolation using exponentially weighted recursive least squares. More recently, (20) propose a prediction approach similar to Gulunay's method but using

[*]**e-mail:** sergey.fomel@beg.utexas.edu

the curvelet transform instead of the Fourier transform. (2) compare the performance of several different interpolation methods.

Correcting irregular spatial sampling is another application for seismic data interpolation algorithms. A variety of interpolation methods have been published in the recent years. One approach is to estimate the PEF on multiple rescaled copies of the irregular data (8), where the data are rescaled with a number of progressively-larger bin sizes. (9) further improves the rescaling method by introducing multiple scales of the data where the location of the grid cells are varied in addition to the size of the cells. (11) use pseudo-primary data by crosscorrelating multiples and primaries to estimate nonstationary PEF and then interpolated missing near offsets. (21) propose autoregressive spectral estimates to reconstruct aliased data and data with gaps.

Seismic data are nonstationary. The standard PEF is designed under the assumption of stationary data and becomes less effective when this assumption is violated (3). Cutting data into overlapping windows (patching) is a common method to handle nonstationarity (5), although it occasionally fails in the presence of variable dips. (7) propose smoothly-varying nonstationary PEFs with "micropatches" and radial smoothing, which typically produces better results than the rectangular patching approach. (13) develops a nonstationary plane-wave destruction (PWD) filter as an alternative to t-x PEF (3) and applies the PWD operator to trace interpolation. The PWD method depends on the assumption of a small number of smoothly variable seismic dips. (8) uses Laplacian and radial rougheners to ensure a nonstationary PEF that varies smoothly in space, which specifies an appropriate regularization operator.

In this paper, we use the two-step strategy, similar to that of (3) and (7), but calculate the adaptive PEF by using regularized nonstationary autoregression (15) to handle both nonstationarity and aliasing. The key idea is the use of shaping regularization (14) to constrain the spatial smoothness of filter coefficients. We provide an approach to nonstationary data interpolation, which has an intuitive selection of parameters and fast iteration convergence. We test the new method by using several benchmark synthetic examples. Results of applying the proposed method to a field data example demonstrate that it can be effective in trace interpolation problems, even in the presence of multiple strongly variable slopes.

THEORY

A common constraint for interpolating missing seismic traces is to ensure that the interpolated data, after specified filtering, have minimum energy (3). Filtering is equivalent to spectral multiplication. Therefore, specified filtering is a way of prescribing a spectrum for the interpolated data. A sensible choice is a spectrum of the recorded data, which can be captured by finding the data's PEF (23; 6). The PEF, also known as the autoregression filter, plays the role of the 'inverse-covariance matrix' in statistical estimation theory. A signal is regressed on itself in the estimation of PEF. The PEF can be implemented in either t-x (time-space) or f-x (frequency-space)

domain. Time-space PEFs are less likely to create spurious events in the presence of noise than f-x PEFs (1; 6). When data interpolation is cast as an inverse problem, a PEF can be used to find missing data. This involves a two-step approach. In the first step, a PEF is estimated by minimizing the output of convolution of known data with an unknown PEF. In the second step, the missing data is found by minimizing the convolution of the recently calculated PEF with the unknown model, which is constrained where the data are known (9).

Step 1: Adaptive PEF estimation

Regular trace interpolation

An important property of PEFs is scale invariance, which allows estimation of PEF coefficients A_n (including the leading "-1" and prediction coefficients B_n) for incomplete aliased data $S(t,x)$ that include known traces $S_{known}(t,x_k)$ and unknown or zero traces $S_{zero}(t,x_z)$. For trace decimation, zero traces interlace known traces. To avoid zeroes that influence filter estimation, we interlace the filter coefficients with zeroes. For example, consider a 2-D PEF with seven prediction coefficients:

$$\begin{array}{ccccc} B_3 & B_4 & B_5 & B_6 & B_7 \\ \cdot & \cdot & -1 & B_1 & B_2 \end{array} \tag{1}$$

Here, the horizontal axis is time, the vertical axis is space, and "$\cdot$" denotes zero. Rescaling both time and spatial axes assumes that the dips represented by the original filter in equation 1 are the same as those represented by the scaled filter (3):

$$\begin{array}{ccccccccc} B_3 & \cdot & B_4 & \cdot & B_5 & \cdot & B_6 & \cdot & B_7 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & -1 & \cdot & B_1 & \cdot & B_2 \end{array} \tag{2}$$

For nonstationary situations, we can also assume locally stationary spectra of the data because trace decimation makes the space between known traces small enough, thus making adaptive PEFs locally scale-invariant. For estimating adaptive PEF coefficients, nonstationary autoregression allows coefficients B_n to change with both t and x. The new adaptive filter can look something like

$$\begin{array}{ccccccccc} B_3(t,x) & \cdot & B_4(t,x) & \cdot & B_5(t,x) & \cdot & B_6(t,x) & \cdot & B_7(t,x) \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & -1 & \cdot & B_1(t,x) & \cdot & B_2(t,x) \end{array} \tag{3}$$

In other words, prediction coefficients $B_n(t,x)$ are obtained by solving the least-squares problem,

$$\begin{aligned} \widehat{B_n}(t,x) &= \arg\min_{B_n} \|S(t,x) - \sum_{n=1}^{N} B_n(t,x) S_n(t,x)\|_2^2 \\ &\quad + \epsilon^2 \sum_{n=1}^{N} \|\mathbf{D}[B_n(t,x)]\|_2^2 \,, \end{aligned} \tag{4}$$

where $S_n(t,x)$=$S(t-m\,i\,\Delta\,t, x-m\,j\,\Delta\,x)$, which represents the causal translation of $S(t,x)$, with time-shift index i and spatial-shift index j scaled by decimation interval m. Note that predefined constant m uses the interlacing value as an interval; i.e., the shift interval equals 2 in equation 3. Subscript n is the general shift index for both time and space, and the total number of i and j is N. $\mathbf{D}$ is the regularization operator, and ϵ is a scalar regularization parameter. All coefficients $B_n(t,x)$ are estimated simultaneously in a time/space variant manner. This approach was described by (15) as regularized nonstationary autoregression (RNA). If $\mathbf{D}$ is a linear operator, least-squares estimation reduces to linear inversion

$$\mathbf{b} = \mathbf{A}^{-1}\,\mathbf{d}\;, \tag{5}$$

where

$$\mathbf{b} = \left[\begin{array}{cccc} B_1(t,x) & B_2(t,x) & \cdots & B_N(t,x) \end{array}\right]^T\;, \tag{6}$$

$$\mathbf{d} = \left[\begin{array}{cccc} S_1(t,x)\,S(t,x) & S_2(t,x)\,S(t,x) & \cdots & S_N(t,x)\,S(t,x) \end{array}\right]^T\;, \tag{7}$$

and the elements of matrix $\mathbf{A}$ are

$$A_{nk}(t,x) = S_n(t,x)\,S_k(t,x) + \epsilon^2\,\delta_{nk}\,\mathbf{D}^T\,\mathbf{D}\;. \tag{8}$$

Shaping regularization (14) incorporates a shaping (smoothing) operator $\mathbf{G}$ instead of $\mathbf{D}$ and provides better numerical properties than Tikhonov's regularization (24) in equation 2 (15). Inversion using shaping regularization takes the form

$$\mathbf{b} = \widehat{\mathbf{A}}^{-1}\,\widehat{\mathbf{d}}\;, \tag{9}$$

where

$$\widehat{\mathbf{d}} = \left[\begin{array}{cccc} \mathbf{G}\left[S_1(t,x)\,S(t,x)\right] & \mathbf{G}\left[S_2(t,x)\,S(t,x)\right] & \cdots & \mathbf{G}\left[S_N(t,x)\,S(t,x)\right] \end{array}\right]^T\;, \tag{10}$$

the elements of matrix $\widehat{\mathbf{A}}$ are

$$\widehat{A}_{nk}(t,x) \;=\; \lambda^2\,\delta_{nk} + \mathbf{G}\left[S_n(t,x)\,S_k(t,x) - \lambda^2\,\delta_{nk}\right]\;, \tag{11}$$

and λ is a scaling coefficient. One advantage of the shaping approach is the relative ease of controlling the selection of λ and $\mathbf{G}$ in comparison with ϵ and $\mathbf{D}$. We define $\mathbf{G}$ as Gaussian smoothing with an adjustable radius, which is designed by repeated application of triangle smoothing (14), and choose λ to be the mean value of $S_n(t,x)$.

Coefficients $B_n(t,x_z)$ at zero traces $S_{zero}(t,x_z)$ get constrained (effectively smoothly interpolated) by regularization. The required parameters are the size and shape of the filter $B_n(t,x)$ and the smoothing radius in $\mathbf{G}$.

Missing data interpolation

Irregular gaps occur in the recorded data for many different reasons, and prediction-error filters are known to be a powerful method for interpolating missing data. Missing data interpolation is a particular case of data regularization, where the input data are already given on a regular grid, and one needs to reconstruct only the missing values in empty bins (12). One can use existing traces to directly estimate adaptive PEF coefficients instead of scaling the filter as in regular trace interpolation problem. However, finding the adaptive PEF needs to avoid using any regression equations that involve boundaries or missing data. This can be achieved by creating selection mask operator $K(t,x)$, a diagonal matrix with ones at the known data locations and zeros elsewhere, for both causal translations and input data (5).

Analogously to the stationary prediction-error filter (1), adaptive PEF coefficients $B_n(t,x)$ use the unscaled format and appear as

$$\begin{array}{ccccc} B_3(t,x) & B_4(t,x) & B_5(t,x) & B_6(t,x) & B_7(t,x) \\ \cdot & \cdot & -1 & B_1(t,x) & B_2(t,x) \end{array} \tag{12}$$

The nonstationary coefficients $B_n(t,x)$ can be obtained by solving the least-squares problem

$$\begin{aligned} \widehat{B_n}(t,x) &= \arg\min_{B_n} \|K(t,x)[S(t,x) - \sum_{n=1}^{N} B_n(t,x) S_n(t,x)]\|_2^2 \\ &\quad + \epsilon^2 \sum_{n=1}^{N} \|\mathbf{D}[B_n(t,x)]\|_2^2 \ . \end{aligned} \tag{13}$$

where $S_n(t,x) = S(t-i, x-j)$. By using shaping regularization, adaptive PEF coefficients are smoothly filled at missing trace locations.

Step 2: Data interpolation with adaptive PEF

In the second step, a similar problem is solved, except that the filter is known, and the missing traces are unknown. In the decimated-trace interpolation problem, we squeeze (by throwing away alternate zeroed rows and columns) the filter in equation 3 to its original size and then formulate the least-squares problem,

$$\widehat{S}(t,x) = \arg\min_{S} \|S(t,x) - \sum_{n=1}^{N} \widehat{B_n}(t,x) S_n(t,x)\|_2^2 \ , \tag{14}$$

subject to

$$\widehat{S}(t,x_k) = S_{known}(t,x_k) \ , \tag{15}$$

where $\widehat{S}(t,x)$ represents the interpolated output, and i and j use the original shift as the interval; i.e., the shift interval equals 1.

We carry out the minimization in equations 2, 13, and 14 by the conjugate gradient method (17). The constraint condition (equation 15) is used as the initial model and constrains the output by using the known traces for each iteration in the conjugate-gradient scheme. The computational cost is proportional to $N_{iter} \times N_f \times N_t \times N_x$, where N_{iter} is the number of iterations, N_f is the filter size, and $N_t \times N_x$ is the data size. In our tests, N_f and N_{iter} were approximately equal to 100. Increasing the smoothing radius in shaping regularization decreases N_{iter} in the filter estimation step.

SYNTHETIC DATA TESTS

Aliasing decimated-trace interpolation test

We start with a strongly aliased synthetic example from (4). The sparse spatial sampling makes the gather severely aliased, especially at the far offset positions (Figure 1a). For comparison, we used PWD (13) to interpolate the traces (Figure 1b). Interpolation with PWD depends on dip estimation. In this example, the true dip is non-negative everywhere and is easily distinguished from the aliased one. Therefore, the PWD method recovers the interpolated traces well. However, in the more general case, an additional interpretation may be required to determine which of the dip components is contaminated by aliasing. According to the theory described in the previous section, the PEF-based methods use the lower (less aliased) frequencies to estimate PEF coefficients, and then interpolate the decimated traces (high-frequency information) by minimizing the convolution of the scale-invariant PEF with the unknown model, which is constrained where the data is known. We designed adaptive PEFs using 10 (time) $\times$ 2 (space) coefficients for each sample and a 50-sample (time) $\times$ 2-sample (space) smoothing radius and then applied them so as to interpolate the aliased trace. The nonstationary autoregression algorithm effectively removes all spatial aliasing artifacts (Figure 1c). The proposed method compares well with the PWD method. The CPU times, for single 2.66GHz CPU used in this example, are 20 seconds for adaptive PEF estimation (step 1) and 2 seconds for data interpolation (step 2).

Abma decimated-trace interpolation tests

A benchmark example created by Raymond Abma (personal communication) shows a simple curved event (Figure 2a). The challenge in this example is to account for both nonstationarity and aliasing. Figure 2b shows the interpolated result using Claerbout's stationary t-x PEF, which was estimated and applied in one big window, with each PEF coefficient B_n constant at every data location. Note that the t-x PEF method can recover the aliasing trace only in the dominant slope range. The trace-interpolating result using regularized nonstationary autoregression is shown in Figure 2c. The adaptive PEF has 20 (time) $\times$ 3 (space) coefficients for each sample

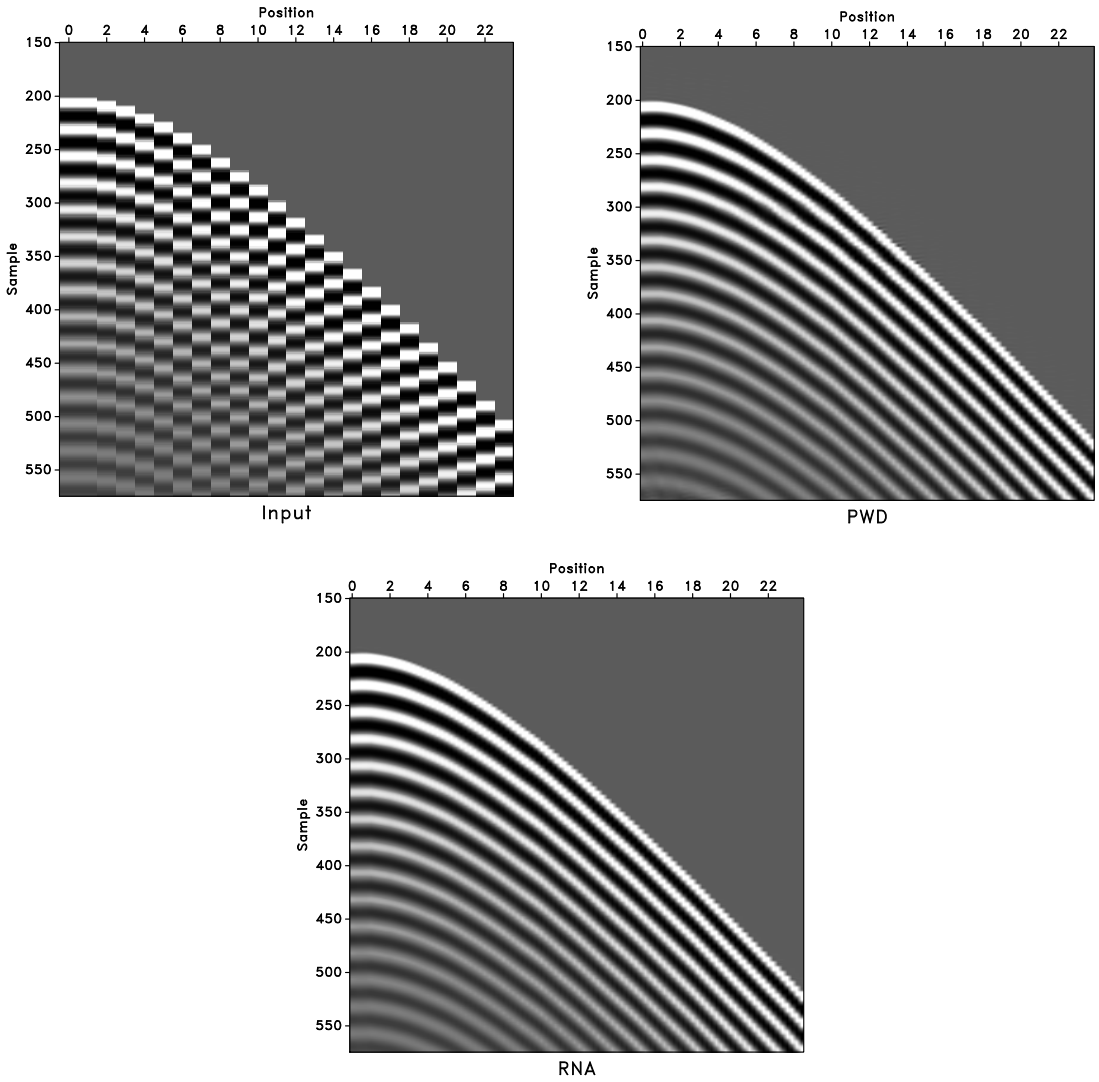


Figure 1: Aliased synthetic data (a), trace interpolation with plane-wave destruction (b), and trace interpolation with regularized nonstationary autoregression (c). Three additional traces were inserted between each of the neighboring input traces. apefint/alias jaliasp,dealias,jamiss

and a 20-sample (time) × 3-sample (space) smoothing radius. The proposed method eliminates all nonstationary aliasing and improves the continuity of the curved event.

(2) present a comparison of several algorithms used for trace interpolation. We chose the most challenging benchmark Marmousi example from Abma and Kabir to illustrate the performance of RNA interpolation. Figure 3a shows a zero-offset section of the Marmousi model, in which curved events violate the assumptions common for most trace-interpolating methods. Figure 3b shows that our method produces reasonable results for both curved and weak events and does not introduce any undesirable noise. The adaptive PEF parameters correspond to 7 (time) × 5 (space) coefficients for each sample and a 40-sample (time) × 30-sample (space) smoothing radius.

Missing-trace interpolation test

A missing trace test is shown in Figure 4a and comes from decimated-trace interpolation result (Figure 2c) after removing 70% of randomly selected traces. The curved event makes it difficult to recover the missing traces. The interpolated result is shown in Figure 4b, which uses a regularized adaptive PEF with 4 (time) × 2 (space) coefficients for each sample and a 50-sample (time) × 10-sample (space) smoothing radius. In the interpolated result, it is visually difficult to distinguish the missing trace locations, which is an evidence of successful interpolation. The filter size along space direction needs to be small in order to generate enough regression equations.

FIELD DATA EXAMPLES

We use a set of marine 2-D shot gathers from a deepwater Gulf of Mexico survey (7; 13) to further test the proposed method. Figure 5 shows the data before and after subsampling in the offset direction. The shot gather has long-period multiples and complicated diffraction events caused by a salt body. Amplitudes of the events are not uniformly distributed. Subsampling by a factor of 2 (Figure 5b) causes visible aliasing of the steeply dipping events. We designed a nonstationary PEF, with 15 (time) × 5 (space) coefficients for each sample and a 50-sample (time) × 20-sample (space) smoothing radius to handle the variability of events. Figure 6 shows the interpolation result and the difference between interpolated traces and original traces plotted at the same clip value. The proposed method succeeds in the sense that it is hard to distinguish interpolated traces from the interpolation result alone. A close-up comparison between the original and interpolated traces (Figure 7) shows some small imperfections. Some energy of the steepest events is partly missing. Coefficients of the adaptive PEF are illustrated in Figure 8, which displays the first coefficient (B_1) and the mean coefficient of B_n, respectively. The filter coefficients vary in time and space according to the curved events. The interpolated results are relatively insensitive to the smoothing parameters.

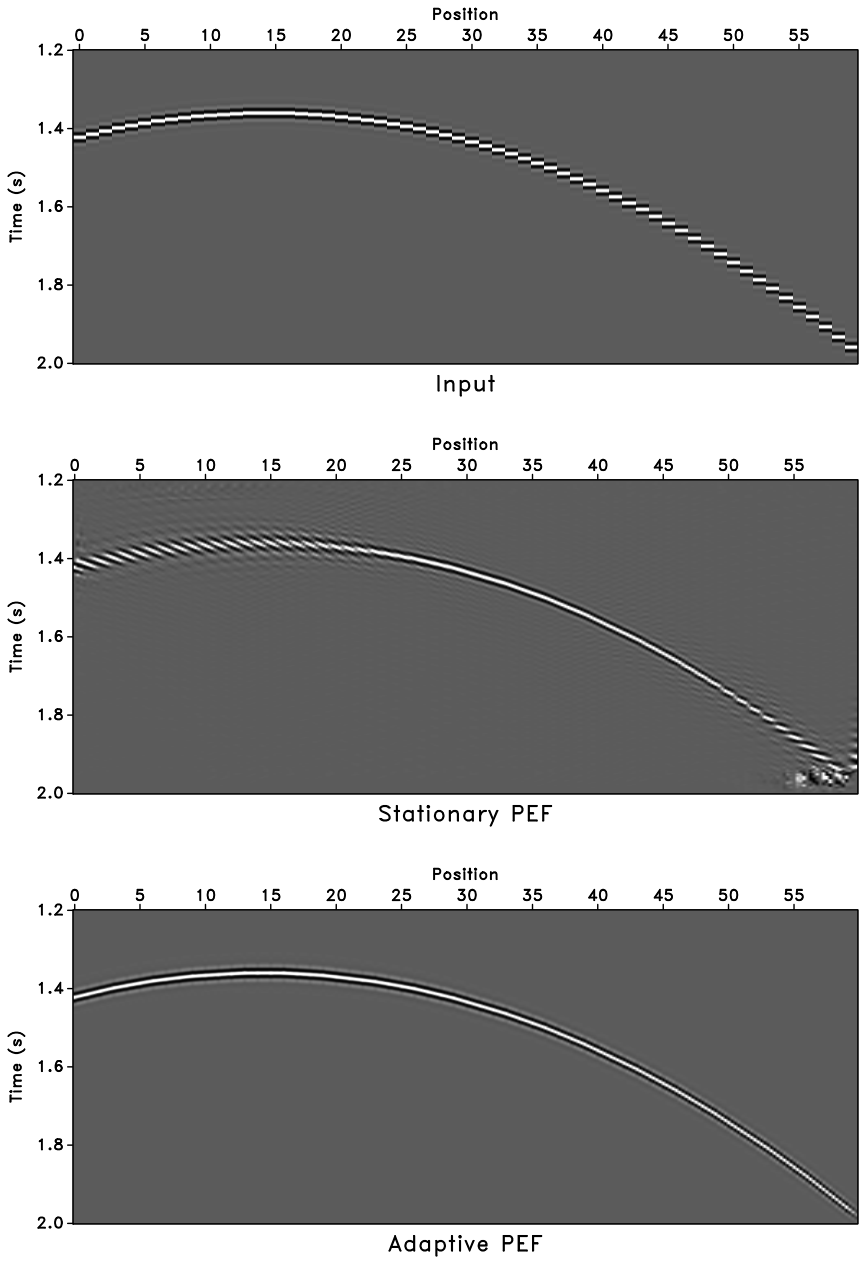


Figure 2: Curve model (a), trace interpolation with stationary PEF (b), and trace interpolation with adaptive PEF (c). apefint/ray jcurve,jcscov,jcacov

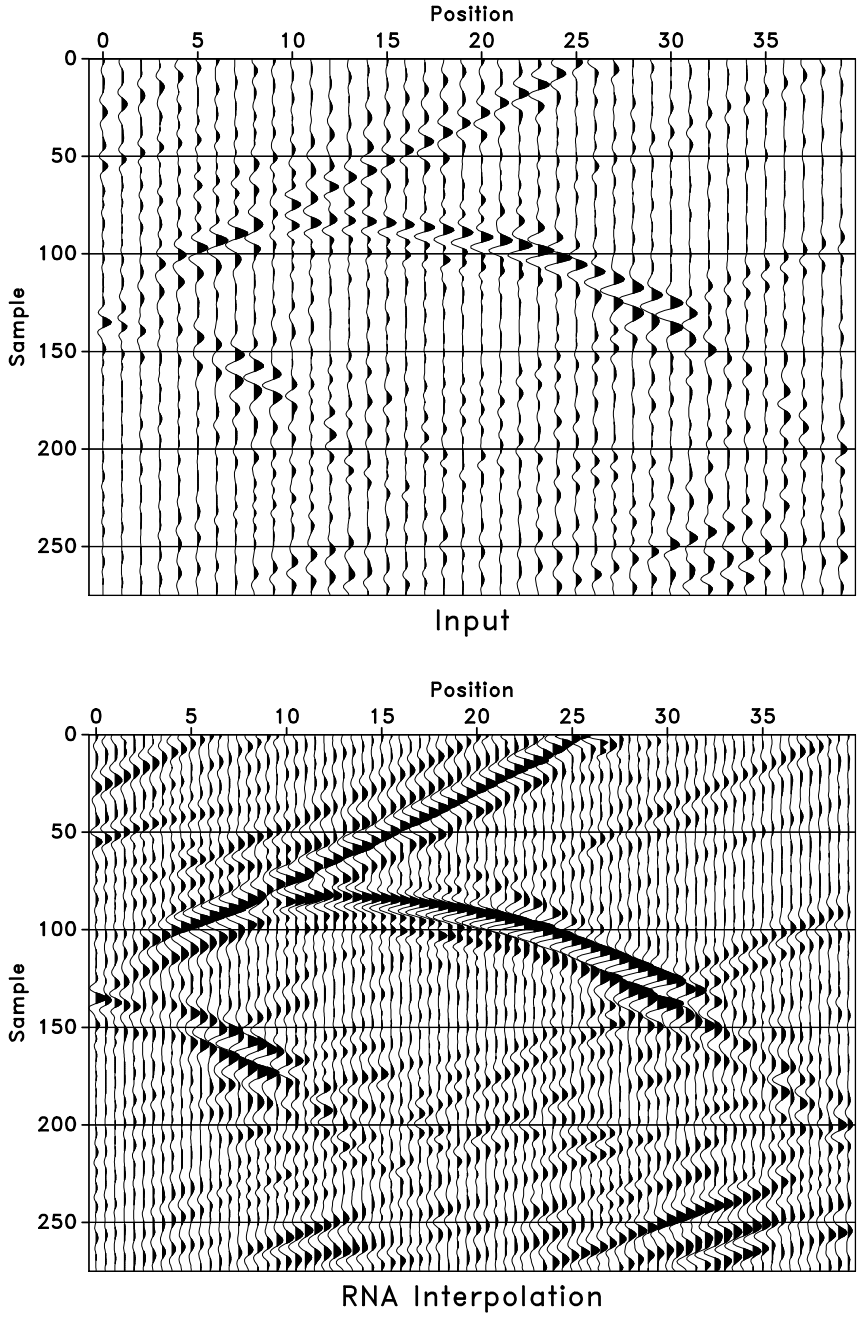


Figure 3: Marmousi model (a) and trace interpolation with regularized nonstationary autoregression (b). apefint/ray jmarm,jmacov

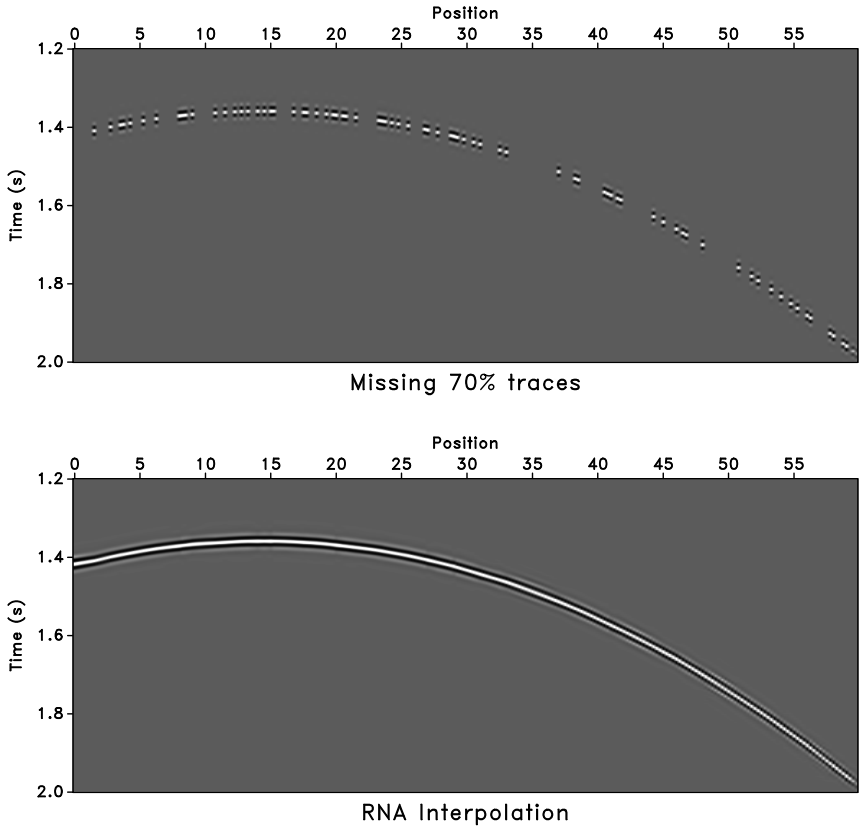


Figure 4: Curved model (Figure 2c) with 70% randomly selected traces removed (a) and trace interpolation with regularized nonstationary autoregression (b). apefint/misscurv zero,jamiss

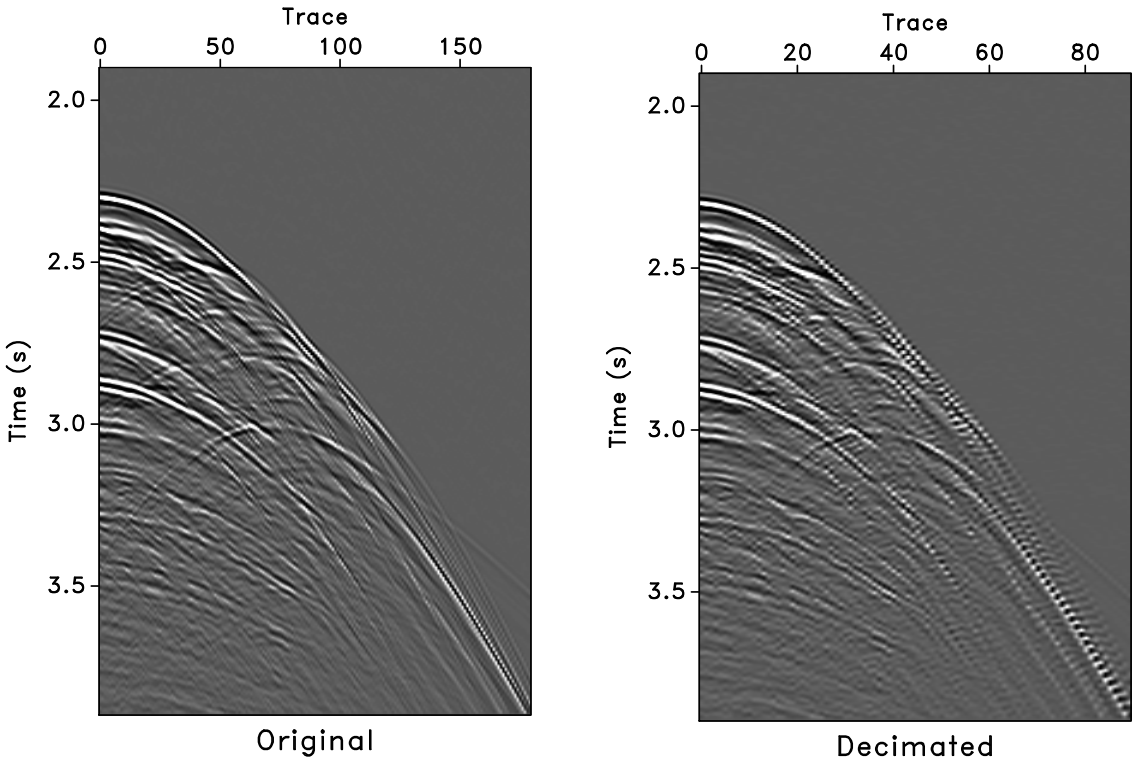


Figure 5: A 2-D marine shot gather. Original input (a) and input subsampled by a factor of 2 (b). apefint/sean sean,sean2

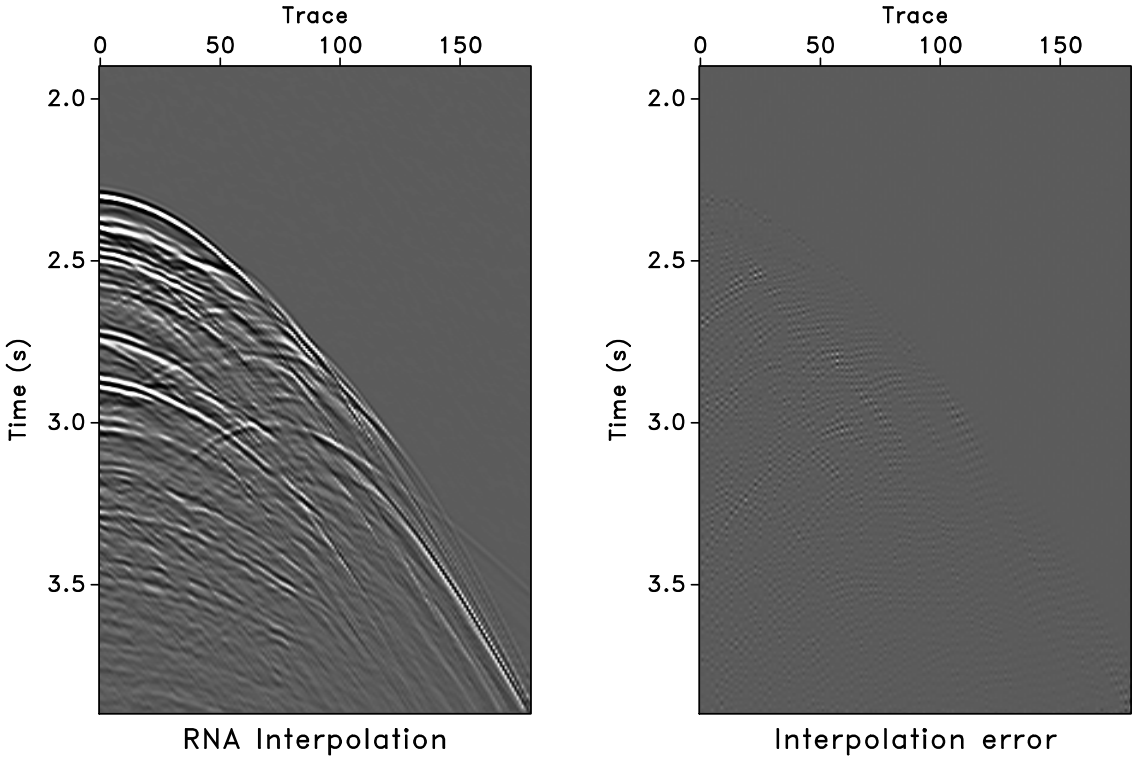


Figure 6: Shot gather after trace interpolation (adaptive PEF with 15 $\times$ 5) (a) and difference between original gather (Figure 5a) and interpolated result (Figure 6a) (b).
apefint/sean samiss,serr

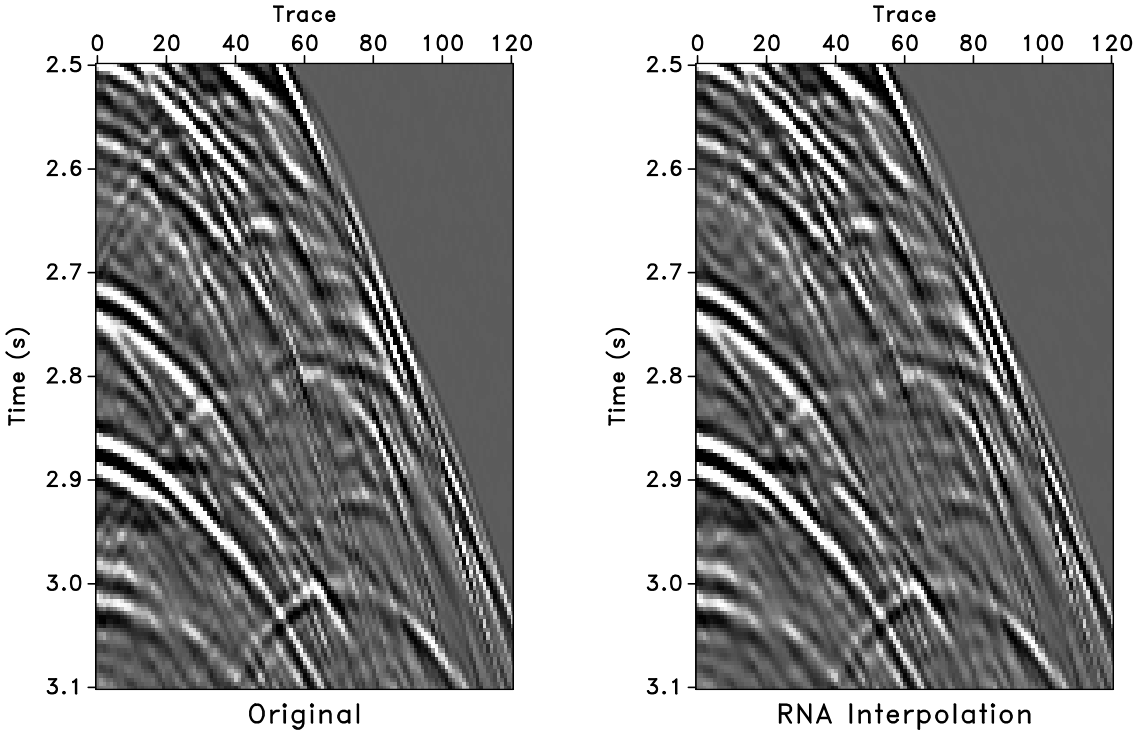


Figure 7: Close-up comparison of original data (a) and interpolated result by RNA (b). apefint/sean win,win2

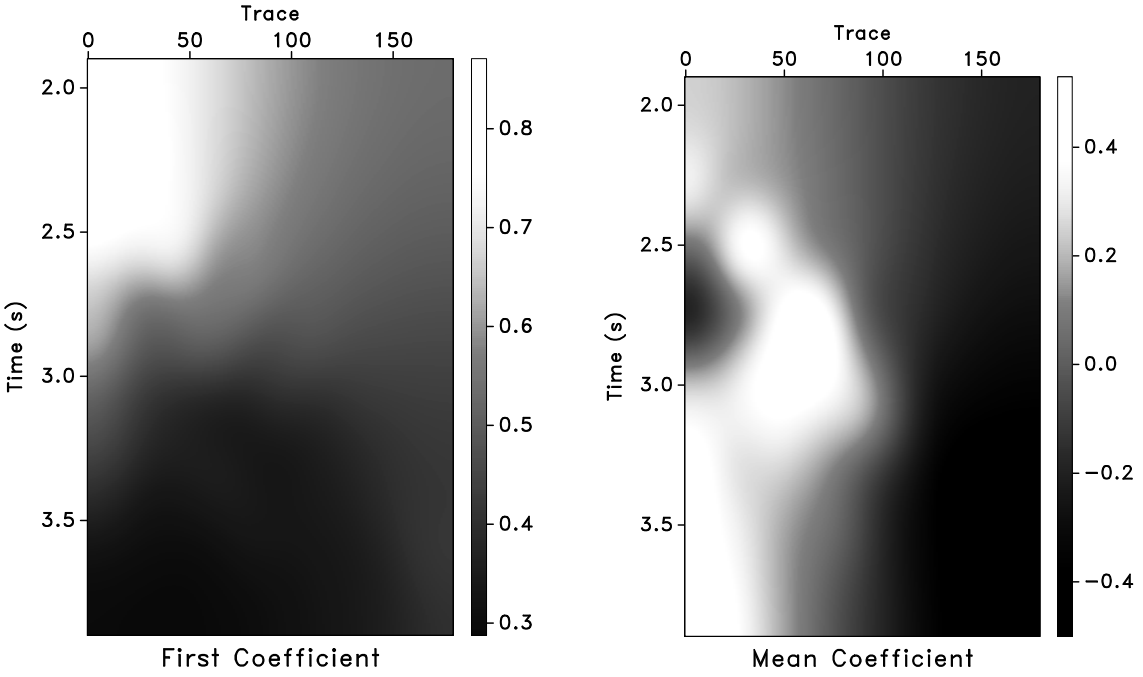


Figure 8: Adaptive PEF coefficients. First coefficient B_1 (a) and mean coefficient of B_n (b). apefint/sean sfcoe,smcoe

For a missing-trace interpolation test (Figure 9a), we removed 40% of randomly selected traces from the input data (Figure 5a). Furthermore, the first five traces were also removed to simulate traces missing at near offset. The adaptive PEF can only use a small number of coefficients in the spatial direction because of a small number of fitting equations (where the adaptive PEF lies entirely on known data). However, it also limits the ability of the proposed method to interpolate dipping events. We used a nonstationary PEF with 4 (time) $\times$ 3 (space) coefficients for each sample and a 50-sample (time) $\times$ 10-sample (space) smoothing radius to handle the missing trace recovery. The result is shown in Figure 9b. By comparing the results with the original input (Figure 5a), the missing traces are interpolated reasonably well except for weaker amplitude of the steeply dipping events.

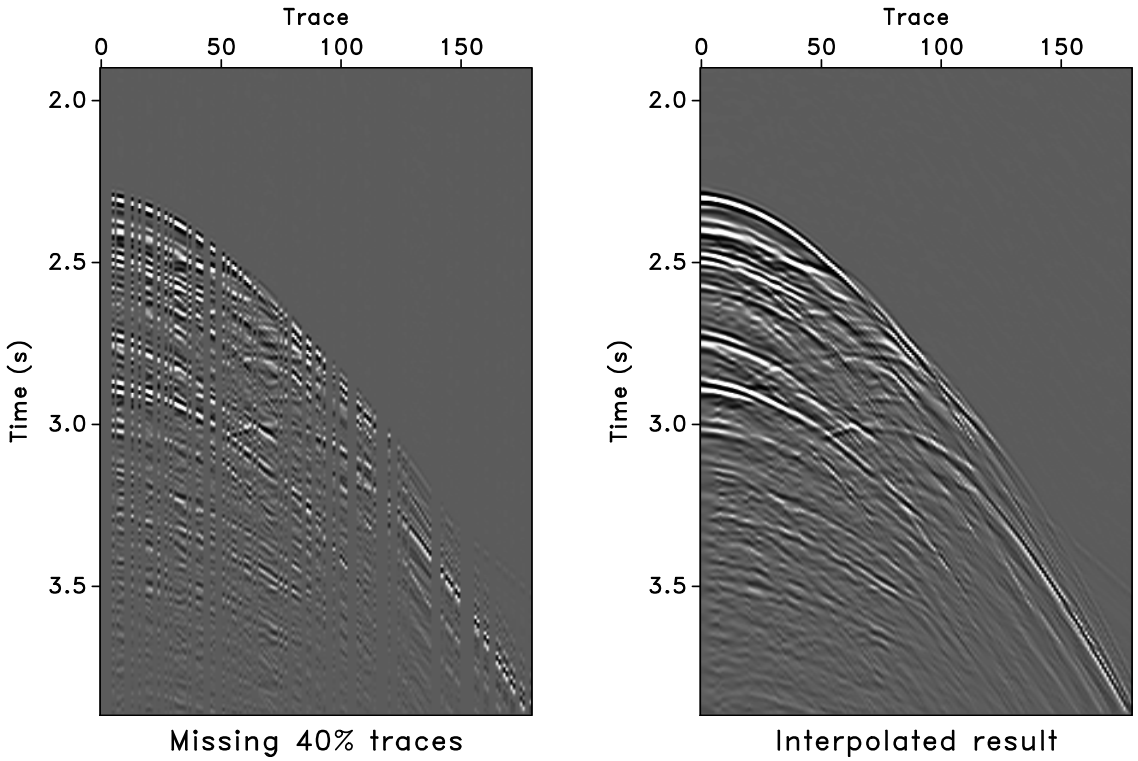


Figure 9: Field data with 40% randomly missing traces (a), and reconstructed data using RNA (b). apefint/missing zero,ramiss

An extension of the method to 3-D is straightforward and follows a two-step least-squares method with 3-D adaptive PEF estimation. We use a set of shot gathers as the input data volume to further test our method (Figure 10a). We removed 50% of randomly selected traces and five near offset traces for all shots (Figure 10b). For comparison, we used PWD to recover the missing traces (Figure 11a). The PWD method produces a reasonable result after carefully estimating dip information, but the interpolated error is slightly larger in the diffraction locations. (Figure 11b). The additional direction provided more information for interpolation but also increased the number of zeros in the mask operator $K(t, x)$, which constrains enough fitting

equations in equation 13. To use the available fitting equations for adaptive PEF estimation, we chose a smaller number of coefficients in the spatial direction. The proposed method is able to handle conflicting dips, although it does not appear to improve the dipping-event recovery compared to the 2-D case. This characteristic partly limits the application of RNA in 3-D case. We used a 3-D nonstationary PEF with 4 (time) $\times$ 2 (space) $\times$ 2 (space) coefficients for each sample and a 50-sample (time) $\times$ 10-sample (space) $\times$ 10-sample (space) smoothing radius was selected. Similar to the result in the 2-D example, Figure 11c shows the interpolation result, in which only steeply-dipping low-amplitude diffraction events with are lost (Figure 11d).

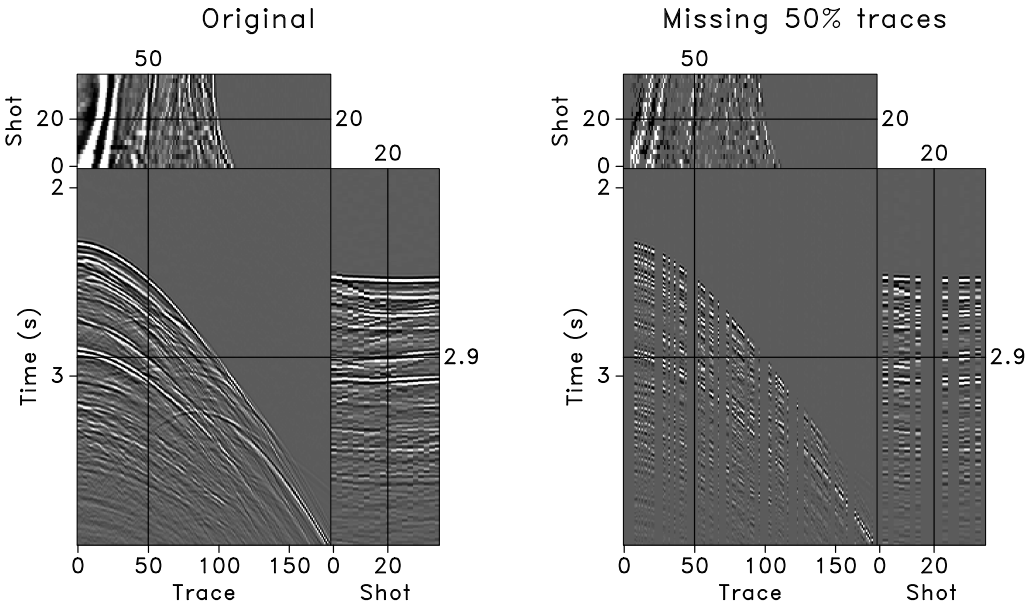


Figure 10: A 3-D field data volume (a) and data with 50% randomly missing traces (b). apefint/missing sean3,zero3

CONCLUSIONS

We have introduced a new approach to adaptive prediction-error filtering for seismic data interpolation. Our approach uses regularized nonstationary autoregression to handle time-space variation of nonstationary seismic data. We apply this method to interpolating seismic traces beyond aliasing and to reconstructing data with missing and decimated traces. Experiments with benchmark synthetic examples and field data tests show that the proposed filters can depict nonstationary signal variation and provide a useful description of complex wavefields having multiple curved events. These properties are useful for applications such as seismic data interpolation and regularization. Other possible applications may include seismic noise attenuation.

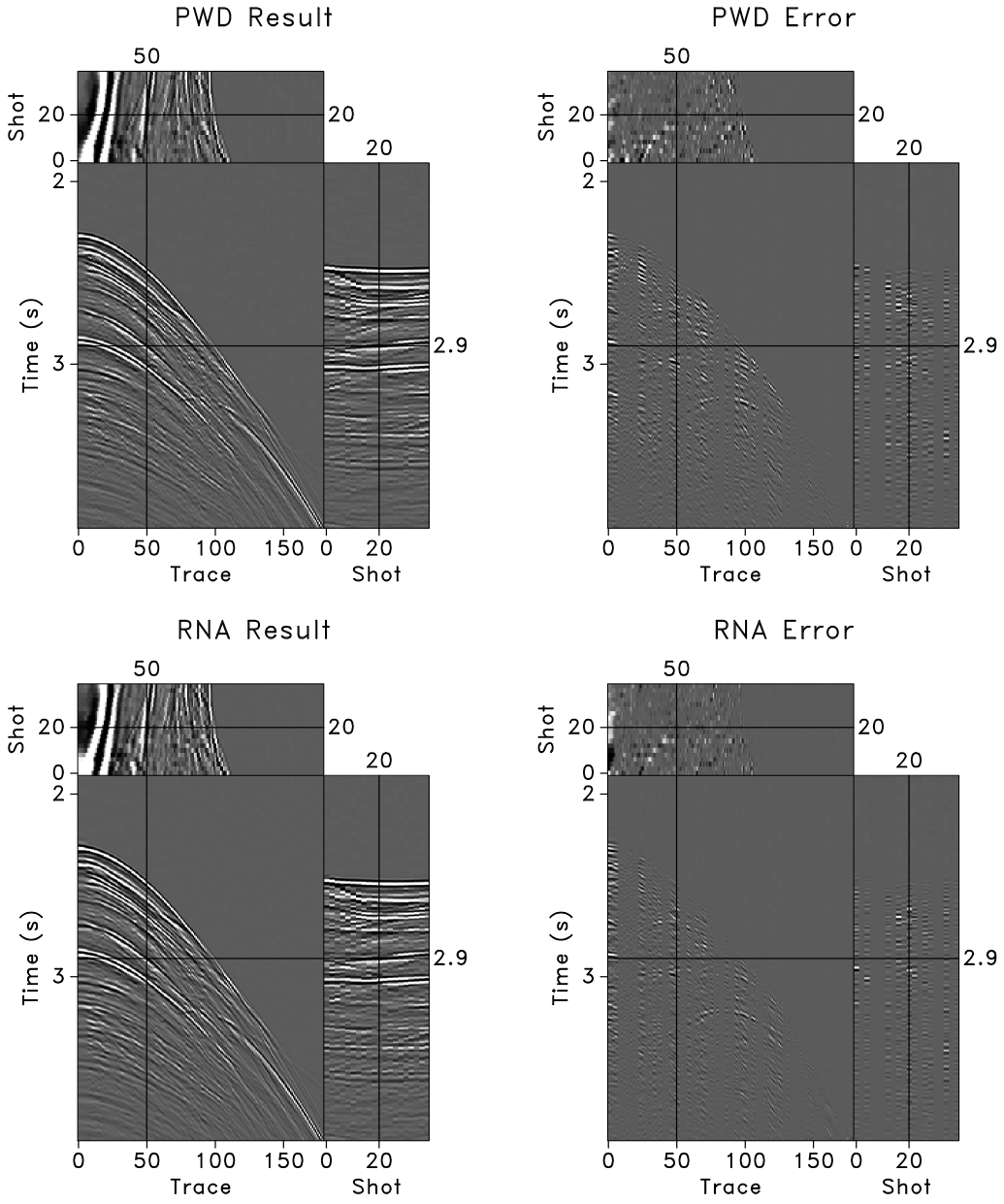


Figure 11: Reconstructed data volume using 3-D plane-wave destruction (a), difference between original input (Figure 10a) and interpolated result (Figure 11a), reconstructed data volume using 3-D regularized nonstationary autogression (c), and difference between original input (Figure 10a) and interpolated result (Figure 11c) (d). apefint/missing miss3,diff3,amiss3,adiff3

ACKNOWLEDGMENTS

We are grateful to Raymond Abma for providing benchmark data sets. We thank Eric Verschuur, Bill Curry, and two anonymous reviewers for helpful suggestions, which improved the quality of the paper.

This publication was authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Abma, R., 1995, Least-squares separation of signal and noise with multidimensional filters: PhD thesis, Stanford University.

[2] Abma, R., and N. Kabir, 2005, Comparisons of interpolation methods: The Leading Edge, **24**, 984–989.

[3] Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.

[4] ———, 2009, Basic earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

[5] ———, 2010, Image estimation by example: Geophysical soundings image construction - Multidimensional autoregression: Stanford Exploration Project, http://sepwww.stanford.edu/sep/prof/.

[6] Crawley, S., 2000, Seismic trace interpolation with nonstationary prediction-error filters: PhD thesis, Stanford University.

[7] Crawley, S., J. Claerbout, and R. Clapp, 1999, Interpolation with smoothly nonstationary prediction-error filters: 69th Annual International Meeting, SEG, Expanded Abstracts, 1154–1157.

[8] Curry, W., 2003, Interpolation of irregularly sampled data with nonstationary, multiscale prediction-error filters: 73th Annual International Meeting, SEG, Expanded Abstracts, 1913–1916.

[9] ———, 2004, Interpolation with multi-shifted-scale prediction-error filters: 74th Annual International Meeting, SEG, Expanded Abstracts, 2005–2008.

[10] ———, 2006, Interpolating diffracted multiples with prediction-error filters: 76th Annual International Meeting, SEG, Expanded Abstracts, 2709–2712.

[11] Curry, W., and G. Shan, 2008, Interpolation of near offsets using multiples and prediction-error filters: 78th Annual International Meeting, SEG, Expanded Abstracts, 2421–2424.

[12] Fomel, S., 2001, Three-dimensinal seismic data regularization: PhD thesis, Ph.D. thesis, Stanford University.

[13] ———, 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

[14] ———, 2007, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[15] ———, 2009, Adaptive multiple subtraction using regularized nonstationary regression: Geophysics, **74**, V25–V33.

[16] Gulunay, N., 2003, Seismic trace interpolation in the Fourier transform domain: Geophysics, **68**, 355–369.
[17] Hestenes, M. R., and E. Stiefel, 1952, Methods of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**, 409–436.
[18] Huard, I., S. Medina, and S. Spitz, 1996, F-XY wavefiedl de-aliasing for acquisition configurations leading to coarse sampling: 58th Annual International Meeting, EAGE, Expanded Abstracts, B039.
[19] Naghizadeh, M., and M. D. Sacchi, 2009, f-x adaptive seismic-trace interpolation: Geophysics, **74**, V9–V16.
[20] ———, 2010a, Beyond alias hierarchical scale curvelet interpolation of regularly and irregularly sampled seismic data: Geophysics, **75**, WB189–WB202.
[21] ———, 2010b, Robust reconstruction of aliased data using autoregressive spectral estimates: Geophysical Prospecting, **58**, 1049–1062.
[22] Porsani, M. J., 1999, Seismic trace interpolation using half-step prediction filters: Geophysics, **64**, 1461–1467.
[23] Spitz, S., 1991, Seismic trace interpolation in the F-X domain: Geophysics, **56**, 785–794.
[24] Tikhonov, A. N., 1963, Solution of Incorrectly Formulated Problems and the Regularization Method: Soviet Mathematics – Doklady.
[25] Wang, Y., 2002, Seismic trace interpolation in the f-x-y domain: Geophysics, **67**, 1232–1239.

Local seismic attributes

*Sergey Fomel**

ABSTRACT

Local seismic attributes measure seismic signal characteristics not instantaneously at each signal point and not globally across a data window but locally in the neighborhood of each point. I define local attributes with the help of regularized inversion and demonstrate their usefulness for measuring local frequencies of seismic signals and local similarity between different datasets. I use shaping regularization for controlling the locality and smoothness of local attributes. A multicomponent image registration example from a nine-component land survey illustrates practical applications of local attributes for measuring differences between registered images.

INTRODUCTION

Seismic attribute is defined by (13) as a "measurement derived from seismic data". Such a broad definition allows for many uses and abuses of the term. Countless attributes have been introduced in the practice of seismic exploration, (3; 4), which led (7) to talk about "attribute explosion". Many of these attributes play an exceptionally important role in interpreting and analyzing seismic data (5).

In this paper, I consider two particular attribute applications:

1. *Measuring local frequency content in a seismic image* is important both for studying the phenomenon of seismic wave attenuation and for processing of attenuated signals.

2. *Measuring local similarity between two seismic images* is useful for seismic monitoring, registration of multicomponent data, and analysis of velocities and amplitudes.

Some of the best known seismic attributes are instantaneous attributes such as instantaneous phase or instantaneous dip (14; 1; 2). Such attributes measure seismic frequency characteristics as being attached instantaneously to each signal point. This measure is notoriously noisy and may lead to unphysical values such as negative frequencies (15).

*e-mail: sergey.fomel@beg.utexas.edu

In this paper, I introduce a concept of *local attributes*. Local attributes measure signal characteristics not instantaneously at each data point but in a local neighborhood around the point. According to the Fourier uncertainty principle, frequency is essentially an uncertain characteristic when applied to a local region in the time domain. Therefore, local frequency is more physically meaningful than instantaneous frequency. The idea of locality extends from local frequency to other attributes, such as the correlation coefficient between two different datasets, that are conventionally evaluated in sliding windows.

The paper starts with reviewing the definition of instantaneous frequency. I modify this definition to that of local frequency by recognizing it as a form of regularized inversion and by changing regularization to constrain the continuity and smoothness of the output. The same idea is extended next to define local correlation . I illustrate a practical application of local attributes using an example from multicomponent seismic image registration in a nine-component land survey.

MEASURING LOCAL FREQUENCIES

Definition of instantaneous frequency

Let $f(t)$ represent seismic trace as a function of time t. The corresponding complex trace $c(t)$ is defined as

$$c(t) = f(t) + i\, h(t)\ , \tag{1}$$

where $h(t)$ is the Hilbert transform of the real trace $f(t)$. One can also represent the complex trace in terms of the envelope $A(t)$ and the instantaneous phase $\phi(t)$, as follows:

$$c(t) = A(t)\, e^{i\,\phi(t)}\ . \tag{2}$$

By definition, instantaneous frequency is the time derivative of the instantaneous phase (14)

$$\omega(t) = \phi'(t) = Im\left[\frac{c'(t)}{c(t)}\right] = \frac{f(t)\, h'(t) - f'(t)\, h(t)}{f^2(t) + h^2(t)}\ . \tag{3}$$

Different numerical realizations of equation 3 produce slightly different algorithms (1).

Note that the definition of instantaneous frequency calls for division of two signals. In a linear algebra notation,

$$\mathbf{w} = \mathbf{D}^{-1}\,\mathbf{n}\ , \tag{4}$$

where $\mathbf{w}$ represents the vector of instantaneous frequencies $\omega(t)$, $\mathbf{n}$ represents the numerator in equation 3, and $\mathbf{D}$ is a diagonal operator made from the denominator of equation 3. A recipe for avoiding division by zero is adding a small constant ϵ to the denominator (12). Consequently, equation 4 transforms to

$$\mathbf{w}_{inst} = \left(\mathbf{D} + \epsilon\,\mathbf{I}\right)^{-1}\,\mathbf{n}\ , \tag{5}$$

where **I** stands for the identity operator. Stabilization by ϵ does not, however, prevent instantaneous frequency from being a noisy and unstable attribute. The main reason for that is the extreme locality of the instantaneous frequency measurement, governed only by the phase shift between the signal and its Hilbert transform.

Figure 9(a) shows three test signals for comparing frequency attributes. The first signal is a synthetic chirp function with linearly varying frequency. Instantaneous frequency shown in Figure 1 correctly estimates the modeled frequency trend. The second signal is a piece of a synthetic seismic trace obtained by convolving a 40-Hz Ricker wavelet with synthetic reflectivity. The instantaneous frequency (Figure 1b) shows many variations and appears to contain detailed information. However, this information is useless for characterizing the dominant frequency content of the data, which remains unchanged due to stationarity of the seismic wavelet. The last test example (Figure 9(a)c) is a real trace extracted from a seismic image. The instantaneous frequency (Figure 1c) appears noisy and even contains physically unreasonable negative values. Similar behavior was described by (15).

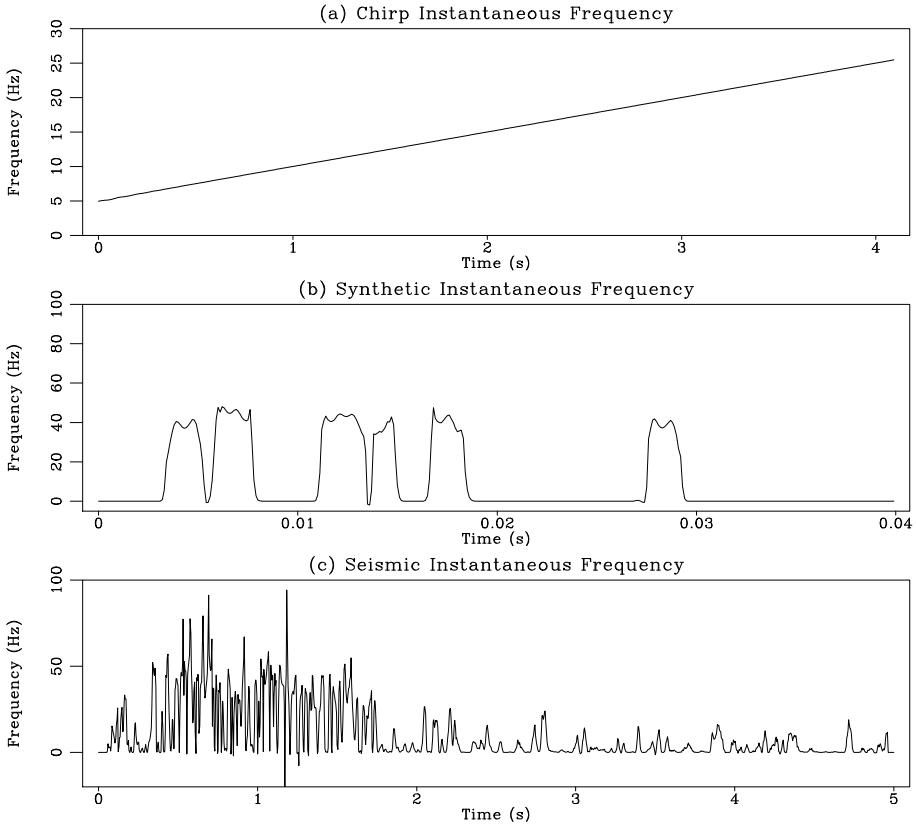


Figure 1: Instantaneous frequency of test signals from Figure 9(a). attr/attr inst

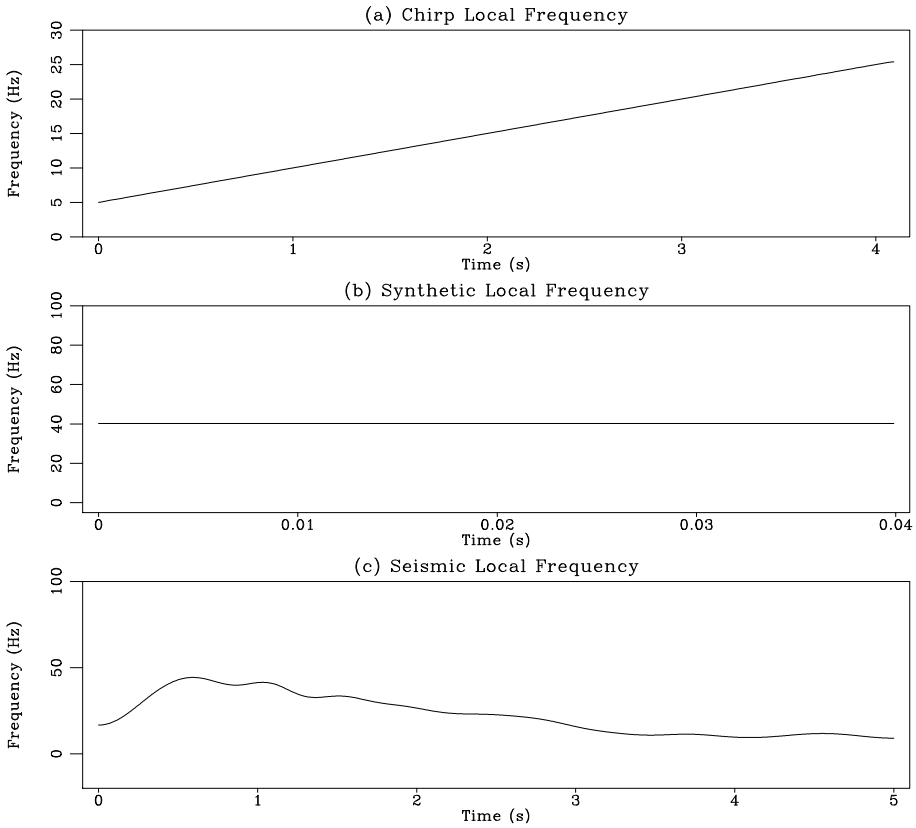


Figure 2: Local frequency of test signals from Figure 9(a). attr/attr locl

Definition of local frequency

The definition of the local frequency attribute starts by recognizing equation 5 as a regularized form of linear inversion. Changing regularization from simple identity to a more general regularization operator $\mathbf{R}$ provides the definition for local frequency as follows:

$$\mathbf{w}_{loc} = (\mathbf{D} + \epsilon\,\mathbf{R})^{-1}\ \mathbf{n}\ , \tag{6}$$

The role of the regularization operator is ensuring continuity and smoothness of the local frequency measure. A different approach to regularization follows from the shaping method (8). Shaping regularization operates with a smoothing (shaping) operator $\mathbf{S}$ by incorporating it into the inversion scheme as follows:

$$\mathbf{w}_{loc} = \left[\lambda^2\,\mathbf{I} + \mathbf{S}\ \left(\mathbf{D} - \lambda^2\,\mathbf{I}\right)\right]^{-1}\ \mathbf{S}\,\mathbf{n}\ , \tag{7}$$

Scaling by λ preserves physical dimensionality and enables fast convergence when inversion is implemented by an iterative method. A natural choice for λ is the least-squares norm of $\mathbf{D}$.

Figure 2 shows the results of measuring local frequency in the test signals from Figure 9(a). I used the shaping regularization formulation 7 with the shaping operator $\mathbf{S}$ defined as a triangle smoother. The chirp signal frequency (Figure 2a) is correctly recovered. The dominant frequency of the synthetic signal (Figure 2b) is correctly estimated to be stationary at 40 Hz. The local frequency of the real trace (Figure 2c) appears to vary with time according to the general frequency attenuation trend.

This example highlights some advantages of the local attribute method in comparison with the sliding window approach:

- Only one parameter (the smoothing radius) needs to be specified as opposed to several (window size, overlap, and taper) in the windowing approach. The smoothing radius directly reflects the locality of the measurement.
- The local attribute approach continues the measurement smoothly through the regions of absent information such as the zero amplitude regions in the synthetic example, where the signal phase is undefined. This effect is impossible to achieve in the windowing approach unless the window size is always larger than the information gaps in the signal.

Figure **??** shows seismic images from compressional (PP) and shear (SS) reflections obtained by processing a land nine-component survey. Figure **??** shows local frequencies measured in PP and SS images after warping the SS image into PP time. The term "image warping" comes from medical imaging (16) and refers, in this case, to squeezing the SS image to PP reflection time to make the two images display in the same coordinate system. We can observe a general decay of frequency with time caused by seismic attenuation. After mapping (squeezing) to PP time, the SS

image frequency appears higher in the shallow part of the image because of a relatively low S-wave velocity but lower in the deeper part of the image because of the apparently stronger attenuation of shear waves. A low-frequency anomaly in the PP image might be indicative of gas presence. Identifying and balancing non-stationary frequency variations of multicomponent images is an essential part of the multistep image registration technique (9; 10).

MEASURING LOCAL SIMILARITY

Consider the task of measuring similarity between two different signals $a(t)$ and $b(t)$. One can define similarity as a global correlation coefficient and then, perhaps, measure it in sliding windows across the signal. The local construction from the previous section suggests approaching this problem in a more elegant way.

Definition of global correlation

Global correlation coefficient between $a(t)$ and $b(t)$ can be defined as the functional

$$\gamma = \frac{< a(t), b(t) >}{\sqrt{< a(t), a(t) >< b(t), b(t) >}} \; , \tag{8}$$

where $< x(t), y(t) >$ denotes the dot product between two signals:

$$< x(t), y(t) >= \int x(t)\, y(t)\, dt \; .$$

According to definition 8, the correlation coefficient of two identical signals is equal to one, and the correlation of two signals with opposite polarity is minus one. In all the other cases, the correlation will be less then one in magnitude thanks to the Cauchy-Schwartz inequality.

The global measure 8 is inconvenient because it supplies only one number for the whole signal. The goal of local analysis is to turn the functional into an operator and to produce local correlation as a variable function $\gamma(t)$ that identifies local changes in the signal similarity.

Definition of local correlation

In a linear algebra notation, the squared correlation coefficient γ from equation 8 can be represented as a product of two least-squares inverses

$$\gamma^2 = \gamma_1 \, \gamma_2 \; , \tag{9}$$

$$\gamma_1 = \left(\mathbf{a}^T \, \mathbf{a}\right)^{-1} \left(\mathbf{a}^T \, \mathbf{b}\right) \; , \tag{10}$$

$$\gamma_2 = \left(\mathbf{b}^T \, \mathbf{b}\right)^{-1} \left(\mathbf{b}^T \, \mathbf{a}\right) \; , \tag{11}$$

where $\mathbf{a}$ is a vector notation for $a(t)$, $\mathbf{b}$ is a vector notation for $b(t)$, and $\mathbf{x}^T\,\mathbf{y}$ denotes the dot product operation. Let $\mathbf{A}$ be a diagonal operator composed from the elements of $\mathbf{a}$ and $\mathbf{B}$ be a diagonal operator composed from the elements of $\mathbf{b}$. Localizing equations 10-11 amounts to adding regularization to inversion. Scalars γ_1 and γ_2 turn into vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ defined, using shaping regularization , as

$$\mathbf{c}_1 = \left[\lambda^2\,\mathbf{I} + \mathbf{S}\,\left(\mathbf{A}^T\,\mathbf{A} - \lambda^2\,\mathbf{I}\right)\right]^{-1}\,\mathbf{S}\,\mathbf{A}^T\,\mathbf{b}\,, \tag{12}$$

$$\mathbf{c}_2 = \left[\lambda^2\,\mathbf{I} + \mathbf{S}\,\left(\mathbf{B}^T\,\mathbf{B} - \lambda^2\,\mathbf{I}\right)\right]^{-1}\,\mathbf{S}\,\mathbf{B}^T\,\mathbf{a}\,. \tag{13}$$

To define a local similarity measure, I apply the component-wise product of vectors $\mathbf{c}_1$ and $\mathbf{c}_2$. It is interesting to note that, if one applies an iterative conjugate-gradient inversion for computing the inverse operators in equations 7 and 8, the output of the first iteration will be the smoothed product of the two signals $\mathbf{c}_1 = \mathbf{c}_2 = \mathbf{S}\,\mathbf{A}^T\,\mathbf{b}$, which is equivalent, with an appropriate choice of $\mathbf{S}$, to the algorithm of fast local cross-correlation proposed by (11).

The local similarity attribute is useful for solving the problem of multicomponent image registration. After an initial registration using interpreter's "nails" (6) or velocities from seismic processing, a useful registration indicator is obtained by squeezing and stretching the warped shear-wave image while measuring its local similarity to the compressional image. Such a technique was named *residual γ scan* and proposed by (10). Figure **??** shows a residual scan for registration of multicomponent images from Figure **??**. Identifying and picking points of high local similarity enables multicomponent registration with high-resolution accuracy. The registration result is visualized in Figure **??**, which shows interleaved traces from PP and SS images before and after registration. The alignment of main seismic events is an indication of successful registration.

CONCLUSIONS

I have introduced a concept of local seismic attributes and specified it for such attributes as local frequency and local similarity. Local attributes measure signal characteristics not instantaneously at each signal point and not globally across a data window but locally in the neighborhood of each point. They find applications in different steps of multicomponent seismic image registration. One can extend the idea of local attributes to other applications.

ACKNOWLEDGMENTS

I would like to thank Bob Hardage, Milo Backus, and Bob Graebner for introducing me to the problem of multicomponent data registration and for many useful discussions. The data example was kindly provided by Vecta. This work was partially supported by the U.S. Department of Energy through Program DE-PS26-04NT42072

and is authorized for publication by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Barnes, A. E., 1992, The calculation of instantaneous frequency and instantaneous bandwidth (short note): Geophysics, **57**, 1520–1524.

[2] ———, 1993, Instantaneous spectral bandwidth and dominant frequency with applications to seismic reflection data: Geophysics, **58**, 419–428.

[3] Brown, A. R., 1996, Interpreter's corner - Seismic attributes and their classification: The Leading Edge, **15**, 1090.

[4] Chen, Q., and S. Sidney, 1997, Seismic attribute technology for reservoir forecasting and monitoring: The Leading Edge, **16**, 445–456.

[5] Chopra, S., and K. J. Marfurt, 2005, Seismic attributes – A historical perspective: Geophysics, **70**, 3SO–28SO.

[6] DeAngelo, M. V., R. Remington, P. Murray, B. A. Hardage, R. Graebner, and K. Fouad, 2004, Multicomponent seismic technology for imaging deep gas prospects: The Leading Edge, **23**, 1270–1281.

[7] Eastwood, J., 2002, The attribute explosion: The Leading Edge, **21**, 994.

[8] Fomel, S., 2006, Shaping regularization in geophysical estimation problems: Geophysics, accepted for publication.

[9] Fomel, S., and M. Backus, 2003, Multicomponent seismic data registration by least squares: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 781–784.

[10] Fomel, S., M. Backus, K. Fouad, B. Hardage, and G. Winters, 2005, A multistep approach to multicomponent seismic image registration with application to a West Texas carbonate reservoir study: 75th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1018–1021.

[11] Hale, D., 2006, Fast local cross-correlations of images: 76th Ann. Internat. Mtg, Soc. of Expl. Geophys., 3160–3163.

[12] Matheney, M. P., and R. L. Nowack, 1995, Seismic attenuation values obtained from instantaneous frequency matching and spectral ratios: Geohysical Journal International, **123**, 1–15.

[13] Sheriff, R. E., 1991, Encyclopedic Dictionary of Exploration Geophysics: Soc. of Expl. Geophys. (How it was done? - see TLE-10-4-41-42).

[14] Taner, M. T., F. Koehler, and R. E. Sheriff, 1979, Complex seismic trace analysis: Geophysics, **44**, 1041–1063. (Errata in GEO-44-11-1896; Discussion in GEO-45-12-1877-1878; Reply in GEO-45-12-1878-1878).

[15] White, R. E., 1991, Properties of instantaneous seismic attributes: The Leading Edge, **10**, 26–32. (Discussion and reply in TLE-11-8-45-46 and TLE-11-10-10-12).

[16] Wolberg, G., 1990, Digital image warping: IEEE Computer Society.

Shaping regularization in geophysical estimation problems

*Sergey Fomel**

ABSTRACT

Regularization is a required component of geophysical estimation problems that operate with insufficient data. The goal of regularization is to impose additional constraints on the estimated model. I introduce shaping regularization, a general method for imposing constraints by explicit mapping of the estimated model to the space of admissible models. Shaping regularization is integrated in a conjugate-gradient algorithm for iterative least-squares estimation. It provides the advantage of a better control on the estimated model in comparison with traditional regularization methods and, in some cases, leads to a faster iterative convergence. Simple data interpolation and seismic velocity estimation examples illustrate the concept[a].

[a]Presented at the SEG Annual Meeting in 2005

INTRODUCTION

A great number of geophysical estimation problems are mathematically ill-posed because they operate with insufficient data (16). Regularization is a technique for making the estimation problems well-posed by adding indirect constraints on the estimated model (8; 28). Developed originally by (22) and others, the method of regularization has become an indispensable part of the inverse problem theory and has found many applications in geophysical problems: traveltime tomography (1; 5), migration velocity analysis (27; 29), high-resolution Radon transform (23), spectral decomposition (18), etc.

While the main goal of inversion is to fit the observed data, Tikhonov's regularization adds another goal of fitting the estimated model to a priorly assumed behavior. The contradiction between the two goals often leads to a slow convergence of iterative estimation algorithms (14). The speed can be improved considerably by an appropriate model reparameterization or preconditioning (11). However, the difficult situation of trying to satisfy two contradictory goals simultaneously leads sometimes to an undesirable behavior of the solution at the early iterations of an iterative optimization scheme.

*__e-mail:__ sergey.fomel@beg.utexas.edu

In this paper, I introduce *shaping regularization*, a new general method of imposing regularization constraints. A shaping operator provides an explicit mapping of the model to the space of acceptable models. The operator is embedded in an iterative optimization scheme (the conjugate-gradient algorithm) and allows for a better control on the estimation result. Shaping into the space of smooth functions can be accomplished with efficient lowpass filtering. Depending on the desirable result, it is also possible to shape the model into a piecewise-smooth function, a function following geological structure, or a function representable in a predefined basis. I illustrate the shaping concept with simple numerical experiments of data interpolation and seismic velocity estimation.

REVIEW OF TIKHONOV'S REGULARIZATION

If the data are represented by vector $\mathbf{d}$, model parameters by vector $\mathbf{m}$, and their functional relationship is defined by the forward modeling operator $\mathbf{L}$, the least-squares optimization approach amounts to minimizing the least-squares norm of the residual difference $\mathbf{L}\,\mathbf{m}-\mathbf{d}$. In Tikhonov's regularization approach, one additionally attempts to minimize the norm of $\mathbf{D}\,\mathbf{m}$, where $\mathbf{D}$ is the *regularization* operator. Thus, we are looking for the model $\mathbf{m}$ that minimizes the least-squares norm of the compound vector $\left[\begin{array}{cc}\mathbf{Lm}-\mathbf{d} & \epsilon\mathbf{Dm}\end{array}\right]^T$, where ϵ is a scalar scaling parameter. The formal solution has the well-known form

$$\widehat{\mathbf{m}} = \left(\mathbf{L}^T\,\mathbf{L}+\epsilon^2\,\mathbf{D}^T\,\mathbf{D}\right)^{-1}\,\mathbf{L}^T\,\mathbf{d}\;, \tag{1}$$

where $\widehat{\mathbf{m}}$ denotes the least-squares estimate of $\mathbf{m}$, and $\mathbf{L}^T$ denotes the adjoint operator. One can carry out the optimization iteratively with the help of the conjugate-gradient method (15) or its analogs. Iterative methods have computational advantages in large-scale problems when forward and adjoint operators are represented by sparse matrices and can be computed efficiently (19; 25).

In an alternative approach, one obtains the regularized estimate by minimizing the least-squares norm of the compound vector $\left[\begin{array}{cc}\mathbf{p} & \mathbf{r}\end{array}\right]^T$ under the constraint

$$\epsilon\mathbf{r} = \mathbf{d}-\mathbf{Lm} = \mathbf{d}-\mathbf{LPp}\;. \tag{2}$$

Here $\mathbf{P}$ is the *model reparameterization* operator that translates vector $\mathbf{p}$ into the model vector $\mathbf{m}$, $\mathbf{r}$ is the scaled residual vector, and ϵ has the same meaning as before. The formal solution of the preconditioned problem is given by

$$\widehat{\mathbf{m}} = \mathbf{P}\,\widehat{\mathbf{p}} = \mathbf{P}\,\mathbf{P}^T\,\mathbf{L}^T\,\left(\mathbf{L}\,\mathbf{P}\,\mathbf{P}^T\,\mathbf{L}^T+\epsilon^2\,\mathbf{I}\right)^{-1}\,\mathbf{d}\;, \tag{3}$$

where $\mathbf{I}$ is the identity operator in the data space. Estimate 3 is mathematically equivalent to estimate 1 if $\mathbf{D}^T\,\mathbf{D}$ is invertible and

$$\left(\mathbf{D}^T\,\mathbf{D}\right)^{-1} = \mathbf{P}\,\mathbf{P}^T = \mathbf{C}\;. \tag{4}$$

Statistical theory of least-squares estimation connects $\mathbf{C}$ with the model covariance operator (21). In a more general case of reparameterization, the size of $\mathbf{p}$ may be different from the size of $\mathbf{m}$, and $\mathbf{C}$ may not have the full rank. In iterative methods, the preconditioned formulation often leads to faster convergence. (11) suggest constructing preconditioning operators in multi-dimensional problems by recursive helical filtering.

SMOOTHING BY REGULARIZATION

Let us consider an application of Tikhonov's regularization to one of the simplest possible estimation problems: smoothing. The task of smoothing is to find a model $\mathbf{m}$ that fits the observed data $\mathbf{d}$ but is in a certain sense smoother. In this case, the forward operator $\mathbf{L}$ is simply the identity operator, and the formal solutions 1 and 3 take the form

$$\widehat{\mathbf{m}} = \left(\mathbf{I} + \epsilon^2 \, \mathbf{D}^T \, \mathbf{D}\right)^{-1} \, \mathbf{d} = \mathbf{C} \, \left(\mathbf{C} + \epsilon^2 \, \mathbf{I}\right)^{-1} \, \mathbf{d} \; . \tag{5}$$

Smoothness is controlled by the choice of the regularization operator $\mathbf{D}$ and the scaling parameter ϵ.

Figure 7 shows the impulse response of the regularized smoothing operator in the 1-D case when $\mathbf{D}$ is the first difference operator. The impulse response has exponentially decaying tails. Repeated application of smoothing in this case is equivalent to applying an implicit Euler finite-difference scheme to the solution of the diffusion equation

$$\frac{\partial \mathbf{m}}{\partial t} = -\mathbf{D}^T \, \mathbf{D} \, \mathbf{m} \tag{6}$$

The impulse response converges to a Gaussian bell-shape curve in the physical domain, while its spectrum converges to a Gaussian in the frequency domain.

As far as the smoothing problem is concerned, there are better ways to smooth signals than applying equation 5. One example is triangle smoothing (3). To define triangle smoothing of one-dimensional signals, start with box smoothing, which, in the Z-transform notation, is a convolution with the filter

$$B_k(Z) = \frac{1}{k} \, \left(1 + Z + Z^2 + \cdots + Z^k\right) = \frac{1}{k} \, \frac{1 - Z^{k+1}}{1 - Z} \; , \tag{7}$$

where k is the filter length. Form a triangle smoother by correlation of two boxes

$$T_k(Z) = B_k(1/Z) \, B_k(Z) \tag{8}$$

Triangle smoothing is more efficient than regularized smoothing, because it requires twice less floating point multiplications. It also provides smoother results while having a compactly supported impulse response (Figure 2). Repeated application of triangle smoothing also makes the impulse response converge to a Gaussian shape but at a significantly faster rate. One can also implement smoothing by Gaussian filtering in the frequency domain or by applying other types of bandpass filters.

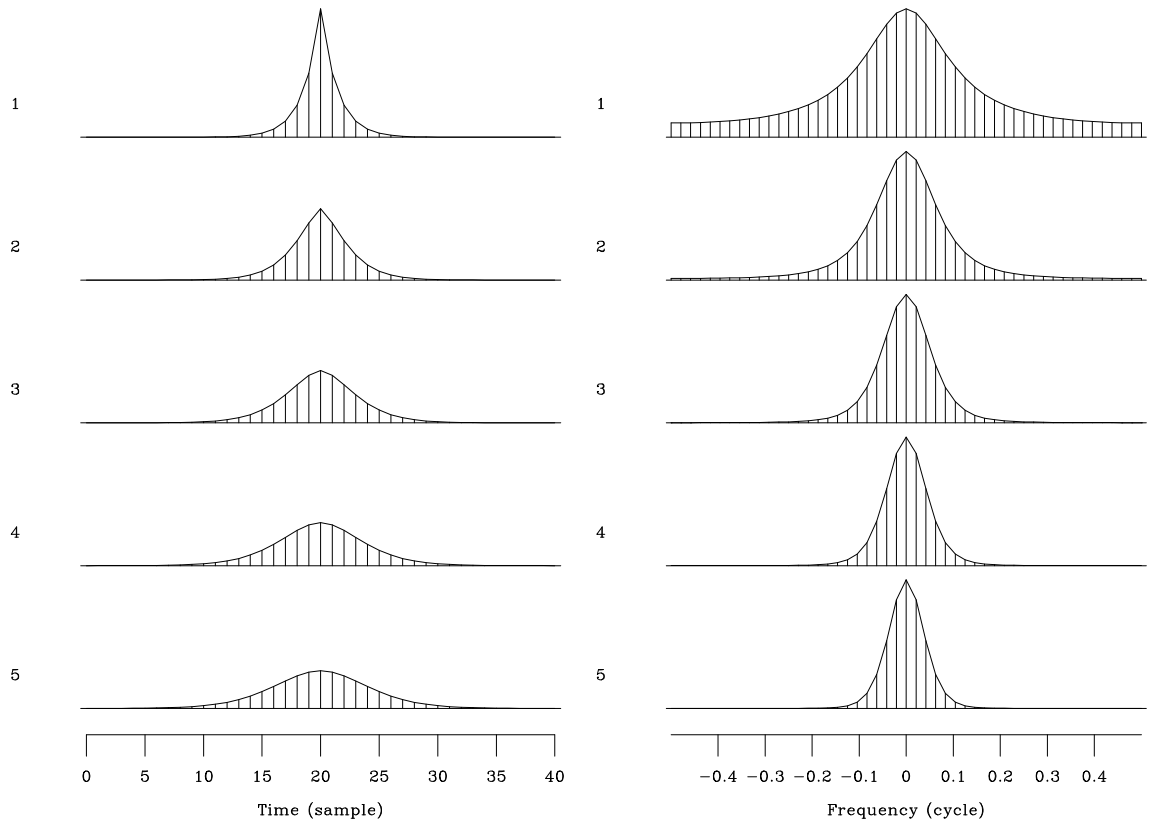


Figure 1: Left: impulse response of regularized smoothing. Repeated smoothing converges to a Gaussian bell shape. Right: frequency spectrum of regularized smoothing. The spectrum also converges to a Gaussian. shape/smoo exp

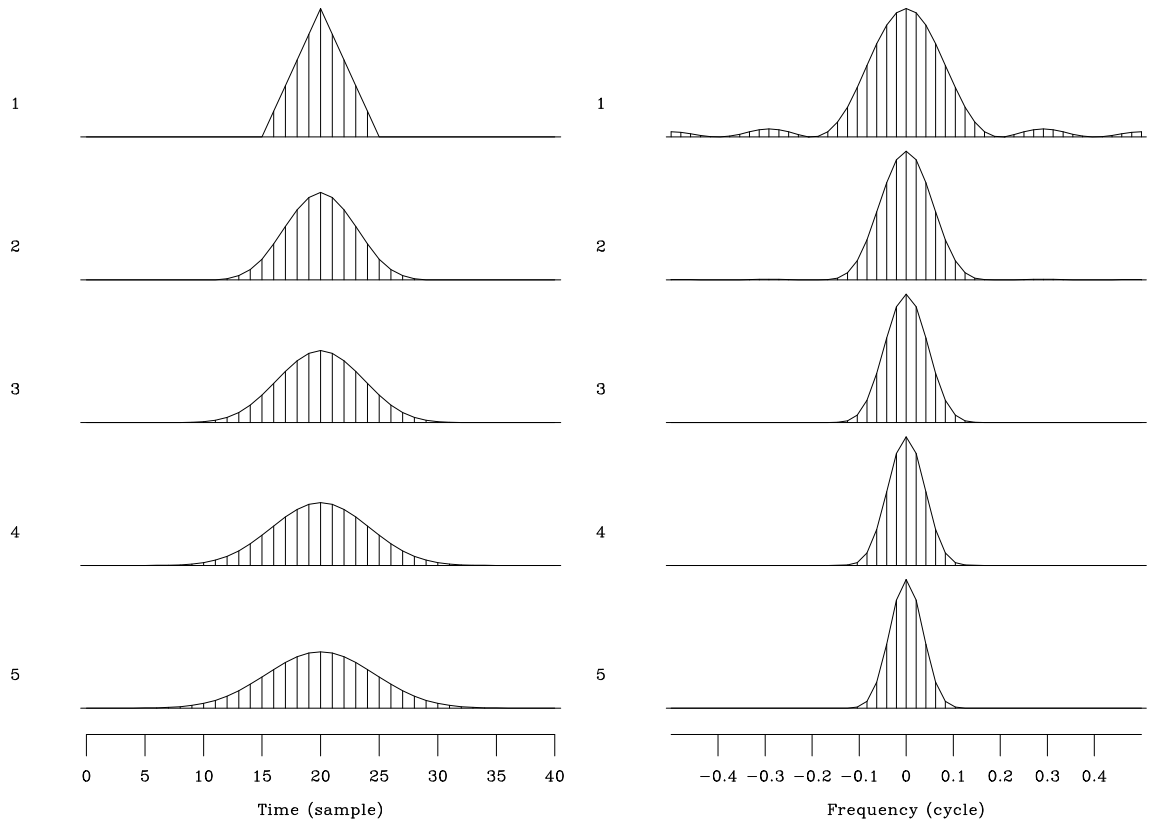


Figure 2: Left: impulse response of triangle smoothing. Repeated smoothing converges to a Gaussian bell shape. Right: frequency spectrum of triangle smoothing. Convergence to a Gaussian is faster than in the case of regularized smoothing. Compare to Figure 7. shape/smoo tri

SHAPING REGULARIZATION IN THEORY

The idea of shaping regularization starts with recognizing smoothing as a fundamental operation. In a more general sense, smoothing implies mapping of the input model to the space of admissible functions. I call the mapping operator *shaping*. Shaping operators do not necessarily smooth the input but they translate it into an acceptable model.

Taking equation 5 and using it as the definition of the regularization operator $\mathbf{D}$, we can write

$$\mathbf{S} = \left(\mathbf{I} + \epsilon^2 \, \mathbf{D}^T \, \mathbf{D}\right)^{-1} \tag{9}$$

or

$$\epsilon^2 \, \mathbf{D}^T \, \mathbf{D} = \mathbf{S}^{-1} - \mathbf{I} \; . \tag{10}$$

Substituting equation 10 into 1 yields a formal solution of the estimation problem regularized by shaping:

$$\widehat{\mathbf{m}} = \left(\mathbf{L}^T \, \mathbf{L} + \mathbf{S}^{-1} - \mathbf{I}\right)^{-1} \, \mathbf{L}^T \, \mathbf{d} = \left[\mathbf{I} + \mathbf{S} \, \left(\mathbf{L}^T \, \mathbf{L} - \mathbf{I}\right)\right]^{-1} \, \mathbf{S} \, \mathbf{L}^T \, \mathbf{d} \; . \tag{11}$$

The meaning of equation 11 is easy to interpret in some special cases:

- If $\mathbf{S} = \mathbf{I}$ (no shaping applied), we obtain the solution of unregularized problem.
- If $\mathbf{L}^T \, \mathbf{L} = \mathbf{I}$ ($\mathbf{L}$ is a unitary operator), the solution is simply $\mathbf{S} \, \mathbf{L}^T \, \mathbf{d}$ and does not require any inversion.
- If $\mathbf{S} = \lambda \, \mathbf{I}$ (shaping by scaling), the solution approaches $\lambda \, \mathbf{L}^T \, \mathbf{d}$ as λ goes to zero.

The operator $\mathbf{L}$ may have physical units that require scaling. Introducing scaling of $\mathbf{L}$ by $1/\lambda$ in equation 11, we can rewrite it as

$$\widehat{\mathbf{m}} = \left[\lambda^2 \, \mathbf{I} + \mathbf{S} \, \left(\mathbf{L}^T \, \mathbf{L} - \lambda^2 \, \mathbf{I}\right)\right]^{-1} \, \mathbf{S} \, \mathbf{L}^T \, \mathbf{d} \; . \tag{12}$$

The λ scaling in equation 12 controls the relative scaling of the forward operator $\mathbf{L}$ but not the shape of the estimated model, which is controlled by the shaping operator $\mathbf{S}$.

Iterative inversion with the conjugate-gradient algorithm requires symmetric positive definite operators (15). The inverse operator in equation 12 can be symmetrized when the shaping operator is symmetric and representable in the form $\mathbf{S} = \mathbf{H}\mathbf{H}^T$ with a square and invertible $\mathbf{H}$. The symmetric form of equation 12 is

$$\widehat{\mathbf{m}} = \mathbf{H} \, \left[\lambda^2 \, \mathbf{I} + \mathbf{H}^T \, \left(\mathbf{L}^T \, \mathbf{L} - \lambda^2 \, \mathbf{I}\right) \, \mathbf{H}\right]^{-1} \, \mathbf{H}^T \, \mathbf{L}^T \, \mathbf{d} \; . \tag{13}$$

When the inverted matrix is positive definite, equation 13 is suitable for an iterative inversion with the conjugate-gradient algorithm. Appendix A contains a complete algorithm description.

FROM TRIANGLE SMOOTHING TO TRIANGLE SHAPING

The idea of triangle smoothing can be generalized to produce different shaping operators for different applications. Let us assume that the estimated model is organized in a sequence of records, as follows: $\mathbf{m} = \left[\begin{array}{cccc} \mathbf{m}_1 & \mathbf{m}_2 & \vdots & \mathbf{m}_n \end{array} \right]^T$. Depending on the application, the records can be samples, traces, shot profiles, etc. Let us further assume that, for each pair of neighboring records, we can design a prediction operator $\mathbf{Z}_{k\to k+1}$, which predicts record $k+1$ from record k. A global prediction operator is then

$$\mathbf{Z} = \left[\begin{array}{cccccc} 0 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{Z}_{1\to 2} & 0 & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{Z}_{2\to 3} & 0 & \cdots & 0 & 0 \\ 0 & 0 & \mathbf{Z}_{3\to 4} & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \cdots & \mathbf{Z}_{n-1\to n} & 0 \end{array} \right] . \tag{14}$$

The operator $\mathbf{Z}$ effectively shifts each record to the next one. When local prediction is done with identity operators, this operation is completely analogous to the Z operator used in the theory of digital signal processing. The $\mathbf{Z}$ operator can be squared, as follows:

$$\mathbf{Z}^2 = \left[\begin{array}{cccccc} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \mathbf{Z}_{2\to 3}\,\mathbf{Z}_{1\to 2} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \mathbf{Z}_{3\to 4}\,\mathbf{Z}_{2\to 3} & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{Z}_{n-1\to n}\,\mathbf{Z}_{n-2\to n-1} & 0 & 0 \end{array} \right] . \tag{15}$$

In a shorter notation, we can denote prediction of record j from record i by $\mathbf{Z}_{i\to j}$ and write

$$\mathbf{Z}^2 = \left[\begin{array}{cccccc} 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 \\ \mathbf{Z}_{1\to 3} & 0 & \cdots & 0 & 0 & 0 \\ 0 & \mathbf{Z}_{2\to 4} & \cdots & 0 & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \mathbf{Z}_{n-2\to n} & 0 & 0 \end{array} \right] . \tag{16}$$

Subsequently, the prediction operator $\mathbf{Z}$ can be taken to higher powers. This leads immediately to an idea on how to generalize box smoothing: predict each record from the record immediately preceding it, the record two steps away, etc. and average all those predictions and the actual records. In mathematical notation, a box shaper of length k is then simply

$$\mathbf{B}_k = \frac{1}{k}\left(\mathbf{I} + \mathbf{Z} + \mathbf{Z}^2 + \cdots + \mathbf{Z}^k\right) , \tag{17}$$

which is completely analogous to equation 7. Implementing equation 17 directly requires many computational operations. Noting that

$$(\mathbf{I}-\mathbf{Z})\ \mathbf{B}_k = \frac{1}{k}\ \left(\mathbf{I}-\mathbf{Z}^{k+1}\right)\ , \tag{18}$$

we can rewrite equation 17 in the compact form

$$\mathbf{B}_k = \frac{1}{k}\ (\mathbf{I}-\mathbf{Z})^{-1}\ \left(\mathbf{I}-\mathbf{Z}^{k+1}\right)\ , \tag{19}$$

which can be implemented economically using recursive inversion of the lower triangular operator $\mathbf{I}-\mathbf{Z}$. Finally, combining two generalized box smoothers creates a symmetric generalized triangle shaper

$$\mathbf{T}_k = \mathbf{B}_k^T\,\mathbf{B}_k\ , \tag{20}$$

which is analogous to equation 8. A triangle shaper uses local predictions from both the left and the right neighbors of a record and averages them using triangle weights.

Figure 3 illustrates generalized triangle shaping by constructing a non-stationary smoothing operator that follows local structural dips. Figure 3a shows a synthetic image from (4). Figure 3b is a local dip estimate obtained with plane-wave destruction (9). Figure 3c is the result of applying triangle smoothing oriented along local dip directions to a field of random numbers. Oriented smoothing generates a pattern reflecting the structural composition of the original image. This construction resembles the method of (2). Figure 3d shows the impulse responses of oriented smoothing for several distinct locations in the image space. As illustrated later in this paper, oriented smoothing can be applied for generating geophysical Earth models that are compliant with the local geological structure (20; 26; 5).

Appendix B describes general rules for combining elementary shaping operators.

EXAMPLES

Two simple examples in data regularization and seismic velocity estimation illustrate the method of shaping.

1-D inverse data interpolation

I start with a simple 1-D example: a benchmark data regularization test used previously by (11).

The input synthetic data are irregular samples from a sinusoidal signal (Figure 4). The task of data regularization is to reconstruct the data on a regular grid. The forward operator $\mathbf{L}$ in this case is forward interpolation from a regular grid using linear (two-point) interpolation.

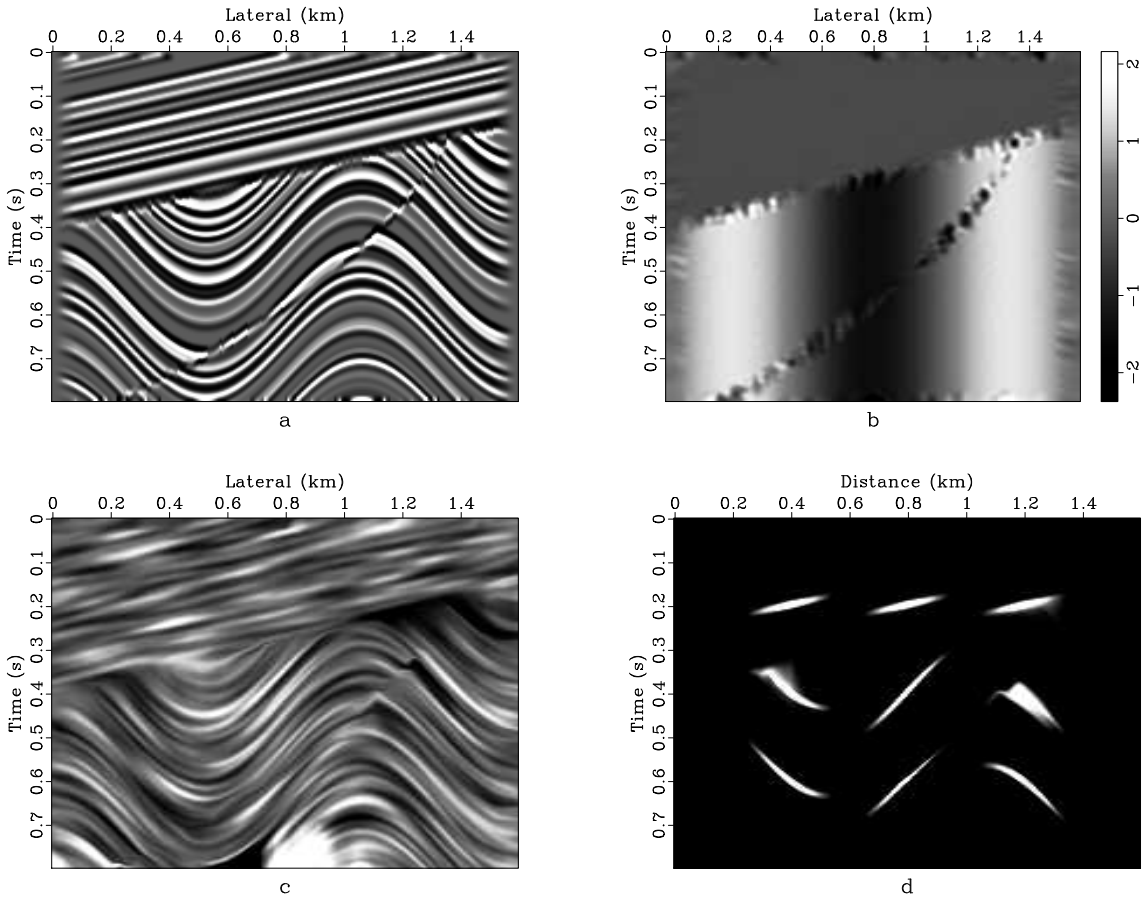


Figure 3: Shaping by smoothing along local dip directions according to operator $\mathbf{T}_k$ from equation 20. a: an example image, b: local dip estimation, c: smoothing random numbers along local dips, d: impulse responses of oriented smoothing for nine different locations in the image space. shape/lomo tris

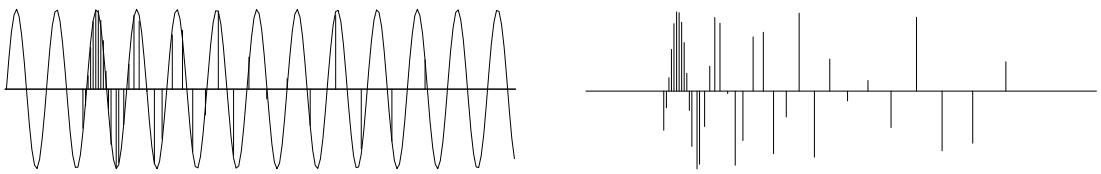


Figure 4: The input data (b) are irregular samples from a sinusoid (a). shape/int1 both2,data2

Figure 5 shows some of the first iterations and the final results of inverse interpolation with Tikhonov's regularization using equation 1 and with model preconditioning using equation 3. The regularization operator $\mathbf{D}$ in equation 1 is the first finite difference, and the preconditioning operator $\mathbf{P}$ in 3 is the inverse of $\mathbf{D}$ or causal integration. The preconditioned iteration converges faster but its very first iterations produce unreasonable results. This type of behavior can be dangerous in real large-scale problems, when only few iterations are affordable.

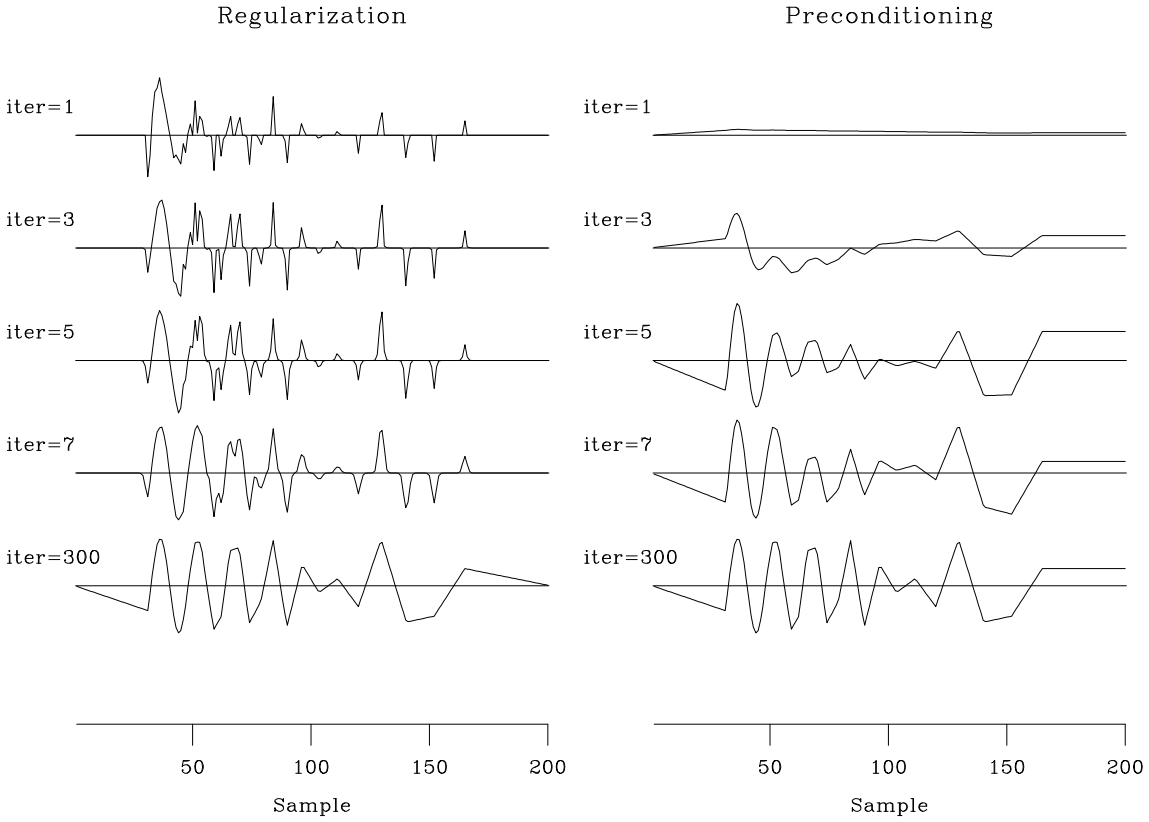


Figure 5: The first iterations and the final result of inverse interpolation with Tikhonov's regularization using equation 1 (left) and with model preconditioning using equation 3 (right). The regularization operator $\mathbf{D}$ is the first finite difference. The preconditioning operator $\mathbf{P} = \mathbf{D}^{-1}$ is causal integration. The number of iterations is indicated in the plot labels. shape/int1 if

The left side of Figure 6 shows some of the first iterations and the final result of inverse interpolation with shaping regularization, where the shaping operator $\mathbf{S}$ was chosen to be Gaussian smoothing with the impulse response width of about 10 samples. The final result is smoother, and the iteration is both fast-converging and producing reasonable results at the very first iterations. Thanks to the fact that the smoothing operation is applied at each iteration, the estimated model is guaranteed to have the prescribed shape.

Examining the spectrum of the final result (Figure 9), one can immediately notice

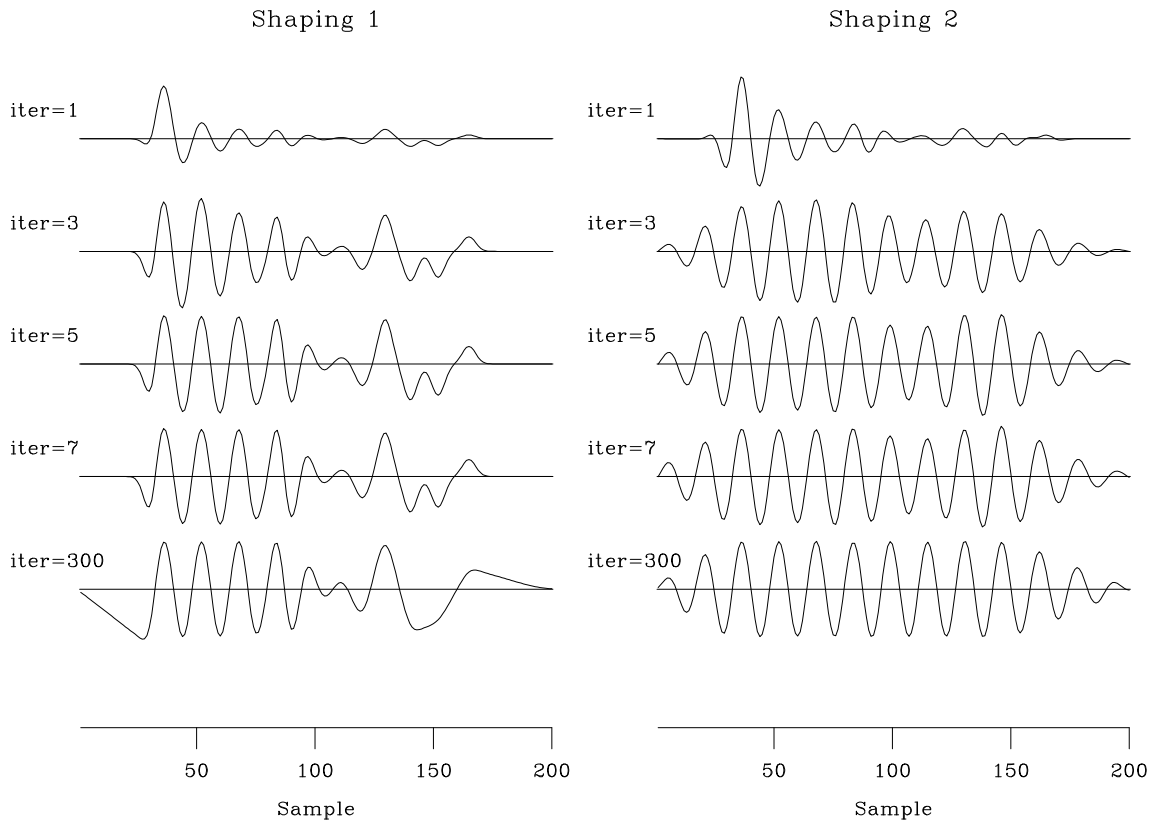


Figure 6: The first iterations and the final result of inverse interpolation with shaping regularization using equation 13. Left: the shaping operator **H** is lowpass filtering with a Gaussian smoother. Right: the shaping operator **H** is bandpass filtering with a shifted Gaussian. Shaping by bandpass filtering recovers the sinusoidal shape of the estimated model. The number of iterations is indicated in the plot labels. shape/int1 sz

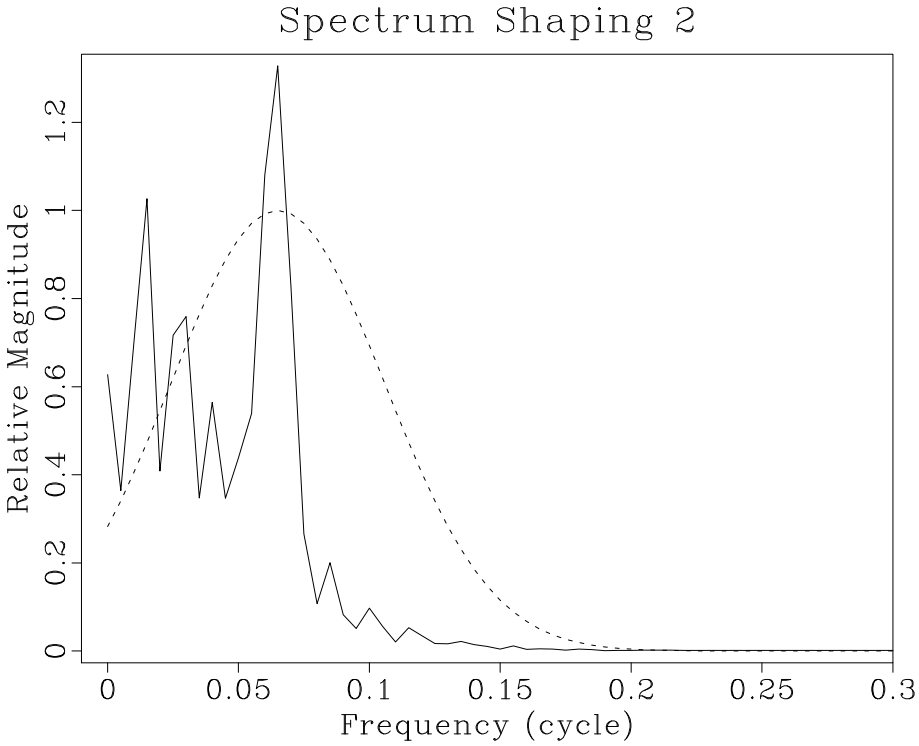


Figure 7: Spectrum of the estimated model (solid curve) fitted to a shifted Gaussian (dashed curve). The Gaussian band-limited filter defines a shaping operator for recovering a band-limited signal. shape/int1 spec

the peak at the dominant frequency of the initial sinusoid. Fitting a Gaussian shape to the peak defines a data-adaptive shaping operator as a bandpass filter implemented in the frequency domain (dashed curve in Figure 9). Inverse interpolation with the estimated shaping operator recovers the original sinusoid (right side of Figure 6). Analogous ideas in the model preconditioning context were proposed by (17).

Velocity estimation

The second example is an application of shaping regularization to seismic velocity estimation. Figure 8 shows a time-migrated image from a historic Gulf of Mexico dataset (4). The image was obtained by velocity continuation (10). The corresponding migration velocity is shown in the right plot of Figure 8. Shaping regularization was used for picking a smooth velocity profile from semblance gathers obtained in the process of velocity continuation.

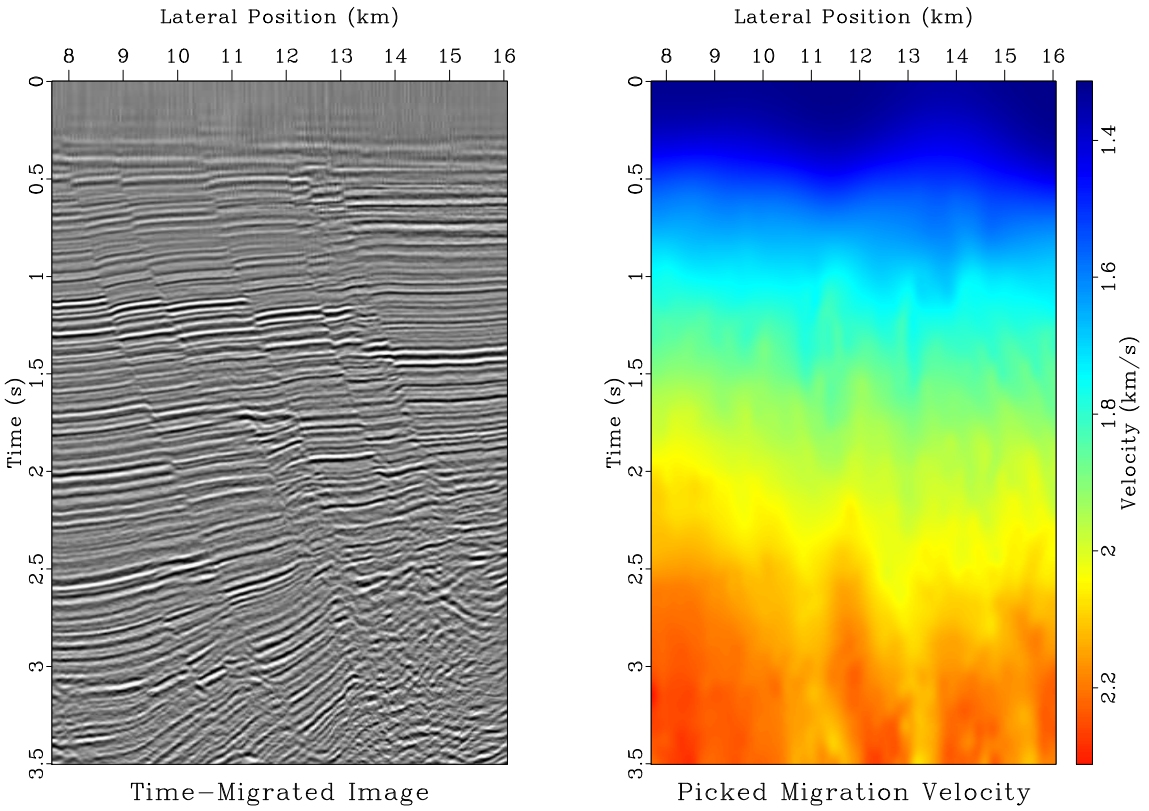


Figure 8: Left: time-migrated image. Right: The corresponding migration velocity from automatic picking. shape/beivc bei-fmg2

The task of this example is to convert the time migration velocity to the interval velocity. I use the simple approach of Dix inversion (7) formulated as a regularized inverse problem (24). In this case, the forward operator $\mathbf{L}$ in equation 11 is a weighted time integration. There is a choice in choosing the shaping operator $\mathbf{H}$.

Figure 9 shows the result of inversion with shaping by triangle smoothing. While the interval velocity model yields a good prediction of the measured velocity, it may not appear geologically plausible because the velocity structure does not follow the structure of seismic reflectors as seen in the migrated image.

Following the ideas of steering filters (6; 5) and plane-wave construction (12), I estimate local slopes in the migration image using the method of plane-wave destruction (9) and define a triangle plane-wave shaping operator $\mathbf{H}$ using the method of the previous section. The result of inversion, shown in Figures 10 and 11, makes the estimated interval velocity follow the geological structure and thus appear more plausible for direct interpretation. Similar results were obtained by (12) using model parameterization by plane-wave construction but at a higher computational cost. In the case of shaping regularization, about 25 efficient iterations were sufficient to converge to the machine precision accuracy.

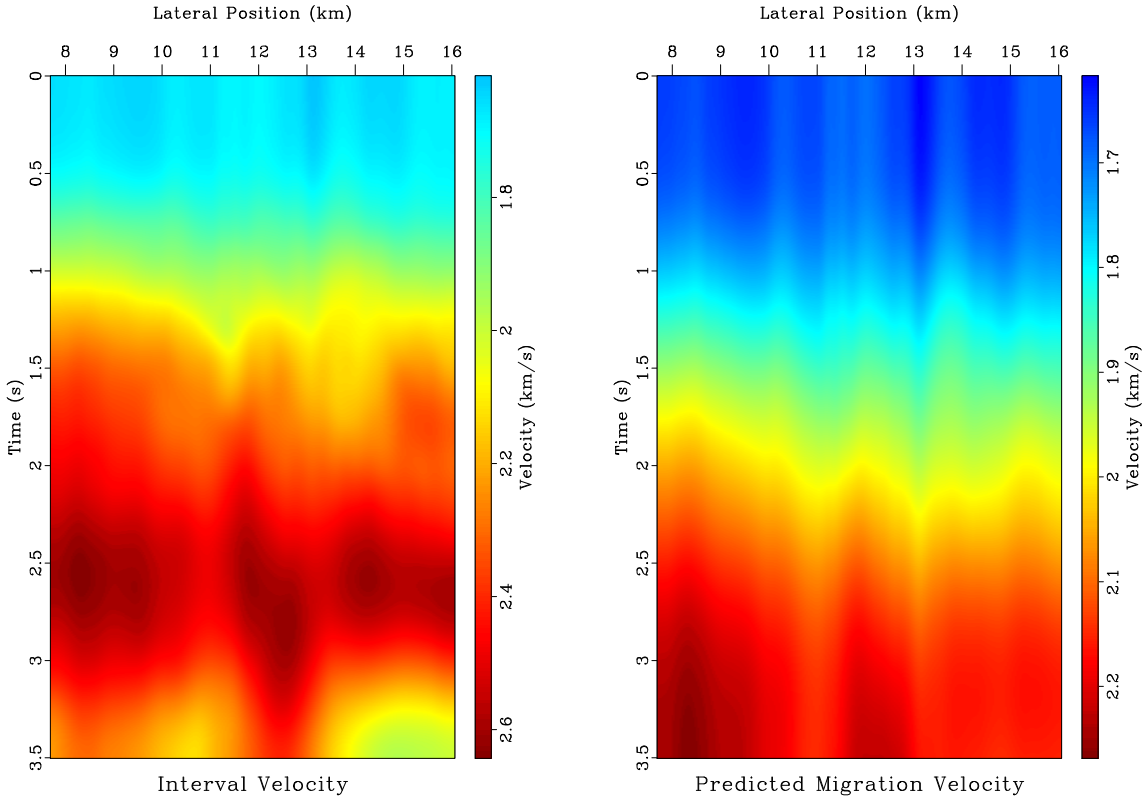


Figure 9: Left: estimated interval velocity. Right: predicted migration velocity. Shaping by triangle smoothing. shape/beivc bei-dix

CONCLUSIONS

Shaping regularization is a new general method for imposing regularization constraints in estimation problems. The main idea of shaping regularization is to rec-

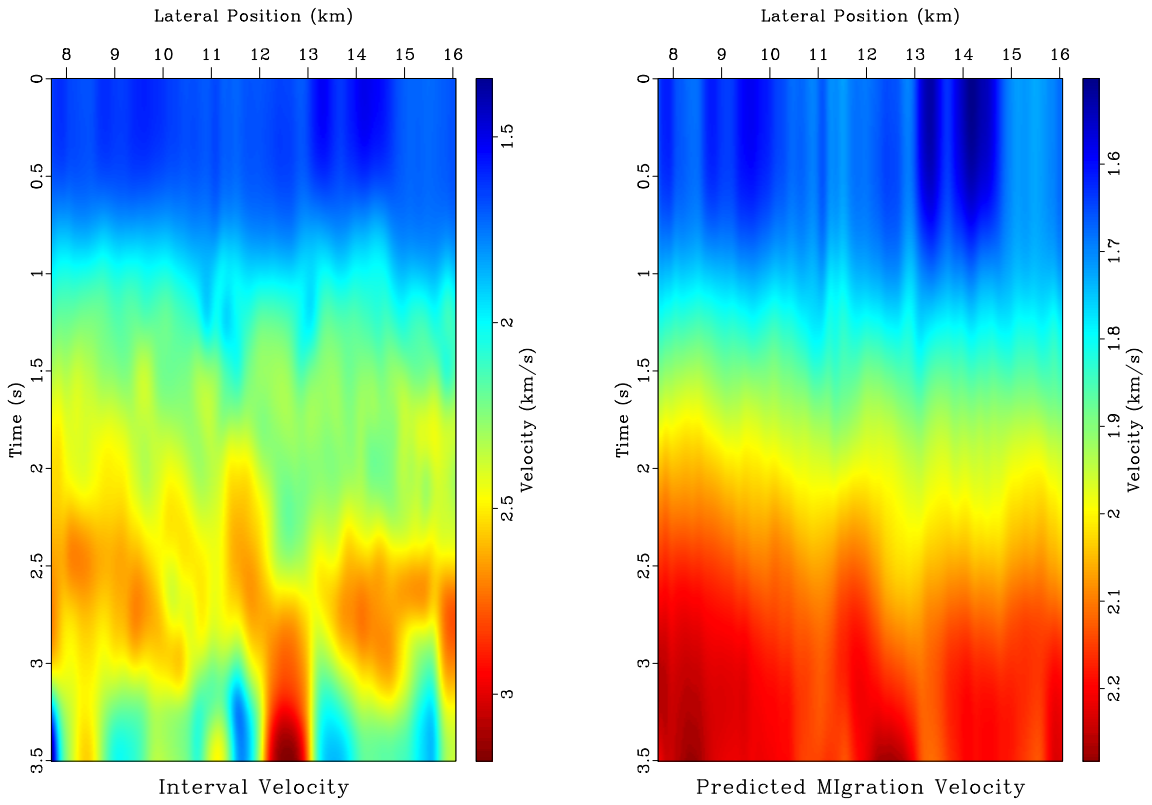


Figure 10: Left: estimated interval velocity. Right: predicted migration velocity. Shaping by triangle local plane-wave smoothing creates a velocity model consistent with the reflector structure. shape/beivc bei-shp

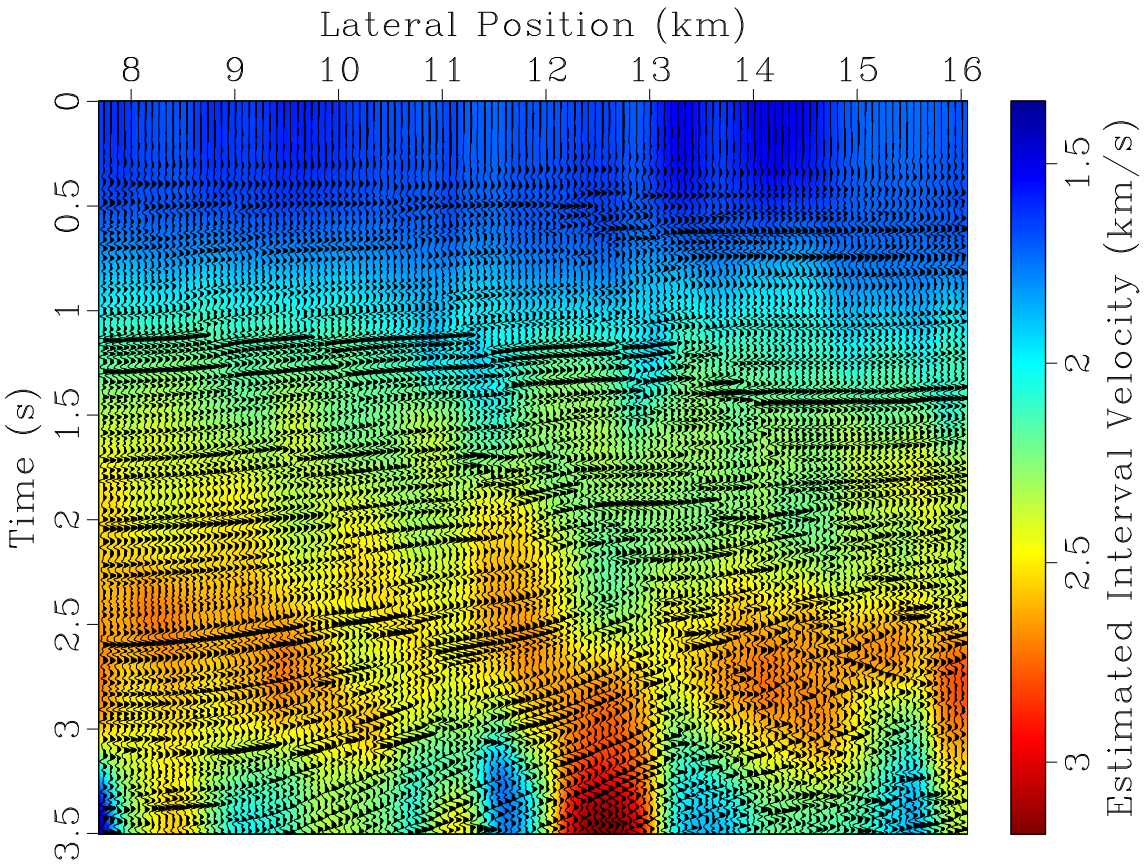


Figure 11: Seismic image from Figure 8 overlaid on top of the interval velocity model estimated with triangle plane-wave shaping regularization. shape/beivc bei-shpw

ognize shaping (mapping to the space of acceptable functions) as a fundamental operation and to incorporate it into iterative inversion.

There is an important difference between shaping regularization and conventional (Tikhonov's) regularization from the user prospective. Instead of trying to find and specify an appropriate regularization operator, the user of the shaping regularization algorithm specifies a shaping operator, which is often easier to design. Shaping operators can be defined following a triangle construction from local predictions or by combining elementary shapers.

I have shown two simple illustrations of shaping applications. The examples demonstrate a typical behavior of the method: enforced model compliance to the specified shape at each iteration and, in many cases, fast iterative convergence of the conjugate gradient iteration. The model compliance behavior follows from the fact that shaping enters directly into the iterative process and provides an explicit control on the shape of the estimated model.

ACKNOWLEDGMENTS

I would like to thank Pierre Hardy and Mauricio Sacchi for inspiring discussions, TOTAL for partially supporting this research, and three reviewers for helpful suggestions.

REFERENCES

[1] Bube, K., and R. Langan, 1999, On a continuation approach to regularization for crosswell tomography: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1295–1298.

[2] Claerbout, J., and M. Brown, 1999, Two-dimensional textures and prediction-error filters: 61st Mtg., Eur. Assn. Geosci. Eng., Session:1009.

[3] Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.

[4] ———, 2006, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

[5] Clapp, R. G., B. Biondi, and J. F. Claerbout, 2004, Incorporating geologic information into reflection tomography: Geophysics, **69**, 533–546.

[6] Clapp, R. G., B. L. Biondi, S. B. Fomel, and J. F. Claerbout, 1998, Regularizing velocity estimation using geologic dip information: 68th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1851–1854.

[7] Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68–86.

[8] Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.

[9] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

[10] ———, 2003, Time-migration velocity analysis by velocity continuation: Geophysics, **68**, 1662–1672.
[11] Fomel, S., and J. Claerbout, 2003, Multidimensional recursive filter preconditioning in geophysical estimation problems: Geophysics, **68**, 577–588.
[12] Fomel, S., and A. Guitton, 2006, Regularizing seismic inverse problems by model re-parameterization using plane-wave construction: Geophysics, **71**, S209–S217.
[13] Fomel, S., P. Sava, J. Rickett, and J. Claerbout, 2003, The Wilson-Burg method of spectral factorization with application to helical filtering: Geophysical Prospecting, **51**, 409–420.
[14] Harlan, W. S., 1995, Regularization by model redefinition: http://billharlan.com/pub/papers/regularization.pdf.
[15] Hestenes, M. R., and E. Steifel, 1952, Methods of conjugate gradients for solving linear systems: J. Res. Nat. Bur. Stand., **49**, 409–436.
[16] Jackson, D. D., 1972, Interpretation of inaccurate, insufficient and inconsistent data: Geophys. J. R. Astron. Soc., **28**, 97–109.
[17] Liu, B., and M. Sacchi, 2001, Minimum weighted norm interpolation of seismic data with adaptive weights: 71st Ann. Internat. Mtg, Soc. of Expl. Geophys., 1921–1924.
[18] Portniaguine, O., and J. Castagna, 2004, Inverse spectral decomposition, *in* 74th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1786–1789.
[19] Saad, Y., 2003, Iterative methods for sparse linear systems: SIAM.
[20] Sinoquet, D., 1993, Modeling a priori information on the velocity field in reflection tomography: 63rd Ann. Internat. Mtg, Soc. of Expl. Geophys., 591–594.
[21] Tarantola, A., 2004, Inverse problem theory and methods for model parameter estimation: SIAM.
[22] Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Sovet Math. Dokl., 1035–1038.
[23] Trad, D., T. Ulrych, and M. Sacchi, 2003, Latest views of the sparse Radon transform: Geophysics, **68**, 386–399.
[24] Valenciano, A. A., M. Brown, A. Guitton, and M. D. Sacchi, 2004, Interval velocity estimation using edge-preserving regularization, *in* 74th Ann. Internat. Mtg: Soc. of Expl. Geophys.
[25] van der Vorst, H. A., 2003, Iterative Krylov methods for large linear systems: Cambridge University Press.
[26] Versteeg, R., and W. W. Symes, 1993, Geometric constraints on seismic inversion: 63rd Ann. Internat. Mtg, Soc. of Expl. Geophys., 595–598.
[27] Woodward, M. J., P. Farmer, D. Nichols, and S. Charles, 1998, Automated 3-D tomographic velocity analysis of residual moveout in prestack depth migrated common image point gathers: 68th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1218–1221.
[28] Zhdanov, M. S., 2002, Geophysical inverse theory and regularization problems: Elsevier Science.
[29] Zhou, H., S. Gray, J. Young, D. Pham, and Y. Zhang, 2003, Tomographic residual curvature analysis: The process and its components: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 666–669.

APPENDIX A

CONJUGATE-GRADIENT ALGORITHM

A complete algorithm for conjugate-gradient iterative inversion with shaping regularization is given below. The algorithm follows directly from combining equation 13 with the classic conjugate-gradient algorithm of (15).

CONJUGATE GRADIENTS WITH SHAPING($\mathbf{L}, \mathbf{H}, \mathbf{d}, \lambda, tol, N$)

$\mathbf{p} \leftarrow \mathbf{0}$
$\mathbf{m} \leftarrow \mathbf{0}$
$\mathbf{r} \leftarrow -\mathbf{d}$
for $n \leftarrow 1, 2, \ldots, N$
  do
    $\mathbf{g}_m \leftarrow \mathbf{L}^T \mathbf{r} - \lambda \mathbf{m}$
    $\mathbf{g}_p \leftarrow \mathbf{H}^T \mathbf{g}_m + \lambda \mathbf{p}$
    $\mathbf{g}_m \leftarrow \mathbf{H}\, \mathbf{g}_p$
    $\mathbf{g}_r \leftarrow \mathbf{L}\, \mathbf{g}_m$
    $\rho \leftarrow \mathbf{g}_p^T \mathbf{g}_p$
    if $n = 1$
      then $\beta \leftarrow 0$
        $\rho_0 \leftarrow \rho$
      else $\beta \leftarrow \rho / \hat{\rho}$
        if $\beta < tol$ or $\rho / \rho_0 < tol$
          then return m

$$\begin{bmatrix} \mathbf{s}_p \\ \mathbf{s}_m \\ \mathbf{s}_r \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{g}_p \\ \mathbf{g}_m \\ \mathbf{g}_r \end{bmatrix} + \beta \begin{bmatrix} \mathbf{s}_p \\ \mathbf{s}_m \\ \mathbf{s}_r \end{bmatrix}$$

    $\alpha \leftarrow \rho / \left[\mathbf{s}_r^T \mathbf{s}_r + \lambda \left(\mathbf{s}_p^T \mathbf{s}_p - \mathbf{s}_m^T \mathbf{s}_m \right) \right]$

$$\begin{bmatrix} \mathbf{p} \\ \mathbf{m} \\ \mathbf{r} \end{bmatrix} \leftarrow \begin{bmatrix} \mathbf{p} \\ \mathbf{m} \\ \mathbf{r} \end{bmatrix} - \alpha \begin{bmatrix} \mathbf{s}_p \\ \mathbf{s}_m \\ \mathbf{s}_r \end{bmatrix}$$

    $\hat{\rho} \leftarrow \rho$
return m

The iteration terminates after N iterations or upon reaching convergence to the specified tolerance tol. It uses auxiliary vectors $\mathbf{p}$, $\mathbf{r}$, $\mathbf{s}_p$, $\mathbf{s}_m$, $\mathbf{s}_r$, $\mathbf{g}_p$, $\mathbf{g}_m$, $\mathbf{g}_r$ and applies operators $\mathbf{L}$, $\mathbf{H}$ and their adjoints only once per each iteration.

APPENDIX B

COMBINING SHAPING OPERATORS

General rules can be developed to combine two or more shaping operators for the cases when there are several features in the model that need to be characterized simultaneously. A general rule for combining two different shaping operators $\mathbf{S}_1$ and $\mathbf{S}_2$ can have the form

$$\mathbf{S}_{12} = \mathbf{S}_1 + \mathbf{S}_2 - \mathbf{S}_1 \, \mathbf{S}_2 \ , \tag{B-1}$$

where one adds the responses of the two shapers and then subtracts their overlap. An example is shown in Figure B-1, where an impulse response for oriented smoothing in two different directions is constructing from smoothing in each of the two directions separately.

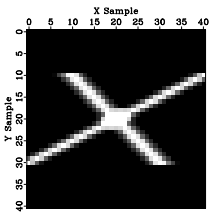


Figure B-1: Impulse response for a combination of two shaping operators smoothing in two different directions. shape/smoo test12

Combining two operators that work in orthogonal directions can be accomplished with a simple tensor product, as follows:

$$\mathbf{S}_{xy} = \mathbf{S}_x \, \mathbf{S}_y \ , \tag{B-2}$$

where $\mathbf{S}_x$ and $\mathbf{S}_y$ are shaping operators that apply in orthogonal x- and y-directions, and $\mathbf{S}_{xy}$ is a combined operator that works in both directions. An example is shown in Figure 3(a), where two two-dimensional shapers working in orthogonal directions are combined to produce an impulse response of 3-D shaping operator that applies smoothing along a three-dimensional plane.

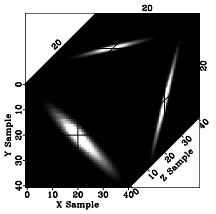


Figure B-2: 3-D impulse response for a combination of two 2-D shaping operators smoothing in in-line and cross-line directions. shape/smoo plane

Constructing multidimensional recursive filters for helical preconditioning (11) is significantly more difficult. It involves helical spectral factorization, which may create long inefficient filters (13).

Kirchhoff migration using eikonal-based computation of traveltime source-derivatives

*Siwei Li and Sergey Fomel**

ABSTRACT

The computational efficiency of Kirchhoff-type migration can be enhanced by employing accurate traveltime interpolation algorithms. We address the problem of interpolating between a sparse source sampling by using the derivative of traveltime with respect to the source location. We adopt a first-order partial differential equation that originates from differentiating the eikonal equation to compute the traveltime source-derivatives efficiently and conveniently. Unlike methods that rely on finite-difference estimations, the accuracy of the eikonal-based derivative does not depend on input source sampling. For smooth velocity models, the first-order traveltime source-derivatives enable a cubic Hermite traveltime interpolation that takes into consideration the curvatures of local wave-fronts and can be straight-forwardly incorporated into Kirchhoff anti-aliasing schemes. We provide an implementation of the proposed method to first-arrival traveltimes by modifying the fast-marching eikonal solver. Several simple synthetic models and a semi-recursive Kirchhoff migration of the Marmousi model demonstrate the applicability of the proposed method.

INTRODUCTION

Over the years, there have been significant efforts and progress in traveltime computations. The quality of traveltimes has a direct influence on Kirchhoff-type migrations since it determines the kinematic behaviors of the imaged wavefields. One can use either ray-tracing approaches or finite-difference solutions of the eikonal equation. The first option naturally handles multi-arrivals and can be extended to other wavefield approximations, such as Gaussian beams (17; 18; 2; 16), but is at the same time usually subject to the necessities of ray-coordinate and migration-grid mapping and irregular interpolation between rays in the presence of shadow zones in complex velocity media (24). Two popular methods from the second option are the *fast-marching method* (FMM) (25; 26) and the *fast-sweeping method* (FSM) (35). They both rely on an ordered update to recover the causality behind expanding wave-fronts in a general medium, and are thus limited to first-arrival computations. Several works attempt to overcome the single-arrival drawback of the finite-difference eikonal solvers, for exam-

*e-mail: sergey.fomel@beg.utexas.edu

ple multi-phase computation (10), phase-space escape equations (13), and slowness marching (28).

In practice, traveltime tables can be pre-computed on coarse grids and saved on disk, then serve as a dictionary when read by Kirchhoff migration algorithms. It is common to carry out a certain interpolation in this process in order to satisfy the needs of depth migration for fine-gridded traveltime tables at a large number of source locations (20; 30; 4). Kirchhoff migrations with traveltime tables computed on the fly face the same issue. During the traveltime computation stage, accuracy requirements from eikonal solvers may lead to a fine model sampling. Combined with a large survey, traveltime computation for each shot can be costly. Because all traveltime computations handle one shot at a time, the overall cost increases linearly with the number of sources. Moreover, we need to store a large amount of traveltimes out of a dense source sampling. Therefore a sparse source sampling is preferred. In this paper, we try to address the problem of traveltime table interpolation between sparse source samples. The traveltime table estimated with simple nearest-neighbor or linear interpolation could not provide satisfying accuracy unless the velocity model has small variations. One possible improvement is to include derivatives in interpolation. During ray tracing, traveltime source-derivatives are directly connected to the slowness vector at the source and stay constant along individual rays, thus could be outputted as a by-product of traveltimes. For finite-difference eikonal solvers, such a convenience is not easily available. In these cases, we would like to avoid an extra differentiation on traveltime tables along the source dimension to compute such derivatives (30), because its accuracy in turn relies on a dense source sampling and induces additional computations. (5) derived an equation for the traveltime perturbation with respect to the source location changes. The governing equation is a first-order partial differential equation (PDE) that describes traveltime source-derivatives in a relative coordinate moving along with the source. In this paper, we show that the traveltime source-derivative desired by interpolation is related to this relative-coordinate quantity by a simple subtraction with the slowness vector. Unlike a finite-difference approach, traveltime source-derivatives computed by the PDE method are source-sampling independent. The extra costs are rather inexpensive. In this paper, we apply this method to Kirchhoff migration with first-arrival traveltimes computed by the FMM eikonal solver.

The paper is organized as follows. In the first section, we review the theory and implementation of the eikonal-based traveltime source-derivatives. Next, we use both simple and complex synthetic models to demonstrate the accuracy of a cubic Hermite traveltime table interpolation using the source-derivatives, and show effects of incorporating such an interpolation into Kirchhoff migration. We focus mainly on the kinematics in these experiments by neglecting possible true-amplitude weights in Kirchhoff migration. Finally, we discuss limitations and possible extensions of the proposed method.

THEORY AND IMPLEMENTATION

We consider the isotropic eikonal equation:

$$\nabla T(\mathbf{x}) \cdot \nabla T(\mathbf{x}) = \frac{1}{v^2(\mathbf{x})} \equiv W(\mathbf{x}) \ , \tag{1}$$

where $\mathbf{x}$ is a point in space, $T(\mathbf{x})$ is the traveltime and $v(\mathbf{x})$ is the velocity. For 2D models, $\mathbf{x}$ is a vector containing the depth and the inline position. For 3D models, $\mathbf{x}$ also includes the crossline position. For conciseness, we define $W(\mathbf{x})$ as slowness-squared. Equation 1 can be derived by inserting the ray-theory series into the wave-equation and setting the coefficient of the leading-order term to zero (8). We are interested in particular in point-source solutions of the eikonal equation, i.e. with the initial condition $T(\mathbf{x_s}) = 0$ where $\mathbf{x_s}$ denotes the source location.

Traveltime Source-derivative

The point-source traveltime $T(\mathbf{x})$ clearly depends on the source location $\mathbf{x_s}$. To explicitly show such a dependency in the eikonal equation, we define a relative coordinate $\mathbf{q}$ as

$$\mathbf{q} = \mathbf{x} - \mathbf{x_s} \ , \tag{2}$$

and use $\hat{T}(\mathbf{q}; \mathbf{x_s})$ to denote traveltime in the relative coordinates. After inserting this definition into equation 1, we obtain

$$\nabla_{\mathbf{q}} \hat{T} \cdot \nabla_{\mathbf{q}} \hat{T} = W(\mathbf{q} + \mathbf{x_s}) \ . \tag{3}$$

Here the differentiation $\nabla_{\mathbf{q}}$ stands for gradient operator in the relative coordinate $\mathbf{q}$ and is taken with a fixed source location $\mathbf{x_s}$. In 3D, if $\mathbf{q} = (q_1, q_2, q_3)$ and denoting $\mathbf{e}_i$ with $i = \{1, 2, 3\}$ to be the unit vector in depth, inline and crossline directions, respectively, then

$$\nabla_{\mathbf{q}} \equiv \frac{\partial}{\partial q_1}\mathbf{e}_1 + \frac{\partial}{\partial q_2}\mathbf{e}_2 + \frac{\partial}{\partial q_3}\mathbf{e}_3 \ . \tag{4}$$

Since we are interested in the traveltime derivative with respect to the source, i.e. $\partial T/\partial \mathbf{x_s}$, we take directional derivative $\partial/\partial \mathbf{x_s}$ to $\hat{T}(\mathbf{q}; \mathbf{x_s})$ and apply the chain-rule according to equation 2:

$$\frac{\partial T}{\partial \mathbf{x_s}} \equiv \frac{\partial \hat{T}}{\partial \mathbf{x_s}} = \frac{\partial \hat{T}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{x_s}} + \frac{\partial \hat{T}}{\partial \mathbf{q}}\frac{\partial \mathbf{q}}{\partial \mathbf{x_s}} = \frac{\partial \hat{T}}{\partial \mathbf{x}} - \frac{\partial \hat{T}}{\partial \mathbf{q}} \ . \tag{5}$$

Equation 14 results in a vector that contains the traveltime source-derivatives in depth, inline and crossline directions. In accordance with $\partial/\partial \mathbf{x_s}$, $\partial/\partial \mathbf{x}$ and $\partial/\partial \mathbf{q}$ are also directional derivatives. All numerical examples in this paper are based on a typical 2D acquisition, where we assume a constant source depth and thus only the inline traveltime source-derivative is of interest. The quantity $\partial \hat{T}/\partial \mathbf{q}$ coincides

with the slowness vector of the ray that originates from $\mathbf{x_s}$. For a finite-difference eikonal solver such as FMM and FSM, it is usually estimated by an upwind scheme during traveltime computations at each grid point and thus can be easily extracted. Applying $\partial/\partial\mathbf{x}$ to both sides of equation 3, we find

$$\nabla_{\mathbf{q}}\hat{T}\cdot\nabla_{\mathbf{q}}\frac{\partial\hat{T}}{\partial\mathbf{x}}=\frac{1}{2}\frac{\partial W}{\partial\mathbf{x}}\ . \tag{6}$$

Equation 6 has the form of the linearized eikonal equation (3) and was previously derived, in a slightly different notation, by (5). It implies that $\partial\hat{T}/\partial\mathbf{x}$, as needed by equation 14, can be determined along the characteristics of $\hat{T}$. Since the right-hand side contains a slowness-squared derivative, the velocity model must be differentiable, as is usually required by traveltime computations. The derivation also indicates that the accuracy of an eikonal-based traveltime source-derivative is source-sampling independent but model-sampling dependent, as from equations 14 and 6 $\partial/\partial\mathbf{x_s}$ relies on $\hat{T}$, $\partial/\partial\mathbf{q}$ and $\partial/\partial\mathbf{x}$. The accuracy from a direct finite-difference estimation on $\partial/\partial\mathbf{x_s}$, in comparison, is both source- and model-sampling dependent.

Continuing applying differentiation and the chain-rule to equation 14 will result in higher-order traveltime source-derivatives. For example, the second-order derivative reads:

$$\begin{aligned}\frac{\partial^2 T}{\partial\mathbf{x_s^2}}\equiv\frac{\partial^2\hat{T}}{\partial\mathbf{x_s^2}} &= \frac{\partial}{\partial\mathbf{x}}\frac{\partial\hat{T}}{\partial\mathbf{x}}\cdot\frac{\partial\mathbf{x}}{\partial\mathbf{x_s}}+\frac{\partial}{\partial\mathbf{q}}\frac{\partial\hat{T}}{\partial\mathbf{x}}\cdot\frac{\partial\mathbf{q}}{\partial\mathbf{x_s}}-\frac{\partial}{\partial\mathbf{x}}\frac{\partial\hat{T}}{\partial\mathbf{q}}\cdot\frac{\partial\mathbf{x}}{\partial\mathbf{x_s}}-\frac{\partial}{\partial\mathbf{q}}\frac{\partial\hat{T}}{\partial\mathbf{q}}\cdot\frac{\partial\mathbf{q}}{\partial\mathbf{x_s}}\\ &= \frac{\partial^2\hat{T}}{\partial\mathbf{x}^2}-2\frac{\partial^2\hat{T}}{\partial\mathbf{x}\partial\mathbf{q}}+\frac{\partial^2\hat{T}}{\partial\mathbf{q}^2}\ .\end{aligned} \tag{7}$$

Further, differentiating equation 6 once more by $\mathbf{x}$ provides

$$\nabla_{\mathbf{q}}\frac{\partial\hat{T}}{\partial\mathbf{x}}\cdot\nabla_{\mathbf{q}}\frac{\partial\hat{T}}{\partial\mathbf{x}}+\nabla_{\mathbf{q}}\hat{T}\cdot\nabla_{\mathbf{q}}\frac{\partial^2\hat{T}}{\partial\mathbf{x}^2}=\frac{1}{2}\frac{\partial^2 W}{\partial\mathbf{x}^2}\ . \tag{8}$$

It is easy to verify that any order of the traveltime source-derivative will require the corresponding order of the slowness-squared derivative. An approximation based on Taylor expansions of the traveltime around a fixed source location can make use of these derivatives. For example, (29) and (7) introduced parabolic and hyperbolic traveltime approximations with the first- and second-order derivatives. Notice that the need for slowness-squared derivatives may cause instability unless the velocity model is sufficiently smooth. (5) also proved the following relationship between $\partial W/\partial\mathbf{x}$ and $\partial\hat{T}/\partial\mathbf{q}$:

$$\nabla_{\mathbf{q}}\hat{T}\cdot\nabla_{\mathbf{q}}\frac{\partial\hat{T}}{\partial(\mathbf{q}+\mathbf{x_s})}=\frac{1}{2}\frac{\partial W}{\partial\mathbf{x}}\ , \tag{9}$$

which implies that the traveltime source-derivative can be computed from the given traveltime tables only. However, the velocity smoothness is still implicitly assumed as the second-order spatial derivatives of traveltimes appear in the equation. For this reason, we restrict our current implementation to the first-order derivative only.

In a ray-tracing eikonal solver, $\partial T/\partial \mathbf{x_s}$ is the slowness vector of a particular ray at $\mathbf{x_s}$ and holds constant along the trajectory. As it may also require irregular coordinate mappings, one may use the same strategy as for the traveltime tables. In this way, there is no necessity for any additional effort. On the other hand, equations 14 and 6 and their second-order extensions provide important attributes for use in Gaussian beams, which are commonly calculated by the dynamic ray tracing (32). They might be alternatively estimated by the eikonal-based source-derivative formulas but with the traveltime tables from a finite-difference eikonal solver. However, this application is beyond the scope of this paper. In the following sections, we consider only the source-derivative estimation from traveltimes computed by a finite-difference eikonal solver.

Numerical Implementation

Equation 6 is a linear first-order PDE suitable for upwind numerical methods (14). Since it does not change the non-linear nature of the eikonal equation, the resulting traveltime source-derivative can be related to any branch of multi-arrivals, if one supplies the corresponding traveltime in $\hat{T}$. The source-derivatives can be computed either along with traveltimes or separately. In Appendix A, we describe a first-arrival implementation based on a modification of FMM (25).

The first-order traveltime source-derivative enables a cubic Hermite interpolation (22). Geometrically, such an interpolation is valid only when the selected wave-front in the interpolation interval is smooth and continuous. For a 2D model and a source interpolation along the inline direction only, the Hermite interpolation reads:

$$\begin{aligned} T(z,x;z_s,x_s+\alpha\Delta x_s) &= (2\alpha^3-3\alpha^2+1)\,T(z,x;z_s,x_s) \\ &+ (\alpha^3-2\alpha^2+\alpha)\,\tfrac{\partial T}{\partial x_s}(z,x;z_s,x_s) \\ &+ (-2\alpha^3+3\alpha^2)\,T(z,x;z_s,x_s+\Delta x_s) \\ &+ (\alpha^3-\alpha^2)\,\tfrac{\partial T}{\partial x_s}(z,x;z_s,x_s+\Delta x_s)\;, \end{aligned} \tag{10}$$

where $\alpha \in [0,1]$ controls the source position to be interpolated between known values at (z_s,x_s) and $(z_s,x_s+\Delta x_s)$. For comparison, the linear interpolation can be represented by:

$$T(z,x;z_s,x_s+\alpha\Delta x_s) = (1-\alpha)\,T(z,x;z_s,x_s) + \alpha\,T(z,x;z_s,x_s+\Delta x_s)\;. \tag{11}$$

The linear interpolation fixes the subsurface image point (z,x). A possible improvement is to instead fix the vector that links the source with the image, such that on the right-hand side the traveltimes are taken at shifted image locations:

$$\begin{aligned} T(z,x;z_s,x_s+\alpha\Delta x_s) &= (1-\alpha)\,T(z,x-\alpha\Delta x_s;z_s,x_s) \\ &+ \alpha\,T(z,x+(1-\alpha)\Delta x_s;z_s,x_s+\Delta x_s)\;. \end{aligned} \tag{12}$$

We will refer to scheme 12 as shift interpolation. According to our definition of the relative coordinate $\mathbf{q}$ in equation 2, shift interpolation amounts to a linear interpolation in $\hat{T}(\mathbf{q};\mathbf{x_s})$. It is easy to verify that, for a constant-velocity medium, both

Hermite and shift interpolations are accurate, while the linear interpolation is not. However, the accuracy of shift interpolation deteriorates with increasing velocity variations, as it assumes that the wave-front remains invariant in the relative coordinate. Equations 10-12 can be generalized to 3D by cascading the inline and crossline interpolations (for example equation A-1 in 3D case becomes bilinear interpolation). The interpolated source does not need to lie collinear with source samples.

The derivatives themselves can also be directly used for Kirchhoff anti-aliasing (19; 1; 11). Equations 10, A-1 and 12 give rise to their corresponding source-derivative interpolations after applying the following chain-rule to both sides:

$$\frac{\partial}{\partial(x_s+\alpha\Delta x_s)} = \frac{\partial}{\partial\alpha}\frac{\partial\alpha}{\partial(x_s+\alpha\Delta x_s)} = \frac{1}{\Delta x_s}\frac{\partial}{\partial\alpha}\ . \tag{13}$$

The anti-aliasing application is summarized in Appendix B.

NUMERICAL EXAMPLES

Constant-velocity-gradient Model

In a 2D medium of linearly changing velocities, $v(z,x) = v_0 + ax + bz$ where x is the lateral position and z is the depth, the traveltimes and source-derivatives have analytical solutions (27). Figure 1(a) shows the model used in our numerical test and the analytical source-derivative for a source located at $(0,0)$ km. The domain is of size 4km $\times$ 4km with grid spacing 0.01 km in both directions. We solve for the traveltime tables at five sources of uniform spacing 1 km along the top domain boundary by FMM and their associated source-derivatives using the method described in Appendix A. Figure 6(e) compares the errors in computed source-derivative between the proposed approach and a centered second-order finite-difference estimation for the same source shown in Figure 1(a). The proposed method is sufficiently accurate except for the small region around the source. This is due to the source singularity of the eikonal equation and can be improved by adaptive or high-order upwind finite-difference methods (23) or by factoring the singularity (12). Since we are aiming at using the interpolated traveltime tables for migration purposes and the reflection energy around the sources is usually low, these errors in current implementation can be neglected. In Figure 3, we interpolate the traveltime table for a source at location $(0,0.25)$ km from the nearby source samples at $(0,0)$ km and $(0,1)$ km by the cubic Hermite, linear and shift interpolations. We use the eikonal-based source-derivative in the cubic Hermite interpolation. The shift interpolation is not applicable for some $\mathbf{q}$ and $\mathbf{x_s}$ if $\mathbf{x} = \mathbf{q} + \mathbf{x_s}$ is beyond the computational domain. In these regions, we use a linear interpolation to fill the traveltime table. As expected, the cubic Hermite interpolation achieves the best result, while its misfits near the source are related to the errors in source-derivatives. The shift interpolation performs generally better than the linear interpolation, especially in the regions close to the source where the wave-fronts are simple.

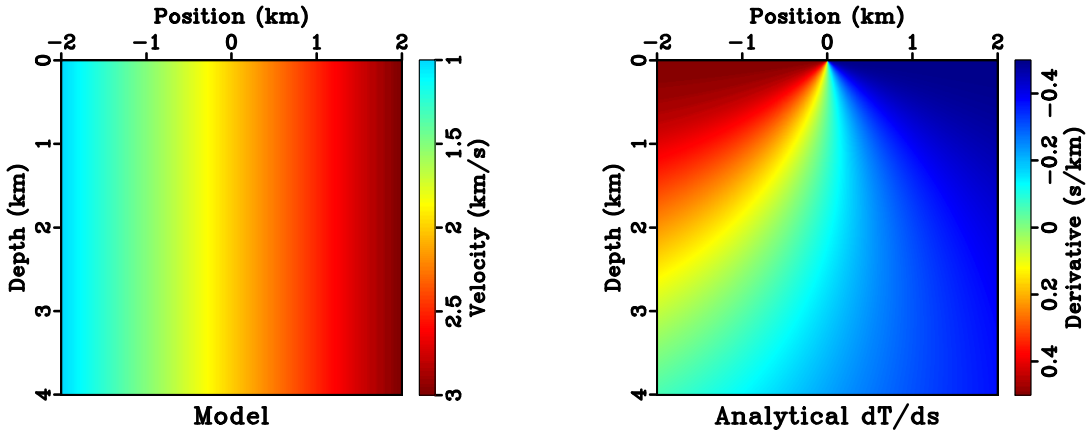


Figure 1: (Left) a constant-velocity-gradient model $v(z,x) = 2 + 0.5x$ km/s and (right) its analytical traveltime source-derivative for a source at origin $\mathbf{x_s} = (0,0)$ km. eikods/check model

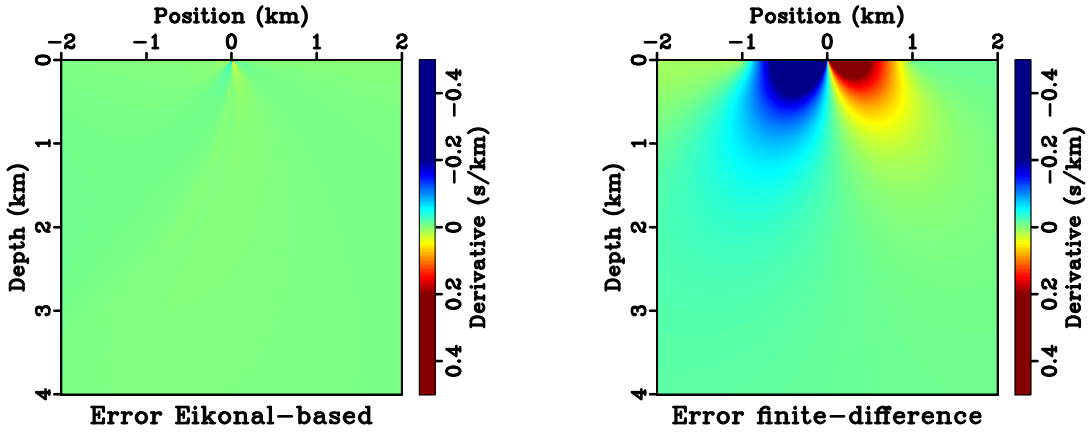


Figure 2: Comparison of error in computed source-derivative by (left) the proposed method and (right) a centered second-order finite-difference estimation based on traveltime tables. The maximum absolute errors are 0.15 s/km and 0.56 s/km, respectively. eikods/check diff

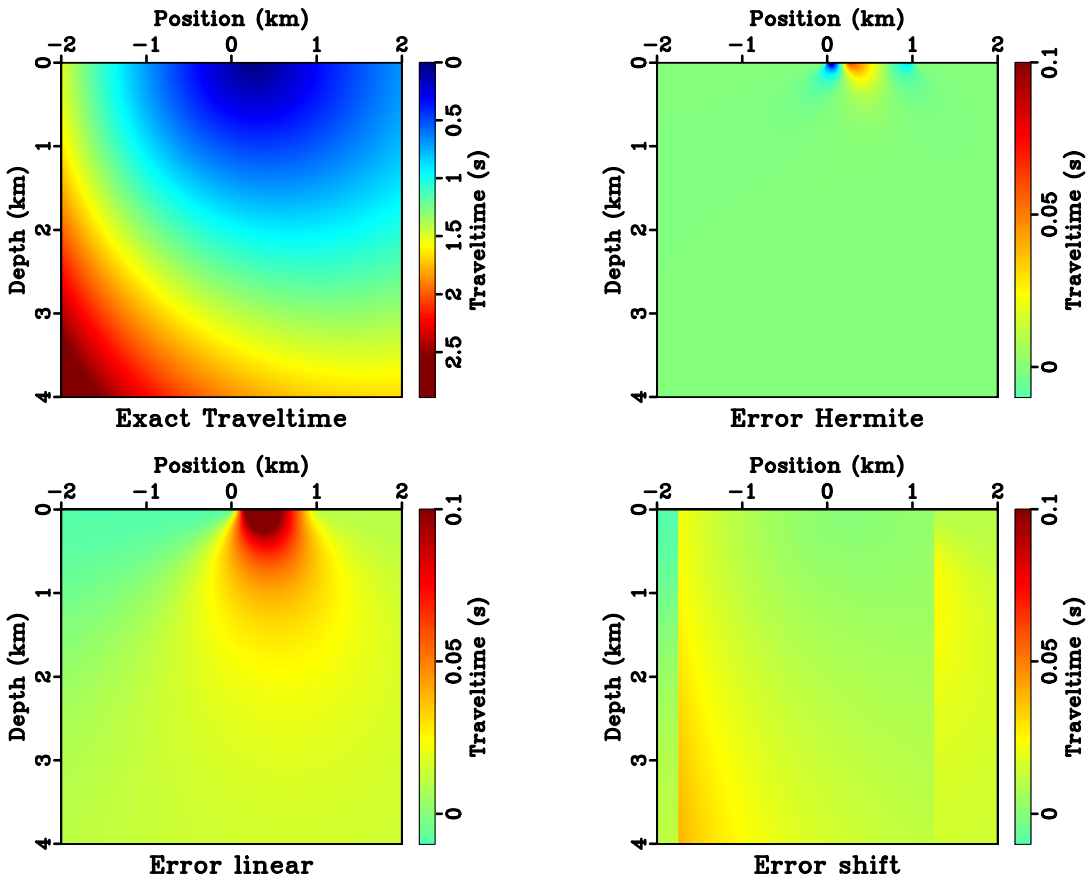


Figure 3: Traveltime interpolation error of three different schemes: (top left) the analytical traveltime of a source at location $(0, 0.25)$ km; (top right) error of the cubic Hermite interpolation; (bottom left) error of the linear interpolation; (bottom right) error of the shift interpolation. Using derivatives in interpolation enables a significantly higher accuracy. The $l2$ norm of the error are 1.5 s, 9.2 s and 6.0 s respectively. eikods/check ierror

The difference between a cubic Hermite interpolation and a linear or shift one is in the usage of source-derivatives. In this regard, one may think of supplying the finite-difference estimated derivatives to the interpolation. Indeed, a refined source sampling and higher-order differentiation may lead to more accurate derivatives. However the additional computation is considerable. For the same model in Figure 1(a), we carry out both a source sampling refinement experiment and a model grid spacing refinement experiment. The results are shown in Figures 4 and 5. Both figures are plotted for the traveltime at subsurface location $(1.5, -0.5)$ km for the source at location $(0, 0)$ km. Although the curves vary for different locations, the source sampling refinement experiment suggests the general need for approximately three times finer source-sampling than that of Figure 6(e) to achieve the same level of accuracy.

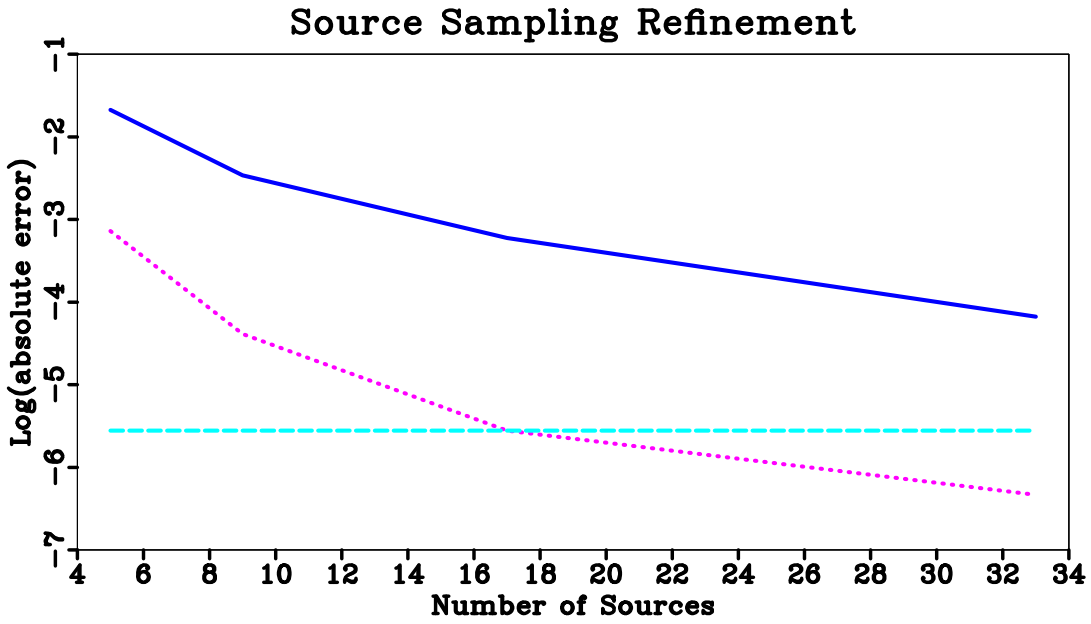


Figure 4: Source-sampling refinement experiment. The plot shows, at a fixed model grid sampling of 0.01 km and increasing source sampling, the error in source-derivative estimated by a first-order finite-difference (solid) and a centered second-order finite-difference scheme (dotted) decrease. The horizontal axis is the number of sources and the source sampling is uniform. The vertical axis is the natural logarithm of the absolute error. The flat line (dash) is from the proposed eikonal-based method and is source-sampling independent. eikods/check sfddiff

Kirchhoff migration can use traveltime source-derivatives in two ways: for traveltime interpolation when the source and receiver of a trace does not lie on the source grid of pre-computed traveltime tables, and for anti-aliasing. Figure 5 shows a synthetic model of constant-velocity-gradient with five dome-shaped reflectors. The model has a 0.01 km grid spacing in both directions. We solve for traveltimes and source-derivatives by the modified FMM introduced in Appendix A at 21 sparse shots of uniform spacing 0.5 km, and migrate synthetic zero-offset data. The interpolation

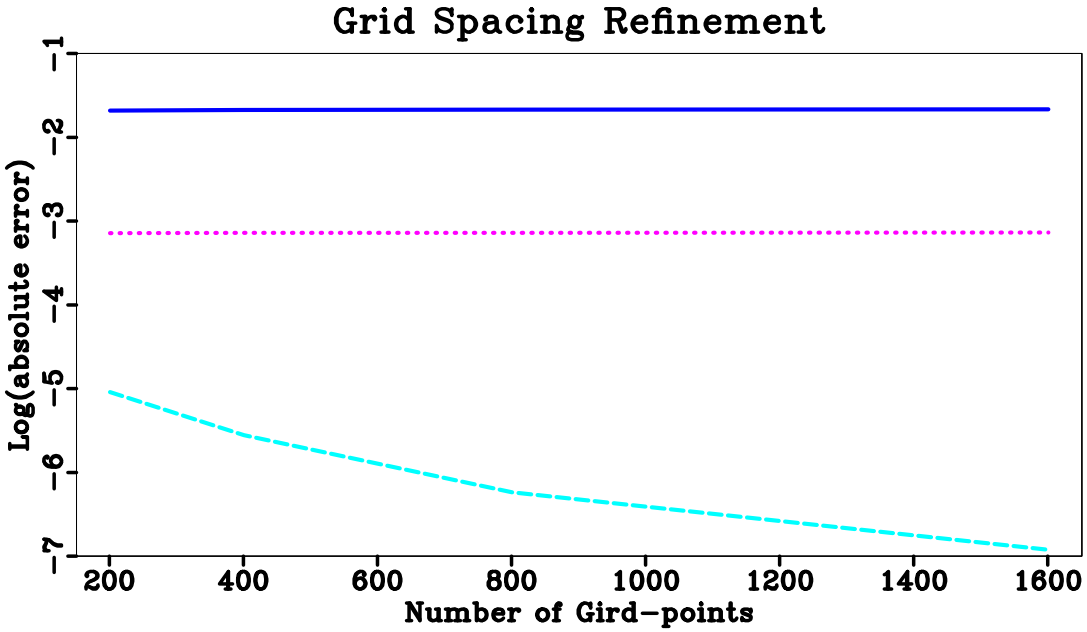


Figure 5: Gird-spacing refinement experiment. The plot shows, at a fixed source sampling of 1 km and increasing model grid sampling, the error in source-derivative estimated by the proposed eikonal-based method decreases. Meanwhile, the errors of both first- and second-order finite-difference estimations do not improve noticeably. The horizontal axis is the number of grid points in both directions and the grid sampling is uniform. See Figure 4 for descriptions of the vertical axis and the lines. eikods/check gfddiff

of source-derivative for the anti-aliasing purpose follows the method described in Appendix B. 48 interpolations are carried out within each sparse source sampling interval. Figures 7 and 8 compare the images obtained by three different interpolations and the effect of anti-aliasing. All images are plotted at the same scale. We do not limit migration aperture for all cases and adopt the anti-aliasing criteria suggested by (1) to filter the input trace before mapping a sample to the image, where the source-derivative and receiver-derivative (in the zero-offset case they coincide) determine the filter coefficients. As expected, the cubic Hermite interpolation with anti-aliasing leads to the most desirable image. The image could be further improved by considering not only the kinematics predicted by the traveltimes but also the amplitude factors (9; 31).

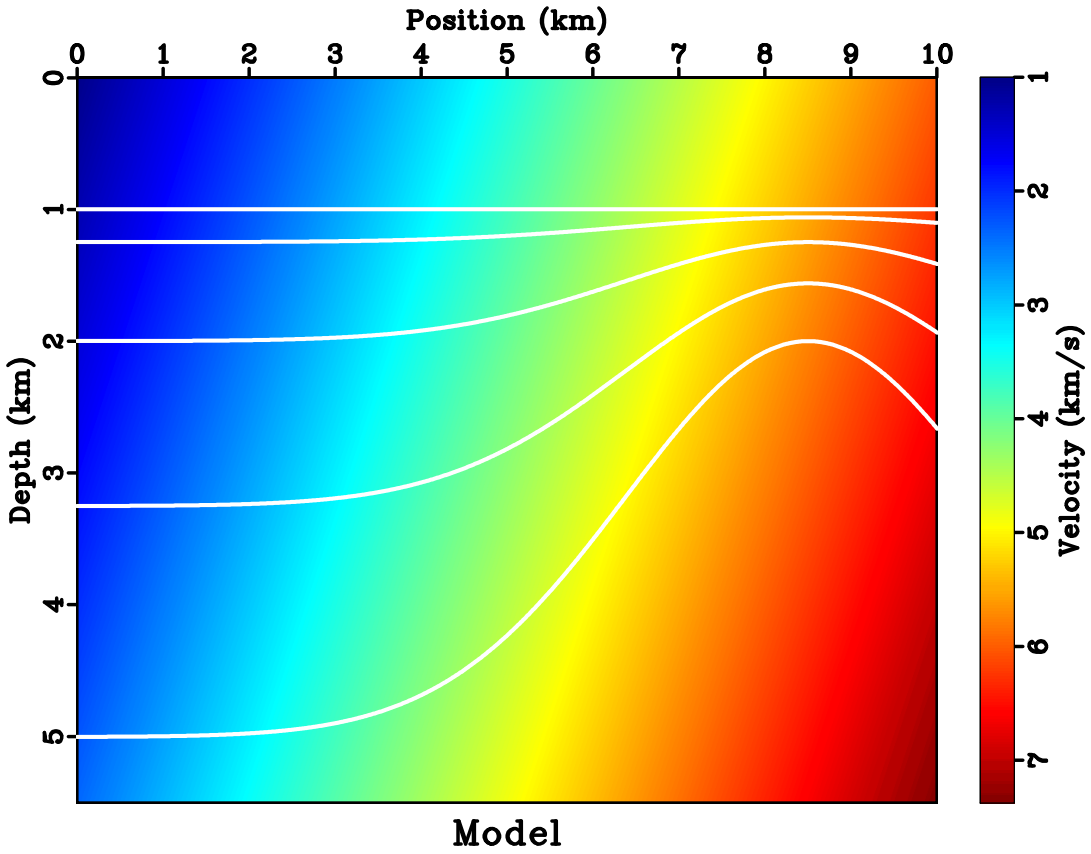


Figure 6: Constant-velocity-gradient background model $v(z,x) = 1.5 + 0.25z + 0.25x$ km/s with dome shaped reflectors. eikods/migration modl

Marmousi Model

The Marmousi model (33) has large velocity variations and is challenging for Kirchhoff migration with first-arrivals (15). We apply a single-fold 2D triangular smoothing of radius 20 m to the original model (see Figure 13(a)) to remove only sharp veloc-

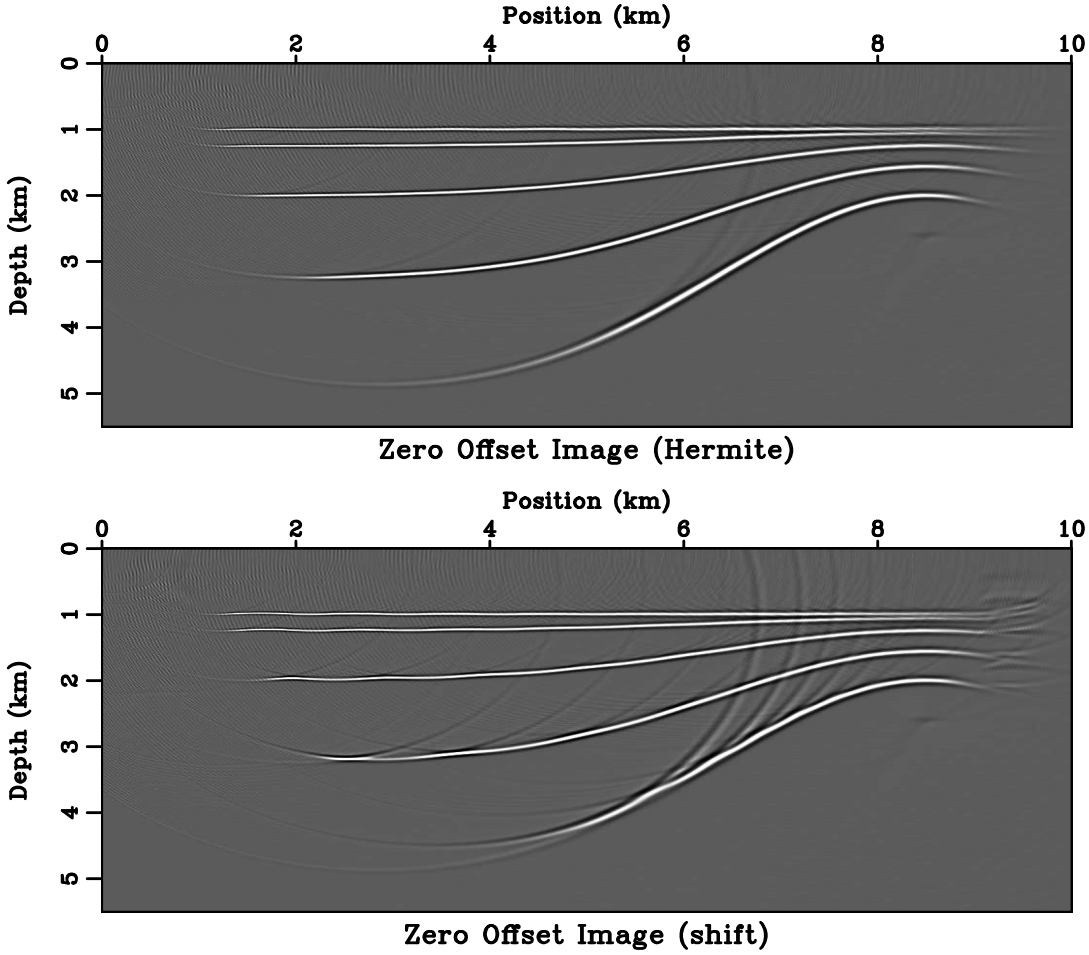


Figure 7: Zero-offset Kirchhoff migration image with (top) the cubic Hermite interpolation and (bottom) the shift interpolation. eikods/migration hzodmig

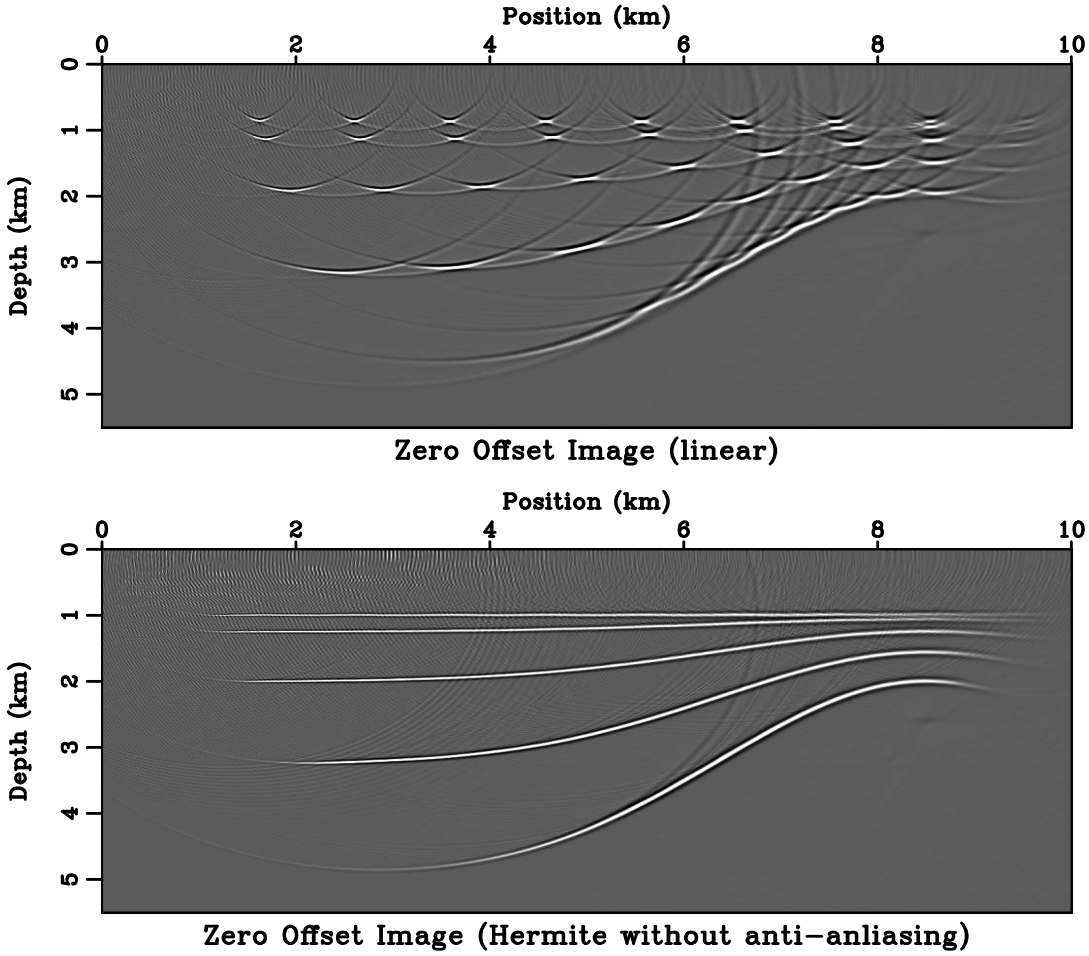


Figure 8: Zero-offset Kirchhoff migration image with (top) the linear interpolation and (bottom) the cubic Hermite interpolation without anti-aliasing. eikods/migration lzodmig

ity discontinuities but retain the complex velocity structures. Because wave-fronts change shapes rapidly, the traveltime interpolation may be subject to inaccurate source-derivatives and provide less satisfying accuracy compared to that in a simple model. Although the derivative computation in the proposed eikonal-based method is source-sampling independent, in practice we should limit the interpolation interval to be sufficiently small, so that the traveltime curve could be well represented by a cubic spline. For the smoothed Marmousi model, we use a sparse source sampling of 0.2 km based on observations of the horizontal width of major velocity structures. Figures 13(a) and 10(a) compare the traveltime interpolation errors of three methods as in Figure 3 for a source located at $(0, 3.1)$ km from nearby source samples at $(0, 3)$ km and $(0, 3.2)$ km. Figure 11 plots a reference traveltime curve for the fixed subsurface location $(2, 3.3)$ km computed by a dense eikonal solving of 4 m source spacing against curves produced by the interpolations. While these comparisons vary between different source intervals and subsurface locations, the cubic Hermite interpolation out-performs the linear and the shift interpolations except for the source singularity region. However in Figure 13(a) the errors are relatively large in the upper-left region. These errors occur due to the collapse of overlapping branches of the traveltime field (34) that causes wave-front discontinuities and undermines the assumptions of the proposed method.

One strategy for imaging multi-arrival wavefields with first-arrival traveltimes is the semi-recursive Kirchhoff migration (6). It breaks the image into several depth intervals, applies Kirchhoff redatuming to the next interval, performs Kirchhoff migration from there, and so on. The small redatuming depth effectively limits the maximum traveltime and the evolving of complex waveforms before the most energetic arrivals separate from first-arrivals. Since Kirchhoff redatuming also relies on traveltimes between datum levels, our method can be fully incorporated into the whole process. Again, for simplicity, we do not consider amplitude factors during migration. We use the Marmousi dataset with a source/receiver sampling of 25 m. Due to the source and receiver reciprocity, the receiver side interpolations are equivalent to those on the source side. Figure 12 is the result of a Kirchhoff migration with eikonal solvings at each source/receiver location, i.e. no interpolation performed. Only the upper portion is well imaged. Figure 13 shows the image after employing the cubic Hermite interpolation with a 0.2 km sparse source/receiver sampling, which means 7 source interpolations within each interval. Even though a 7 times speed-up is not attainable in practice due to the extra computations in source-derivative and interpolation, we are still able to gain an approximately 5-fold cost reduction in traveltime computations, while keeping the image quality comparable between Figures 12 and 13. Next, following (6), we downward continue the data to a depth of 1.5 km in three datuming steps. The downward continued data are then Kirchhoff migrated and combined with the upper portion of Figure 13. We keep the same 0.2 km sparse source/receiver sampling whenever eikonal solvings are required in this process. Figure 14 shows the image obtained by the semi-recursive Kirchhoff migration. The target zone at approximately $(2.5, 6.5)$ km appears better imaged.

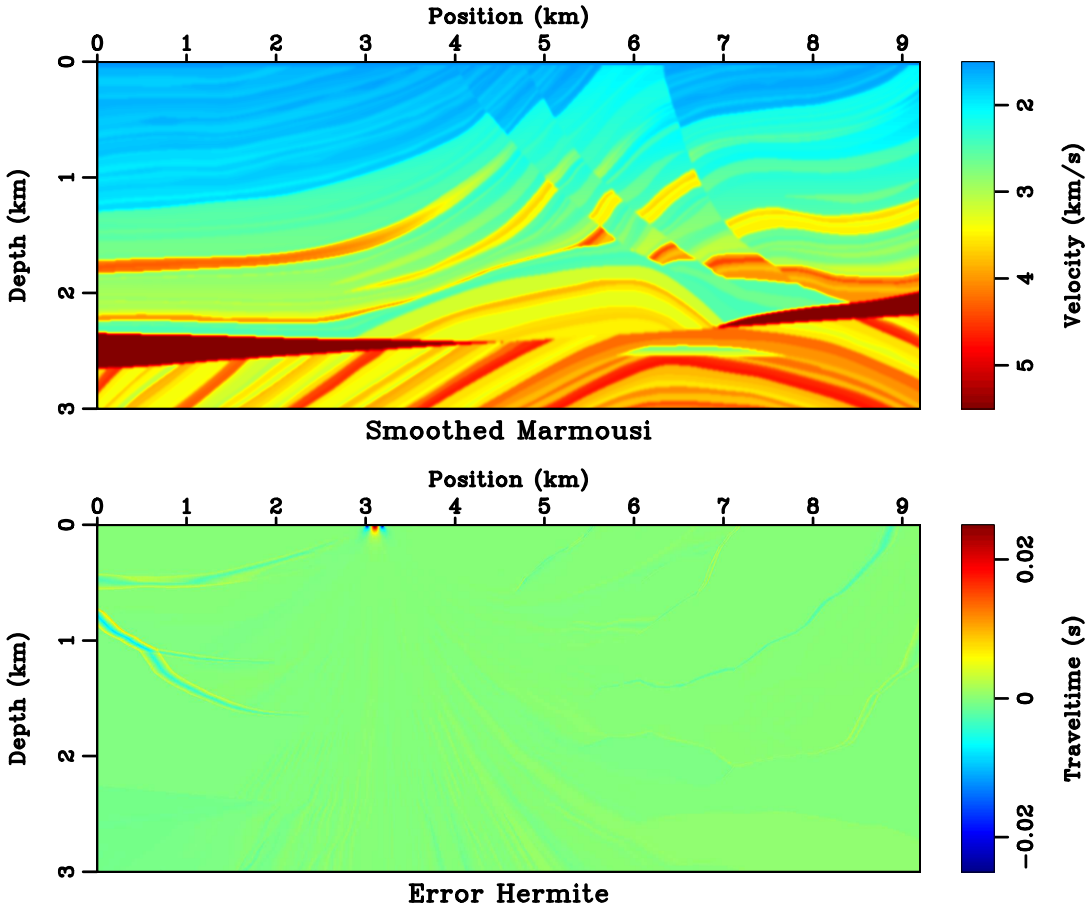


Figure 9: (Top) the smoothed Marmousi model. The model has a 4 m fine grid. (Bottom) the traveltime error by the cubic Hermite interpolation. eikods/marm vel

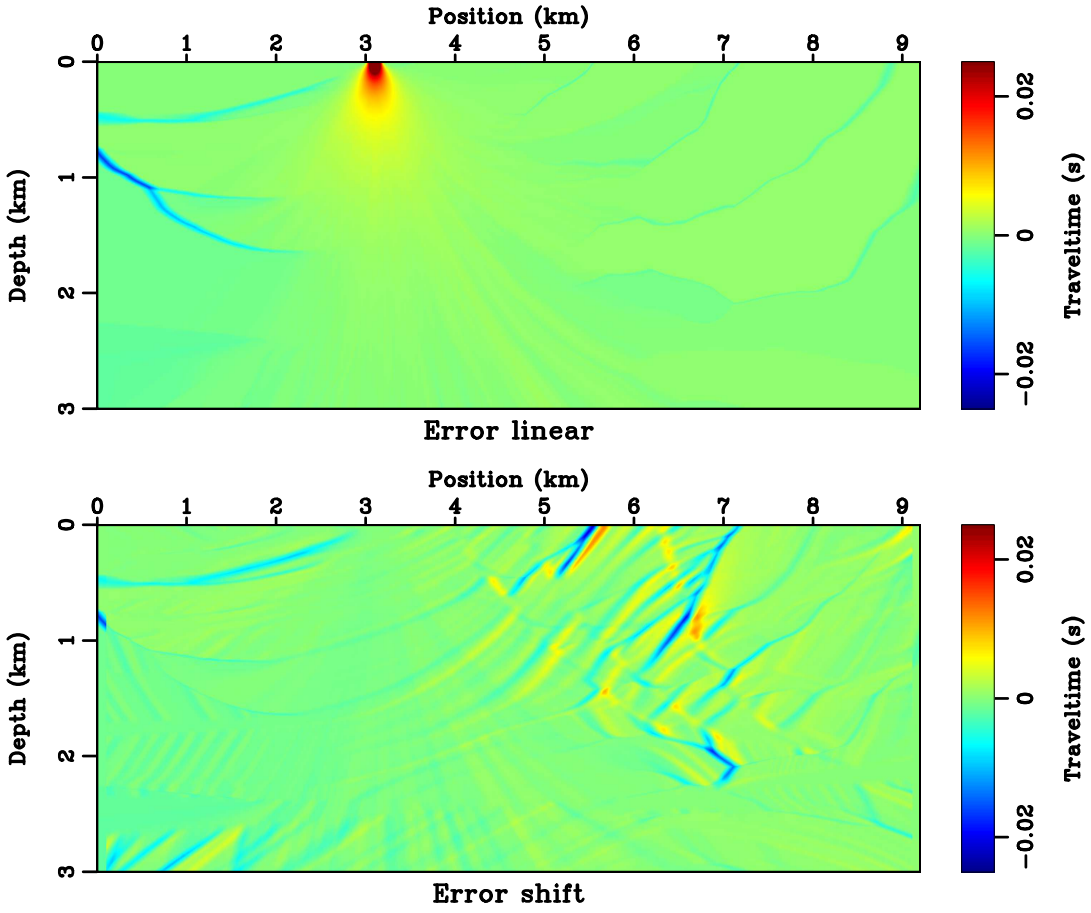


Figure 10: The traveltime error by (top) the linear interpolation and (bottom) the shift interpolation. eikods/marm slice

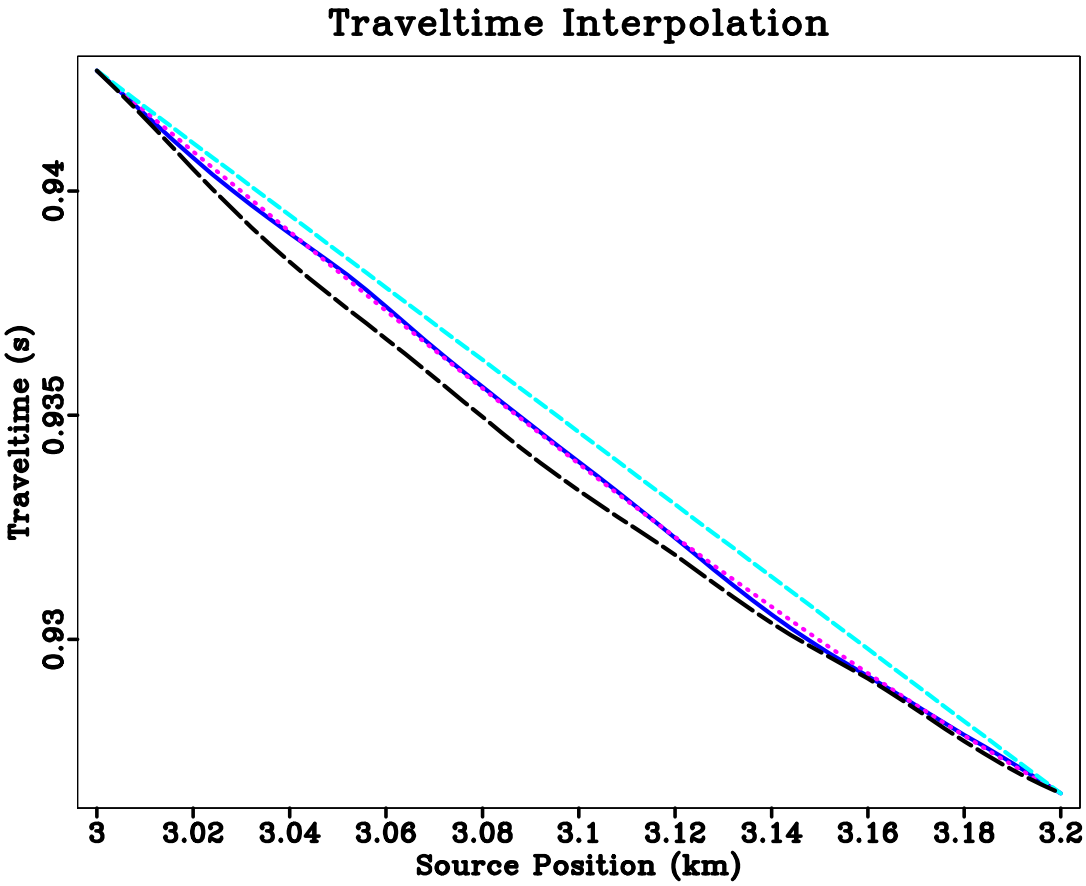


Figure 11: Traveltime interpolation for a fixed subsurface location. Compare between the result from a dense source sampling (solid blue), cubic Hermite interpolation (dotted magenta), linear interpolation (dashed cyan) and shift interpolation (dashed black). The $l2$ norm of the error (against the dense source sampling results) of 49 evenly interpolated sources between interval $(0, 3)$ km and $(0, 3.2)$ km for all locations but the top 100 m source singularity region are 3.9 s, 9.2 s and 11.6 s respectively. eikods/marm curve

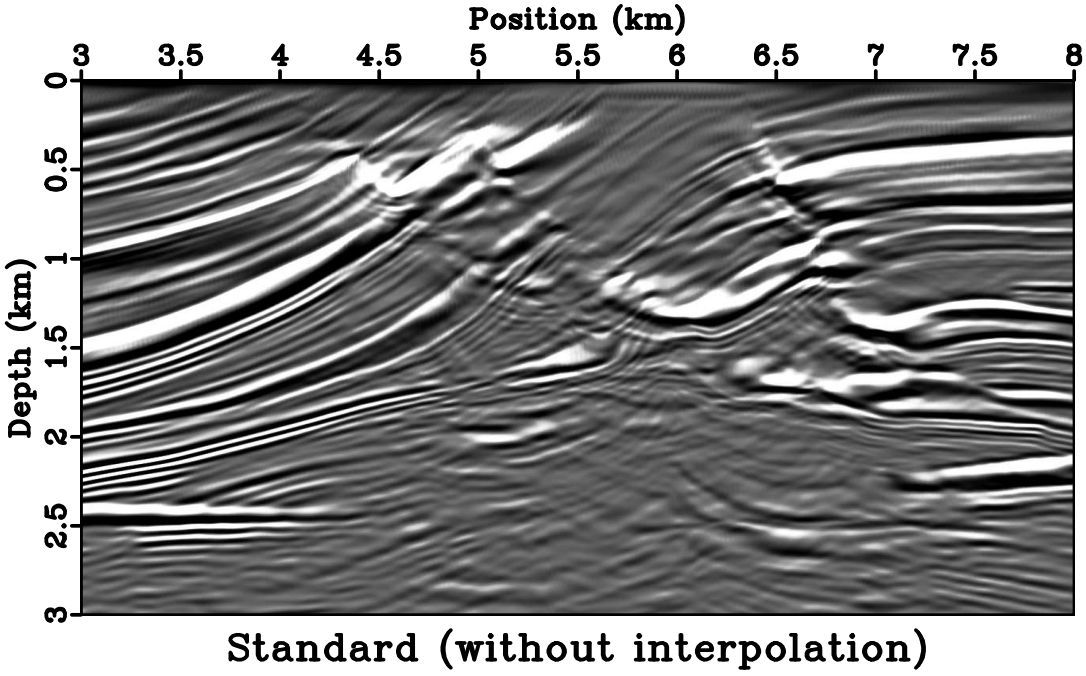


Figure 12: Image of Kirchhoff migration with first-arrivals (no interpolation). eikods/marm dmig0d

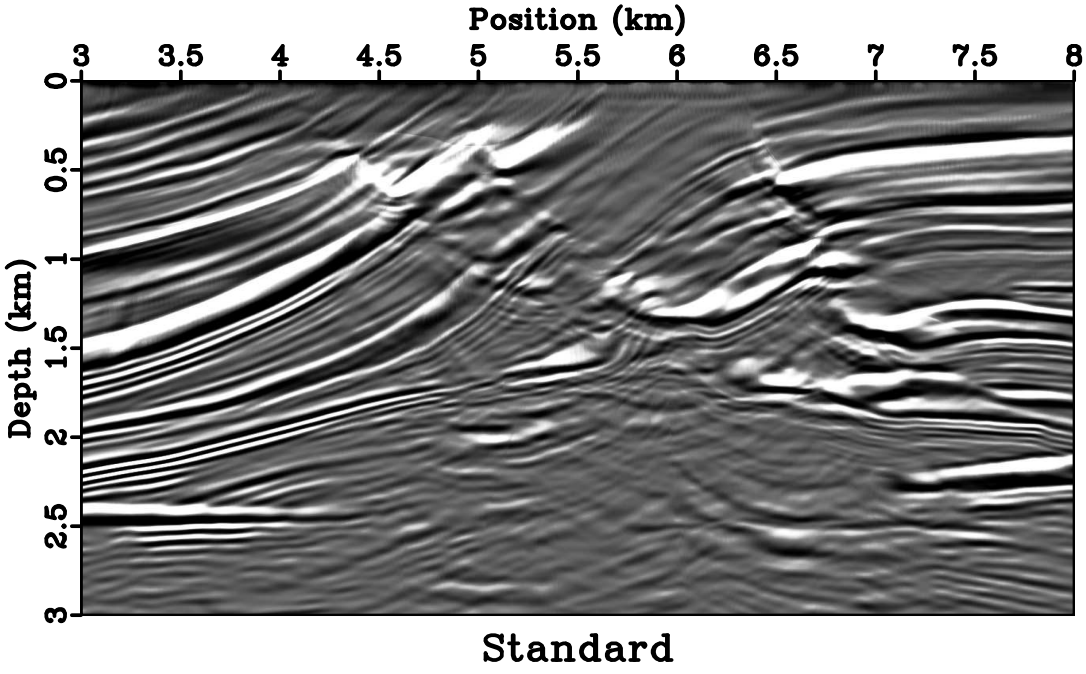


Figure 13: Image of Kirchhoff migration with first-arrivals and a sparse source/receiver sampling. eikods/marm dmig0

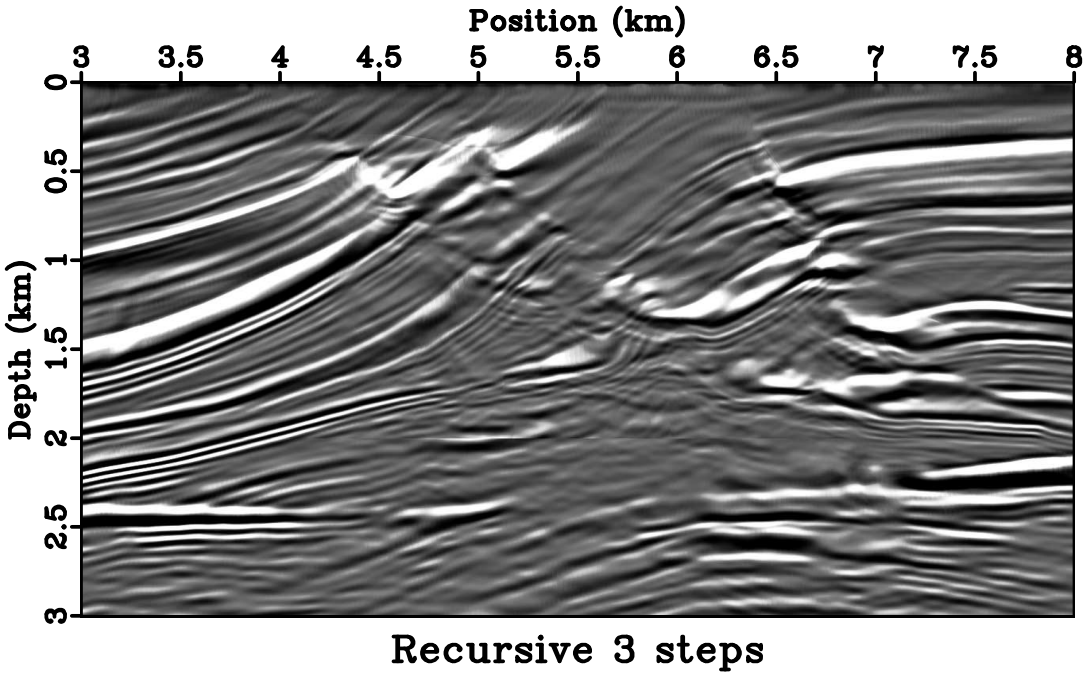


Figure 14: Image of semi-recursive Kirchhoff migration with a three-step redatuming from top surface to 1.5 km depth and a 0.5 km interval each time. The sparse source/receiver sampling is the same as in Figure 13. eikods/marm dmig2

DISCUSSION

The proposed approach could be implemented either along with a finite-difference eikonal solver or separately. Our current implementation outputs both traveltime and source-derivative at the same time, with a roughly 30% extra cost per eikonal solve compared to a FMM solver without the source-derivative functionality. An interpolation with these source-derivatives is superior to the ones without and thus enables an accurate traveltime-table compression. For 3D datasets, as both inline and crossline directions may benefit from the source-derivative and interpolation, the overall data compression could be significant. For instance, interpolating 10 shots within each sparse source sampling interval in both inline and crossline directions leads to an approximately 100-fold savings in traveltime storage. The method could be further combined with an interpolation within each source, for example from a coarse grid to a fine grid, for a greater data compression.

While our implementation is for first-arrivals only, the governing equations are valid also for other characteristic branches, for example the most energetic arrivals. However, an underlying assumption of the proposed method is a continuous change in the wave-front of selected arrivals within individual sources. For first-arrivals, this condition always holds valid. However, the most energetic wave-front can be more complicated than that of first-arrival, for example only piece-wise continuous, which

may lead to a potential degradation in accuracy. For example, (21) showed the most energetic wave-fronts in the Marmousi model. Another assumption is that the traveltime source-derivatives are continuous between nearby sources. This condition breaks down when multi-pathing takes place. (30) suggested to smooth traveltimes around the discontinuities in order to overcome this limitation. In theory, one can try to identify the discontinuities and only perform interpolation within individual continuous pieces by using the eikonal-based source-derivatives. By doing so, one should be able to recover branch jumping in interpolated traveltimes, but only for those locations within the identified continuous pieces. For the discontinuities themselves as well as the gaps between them, additional eikonal solving may be required. An efficient implementation of this strategy remains open for future research.

CONCLUSION

We have shown an application of computing traveltime source-derivatives in Kirchhoff migration. For first-arrivals, a cubic Hermite traveltime interpolation using the first-order source-derivatives speeds up computation and reduces storage without noticeably sacrificing accuracy. Anti-aliasing is another direct application of traveltime source-derivatives that can be easily incorporated into Kirchhoff migration.

Generalization of the method to 3D is straightforward. The computed derivative attributes may benefit other areas besides the kinematic-only Kirchhoff migration shown in this paper. An extension to multi-arrival traveltimes needs further investigation.

ACKNOWLEDGMENTS

We thank the editors, Samuel Gray, and three anonymous reviewers for constructive suggestions that helped improving the quality of the manuscript. We thank Tariq Alkhalifah and Alexander Vladimirsky for useful discussions and sponsors of the Texas Consortium for Computational Seismology (TCCS) for financial support of this research. This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

FMM IMPLEMENTATION OF SOURCE-DERIVATIVES

The FMM is a non-iterative eikonal solver with $O(N \log N)$ complexity, where N is the total number of grid points of the discretized domain. It relies on a heap data structure to keep the updating sequence, and a local one-sided upwind finite-difference scheme for ensuring the causality (25). Consider in 3D a cubic domain discretized

into Cartesian grids, with uniform grid size of $(\Delta x, \Delta y, \Delta z)$. Let $\hat{T}^k_{i,j}$ be the traveltime value at vertices $\mathbf{x}^k_{i,j} = (x_i, y_j, z_k)$ and define difference operator $D^{\pm}_x$ for x direction as

$$D^{\pm}_x \hat{T}^k_{i,j} = \pm \frac{\hat{T}^k_{i\pm1,j} - \hat{T}^k_{i,j}}{\Delta x} \ , \tag{A-1}$$

The causality condition requires picking an upwind neighbor in all directions at $\mathbf{x}^k_{i,j}$.

$$\hat{D}_x \hat{T}^k_{i,j} = \max\left(D^-_x \hat{T}^k_{i,j}, -D^+_x \hat{T}^k_{i,j}, 0 \right) \ . \tag{A-2}$$

After similar definitions for $\hat{D}_y$ and $\hat{D}_z$, the local upwind scheme in FMM for equation 3 reads

$$\left(\hat{D}_x \hat{T}^k_{i,j}\right)^2 + \left(\hat{D}_y \hat{T}^k_{i,j}\right)^2 + \left(\hat{D}_z \hat{T}^k_{i,j}\right)^2 = W^k_{i,j} \ . \tag{A-3}$$

For $\partial\hat{T}/\partial\mathbf{x}$ in equation 6 and $\partial\hat{T}/\partial\mathbf{q}$ in equation 14, we can apply the same upwind strategy:

$$\hat{D}_x \hat{T}^k_{i,j} \cdot \hat{D}_x \left(\frac{\partial \hat{T}}{\partial \mathbf{x}}\right)^k_{i,j} + \hat{D}_y \hat{T}^k_{i,j} \cdot \hat{D}_y \left(\frac{\partial \hat{T}}{\partial \mathbf{x}}\right)^k_{i,j} + \hat{D}_z \hat{T}^k_{i,j} \cdot \hat{D}_z \left(\frac{\partial \hat{T}}{\partial \mathbf{x}}\right)^k_{i,j} = \frac{1}{2} \left(\frac{\partial W}{\partial \mathbf{x}}\right)^k_{i,j} \ , \tag{A-4}$$

$$\left(\frac{\partial \hat{T}}{\partial \mathbf{q}}\right)^k_{i,j} = \hat{D}_{\mathbf{q}} \hat{T}^k_{i,j}, \ \mathbf{q} = (x, y, z) \ . \tag{A-5}$$

where in equation A-4 $\hat{D}_x$, $\hat{D}_y$ and $\hat{D}_z$ are chosen according to $\hat{T}^k_{i,j}$, regardless of $\partial\hat{T}/\partial\mathbf{x}$. Finally,

$$\left(\frac{\partial T}{\partial \mathbf{x_s}}\right)^k_{i,j} = \left(\frac{\partial \hat{T}}{\partial \mathbf{x}}\right)^k_{i,j} - \left(\frac{\partial \hat{T}}{\partial \mathbf{q}}\right)^k_{i,j} \ . \tag{A-6}$$

To incorporate the computation of traveltime source-derivatives into FMM, one only needs to add equations A-4, A-5 and B-5 after A-3. An extra upwind sorting and solving after pre-computing $\hat{T}$ is not necessary. The total complexity of FMM with the auxiliary output of traveltime source-derivative remains $O(N \log N)$.

APPENDIX B

INTERPOLATION OF SOURCE-DERIVATIVES

Applying the chain-rule 13 to equation 10, we arrive at the interpolation equation for source-derivatives in the cubic Hermite scheme:

$$\begin{aligned} \Delta x_s \, \tfrac{\partial T(z,x;z_s,x_s+\alpha\Delta x_s)}{\partial(x_s+\alpha\Delta x_s)} &= (6\alpha^2 - 6\alpha)\, T(z,x;z_s,x_s) \\ &+ (3\alpha^2 - 4\alpha + 1)\, \tfrac{\partial T}{\partial x_s}(z,x;z_s,x_s) \\ &+ (-6\alpha^2 + 6\alpha)\, T(z,x;z_s,x_s+\Delta x_s) \\ &+ (3\alpha^2 - 2\alpha)\, \tfrac{\partial T}{\partial x_s}(z,x;z_s,x_s+\Delta x_s) \ . \end{aligned} \tag{B-1}$$

Analogously, the interpolation of source-derivatives in the linear scheme A-1 reads:

$$\Delta x_s \, \frac{\partial T(z,x;z_s,x_s+\alpha\Delta x_s)}{\partial(x_s+\alpha\Delta x_s)} = -T(z,x;z_s,x_s) + T(z,x;z_s,x_s+\Delta x_s) \; . \tag{B-2}$$

which is a simple first-order finite-difference estimation. Finally, in the case of shift scheme 12, the partial derivative $\partial/\partial\alpha$ must be applied to the shifted traveltime terms at the same time:

$$\begin{aligned} \Delta x_s \, \frac{\partial T(z,x;z_s,x_s+\alpha\Delta x_s)}{\partial(x_s+\alpha\Delta x_s)} &= -T(z,x-\alpha\Delta x_s;z_s,x_s) \\ &- (1-\alpha)\Delta x_s \, \frac{\partial T(z,x-\alpha\Delta x_s;z_s,x_s)}{\partial(x-\alpha\Delta x_s)} \\ &+ T(z,x+(1-\alpha)\Delta x_s;z_s,x_s+\Delta x_s) \\ &- \alpha\Delta x_s \, \frac{\partial T(z,x+(1-\alpha)\Delta x_s;z_s,x_s+\Delta x_s)}{\partial(x+(1-\alpha)\Delta x_s)} \; . \end{aligned} \tag{B-3}$$

The required spatial derivatives can be estimated from the traveltime table by means of finite-differences, for example by using the upwind approximation C-2.

REFERENCES

[1] Abma, R., J. Sun, and N. Bernitsas, 1999, Antialiasing methods in Kirchhoff migration: Geophysics, **64**, 1783–1792.

[2] Albertin, U., D. Yingst, P. Kitchenside, and V. Tcheverda, 2004, True-amplitude beam migration: 68th Annual International Meetings, SEG Expanded Abstracts, 949–952.

[3] Aldridge, D. F., 1994, Linearization of the eikonal equation (short note): Geophysics, **59**, 1631–1632.

[4] Alkhalifah, T., 2011, Efficient traveltime compression for 3-D prestack Kirchhoff migration: Geophysical Prospecting, **59**, 1–9.

[5] Alkhalifah, T., and S. Fomel, 2010, An eikonal based formulation for traveltime perturbation with respect to the source location: Geophysics, **75**, no. 6, T175–T183.

[6] Bevc, D., 1997, Imaging complex structures with semirecursive Kirchhoff migration: Geophysics, **62**, 577–588.

[7] Bortfeld, R., 1989, Geometrical ray theory: rays and traveltimes in seismic systems (second-order approximation of traveltimes): Geophysics, **54**, 342–349.

[8] Chapman, C., 2004, Fundamentals of seismic wave propagation: Cambridge University Press.

[9] Dellinger, J. A., S. H. Gray, G. E. Murphy, and J. T. Etgen, 2000, Efficient 2.5-D true-amplitude migration: Geophysics, **65**, 943–950.

[10] Engquist, B., and O. Runborg, 1996, Multi-phase computations in geometrical optics: Journal of Computational and Applied Mathematics, **74**, 175–192.

[11] Fomel, S., 2002, Antialiasing of Kirchhoff operators by reciprocal parameterization: Journal of Seismic Exploration, **10**, 293–310.

[12] Fomel, S., S. Luo, and H. Zhao, 2009, Fast sweeping method for the factored eikonal equation: Journal of Computational Physics, **228**, 6440–6455.

[13] Fomel, S., and J. A. Sethian, 2002, Fast phase-space computation of multiple arrivals: Proceedings of the National Academy of Sciences, **99**, 7329–7334.

[14] Franklin, J. B., and J. M. Harris, 2001, A high-order fast marching scheme for the linearized eikonal equation: Journal of Computational Acoustics, **9**, 1095–1109.
[15] Geoltrain, S., and J. Brac, 1993, Can we image complex structures with first-arrival traveltime?: Geophysics, **58**, 564–575.
[16] Gray, S. H., 2005, Gaussian beam migration of common-shot records: Geophysics, **70**, S71–S77.
[17] Hill, N. R., 1990, Gaussian beam migration: Geophysics, **55**, 1416–1428.
[18] ———, 2001, Prestack Gaussian beam depth migration: Geophysics, **66**, 1240–1250.
[19] Lumley, D. E., J. F. Claerbout, and D. Bevc, 1994, Anti-aliased Kirchhoff 3-D migration: 64th Annual International Meetings, SEG Expanded Abstracts, 1282–1285.
[20] Mendes, M., 2000, Green's function interpolation for prestack imaging: Geophysical Prospecting, **48**, 49–62.
[21] Nichols, D. E., 1994, Imaging complex structures using band-limited Green's functions: PhD thesis, Stanford University.
[22] Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2007, Numerical recipes: The art of scientific computing: Cambridge University Press.
[23] Qian, J., and W. W. Symes, 2002, An adaptive finite-difference method for traveltimes and amplitudes: Geophysics, **67**, 167–176.
[24] Sava, P., and S. Fomel, 1998, Huygens wavefront tracing: A robust alternative to ray tracing: 62nd Annual International Meetings, SEG Expanded Abstracts, 1961–1964.
[25] Sethian, J. A., 1996, Level set methods: Evolving interfaces in geometry, fluid mechanics, computer vision, and materials science: Cambridge University Press.
[26] Sethian, J. A., and A. M. Popovici, 1999, 3-D imaging using higher order fast marching traveltimes: Geophysics, **64**, 516–523.
[27] Slotnick, M. M., 1959, Lessons in seismic computing: Society of Exploration Geophysics.
[28] Symes, W. W., and J. Qian, 2003, A slowness matching Eulerian method for multivalued solutions of eikonal equations: Journal of Scientific Computing, **19**, 501–526.
[29] Ursin, B., 1982, Quadratic wavefront and traveltime approximations in inhomogeneous layered media with curved interfaces: Geophysics, **47**, 1012–1021.
[30] Vanelle, C., and D. Gajewski, 2002, Second-order interpolation of traveltimes: Geophysical Prospecting, **50**, 73–83.
[31] Vanelle, C., M. Spinner, T. Hertweck, C. Jäger, and D. Gajewski, 2006, Traveltime-based true-amplitude migration: Geophysics, **71**, no. 6, S251–S259.
[32] Červený, V., 2001, Seismic ray theory: Cambridge University Press.
[33] Versteeg, R., 1994, The Marmousi experience: Velocity model determination on a synthetic complex data set: The Leading Edge, **13**, 927–936.
[34] Xu, S., H. Chauris, G. Lambaré, and M. Noble, 2001, Common-angle migration: A strategy for imaging complex media: Geophysics, **66**, 1877–1894.
[35] Zhao, H. K., 2005, A fast sweeping method for eikonal equations: Mathematics of Computation, **74**, 603–627.

Seismic data analysis using local time-frequency decomposition

Yang Liu[*†], *Sergey Fomel*[†*]

ABSTRACT

Many natural phenomena, including geologic events and geophysical data, are fundamentally nonstationary - exhibiting statistical variation that changes in space and time. Time-frequency characterization is useful for analyzing such data, seismic traces in particular.

We present a novel time-frequency decomposition, which aims at depicting the nonstationary character of seismic data. The proposed decomposition uses a Fourier basis to match the target signal using regularized least-squares inversion. The decomposition is invertible, which makes it suitable for analyzing nonstationary data. The proposed method can provide more flexible time-frequency representation than the classical S transform. Results of applying the method to both synthetic and field data examples demonstrate that the local time-frequency decomposition can characterize nonstationary variation of seismic data and be used in practical applications, such as seismic ground-roll noise attenuation and multicomponent data registration.

INTRODUCTION

Geological events and geophysical data often exhibit fundamentally nonstationary variations. Therefore, time-frequency characterization of seismic traces is useful for geophysical data analysis. A widely used method of time-frequency analysis is the short-time Fourier transform (STFT) (1). However, the window function limits the time-frequency resolution of STFT (6). An alternative is the wavelet transform, which expands the signal in terms of wavelet functions that are localized in both time and frequency (4). However, because a wavelet family is built by restricting its frequency parameter to be inversely proportional to the scale, expansion coefficients in a wavelet frame may not provide precise enough estimates of the frequency content of waveforms, especially at high frequencies (28). Therefore, (24; 23) developed a time-frequency continuous-wavelet transform (TFCWT) to describe time-frequency map more accurately than the conventional continuous-wavelet transform (CWT). The S transform (26) is another generalization of STFT, which extends CWT and overcomes some of its disadvantages. (22) developed a general version of the S transform by employing windows of arbitrary and varying shape. The clarity of the S transform

[*]**e-mail:** sergey.fomel@beg.utexas.edu

is worse than the Wigner-Ville distribution function (30), which achieves a higher resolution but is seldom used in practice because of its well-known drawbacks, such as interference and aliasing. For this reason, (17) provided a smoothed Wigner-Wille distribution (SWVD) to reduce the interference caused by the cross-term interference. The matching pursuit method is yet another approach to representing the time-frequency signature (21; 28; 29). Matching pursuit involves several parameters and is a relatively expensive method. There are some other approaches to spectral decomposition. (3) compare several different spectral-decomposition methods.

(18; 19) recently proposed a new method of time-varying frequency characterization of nonstationary seismic signals that is based on regularized least-squares inversion. In this paper, we expand the method of (19) by designing an invertible nonstationary time-frequency decomposition — *local time-frequency (LTF) decomposition* and its extensions — *local time-frequency-wavenumber (LTFK)* and *local space-frequency-wavenumber (LXFK) decompositions*. The key idea is to minimize the error between the input signal and all its Fourier components simultaneously using regularized nonstationary regression (12) with control on time resolution. This approach is generic, in the sense that it is possible to combine other basis functions, eg., fractional splines, with regularization (16). Although there is an iterative inversion inside the algorithm, one can use LTF decomposition as an invertible "black box" transform from time to time-frequency, similar in properties to the S transform. The proposed decompositions can provide local time-frequency or space-wavenumber representations for common seismic data-processing tasks. We test the new method and compare it with the S transform by using a classical benchmark signal with two crossing chirps. The proposed LTF decomposition appears to provide higher resolution in both time and frequency when an appropriate parameters of the shaping regularization operator (11) are used to constrain the time resolution. Examples of ground-roll attenuation and multicomponent image registration demonstrate that the method can be effective in practical applications.

THEORY

Local time-frequency (LTF) decomposition

The Fourier series is by definition an expansion of a function in terms of a sum of sines and cosines. Letting a causal signal, $f(x)$, be in range of $[0, L]$, the Fourier series of the signal is given by

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos\left(\frac{2\pi n x}{L}\right) + b_n \sin\left(\frac{2\pi n x}{L}\right) \right] . \tag{1}$$

The notion of a Fourier series can also be extended to complex coefficients as

follows:

$$f(x) = \sum_{n=-\infty}^{\infty} A_n \Psi_n(x) \ , \tag{2}$$

where A_n are the Fourier coefficients and $\Psi_n(x) = e^{i(2\pi nx/L)}$.

Nonstationary regression allows the coefficients A_n to change with x. In the linear notation, $A_n(x)$ can be obtained by solving the least-squares minimization problem

$$\min_{A_n} \|f(x) - \sum_n A_n(x)\Psi_n(x)\|_2^2 \ . \tag{3}$$

The minimization problem is ill posed because there are a lot more unknown variables than constraints. Our solution is to include additional constraints in the form of regularization, which limits the allowed variability of the estimated coefficients (12). Tikhonov's regularization (27) can modify the objective function to

$$\widetilde{A_n}(x) = \arg\min_{A_n} \|f(x) - \sum_n A_n(x)\Psi_n(x)\|_2^2 + \epsilon^2 \sum_n \|\mathbf{D}[A_n(x)]\|_2^2 \ , \tag{4}$$

where $\mathbf{D}$ is the *regularization* operator and ϵ is a scaling parameter. One can define $\mathbf{D}$, for example, as a gradient operator that penalizes roughness of $A_n(x)$.

We use shaping regularization (11) instead of Tikhonov's regularization to constrain the least-squares inversion. Shaping is a general method for imposing constraints by explicit mapping the estimated model to the desired model, eg., smooth model. Instead of trying to find and specify an appropriate regularization operator, the user of the shaping-regularization algorithm specifies a shaping operator, which is often easier to design.

The absolute value of time-varying coefficients $|A_n(x)|$ provides a time-frequency representation, and equation 2 provides the inverse calculation. In the discrete form, a range of frequencies can be decided by the Nyquist frequency (6) or by the user's assignment. In a somewhat different approach, (18) minimized the error between the input signal and each frequency component independently. Their algorithm and the proposed algorithm are equivalent when the decomposition is stationary (or using a very large shaping radius), because they both reduce to the regular Fourier transform. In the case of nonstationarity, their approach does not guarantee invertability, because it processes each frequency independently.

Local t-f-k (LTFK) and local x-f-k (LXFK) decompositions

(2) proposed t-f-k and x-f-k transforms that are based on the S transform. We define analogous LTFK and LXFK decompositions by using the LTF decomposition. The new decompositions can be used to design general time-varying or space-varying FK filters. The key steps of the algorithm are illustrated schematically in Figure 1.

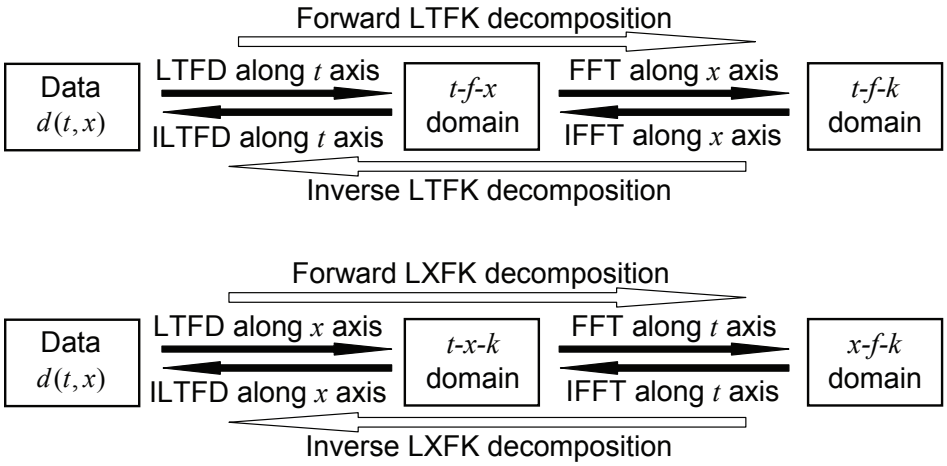


Figure 1: Schematic illustration of LTFK decomposition (a) and LXFK decomposition (b) by using the LTF decomposition. ltft/. LTFK,LXFK

Example of Time-frequency Characterization

A simple 1-D example is shown in Figure 2. The input signal includes two crossing chirp signals and displays nonstationary characteristics (Figure 2a). We applied the LTF decomposition to obtain a time-frequency distribution. Figure 3a shows that the proposed method recovers the linear frequency trend with high resolution in both time and frequency. In comparison, the S transform has high resolution near low frequencies but loses resolution at high frequencies (Figure 2b). Figure 3b displays the LTF decomposition using a different smoothing parameter (14 points) to demonstrate adjustable time-frequency characteristics of the LTF decomposition.

APPLICATION TO GROUND-ROLL ATTENUATION

Seismic data always consist of signal and noise components. The time-frequency denoising algorithm is an effective method for handling noise problems (9). Ground roll is the main type of coherent noise in land seismic surveys and is characterized by low frequencies and high amplitudes. Current processing techniques for attenuating ground roll include frequency filtering, FK filtering (33), radon transform (20), wavelet transform (8), and the curvelet transform (32). (2) applied the S transform to ground-roll attenuation. Here, we propose a similar strategy, except that we are applying the proposed local time-frequency decomposition instead of the S transform.

We applied our methods to a land shot gather contaminated by nearly radial ground roll (Figure 4a). All time-domain images are obtained after automatic gain control (AGC). We applied the forward LTF decomposition to each trace to generate a time-frequency cube (Figure 5a). Note that the ground roll is distributed at localized

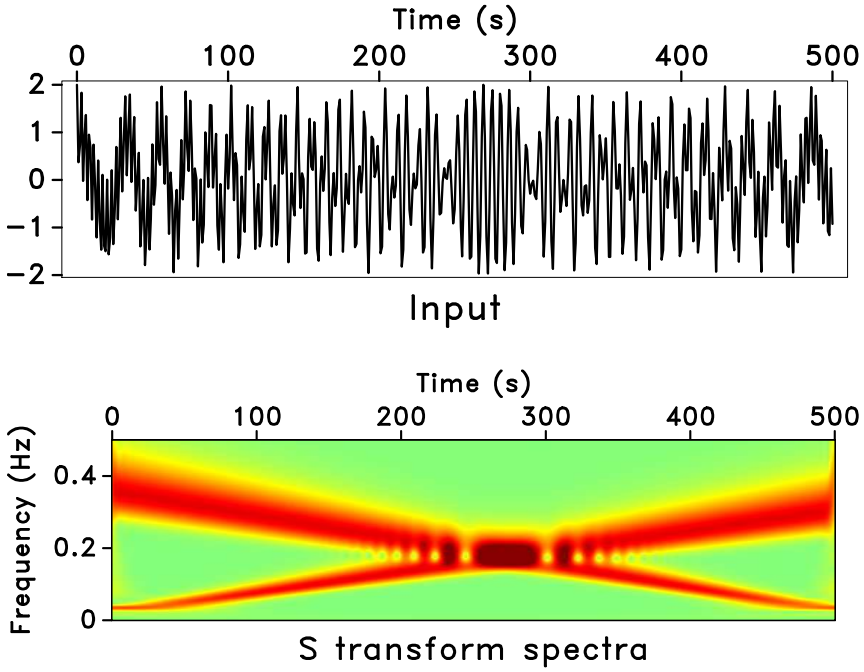


Figure 2: Synthetic signal with two crossing chirps (a) and time-frequency spectra from S transform (b). ltft/timefreq cchirps,st1

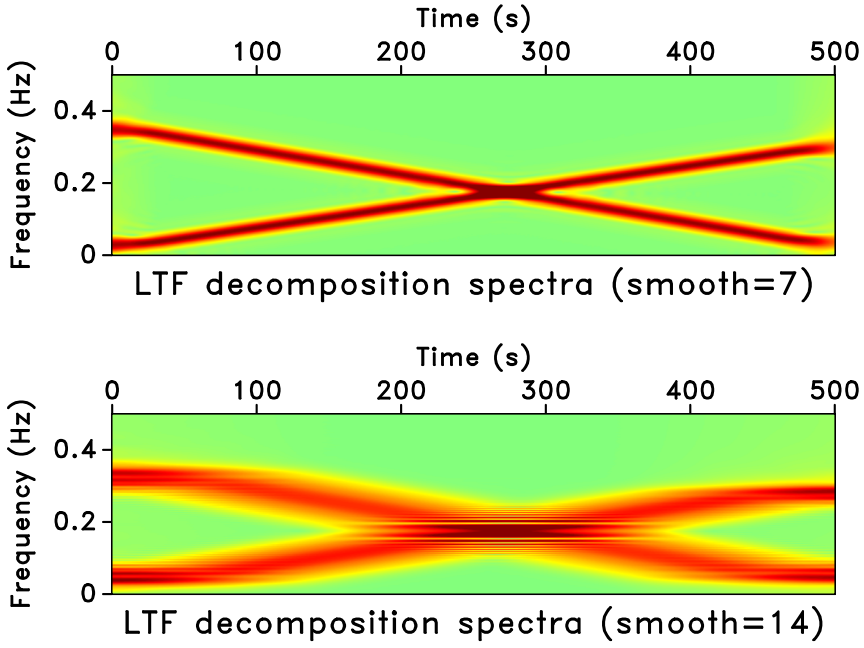


Figure 3: Time-frequency spectra from LTF decomposition with different sizes of the smoothing radius. Smoothing radius of 7 points (a) and smoothing radius of 14 points (b). ltft/timefreq ltft1,ltft12

time-space (left-down section of Figure 5a) and time-frequency (right-down section of Figure 5a) positions. The LTF decomposition is flexible, due to its adjustable time-frequency resolution. Therefore, we designed a simple muting filter to remove the noise components localized in both frequency and space (Figure 5b). The inverse LTF decomposition brings the separated signal back to the original domain (Figure 4). Figure 4c shows the difference between raw data (Figure 4a) and denoised result using LTF decomposition (Figure 4b). It is possible to design more complicated but more powerful masks. Without a time-space mask, our method of simply muting selected frequencies would reduce to band-pass filtering.

The LTFK and LXFK decompositions generate data in different domains (Figure 6a and 7a), which show the trend of ground-roll noise in the frequency-wavenumber sections. Simple frequency-wavenumber masks (Figure 6b and 7b) can eliminate ground-roll noise in the decomposition domains. The recovered signals using the inverse LTFK and LXFK decompositions produce similar results (Figure 8a and 8b, respectively). Furthermore, different decompositions can be cascaded to improve their denoising abilities. For comparison, we used a simple high-pass filter. Figure 9a shows that the high-pass filter fails in removing noise, a larger filter window can damage the seismic signal. Another choice is FK filtering (Figure 9b), which cannot remove the low-frequency part of ground-roll noise. The result is similar to that of the LXFK decomposition (Figure 8b), but the proposed method tends to remove more noise than the standard FK filter (especially near location of time 2.7s and offset 1.2km in Figure 8b and 9b) because of the decomposition's locality and its more flexible design. Radial trace (RT) transform is another approach to deal with ground-roll noise, which is a simple geometric re-mapping method of a seismic trace gather. Idealized ground roll is transformed to small temporal frequency by the RT transform and can be eliminated by applying the RT transform, followed by high-pass filtering and the inverse RT transform (5; 15). Figure 10a shows that the RT transform performs better than the high-pass filter or the FK filter. However, it still has trouble separating signal and noise near the source. Figure 10b shows the denoised result after cascading the proposed LXFK and LTF decompositions, which achieved the best result in this case (especially at locations around the bottom left corner).

APPLICATION TO MULTICOMPONENT DATA REGISTRATION

Multicomponent seismic data provide additional information about subsurface physical characteristics (25). Joint interpretation of multiple image components depends on our ability to identify and register reflection events from similar reflectors. (13) and (14) proposed a multistep approach for registering PP and PS images, and identified spectral differences between PP and PS images as a major problem that prevents an easy automatic registration. The new LTF decomposition can provide a natural domain for nonstationary spectral balancing of multicomponent images.

Figure 11a and b show seismic images from compressional (PP) and shear (SS)

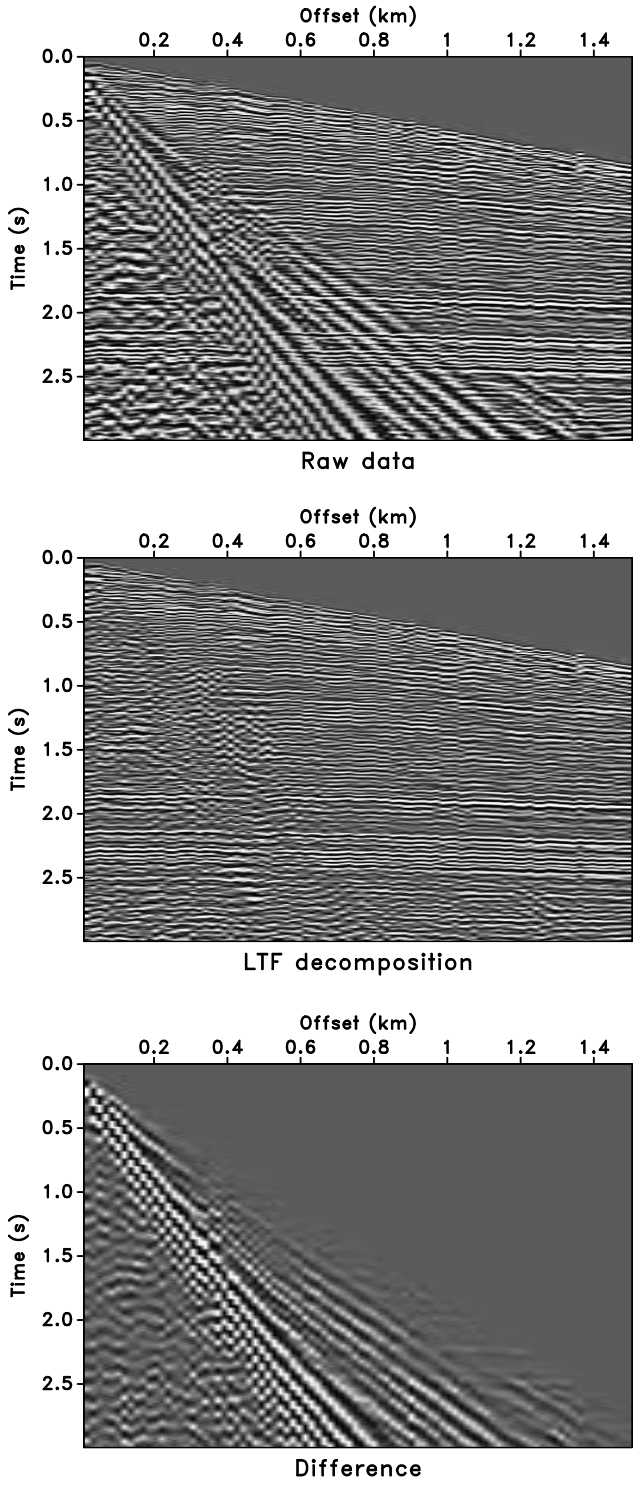


Figure 4: Field land data (a), denoised result using LTF decomposition (b), and difference between raw data (Figure 4a) and denoised result using LTF decomposition (Figure 4b) (c). ltft/groll dat,siltft,iltft

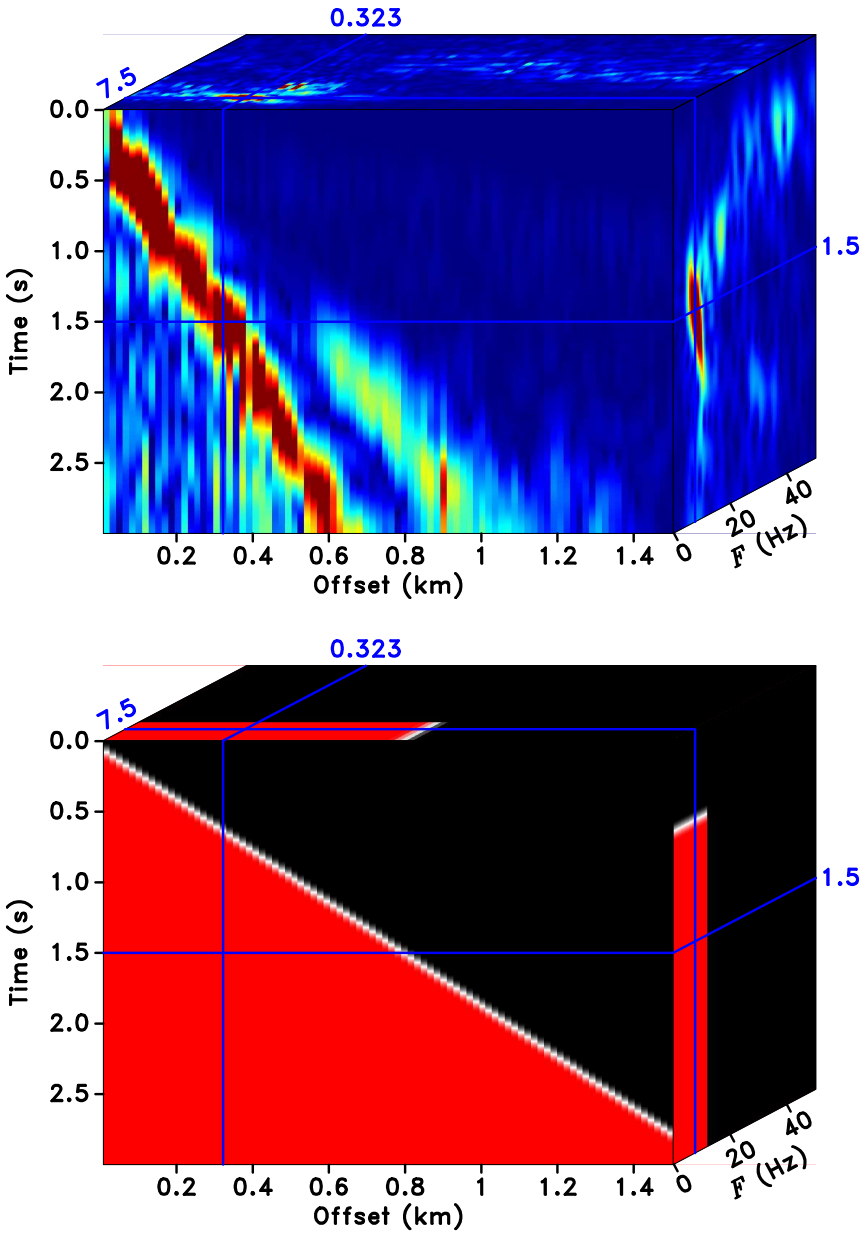


Figure 5: Local T-X-F spectra (a) and filter mask in T-X-F domain (b). ltft/groll ltft,mask

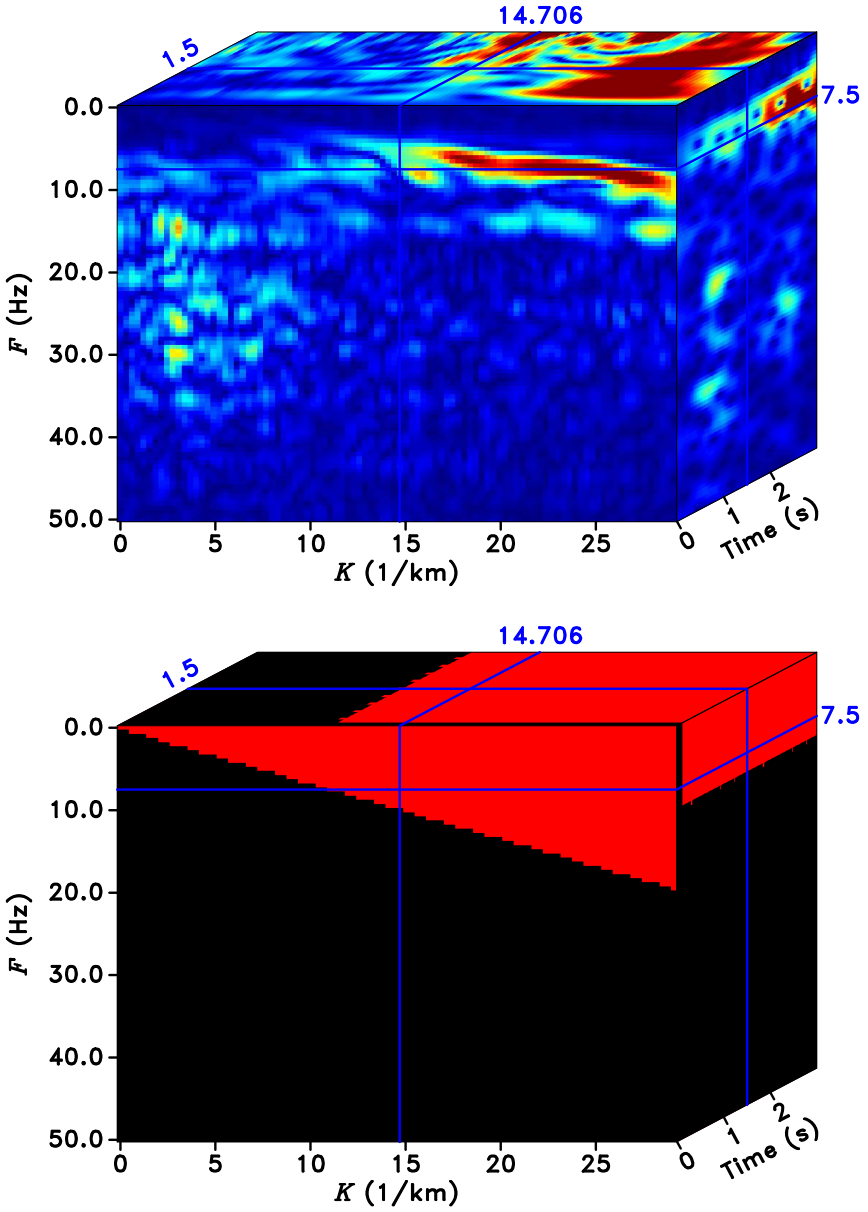


Figure 6: Local F-K-T spectra (a) and filter mask in F-K-T domain (b). ltft/groll tfk,masktfk

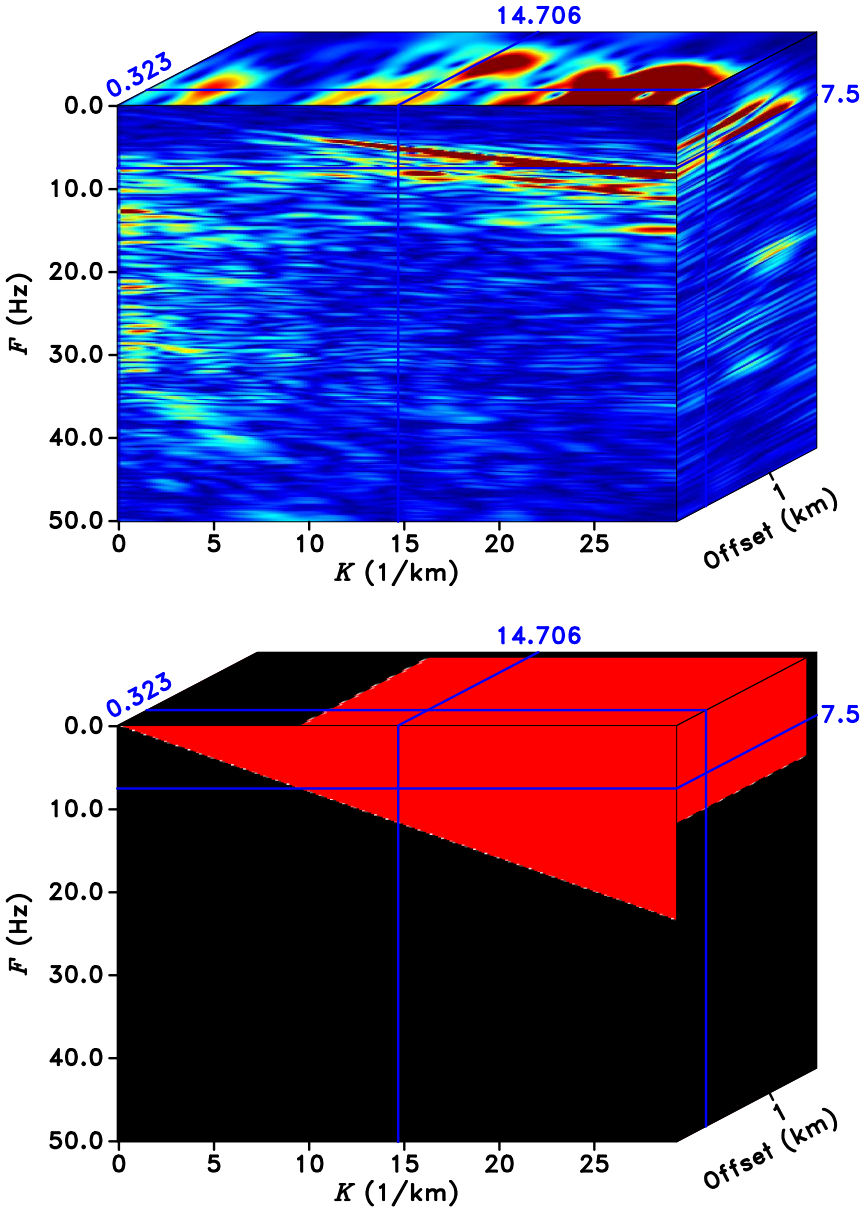


Figure 7: Local F-K-X spectra (a) and filter mask in F-K-X domain (b). ltft/groll fxk,maskfxk

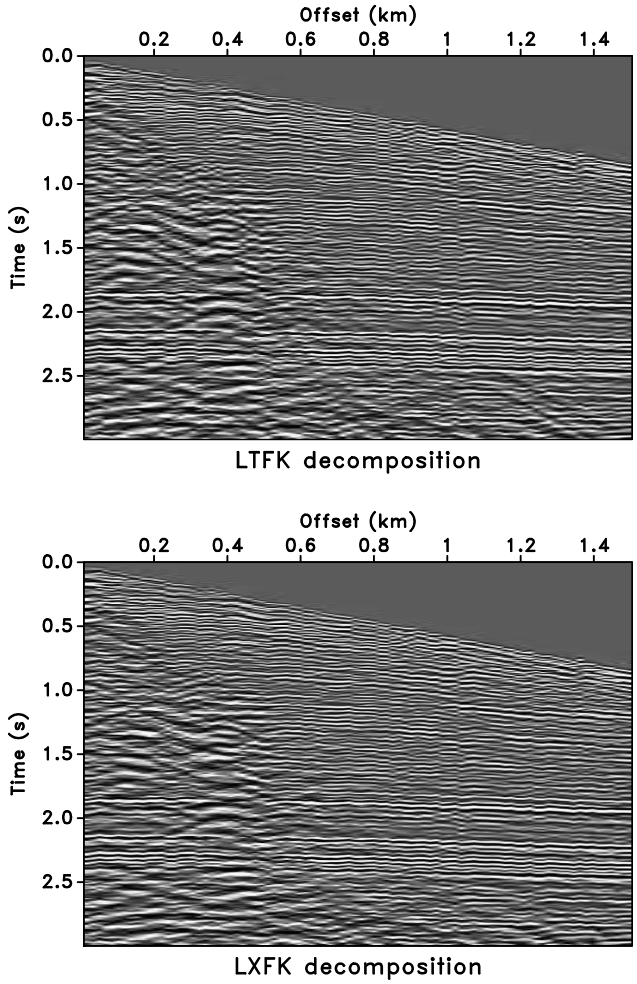


Figure 8: Denoised results using different local decompositions. LTFK decomposition (a) and LXFK decomposition (b). ltft/groll sitfk,sifxk

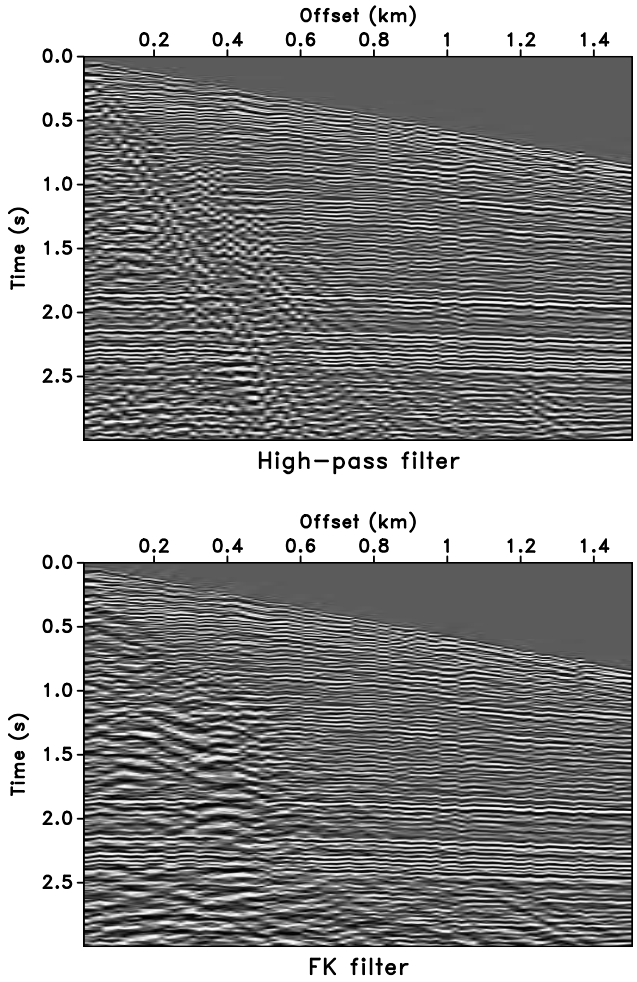


Figure 9: Denoised data using different methods (shown for comparison). High-pass filter (a) and FK filter (b). ltft/groll brshot,ifk

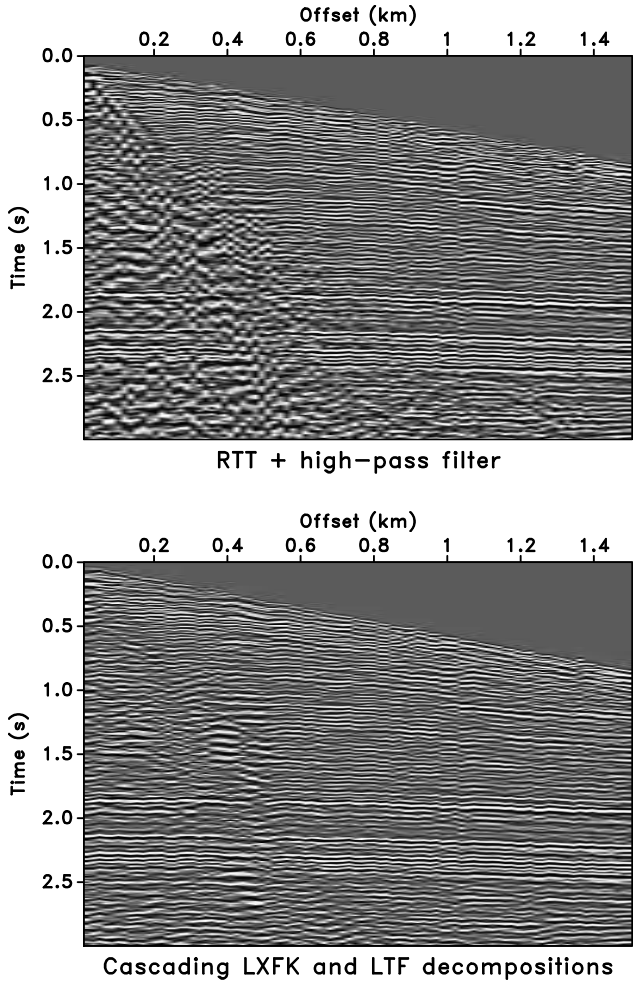


Figure 10: Denoised result by using RT transform with high-pass filter (a) and cascading LXFK and LTF decompositions (b). ltft/groll resu8,csiltft

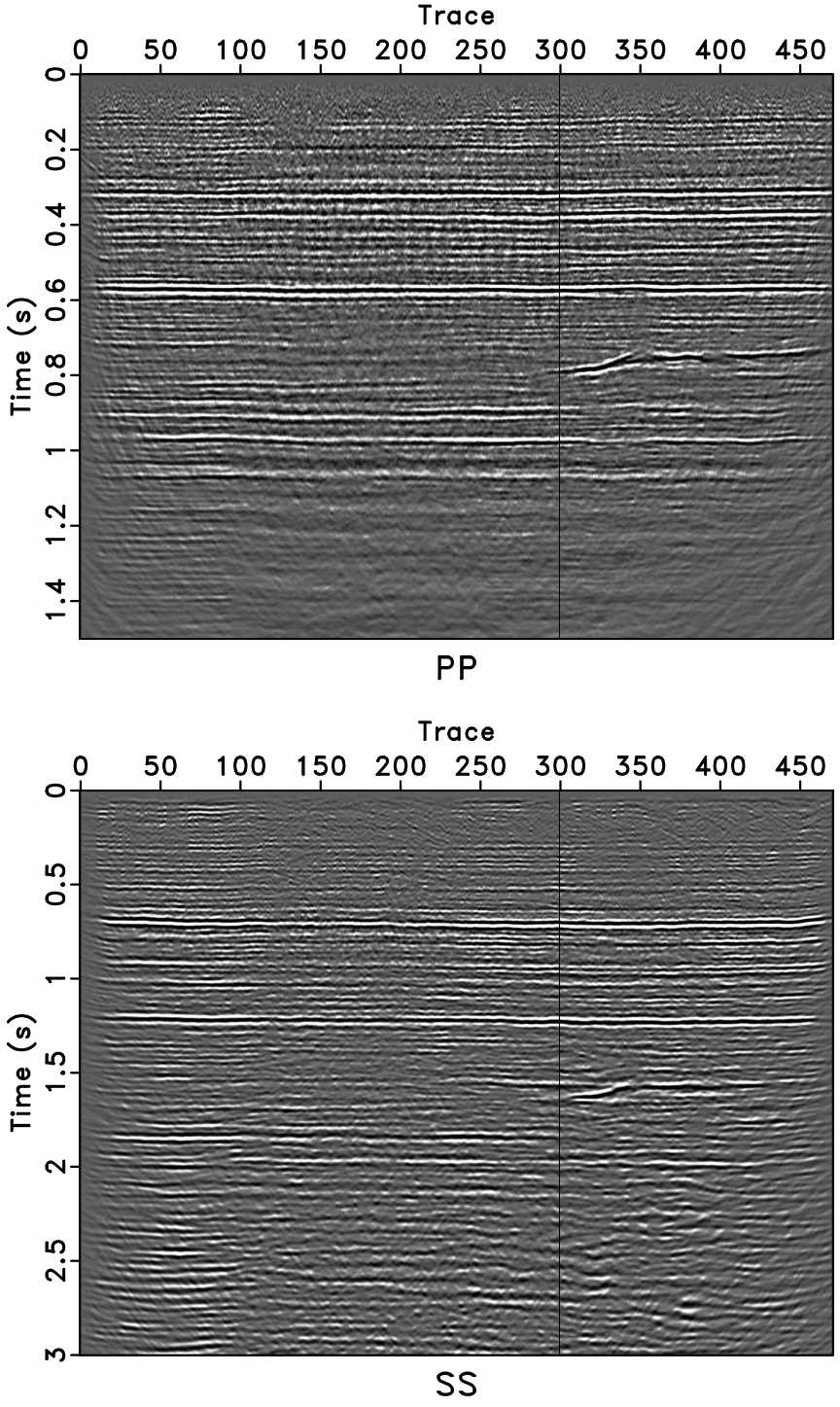


Figure 11: PP (a) and SS (b) images from a nine-component land survey. ltft/vecta vpp,vss

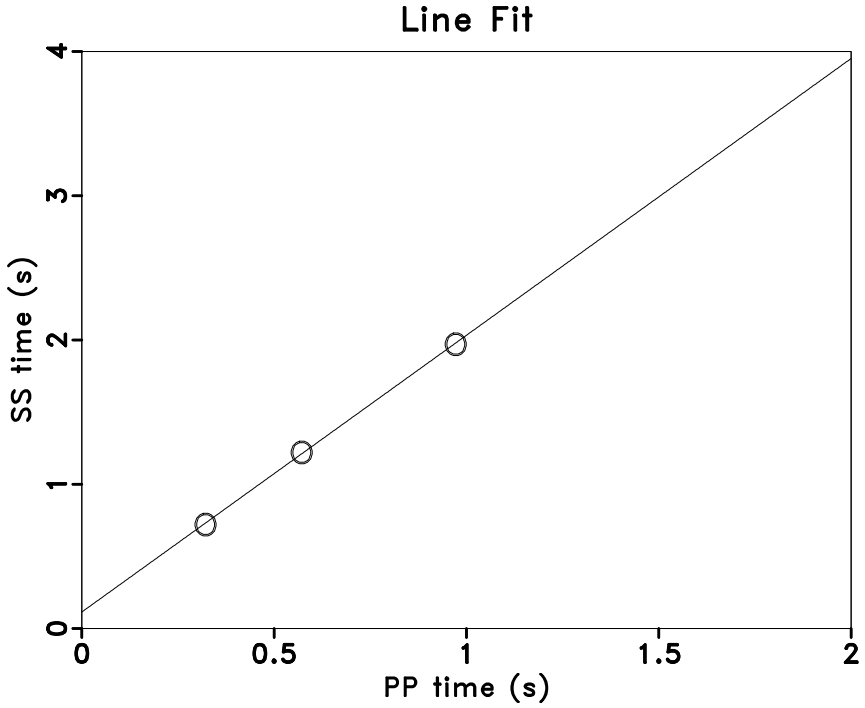


Figure 12: Three "nails" for PP and SS time correlation identified by initial image interpretation and fitted to a straight line. ltft/vecta vnails

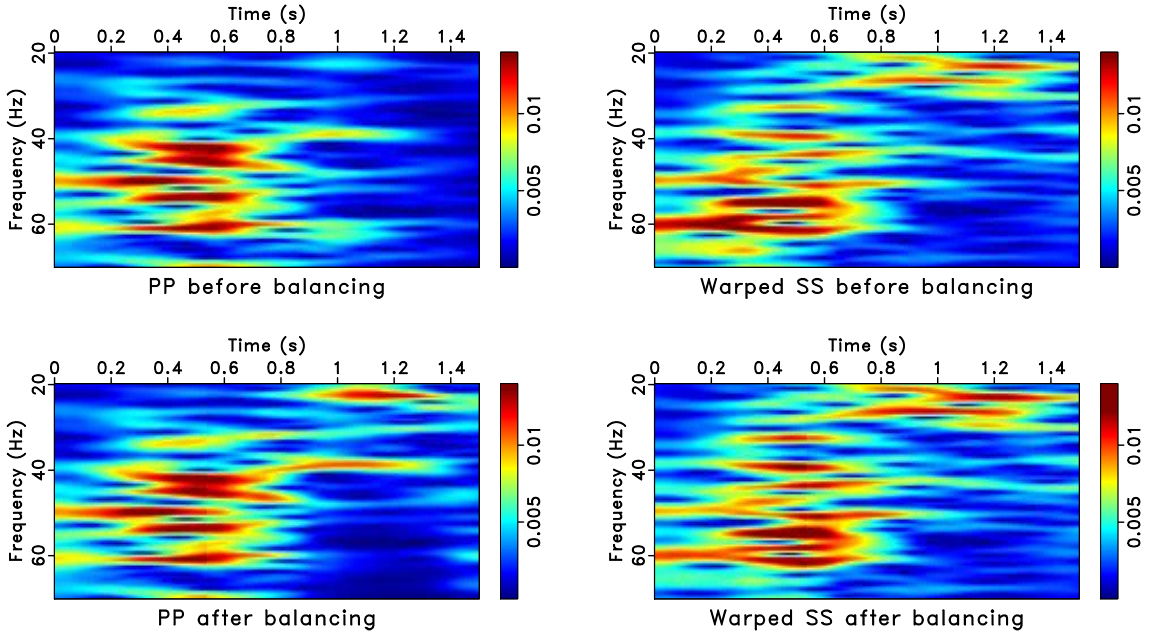


Figure 13: Time-frequency spectra in LTF decomposition domain. PP before balancing (a), SS after initial warping (b), PP after balancing (c), and warped SS after balancing (d). ltft/vecta vect-ppft-0a,vect-psft-0a,vect-ppft-0c,vect-psft-0c

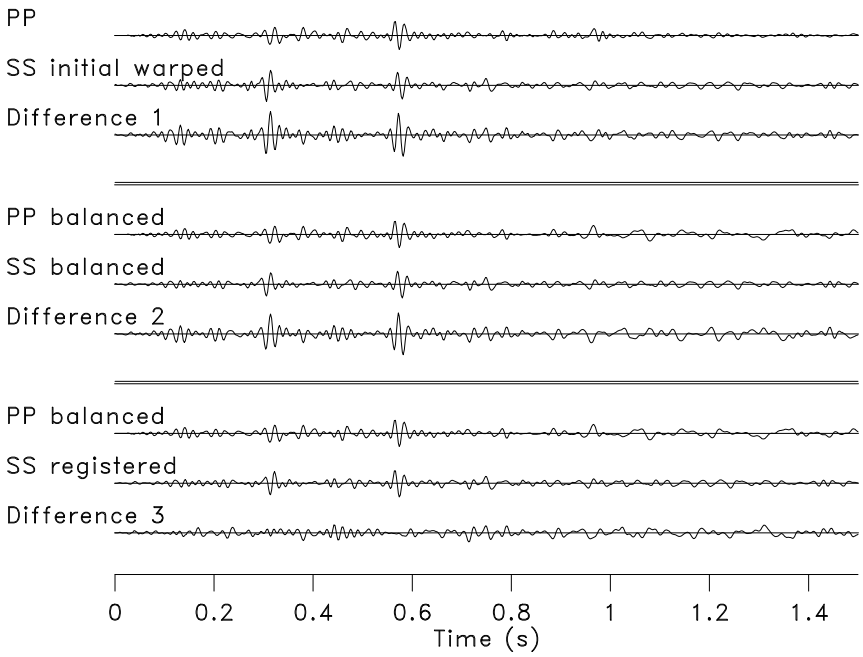


Figure 14: Three stages for PP and SS registration. Initial warping (top), nonstationary spectral balancing (middle), and final registration after warping scan (bottom). ltft/vecta vect-psw-0c

reflections obtained by processing a land nine-component survey (10). One can use "image warping" (31) to squeeze the SS image to PP reflection time and make the two images display in the same coordinate system. Using initial interpretation and well-log analysis, we identified three individual correlation "nails" in the terminology of (7). Fitting a straight line through the nails suggests a constant initial V_P/V_S ratio (Figure 12). For illustration of spectral balancing, we select the 300th trace in the PP and SS images and then warp (squeeze) SS time to PP time by using the initial V_P/V_S ratio. The corresponding local time-frequency spectra are shown in Figures 13a and b. The SS-trace frequency appears higher in the shallow part of the image because of a relatively low S-wave velocity but lower in the deeper part of the image because of the apparently stronger attenuation of shear waves. Spectral balancing essentially smoothes the high-frequency image to match the low-frequency image. The LTF decompositions provide a nonstationary domain for time-varying spectral balancing. Our spectral balancing works as follows. For each time slice in LTF domains, we use three steps:

1. Match the PP and SS spectra by least-squares fitting with Ricker spectra

$$R_i(f) = A_i^2 \frac{f^2}{f_i^2}\, e^{-f^2/f_i^2}\ , \tag{5}$$

where f is frequency axis and get the dominant frequencies f_1 and f_2 ($f_2 > f_1$) and the corresponding amplitudes A_1 and A_2.

2. Use the estimated Ricker parameters to design a matching Gaussian filter

$$G(f) = \frac{A_1 f_2^2}{A_2 f_1^2} e^{f^2(1/f_2^2 - 1/f_1^2)} . \quad (6)$$

3. Shrink the high-frequency spectra to match the low-frequency spectra by applying the Gaussian filter.

The LTF spectra of PP and warped SS trace after nonstationary spectral balancing are shown in Figure 13c and d, respectively, which shows a reasonable similarity between the PP and SS traces for both shallow and deep parts. The inverse LTF decomposition reconstructs balanced PP and SS waveforms in the time domain. Figure 14 displays PP trace, SS trace, and the difference between the two traces in time domain, which are compared for three stages of automatic data registration (14). Residual γ scan is an algorithm for rapid scanning of the field of possible registrations. After applying residual γ scan to update V_P/V_S ratio, the difference between balanced PP and registered SS traces is substantially reduced compared to the initial registration. The final registration result is visualized in Figure 15, which shows interleaved traces from PP and SS images before and after registration. The alignment of main seismic events (especially those at locations "A" and "B") is an indication of successful registration.

CONCLUSION

We have introduced a new time-frequency decomposition that uses regularized nonstationary regression with Fourier bases to represent the time-frequency variation of nonstationary signals. The decomposition is invertible and provides an explicit control on the time and frequency resolution of the time-frequency representation. Experiments with synthetic and field data show that the proposed local time-frequency decomposition can depict nonstationary variation and provide a useful domain for practical applications, such as ground-roll noise attenuation and multicomponent image registration.

ACKNOWLEDGMENTS

We thank Guochang Liu and Mirko van der Baan for stimulating discussions. We thank Partha Routh, one anonymous associate editor, and one anonymous reviewer for helpful suggestions, which improved the quality of the paper. This work is supported in part by National Natural Science Foundation of China (Grant No. 41004041) and 973 Programme of China (Grant No. 2009CB219301). This publication was authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

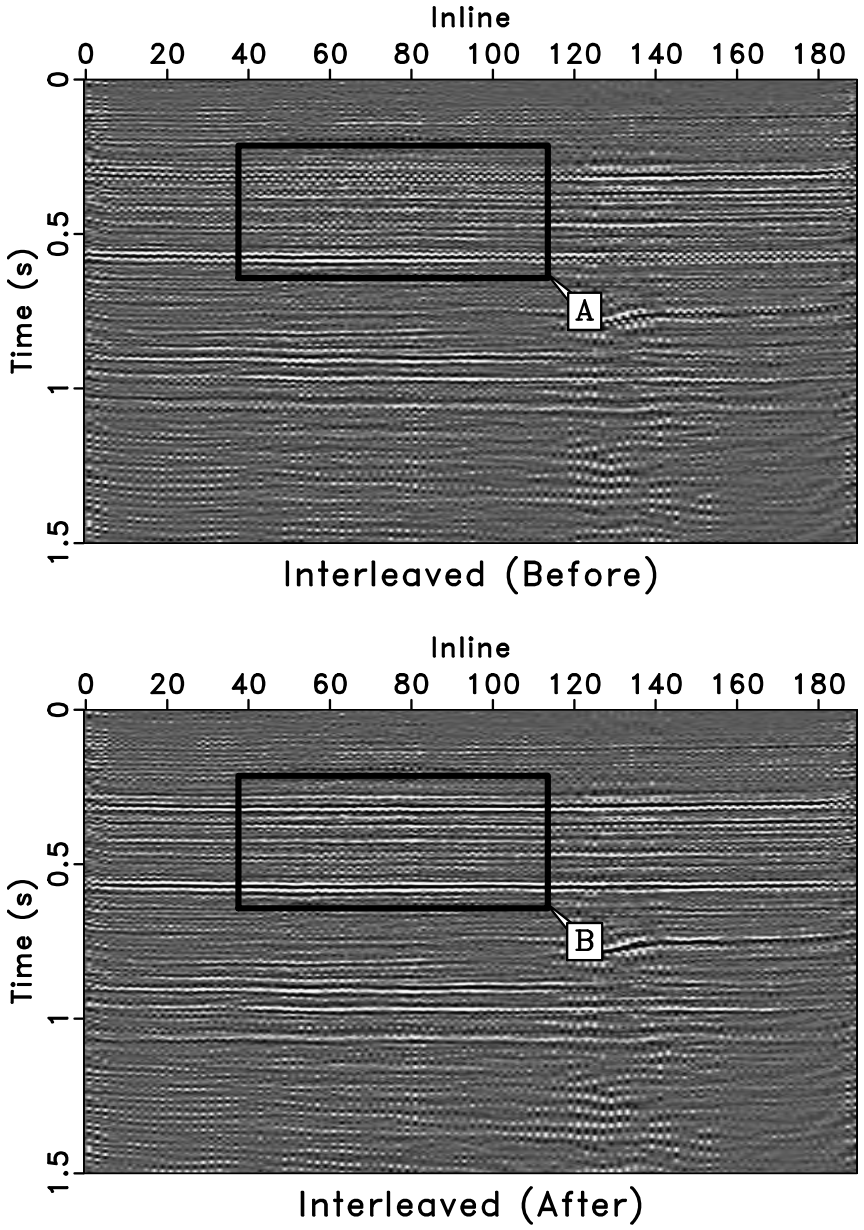


Figure 15: Interleaved traces from PP and SS images before (a) and after (b) multi-component registration. ltft/vecta before,after

REFERENCES

[1] Allen, J. B., 1977, Short term spectral analysis, synthetic and modification by discrete Fourier transform: IEEE Transactions on Acoustic, Speech, Signal Processing, **25**, 235–238.

[2] Askari, R., and H. R. Siahkoohi, 2008, Ground roll attenuation using the S and x-f-k transforms: Geophysical Prospecting, **56**, 105–114.

[3] Castagna, J. P., and S. Sun, 2006, Comparison of spectral decomposition methods: First break, **24**, 75–79.

[4] Chakraborty, A., and D. Okaya, 1995, Frequency-time decomposition of seismic data using wavelet-based method: Geophysics, **60**, 1906–1916.

[5] Claerbout, J. F., 1983, Ground roll and radial traces: Stanford Exploration Project, **SEP-35**, 43–54.

[6] Cohen, L., 1995, Time-frequency analysis: Prentice Hall, Inc.

[7] DeAngelo, M. V., M. Backus, B. A. Hardage, P. Murray, and S. Knapp, 2003, Depth registration of P-wave and C-wave seismic data for shallow marine sediment characterization, Gulf of Mexico: The Leading Edge, **22**, 96–105.

[8] Deighan, A. J., and D. R. Watts, 1997, Ground-roll suppression using the wavelet transform: Geophysics, **62**, 1896–1903.

[9] Elboth, T., I. V. Presterud, and D. Hermansen, 2010, Time-frequency seismic data de-noising: Geophysical Prospecting, **58**, 441–453.

[10] Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.

[11] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[12] ———, 2009, Adaptive multiple subtraction using regularized nonstationary regression: Geophysics, **74**, V25–V33.

[13] Fomel, S., and M. Backus, 2003, Multicomponent seismic data registration by least squares: 73rd Annual International Meeting, SEG, Expanded Abstracts, 781–784.

[14] Fomel, S., M. Backus, K. Fouad, B. Hardage, and G. Winters, 2005, A multistep approach to multicomponent seismic image registration with application to a West Texas carbonate reservoir study: 75th Annual International Meeting, SEG, Expanded Abstracts, 1018–1021.

[15] Henley, D., 1999, The radial trace transform: An effective domain for coherent noise attenuation and wavefield separation: 69nd Annual International Meeting, SEG, Expanded Abstracts, 1204–1207.

[16] Herrmann, F. J., 2001, Fractional spline matching pursuit: A quantitative tool for seismic stratigraphy: 71st Annual International Meeting, SEG, Expanded Abstracts, 1965–1968.

[17] Li, Y., and X. Zheng, 2008, Spectral decomposition using wigner-ville distribution with applications to carbonate reservoir characterization: The Leading Edge, **27**, 1050–1057.

[18] Liu, G., S. Fomel, and X. Chen, 2009, Time-frequency characterization of seismic data using local attributes: 79nd Annual International Meeting, SEG, Expanded Abstracts, 1825–1829.

[19] ———, 2011, Time-frequency characterization of seismic data using local attributes: Geophysics, **76**, P23–P34.
[20] Liu, J., and K. J. Marfurt, 2004, 3-d high resolution radon transforms applied to ground roll suppression in orthogonal seismic surveys: 74th Annual International Meeting, SEG, Expanded Abstracts, 2144–2147.
[21] ———, 2007, Instantaneous spectral attributes to detect channels: Geophysics, **72**, P23–P31.
[22] Pinnegar, C., and L. Mansinha, 2003, The S-transform with windows of arbitrary and varying shape: Geophysics, **68**, 381–385.
[23] Sinha, S., P. S. Routh, and P. Anno, 2009, Instantaneous spectral attributes using scales in continuous-wavelet transform: Geophysics, **74**, WA137–WA142.
[24] Sinha, S., P. S. Routh, P. Anno, and J. P. Castagna, 2005, Spectral decomposition of seismic data with continuous-wavelet transform: Geophysics, **70**, P19–P25.
[25] Stewart, R. R., J. Gaiser, R. J. Brown, and D. C. Lawton, 2003, Converted-wave seismic exploration: Applications: Geophysics, **68**, 40–57.
[26] Stockwell, R. G., L. Mansinha, and R. P. Lowe, 1996, Localization of the complex spectrum: the S transform: IEEE Transactions on Signal Processing, **44**, 998–1001.
[27] Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Soviet Mathematics – Doklady.
[28] Wang, Y., 2007, Seismic time-frequency spectral decomposition by matching pursuit: Geophysics, **72**, V13–V20.
[29] ———, 2010, Multichannel matching pursuit for seismic trace decomposition: Geophysics, **75**, V61–V66.
[30] Wigner, W., 1932, On the quantum correction for thermodynamic equilibrium: Physical Review, **40**, 749–759.
[31] Wolberg, G., 1990, Digital image warping: IEEE Computer Society.
[32] Yarham, C., and F. J. Herrmann, 2008, Bayesian ground-roll separation by curvelet-domain sparsity promotion: 78th Annual International Meeting, SEG, Expanded Abstracts, 2576–2580.
[33] Yilmaz, Ö., 2001, Seismic data analysis: Processing, inversion and interpretation of seismic data: Society of Exploration Geophysics.

Time-frequency analysis of seismic data using synchrosqueezing wavelet transform

Yangkang Chen, Tingting Liu†, Xiaohong Chen†, Jingye Li† and Erying Wang‡**

ABSTRACT

Time-frequency (TF) decomposition is used for characterizing the non-stationary relation between time and instantaneous frequency, which is very important in the processing and interpretation of seismic data. The conventional time-frequency analysis approaches suffer from the contradiction between time resolution and frequency resolution. A new time-frequency analysis approach is proposed based on the synchrosqueezing wavelet transform (SSWT). The SSWT is an empirical-mode-decomposition-like tool but uses a different approach in constructing the components. With the help of the synchrosqueezing techniques, the SSWT can obtain obvious higher time and frequency resolution. Synthetic examples show that the SSWT based TF analysis can exactly capture the variable frequency components. Field data tests show the potential of the proposed approach in detecting anomalies of high-frequency attenuation and detecting the deep-layer weak signal.

KEYWORDS

Time-frequency analysis, synchrosqueezing wavelet transform, low-frequency anomalies, deep-layer weak signal

INTRODUCTION

Time-frequency (TF) analysis solves the problems of identifying and quantifying the oscillatory components presented in the signal, which has been exclusively utilized by the field of exploration geophysics over the past two decades (7; 1; 8; 6). One of the most common use of TF analysis is that deeper channels are usually stronger at lower frequency and the shallower flank of the channel has stronger amplitudes at higher frequencies. In addition, TF analysis can be used to estimating attenuation, pore-pressure prediction, and seismic unconformities, and some implementation of seismic chronostratigraphy (5). Many approaches have been proposed for TF analysis, such as wavelet transform (WT), shot-time Fourier transform (STFT), Wigner-Ville distribution (WVD), S transform and recently proposed local attributes based TF

***e-mail:** sergey.fomel@beg.utexas.edu

(6). All of the mentioned approaches have their own advantages, however, they all have limited resolution either in time or in frequency.

The empirical mode decomposition (EMD) (4) algorithm can separate a signal into locally-constant frequency components, and have been shown to have a high resolution both in time and frequency with some types of extensions, like ensemble empirical mode decomposition (EEMD) and complete ensemble empirical mode decomposition (CEEMD). However, the EMD algorithm is still remaining heuristic because of the lack of mathematical support. The newly proposed synchrosqueezing wavelet transform (SSWT) capture the flavor and philosophy of the EMD approach, but with a mathematical way in constructing the components (2). Because of the high-resolution property of SSWT, it is becoming more and more popular for characterizing non-stationary property in signal analysis field recently. In the exploration field, SSWT has been successfully used for removing ground rolls (9). In this abstract, we use one benchmark non-stationary synthetic model for showing SSWT's high resolution both in time and frequency compared with other two robust TF decomposition approaches. We also applied SSWT onto two field data examples and show its potential in detecting anomalies of high-frequency attenuation and detecting deep-layer weak signal.

SYNCHROSQUEEZING WAVELET TRANSFORM

The well-known continuous wavelet transform is defined by:

$$W_x(a,t) = \int_{-\infty}^{\infty} x(\tau) a^{-1/2} \psi^* \left(\frac{\tau - t}{a} \right) d\tau, \tag{1}$$

where $\psi(t)$ is the chosen mother wavelet, ψ^* denotes the complex conjugate of ψ. The instantaneous frequency of $w_x(a,t)$ for signal $x(t)$ can be got by:

$$w_x(a,t) = -i(W_x(a,t))^{-1} \frac{\partial}{\partial t} W_x(a,t). \tag{2}$$

The synchrosqueezed transform $T_x(w,t)$ can be determined only at the centers w_l of the successive bins $[w_l - \frac{1}{2}\Delta w, w_l + \frac{1}{2}\Delta w]$, with $w_l - w_{l-1} = \Delta w$, by summing different contributions (2):

$$T_x(w_l,t) = (\Delta w)^{-1} \sum_{a_k : |w(a_k,t) - w_l| \le \Delta w/2} W_x(a_k,t) a_k^{-3/2} (\Delta a)_k. \tag{3}$$

The SSWT is invertible and the original signal can be obtained by:

$$\begin{aligned} x(t) &= \mathcal{R}\rceil \left[C_\psi^{-1} \int_0^{\infty} W_x(a,t) a^{-3/2} da \right] \\ &= \mathcal{R}\rceil \left[C_\psi^{-1} \sum_k (a_k,t) a_k^{-3/2} (\Delta a)_k \right] \\ &= \mathcal{R}\rceil \left[C_\psi^{-1} \sum_l T_x(w_l,t)(\Delta w) \right]. \end{aligned} \tag{4}$$

EXAMPLES

Benchmark synthetic example

We first use a synthetic example to demonstrate the SSWT's high resolution both in time and frequency. The synthetic data is borrowed from (6). It is composed of three constant-frequency components and two spikes, as shown in Figure 1(a). We compare the TF analysis result of the SSWT with that of the S transform and local-attributes based TF analysis. The differences among the three TF analysis approaches are obvious. The SSWT exactly capture the three constant-frequency component and get a higher time resolution for the two spikes compared with the other two approaches. In the next part, we apply the SSWT to two field datasets to show its application in detecting low-frequency anomalies and deep-layer weak signal.

Field examples

The first field data (Figure 9(a)) is a land seismic survey used previously by (3). We extract two frequency slices (30 Hz and 60 Hz) from the TF decomposed result and show them in Figures 2(b) and 2(c). It's obvious that the high-frequency slice has some anomalies pointed out by the labels, which may comes from the high-frequency attenuation because of the trapped oil & gas. The results are consistent from that found by (3) using a different technique. The TF cube of the real data corresponding to three different TF analysis approaches are shown in Figure 3. The TF map of the S transform seems blurs together. While the local attributes based TF analysis can have good time resolution, the SSWT can get a better result. The deep-layer of the TF cube show coherent frequency components, indicating the potential week signal which can not be sensed visually or by the other two TF analysis approaches. The detection of deep-layer week signal is significant because these weak signals usually indicate the existence of layers that has obvious impedance difference compared with their overburden layers. Especially when the detected signals are mainly existing in the low frequency band, the detection will indicate the existence of residual oil&gas traps.

The second field data example is a post-stack field data set, which is also used in the literature by (6). The data is shown in Figure 11. In this example, we extract three constant-frequency slices corresponding to 20Hz, 40Hz, and 60 Hz, using three different TF decomposition approaches: TF decomposition using the S transform, TF decomposition using local attributes and TF decomposition using the SSWT. We also extract a time-frequency spectral map for the 80th trace in the field data shown in Figure 11. All the constant-frequency slices and TF spectral map are shown in Figure 5.

From the constant-frequency slices shown in Figure 5, we can observe that the S transform has the lowest time resolution, followed by the local-attributes related TF decomposition. The constant-frequency slices and the TF spectral map from

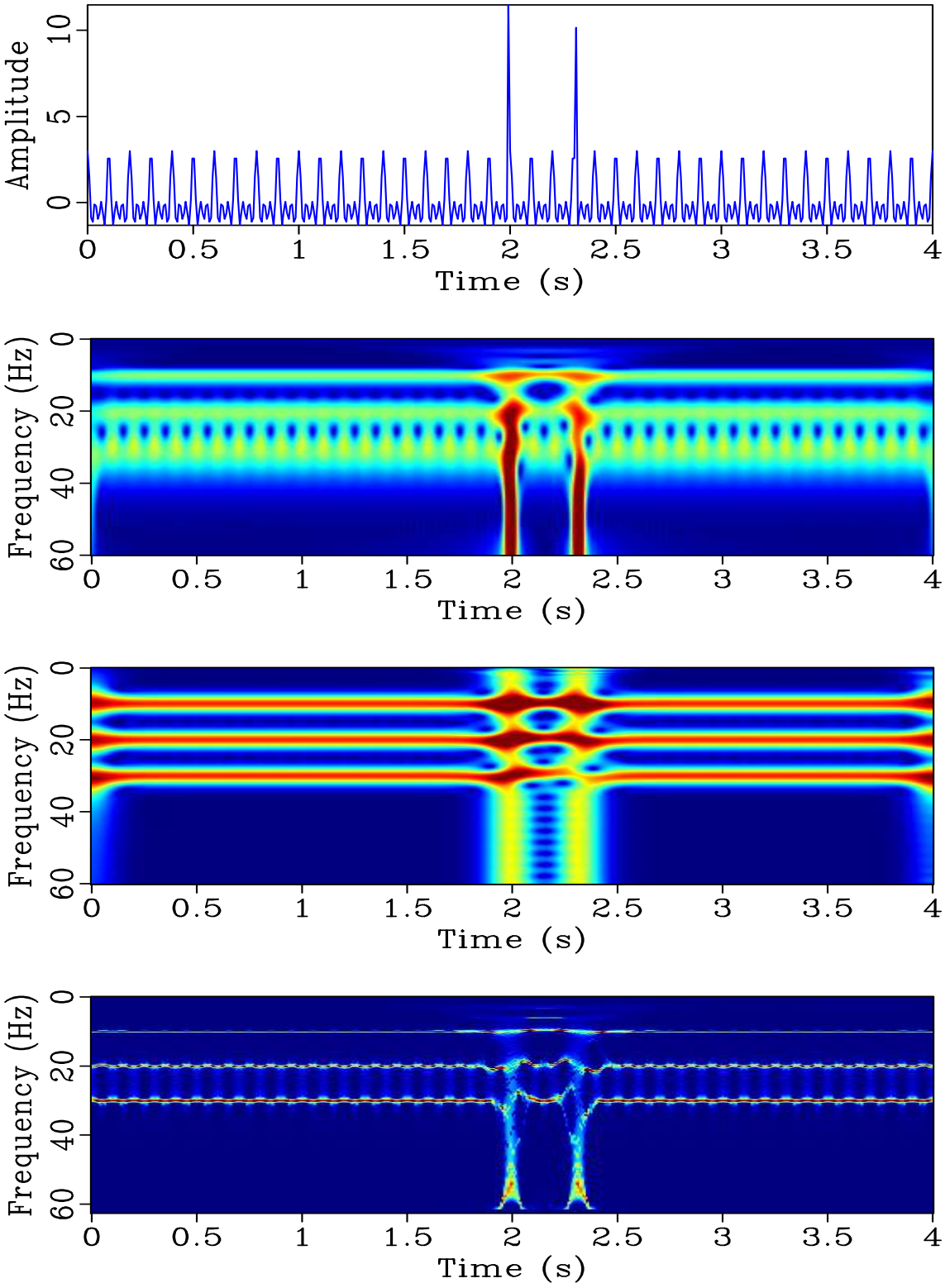


Figure 1: (a) Synthetic signal with three constant-frequency components and two spikes, borrowed from (6). (b) Time-frequency map of the S transform. (c) Time-frequency map using local attributes. (d) Time-frequency map using the synchrosqueezing wavelet transform. synwav/gch1 s-0,st-0,proj-0,gch1-tfsswt

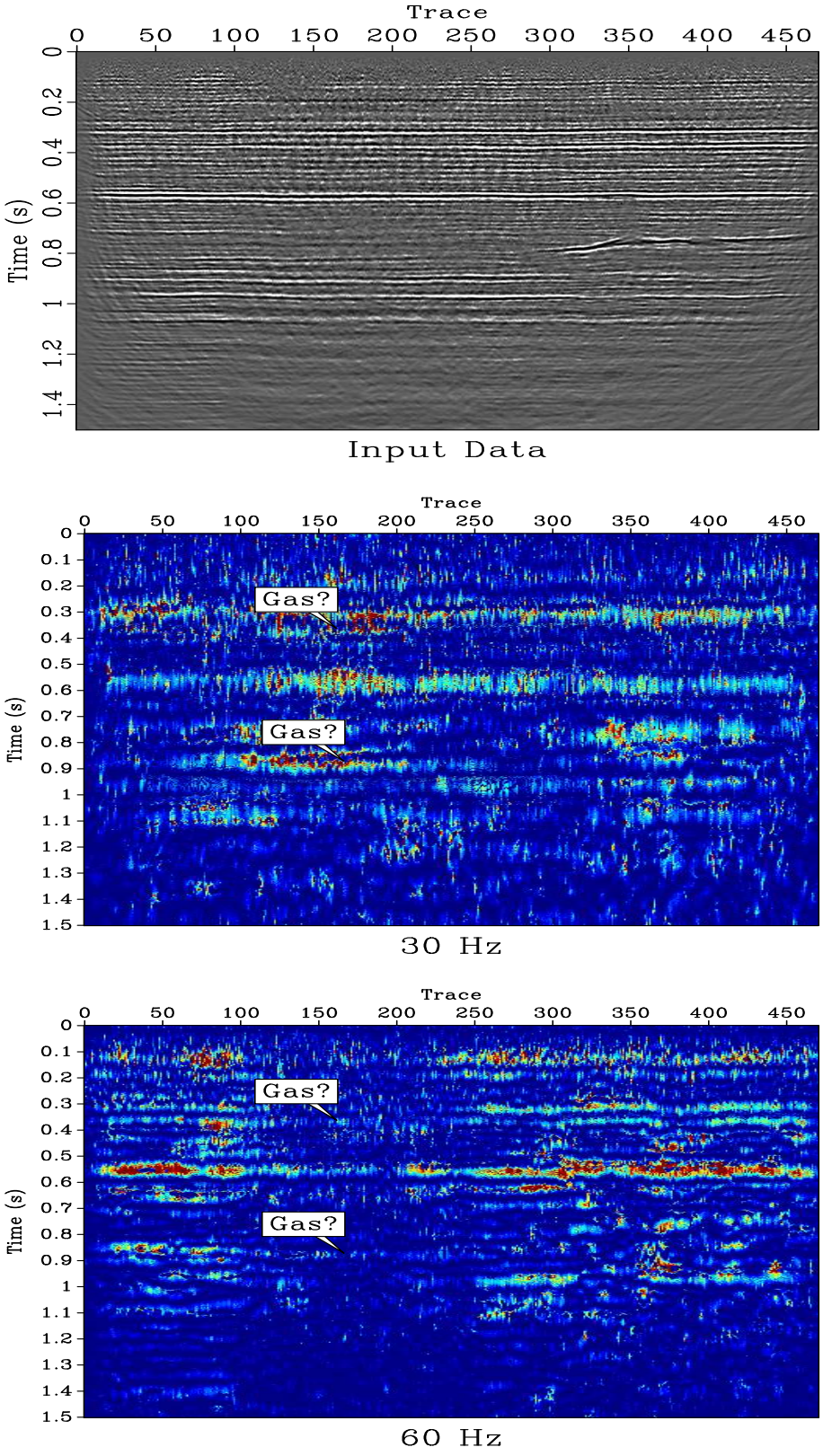


Figure 2: (a) Input data. (a) Frequency slice corresponding to 30 Hz. (b) Frequency slice corresponding to 60 Hz. synwav/pp data,lowf,higf

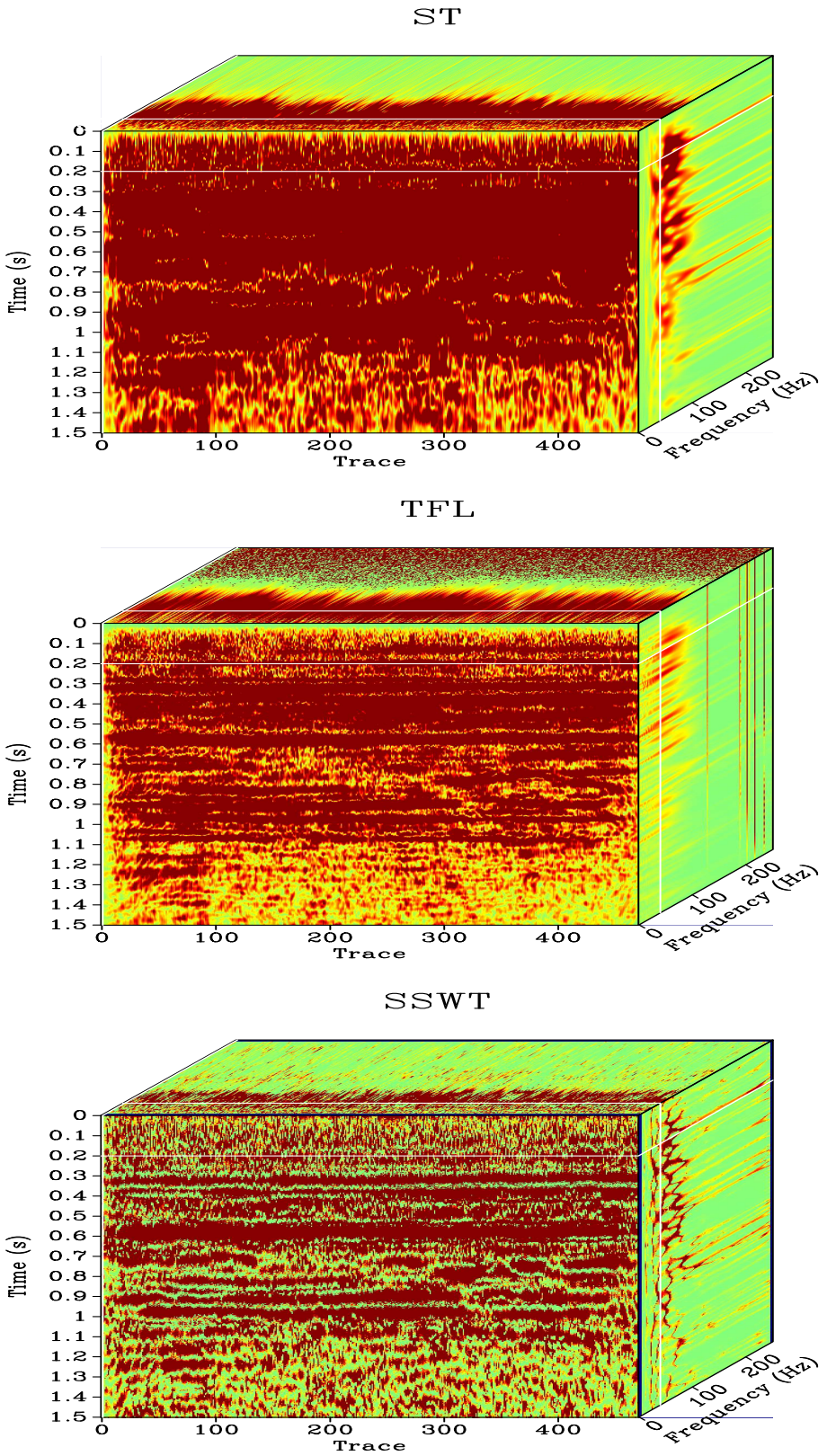


Figure 3: (a) Time-frequency map of the S transform. (b) Time-frequency map using local attributes. (c) Time-frequency map using the synchrosqueezing wavelet transform. synwav/pp timefreq-st,timefreq,pp-tfsswt

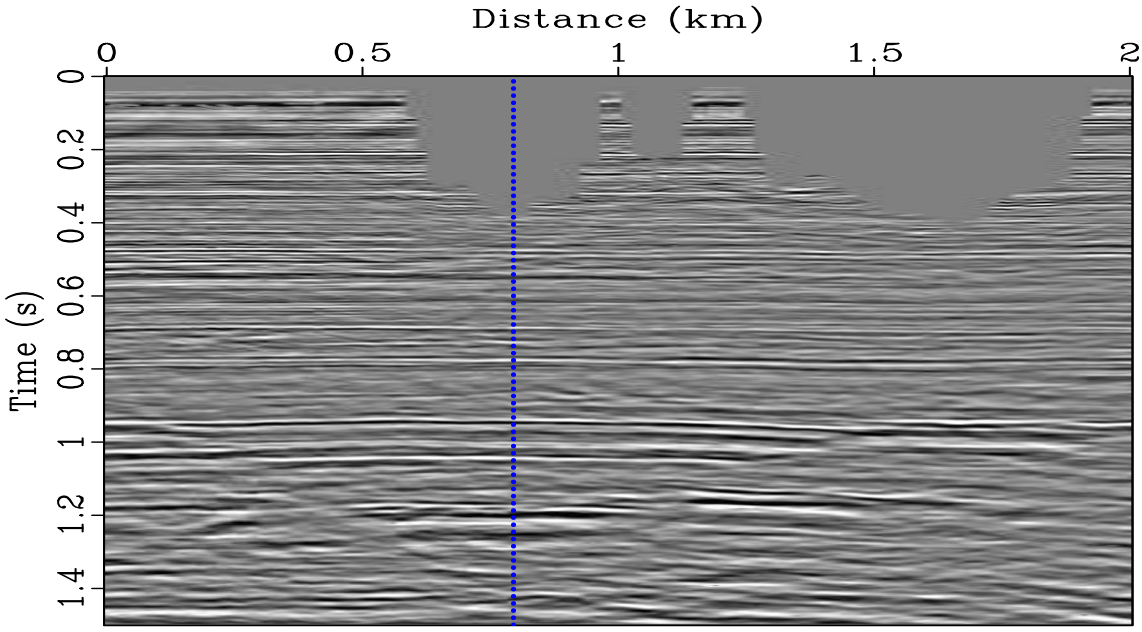


Figure 4: Field data example for detecting low frequency anomalies. synwav/lowf old-1

the SSWT both show that the TF decomposed results have the largest resolution both in time and frequency. From the TF spectral map, the S transform has a slightly lower frequency resolution. By the SSWT based TF decomposition, we can detect several low-frequency anomalies. In this example, we just emphasize four clearest anormalies, and point out them by four frame boxes with different colors. To better illustrate the result, let us first focus on Figure 5(i). The four frame boxes pointed out four strong-amplitude areas, which do not show in its corresponding higher-frequency slices. Thus, these four anomalies may indicate potential oil&gas traps with high probability. Even though from the constant-frequency slices using other two approaches have more or less some anomalies, these anomalies are not very easy to detect. Some highly-potential sweet spots have been lost by using the other two approaches, e.g., area pointed out by the four frame boxes. The best strategy for implementing the SSWT based TF decomposition onto the interpretation for subsurface phenomenons is to combine the SSWT based TF decomposition with the conventional approaches. While the SSWT based TF decomposition can point out the potential anomalies with high fidelity, a more convincing conclusion can be obtained by considering the results from the other approaches.

CONCLUSIONS

We have proposed a new TF analysis approach for seismic data using the synchrosqueezing wavelet transform. The proposed approach can achieve higher reso-

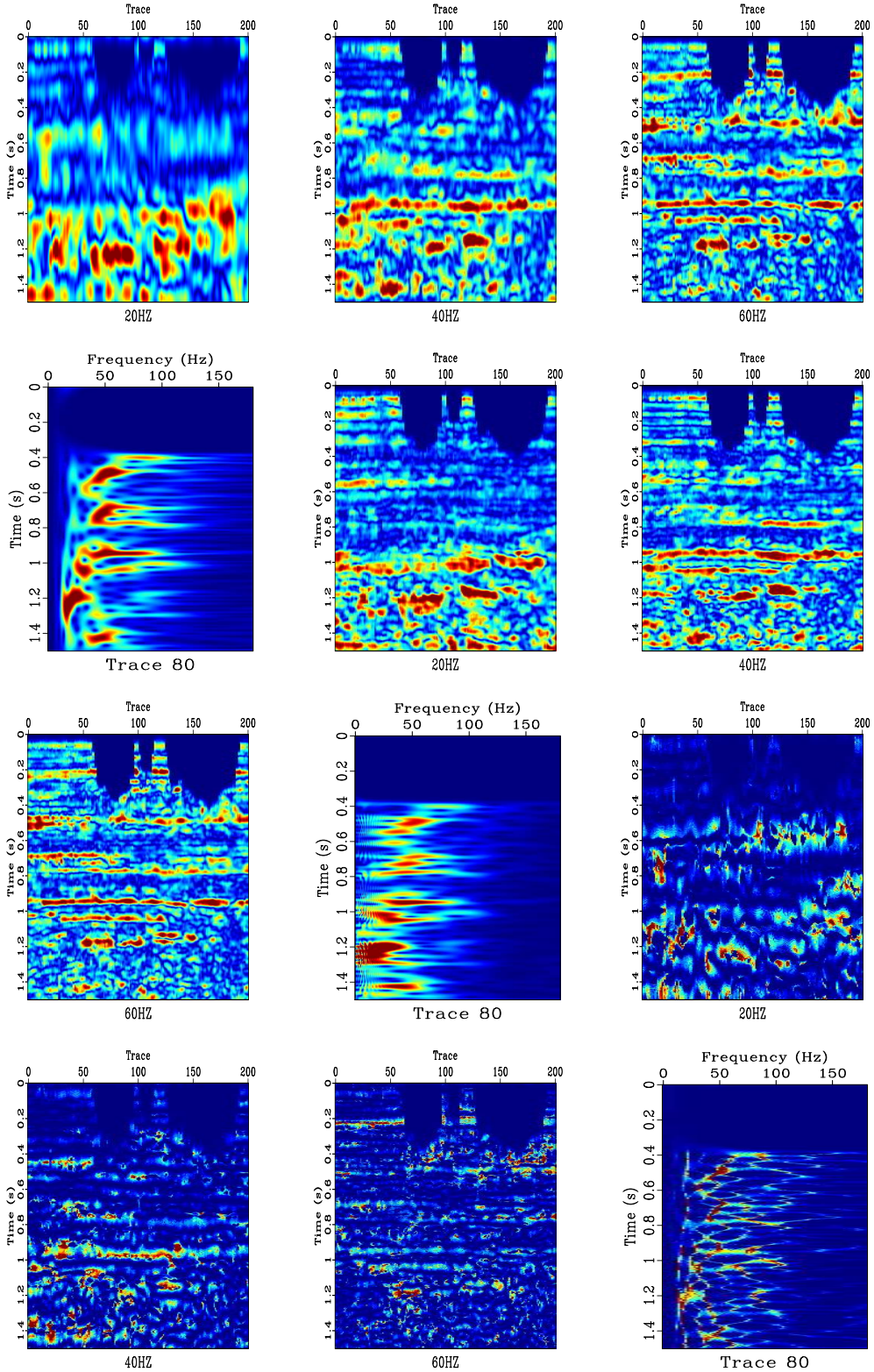


Figure 5: (a)-(c) Frequency slices of 20Hz, 40Hz, and 60 Hz, using the S transform. (d) Time-frequency spectral map using the S transform for 80th trace of the seismic profile shown in Figure 11. (e)-(g) Frequency slices of 20Hz, 40Hz, and 60 Hz, using the local-attributes related TF decomposition. (h) Time-frequency spectral map using the local-attributes related TF decomposition for 80th trace of the seismic profile shown in Figure 11. (i)-(k) Frequency slices of 20Hz, 40Hz, and 60 Hz, using the SSWT. (l) Time-frequency spectral map using the SSWT for 80th trace of the seismic

lution both in time and frequency compared with the S transform and the local-attributes based TF decomposition. Numerical examples show that the SSWT can capture the variable-frequency components embedded in the non-stationary signal exactly. We show the potential of the SSWT in detecting anomalies of high-frequency attenuation, which is crucial for finding oil & gas traps. Field data tests also show that the proposed TF analysis approach can also help to detect deep-layer weak signal that is usually smeared in seismic data, which in turns can help to find deserted residual oil & gas.

ACKNOWLEDGMENTS

We would like to thank Guochang Liu for making the benchmark synthetic example publicly available, and developers of Madagascar software package for providing corresponding codes for testing the algorithms and preparing the figures.

REFERENCES

[1] Castagna, J., S. Sun, and R. W. Siegfried, 2003, Instantaneous spectral analysis: Detection of low-frequency shadows associated with hydrocarbons: The Leading Edge, **22**, 120–127.

[2] Daubechies, I., J. Lu, and H.-T. Wu, 2011, Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool: Applied and Computational Harmonic Analysis, **30**, 243–261.

[3] Fomel, S., 2013, Seismic data decomposition into spectral components using regularized nonstationary autoregression: Geophysics, **78**, O69–O76.

[4] Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: Proceeding of the Royal Society of London Series A, **454**, 903–995.

[5] Lin, T., Z. Wan, B. Zhang, S. Guo, K. Marfurt, and Y. Guo, 2013, Spectral decomposition of time- vs. depth-migrated data: SEG expanded abstracts:83rd Annual international meeting, 1384–1388.

[6] Liu, G., S. Fomel, and X. Chen, 2011, Time-frequency analysis of seismic data using local attributes: Geophysics, **76**, P23–P34.

[7] Partyka, G. A., J. M. Gridley, and J. Lopez, 1998, Interpretational applications of spectral decomposition in reservoir characterization: The Leading Edge, **18**, 353–360.

[8] Reine, C., M. van der Baan, and R. Clark, 2009, The robustness of seismic attenuation measurements using fixed- and variable-window time-frequency transforms: Geophysics, **74**, 123–135.

[9] Shang, S., L. Han, and W. Hu, 2013, Seismic data analysis using synchrosqueezing wavelet transform: SEG expanded abstracts: 83rd Annual international meeting, 4330–4334.

Fourier finite-difference wave propagation

*Xiaolei Song and Sergey Fomel**

ABSTRACT

We introduce a novel technique for seismic wave extrapolation in time. The technique involves cascading a Fourier Transform operator and a finite difference operator to form a chain operator: Fourier Finite Differences (FFD). We derive the FFD operator from a pseudo-analytical solution of the acoustic wave equation. 2-D synthetic examples demonstrate that the FFD operator can have high accuracy and stability in complex velocity media. Applying the FFD method to the anisotropic case overcomes some disadvantages of other methods, such as the coupling of qP-waves and qSV-waves. The FFD method can be applied to enhance accuracy and stability of seismic imaging by reverse-time migration.

INTRODUCTION

The wavefield extrapolation problem refers to advancement of a wavefield through space or time. Both extrapolation in depth and extrapolation in time can be used in seismic modeling and seismic migration. Reverse time migration, or RTM (4; 20; 30; 17), involves wave extrapolation forward and backward in time. RTM is useful for accurate imaging in complex areas and is drawing more and more attention as the most powerful depth-imaging method (32; 27; 13; 16).

Reverse-time migration can correctly handle complex velocity models without dip limitations on the image. However, it has large memory requirements and needs a significant amount of computation. The most popular and straightforward way to implement reverse-time migration is the method of explicit finite differences, which is only conditionally stable because of the limit on time-step size. Finite-difference methods also suffer from numerical dispersion problems, which can be overcome either by decreasing the time step or by high-order schemes (31; 19). Several alternative algorithms have been developed for seismic wave extrapolation in variable velocity media. (26) introduced an algorithm based on a high-order differential operator, which allows a large extrapolation time step by solving a coefficient optimization problem. (34) proposed one-step extrapolation method by introducing a square-root operator. This method can formulate the two-way wave equation as a first-order partial differential equation in time similar to the one-way wave equation. (11) modified the Fourier Transform of the Laplacian operator to compensate exactly for the error

*e-mail: sergey.fomel@beg.utexas.edu

resulting from the second-order time marching scheme used in conventional pseudo spectral methods (22). (16) provided an accurate VTI P-wave modeling method with coupled second-order pseudo-acoustic equations. (21) presented an application of Rapid Expansion Method (REM) (28) for forward modeling with one-step time evolution algorithm and RTM with recursive time stepping algorithm.

In this paper, we present a new wave extrapolator derived from the pseudo-analytical approach of (11). Our method combines FFT and finite differences. We call it the Fourier Finite Difference method because it is analogous to the concept introduced previously for one-way wave extrapolation by (24).

As a chain operator of Fast Fourier Transform and Finite Difference operators, the proposed extrapolator can be as accurate as the parameter interpolation approach employed by (11) but at a cost of only one Fast Fourier Transform (FFT) and inverse Fast Fourier Transform (IFFT) operation. The advantages of the FFD operator are even more apparent in the anisotropic case: no need for several interpolations for different parameters with the corresponding computational burden of several FFTs and IFFTs. In addition, the operator can overcome the coupling of qP-waves and qSV-waves (33). We demonstrate the method on synthetic examples and propose to incorporate FFD into reverse-time migration in order to enhance migration accuracy and stability.

THEORY

The acoustic wave equation is widely used in forward seismic modeling and reverse-time migration (5; 12):

$$\frac{\partial^2 p}{\partial t^2} = v(\mathbf{x})^2 \, \nabla^2 p \; , \tag{1}$$

where $p(\mathbf{x}, t)$ is the seismic pressure wavefield, and $v(\mathbf{x})$ is the propagation velocity. Assuming a constant velocity v, after Fourier transform in space, we can get the following explicit expression:

$$\frac{d^2 \hat{p}}{dt^2} = -v^2 |\mathbf{k}|^2 \hat{p} \; , \tag{2}$$

where

$$\hat{p}(\mathbf{k}, t) = \int_{-\infty}^{+\infty} p(\mathbf{x}, t) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \; . \tag{3}$$

Equation 2 has the following solution:

$$\hat{p}(\mathbf{k}, t + \Delta t) = e^{\pm i |\mathbf{k}| v \Delta t} \hat{p}(\mathbf{k}, t) \; . \tag{4}$$

A second-order time-marching scheme and the inverse Fourier transform lead to the well-known expression (10; 26):

$$p(\mathbf{x}, t+\Delta t) + p(\mathbf{x}, t-\Delta t) - 2p(\mathbf{x}, t) = \\ 2\int_{-\infty}^{+\infty} \hat{p}(\mathbf{k}, t)(\cos(|\mathbf{k}|v\Delta t) - 1)e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \; . \tag{5}$$

Equation 5 provides an elegant and efficient solution in the case of a constant-velocity medium with the aid of FFT. In the case of a variable-velocity medium, equation 5 can provide an approximation by replacing v with $v(\mathbf{x})$. However, FFT can no longer be applied directly for the inverse Fourier transform from the wavenumber domain back to the space domain. To overcome this problem, (11) propose a velocity interpolation method. They present an implementation for isotropic, VTI (vertical transversely isotropic) and TTI (tilted transversely isotropic) media. In the isotropic case, two FFTs can be sufficient. For anisotropic media, more than one velocity parameter must be used. Therefore, it is necessary to perform velocity interpolation by combining different parameters and computing the corresponding forward and inverse FFTs for each of the velocity parameters, thus increasing the computational burden. Other FFT-based solutions include the optimized separable approximation or OSA (25; 18; 34; 9) and the lowrank approximation (15).

We propose an alternative approach. First, we adopt the following form of the right-hand side of equation 5 in the variable velocity case:

$$2\left[\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1\right] = \\ 2\left[\cos(v_0|\mathbf{k}|\Delta t) - 1\right]\left[\frac{\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1}{\cos(v_0|\mathbf{k}|\Delta t) - 1}\right] \; , \tag{6}$$

where v_0 is the reference velocity, such as the RMS (root-mean-square) velocity of the medium. After that, we apply the following approximation:

$$\frac{\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1}{\cos(v_0|\mathbf{k}|\Delta t) - 1} \approx a + 2\sum_{n=1}^{3} b_n \cos(k_n \Delta x_n) \; , \tag{7}$$

where coefficients a and b_n are defined using the Taylor expansion around $k = 0$,

$$a = \frac{v^2(\mathbf{x})}{v_0^2}\left[1 + \frac{(\Delta t)^2(v_0^2 - v^2(\mathbf{x}))(\Delta x_1^2\Delta x_2^2 + \Delta x_2^2\Delta x_3^2 + \Delta x_3^2\Delta x_1^2)}{6\Delta x_1^2\Delta x_2^2\Delta x_3^2}\right] \\ b_n = \frac{(\Delta t)^2 v^2(\mathbf{x})(v^2(\mathbf{x}) - v_0^2)}{12(\Delta x_n^2)v_0^2} \; , \tag{8}$$

and Δx_n is the sampling in the n-th direction. We only need to calculate these coefficients once. After completing the calculation, they can be used at each time step during the wave extrapolation process.

Equation 6 consists of two terms: the first term is independent of $\mathbf{x}$ and only depends on $\mathbf{k}$. For this part, we use inverse FFT to return to the space domain from the

wavenumber domain. For the remaining part, however, we can avoid phase shift in the wavenumber domain by implementing space shifts through finite differences (approximation 7) with coefficients provided by equation 8. This approach is analogous to the FFD method proposed by (24) for one-way extrapolation in depth.

Figure 1 (a) shows approximations for $[\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1]$ by the 4th-order FD method (dash line) and pseudo-spectral method (dotted line). Figure 1 (b) shows approximations by the FFD method (2nd-order: dash line, 4th-order: dotted line). The solid lines stand for the exact values for function $[\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1]$ with true velocity: $v = 4.0km/s$ (bottom solid line) and $v_0 = 2.0km/s$ (top solid line), which indicates a significant velocity contrast (100% difference). In this situation, all the approximations deviate from the exact solution as the wavenumber $|k|$ becomes large. However, the 4th-order FFD method approximates the exact solution with the most accuracy, as shown in the error plot (Figure 2). In order to enhance the stability, one can suppress the wavefield at high wavenumbers for both pseudo-spectral and the FFD method.

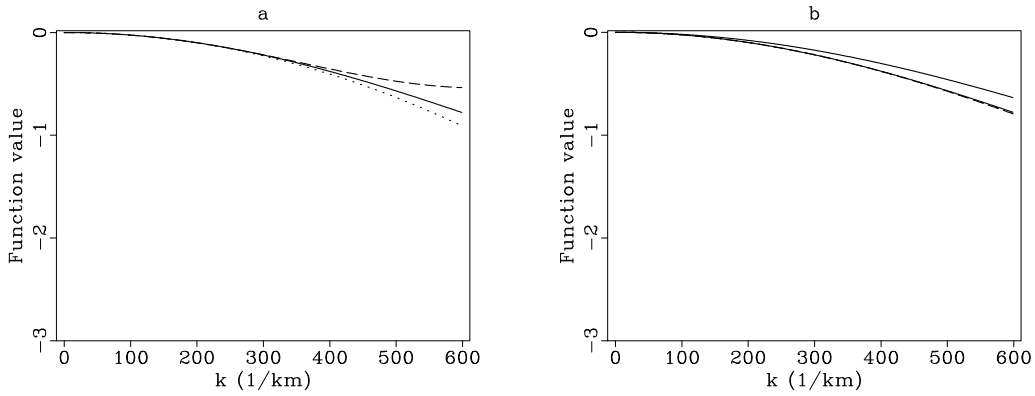


Figure 1: Different approximations for $\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1$. Solid lines: exact solution $(\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1)$ for $v = 4.0$ km/s and $v_0 = 2.0$ km/s. (a) Dash line: the 4th-order FD. Dotted line: pseudo-spectral method. (b) Dash line: the 2nd-order FFD method. Dotted line: the 4th-order FFD with v_0 as reference velocity. $\Delta t = 0.001$ s. $\Delta x = 0.005\,km$. ffd/cos cos-side1

Assuming that $p(\mathbf{x}, t - \Delta t)$ and $p(\mathbf{x}, t)$ are already known, the time-marching algorithm can be specified as follows:

1. Transform $p(\mathbf{x}, t)$ to $\hat{p}(\mathbf{k}, t)$ by 3-D FFT;

2. Multiply $\hat{p}(\mathbf{k}, t)$ by $2\,[\cos(v_0|\mathbf{k}|\Delta t) - 1]$ to get $\hat{q}(\mathbf{k}, t)$;

3. Transform $\hat{q}(\mathbf{k}, t)$ to $q(\mathbf{x}, t)$ by inverse FFT;

4. Apply finite differences to $q(\mathbf{x}, t)$ with coefficients in equation 8 to get $q(\mathbf{x}, t +$

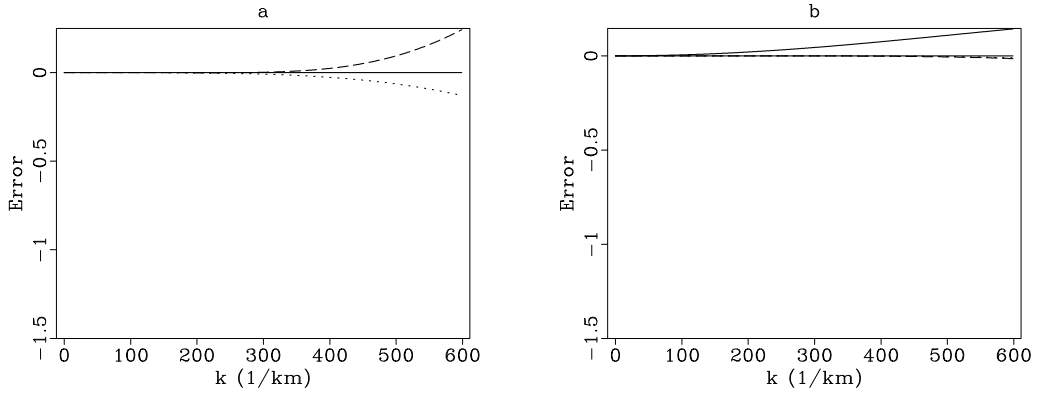


Figure 2: Errors for different approximations for $\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1$. Solid lines: exact solution ($\cos(v(\mathbf{x})|\mathbf{k}|\Delta t) - 1$) for $v = 4.0$ km/s and $v_0 = 2.0$ km/s. (a) Dash line: the 4th-order FD. Dotted line: pseudo-spectral method. (b) Dash line: the 2nd-order FFD method. Dotted line: the 4th-order FFD with v_0 as reference velocity. $\Delta t = 0.001$ s. $\Delta x = 0.005\, km$. ffd/cos diff1

Δt). Namely,

$$\begin{aligned} q^{i,j,k}(t+\Delta t) = a^{i,j,k} q^{i,j,k}(t) \\ +b_1^{i,j,k}(q^{i-1,j,k}(t) + q^{i+1,j,k}(t)) \\ +b_2^{i,j,k}(q^{i,j-1,k}(t) + q^{i,j+1,k}(t)) \\ +b_3^{i,j,k}(q^{i,j,k-1}(t) + q^{i,j,k+1}(t)) \,, \end{aligned} \tag{9}$$

where i is the grid index of x_i direction;

5. $p(\mathbf{x}, t+\Delta t) \leftarrow q(\mathbf{x}, t+\Delta t) + 2p(\mathbf{x}, t) - p(\mathbf{x}, t-\Delta t)$.

Here $q(\mathbf{x}, t)$ and $\hat{q}(\mathbf{k}, t)$ are temporary functions.

The FFD approach is not limited to the isotropic case. In the case of transversally isotropic (TTI) media, the term $v(\mathbf{x})\,|\mathbf{k}|$ on the left-hand side of equation 7, can be replaced with the acoustic approximation (1; 2; 14),

$$f(\mathbf{v}, \hat{\mathbf{k}}, \eta) = \sqrt{\frac{1}{2}(v_1^2\,\hat{k}_1^2 + v_2^2\,\hat{k}_2^2) + \frac{1}{2}\sqrt{(v_1^2\,\hat{k}_1^2 + v_2^2\,\hat{k}_2^2)^2 - \frac{8\eta}{1+2\eta} v_1^2 v_2^2\,\hat{k}_1^2\,\hat{k}_2^2}} \,, \tag{10}$$

where v_1 is the P-wave phase velocity in the symmetry plane, v_2 is the P-wave phase velocity in the direction normal to the symmetry plane, η is the anisotropic elastic parameter (3) related to Thomsen's elastic parameters ϵ and δ (29) by

$$\frac{1+2\delta}{1+2\epsilon} = \frac{1}{1+2\eta} \,;$$

and $\hat{k}_1$ and $\hat{k}_2$ stand for the wavenumbers evaluated in a rotated coordinate system aligned with the symmetry axis:

$$\begin{aligned} \hat{k}_1 &= k_1 \cos\phi + k_2 \sin\phi \\ \hat{k}_2 &= -k_1 \sin\phi \cos\theta + k_2 \cos\phi \cos\theta + k_3 \sin\theta \\ \hat{k}_3 &= k_1 \sin\phi \sin\theta - k_2 \cos\phi \sin\theta + k_3 \cos\theta \ , \end{aligned} \tag{11}$$

where θ is the tilt angle measured with respect to vertical and ϕ is the angle between the projection of the symmetry axis in the horizontal plane and the original X-coordinate. The symmetry axis has the direction of $\{-\sin\theta\sin\phi, -\sin\theta\cos\phi, \cos\theta\}$.

Using these definitions, we develop a finite-difference approximation analogous to equation 7 for FFD in TTI media. The details of the derivation are given in the appendix. For the 2D TTI case, the corresponding FFD algorithm is as following:

1. Transform $p(\mathbf{x},t)$ to $\hat{p}(\mathbf{k},t)$ by FFT;
2. Multiply $\hat{p}(\mathbf{k},t)$ by $\dfrac{2[\cos(f(\mathbf{v_0},\hat{\mathbf{k}},\eta_0)\Delta t)-1]}{|\hat{\mathbf{k}}|^2}$ to get $\hat{q}(\mathbf{k},t)$;
3. Transform $\hat{q}(\mathbf{k},t)$ to $q(\mathbf{x},t)$ by inverse FFT;
4. Apply finite differences to $q(\mathbf{x},t)$ with coefficients in Table 1 to get $q(\mathbf{x},t+\Delta t)$. Namely,

$$\begin{aligned} q^{i,j,t+\Delta t} = {} & a^{i,j} q^{i,j,t} \\ & + b_1^{i,j}(q^{i-1,j,t} + q^{i+1,j,t}) \\ & + b_2^{i,j}(q^{i,j-1,t} + q^{i,j+1,t}) \\ & + d_1^{i,j}(q^{i-2,j,t} + q^{i+2,j,t}) \\ & + d_2^{i,j}(q^{i,j-2,t} + q^{i,j+2,t}) \\ & + c^{i,j}(q^{i-1,j-1,t} + q^{i-1,j+1,t} + q^{i+1,j-1,t} + q^{i+1,j+1,t}) \ . \end{aligned} \tag{12}$$

   where i is the grid index of x_i direction;
5. $p(\mathbf{x},t+\Delta t) \leftarrow q(\mathbf{x},t+\Delta t) + 2p(\mathbf{x},t) - p(\mathbf{x},t-\Delta t)$.

Here, $q(\mathbf{x},t)$ and $\hat{q}(\mathbf{k},t)$ are temporary functions.

NUMERICAL EXAMPLES

Our first example is a comparison of four methods: the 4th-order finite differences, pseudo-spectral method, velocity interpolation, and the FFD method in a velocity model with smooth variation, formulated as

$$v(x,z) \ = \ 550 + 1.5 \times 10^{-4}(x-800)^2 + 10^{-4}(z-500)^2;$$

$$0 \le x \le 2560,\ 0 \le z \le 2560.$$

The velocity is between 550 m/s and 1439 m/s. A Ricker-wavelet source with maximum frequency 70 Hz is located at the center of the model. For all the numerical simulations based on this model, we use the same grid size: $\Delta x = 5$ m and $\Delta t = 2$ ms.

Figure 3(a) shows an obvious numerical dispersion from the snapshot of the acoustic wavefield computed by the 4th-order finite difference method. Figure 3(b) shows a slight dispersion from the snapshot computed by pseudo-spectral method (23). Figure 3(c) shows the corresponding snapshot of the velocity-interpolation method (11; 8), calculated using two reference velocities. It is practically free of dispersion thanks to spectral compensation. Figure 3(d) shows a snapshot of the proposed FFD method. It is almost exactly the same as Figure 3(c); however, only one reference velocity is used instead of two. As comparison between Figure 3(d) and Figure 3(c) implies, the FFD method has practically the same accuracy as the velocity interpolation method while having only one reference velocity and therefore replacing the cost of one additional FFT with the cost of a low-order finite-difference operator.

Our next example is a snapshot of the acoustic wavefield calculated by FFD in the BP model (6). We use a Ricker-wavelet at a point source. The maximum frequency is 50 Hz. The horizontal grid size Δx is 37.5 m, the vertical grid size Δz is 12.5 m and the time step is 1 ms. Figure 13(a) shows a part of the model with a salt body. Figure 5 shows a wavefield snapshot confirming that the FFD method can work in a complex-velocity medium as well.

Next we apply our FFD algorithm to RTM with a simple exponential decaying boundary condition (7). The dominant frequency is 27 Hz. The space grid is 25 m and the time step is 1.5 ms. Figure 5(b) shows the output image. The inner and outer flanks of the salt body are clearly imaged.

The cost advantage of FFD is even more appealing in anisotropic (TTI) media, which require multiple velocity parameters and increase the cost of velocity interpolation. Figure 7 shows the impulse response of a 4th-order FFD operator in a TTI model with the tilt of 45 ° and a smooth velocity variation (v_x: 800-1225.41 m/s, v_z: 700-883.6 m/s). The space grid size is 5 m and the time step is 1 ms. Note no coupling of qP-waves and qSV-waves (33) in the figure, thanks to the Fourier construction of the operator.

For this paper, we implemented a 2nd-order 5-point FD scheme for 2D isotropic case. One can observe little dispersion in isotropic examples. For TTI media, we use a 13-point scheme which minimizes the error in the symmetry plane and in the direction normal to the symmetry plane. One can still see some dispersion in the corresponding snapshot; which indicates that a higher order FD scheme might be required to further supress the dispersion.

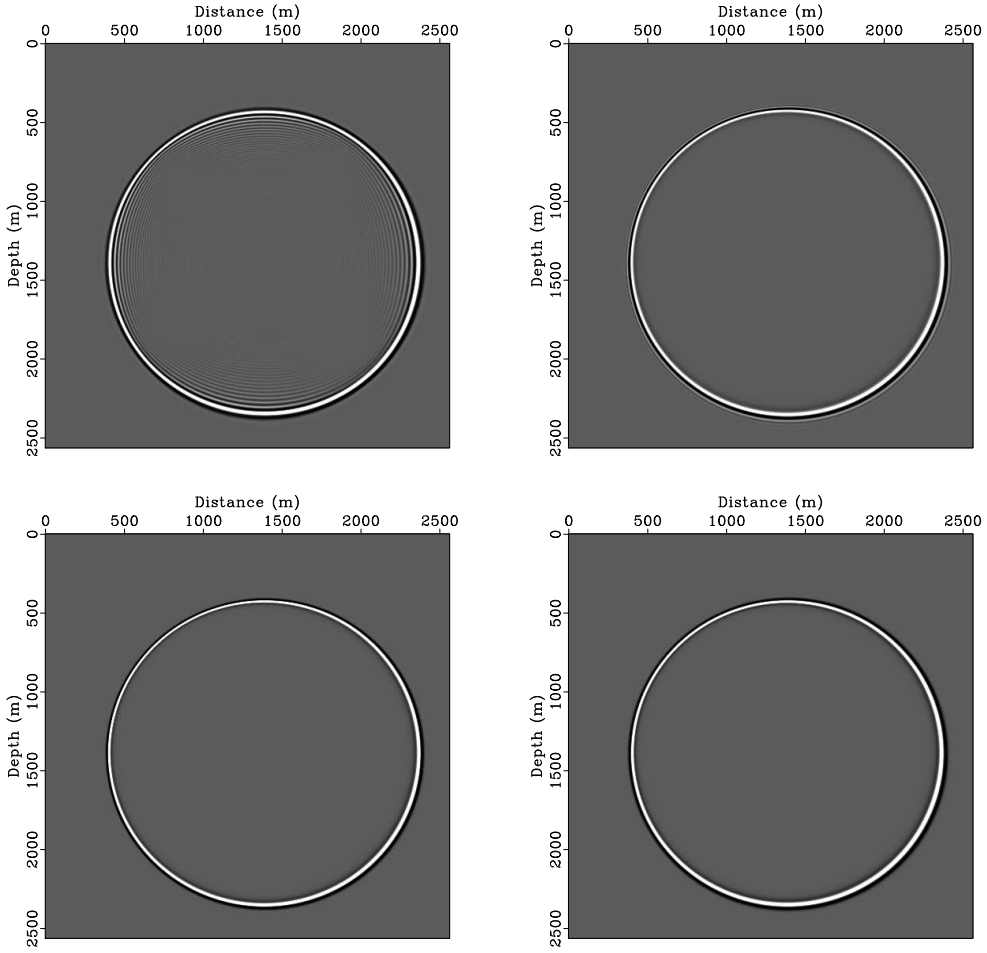


Figure 3: Acoustic wavefield snapshot by: (a) 4th-order Finite Difference method; (b) pseudo-spectral method; (c) velocity interpolation method with 2 reference velocities; (d) FFD method with RMS velocity. ffd/ffd wavfd,wavsp,wavpspi,wavffd

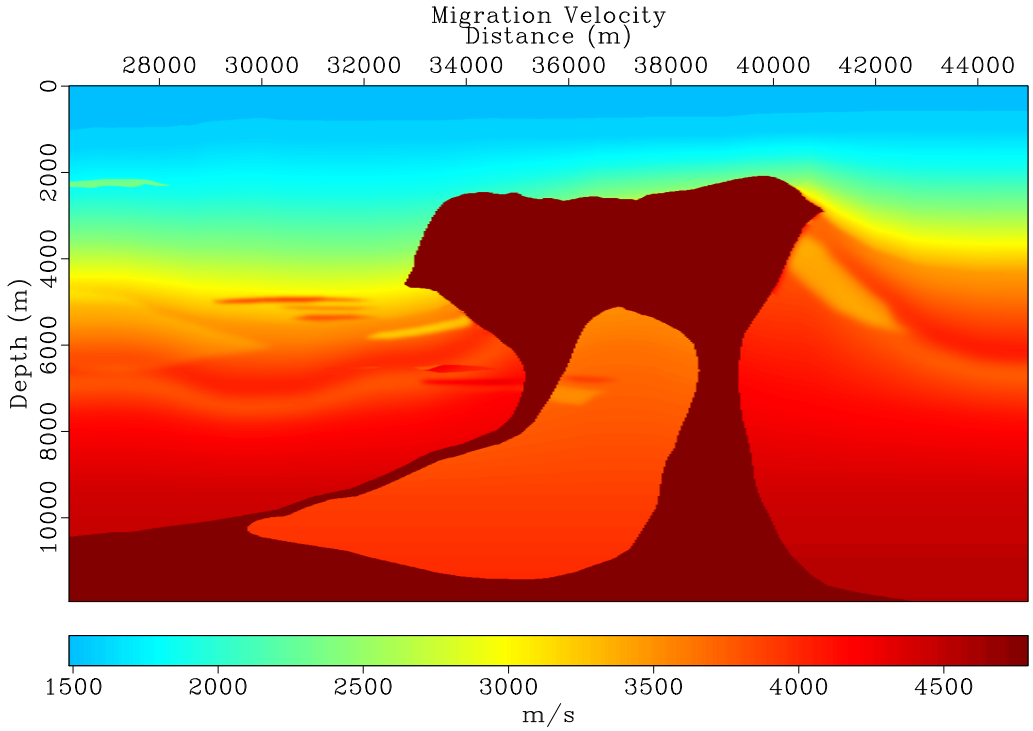


Figure 4: Portion of BP 2004 synthetic velocity model. ffd/bpmodel vel

Our last example is qP-wave simulation in the BP 2D TTI model. Figure 8(a)-8(d) shows parameters for part of the model. The maximum frequency is 50 Hz. The space grid size is 12.5 m and the time step is 1 ms. The snapshot of the acoustic wavefield in Figure 9 demonstrates the stability of our approach in a complicated anisotropic model. Some small dispersion is present in the TTI examples, pointing to a possible need to extend the FD part of the FFD scheme from second order to higher orders.

CONCLUSIONS

Accurate and efficient numerical wave propagation in variable velocity media is crucial for seismic modeling and seismic migration, particularly for reverse-time migration. The FFD technique proposed in this paper promises higher accuracy than that of the conventional, explicit finite-difference method, at the cost of only one forward and inverse Fast Fourier Transform. Results in synthetic isotropic and anisotropic models illustrate FFD's stability in complicated velocity models. The method can be used in reverse-time migration to enhance its accuracy and stability.

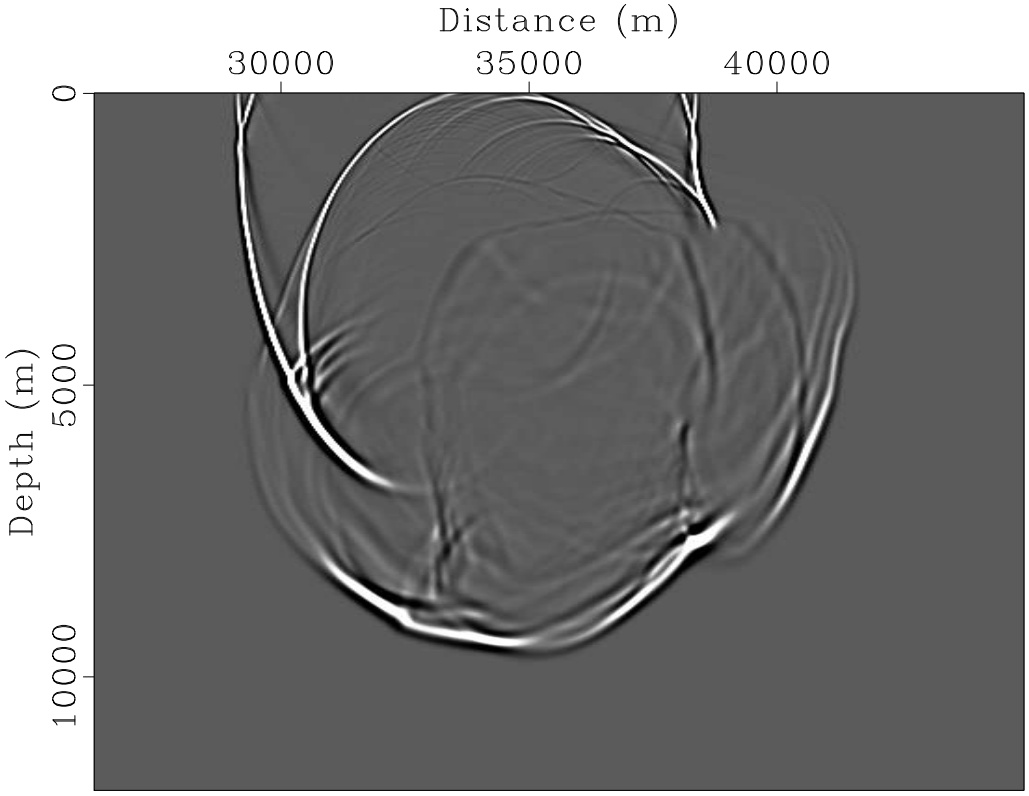


Figure 5: Wavefield snapshot in the BP Model shown in Figure 13(a). ffd/bpmodel wavsnap

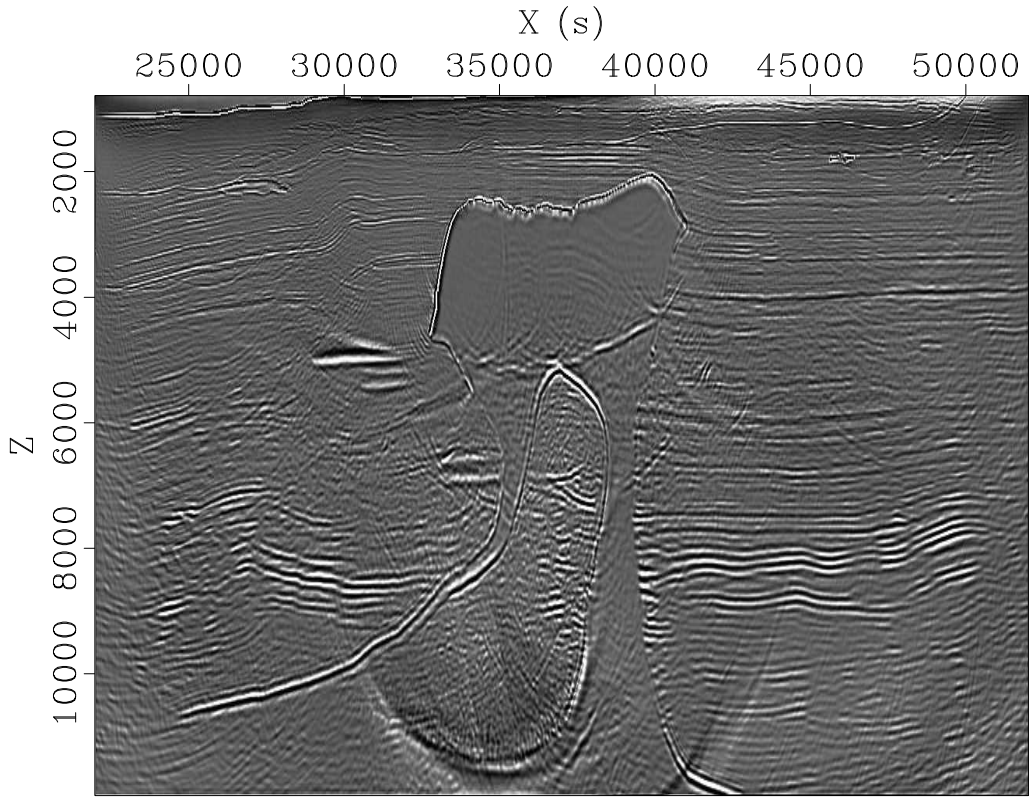


Figure 6: RTM image of BP salt model. ffd/. img

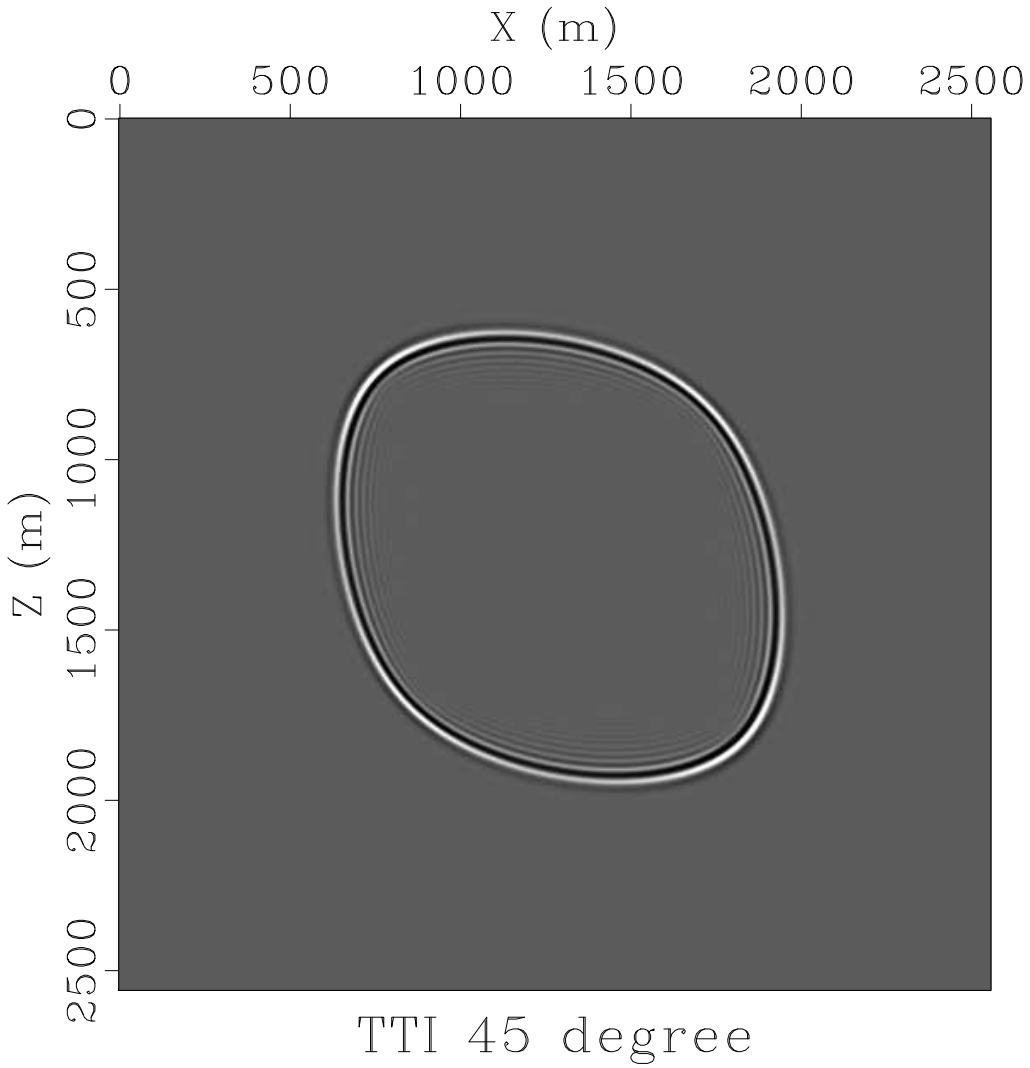


Figure 7: Wavefield snapshot in a TTI medium with tilt of 45 degrees. $v_x(x,z) = 800 + 10^{-4}(x-1000)^2 + 10^{-4}(z-1200)^2$; $v_z(x,z) = 700 + 10^{-4}(z-1200)^2$; $\eta = 0.3$; $\theta = 45^\circ$. ffd/anisotropic TTI-snapshot

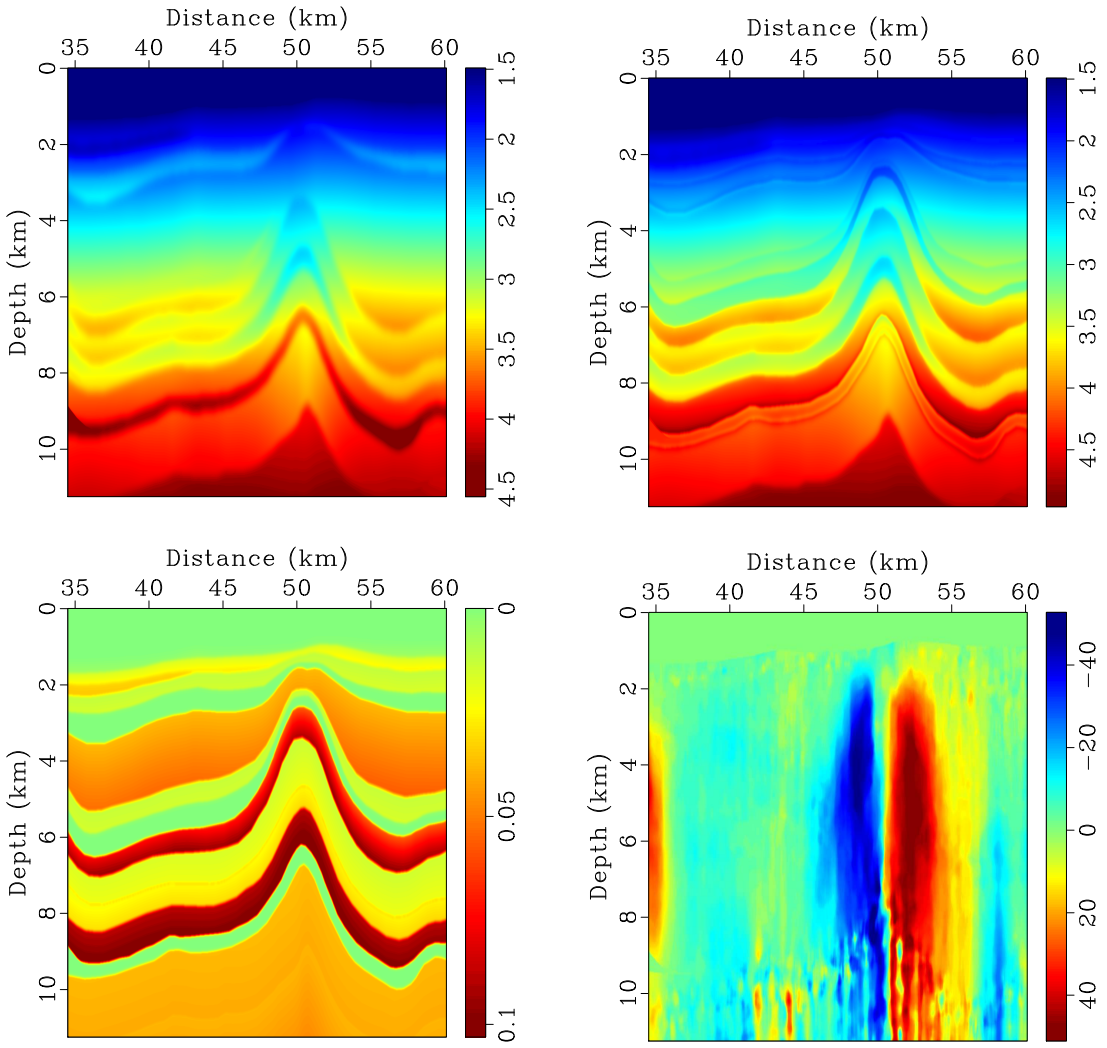


Figure 8: Partial region of the 2D BP TTI model. a: v_z. b: v_x. c: η. d:θ. ffd/bptti vp2,vx2,yita2,theta2

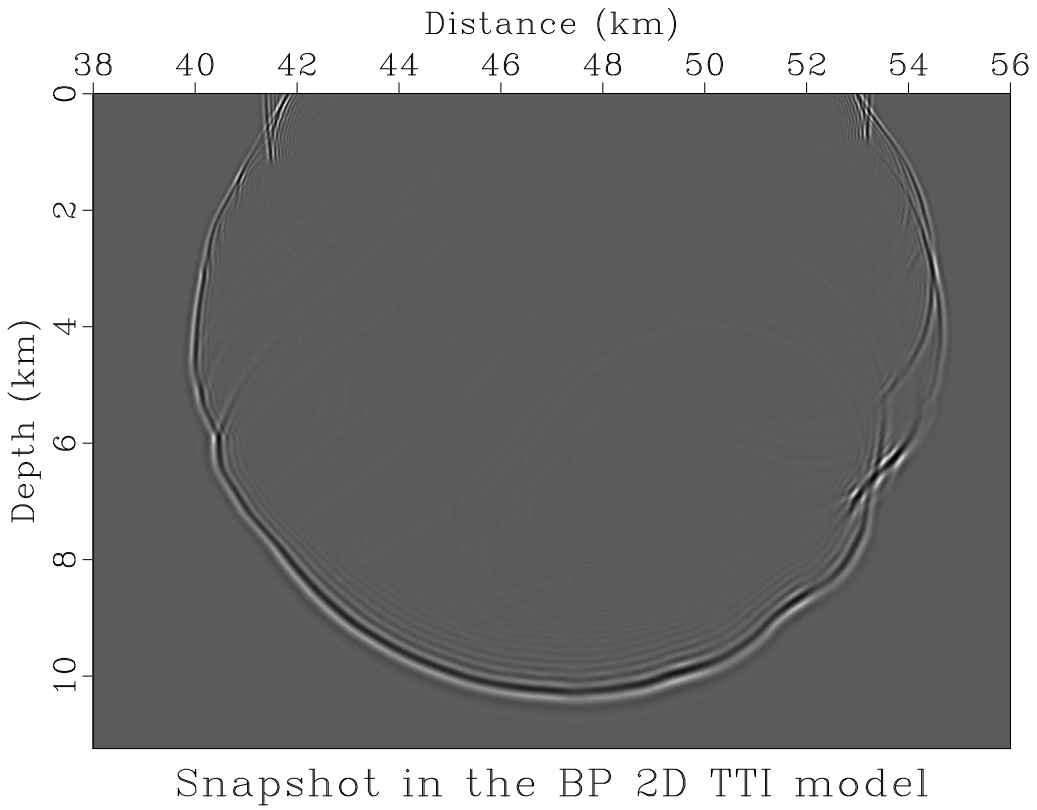


Figure 9: Scalar wavefield snapshot in the 2D BP TTI model, shown in Figure 8. ffd/bptti ttisnapshot

ACKNOWLEDGEMENT

We thank BP for releasing benchmark synthetic models, Bjorn Enquist, Paul Fowler and Lexing Ying for useful discussions, and John Etgen, Stig Hestholm, Erik Saenger and two anonymous reviewers for helpful reviews.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

FFD FOR TTI MEDIA

To develop a 25-point finite-difference scheme analogous to equation 7 for FFD in 3D TTI media, we first apply the following approximation:

$$\begin{aligned}\frac{\cos(f(\mathbf{v},\hat{\mathbf{k}},\eta)\Delta t)-1}{\cos(f(\mathbf{v_0},\hat{\mathbf{k_0}},\eta_0)\Delta t)-1}|\hat{\mathbf{k}}|^2 \approx \\ a+2\sum_{n=1}^{3}\left(b_n\cos(k_n\Delta x_n)+d_n\cos(2k_n\Delta x_n)\right) \\ +2\sum_{n=1}^{3}c_n[\cos(k_i\Delta x_i+k_j\Delta x_j)+\cos(k_i\Delta x_i-k_j\Delta x_j)]\ ,\end{aligned} \tag{A-1}$$

where $i,j=1,2,3$; $i\neq j$; $i,j\neq n$.

In approximation A-1, $f(\mathbf{v},\hat{\mathbf{k}},\eta)$ is a function as in expression 10 and a, b_n, c_n and d_n are coefficients determined from the Taylor expansion around $k=0$.

Notice that we multiply the left-hand side with $|\hat{\mathbf{k}}|^2$, so one needs to multiply $\hat{p}(\hat{\mathbf{k}},\mathbf{t})$ with $\dfrac{2[\cos(f(\mathbf{v_0},\mathbf{k},\eta_0)\Delta t)-1]}{|\hat{\mathbf{k}}|^2}$.

The coefficients for Equation A-1 are derived in Table 1 and Table 2. w_{n0}, h_{n0}, p_{n0} and q_{n0} have similar expressions as above in Table 2 with $\mathbf{v}$, η and θ substited by the corresponding reference values: RMS velocity $\mathbf{v}_0$, average anisotropic parameter η_0 and average tilt angles θ_0 and ϕ_0.

REFERENCES

[1] Alkhalifah, T., 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.

[2] ———, 2000, Acoustic approximations for processing in transversely isotropic media: Geophysics, **65**, 1239–1250.

[3] Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

[4] Baysal, E., D. D. Kosloff, and J. W. C. Sherwood, 1983, Reverse time migration: Geophysics, **48**, 1514–1524.

a	$a = -2b_1 - 2b_2 - 2b_3 - 4c_1 - 4c_2 - 4c_3 - 2d_1 - 2d_2 - 2d_3$
b	$b_1 = -2c_2 - 2c_3 - 4d_1 - \dfrac{w_1 + h_1}{\Delta x_1^2 (w_{10} + h_{10})}$
c	$b_2 = -2c_1 - 2c_3 - 4d_2 - \dfrac{w_2 + h_2}{\Delta x_2^2 (w_{20} + h_{20})}$
d	$b_3 = -2c_1 - 2c_2 - 4d_3 - \dfrac{w_3 + h_3}{\Delta x_3^2 (w_{30} + h_{30})}$
e	$d_1 = \dfrac{(w_1 + h_1)(2x_1^2 + \Delta t^2 (w_{10} + h_{10} - w_1 - h_1))}{24 \Delta x_1^4 (w_{10} + h_{10})}$
f	$d_2 = \dfrac{(w_2 + h_2)(2x_2^2 + \Delta t^2 (w_{20} + h_{20} - w_2 - h_2))}{24 \Delta x_2^4 (w_{20} + h_{20})}$
g	$d_3 = \dfrac{(w_3 + h_3)(2x_3^2 + \Delta t^2 (w_{30} + h_{30} - w_3 - h_3))}{24 \Delta x_3^4 (w_{30} + h_{30})}$
h	$c_1 = \dfrac{1}{12 \Delta x_2^2 \Delta x_3^2} \left[\dfrac{\Delta x_2^2 (w_2 + h_2)}{w_{20} + h_{20}} + \dfrac{\Delta x_3^2 (w_3 + h_3)}{w_{30} + h_{30}} \right] - d_2 \dfrac{\Delta x_2^2}{\Delta x_3^2} - d_3 \dfrac{\Delta x_3^2}{\Delta x_2^2} + \dfrac{\Delta t^2 (p_1 + q_1)(p_1 0 + q_1 0 - p1 - q1)}{12 \Delta x_2^2 \Delta x_3^2 (p_1 0 + q_1 0)}$
i	$c_2 = \dfrac{1}{12 \Delta x_1^2 \Delta x_3^2} \left[\dfrac{\Delta x_1^2 (w_1 + h_1)}{w_{10} + h_{10}} + \dfrac{\Delta x_3^2 (w_3 + h_3)}{w_{30} + h_{30}} \right] - d_1 \dfrac{\Delta x_1^2}{\Delta x_3^2} - d_3 \dfrac{\Delta x_3^2}{\Delta x_1^2} + \dfrac{\Delta t^2 (p_2 + q_2)(p_2 0 + q_2 0 - p2 - q2)}{12 \Delta x_1^2 \Delta x_3^2 (p_2 0 + q_2 0)}$
j	$c_3 = \dfrac{1}{12 \Delta x_1^2 \Delta x_2^2} \left[\dfrac{\Delta x_1^2 (w_1 + h_1)}{w_{10} + h_{10}} + \dfrac{\Delta x_2^2 (w_2 + h_2)}{w_{20} + h_{20}} \right] - d_1 \dfrac{\Delta x_1^2}{\Delta x_2^2} - d_2 \dfrac{\Delta x_2^2}{\Delta x_1^2} + \dfrac{\Delta t^2 (p_3 + q_3)(p_3 0 + q_3 0 - p3 - q3)}{12 \Delta x_1^2 \Delta x_2^2 (p_3 0 + q_3 0)}$

Table 1: Coefficients for equation A-1.

a	$w_1 = v_1^2 \cos^2\phi + \sin^2\phi(v_1^2\cos^2\theta + v_2^2\sin^2\theta)$
b	$w_2 = (v_1^2 + v_2^2)\cos^2\phi\cos^2\theta + v_1^2\sin^2\theta$
c	$w_3 = v_1^2\sin^2\theta + v_2^2\cos^2\theta$
d	$h_1 = \sqrt{w1^2 - \dfrac{8\eta v_1^2 v_2^2\sin^2\phi(\cos^2\phi + \sin^2\phi\cos^2\theta)\sin^2\theta}{1+2\eta}}$
e	$h_2 = \sqrt{w2^2 - \dfrac{8\eta v_1^2 v_2^2\cos^2\phi\cos^2\theta(\cos^2\phi\cos^2\theta + \sin^2\phi)}{1+2\eta}}$
f	$h_3 = \sqrt{w3^2 - \dfrac{8\eta v_1^2 v_2^2\cos^2\theta\sin^2\theta}{1+2\eta}}$
g	$p_1 = w_2 + w_3 + v_1^2\cos\phi\sin 2\theta - 2v_2^2\cos\phi\cos^2\theta$
h	$q_1 = \sqrt{p_1^2 - \dfrac{32\eta v_1^2 v_2^2\cos^2\theta\sin^4\frac{\phi}{2}(\cos^2\phi\cos^2\theta + \sin^2\theta + \sin^2\phi + \cos\phi\sin 2\theta)}{1+2\eta}}$
i	$p_2 = w_1 + w_3 + (v_2^2 - v_1^2)\sin\phi\sin 2\theta$
j	$q_2 = \sqrt{p_2^2 - \dfrac{8\eta v_1^2 v_2^2(\cos\theta + \sin\phi\sin\theta)^2(\cos^2\phi + (\cos\theta\sin\phi - \sin\theta)^2)}{1+2\eta}}$
k	$p_3 = w_1 + w_2 + v_1^2\sin^2\theta\sin 2\phi + \frac{1}{2}v_2^2\sin 2\phi\sin 2\theta$
l	$q_3 = \sqrt{p_3^2 - \dfrac{4\eta v_1^2 v_2^2\cos^2(\phi+\theta)(\sin 2\phi\cos 2\theta - 3 - \cos 2\theta - \sin 2\phi)}{1+2\eta}}$

Table 2: Coefficients for Table 1.

[5] Bednar, J., 2005, A brief history of seismic migration: Geophysics, **70**, 3MJ–20MJ.
[6] Billette, F. J., and S. Brandsberg-Dahl, 2004, The 2004 BP velocity benchmark: 67th Annual EAGE Meeting, EAGE, Expanded Abstracts, B305.
[7] Cerjan, C., D. Kosloff, R. Kosloff, and M. Reshef, 1985, A nonreflecting boundary condition for discrete acoustic and elastic wave equations: Geophysics, **50**, 705–708.
[8] Crawley, S., S. Brandsberg-Dahl, and J. McClean, 2010, 3d TTI RTM using the pseudo-analytic method: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3216–3220.
[9] Du, X., R. P. Fletcher, and P. J. Fowler, 2010, Pure P-wave propagators versus pseudo-acoustic propagators for RTM in VTI meida: 72nd Annual EAGE Meeting, EAGE, Expanded Abstracts, Accepted.
[10] Etgen, J., 1989, Accurate wave equation modeling, *in* SEP-60: Stanford Exploration Project, 131–148.
[11] Etgen, J., and S. Brandsberg-Dahl, 2009, The pseudo-analytical method: application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation: 79nd Annual International Meeting, SEG, Expanded Abstracts, 2552–2556.
[12] Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: Geophysics, **74**, WCA5–WCA17.
[13] Fletcher, R. P., X. Du, and P. J. Fowler, 2009, Reverse time migration in tilted transversely isotropic (TTI) media: Geophysics, **74**, WCA179–WCA187.
[14] Fomel, S., 2004, On anelliptic approximations for qP velocities in VTI media: Geophysical Prospecting, **52**, 247–259.
[15] Fomel, S., L. Ying, and X. Song, 2010, Seismic wave extrapolation using a lowrank symbol approximation: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3092–3096.
[16] Fowler, P. J., X. Du, and R. P. Fletcher, 2010, Coupled equations for reverse time migration in transversely isotropic media: Geophysics, **75**, S11–S22.
[17] Levin, S. A., 1984, Principle of reverse time migration: Geophysics, **49**, 581–583.
[18] Liu, F., S. A. Morton, S. Jiang, L. Ni, and J. P. Leveille, 2009, Decoupled wave equations for P and SV waves in an acoustic VTI media: 79th Ann. Internat. Mtg., Soc. Expl. Geophys., 2844–2848.
[19] Liu, Y., and M. K. Sen, 2009, A new time-space domain high-order finite-difference method for the acoustic wave equation: Journal of computational Physics, **228**, 8779–8806.
[20] McMechan, G. A., 1983, Migration by extrapolation of time-dependent boundary values: Geophys. Prosp., **31**, 413–420.
[21] Pestana, R. C., and P. L. Stoffa, 2010, Time evolution of the wave equation using rapid expansion method: Geophysics, **75**, T121–T131.
[22] Reshef, M., D. Kosloff, M. Edwards, and C. Hsiung, 1988a, Three-dimensional acoustic modeling by the Fourier method: Geophysics, **53**, 1175–1183.
[23] ———, 1988b, Three-dimensional acoustic modeling by the fourier method: Geophysics, **53**, 1175–1183.
[24] Ristow, D., and T. Ruhl, 1994, Fourier finite-difference migration: Geophysics, **59**, 1882–1893.
[25] Song, J., 2001, The optimized expression of a high dimensional function/manifold

in a lower dimensional space: Chinese Scientific Bulletin, **46**, 977–984.
[26] Soubaras, R., and Y. Zhang, 2008, Two-step explicit marching method for reverse time migration: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2272–2276.
[27] Symes, W. W., 2007, Reverse time migration with optimal checkpointing: Geophysics, **72**, SM213–SM221.
[28] Tal-Ezer, H., D. Kosloff, and Z. Koren, 1987, An accurate scheme for seismic forward modeling: Geophysical Prosecting, **35**, 479–490.
[29] Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.
[30] Whitmore, N. D., 1983, Iterative depth migration by backward time propagation: 53rd Ann. Internat. Mtg., Soc. Expl. Geophys., 382–385.
[31] Wu, W., L. R. Lines, and H. Lu, 1996, Analysis of high-order, finite-difference schemes in 3-D reverse-time migration: Geophysics, **61**, 845–856.
[32] Yoon, K., K. J. Marfurt, and W. Starr, 2004, Challenges in reverse-time migration: 74th Ann. Internat. Mtg., Soc. Expl. Geophys., 1057–1061.
[33] Zhang, H., G. Zhang, and Y. Zhang, 2009, Removing S-wave noise in TTI reverse time migration: 79th Ann. Internat. Mtg., Soc. Expl. Geophys., 2849–2853.
[34] Zhang, Y., and G. Zhang, 2009, One-step extrapolation method for reverse time migration: Geophysics, **74**, A29–A33.

Omnidirectional plane-wave destruction

Zhonghuan Chen, Sergey Fomel†, and Wenkai Lu***

ABSTRACT

Steep structures in seismic data may bring directional aliasing, thus plane-wave destruction (PWD) filter can not obtain an accurate dip estimation. We propose to interpret plane-wave construction (PWC) filter as a line-interpolating operator and introduce a novel circle-interpolating model. The circle-interpolating PWC can avoid phase-wrapping problems, and the related circle-interpolating plane-wave destruction (PWD) can avoid aliasing problems. We design a 2D maxflat fractional delay filter to implement the circle interpolation, and prove that the 2D maxflat filter is separable in each direction. Using the maxflat fractional delay filter in the circle interpolation, we propose the omnidirectional plane-wave destruction (OPWD). The omnidirectional PWD can handle both vertical and horizontal structures. With a synthetic example, we show how to obtain an omnidirectional dip estimation using the proposed omnidirectional PWD. An application of the omnidirectional PWD to a field dataset improves the results of predictive painting and event picking, as compared to conventional PWD.

INTRODUCTION

Seismic wavefields can be approximated by local plane waves, and the local slope field of these plane waves has been a significant parameter in seismic analysis. Plane-wave destruction (PWD) is one of the most popular tools to estimate the slope field. The linear PWD proposed by (4) uses explicit finite differences and obtains the slope field by least-squares estimation. The nonlinear PWD proposed by (6) uses the maxflat fractional delay filter (18; 20) to approximate a linear phase operator (or phase shift operator) and to provide a polynomial equation for the local slope (3). When the nonlinear PWD is applied iteratively, its first iteration corresponds to the linear PWD.

Local slope field has been widely applied in seismic applications. In applications such as model parameterization (9; 10), trace interpolation (1) and wavefield separation and denoising (12) in prestack datasets, seismic events usually have moderate slopes. However, in other applications, there might be steep or even vertical structures in the data and the slopes might be large or even infinite. Some common examples are: (1) migrated datasets, where geological structures can be steeply dipping, and attribute analysis may require the slope field (14); (2) time slices, where azimuths can follow any directions (13, Figure 2a is a good example); (3) profiles with insufficient

***e-mail:** sergey.fomel@beg.utexas.edu

horizontal sampling interval, causing dip aliasing problems (2). (11) has shown that neither linear nor nonlinear PWD can cope with these steep structures well. In this case, the PWD-based slope estimation may be inaccurate.

In order to handle steep structures, several methods have been proposed to improve the linear PWD: (5) and (15) introduced other finite-difference methods to obtain a better phase-shift response. (11) applied the linear PWD in both horizontal and vertical directions, to construct the directional Laplacian operator. (16) proposed total least-squares estimation to improve the least-squares estimation.

In this paper, we propose to interpret the phase shift operator in the nonlinear PWD as a 1D wavefield interpolator (vertical interpolation in a seismic trace). In order to handle omnidirectional structures, we introduce a 2D interpolator, which interpolates the wavefield along a circle instead of a vertical line. Circle interpolation can avoid aliasing problems and enable efficient modeling of steep structures.

We design a 2D maxflat fractional delay filter to implement circle interpolation. This 2D filter can be decoupled into a cascade of two 1D filters applied in different directions. Using the polynomial design of the maxflat fractional delay filter (3), we propose an omnidirectional plane-wave destruction (OPWD). We use a synthetic example to demonstrate the omnidirectional modeling ability and apply OPWD to improve events picking for a field dataset.

FROM LINE INTERPOLATION TO CIRCLE INTERPOLATION

Considering the wavefield $u(x_1, x_2)$ observed in a 2D sampling system and following (6), plane-wave destruction can be represented in the Z-transform domain as

$$(1 - Z_2 Z_1^p) U(Z_1, Z_2) = 0 \, , \tag{1}$$

where Z_1, Z_2 are the unit shift operators in the first and second dimensions. $U(Z_1, Z_2)$ (or U for convenience) denotes the Z transform of $u(x_1, x_2)$. p is the local slope. We call $Z_2 Z_1^p$ and $1 - Z_2 Z_1^p$ plane-wave constructor and destructor, respectively. The slope p has the following relationship with the dip angle θ: $p = \tan\theta$.

Applying $Z_2 Z_1^p$ at one point, for example, point O in Figure 1, PWC obtains the wavefield at the point with a unit shift in the second dimension and p unit shifts in the first dimension, denoted by $A(x_1 + p\Delta x_1, x_2 + \Delta x_2)$. As $-\pi/2 \le \theta \le \pi/2$, p can be any value from $-\infty$ to $+\infty$. That is to say, the forward plane-wave constructor $Z_2 Z_1^p$ interpolates the wavefield along the vertical line at $x_2 + \Delta x_2$. Similarly, the backward PWC interpolates the wavefield along the vertical line at $x_2 - \Delta x_2$.

In order to handle both vertical and horizontal structures, we propose to modify the plane-wave destruction in equation 3 into the following form:

$$(1 - Z_2^{p_2} Z_1^{p_1}) U = 0, \tag{2}$$

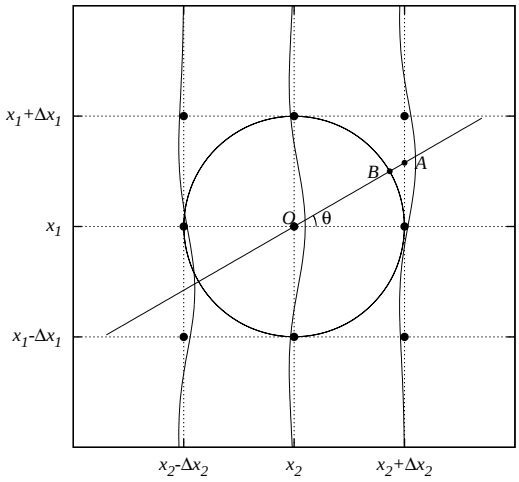


Figure 1: Interpolation in plane-wave construction: line-interpolating PWC interpolates the wavefield at point A, while circle-interpolating PWC interpolates at point B. opwd/Gnuplot opwd

where p_1, p_2 are parameters related to the trial dip angle, as follows: $p_1 = r \sin\theta$, $p_2 = r \cos\theta$.

In other words, we consider a circle in polar coordinates, parameterized by the radius r and the dip angle θ. Applying the new PWC $Z_2^{p_2} Z_1^{p_1}$ at point O, it obtains the wavefield at the point with p_1 unit shifts in the first dimension and p_2 unit shifts in the second dimension. That is point $B(x_1 + p_1\Delta x_1, x_2 + p_2\Delta x_2)$. As θ changes, the new PWC $Z_1^{p_1} Z_2^{p_2}$ interpolates the wavefield along a circle with radius r. We draw the interpolating circle with $r = 1$ in Figure 1. The circle-interpolating PWC $Z_1^{p_1} Z_2^{p_2}$ corresponds to a 2D interpolation. Equation 3 can also be seen as a special case of equation 2 when $p_2 = 1$. Compared with the 1D line-interpolating method, the main benefit of circle interpolation is its antialiasing ability.

Anti-aliasing ability

We compare the line-interpolating and circle-interpolating PWD operators in the frequency domain. At different dip angles, the magnitude responses of $1 - Z_2 Z_1^p$ and $1 - Z_1^{p_1} Z_2^{p_2}$ are shown in Figure 2: When dip angle $\theta = 20°$, the two operators have similar responses (Figure 2a and 2d); when $\theta = 50°$, the line-interpolating PWD become slightly aliased (Figure 2b), while the circle-interpolating PWD is not aliased (Figure 2e); as θ increases to 80°, the former is badly aliased (Figure 2c), and the latter is still not aliased (Figure 2f).

In summary, the line-interpolating PWD has different frequency responses for

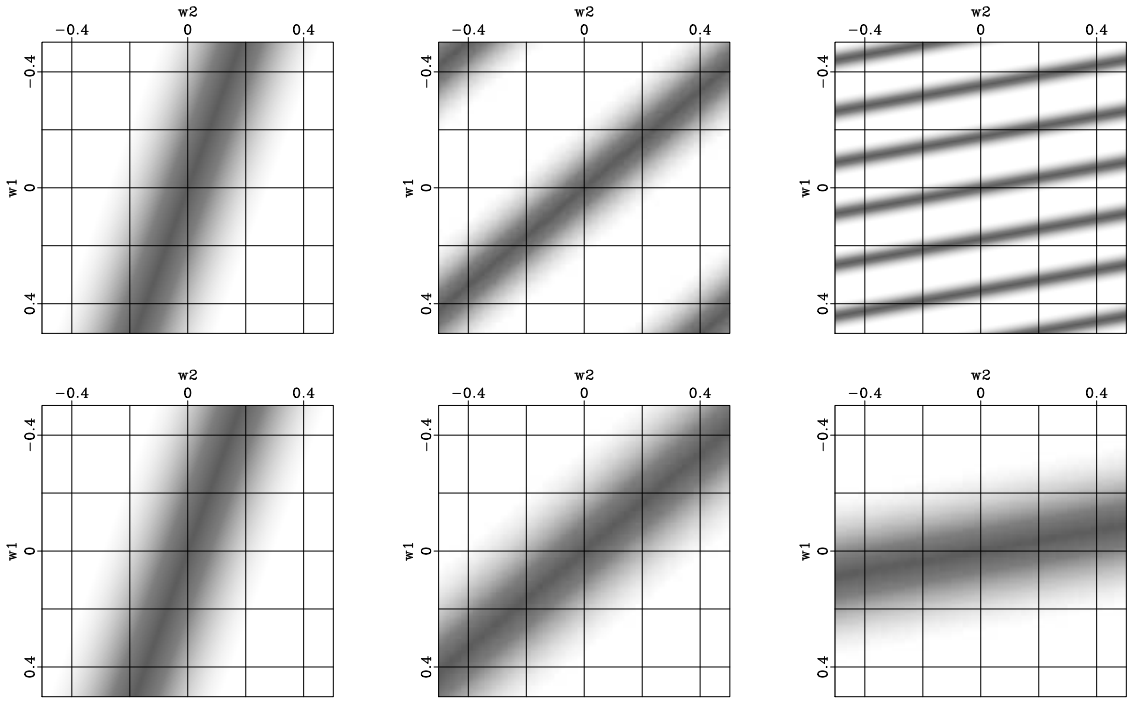


Figure 2: Magnitude responses of the line-interpolating PWD $1 - Z_1 Z_2^p$ (top) and the circle-interpolating PWD $1 - Z_1^{p_1} Z_2^{p_2}$ with $r = 1$ (bottom): from left to right $\theta = 20°, 50°, 80°$. Here, $w1$ and $w2$ are normalized frequencies (0.5 denotes Nyquist frequency, half the sampling frequency) in vertical and horizontal directions. White color denotes one and dark denotes zero. opwd/freq lidl20,lidl50,lidl80,oidl20,oidl50,oidl80

different dip angles. It may become aliased when the slope is large. The circle-interpolating PWD avoids aliasing for both small and large dip angles.

In line-interpolating PWD, we must design a digital filter to approximate the linear phase operator (or phase shift operator) Z_1^p. The slope has an infinite range $[-\infty, +\infty]$. In circle-interpolating PWD, there are two linear phase operators $Z_1^{p_1}$ and $Z_2^{p_2}$, related to the respective directions. Both the slopes p_1, p_2 have a finite range $[-r, r]$.

Following (6), the phase shift operators can be approximated by the following maxflat fractional delay filter (18):

$$H_1(Z_1) = \frac{B(1/Z_1)}{B(Z_1)} \approx Z_1^p, \tag{3}$$

where

$$B(Z_1) = \sum_{k_1=-N}^{N} b_{k_1} Z_1^{-k_1}, \tag{4}$$

N is the filter order and coefficients b_{k_1} are polynomial functions of local slopes p (3):

$$b_{k_1}(p) = \frac{(2N)!(2N)!}{(4N)!(N+k_1)!(N-k_1)!} \prod_{m=0}^{N-1-k_1} (m - 2N + p) \prod_{m=0}^{N-1+k_1} (m - 2N - p). \tag{5}$$

In Figure 3, we show the phase approximating performances of the maxflat fractional delay filters for different slopes. For small slope $p = 0.2$, the approximations are good, but when the slopes become large, the phases get wrapped. It is obvious that the phase wrapping comes when and only when $p > 1$. The larger the slope p, the more narrow the linear-phase frequency bands become.

As mentioned above, in line-interpolating PWC, the slope p is in the infinite interval $[-\infty, +\infty]$. For steep structures, where the slope p becomes larger than 1, there may be phase wrapping in the linear phase approximator. However, in circle-interpolating PWC, the ranges of p_1, p_2 can be easily controlled by the radius r. If we choose $r \leq 1$, the circle-interpolating can avoid phase wrapping completely for all dip angles.

2D LINEAR PHASE APPROXIMATION

Similar to the line-interpolating PWD, the 2D phase shift operator $Z_1^{p_1} Z_2^{p_2}$ can be approximated by the following 2D allpass system:

$$H_2(Z_1, Z_2) = \frac{F(1/Z_1, 1/Z_2)}{F(Z_1, Z_2)}, \tag{6}$$

where

$$F(Z_1, Z_2) = \sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} f_{k_1 k_2} Z_1^{-k_1} Z_2^{-k_2}. \tag{7}$$

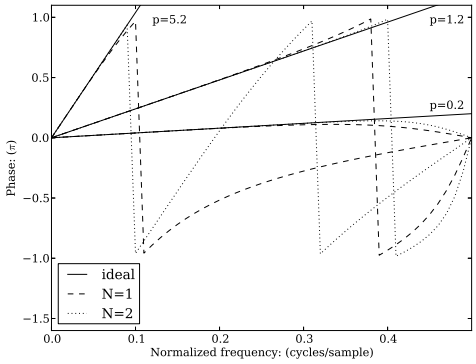


Figure 3: Phase approximating performances of the maxflat fractional delay filter $H_1(Z_1)$ when $p = 0.2, 1.2, 5.2$: dash lines denote first-order filter and dotted lines denote second-order filter. opwd/wrap wrap

2D maxflat condition

The frequency response of the objective phase shift operator is $Z_1^{p_1} Z_2^{p_2} = e^{\mathrm{j}(w_1 p_1 + w_2 p_2)}$, where, w_1, w_2 are frequencies in radius in vertical and horizontal directions. We must design the coefficients $f_{k_1 k_2}$ so that the allpass system $H_2(Z_1, Z_2)$ can obtain a similar linear phase response. The frequency response of $H_2(Z_1, Z_2)$ is

$$H_2(e^{\mathrm{j}w_1}, e^{\mathrm{j}w_2}) = \frac{F^*(e^{\mathrm{j}w_1}, e^{\mathrm{j}w_2})}{F(e^{\mathrm{j}w_1}, e^{\mathrm{j}w_2})} = e^{-\mathrm{j}2\theta_F(w_1, w_2)}, \tag{8}$$

where $\theta_F(w_1, w_2)$ is the phase of $F(Z_1, Z_2)$, which takes the following form:

$$\theta_F(w_1, w_2) = -\tan^{-1} \frac{\sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} f_{k_1 k_2} \sin(k_1 w_1 + k_2 w_2)}{\sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} f_{k_1 k_2} \cos(k_1 w_1 + k_2 w_2)}. \tag{9}$$

The phase approximating error is $w_1 p_1 + w_2 p_2 + 2\theta_F(w_1, w_2)$. In order to obtain an analytical $f_{k_1 k_2}$, we remove $\tan^{-1}$ and redefine the phase approximating error as

$$\begin{aligned} \epsilon(w_1, w_2) &= \tan(\frac{w_1 p_1 + w_2 p_2}{2}) - \frac{\sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} f_{k_1 k_2} \sin(k_1 w_1 + k_2 w_2)}{\sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} f_{k_1 k_2} \cos(k_1 w_1 + k_2 w_2)} \\ &= \frac{\sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} f_{k_1 k_2} \sin[w_1(\frac{p_1}{2} - k_1) + w_2(\frac{p_2}{2} - k_2)]}{\cos(\frac{w_1 p_1 + w_2 p_2}{2}) \sum_{k_1=-N}^{N} \sum_{k_2=-N}^{N} f_{k_1 k_2} \cos(k_1 w_1 + k_2 w_2)}. \end{aligned} \tag{10}$$

The sine function in the numerator can be expressed by 2D Taylor's expansion as

$$\begin{aligned} & \sin(w_1(\frac{p_1}{2} - k_1) + w_2(\frac{p_2}{2} - k_2)) \\ = & \sum_{j_1=0}^{\infty} \sum_{j_2=0}^{\infty} (-1)^{j_1+j_2} \frac{(\frac{p_1}{2} - k_1)^{2j_1+1} (\frac{p_2}{2} - k_2)^{2j_2+1}}{(2j_1+1)!(2j_2+1)!} w_1^{2j_1+1} w_2^{2j_2+1}. \end{aligned}$$

We use the maxflat phase criterion (18), which means that the filter has a phase response as flat as the desired linear response. In the 2D case, the criterion is equivalent to the mathematical expression that the partial derivatives of the error function should be as small as possible. We set them to be zero:

$$\frac{\partial^{j_1+j_2} \epsilon(w_1, w_2)}{\partial w_1^{j_1} \partial w_2^{j_2}} = 0 \qquad j_1, j_2 = 0, 1, \dots , \tag{11}$$

which is equivalent to the following 2D maxflat condition:

$$\sum_{k_1=-N}^{N}\sum_{k_2=-N}^{N}(\frac{p_1}{2}-k_1)^{2j_1+1}(\frac{p_2}{2}-k_2)^{2j_2+1}f_{k_1k_2}=0. \tag{12}$$

Additional constraint

The above maxflat condition is a linear system with $(2N+1)^2$ unknown variables $f_{k_1k_2}$. In order to obtain a non-zero solution, we use only the first $4N^2+4N$ equations, which zeros the partial derivatives with order less than $(2N+1)^2$, and impose an additional constraint on the filter coefficients,

$$\sum_{k_1=-N}^{N}\sum_{k_2=-N}^{N}f_{k_1k_2}=1. \tag{13}$$

This additional constraint yields a unit amplitude response at zero frequency, similarly to the 1D maxflat approximation (3).

Directional separability

The combined linear system can be solved analytically. It is based on the property that $f_{k_1k_2}$ can be decoupled into the product of two terms, as shown in the appendix,

$$f_{k_1k_2}(p_1,p_2)=b_{k_1}(p_1)b_{k_2}(p_2), \tag{14}$$

where $b_{k_1}(p_1)$ and $b_{k_2}(p_2)$ denote coefficients of the 1D maxflat fractional delay filter used to approximate $Z_1^{p_1}$ and $Z_2^{p_2}$ respectively. Equation 14 implies that the 2D maxflat filter $H_2(Z_1,Z_2)$ is equivalent to a cascade of two 1D maxflat filters,

$$H_2(Z_1,Z_2)=\frac{B_1(1/Z_1)}{B_1(Z_1)}\frac{B_2(1/Z_2)}{B_2(Z_2)}. \tag{15}$$

$B_1(Z_1)$ and $B_2(Z_2)$ can be designed in the same way as for the line-interpolating PWD, whose coefficients are given by equation 5. $H_2(Z_1,Z_2)$ can be implemented by applying the 1D maxflat filter on each direction independently.

This separability of the maxflat linear phase filter extends to 3D and higher dimensions.

Frequency responses

We show the magnitude responses of the infinite impulse response (IIR) omnidirectional PWD (OPWD) $1-H_2(Z_1,Z_2)$ when $N=2$ in Figure 4d-f. Compared with the

ideal responses in Figure 2, it has good approximations for all the three dip angles in most of the frequency band. There are distortions in the high frequency bands, due to phase approximation errors in $H_2(Z_1, Z_2)$ for high frequencies. They are not significant in practice, because the frequency band of seismic data is temporally limited.

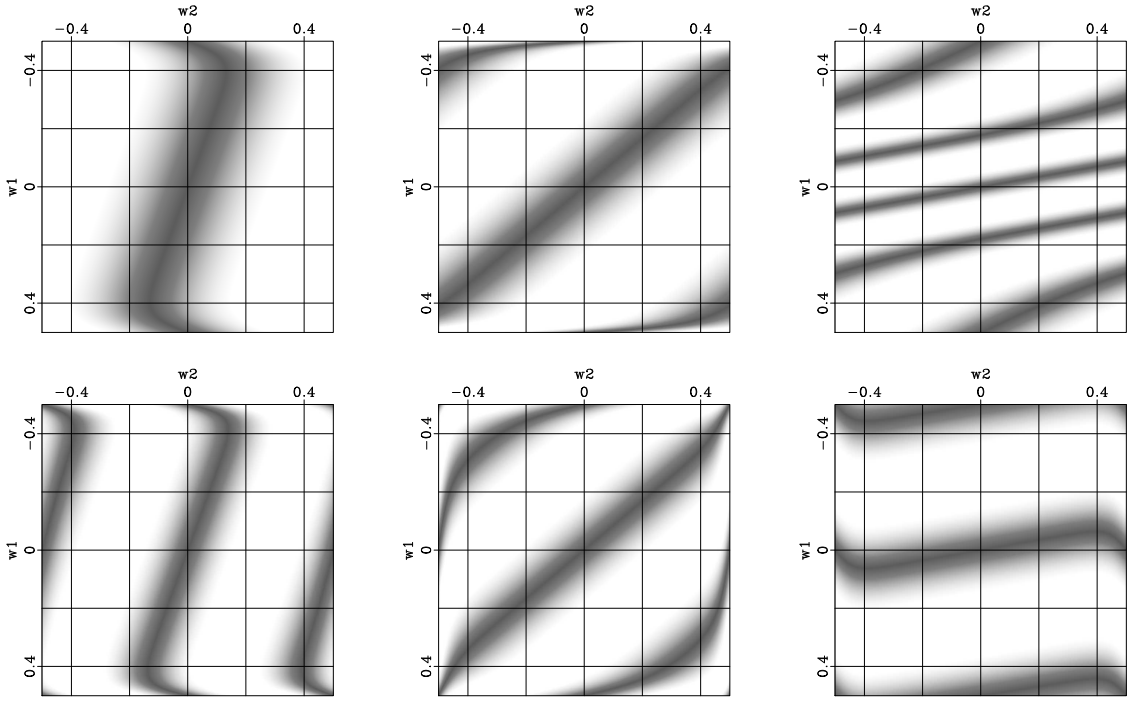


Figure 4: Magnitude responses of the IIR implementation of LPWD $1 - Z_2H_1(Z_1)$ (top) and OPWD $1 - H_2(Z_1, Z_2)$ (bottom): from left to right $\theta = 20°, 50°, 80°$. opwd/freq liir20,liir50,liir80,oiir20,oiir50,oiir80

When $\theta = 20°$ or $50°$, the omnidirectional PWD has a low-frequency response similar to that of line-interpolating PWD (LPWD) shown in Figure 4a-b. When $\theta = 80°$, the LPWD has aliasing in low-frequency bands (Figure 4c), while the OPWD exhibits a more desirable low-frequency response (Figure 4f).

Following (6), the IIR LPWD $1 - Z_2H_1(Z_1)$ can be approximated by the following finite impulse response (FIR) filter:

$$H_3(Z_1, Z_2) = B(Z_1) - Z_2B(1/Z_1). \tag{16}$$

Similarly, the IIR OPWD filter $1 - H_2(Z_1, Z_2)$ can also be approximated by an FIR implementation:

$$H_4(Z_1, Z_2) = F(Z_1, Z_2) - F(1/Z_1, 1/Z_2). \tag{17}$$

The frequency responses of these two FIR approximations are shown in Figure 5. Similar to the IIR implementation, the FIR OPWD (Figure 5d-f) can obtain

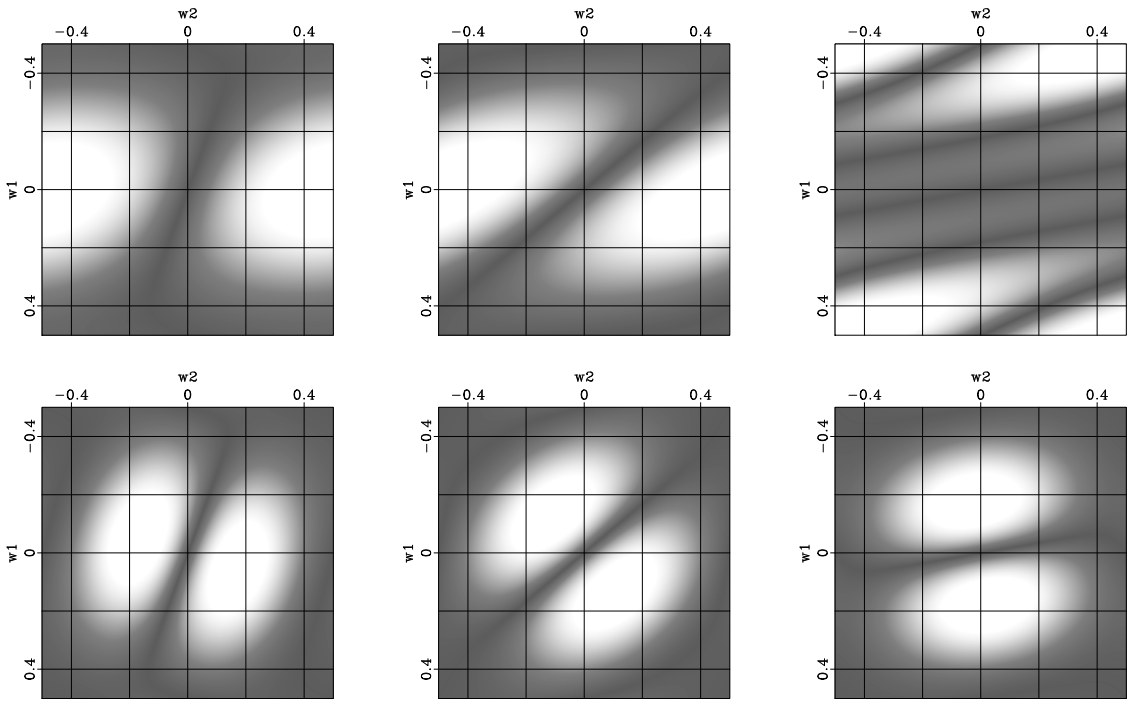


Figure 5: Magnitude response of the FIR approximation of LPWD $H_3(Z_1, Z_2)$ (top) and OPWD $H_4(Z_1, Z_2)$ (bottom): from left to right $\theta = 20°, 50°, 80°$. opwd/freq lfir20,lfir50,lfir80,ofir20,ofir50,ofir80

an expected low-frequency response for all three dip angles. While the FIR LPWD cannot obtain a good low-frequency response when $\theta = 80°$ (Figure 5c). Compared with the IIR implementations, the FIR approximations of both LPWD and OPWD have less desirable high-frequency responses.

APPLICATIONS

Omnidirectional dip estimation

Dip estimation by OPWD is equivalent to the parameter estimation of the OPWD filter in equation 17. The desired parameters p_1, p_2 minimize the predictive error:

$$H_4(Z_1, Z_2, p_1, p_2)U \approx 0, \tag{18}$$

with

$$p_1^2 + p_2^2 = r^2. \tag{19}$$

We solve the above equation set by Newton's search. At each iteration, the increments $\Delta p_1, \Delta p_2$ are calculated from the following linearization:

$$\begin{bmatrix} \frac{\partial H_4}{\partial p_1}U & \frac{\partial H_4}{\partial p_2}U \\ 2p_1 & 2p_2 \end{bmatrix} \begin{bmatrix} \Delta p_1 \\ \Delta p_2 \end{bmatrix} \approx \begin{bmatrix} -H_4U \\ r^2 - p_1^2 - p_2^2 \end{bmatrix}. \tag{20}$$

Both p_1 and p_2 are updated until the residual of OPWD becomes less than a specified tolerance. Similarly to the line-interpolating PWD, when we are solving for Δp_1 and Δp_2, either Tikhonov's regularization (6) or shaping regularization (7) can be applied to obtain a robust estimation.

In order to test the performance of OPWD, we generate a 2D synthetic image in polar coordinates using the following equation:

$$u(r) = e^{-\alpha r} \sin(2\pi(f_0 + \beta r)r). \tag{21}$$

Structures in this image have constant dip angles along radial directions, but the frequency and magnitude vary with radius. Figure 6a shows a test example with $f_0 = 0.02$, $\alpha = 0.01, \beta = 0.0005$. We apply the first-order ($N = 1$) OPWD filter to this test image. With the starting dip $\theta_0 = \frac{\pi}{4}$, we obtain an estimation result in 10 iterations, as shown in Figure 6b. Obviously, dip estimation by OPWD is effective for both horizontal and vertical structures.

We compare the proposed OPWD with LPWD by estimating errors in Figures 6c and 6d. As the test image is noise-free, we do not need to use any smoothing regularization in both of these methods. In this case, the modeling errors are the dominate factor of the final estimating errors. Obviously, the LPWD cannot detect the vertical structure accurately, while the OPWD can.

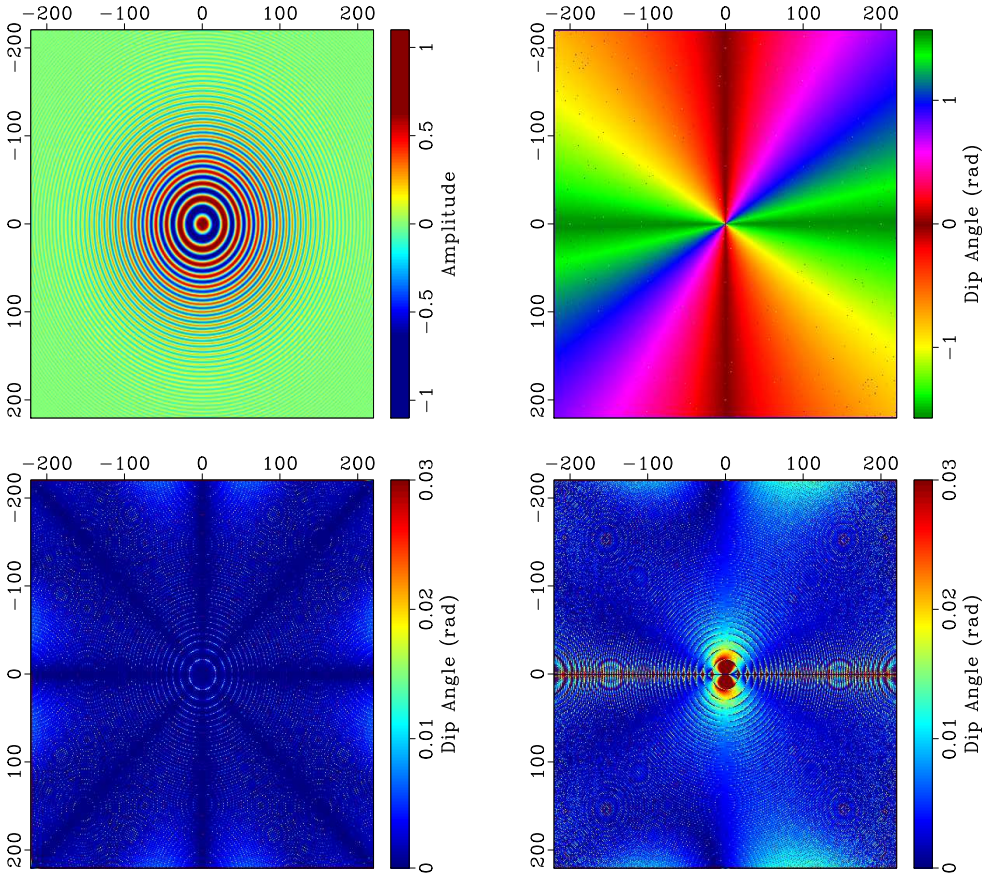


Figure 6: An omnidirectional dip estimation example: (a) test image; (b) dip angles estimated by the OPWD filter; (c) estimation error of OPWD; (d) estimation error of LPWD. opwd/circle circle,odip,oerr,lerr

Improved predictive picking

Event picking is a useful tool for interpretation. However, when there are salt bodies and buried hills in our images, picking becomes difficult. Using some additional information, such as cross-correlation analysis (19), high-order statistics (17) or local slope fields (8), automatic picking tools can be improved. A challenge for automatic picking is therefore the robust estimation of the statistics or slope fields. This challenge is especially difficult near boundaries of salt bodies and buried hills because of steep structures. In this paper, we use the slope-based predictive painting method (8) to pick events. We use the proposed OPWD to estimate the slope field, and show how it improves event-picking results.

In Figure 9(a), we show a profile of a post-stack dataset from an oilfield in China. We can see a buried hill in this area, and there are several steep boundaries and faults near the buried hill. The slope field estimated by OPWD is shown in Figure 8a. Compared with the slope by LPWD in Figure 8b, the OPWD yields a larger estimated slope near the 360-th trace and the 475-th trace. As shown in Figure 9(a), there exist faults and steep buried-hill boundaries at these two positions, respectively. In this application, we use the shaping regularization with 25-points window size. We choose the 400-th and 560-th traces as two reference traces and use the two slope fields in 2D predictive painting. Painting results are shown in Figure 9a and Figure 9b. Meanwhile, we pick some of the events, as shown in Figure 10. Compared with the painting and picking results based on the LPWD, the OPWD provides more reasonable results near the fault (360-th trace) and near the buried-hill boundaries (475-th trace).

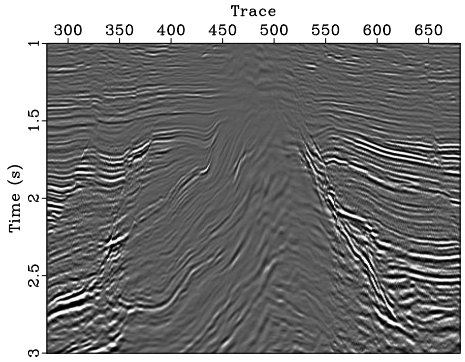


Figure 7: Field data example: steep structures and faults are rich in this selected area. opwd/pick data

CONCLUSIONS

We have addressed the problem of steep slope estimation by introducing a novel circle-interpolating method in plane-wave construction, which leads to omnidirectional plane-wave destruction. Frequency analysis demonstrates the ability of the

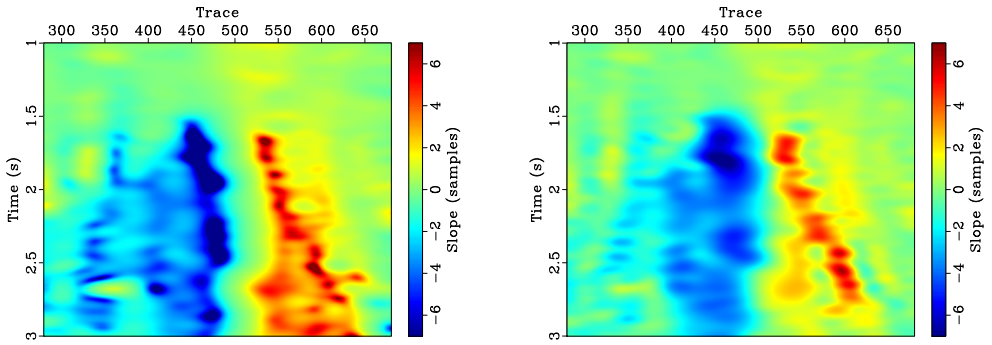


Figure 8: Slope field estimated by OPWD (a) and LPWD (b). In both the two methods, we choose $N = 2$ and use 25 points regularization window in each directions. opwd/pick odip,dip

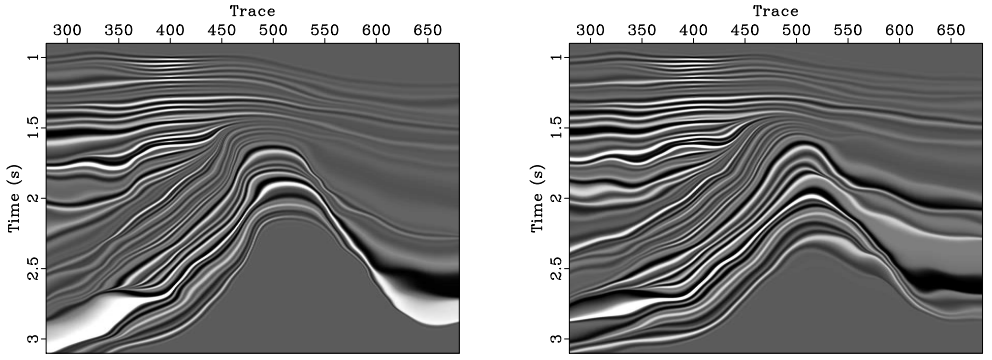


Figure 9: Predictive painting results using the OPWD slope (a) and LPWD slope (b): the reference traces are the 400-th and 560-th traces. opwd/pick pred-odip,pred-dip

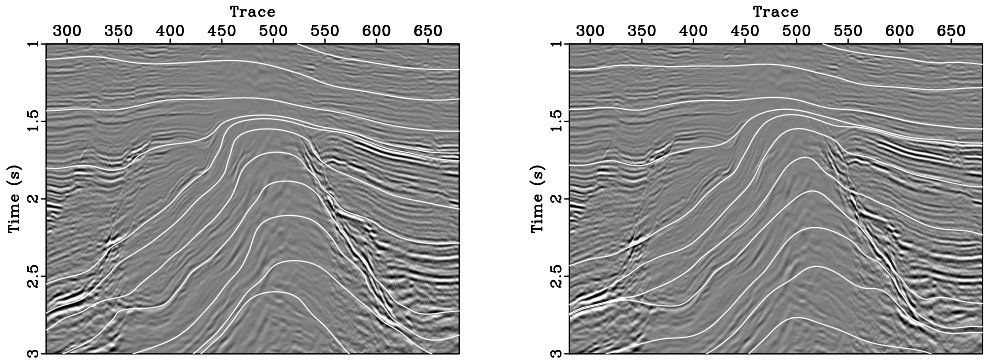


Figure 10: Predictive picking results using the OPWD slope (a) and LPWD slope (b). opwd/pick pick-odip,pick-dip

proposed method to handle steep structures. We presented both synthetic and field data examples to show the performance of the omnidirectional dip estimation. The new estimation method can also be used in other applications that require slope fields for steep structures, such as attribute analysis on a migrated image or a time slice.

As a byproduct, we proved that the high-dimensional maxflat fractional delay filter is separable in all directions. A typical $m-$dimensional $n-$th order ($n = 2N + 1$ in this paper) linear-phase approximating system is supported by n^m coefficients. Using the separability property, the maxflat linear-phase approximating system has only mn coefficients. It is useful in applications that require high-dimensional linear-phase approximations.

ACKNOWLEDGMENT

The joint research is sponsored by by China Scholarship Council and China State Key Science and Technology Projects (2011ZX05023-005-007). We thank the reviewers and the Associate Editor for their helpful suggestions.

APPENDIX: SEPARABILITY OF THE 2D MAXFLAT FILTER

Following (18), the maxflat condition of 1D fractional delay filter $B_1(1/Z_1)/B_1(Z_1)$ is expressed as

$$\sum_{k_1=-N}^{N} (\frac{p_1}{2} - k_1)^{2j_1+1} b_{k_1}(p_1) = 0 \ , \qquad j_1 = 0, \ldots, 2N-1. \tag{A-1}$$

With the additional constraint (3)

$$\sum_{k_1=-N}^{N} b_{k_1}(p_1) = 1, \tag{A-2}$$

we can obtain the unique $b_{k_1}(p_1)$. Similarly, $b_{k_2}(p_2)$ satisfies the following linear system:

$$\begin{cases} \displaystyle\sum_{k_2=-N}^{N} (\frac{p_2}{2} - k_2)^{2j_2+1} b_{k_2}(p_2) = 0 \ , \quad j_2 = 0, \ldots, 2N-1 \\ \displaystyle\sum_{k_2=-N}^{N} b_{k_2}(p_2) = 1 \end{cases} . \tag{A-3}$$

Substituting $b_{k_1}(p_1) b_{k_2}(p_2)$ into the 2D maxflat equation 12, for all $j_1, j_2 = 0, \ldots, 2N-1$, we obtain

$$\sum_{k_1=-N}^{N} (\frac{p_1}{2} - k_1)^{2j_1+1} b_{k_1}(p_1) \sum_{k_2=-N}^{N} (\frac{p_2}{2} - k_2)^{2j_2+1} b_{k_2}(p_2) = 0. \tag{A-4}$$

Also, for $j_1 = 2N$, and $j_2 = 0, \ldots, 2N-1$ or $j_2 = 2N$, and $j_1 = 0, \ldots, 2N-1$, equation 12 still holds true.

Substituting $b_{k_1}(p_1)b_{k_2}(p_2)$ into the additional constraint in Equation 13,

$$\sum_{k_1=-N}^{N} b_{k_1}(p_1) \sum_{k_2=-N}^{N} b_{k_2}(p_2) = 1. \tag{A-5}$$

In other words, $b_{k_1}(p_1)b_{k_2}(p_2)$ is a solution of the combined linear system of equation 12 and 13. This linear system must have a unique solution, therefore

$$f_{k_1k_2}(p_1, p_2) = b_{k_1}(p_1)b_{k_2}(p_2). \tag{A-6}$$

REFERENCES

[1] Bardan, V., 1987, Trace interpolation in seismic data processing: Geophysical Prospecting, **35**, 343–358.
[2] Barnes, A. E., 1996, Theory of 2-D complex seismic trace analysis: Geophysics, **61**, 264–272.
[3] Chen, Z., S. Fomel, and W. Lu, 2013, Accelerated plane-wave destruction: Geophysics, **78**, V1–V9.
[4] Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications.
[5] Davis, S., 1991, Low-dispersion finite difference methods for acoustic waves in a pipe: The Journal of the Acoustical Society of America, **90**, 2775–2781.
[6] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.
[7] ———, 2007, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.
[8] ———, 2010, Predictive painting of 3D seismic volumes: Geophysics, **75**, A25–A30.
[9] Fomel, S., and A. Guitton, 2006, Regularizing seismic inverse problems by model reparameterization using plane-wave construction: Geophysics, **71**, A43–A47.
[10] Fomel, S., E. Landa, and M. T. Taner, 2007, Poststack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.
[11] Hale, D., 2007, Local dip filtering with directional Laplacians: CWP Report, **567**.
[12] Harlan, W. S., J. F. Claerbout, and F. Rocca, 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, 1869–1880.
[13] Marfurt, K. J., 2006, Robust estimates of 3D reflector dip and azimuth: Geophysics, **71**, P29–P40.
[14] Marfurt, K. J., V. Sudhaker, A. Gersztenkorn, K. D. Crawford, and S. E. Nissen, 1999, Coherency calculations in the presence of structural dip: Geophysics, **64**, 104–111.
[15] Noye, B., 2000, Explicit finite difference methods for variable velocity advection in the presence of a source: Computers & Fluids, **29**, 385 – 399.

[16] Schleicher, J., J. C. Costa, L. T. Santos, A. Novais, and M. Tygel, 2009, On the estimation of local slopes: Geophysics, **74**, P25–P33.
[17] Srinivasan, K., and L. T. Ikelle, 2001, Short note: Applications of third-order statistics for the automatic time picking of seismic events: Geophysical Prospecting, **49**, 155–158.
[18] Thiran, J., 1971, Recursive digital filters with maximally flat group delay: IEEE Transactions on Circuit Theory, **18**, 659–664.
[19] Yung, S. K., and L. T. Ikelle, 1997, An example of seismic time picking by third-order bicoherence: Geophysics, **62**, 1947–1952.
[20] Zhang, X., 2009, Maxflat fractional delay IIR filter design: IEEE Transactions on Signal Processing, **57**, 2950 –2956.

Time-to-depth conversion and seismic velocity estimation using time-migration velocity

Maria Cameron[*], *Sergey Fomel*[†], *and James Sethian*[‡*]

ABSTRACT

The objective of this work is to build an efficient algorithm (a) to estimate seismic velocity from time-migration velocity, and (b) to convert time-migrated images to depth. We establish theoretical relations between the time-migration velocity and the seismic velocity in 2-D and 3-D using paraxial ray tracing theory. The relation in 2-D implies that the conventional Dix velocity is the ratio of the interval seismic velocity and the geometrical spreading of the image rays. We formulate an inverse problem of finding seismic velocity from the Dix velocity and develop a numerical procedure for solving it. This procedure consists of two steps: (1) computation of the geometrical spreading of the image rays and the true seismic velocity in the time-domain coordinates from the Dix velocity; (2) conversion of the true seismic velocity from the time domain to the depth domain and computation of the transition matrices from time-domain coordinates to depth.

For step 1, we derive a partial differential equation (PDE) in 2-D and 3-D relating the Dix velocity and the geometrical spreading of the image rays to be found. This is a nonlinear elliptic PDE. The physical setting allows us to pose a Cauchy problem for it. This problem is ill-posed. However we are able to solve it numerically in two ways on the required interval of time. One way is a finite difference scheme inspired by the Lax-Friedrichs method. The second way is a spectral Chebyshev method. For step 2, we develop an efficient Dijkstra-like solver motivated by Sethian's Fast Marching Method.

We test our numerical procedures on a synthetic data example and apply them to a field data example. We demonstrate that our algorithms give significantly more accurate estimate of the seismic velocity than the conventional Dix inversion. Our velocity estimate can be used as a reasonable first guess in building velocity models for depth imaging.

[*]**e-mail:** sergey.fomel@beg.utexas.edu

INTRODUCTION

Time-domain seismic imaging is a robust and efficient process routinely applied to seismic data (19; 15). Rapid scanning and determination of time-migration velocity can be accomplished either by repeated migrations (20) or by velocity continuation (6). Time migration is considered adequate for seismic imaging in areas with mild lateral velocity variations. However, even mild variations can cause structural distortions of time-migrated images and render them inadequate for accurate geological interpretation of subsurface structures.

To remove structural errors inherent in time migration, it is necessary to convert time-migrated images into the depth domain either by migrating the original data with a prestack depth migration algorithm or by depth migrating post-stack data after time demigration (10). Each of these options requires converting the time migration velocity to a velocity model in depth.

The connection between the time- and depth-domain coordinates is provided by the concept of *image ray*, introduced by (8). Image rays are seismic rays that arrive normal to the Earth's surface. Hubral's theory explains how a depth velocity model can be converted to the time coordinates. However, it does not explain how a depth velocity model can be converted to the time-migration velocity. Moreover, image-ray tracing is a numerically inconvenient procedure for achieving uniform coverage of the subsurface. This may explain why simplified image-ray-tracing algorithms (11; 7) did not find widespread application in practice. Other limitations of image rays are related to the inability of time migration to handle large lateral variations in velocity (1; 15).

The objective of the present work is to find an efficient method for building a velocity model from time-migration velocity. We establish new ray-theoretic connections between time-migration velocity and seismic velocity in 2-D and 3-D. These results are based on the image ray theory and the paraxial ray tracing theory (14; 4; 13). Our results can be viewed as an extension of the Dix formula (5) to laterally inhomogeneous media. We show that the Dix velocity is seismic velocity divided by the geometrical spreading of the image rays. Hence, we use the Dix velocity instead of time migration velocity as a more convenient input. We develop a numerical approach to find (a) seismic velocity from the Dix velocity, and (b) transition matrices from the time-domain coordinates to the depth-domain coordinates. We test our approach on synthetic and field data examples.

Our approach is complementary to more traditional velocity estimation methods and can be used as the first step in a velocity model building process.

TIME MIGRATION VELOCITY

Kirchhoff prestack time migration is commonly based on the following travel time approximation (19). Let $\mathbf{s}$ be a source, $\mathbf{r}$ be a receiver, and $\mathbf{x}$ be the reflection sub-

surface point. Then the total travel time from $\mathbf{s}$ to $\mathbf{x}$ and from $\mathbf{x}$ to $\mathbf{r}$ is approximated as

$$T(\mathbf{s},\mathbf{x}) + T(\mathbf{x},\mathbf{r}) \approx \hat{T}(\mathbf{x}_0, t_0, \mathbf{s}, \mathbf{r}) \tag{1}$$

where $\mathbf{x}_0$ and t_0 are effective parameters of the subsurface point $\mathbf{x}$. The approximation $\hat{T}$ usually takes the form of the double-square-root equation

$$\hat{T}(\mathbf{x}_0, t_0, \mathbf{s}, \mathbf{r}) = \sqrt{t_0^2 + \frac{|\mathbf{x}_0 - \mathbf{s}|^2}{v_m^2(\mathbf{x}_0, t_0)}} + \sqrt{t_0^2 + \frac{|\mathbf{x}_0 - \mathbf{r}|^2}{v_m^2(\mathbf{x}_0, t_0)}}, \tag{2}$$

where $\mathbf{x}_0$ and t_0 are the escape location and the travel time of the *image ray* (8) from the subsurface point $\mathbf{x}$. Regarding this approximation, let us list four cases depending on the seismic velocity v and the dimension of the problem:

2-D and 3-D, velocity v is constant. Equation 2 is exact, and $v_m = v$.

2-D and 3-D, velocity v depends only on the depth z. Equation 2 is a consequence of the truncated Taylor expansion for the travel time around the surface point $\mathbf{x}_0$. Velocity v_m depends only on t_0 and is the root-mean-square velocity:

$$v_m(t_0) = \sqrt{\frac{1}{t_0}\int_0^{t_0} v^2(z(t))dt}. \tag{3}$$

In this case, the Dix inversion formula (5) is exact. We formally define the Dix velocity $v_{Dix}(t)$ by inverting equation 3, as follows:

$$v_{Dix}(t) = \sqrt{\frac{d}{d\,t_0}\left(t_0 v_m^2(t_0)\right)}\,. \tag{4}$$

2-D, velocity is arbitrary. Equation 2 is a consequence of the truncated Taylor expansion for the travel time around the surface point x_0. Velocity $v_m(x_0, t_0)$ is a certain kind of mean velocity, and we establish its exact meaning in the next section.

3-D, velocity is arbitrary. Equation 2 is heuristic and is not a consequence of the truncated Taylor expansion. In order to write an analog of travel time approximation 2 for 3-D, we use the relation (9)

$$\mathbf{\Gamma} = [v(\mathbf{x}_0)\mathbf{R}(\mathbf{x}_0, t_0)]^{-1}, \tag{5}$$

where $\mathbf{\Gamma}$ is the matrix of the second derivatives of the travel times from a subsurface point $\mathbf{x}$ to the surface, $\mathbf{R}$ is the matrix of radii of curvature of the emerging wave front from the point source $\mathbf{x}$, and $v(\mathbf{x}_0)$ is the velocity at the surface point $\mathbf{x}_0$. For convenience, we prefer to deal with matrix $\mathbf{K} \equiv \mathbf{\Gamma}^{-1}$, which, according to equation 5 is

$$\mathbf{K}(\mathbf{x}_0, t_0) \equiv v(\mathbf{x}_0)\mathbf{R}(\mathbf{x}_0, t_0). \tag{6}$$

The travel time approximation for 3-D implied by the Taylor expansion is

$$\begin{aligned} & \hat{T}(\mathbf{x}_0, t_0, \mathbf{s}, \mathbf{r}) \\ = \quad & \sqrt{t_0^2 + t_0(\mathbf{x}_0 - \mathbf{s})^T [\mathbf{K}(\mathbf{x}_0, t_0)]^{-1} (\mathbf{x}_0 - \mathbf{s})} \\ + \quad & \sqrt{t_0^2 + t_0(\mathbf{x}_0 - \mathbf{r})^T [\mathbf{K}(\mathbf{x}_0, t_0)]^{-1} (\mathbf{x}_0 - \mathbf{r})}. \end{aligned} \tag{7}$$

The entries of the matrix $\frac{1}{t_0}\mathbf{K}(\mathbf{x}_0, t_0)$ have dimension of squared velocity and can be chosen optimally in the process of time migration. It is possible to show, however, that one needs only the values of

$$\det\left(\frac{\partial}{\partial t_0}\mathbf{K}(\mathbf{x}_0, t_0)\right) \tag{8}$$

to perform the inversion. This means that the conventional 3-D prestack time migration with traveltime approximation 2 provides sufficient input for our inversion procedure in 3-D. The determinant in equation 8 is well approximated by the square of the Dix velocity obtained from the 3-D prestack time migration using the approximation given by equation 2.

One can employ more complex and accurate approximations than the double-square-root equations 2 and 7, i.e. the shifted hyperbola approximation (18). However, other known approximations also involve parameters equivalent to v_m or $\mathbf{K}$.

SEISMIC VELOCITY

In this section, we will establish theoretical relationships between time-migration velocity and seismic velocity in 2-D and 3-D.

The seismic velocity and the Dix velocity are connected through the quantity $\mathbf{Q}$, the geometrical spreading of image rays. $\mathbf{Q}$ is a scalar in 2-D and a 2×2 matrix in 3-D. The simplest way to introduce $\mathbf{Q}$ is the following. Trace an image ray $\mathbf{x}(\mathbf{x}_0, t)$. $\mathbf{x}_0$ is the starting surface point, t is the traveltime. Call this ray *central*. Consider a small tube of rays around it. All these rays start from a small neighborhood $d\mathbf{x}_0$ of the point $\mathbf{x}_0$ perpendicular to the earth surface. Thus, they represent a fragment of a plane wave propagating downward. Consider the fragment of the wave front defined by this ray tube at time t_0. Let $d\mathbf{q}$ be the fragment of the tangent to the front at the point $\mathbf{x}(\mathbf{x}_0, t_0)$ reached by the central ray at time t_0, bounded by the ray tube (Figure 1). Then, in 2-D, Q is the derivative $Q(x_0, t_0) = \frac{dq}{dx_0}$. In 3-D, $\mathbf{Q}$ is the matrix of the derivatives $\mathbf{Q}_{ij}(\mathbf{x}_0, t_0) = \frac{d\mathbf{q}_i}{d\mathbf{x}_{0j}}$, $i, j = 1, 2$, where derivatives are taken along certain mutually orthogonal directions $\mathbf{e}_1$, $\mathbf{e}_2$ (14; 4; 13).

The time evolution of the matrices $\mathbf{Q}$ and $\mathbf{P}$ is given by

$$\frac{d}{dt}\begin{pmatrix} \mathbf{Q} \\ \mathbf{P} \end{pmatrix} = \begin{pmatrix} \mathbf{0} & v_0^2\mathbf{I} \\ -\frac{1}{v_0}\mathbf{V} & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{Q} \\ \mathbf{P} \end{pmatrix}, \tag{9}$$

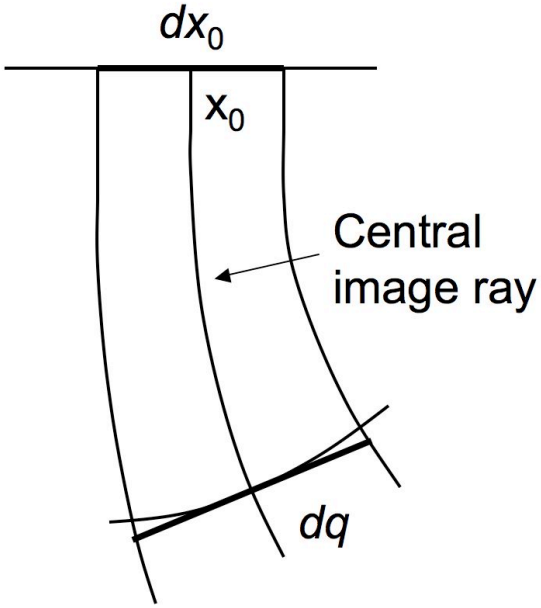


Figure 1: Illustration for the definition of geometrical spreading. time2depth/. Qdef

where v_0 it the velocity at the central ray at time t, $\mathbf{V} = \left(\frac{\partial^2 v}{\partial q_i \partial q_j}\right)_{i,j=1,2}$, and $\mathbf{I}$ is the 2×2 identity matrix. The absolute value of $\det \mathbf{Q}$ has a simple meaning: it is the geometrical spreading of the image rays (14; 4; 13). The matrix $\mathbf{\Gamma}$, introduced in the previous section, relates to $\mathbf{Q}$ and $\mathbf{P}$ as $\mathbf{\Gamma} = \mathbf{P}\mathbf{Q}^{-1}$. Hence, $\mathbf{K} = \mathbf{Q}\mathbf{P}^{-1}$.

In (3), we have proven that

$$v_{Dix}(x_0, t_0) \equiv \sqrt{\frac{\partial}{\partial t_0}\left(t_0 v_m^2(x_0, t_0)\right)} = \frac{v(x(x_0, t_0), z(x_0, t_0))}{|Q(x_0, t_0)|} \tag{10}$$

in 2-D, where $v_m(x_0, t_0)$ is the time-migration velocity, and

$$\frac{\partial}{\partial t_0}\left(\mathbf{K}(\mathbf{x}_0, t_0)\right) = v(\mathbf{x}(\mathbf{x}_0, t_0))\left(\mathbf{Q}(\mathbf{x}_0, t_0)\mathbf{Q}^T(\mathbf{x}_0, t_0)\right)^{-1} \tag{11}$$

in 3-D, $\mathbf{K}$ is defined by equation 6 and can be determined from equation 7.

PARTIAL DIFFERENTIAL EQUATIONS FOR THE GEOMETRICAL SPREADING OF IMAGE RAYS

In this section, we derive the partial differential equations for Q in 2-D and 3-D. From now on, we will denote the square of the Dix velocity by f in 2-D and the

corresponding matrix by $\mathbf{F}$ in 3-D, to avoid the subscript:

$$\mathbf{F} \equiv \frac{\partial}{\partial t_0} \left(\mathbf{K}(\mathbf{x}_0, t_0) \right). \tag{12}$$

Furthermore, we imply that t_0 denotes the one-way traveltime along the image rays. Finally, we assume that our domain does not contain caustics, i.e., the image rays do not cross on the interval of time we consider.

2-D case

Consider a set of image rays coming to the surface. Suppose we are tracing them all backwards in time together with the quantities Q and P. Let us eliminate the unknown velocity v in system 9 using equation 10. Moreover, let us eliminate the differentiation in q using the definition of Q and rewrite it in the time-domain coordinates x_0, t_0). Indeed, $Q = \frac{dq}{dx_0}$, hence $\frac{d}{dq} = \frac{d}{dx_0}\frac{dx_0}{dq} = Q^{-1}\frac{d}{dx_0}$. Therefore, system 9 becomes

$$Q_{t_0} = (fQ)^2 P, \qquad P_{t_0} = -\frac{1}{fQ}\left(\frac{(fQ)_{x_0}}{Q}\right)_{x_0}. \tag{13}$$

Eliminating P in system 13, we get the following partial differential equation (PDE) for Q

$$\left(\frac{Q_{t_0}}{f^2Q^2}\right)_{t_0} = -\frac{1}{fQ}\left(\frac{(fQ)_{x_0}}{Q}\right)_{x_0}. \tag{14}$$

The initial conditions are $Q(x_0, 0) = 1$, $Q_{t_0}(x_0, 0) = 0$. Equation 14 simplifies in terms of the negative reciprocal of Q. Introduce $y = -\frac{1}{Q}$. Then equation 14 becomes

$$\left(\frac{y_{t_0}}{f^2}\right)_{t_0} = \frac{y}{f}\left(\left(\frac{f}{y}\right)_{x_0} y\right)_{x_0}. \tag{15}$$

In the expanded form, equation 15 is

$$\frac{y_{t_0 t_0}}{f^2} - 2\frac{y_{t_0} f_{t_0}}{f^3} = y\frac{f_{x_0 x_0}}{f} - y_{x_0}\frac{f_{x_0}}{f} - y_{x_0 x_0} + \frac{y_{x_0}^2}{y}. \tag{16}$$

3-D case

Equation 11 can be rewritten in the following form

$$v = \sqrt[4]{\det \mathbf{F} (\det \mathbf{Q})^2}, \tag{17}$$

where $\mathbf{F}$ is the left-hand side of equation 11. As in 2-D, we rewrite system 9 in the time-domain coordinates $(\mathbf{x}_0, t_0)$. Then we get

$$\mathbf{Q}_{t_0} = v^2 \mathbf{P}, \tag{18}$$

$$\mathbf{P}_{t_0} = -\frac{1}{v}\mathbf{Q}^{-1}\left[\nabla\left(\mathbf{Q}^{-1}\nabla v\right)^T\right]\mathbf{Q}, \tag{19}$$

where v is given by equation 17, and the gradients are taken with respect to $\mathbf{x}_0$. Then the PDE for $\mathbf{Q}$ is

$$\left(\frac{1}{v^2}\mathbf{Q}_{t_0}\right)_{t_0} = -\frac{1}{v}\mathbf{Q}^{-1}\left[\nabla\left(\mathbf{Q}^{-1}\nabla v\right)^T\right]\mathbf{Q}. \tag{20}$$

The initial conditions are $\mathbf{Q}(\mathbf{x}_0, 0) = \mathbf{I}_2$, $\mathbf{Q}_{t_0}(\mathbf{x}_0, 0) = \mathbf{0}$. The required input $\sqrt{\det \mathbf{F}}$ is well-approximated by the squares of the Dix velocity obtained from the 3-D prestack time migration. We emphasize that despite the fact that $\mathbf{Q}$ is a matrix in 3-D, scalar data are enough for its computation.

Cauchy problem for elliptic equations

Equations 14 and 20 reveal the nature of the instabilities in the problem in hand. These PDEs are elliptic, while the physical setting allows us to pose only a Cauchy problem for them, which is well-known to be ill-posed. Furthermore, the fact that the PDEs involve not only the Dix velocity itself, but also its first and second derivatives, leads to high sensitivity of the solutions to input data.

Nonetheless, we found two ways for solving these PDEs numerically on the required, and relatively short, interval of time: namely, a finite difference scheme inspired by the Lax-Friedrichs method and a spectral Chebyshev method. A detailed analysis of the problem shows that our methods work thanks to

1. the special input v_{Dix}, corresponding to a positive finite seismic velocity;
2. the special initial conditions $Q(x_0, t_0 = 0) = 1$, $Q_t(x_0, t_0 = 0) = 0$ corresponding to the image rays;
3. the fact that our finite difference method contains error terms which damp the high harmonics; truncation of the polynomial series in the spectral Chebyshev method which is similar to truncation of the high harmonics; and
4. the short interval of time, in which we need to compute the solution, so that the growing low harmonics fail to develop significantly.

Items 1 and 2 say that the exact solutions of our PDEs for the hypothetical perfect Dix velocity given by equations 10 and 11 are finite and nonzero. Items 3 and 4 say that the numerical methods take care of the imperfection of the data and computations on a short enough time interval.

INVERSION METHODS

Our numerical reconstruction of true seismic velocity $v(\mathbf{x})$ in depth-domain coordinates from the Dix velocity given in the time-domain coordinates $(\mathbf{x}_0, t_0)$ consists of two steps:

Step 1. Compute the geometrical spreading of the image rays in the time-domain coordinates from the Dix velocity by solving equation 14 in 2-D and 20 in 3-D. Then find $v(\mathbf{x}_0, t_0)$ from equation 10 in 2-D and equation 17 in 3-D.

Step 2. Convert the seismic velocity $v(\mathbf{x}_0, t_0)$ in the time-domain coordinates to the depth-domain coordinates $\mathbf{x}$ using the time-to-depth conversion algorithm, which was presented by (3). It is a fast and robust Dijkstra-like solver motivated by the Fast Marching method (16; 17).

We performed step 1 in two ways: a finite difference method and a spectral Chebyshev method.

Finite difference method

This method was inspired by the Lax-Friedrichs method for hyperbolic conservation laws (12) due to its total variation diminishing property. We use the "Lax-Friedrichs averaging" and the wide 5-point stencil in space. The scheme is given by

$$P_j^{n+1} = \frac{P_{j+1}^n + P_{j-1}^n}{2} - \frac{\Delta t}{4\Delta x}\frac{1}{v_j^n}\left(\frac{v_{j+2}^n - v_j^n}{Q_{j+1}^n} - \frac{v_j^n - v_{j-2}^n}{Q_{j-1}^n}\right), \tag{21}$$

$$-\frac{1}{Q_j^{n+1}} = -\frac{1}{Q_j^n} + \frac{\Delta t}{2}\left((f_j^n)^2 P_j^n + (f_j^{n+1})^2 P_j^{n+1}\right), \tag{22}$$

where $v \equiv fQ$. We impose the following boundary conditions $Q_0^n = Q_{nx-1}^n = 1$, $P_0^n = P_{nx-1}^n = 0$ corresponding the straight boundary rays. We set the initial conditions $Q_j^0 = 1$, $P_j^0 = 0$ corresponding to the initial conditions for the image rays traced backward: $Q = 1$, $P = 0$.

Spectral Chebyshev method

Alternatively, we solve our PDE in the form given by equation 15 by a spectral Chebyshev method (2). Using cubic splines, we define the input data at N_{coef} Chebyshev points. We compute the Chebyshev coefficients and the coefficients of the derivatives in the right-hand side of equation 15. Then we use a smaller number N_{eval} of the coefficients for function evaluation. We need to do such Chebyshev differentiation twice. Finally we perform the time step using the stable third-order Adams-Bashforth method (2), which is

$$u^{n+1} = u^n + \Delta t\left(\frac{23}{12}F^n - \frac{4}{3}F^{n-1} + \frac{5}{12}F^{n-2}\right), \tag{23}$$

where $F^n \equiv F(u^n, x, t^n)$ is the right-hand side. In numerical examples, we tried $N_{coef} \geq 100$ and $N_{eval} \leq 25$. This method allows larger time steps than the finite difference, and has the adjustable parameter N_{eval}.

For step 2, we use a Dijkstra-like solver introduced in (3). It is an efficient time-to-depth conversion algorithm motivated by the Fast Marching Method (16). The input for this algorithm is $v(x_0, t_0)$ and the outputs are the seismic velocity $v(x, z)$ and the transition matrices from time-domain to depth-domain coordinates $x_0(x, z)$ and $t_0(x, z)$. We solve the eikonal equation with an unknown right-hand side coupled with the orthogonality relation

$$|\nabla t_0| = \frac{1}{v(x_0(x,z), t_0(x,z))}, \quad \nabla t_0 \cdot \nabla x_0 = 0. \tag{24}$$

The orthogonality relation means that the image rays are orthogonal to the wavefronts. Such time-to-depth conversion is very fast and produces the outputs directly on the depth-domain grid.

EXAMPLES

Synthetic data example

Figure 2(a) shows a synthetic velocity model. The model contains a high velocity anomaly that is asymmetric and decays exponentially. The corresponding Dix velocity mapped from time to depth is shown in Figure 2(b). There is a significant difference between both the value and the shape of the velocity anomaly recovered by the Dix method and the true anomaly. The difference is explained by taking into account geometrical spreading of image rays. Figure 2(c) shows the velocity recovered by our method and the corresponding family of image rays. An analogous 3-D example is provided in (3).

Field data example

Figure 3, taken from (6), shows a prestack time migrated image from the North Sea and the corresponding time-migration velocity obtained by velocity continuation. The most prominent feature in the image is a salt body which causes significant lateral variations of velocity. Figure 4 compares the Dix velocity converted to depth with the interval velocity model recovered by our method. As in the synthetic example, there is a significant difference between the two velocity caused by the geometrical spreading of image rays. The middle part of the velocity model may not be recovered properly. The true structure should include a salt body visible in the image. The inability of our method to recover it exactly shows the limitation of the proposed approach in the areas of significant lateral velocity variations, which invalidate the assumptions behind time migration (15). Figure 5 compares three images: post-stack depth-migration image using Dix velocity, post-stack depth-migration image using the velocity estimated by our method, and prestack time-migration image converted to depth with our algorithm. The evident structural improvements in Figure 5(b)

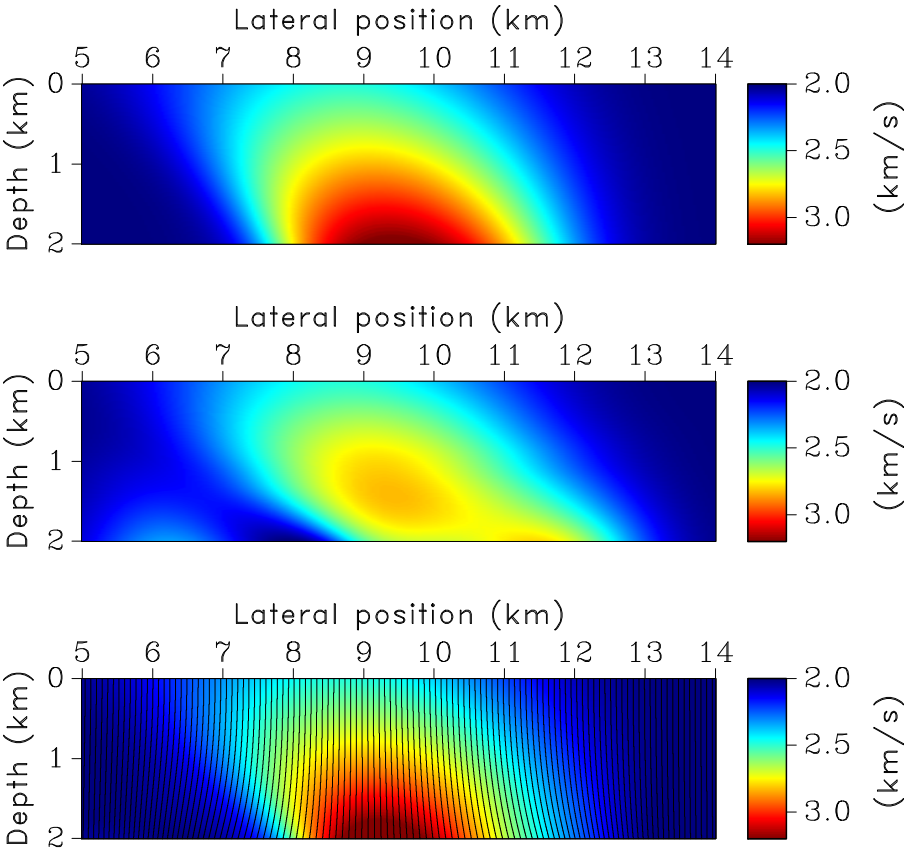


Figure 2: Synthetic test on interval velocity estimation. (a) Exact velocity model. (b) Dix velocity converted to depth. (c) Estimated velocity model and the corresponding image rays. The image ray spreading causes significant differences between Dix velocity and estimated velocity. time2depth/msyn2 ve,vd2,vf2

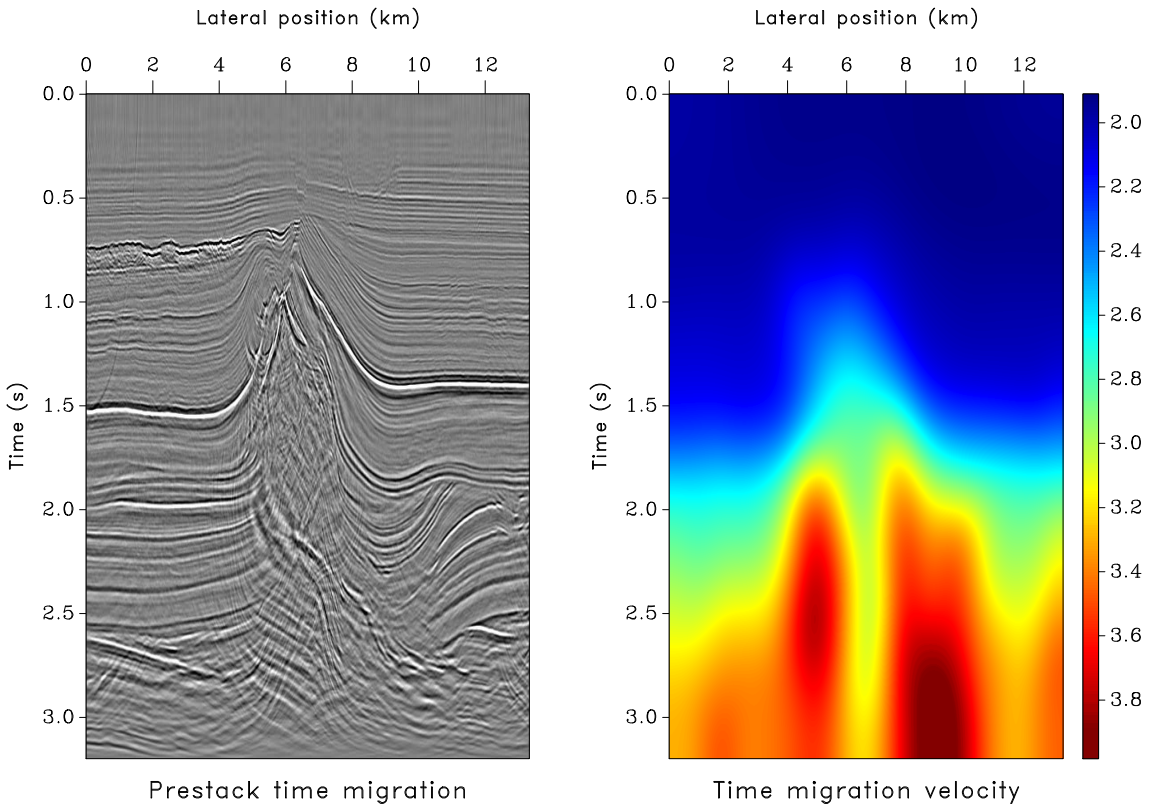


Figure 3: (a) Seismic image from North Sea obtained by prestack time migration using velocity continuation (Fomel, 2003). (b) Corresponding time-migration velocity. time2depth/masha2 elf-fmg2

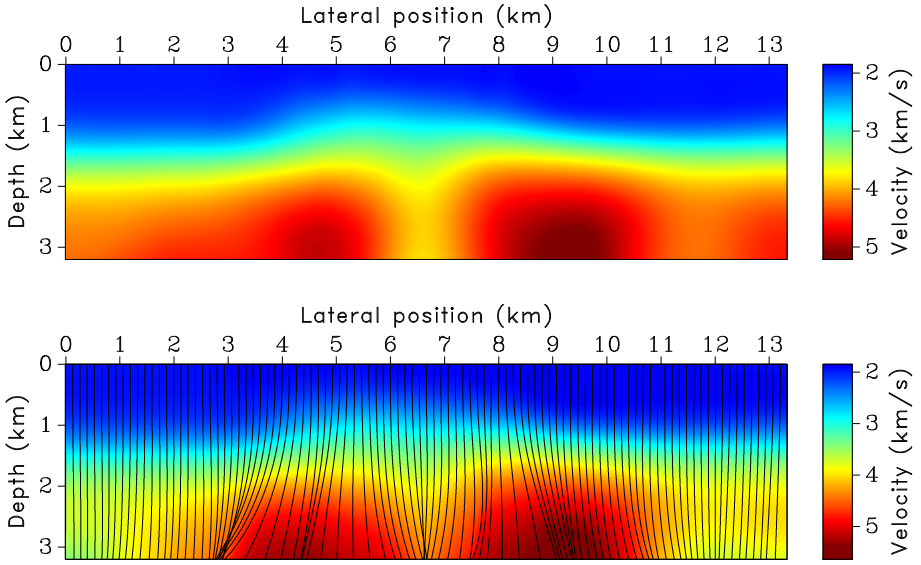


Figure 4: Field data example of interval velocity estimation. (a) Dix velocity converted to depth. (b) Estimated velocity model and the corresponding image rays. The image-ray spreading causes significant differences between Dix velocity and true velocity. time2depth/masha2 vels,elfvxz

in comparison with Figure 5(a), in particular near salt flanks, and a good structural agreement between Figures 5(b) and 5(c) serve as an indirect evidence of the algorithm success. An ultimate validation should come from prestack depth migration velocity analysis, which is significantly more expensive.

CONCLUSIONS

We have applied the recently established theorem that the Dix velocity obtainable from the time-migration velocity is the true interval velocity divided by the geometrical spreading of image rays to pose the corresponding inverse problem. We have suggested a set of numerical algorithms for solving the problem numerically. We have tested these algorithms on a synthetic data example with laterally heterogeneous velocity and demonstrated that they produce significantly better results than simple Dix inversion followed by time-to-depth conversion. Moreover, the Dix velocity may qualitatively differ from the output velocity. We have also tested our algorithm on a field data example and validated it by comparing prestack time migration image mapped to depth with post-stack depth migrated images. Our approach is complementary to velocity estimation methods that work directly in the depth domain. Therefore, it can serve as an efficient first step in seismic velocity model building.

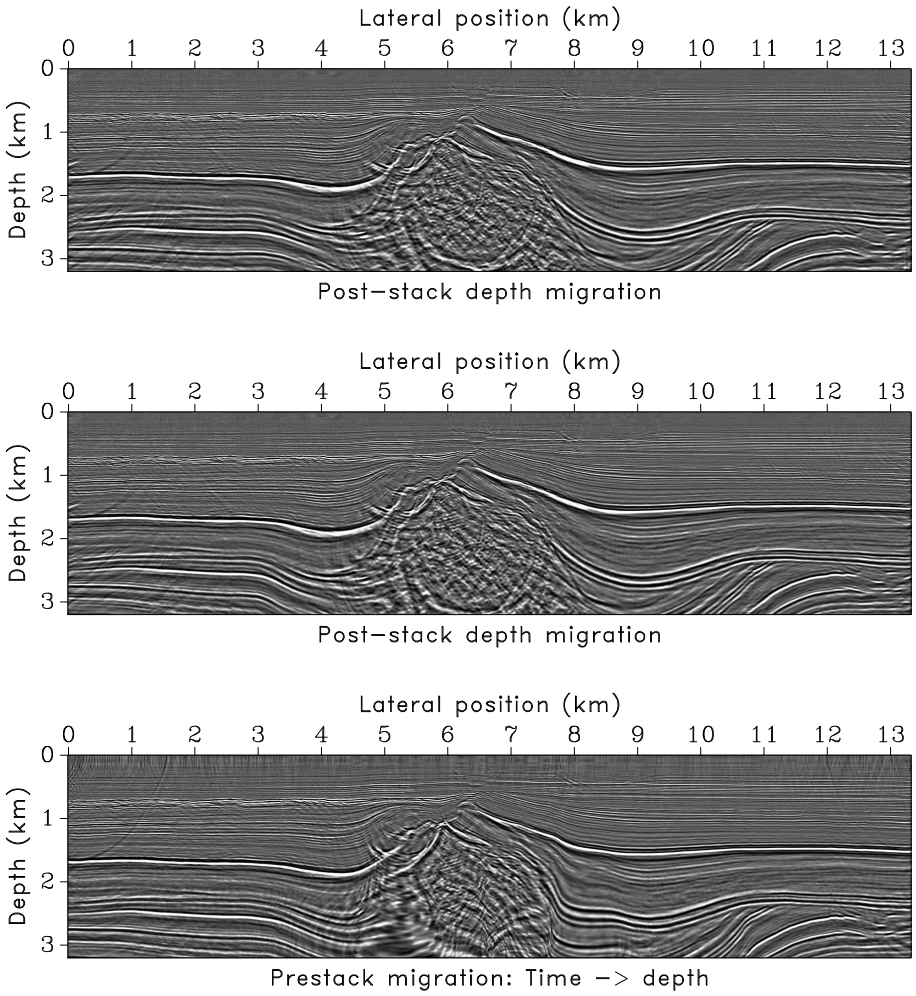


Figure 5: Migrated images of the field data example. (a) Poststack migration using Dix velocity. (b) Poststack migration using estimated velocity. (c) Prestack time migration converted to depth with our algorithm. time2depth/masha2 img0,img,fmg2

REFERENCES

[1] Bevc, D., J. L. Black, and G. Palacharla, 1995, Plumes: Response of time migration to lateral velocity variation: Geophysics, **60**, 1118–1127.

[2] Boyd, J. P., 2001, Chebyshev and Fourier spectral methods, 2nd ed. (revised): Dover Publications.

[3] Cameron, M., S. Fomel, and J. Sethian, 2007, Seismic velocity estimation from time migration: Inverse Problems, **23**, 1329–1369.

[4] Červený, V., 2001, Seismic ray method: Cambridge University Press.

[5] Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68–86.

[6] Fomel, S., 2003, Time-migration velocity analysis by velocity continuation: Geophysics, **68**, 1662–1672.

[7] Hatton, L., L. K. Larner, and B. S. Gibson, 1981, Migration of seismic data from inhomogeneous media: Geophysics, **46**, 751–767.

[8] Hubral, P., 1977, Time migration - Some ray theoretical aspects: Geophysical Prospecting, **25**, 738–745.

[9] Hubral, P., and T. Krey, 1980, Interval velocities from seismic reflection time measurements: Soc. of Expl. Geoph.

[10] Kim, Y. C., W. B. Hurt, L. J. Maher, and P. J. Starich, 1997, Hybrid migration: A cost-effective 3-D depth-imaging technique: Geophysics, **62**, 568–576.

[11] Larner, K. L., L. Hatton, B. S. Gibson, and I. S. Hsu, 1981, Depth migration of imaged time sections: Geophysics, **46**, 734–750.

[12] Lax, P. D., 1954, Weak solutions of hyperbolic equations and their numerical computation: Communications in Pure and Applied Mathematics, **7**, 159–193.

[13] Popov, M. M., 2002, Ray theory and Gaussian beam method for geophysicists: EDUFBA.

[14] Popov, M. M., and I. Pšenčik, 1978, Computation of ray amplitudes in inhomogeneous media with curved interfaces: Studia Geophysica et Geodaetica, **22**, 248–258.

[15] Robein, E., 2003, Velocities, time-imaging and depth-imaging in reflection seismics: EAGE.

[16] Sethian, J., 1996, Fast marching level set method for monotonically advancing fronts: Proceedings of the National Academy of Sciences, **93**, 1591–1595.

[17] ———, 1999, Fast marching methods: SIAM Review, **41**, 199–235.

[18] Siliqi, R., and N. Bousquié, 2000, Anelliptic time processing based on a shifted hyperbola approach: 70th Annual International Meeting, Soc. of Expl. Geophys., 2245–2248.

[19] Yilmaz, O., 2001, Seismic data analysis: Soc. of Expl. Geophys.

[20] Yilmaz, O., I. Tanir, and C. Gregory, 2001, A unified 3-D seismic workflow: Geophysics, **66**, 1699–1713.

A 1-D time-varying median filter for seismic random, spike-like noise elimination

Yang Liu[*†], *Cai Liu*[*], *Dian Wang*[**]

ABSTRACT

Random noise in seismic data affects the signal-to-noise ratio, obscures details, and complicates identification of useful information. We present a new method for reducing random, spike-like noise in seismic data. The method is based on a 1-D stationary median filter (MF) – the *1-D time-varying median filter* (TVMF). We design a threshold value that controls the filter window according to characteristics of signal and random, spike-like noise. In view of the relationship between seismic data and the threshold value, we chose median filters with different time-varying filter windows to eliminate random, spike-like noise. When comparing our method with other common methods, e.g., the band-pass filter and stationary MF, we found that the TVMF strikes a balance between eliminating random noise and protecting useful information. To demonstrate the feasibility of our method in reducing seismic random, spike-like noise, we present results for one synthetic dataset. Results of applying the method to seismic land data from Texas demonstrate that the TVMF method is effective in practice.

INTRODUCTION

Random noise in prestack seismic data can come from various sources, such as wind motion, poorly planted geophones, or electrical noise, and some of this seismic random noise invariably exhibits spike-like characteristics. Although stacking can at least partly suppress random noise in prestack data, residual random noise after stacking will decrease the accuracy of final data interpretation. In recent years, several authors have developed effective methods of eliminating random noise. For example, (5) used the noncausal prediction filter for random-noise attenuation, (12) compared several adaptive filters, which they applied in an attempt to reduce random noise in geophysical data. (6) applied complex-trace analysis to seismic data for random-noise suppression, recommending it for low-fold seismic data, and some transform methods were also used to eliminate seismic random noise, e.g., seislet transform (3; 4), discrete cosine transform (9), and curvelet transform (11).

On the other hand, the median filter, a well-known method that can effectively suppress spike-like noise, refers to nonlinear signal processing. (1) and (2) found the

[*]**e-mail:** sergey.fomel@beg.utexas.edu

method to be both simple and effective for seismic prospecting. More recently, new median filters have been proposed. (10) incorporated median-filter noise reduction into standard Kirchhoff time migration. (13) used a hyperbolic median filter to suppress multiples. (7) advocated random-noise attenuation using the 2-D multistage median filter (MLM).

Because the median filter is a nonlinear filter, filter-window length needs to be adjusted before its characteristics can be changed. The stationary filter, on the other hand, maintains a fixed window length, retaining useful information and random noise at the same scale. An unsuitable filter-window choice would therefore end up in useful information being destroyed or noise remaining. Here we propose a time-varying median filter that adjusts to different filter-window lengths by threshold, making value judgments on useful information versus noise throughout the process. We show that the time-varying window is more powerful than the stationary window.

In this paper, a new nonlinear filter called the *time-varying median filter* (TVMF) is presented, which we designed by defining a threshold value. After adjustment of the filter window in the time domain, this filter eliminates random noise in seismic data. We compare TVMF with other common methods. We use numerical examples, along with synthetic and field data, to demonstrate the validity of the proposed method in practice.

THEORETICAL BASIS

In contrast to useful information, random, spike-like noise in seismic data is neither continuous nor correlative, and a 1-D stationary median filter, having a large filter-window length, can easily remove such noise. However, signal can be damaged by such a filter.

The 1-D TVMF is based on the 1-D stationary median filter. We propose to measure the local noise content of the data and to adjust the filter length adaptively. If a threshold value that judges random noise and estimates noise intensity can be chosen, a filter having a large filter-window length can eliminate random noise while processing useful information by using a small filter-window length. Such a filter can thus effectively eliminate random noise while maximizing preservation of a detailed structure of useful information.

1-D stationary median filter (MF)

The 2-D seismic record can be represented by the following data sequence:

$$x_{i,j} \qquad (i = 1, \cdots, m, \cdots, N_x; j = 1, \cdots, n, \cdots, N_t)\,, \tag{1}$$

where i is the spatial sample index, j is the temporal sample index, and N_x and N_t are the numbers of spatial and temporal samples. When filter-window length, C, of

the stationary median filter is defined (normally C is odd), the result after filtering at the point on the m^{th} trace and the n^{th} sample can be found by

1. Setting the center point at the m^{th} trace and the n^{th} sample, and choosing C samples in the m^{th} trace,

2. Sorting the C samples from smaller to larger, and then

3. Picking the center value, after sequencing, as the output at the point on the m^{th} trace and the n^{th} sample.

Repetition of the process on all data achieves 1-D stationary median filtering of the seismic record. The 1-D stationary median filter can be expressed as $median[x_{i,j}]$.

Signal-to-noise ratio (SNR) estimation using the stack method

A simple definition of the SNR was introduced by (8). Window D, a part of the seismic record, can be chosen for SNR analysis:

$$D = [x_{i,j}]_{M\times N} \qquad (0 < M \leq N_x, 0 < N \leq N_t) \ . \tag{2}$$

Further assumptions are: waveform, amplitude, and phase of seismic wavelet in window D keep stable in respect to distance "i"; noise is "zero mean" randomly distributed, along with survey-line direction being independent (decorrelated) of the signal, so that

$$x_{i,j} = s_j + n_{i,j} \tag{3}$$

$$\sum_{i=1}^{M} n_{i,j} = 0 \ , \tag{4}$$

where s_j is amplitude of signal, and $n_{i,j}$ is amplitude of noise. These assumptions generally imply a limitation to this method, but they can be satisfied if the local window is chosen in the stable signal region of the seismic section. So if the signal energy in the window is

$$E_S = M \sum_{j=1}^{N} s_j^2 = \frac{1}{M} \sum_{j=1}^{N} (\sum_{i=1}^{M} x_{i,j})^2 \ , \tag{5}$$

the noise energy can be calculated by

$$E_N = \sum_{j=1}^{N} \sum_{i=1}^{M} x_{i,j}^2 - E_S \ . \tag{6}$$

Finally, a decibel expression of the SNR is estimated as

$$SNR = \frac{E_S}{E_N} = 10 \ \log_{10} \left(\frac{\sum_{j=1}^{N} \big(\sum_{i=1}^{M} x_{i,j} \big)^2}{M \sum_{j=1}^{N} \sum_{i=1}^{M} x_{i,j}^2 - \sum_{j=1}^{N} \big(\sum_{i=1}^{M} x_{i,j} \big)^2} \right) . \tag{7}$$

1-D time-varying median filter (TVMF)

Using the above definitions, a TVMF can be designed. We use the following three steps to determine its parameters:

1. Choose the reference median filter length.

At point $x_{m,n}$, where the filter-window length of the reference median filter is chosen as C, output can be expressed as

$$Y^C_{m,n} = median[x_{i,j}] \qquad (i = m; j = n - (C-1)/2, \cdots, n + (C-1)/2) \ , \tag{8}$$

where filter-window length C is a large odd number so that random noise could be eliminated as much as possible. C is determined by using the SNR estimation method, which will be discussed later.

2. Choose the threshold value.

Using the reference median filter with its large filter-window length, we processed the seismic data first to find $Y^C_{m,n}$. Then we applied the absolute mean value to calculate the threshold value, which is shown as

$$T = \frac{1}{N_x \times N_t} \sum_{i=1}^{N_x} \sum_{j=1}^{N_t} |Y^C_{i,j}| \ , \tag{9}$$

We can evaluate random-noise data versus useful signal data by using the threshold value. When $|Y^C_{i,j}| < T$, the point is judged to be random noise, whereas when $|Y^C_{i,j}| \geq T$, the point should be signal data. We can therefore use the threshold value as a judgment norm – data in which $|Y^C_{i,j}| \geq T$ should be processed by the median filter having windows smaller than C to protect the detailed signal structure. Data in which $|Y^C_{i,j}| < T$ should be processed by the median filter having windows larger than C to strengthen its ability to eliminate random noise.

3. Choose the time-varying filter windows.

Choices involving time-varying windows abound after the threshold value has been chosen. We can define four scales of windows. Detailed time-varying window length $C_{i,j}$ is defined as

$$C_{i,j} = \begin{cases} C + \alpha, & 0 < |Y^C_{i,j}| < T/2 \\ C + \beta, & T/2 < |Y^C_{i,j}| < T \\ C - \gamma, & T \leq |Y^C_{i,j}| < 2T \\ C - \delta, & |Y^C_{i,j}| \geq 2T \end{cases} \ , \tag{10}$$

where α, β, γ, and δ are constant even numbers, and $\alpha > \beta$ and $\delta > \gamma$. Specific values for these parameters will be discussed in the next section. Using the above definition, we distinguish between random noise and useful signal, such that we can process the seismic data using different filter scales.

SYNTHETIC DATA TESTS

To choose reference filter windows and time-varying filter windows, we analyzed the characteristics of our TVMF formulation using synthetic data. Referring to the velocity model (Table 1), note the synthetic common-shot record with a dominant frequency of 40 Hz shown in Figure 1a. A 5%-density white spike noise was then added to the model. Three different noise peaks (noise amplitude is half, one, and two times the maximum value of the reflections, and the corresponding noise intensity is 1/2, 1, and 2) were chosen to separately test the TVMF. Only noise with twice the maximum value of the reflected wave is displayed in Figure 1b. Energy attenuation has not been taken into account, nor has spherical spreading or AVO. Here, the signal has been drowned out by noise.

Stratum thickness (m)	500	1739	2069	
Stratum velocity (m/s)	2500	3162	4138	4500

Table 1: Velocity model

Definitions of signal energy (equation 8), noise energy (equation 9), and SNR (equation 10) were used to analyze characteristics of the reference median filter. We chose the stationary MF with different filter-window lengths of from 3 to 19 points, comparing various signal energies (Figure 2a), noise energies (Figure 2b), and the SNR's (Figure 2c) after filtering. To meet the assumptions of the SNR model, we chose five traces near the zero offset, where seismic events are approximately invariant across the five traces. Note that in the figures, curves of the three noise intensities (1/2, 1 and 2) appear to have similar tendencies. Both signal energy and noise energy decrease as filter-window length increases, but the SNR displays a wiggle shape. The energy levels of signals are stable after filter-window length reaches 11 points, but the energy levels of noise are stable after filter-window length reaches 15 points. And because the curves show a different rate of descent, the SNR has a different tendency. It reaches a peak at the 5-point filter window and decreases afterward because signal energy attenuates faster than noise energy. The SNR then reaches the minimum near 11 points, where the signal energy is stable, and next, the SNR improves again because the noise energy still decreases. Finally, the SNR reaches stability, however, the signal has been barely damaged.

In the noisy model (Figure 1b), the SNR is -10.7 dB. After stationary MF filtering, even the minimum SNR remains much larger than -10.7 dB, illustrating that all stationary median filters can improve the SNR in the noisy model. We can select large filter windows of reference median filters to keep noise energy to a minimum when threshold T is well able to separate signal from noise, but at the same time we can define time-varying filter windows on the basis of reference filter windows so that the reference filter window will be limited. Three basic principles should be satisfied:

1. The reference filter window can be chosen as a large value to separate signal from noise when noise energy is small.

2. Time-varying filter windows at signal positions should be small values to protect the signal when removing noise.

3. Time-varying filter windows at noise positions should be large values to attenuate noise energy.

In Figure 2, reference filter window C should be larger than the stable signal point (11 points), and time-varying filter windows $C-\gamma$ and $C-\delta$ at signal positions should be in the range of from 5 to 7 points in order to preserve signal energy. Time-varying filter windows $C+\alpha$ and $C+\beta$ should be limited in range from 11 to 13 points in order to attenuate noise energy and save calculation time. To meet all principles, C can be 11 points, with $\alpha = 2$, $\beta = 0$, $\gamma = 4$, and $\delta = 6$. After more tests on synthetic and real data it became clear that these filter parameter choices work for most real data.

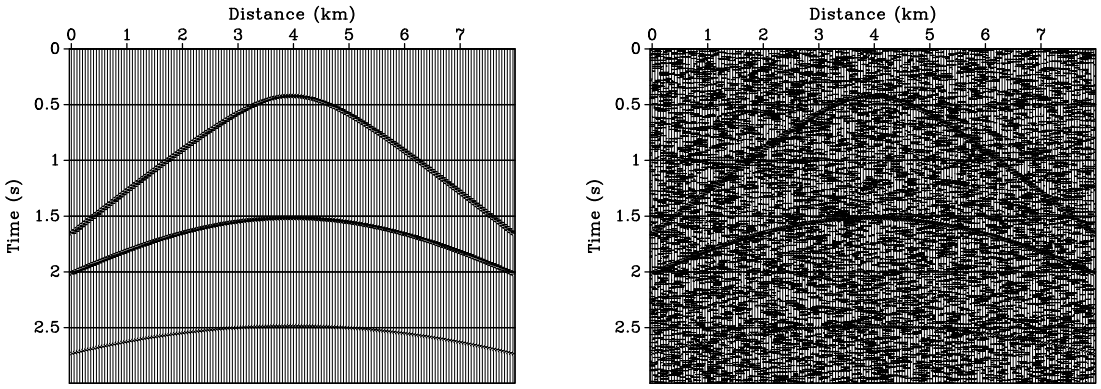


Figure 1: Synthetic model (a) and white-noise model (b). medianfilter/model model,noise0

We used the TVMF, with defined reference and time-varying filter windows, to process the noise model (Figure 1b), the result of which is shown in Figure 3a. When assumptions can be met, the SNR estimation using the stack method can be used to compare results. After TVMF processing, the SNR is 19.7 dB. To compare, we also used the low-pass filter for preserving the signal in the dominant frequency band (Figure 3b, the SNR is -7.4 dB). After TVMF processing, white spike noise attenuated well, but the result after low-pass filtering became a signal with band-limited noise, and a great deal of low-frequency noise remained. Next, we also used the TVMF to process the band-limited noise model (Figure 3b). SNR analysis shows that the parameters of TVMF only change a little, so the same parameters can be used for processing. The TVMF cannot eliminate all band-limited noise, like white noise, but the energy of the noise has been degraded. At the same time, the TVMF introduces a few high-frequency noise components having low energy, but these can be easily removed by using a high-cut filter. Figure 4a shows the result after TVMF and high-cut filtering; the SNR is 0.5 dB. The low-pass filter does nothing about band-limited noise. The 11-point stationary MF (Figure 4b, the SNR is -8.4 dB) can be used

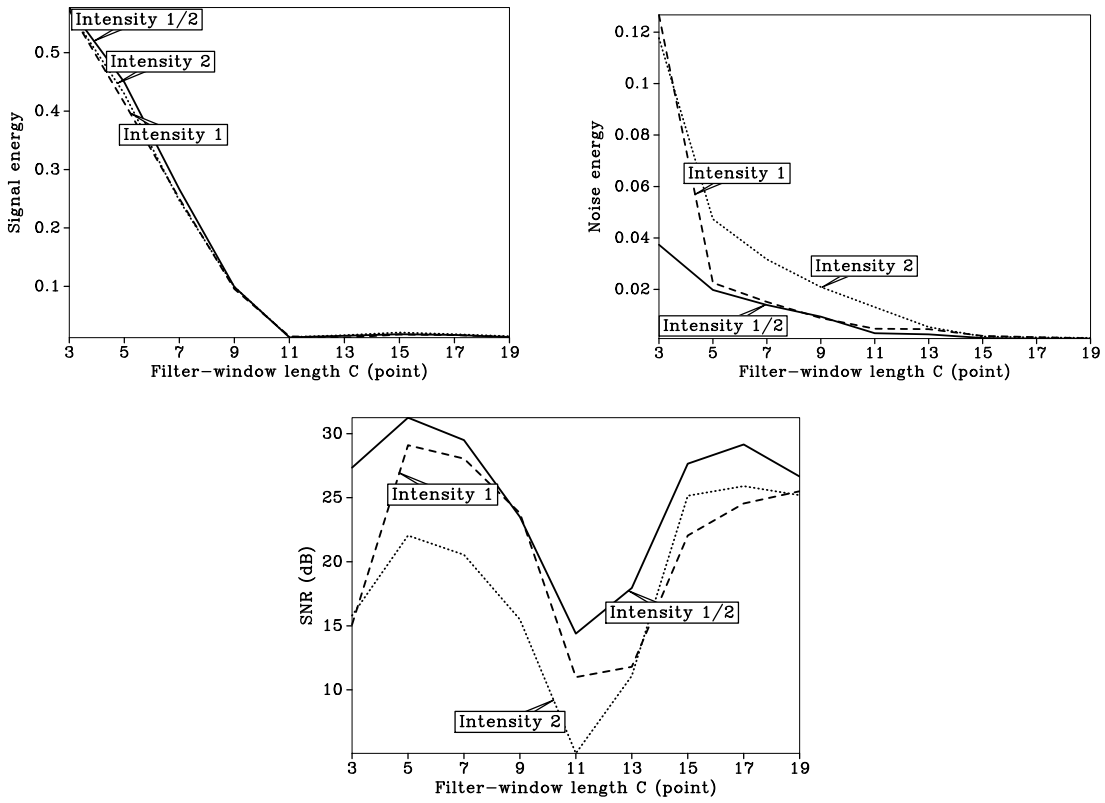


Figure 2: Comparison of different noise levels (Intensity (1/2, 1, and 2) is the amplitude ratio between noise and reflections). Signal energy (a), noise energy (b), and SNR (c). medianfilter/window es,en,snr

to compare with the TVMF. The stationary MF can remove most of the noise, but useful information is also destroyed. After comparing the result of the TVMF with those of the stationary MF and the band-pass filter, we conclude that the TVMF is superior when processing random, spike-like noise.

We can compare results of using different methods by analyzing their spectra as well. We chose spectra of the trace at a distance of 4 km, corresponding to the pertinent parts of Figures 1, 3, and 4. Results in Figure 5 show that spectral values of random noise are larger than those of the signal at every frequency and that reflected waves have been masked by random noise (Figure 5b). The low-pass filter can eliminate noise in the high-frequency band, but it does nothing in the dominant-frequency band (Figure 5c). After the band-limited model is processed by the TVMF using an 11-point reference filter window, most spectral components in the dominant-frequency band can be recovered. An additional high-cut filter was used to remove high-frequency noise introduced by the TVMF (Figure 5d). The corresponding time profile is shown in Figure 4a.

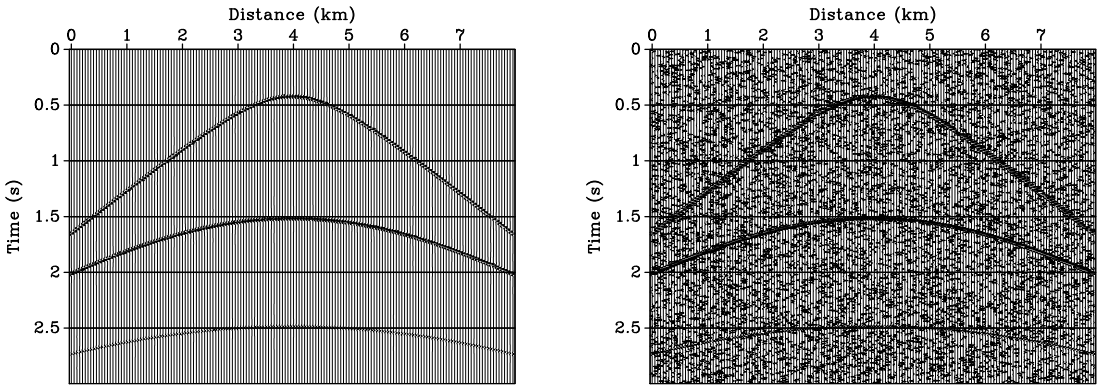


Figure 3: Denoised result after 11-point TVMF filtering (a) and denoised result (band-limited noise model) after low-pass filtering (b). medianfilter/model w11tvmf,noise1

Given the result of model filtering, the TVMF can easily remove spike-like noise, especially noise having a white spectrum. When the TVMF is used to attenuate band-limited, spike-like noise, its filter ability decreases, although it can still work better than other common methods.

PROCESSING OF FIELD DATA

A real-data example involves land prestack data from Texas (Figure 6a), from which we show one common midpoint gather that has 59 traces and an average group interval of 42 m. The sample frequency is 4 ms, and the sample number is 500. In Figure 6a, noise is mainly random noise caused by the surface condition of this area; the 60-Hz electrical interference with high amplitude is especially serious. According

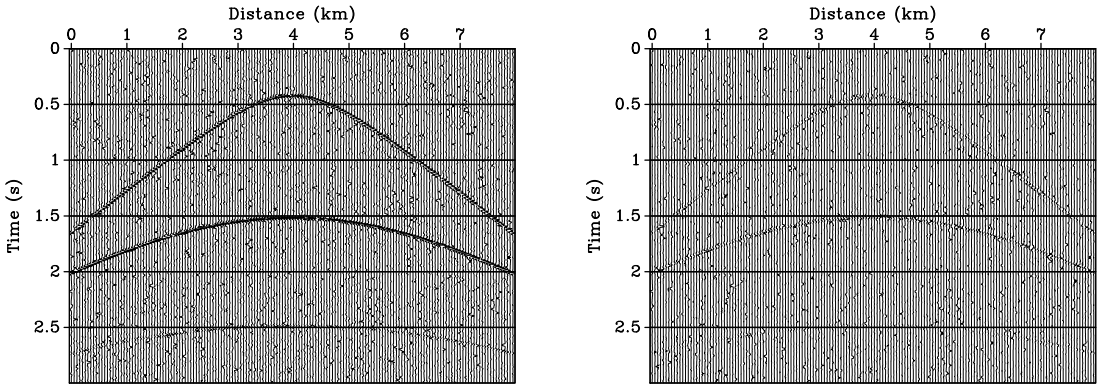


Figure 4: Denoised band-limited noise model using different methods (the result of Figure 3b was input to these tests); 11-point TVMF followed by a high-cut filter (a) and 11-point stationary MF (b). medianfilter/model cb11tvmf,c11mf

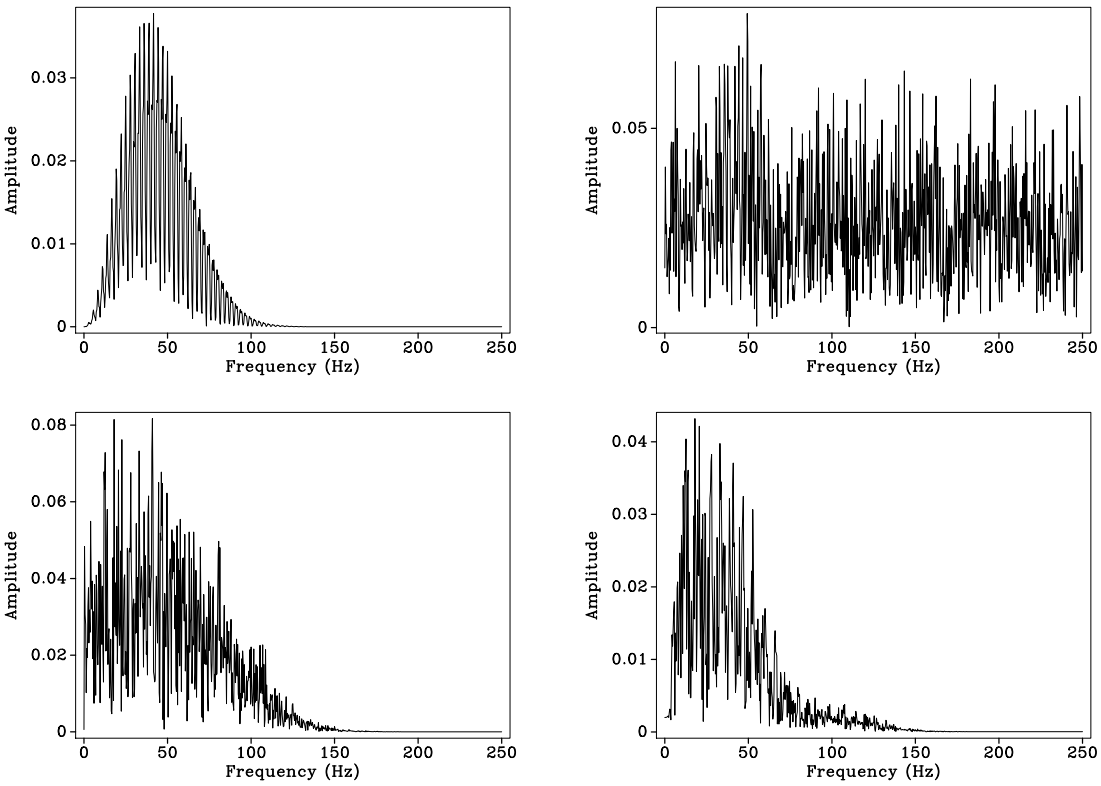


Figure 5: Comparison of amplitude spectra (trace at distance of 4 km) before and after processing. Amplitude spectra in Figure 1a (a), corresponding spectra in Figure 1b (b), corresponding spectra in Figure 3b (c), and corresponding spectra in Figure 4a (d). medianfilter/model smodel,snoise0,snoise1,scb11tvmf

to the amplitude spectra, the dominant frequency of the real data is about 25 Hz; the 60-Hz electrical interference thus exhibits high-frequency, spike-like characteristics when compared with the useful signal. In the time domain (Figure 6a, especially at locations "A," "B," and "C"), this spike-like characteristic of electrical interference can also be observed. To meet the assumptions of SNR estimation using the stack method, a rectangle region "D" that has 9 traces and 125 samples is chosen, and, after calculating, the SNR in the original data is found to be -7.3 dB. First, we apply a low-pass filter to remove the random noise (Figure 6b); the frequency is limited to 60 Hz. The low-pass filter can partly eliminate random noise, and the corresponding SNR is improved to 0.4 dB, but it cannot attenuate the noise in the dominant-frequency band. We can thus still see lots of random noise after filtering. The 11-point stationary MF is also used to compare with the TVMF, and the result is shown in Figure 6c. After processing, there is still some spike-like noise left, and part of the signal has been attenuated. The SNR is 1.9 dB, and a larger filter window will further destroy the useful signal. By using the SNR analysis introduced in the theory section, we are choosing the parameters about the TVMF in window "D." The TVMF with the same parameters as for the synthetic examples, $C = 11$, $\alpha = 2$, $\beta = 0$, $\gamma = 4$, and $\delta = 6$, is applied to the prestack data (Figure 6d). Figure 6e shows the difference section, in which the processed data using the TVMF have been subtracted from the original data. After TVMF processing, continuity of events and information of the reflected waves are all enhanced, and there is little noise left. Note that the 60-Hz electrical interference in particular can be completely removed. The SNR reaches 4.7 dB. In the difference section it can be seen that the TVMF eliminates major spike-like noise and that there is little useful information beyond the spike-like noise, showing that the TVMF can effectively eliminate spike-like noise and protect useful information at the same time. Because the TVMF is based on the stationary MF that can filter noise spikes, the TVMF can effectively eliminate spike-like noise. And because the TVMF can separate useful information from spike-like noise, it can protect useful information and perform better overall than other potential methods in spike-like noise attenuation.

We chose a trace from the original data and corresponding traces in processed data and got the amplitude spectra of four traces (Figure 7). After low-pass filtering, high-frequency noise can be removed thoroughly, but some noise in the dominant-frequency band can remain; electrical interference in particular must be eliminated by other methods (Figure 7b). Stationary MF can eliminate enough noise, but spectra of useful signals have been distorted (Figure 7c). In Figure 7d, the TVMF can attenuate noise in the entire frequency band and protect the useful-information frequency components.

Next we obtained stacked sections corresponding to Figure 6b and 6d; the stacking fold is 59. A rectangular region, "H," was chosen to calculate the local SNR. Events of the stacked section using low-pass processing were discontinuous (Figure 8a), the SNR is -7.8 dB, and the continuity of events in the stacked section using TVMF processing was improved, especially at locations "E," "F," and "G" (Figure 8b), and the corresponding SNR is -3.1 dB. Because the TVMF can eliminate spike-like noise

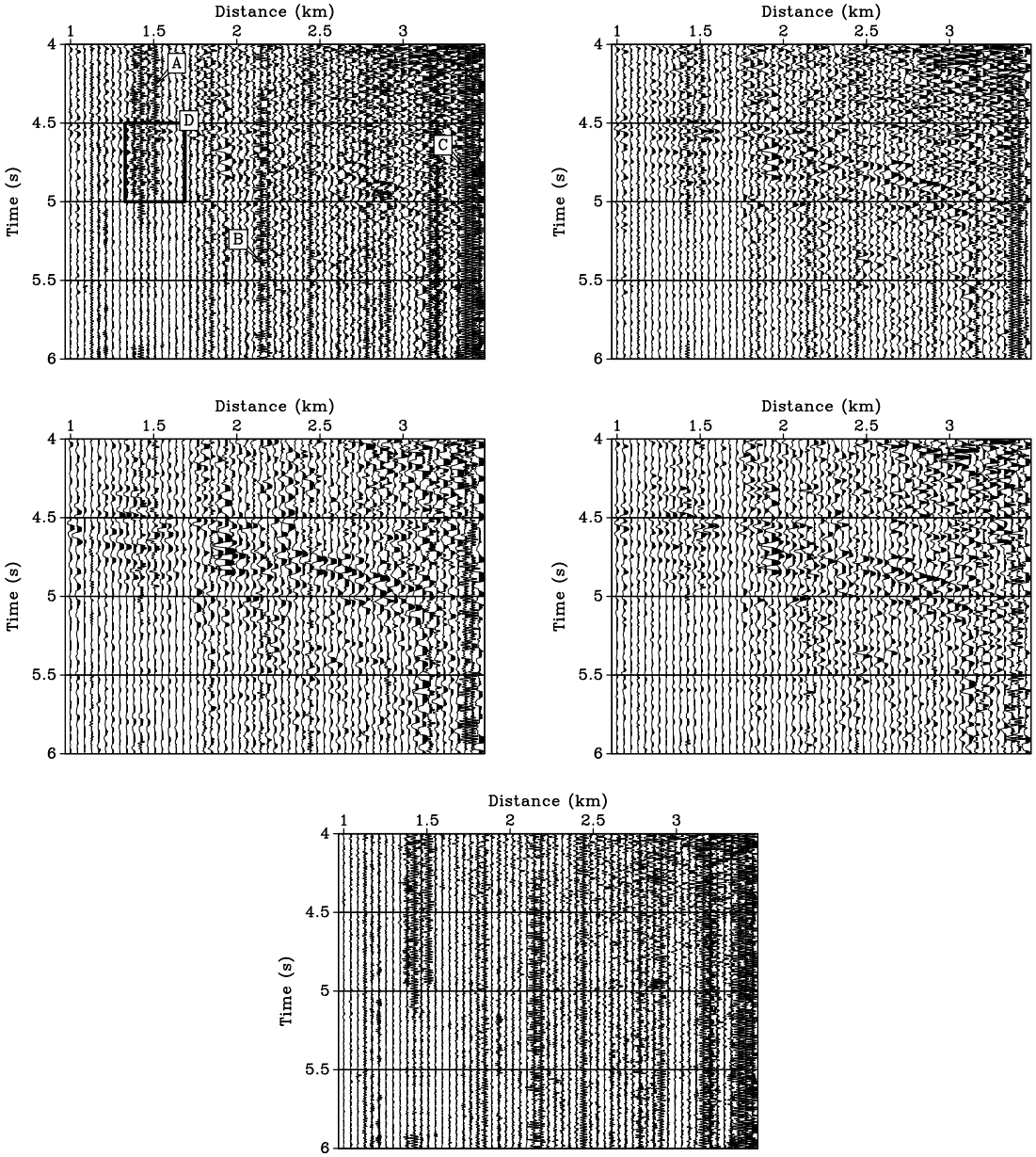


Figure 6: Real land data (a), denoised result after low-pass filtering (b), denoised result after stationary MF filter (c), denoised result after TVMF filtering (d), and difference between original data and result after TVMF processing (e). medianfilter/dragon dragon1,band,11mf,11tvmf,diff

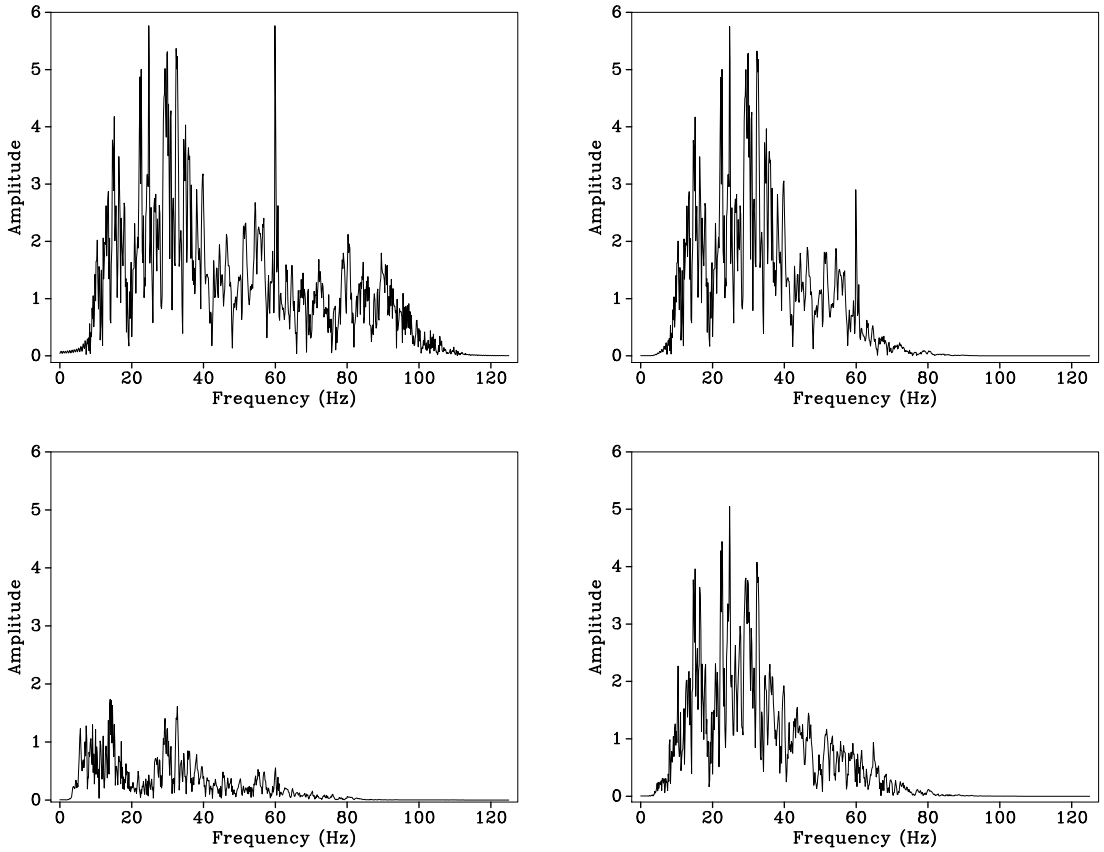


Figure 7: Comparison of amplitude spectra (trace at distance of 1.8 km) before and after processing. Amplitude spectra in Figure 6a (a), corresponding spectra in Figure 6b (b), corresponding spectra in Figure 6c (c), and corresponding spectra in Figure 6d (d). medianfilter/dragon dspec,bspec,mfspec,fspec

in prestack data, it is advantageous to apply to such data.

CONCLUSION

We have proposed and tested the time-varying median filter (TVMF) for attenuating prestack random, spike-like noise. We used a signal-to-noise ratio estimation to design the reference filter window and time-varying filter windows, in which the filter can adapt filter-window length to meet the intensity of the random, spike-like noise. The ability of this method to eliminate random noise while protecting desired signal further attests to the strength of the method. Our experiments show that in field data the TVMF can eliminate random noise enough to enhance the continuity of events. Spectral analysis also shows that the TVMF can effectively suppress random, spike-like noise in the whole frequency band. Comparison of different methods shows that the TVMF is more effective than the stationary median filter for eliminating spike-like noise, and it can enhance results of band-pass filtering by attenuating spike-like noise within the pass band.

ACKNOWLEDGMENTS

The authors gratefully acknowledge Sergey B. Fomel for his suggestions and help. We thank Douglas W. McCowan for fruitful discussions. We sincerely thank assistant editor Vladimir Grechka, associate editor Eric Verschuur, and three anonymous reviewers for their constructive comments and suggestions. We also thank one anonymous reviewer for his suggestion about the possibility of our method on marine swell-noise attenuation. Thanks also to BGP International for financial support of this work. This work is partly supported by The National Basic Research Program of China (973 Program: 2009CB219301). Publication authorized by the Director, Bureau of Economic Geology.

REFERENCES

[1] Bednar, J. B., 1983, Applications of median filtering to deconvolution, pulse estimation, and statistical editing of seismic data: Geophysics, **48**, 1598–1610.

[2] Duncan, G., and G. Beresford, 1995, Some analyses of 2-D median f-k filters: Geophysics, **60**, 1157–1168.

[3] Fomel, S., 2006, Towards the seislet transform: 76th Annual International Meeting, SEG, Expanded Abstracts, 2847–2851.

[4] Fomel, S., and Y. Liu, 2008, Seismic data analysis with one-dimensional seislet frame: 78th Annual International Meeting, SEG, Expanded Abstracts, 2581–2585.

[5] Gülünay, N., 2000, Noncausal spatial prediction filtering for random noise reduction on 3-D poststack data: Geophysics, **65**, 1641–1653.

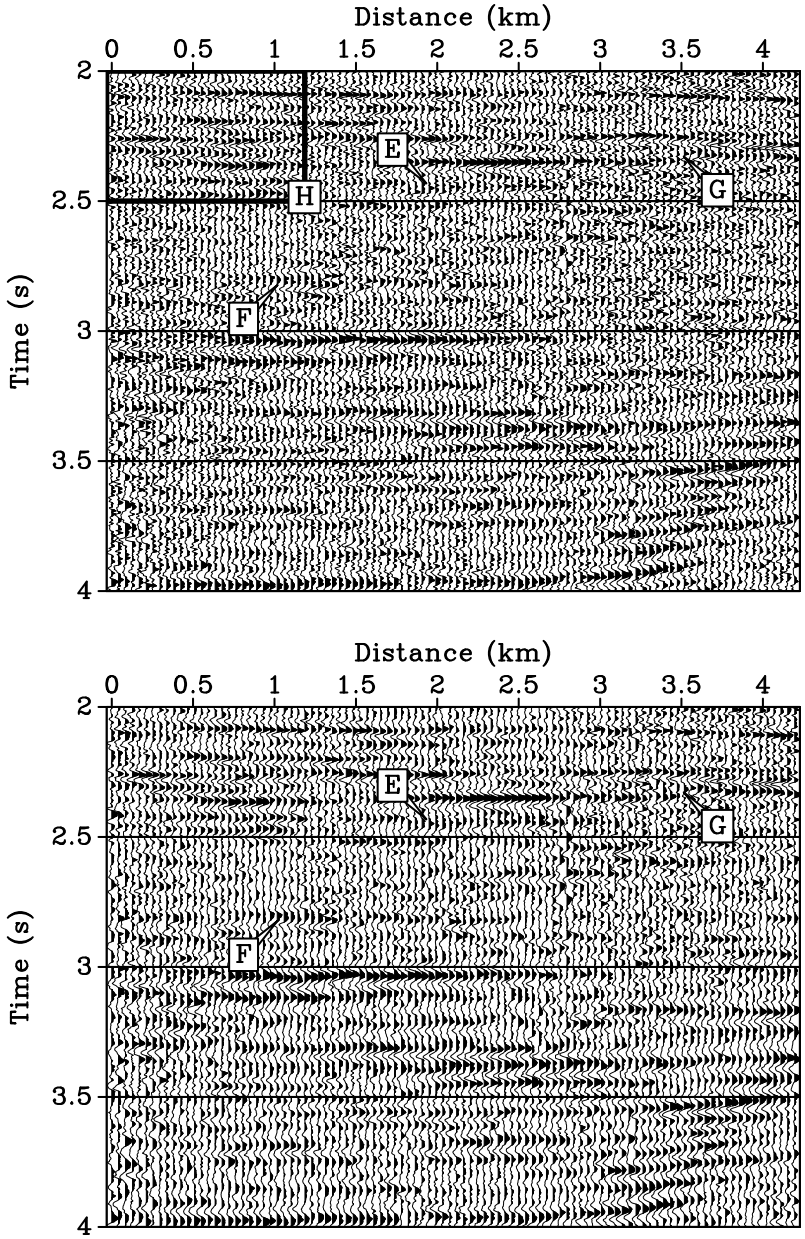


Figure 8: Stacked section of prestack data after low-pass filtering (a) and corresponding section after TVMF processing (b). medianfilter/dragon stack1,stack2

[6] Karsli, H., D. Dondurur, and G. Çifçi, 2006, Application of complex-trace analysis to seismic data for random-noise suppression and temporal resolution improvement: Geophysics, **71**, V79–V86.
[7] Liu, C., Y. Liu, B. Yang, D. Wang, and J. Sun, 2006, A 2D multistage median filter to reduce random seismic noise: Geophysics, **71**, V105–V110.
[8] Liu, Y., and C. Li, 1997, Some methods for estimating the signal/noise ratio of seismic data: Oil Geophysical Prospecting, **32**, 257–262 (In Chinese with English abstract).
[9] Lu, W., and J. Liu, 2007, Random noise suppression based on discrete cosine transform: 77th Annual International Meeting, SEG, Expanded Abstracts, 2668–2672.
[10] Mi, Y., and G. F. Margrave, 2000, Median filtering in Kirchhoff migration for noisy data: 70th Annual International Meeting, SEG, Expanded Abstracts, 822–825.
[11] Neelamani, R., A. I. Baumstein, D. G. Gillard, M. T. Hadidi, and W. I. Soroka, 2008, Coherent and random noise attenuation using the curvelet transform: The Leading Edge, **27**, 240–248.
[12] Ristau, J. P., and W. M. Moon, 2001, Adaptive filtering of random noise in 2-D geophysical data: Geophysics, **66**, 342–349.
[13] Zhang, R., and T. J. Ulrych, 2003, Multiple suppression based on the migration operator and a hyperbolic median filter: 73th Annual International Meeting, SEG, Expanded Abstracts, 1949–1952.

Velocity-independent time-domain seismic imaging using local event slopes

*Sergey Fomel**

ABSTRACT

I show that, by estimating local event slopes in prestack seismic reflection data, it is possible to accomplish all common time-domain imaging tasks, from normal moveout to prestack time migration, without the need to estimate seismic velocities or any other attributes. Local slopes contain complete information about the reflection geometry. Once they are estimated, seismic velocities and all other moveout parameters turn into data attributes and are directly mappable from the prestack data domain into the time-migrated image domain. I develop an analytical theory for this method and demonstrate its applicability on synthetic and field data examples.

INTRODUCTION

The conventional approach to seismic data analysis (38) consists of two main steps: estimating seismic velocities (the subsurface macromodel) and seismic imaging (mapping of the reflected seismic energy to the reflector positions). The two steps can be repeated when the velocity model gets refined by imaging. Estimating velocities remains one of the most labor-intensive and time-consuming tasks in seismic data processing although several approaches have been developed to partially automate and simplify it (1; 33; 25). In time-domain imaging, effective seismic velocities are picked from semblance scans. More than just a single parameter (such as seismic velocity) needs to be picked to use higher-order traveltime approximations such as those suggested by the multifocusing approach (26; 20).

The idea of *velocity-independent time-domain seismic imaging* belongs to (28), who considered decomposing seismic data into a range of local slopes. (37) observed that it is possible to perform moveout analysis by estimating local data slopes in the prestack data domain using an automatic method such as plane-wave destruction (13). In this paper, I extend Ottolini's idea of velocity-independent imaging and show that extracting local event slopes in prestack data is sufficient for accomplishing all common time-domain imaging tasks, from hyperbolic and non-hyperbolic normal moveout corrections to dip moveout and prestack time migration. Rather than being a prerequisite for seismic imaging, seismic velocities turn into data attributes that can be extracted from the input data simultaneously with imaging.

*e-mail: sergey.fomel@beg.utexas.edu

The idea of using local event slopes estimated from prestack seismic data goes back to the work of (30) and (29). It was used later by (36) and extended in the method of stereotomography (4; 3; 24; 23). In the depth imaging context, local data slopes were also utilized by (2) for anti-aliased Kirchhoff migration and by (21) in the method of parsimonious depth migration. In this paper, I extend and formalize the application of these ideas to time-domain imaging. By analogy with the oriented wave equation, which describes wave propagation in the space of local orientations (14), I use the term *oriented* when referring to local slopes.

The analysis proceeds from normal moveout (NMO) to dip moveout (DMO) and prestack time migration. I derive analytical expressions for mapping the data attributes to the imaging domain and demonstrate their use with synthetic and field data tests.

ORIENTED TIME-DOMAIN IMAGING

Let us start with a simple hyperbolic moveout correction and then consider a progression of time-domain imaging operators with increasing complexity.

Normal moveout

Let us consider the classic hyperbolic model of reflection moveout:

$$t(l) = \sqrt{t_0^2 + \frac{l^2}{v_n^2(t_0)}} \; , \tag{1}$$

where t_0 is the zero-offset traveltime, $t(l)$ is the corresponding traveltime recorded at offset l, and $v_n(t_0)$ is the stacking or RMS velocity. As follows from equation 1, the traveltime slope $p = dt/dl$ is given by

$$p = \frac{l}{t\, v_n^2(t_0)} \; . \tag{2}$$

Knowing the local slope $p(t,l)$ in a common-midpoint gather, one can effectively eliminate the velocity $v(t_0)$ from equations 1 and 2 arriving at the velocity-independent moveout equation (28)

$$t_0 = \sqrt{t^2 - t\, p(t,l)\, l} \; . \tag{3}$$

Equation 3 describes a direct mapping from prestack data to the zero-offset time. This equation constitutes an *oriented normal moveout* correction. As follows from equation 2, the velocity is mapped as follows:

$$\frac{1}{v_n^2} = p(t,l)\, \frac{t}{l} \; . \tag{4}$$

Unlike its role in the conventional normal moveout (NMO), velocity is not a requirement for the moveout correction but rather a data attribute derived from the local slopes and mappable directly to the appropriate zero-offset time.

A simple synthetic example is shown in Figure 1. The synthetic data (Figure 1a) were generated by applying inverse NMO with time-variable velocity and represent perfectly hyperbolic events. Figure 1b shows local event slopes measured from the data using the plane-wave destruction algorithm of (13). Plane-wave destruction (6) works by making a prediction of each seismic trace from the neighboring trace along local slopes and then minimizing the prediction error by an iterative regularized least-squares optimization. Regularization controls smoothness of the estimated slope field. In this work, I use shaping regularization (16) for an optimal smoothness control. Figure 1c shows the output of oriented NMO using equation 3. As expected, all events are perfectly flattened after NMO.

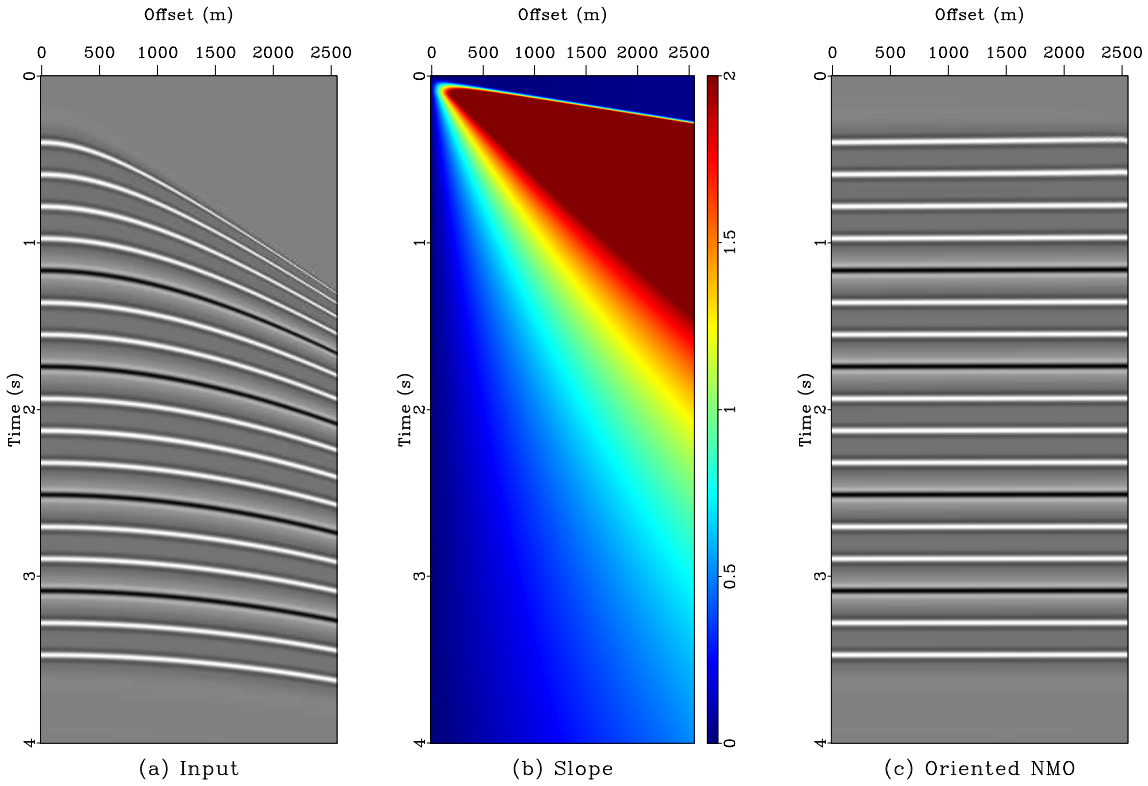


Figure 1: a: A synthetic CMP gather composed of hyperbolas. b: Estimated local event slopes. c: The output of oriented velocity-independent NMO. pmig/nmo synt

In conventional NMO processing, one scans a number of velocities, performs the corresponding number of moveout corrections, and picks the velocity trend from velocity spectra. In oriented processing, velocity becomes, according to equation 4, a data attribute rather than a prerequisite for imaging. Figure 2 shows a comparison between velocity spectra used for picking velocities in the conventional NMO processing and the velocity attribute mapped directly from the data space using equations 3

and 4. A noticeably higher resolution of the oriented map follows from the fact that only the true signal slopes are identified by the slope estimation algorithm. The slope uncertainty, unlike the velocity uncertainty, is not taken into account.

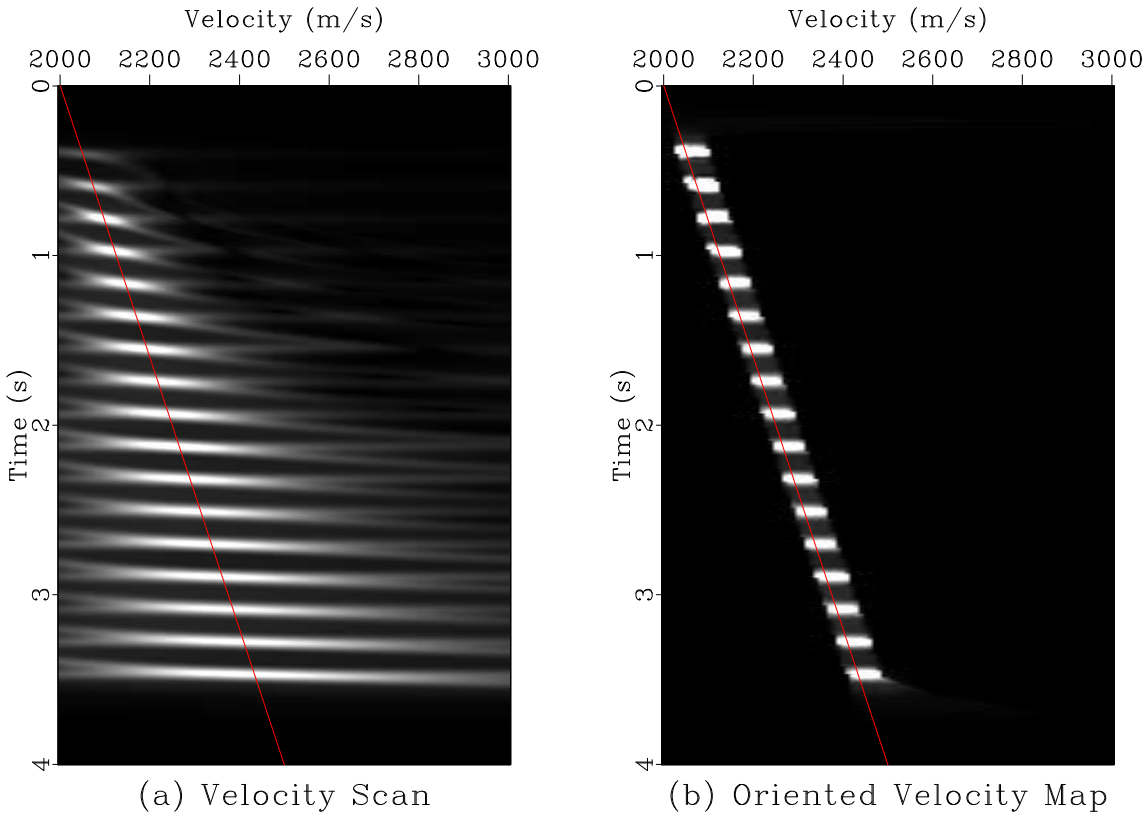


Figure 2: a: Velocity spectra using conventional velocity scanning. In the conventional NMO processing, the velocity trend is picked from velocity spectra prior to NMO. b: Velocity mapped from the data space using oriented NMO. The input for both plots is the synthetic CMP gather shown in Figure 1a. The red line indicates the exact velocity used for generating the synthetic data. pmig/nmo svsc

Figure 3 shows a field data example. The data are taken from a historic Gulf of Mexico dataset (7). Analogously to the synthetic case, a CMP gather (Figure 3a) is properly flattened (Figure 3c) by an application of oriented NMO using the local event slopes (Figure 3b) measured directly from the data. A comparison between conventional velocity spectra and the velocity mapping with oriented NMO is shown in Figure 4. Again, a significantly higher resolution is observed.

τ-p NMO

One can choose to perform normal moveout in the slant-stack (τ-p) domain rather than the original t-x domain. In the τ-p domain, hyperbolas turn into ellipses but the

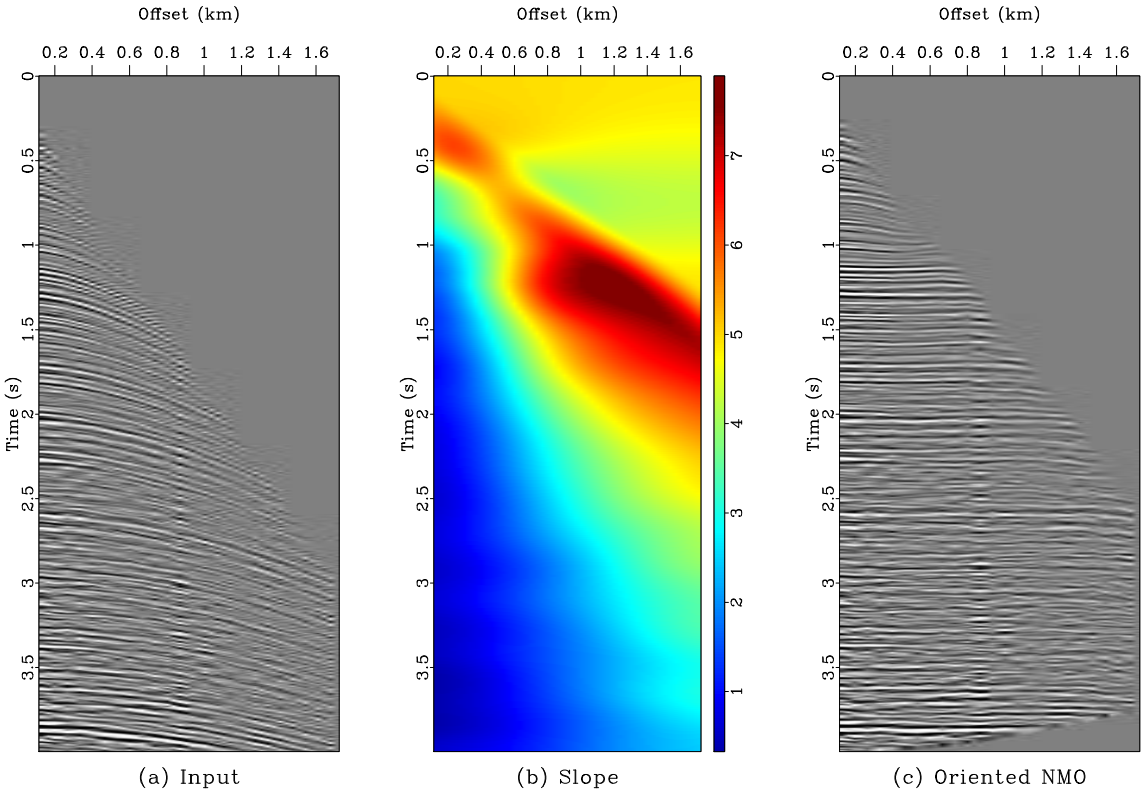


Figure 3: a: A CMP gather from the Gulf of Mexico. b: estimated local event slopes. c: the output of oriented velocity-independent NMO. pmig/mig pnmo

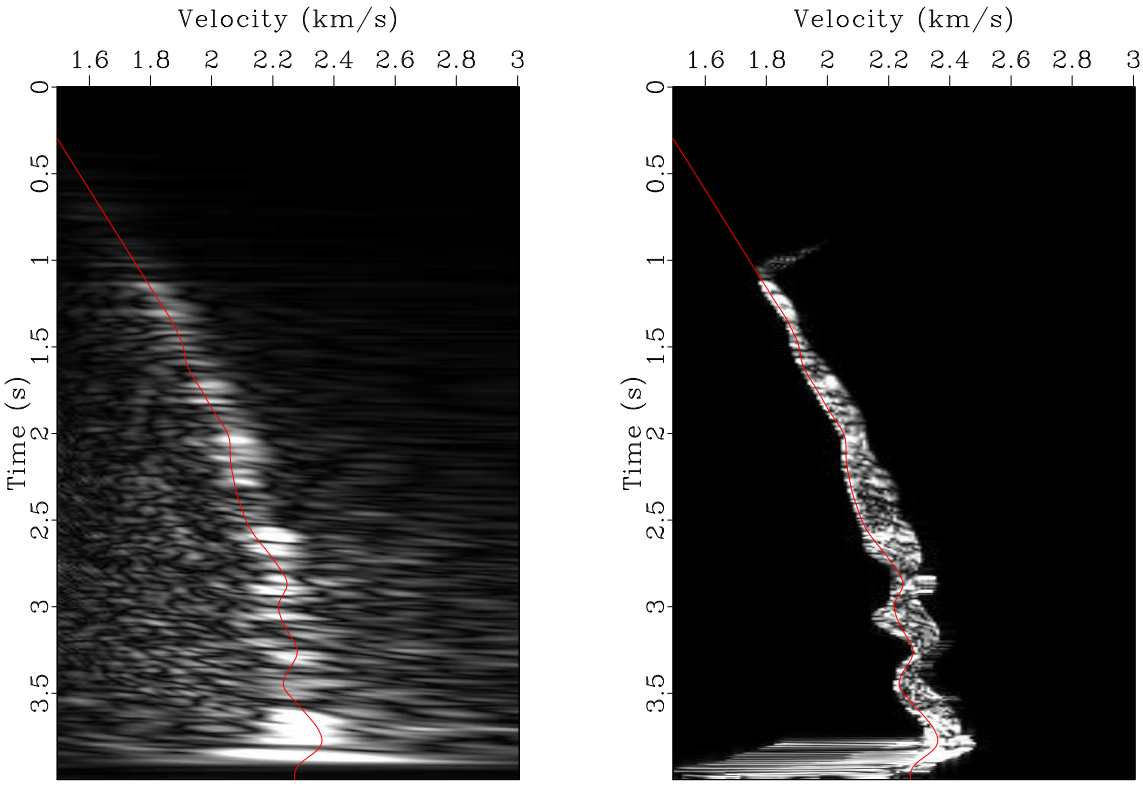


Figure 4: a: Velocity spectra using conventional velocity scanning. In the conventional NMO processing, the velocity trend is picked from velocity spectra prior to NMO. b: Velocity mapped from the data space using oriented NMO. The input for both plots is the CMP gather shown in Figure 3a. The red curve is an automatic pick of the velocity trend shown for comparison. pmig/mig bvsc

moveout correction problem remains (34). Differentiating the τ-p moveout equation

$$\tau(p) = \tau_0 \sqrt{1 - p^2\, v_n^2(\tau_0)} \tag{5}$$

leads to

$$r = \frac{d\tau}{dp} = -\frac{p\, v_n^2(\tau_0)\, \tau_0^2}{\tau} \ , \tag{6}$$

which resolves in the oriented velocity-free moveout equation analogous to equation 3

$$\tau_0 = \sqrt{\tau^2 - \tau\, r(\tau, p)\, p} \ . \tag{7}$$

Combining equations 5 and 7 produces the corresponding velocity mapping equation

$$v_n^2 = \frac{r(\tau, p)}{p^2\, r(\tau, p) - p\, \tau} \ . \tag{8}$$

Non-hyperbolic moveout

The hyperbolic model 1 is not accurate at large offsets in the case of non-hyperbolic moveouts, caused by vertical or lateral heterogeneity, reflector curvature, or anisotropy (17). One popular model for describing non-hyperbolic moveouts was developed by (27) and has the form of a shifted hyperbola (9; 5; 32)

$$t(l) = t_0 \left(1 - \frac{1}{S(t_0)}\right) + \frac{1}{S(t_0)} \sqrt{t_0^2 + S(t_0)\, \frac{l^2}{v_n^2(t_0)}} \ , \tag{9}$$

Equation 9 contains an additional parameter S, which is related to heterogeneity and anisotropy of seismic velocities. We can eliminate this parameter by differentiating the equation twice and defining the second derivative $q = \partial p/\partial l = \partial^2 t/\partial l^2$. Eliminating both v_n and S from equation 9 and equations

$$p = \frac{l}{[t_0 + S(t_0)\ (t - t_0)]\ v_n^2} \ , \tag{10}$$

$$q = \frac{t_0^2}{[t_0 + S(t_0)\ (t - t_0)]^3\ v_n^2} \tag{11}$$

leads to the velocity-independent non-hyperbolic moveout equation

$$t_0 = t - \frac{p\, l}{1 + \sqrt{\frac{q\, l}{p}}} \ . \tag{12}$$

If moveout parameters v_n and S are required for subsequent interpretation, one can easily extract them as special data attributes

$$\frac{1}{v_n^2} = t_0 \sqrt{\frac{p^3}{q\, l^3}} \ , \tag{13}$$

$$S = 1 + \frac{p\,(t - p\, l) - q\, l\, t}{\sqrt{q\, p^3\, l^3}} \ . \tag{14}$$

One could estimate the function $q(t, l)$ in practice by numerically differentiating the local slope field $p(t, l)$.

Dix inversion

Not only moveout velocities but also interval velocities can be estimated by using the local slope information. Employing the Dix inversion approach (10) and defining $p_t = \partial p/\partial t$, one can deduce from equations 3 and 4 an expression for the interval velocity v_i, which becomes another attribute directly mappable from the data, as follows:

$$v_i^2 = \frac{l}{p^2\, t}\, \frac{p\, l(p + t\, p_t) - 2 p_t\, t^2}{2\, t - l\, (p + t\, p_t)} \; . \tag{15}$$

Equation 15 shows that then the interval velocity can be regarded as another attribute mappable directly from the data. The derivation is detailed in Appendix A.

Figure 5 shows the interval velocity mapping according to equation 15 for the synthetic data example shown in Figure 1a. The exact interval velocity profile (red curve) is recovered perfectly. An analogous example for the field dataset from Figure 3 is shown in Figure 6. The field data result is noisy because of the instability of numerical differentiation but clearly shows the overall range of interval velocities.

Equation 15 enables direct mapping from data slope attributes into interval velocities. Thus, it provides an analytical solution to the stereotomography problem (4) for the special case of horizontal reflection layers and vertically variable velocities.

Migration to zero offset

NMO correction accomplishes mapping to zero offset in the case of horizontal reflectors. Taking account of the reflector dip effect requires dip moveout (18). Combined with NMO, dip moveout maps the input data in time-offset-midpoint coordinates $\{t, h, y\}$ to time-midpoint coordinates $\{t_0, y_0\}$ at the zero-offset section. The corresponding oriented mapping, derived in Appendix B, is

$$t_0^2 \;=\; t\, \frac{\left[(t - h\, p_h)^2 - h^2\, p_y^2\right]^2}{(t - h\, p_h)^3} \; , \tag{16}$$

$$y_0 \;=\; y - \frac{h^2\, p_y}{t - h\, p_h} \; , \tag{17}$$

where $h = l/2$, and $p_h = \partial t/\partial h$ and $p_y = \partial t/\partial y$ are prestack slopes.

The full 2-D dataset and the estimated prestack slopes are shown in Figures 6(a) and 8, respectively. The output of migration to zero offset and stack is shown in Figures 9 (before stack) and 10 (after stack). All major events are properly transformed to the appropriate zero-offset positions.

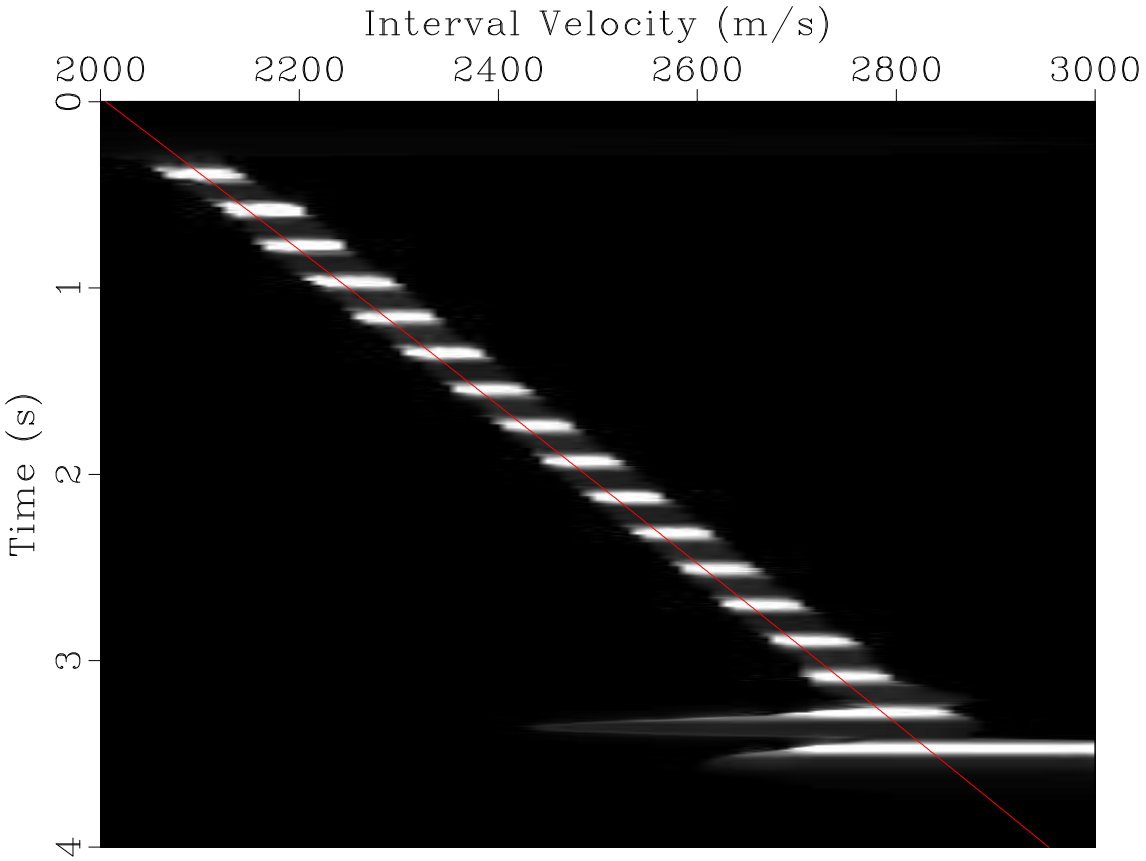


Figure 5: Oriented mapping to interval velocity for the synthetic data set shown in Figure 1a. The exact interval velocity profile is shown by a red curve. pmig/nmo svin

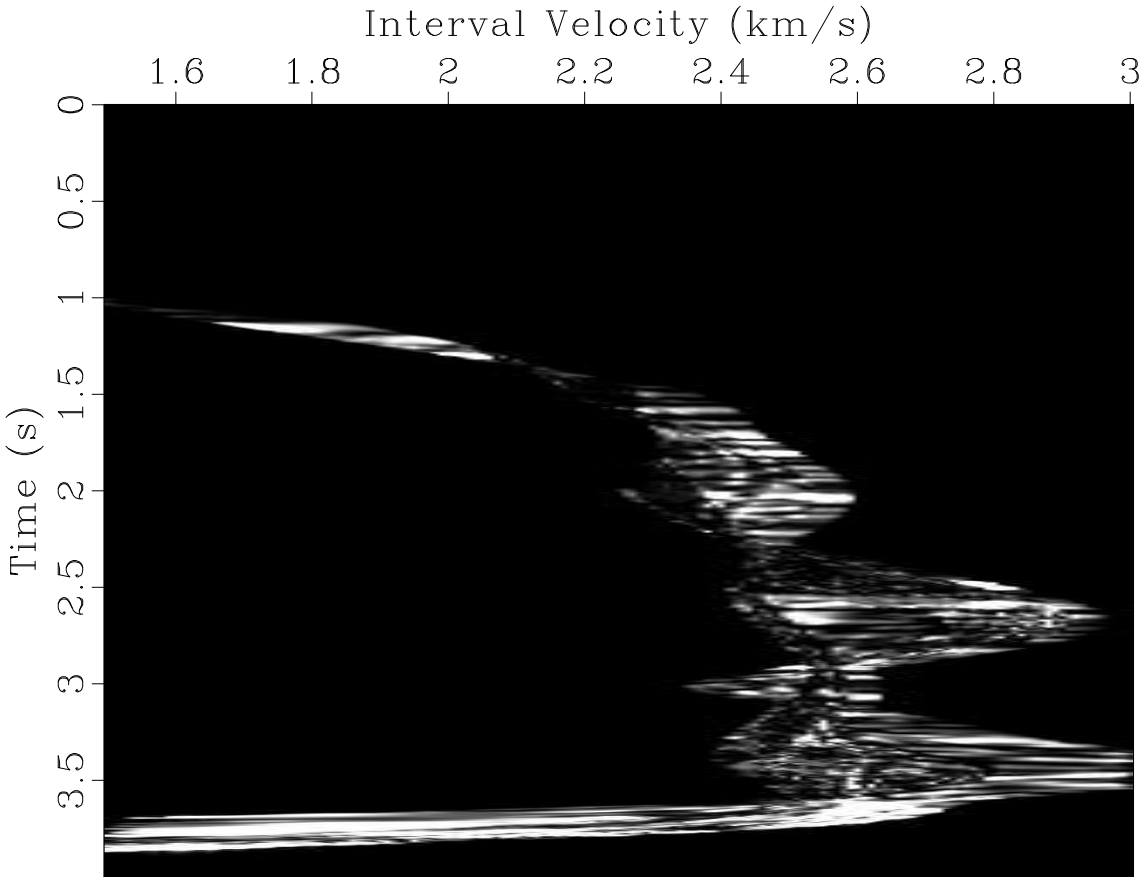


Figure 6: Oriented mapping to interval velocity for the field data set shown in Figure 3a. pmig/mig bvint

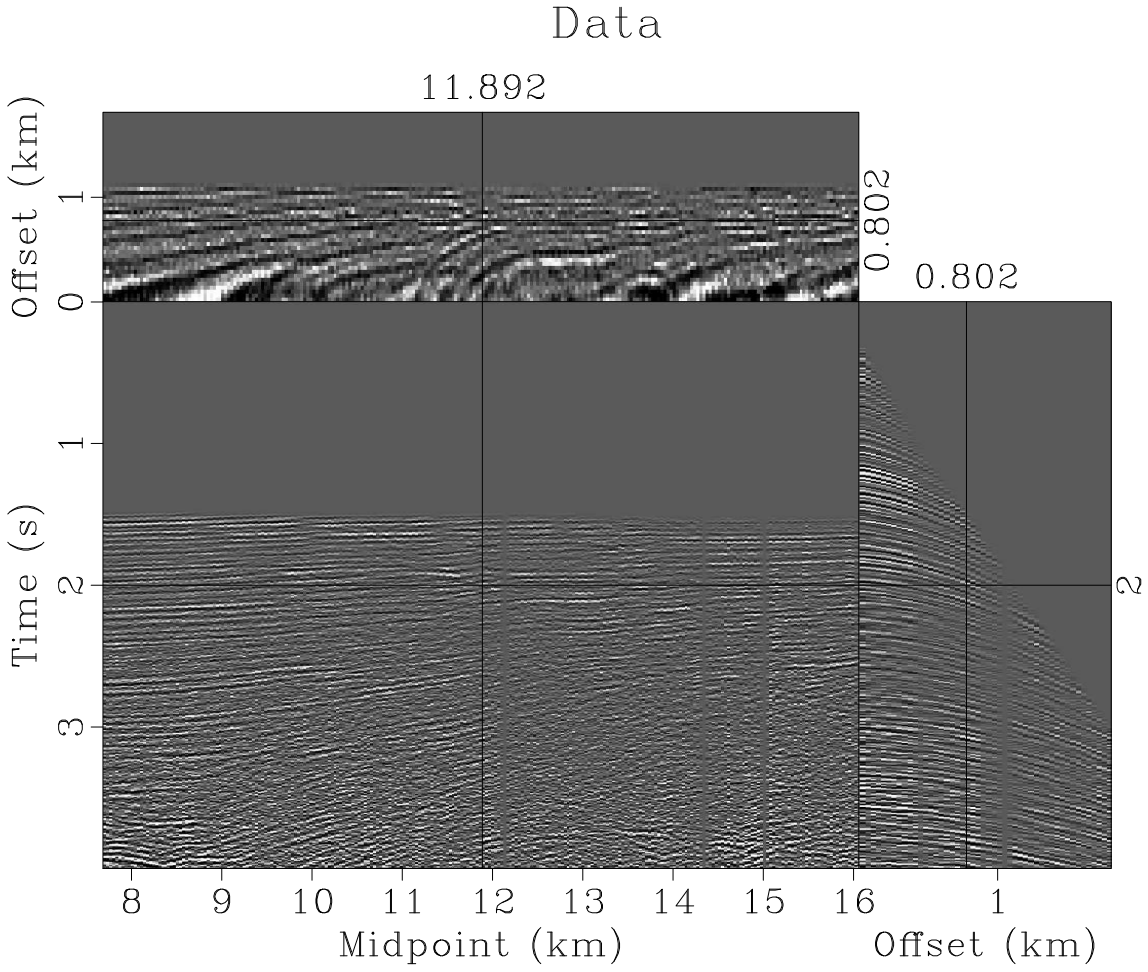


Figure 7: A field 2-D dataset from the Gulf of Mexico used for numerical experiments. The display in this and other 3-D figures is composed of three sections from the cube indicated by vertical and horizontal lines. pmig/mig bei

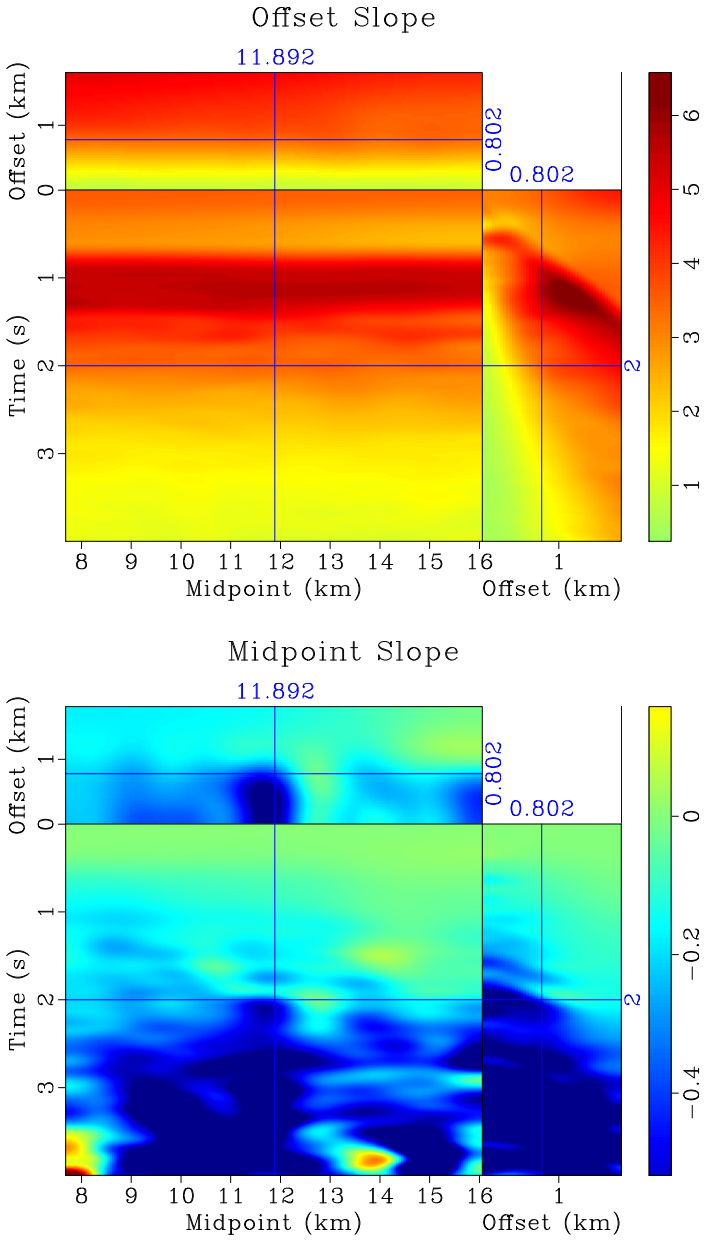


Figure 8: Prestack data slopes estimated from the dataset. a: Offset slope p_h. b: Midpoint slope p_y. pmig/mig hdip,ydip

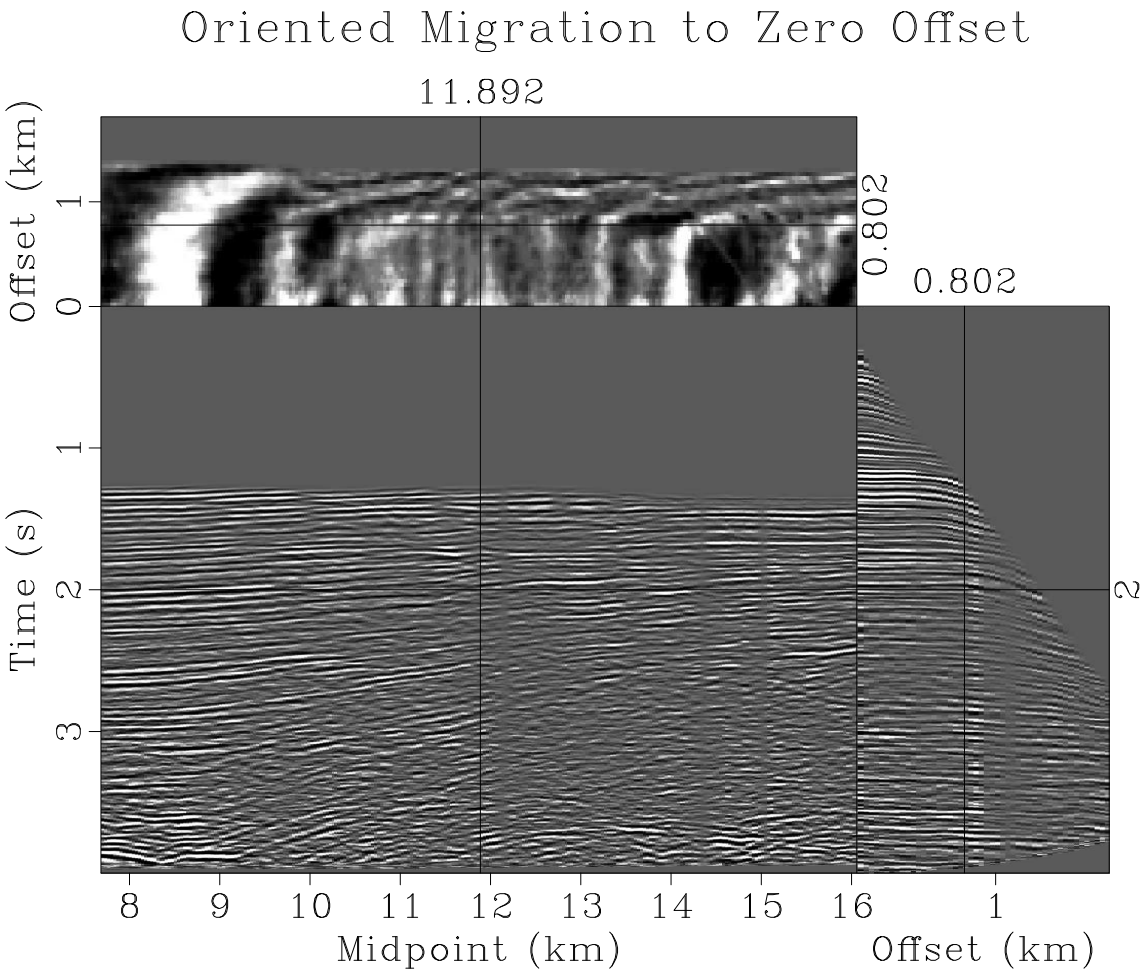


Figure 9: Output of oriented migration to zero offset. pmig/mig bei-mzo

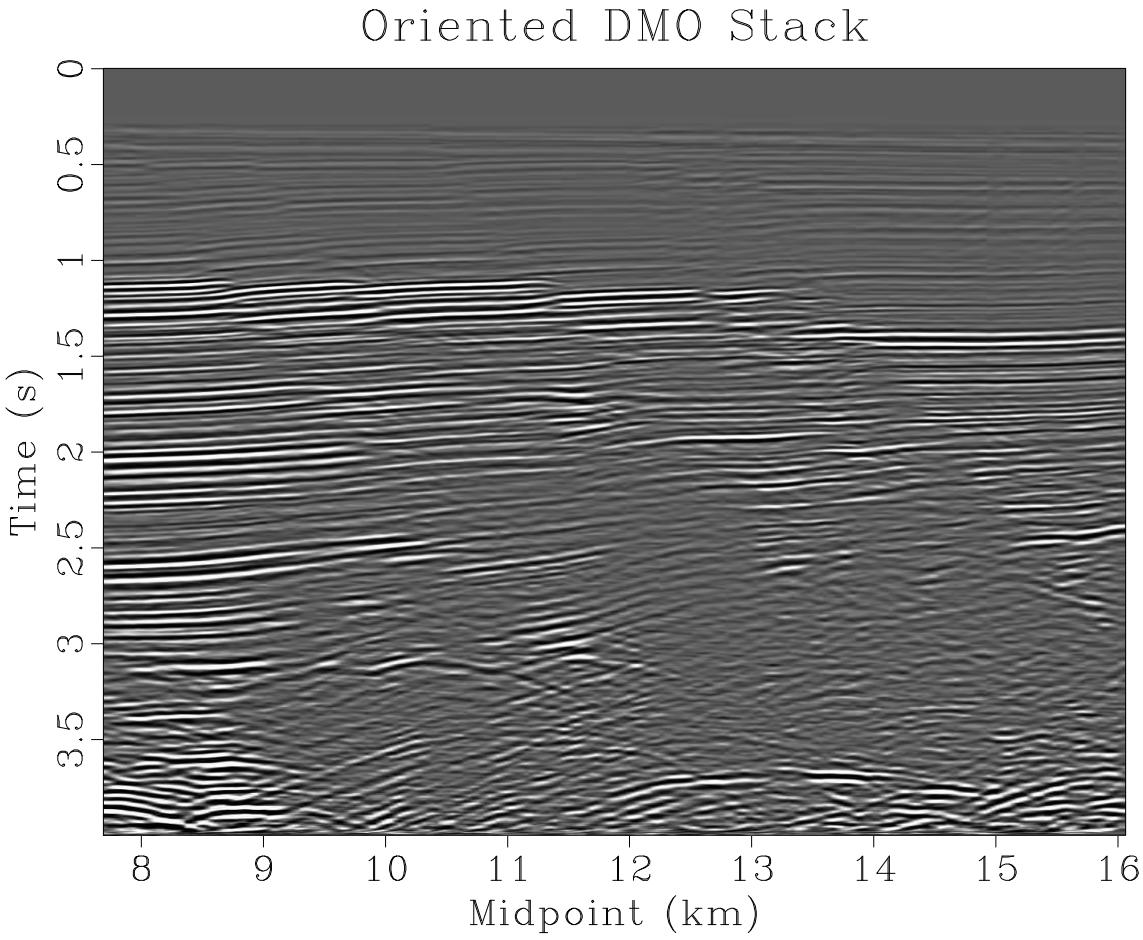


Figure 10: Seismic stack obtained by oriented migration to zero offset. pmig/mig bei-mzo2

Oriented prestack time migration

Mapping prestack data to the reflector position expressed in vertical traveltime is the task of prestack time migration. Analogously to dip moveout, prestack migration moves events not only across offsets but also across different midpoints provided that the reflectors are not flat. Using the oriented approach, mapping from the prestack domain $\{t, h, y\}$ to the time-migrated image domain $\{\tau, x\}$ is defined by the following equations derived in Appendix B:

$$\tau^2 = \frac{t\, p_h \, \left[(t - h\, p_h)^2 - h^2\, p_y^2\right]^2}{(t - h\, p_h)^2 \, \left[t\, p_h + h\, (p_y^2 - p_h^2)\right]} \,, \tag{18}$$

$$x = y - \frac{h\, t\, p_y}{t\, p_h + h\, (p_y^2 - p_h^2)} \,. \tag{19}$$

Here $\{\tau, x\}$ are the time-migrated image coordinates that correspond to the vertical ray traveltime and location (22).

Under the assumption of a hyperbolic diffraction moveout, the migration velocity becomes, analogously to equation 4, a data attribute completely defined by prestack slopes, as follows:

$$\frac{4}{v^2} = \frac{t \, \left[t\, p_h + h\, (p_y^2 - p_h^2)\right]}{h\, (t - h\, p_h)} \,. \tag{20}$$

Equations 18 and 20 transform to equations 3 and 4 when $p_y = 0$, which corresponds to horizontal reflectors. Equations analogous to 18, 19, and 20 were derived previously by (3) in a different way and in the depth imaging context.

Figures 11 and 12 show the output of prestack migration before and after stack. The result is comparable to that of conventional processing (7) but obtained two orders of magnitude faster, since each data point transforms directly to the image space through a one-to-one mapping instead of being spread along a wide migration impulse response, as in conventional prestack time migration.

DISCUSSION

The main advantage of the oriented approach is speed. The cost of velocity scanning in conventional processing (excluding the manual picking labor involved) can be estimated roughly as the number of scanned velocities N_v times the input data size. The cost increases dramatically in the case of non-hyperbolic approximations when more than one parameter needs to be picked. The cost of local slope estimation with the plane-wave destruction method is roughly the data size times the number of iterations N_i times the filter size N_f. Typically, $N_i \approx 10$ and $N_f = 6$, which is approximately equivalent in cost to scanning $N_v = 60$ velocities. The next step, however, is dramatically different. Since each data point is mapped directly to the image instead of being spread into a wide impulse response, we save the factor in cost proportional

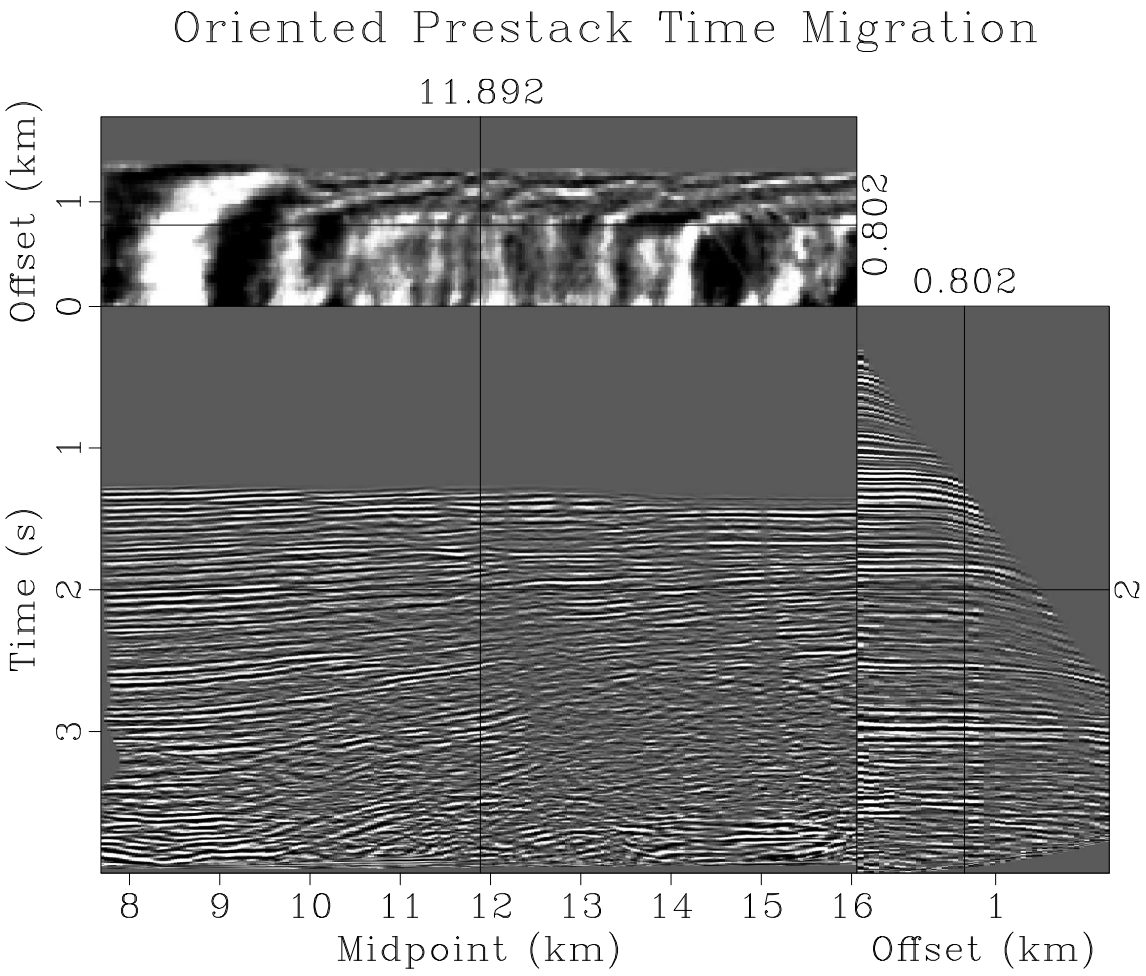


Figure 11: Output of oriented prestack time migration. pmig/mig bei-mig

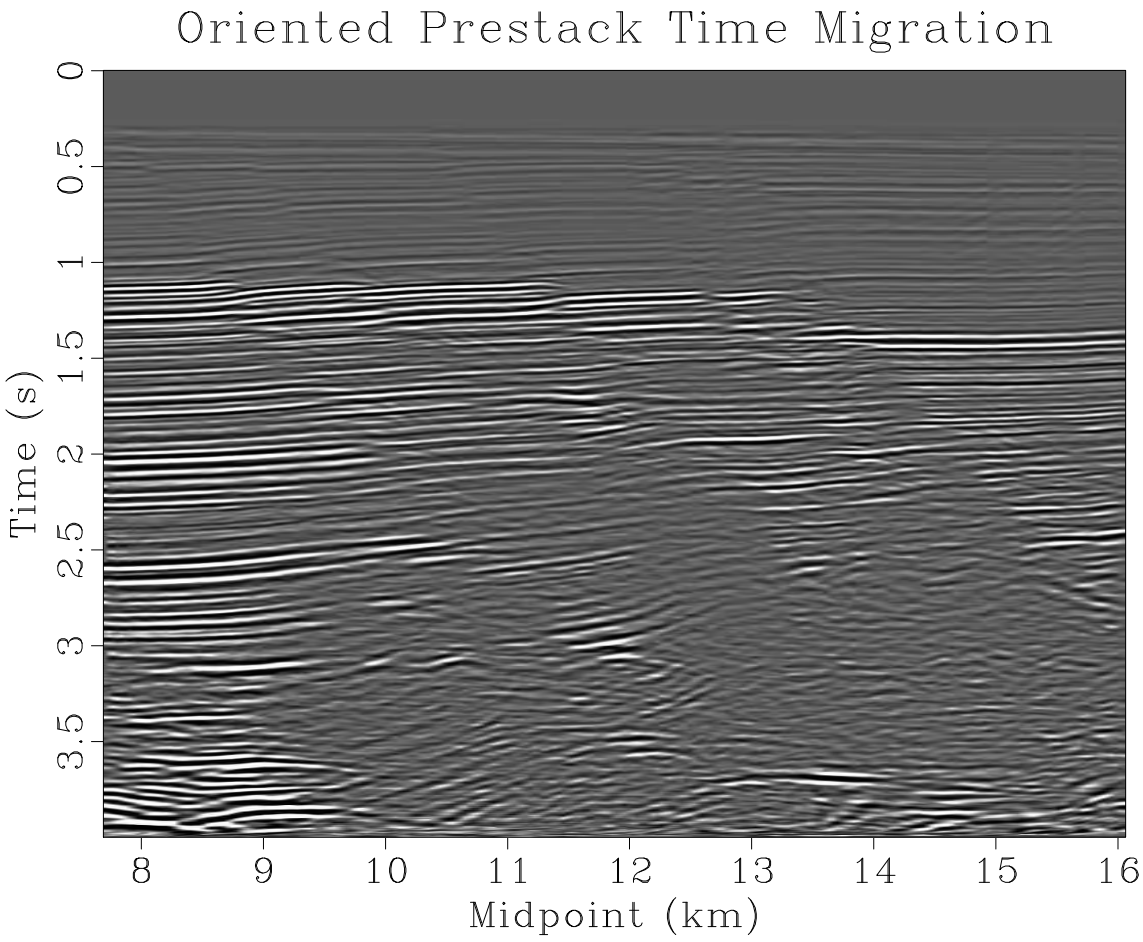


Figure 12: Seismic image obtained by oriented prestack time migration. pmig/mig bei-mig2

to the size (in samples) of the migration impulse response. The prestack migration result shown in Figure 11 was accomplished in under 10 seconds on a single-node PC.

The cost savings will be reduced somewhat if we take into account more than one local slope (crossing reflection events, diffractions, multiple reflections, etc.) The plane-wave destruction algorithm (13) can be applied for estimating several interfering data slopes simultaneously. In order to take full advantage of it, data decomposition into local slope components may be required. The curvelet transform (19; 12) suggests a possible data decomposition approach. To extend the method of curvelet imaging developed by (11), each local slope component would need to be imaged separately by an oriented approach with its contribution stacked into the final image.

Seismic imaging and velocity estimation is inherently an uncertain process because of limitations in the data acquisition geometry and signal bandwidth. In the oriented approach, the uncertainty in the velocity estimation and in the positioning of seismic reflectors comes directly from the uncertainty in estimating local event slopes. Such uncertainty is much easier to estimate and analyze in the oriented rather than in the traditional approach thanks to the explicit time-domain imaging equations 18 and 19 that transform uncertainties in the local event slopes p_h and p_y directly into uncertainties of the image point positioning.

CONCLUSIONS

Local slopes of seismic events contain complete information about the reflection geometry. Once they are estimated, one can turn seismic velocities and all other moveout parameters into data attributes directly mappable from the prestack data domain into the time-migrated image domain. In this paper, I have developed the analytical theory for this transformation and demonstrated its applicability with examples.

Extensions of oriented imaging from time to depth migration are provided in parsimonious migration (21) using a ray tracing formalism and in the theory of the oriented wave equation (14) using a wave equation formalism.

ACKNOWLEDGMENTS

I would like to thank Jon Claerbout and Antoine Guitton for inspiring discussions. Huub Douma, Gilles Lambaré, Isabelle Lecomte, and one anonymous reviewer provided thorough and helpful reviews.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

MATHEMATICAL DERIVATION OF ORIENTED DIX INVERSION

The famous Dix inversion formula (10) can be written in the form

$$v_i^2 = \frac{d}{d\,t_0}\,\left[t_0\,v^2(t_0)\right]\ , \tag{A-1}$$

where v_i is the interval velocity corresponding to the zero-offset traveltime t_0 and $v(t_0)$ is the vertically-variable root-mean-square velocity. By a straightforward application of the chain rule, I rewrite the Dix equation in the form

$$v_i^2 = \frac{d\,\left[t_0(t)\,v^2(t)\right]/d\,t}{d\,t_0/d\,t}. \tag{A-2}$$

Substituting $t_0(t)$ and $v(t)$ dependences from equations 3 and 4 and doing algebraic simplifications yields

$$\frac{d\,\left[t_0(t)\,v^2(t)\right]}{d\,t} = \frac{l}{p^2\,t}\,\frac{p\,l(p+t\,p_t)-2p_t\,t^2}{2\,t_0}\ , \tag{A-3}$$

$$\frac{d\,t_0}{d\,t} = \frac{2\,t-l\,(p+t\,p_t)}{2\,t_0}\ , \tag{A-4}$$

where $p_t = \partial p/\partial t$. Substituting equations A-3 and A-4 into A-2 produces equation 15 in the main text.

APPENDIX B

MATHEMATICAL DERIVATION OF ORIENTED TIME-DOMAIN IMAGING OPERATORS

The mathematical derivation of oriented time-domain imaging operators follows geometrical principles. Consider the reflection ray geometry in Figure B-1. Making a hyperbolic approximation of diffraction traveltimes used in seismic time migration is equivalent to assuming an effective constant-velocity medium and straight-ray geometry. The geometrical connection between the effective dip angle α, the effective reflection angle θ, the effective velocity v, half-offset h, and the reflection traveltime t is given by the equation

$$t = \frac{2\,h}{v}\,\frac{\cos\alpha}{\sin\theta}\ , \tag{B-1}$$

which follows directly from the trigonometry of the reflection triangle (8; 15). Additionally, the two angles are connected with the traveltime derivatives $p_h = \partial t/\partial h$ and

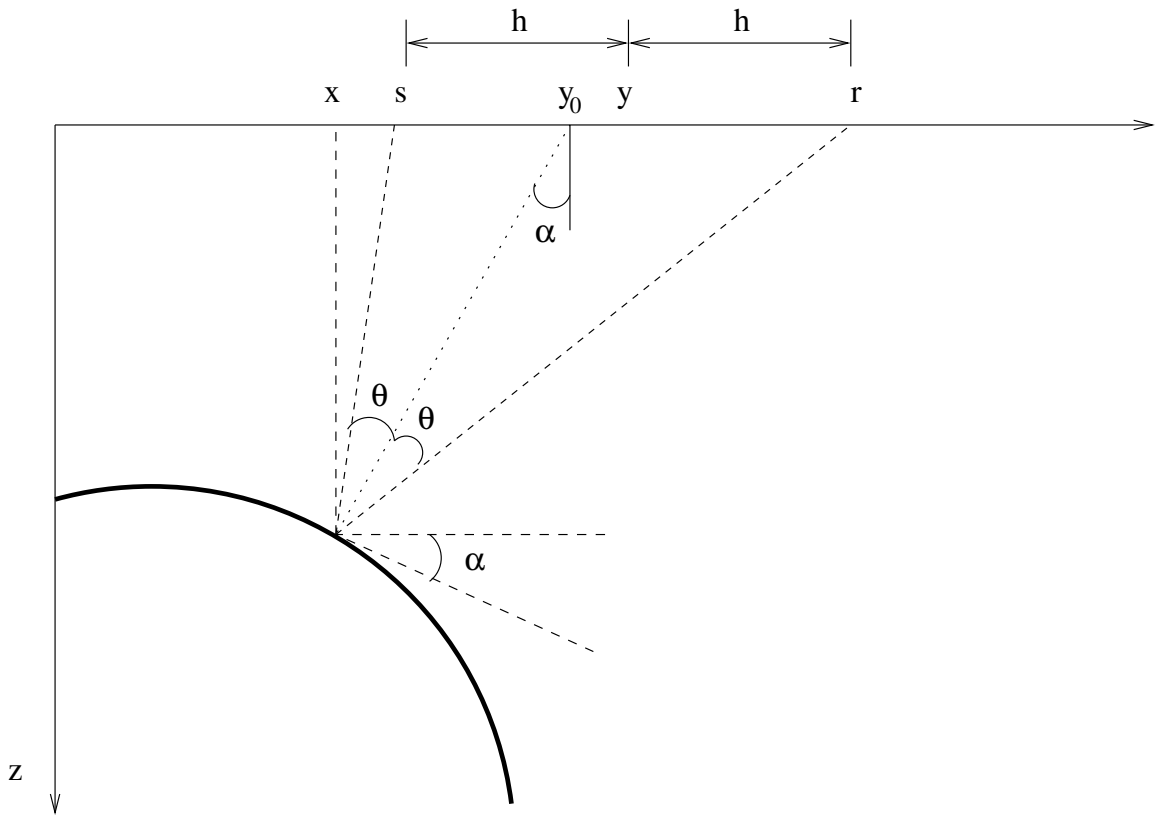


Figure B-1: Reflection ray geometry in an effectively homogeneous medium (a scheme). pmig/XFig rays

$p_y = \partial t/\partial y$ according to equations

$$p_h = \frac{2\,\cos\alpha\,\sin\theta}{v}\,, \tag{B-2}$$

$$p_y = \frac{2\,\sin\alpha\,\cos\theta}{v}\,. \tag{B-3}$$

Using equations B-1, B-2, and 7, one can explicitly solve for the effective parameters α, θ, and v expressing them in terms of the data coordinates t and h and event slopes p_h and p_y. The solution takes the form

$$\tan^2\alpha = \frac{h\,p_y^2}{p_h\,(t - h\,p_h)}\,, \tag{B-4}$$

$$\sin^2\theta = \frac{h\,p_h}{t}\,, \tag{B-5}$$

$$v^2 = \frac{4\,h\,(t - h\,p_h)}{t\,\left[t\,p_h + h\,(p_y^2 - p_h^2)\right]}\,. \tag{B-6}$$

Note that equation A-9 is equivalent to equation 20 in the main text. It reduces to equation 4 in the case of a horizontal reflector ($p_y = 0$).

With the help of equations A-8, B-5, and A-9, one can transform all other geometrical quantities associated with time-domain imaging into data attributes. The vertical two-way time is (31)

$$\tau = t\,\frac{\cos^2\alpha - \sin^2\theta}{\cos\alpha\,\cos\theta}\,, \tag{B-7}$$

which turns, after substituting equations A-8 and B-5, into equation 18 in the main text. The separation between the midpoint and the vertical is (31)

$$y - x = h\,\frac{\sin\alpha\,\cos\alpha}{\sin\theta\,\cos\theta}\,, \tag{B-8}$$

which turns, after substituting equations A-8 and B-5, into equation 19 in the main text. Additionally, the zero-offset traveltime is (using equation B-7)

$$t_0 = \frac{\tau}{\cos\alpha} = t\,\frac{\cos^2\alpha - \sin^2\theta}{\cos^2\alpha\,\cos\theta}\,, \tag{B-9}$$

which turns into equation 16. Finally, the separation between the midpoint and the zero-offset point is (using equations B-1 and B-8)

$$y - y_0 = y - x - \frac{v\,\tau}{2}\,\tan\alpha = h\,\tan\alpha\,\tan\theta\,, \tag{B-10}$$

which turns into equation 17. In the case of a horizontal reflector ($\alpha = 0$), $y = y_0 = x$, $t_0 = \tau$, and the zero-offset traveltime reduces to the NMO-corrected traveltime in equation 3.

Non-hyperbolic and three-dimensional generalizations of this theory are possible.

REFERENCES

[1] Adler, F., and S. Brandwood, 1999, Robust estimation of dense 3-D stacking velocities from automated picking: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1162–1165.

[2] Baina, R., S. Nguyen, M. Noble, and G. Lambare, 2003, Optimal antialiasing for ray-based depth Kirchhoff migration: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 1130–1133.

[3] Billette, F., S. L. Begat, P. Podvin, and G. Lambare, 2003, Practical aspects and applications of 2D sterotomography: Geophysics, **68**, 1008–1021.

[4] Billette, F., and G. Lambaré, 1998, Velocity macromodel estimation from seismic reflection data by stereotomography: Geoph. J. Internat., **135**, 671–680.

[5] Castle, R. J., 1994, Theory of normal moveout: Geophysics, **59**, 983–999.

[6] Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.

[7] ———, 2005, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

[8] Clayton, R. W., 1978, Common midpoint migration, *in* SEP-14: Stanford Exploration Project, 21–36.

[9] de Bazelaire, E., 1988, Normal moveout revisited - Inhomogeneous media and curved interfaces: Geophysics, **53**, 143–157.

[10] Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68–86.

[11] Douma, H., 2006, A hybrid formulation of map migration and wave-equation-based migration using curvelets: PhD thesis, Colorado School of Mines.

[12] Douma, H., and M. de Hoop, 2006, Leading-order seismic imaging using curvelets, *in* 76th Ann. Internat. Mtg.: Soc. of Expl. Geophys, accepted.

[13] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

[14] ———, 2003a, Angle-domain seismic imaging and the oriented wave equation: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 893–898.

[15] ———, 2003b, Velocity continuation and the anatomy of residual prestack time migration: Geophysics, **68**, 1650–1661.

[16] ———, 2007, Shaping regularization in geophysical estimation problems: Geophysics, accepted for publication.

[17] Fomel, S., and V. Grechka, 2001, Nonhyperbolic reflection moveout of P waves. An overview and comparison of reasons, *in* CWP-372: Colorado School of Mines.

[18] Hale, D., 1995, DMO processing: Soc. of Expl. Geophys.

[19] Herrmann, F., 2003, Optimal seismic imaging with curvelets: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 997–1000.

[20] Hertweck, T., C. Jäger, J. Mann, E. Duveneck, and Z. Heilmann, 2004, A seismic reflection imaging workflow based on the Common-Reflection-Surface (CRS) stack: theoretical background and case study, *in* 74th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2032–2035.

[21] Hua, B., and G. A. McMechan, 2003, Parsimonious 2D prestack Kirchhoff dept

migration: Geophysics, **68**, 1043–1051.
[22] Hubral, P., 1977, Time migration - Some ray theoretical aspects: Geophys. Prosp., **25**, 738–745.
[23] Lambaré, G., 2004, Stereotomography: Where we are, and where we should go: 74th Ann. Internat. Mtg., Soc. of Expl. Geophys., 2367–2370.
[24] Lambaré, G., M. Alerini, R. Baina, and P. Podvin, 2004a, Stereotomography: a semi-automatic approach for velocity macromodel estimation: Geophysical Prospecting, **52**, 671–681.
[25] Lambaré, G., M. Alerini, and P. Podvin, 2004b, Stereotomographic picking in practice, *in* 74th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2343–2346.
[26] Landa, E., B. Gurevich, S. Keydar, and P. Trachtman, 1999, Application of multifocusing for subsurface imaging: Applied Geophysics, **42**, 283–300.
[27] Malovichko, A. A., 1978, A new representation of the traveltime curve of reflected waves in horizontally layered media: Applied Geophysics (in Russian), the translation provided by (35), **91**, 47–53.
[28] Ottolini, R., 1983, Velocity independent seismic imaging, *in* SEP-37: Stanford Exploration Project, 59–68.
[29] Riabinkin, L. A., 1957, Fundamentals of resolving power of controlled directional reception (CDR) of seismic waves, *in* Slant-stack processing, 1991: Soc. of Expl. Geophys. (Translated and paraphrased from *Prikladnaya Geofiizika*, **16**, 3-36), 36–60.
[30] Rieber, F., 1936, A new reflection system with controlled directional sensitivity: Geophysics, **01**, 97–106.
[31] Sava, P. C., and S. Fomel, 2003, Angle-domain common-image gathers by wavefield continuation methods: Geophysics, **68**, 1065–1074.
[32] Siliqi, R., and N. Bousquié, 2000, Anelliptic time processing based on a shifted hyperbola approach: 70th Ann. Internat. Mtg, Soc. of Expl. Geophys., 2245–2248.
[33] Siliqi, R., D. L. Meur, F. Gamar, L. Smith, J. Toure, and P. Herrmann, 2003, High-density moveout parameter fields V and Eta, Part 1: Simultaneous automatic picking: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 2088–2091.
[34] Stoffa, P. L., P. Buhl, J. B. Diebold, and F. Wenzel, 1981, Direct mapping of seismic data to the domain of intercept time and ray parameter - A plane-wave decomposition: Geophysics, **46**, 255–267.
[35] Sword, C. H., 1987a, A Soviet look at datum shift, *in* SEP-51: Stanford Exploration Project, 313–316.
[36] ———, 1987b, Tomographic determination of interval velocities from reflection seismic data: The method of controlled directional reception: PhD thesis, Stanford University.
[37] Wolf, K., D. Rosales, A. Guitton, and J. Claerbout, 2004, Robust moveout without velocity picking, *in* 74th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2423–2426.
[38] Yilmaz, O., 2000, Seismic data analysis: Soc. of Expl. Geophys.

Random noise attenuation using local signal-and-noise orthogonalization

*Yangkang Chen and Sergey Fomel**

ABSTRACT

We propose a novel approach to attenuate random noise based on local signal-and-noise orthogonalization. In this approach, we first remove noise using one of the conventional denoising operators, and then apply a weighting operator to the initially denoised section in order to predict the signal-leakage energy and retrieve it from the initial noise section. The weighting operator is obtained by solving a least-squares minimization problem via shaping regularization with a smoothness constraint. Next, the initially denoised section and the retrieved signal are combined to form the final denoised section. The proposed denoising approach corresponds to orthogonalizing the initially denoised signal and noise in a local manner. We evaluate denoising performance by using local similarity. In order to test the orthogonalization property of the estimated signal and noise, we calculate the local similarity map between the denoised signal section and removed noise section. Low values of local similarity indicate a good orthogonalization and thus a good denoising performance. Synthetic and field data examples demonstrate the effectiveness of the proposed approach in applications to noise attenuation for both conventional and simultaneous-source seismic data.

INTRODUCTION

Most random noise attenuation techniques necessitate a control of the trade-off between preservation of useful signal and removal of random noise. F-x deconvolution and projection (5; 26), also known as f-x predictive filtering (3; 8), is a common technique for random noise attenuation thanks to its convenient implementation and high efficiency. Commonly a prediction filter length is modified in order to increase or decrease the ability of f-x deconvolution to remove random noise, at the expense of decreasing or increasing the preservation of useful signal. When the data character is complex, the filter length should be long enough to capture certain amounts of useful signal, but at the same time to attenuate a limited amount of noise. Otherwise, a portion of useful signal can be lost because of a large number of dip components in the data (8). For certain specific types of random noise, such as blending noise in simultaneous-source seismic data (2; 4; 1; 7), median filtering (MF) can be particularly effective (19; 6; 9). However, the ability of MF to remove spike-like blending

*e-mail: sergey.fomel@beg.utexas.edu

noise strongly depends on the length of the filtering window. A longer filter leads to a more powerful attenuation for blending noise, but at the same time causes a possible loss of useful energy. In contrast, when the length of the filtering window is small, MF may eliminate less useful energy but also will be less able to remove spiky noise. Removing spiky noise without losing too much useful signal becomes a difficult trade-off.

Suppose that the seismic data $\mathbf{d}$ is composed of signal $\mathbf{s}_{true}$ and noise $\mathbf{n}_{true}$, a successful random noise attenuation will give accurate signal and noise estimation: $\mathbf{s}_0 \approx \mathbf{s}_{true}$ and $\mathbf{n}_0 \approx \mathbf{n}_{true}$, where $\mathbf{s}_0$ and $\mathbf{n}_0$ are the estimated signal and noise. However, the ideal situation may not occur in practice for two main possible reasons: incorrect parameter selection or inadequacy of denoising assumptions. When those assumptions are not met, the conventional denoising approaches may not achieve the optimal performance. For most noise attenuation approaches, the noise section will contain a certain amount of useful signal, which can be called *signal-leakage energy*. Sometimes, the leakage energy is negligible, whereas in other cases, this loss of energy may result in a decrease in resolution.

Local seismic attributes measure seismic signal characteristics neither instantaneously, at each signal point, nor globally, across a data window, but locally in the neighborhood of each point (10). One of the most useful local attributes is *local similarity*, which has found numerous successful applications in different areas of seismic data processing: multicomponent image registration (14; 10), time-lapse registration (15; 27), time-frequency analysis (22), structure-enhancing filtering (25), phase estimation (17), etc. By using a weighted stacking of seismic data according to local similarity to the reference trace, a seismic image with an increased signal-to-noise ratio can be obtained (23; 21). In this paper, we introduce a new local seismic attribute: *local orthogonalization weight* (LOW) in order to perform local orthogonalization. LOW appeared previously in the definition of local similarity and can be obtained by solving a minimization problem using shaping regularization with a smoothness constraint.

In order to compensate for the loss of useful signal in traditional noise attenuation approaches, because of incorrect parameter selection or inadequacy of denoising assumptions, we employ a weighting operator on the initially denoised section for the retrieval of useful signal from the initial noise section. The new denoising process corresponds to local orthogonalization of signal and noise based on the assumption that final estimated signal and noise should be orthogonal to each other in the time-space domain. The orthogonality assumption is similar as assuming that the signal and noise do not correlate with each other. Thus, the orthogonality assumption is assumed to be valid for all kinds of noise that do not correlate with the useful signal, e.g., random noise. The proposed local-orthogonalization approach can be considered as a specific case of previously proposed nonstationary matching filtering (12) for one-point filter length. The proposed approach is not very sensitive to the parameter selection, and thus is robust in practice. We use two synthetic examples and three field data examples to demonstrate successful performance of the proposed approach

in applications to both conventional and simultaneous-source seismic data.

METHOD

Compensating for the signal-leakage energy by weighting

For many random noise attenuation approaches, the leakage energy is not negligible. We can attempt to retrieve the leaking signal from the noise section by applying a simple nonstationary weighting operator to the initially denoised signal assuming that the leakage energy can be predicted by weighting the useful signal:

$$\mathbf{s}_1 = \mathbf{w} \circ \mathbf{s}_0. \tag{1}$$

Here $\mathbf{s}_1$ is the retrieved signal, $\mathbf{w} \circ \mathbf{s}_0 = diag(\mathbf{w})\mathbf{s}_0 = diag(\mathbf{s}_0)\mathbf{w}$, which denotes Hadamard (or Schur) product, and $diag(\cdot)$ denotes the diagonal matrix composed of an input vector. The weighting vector $\mathbf{w}$ can be estimated by solving the following minimzation problem:

$$\min_{\mathbf{w}} \| \overbrace{\mathbf{d} - \mathbf{P}[\mathbf{d}]}^{\mathbf{n}_0} - \mathbf{w} \circ \overbrace{\mathbf{P}[\mathbf{d}]}^{\mathbf{s}_0} \|_2^2, \tag{2}$$

where $\mathbf{d}$ denotes the observed noisy data, and $\mathbf{P}$ denotes the initial random noise attenuation operator. Equation 2 uses a weighted (scaled) signal $\mathbf{s}_0$ to match the leakage energy in the initial noise section ($\mathbf{n}_0$) in a least-squares sense. In the next section, we will introduce an approach to calculate the weighting vector $\mathbf{w}$ using local orthogonalization.

Local orthogonalization

Based on the assumption that final estimated signal $\mathbf{s}$ and noise $\mathbf{n}$ should be orthogonal to each other, we can orthogonalize the estimated signal and estimated noise by

$$\hat{\mathbf{n}} = \mathbf{n}_0 - w\mathbf{s}_0, \tag{3}$$

$$\hat{\mathbf{s}} = \mathbf{s}_0 + w\mathbf{s}_0, \tag{4}$$

where w is the global orthogonalization weight (GOW) defined by

$$w = \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0}. \tag{5}$$

Here $[\cdot]^T$ denotes transpose. Appendix A provides a demonstration and proof for the global orthogonalization process as denoted by equations 3, 4 and 5. The orthogonality assumption is similar as assuming that the signal and noise do not correlate with each other, which implies the kinds of noise that do not correlate with the useful

signal, e.g., random noise. The orthogonality assumption is assumed to be valid in the original time-space domain, but also has the possibility of being applied in other transformed domains. Instead of using GOW, we propose to use nonstationary local orthogonalization weight (LOW). One possible definition of LOW is:

$$w_m(t) = \frac{\sum_{i=t-m/2}^{t+m/2} s_0(i) n_0(i)}{\sum_{i=t-m/2}^{t+m/2} s_0^2(i)}, \tag{6}$$

where $w_m(t)$ denotes the LOW for each temporal point t with a local window length m. $s_0(t)$ and $n_0(t)$ here denotes the initially estimated signal and noise for each point t.

Solving LOW by shaping regularization

In order to better control the locality and smoothness of LOW, we follow the local-attribute scheme introduced by (10). Simplifing equation 2, and adding a local-smooth constraint to the optimization problem, we obtain:

$$\mathbf{w} = \arg\min_{\mathbf{w}} \parallel \mathbf{n}_0 - \mathbf{S}_0\mathbf{w} \parallel_2^2 + \mathbf{R}(\mathbf{w}). \tag{7}$$

Here, $\mathbf{w}$ is the LOW, $\mathbf{S}_0$ is a diagonal matrix composed of the initial estimated signal $\mathbf{s}_0$: $\mathbf{S}_0 = diag(\mathbf{s}_0)$. $\mathbf{R}$ denotes a smoothing regularization operator. Then we solve the least-squares problem 7 with the help of shaping regularization (11) using a local-smoothness constraint:

$$\mathbf{w} = [\lambda^2\mathbf{I} + \mathcal{T}(\mathbf{S}_0^T\mathbf{S}_0 - \lambda^2\mathbf{I})]^{-1}\mathcal{T}\mathbf{S}_0^T\mathbf{n}_0, \tag{8}$$

where $\mathcal{T}$ is a triangle smoothing operator and λ is a scaling parameter set as $\lambda = \|\mathbf{S}_0^T\mathbf{S}_0\|_2$. Inserting $\mathbf{w}$ into equation 1, we can obtain a new denoising approach:

$$\mathbf{s} = \mathbf{s}_0 + \mathbf{s}_1 = \mathbf{P}[\mathbf{d}] + \mathbf{w} \circ \mathbf{P}[\mathbf{d}] = (\mathbf{I} + diag(\mathbf{w}))\mathbf{P}[\mathbf{d}], \tag{9}$$

$$\mathbf{n} = \mathbf{n}_0 - \mathbf{s}_1 = \mathbf{d} - \mathbf{P}[\mathbf{d}] - \mathbf{w} \circ \mathbf{P}[\mathbf{d}] = \mathbf{d} - (\mathbf{I} + diag(\mathbf{w}))\mathbf{P}[\mathbf{d}]. \tag{10}$$

Here, $\mathbf{s}$ and $\mathbf{n}$ are the final estimated signal and noise.

The controlling parameter for calculating LOW is the width of the triangle smoother, or the smoothing radius (SR). Figure 1 shows the calculated LOWs that correspond to the second synthetic example (Figure 3) with different smoothing radii. The LOW are calculated by solving the equation 7 with the initial denoised signal and initial removed noise set as Figures 3(d) and 3(e), respectively, using the approach shown in equation 8. As we can see from the comparison, LOW is robust with respect to the choice of the smoothing radius (e.g., from 25 samples to 20 samples, or from 10

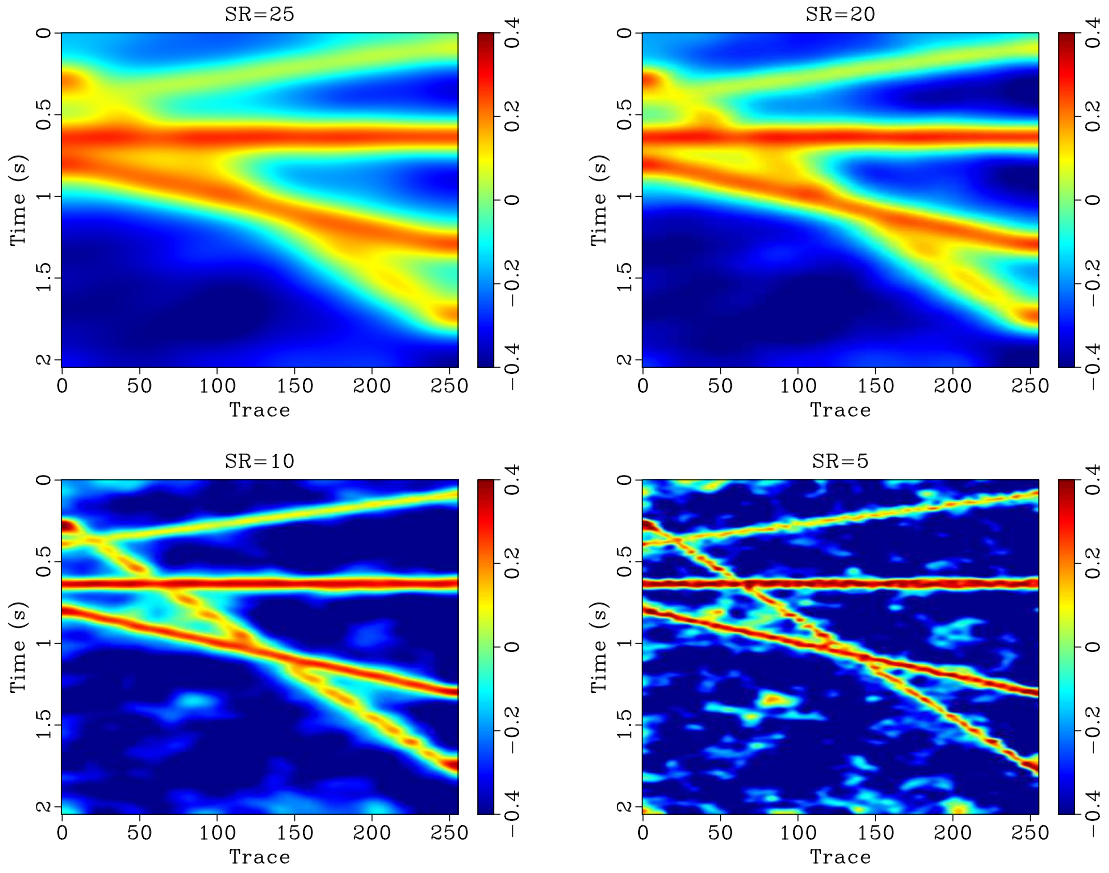


Figure 1: LOWs using different smoothing radii for the example as shown in Figure 3. (a) Vertical and lateral smoothing radii are 25 samples. (b) Vertical and lateral smoothing radii are 20 samples. (c) Vertical and lateral smoothing radii are 10 samples. (d) Vertical and lateral smoothing radii are 5 samples. ortho/orthocomplex low1,low2,low3,low4

samples to 5 samples), but may differ a lot for significantly different smoothing radii (e.g., from 25 samples to 5 samples). We can also conclude from the comparison that as the smoothing radius increases, the temporal and spatial resolutions decrease, but the anti-noise ability increases. Note that the orthogonalization equations 2 and 7 can also be understood as one-point nonstationary matching filtering (12).

SYNTHETIC EXAMPLES

We use two synthetic examples to test the performance of the proposed denoising algorithm. For consistency, we use *denoising* to denote random noise attenuation for regular seismic data, and *deblending* to denote blending noise attenuation for simultaneous-source seismic data. The first example is a deblending test, similar to the one used by (19). The second example is a denoising test that contains four crossing linear events. In order to test the improvement in the signal-to-noise ratio of these two synthetic examples, we utilize the following measurement:

$$SNR = 10 \log_{10} \frac{\|\mathbf{s}_{true}\|_2^2}{\|\mathbf{s}_{true} - \mathbf{s}\|_2^2}. \tag{11}$$

In addition to SNR, we also propose to use local similarity (10) as a convenient measure to evaluate denoising performance. Appendix B gives a brief review of local similarity. After the local similarity map between the denoised data and removed noise is calculated, we can judge from the local similarity if there is leakage energy in the noise. When the clean data is unknown for field data denoising tests, the SNR based evaluation becomes unavailable. Besides, the SNR is not always the best measurement for denoising performance because it does not measure the signal leakage. However, the local similarity map can always be used to evaluate the denoising performance.

Figure 2 shows the deblending performance using a conventional MF and the proposed denoising approach. The original clean, unblended data are shown in Figure 2(a). After blending using a simple random-dithering method (7), we obtained the noisy data containing spike-like blending noise (5(a)). Figure 2(d) demonstrates a denoised section after using MF. The noise section corresponding to MF is shown in Figure 2(e) and contains a certain amount of signal leakage in the form of linear events. Using the proposed denoising approach, we obtained a denoised section with stronger-amplitude linear events. In the corresponding noise section, any coherent linear events are barely noticeable, suggesting a nearly perfect deblending. In this example, because the noise are spike-like blending noise, we did smooth too much during shaping regularization. In this example, the vertical and lateral smoothing radii are both 2 samples. Figures 2(f) and 2(i) show local similarity of the noise section to the denoised section before and after using the proposed approach. After applying the proposed method, the noise section exhibits low similarity with the denoised section. Although structure-oriented median filtering can handle the dipping reflector, they can only work well for relative simpler structures where the dip estimation can be precisely obtained, in which case there will be no need to use the proposed approach.

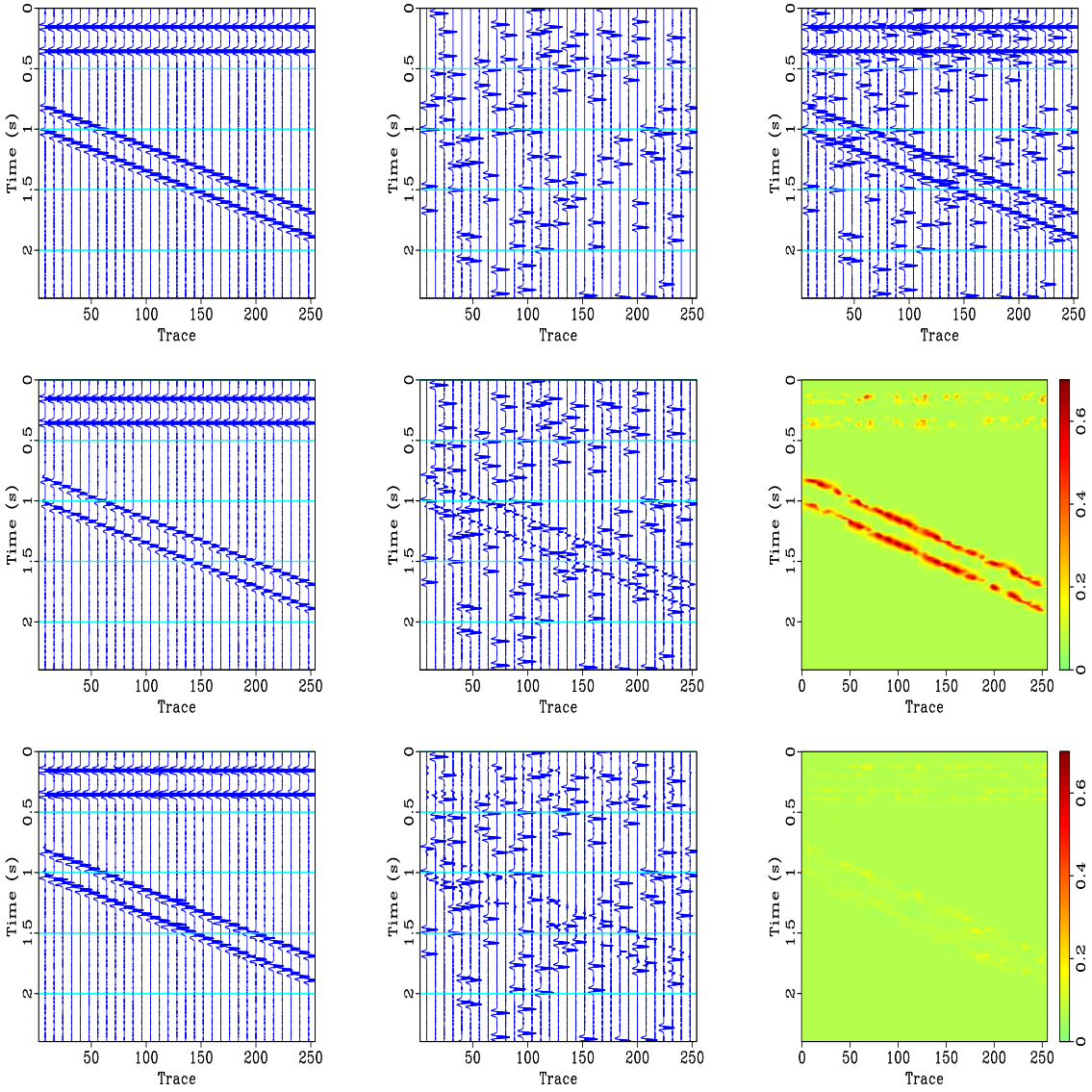


Figure 2: Comparison of using the local-orthogonalization-based random noise attenuation approach for deblending test before and after. (a) Unblended section. (b) Blending noise section. (c) Blended section. (d) Deblended result using MF. (e) Noise section using MF. (f) Local similarity map between (d) and (e). (g) Deblended result using the proposed method. (h) Noise section using the proposed method. (i) Local similarity map between (g) and (h).

ortho/orthocsimul huo,huo-noise,huos,huos-mf,huosdiff-mf,huos-simi,huos-ortho,huosdiff-ortho,huos-si

However, when the dip estimation is not accurate, or there are conflicting dips in a processing window, the structure-oriented median filtering may not work well, in which case we can use the proposed approach for retrieving the leaked useful energy.

Figure 3 shows the denoising performance based on the conventional f-x deconvolution (5). The clean data is shown in Figure 3(a) and the noise section, corresponding to f-x deconvolution, is shown in Figure 3(f); it contains some horizontal and low-dip-angle signals. The coherent leakage energy is also visible in the local similarity map (Figure 3(f)). Applying the proposed denoising approach, we obtained a denoised section with stronger-amplitude horizontal events. The corresponding noise section does not contain any coherent signal. In this example, the vertical and lateral smoothing radii are both 25 samples. Figures 3(f) and 3(i) denote local similarity of the noise section to the denoised section before and after using the proposed approach. The similarity decreases significantly after applying local orthogonalization.

The SNR comparison before and after using local-orthogonalization approach is summarized in Table 1.

Test	Deblending	denoising
Original (dB)	1.173	-1.72
Initially denoised (dB)	17.65	21.21
Orthogonalized (dB)	20.95	25.30

Table 1: Comparison of SNR before and after using the proposed approach.

FIELD EXAMPLES

In this section, we use three field data sets to evaluate the proposed method. The first example is a simulated deblending test and the latter two examples are common denoising tests.

Figure 4(a) shows the unblended data for the first test, which is a deep-water common receiver section containing events that are nearly flat. After numerical blending using the same blending approach as in the first synthetic example, the blended data (shown in Figure 4(c)) demonstrate strong interference. Figure 4(d) shows the deblended data after the use of f-x deconvolution. When checking the noise section in Figure 4(e), we found many coherent signals, suggesting a significant leakage of useful signal energy. The denoised section after using the proposed approach is shown in Figure 4(g). The noise section no longer contains coherent useful signals while the removed blending noise appears nearly unchanged. In this example, we set the vertical and lateral smoothing radii to 25 samples. The similarity maps before and after application of the proposed method are shown in Figure 4(f) and 4(h). The SNR using the previous criteria of the blended data is 0.04. After the initial deblending using f-x deconvolution, SNR improves to 7.44. Further, after the proposed

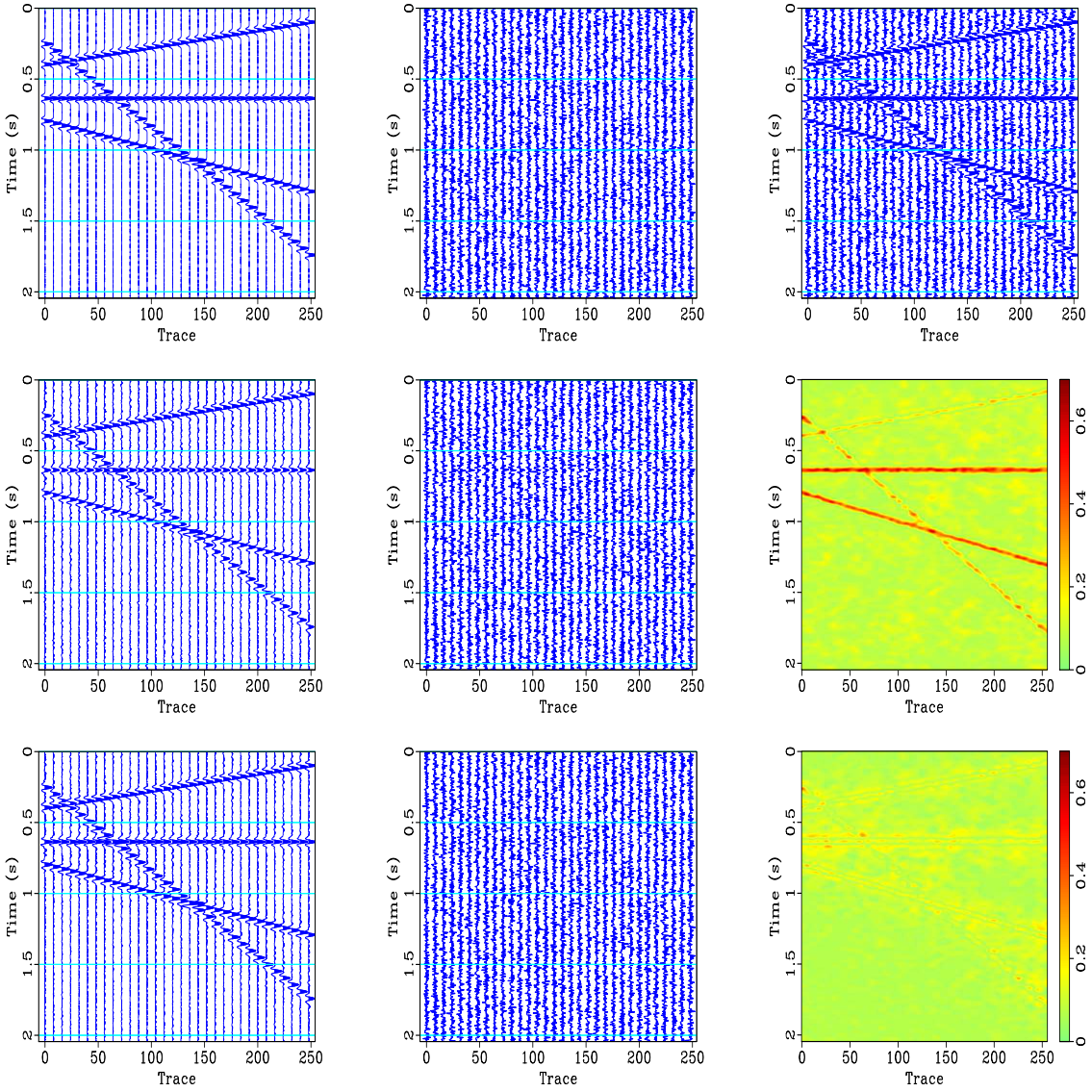


Figure 3: Comparison of using the local-orthogonalization-based random noise attenuation approach for denoising test before and after. (a) Clean section. (b) Random noise section. (c) Noisy section. (d) Denoised result using f-x deconvolution. (e) Noise section using f-x deconvolution. (f) Local similarity map between (d) and (e). (g) Denoised result using the proposed method. (h) Noise section using the proposed method. (i) Local similarity map between (g) and (h).

ortho/orthocomplex c,n,cn,cfx,cdifffx,csimi-difffx,cfx2,cdifffx2,csimi-difffx2

retrieving-signal process, SNR improves to 10.08. It is worth to be mentioned that the preservation of useful signal is of significant importance in deblending, and thus the proposed approach is particularly attractive in its application to deblending.

The second field data test is for a land seismic section, which was used previously by (10) and (24). The input image is shown in Figure 5. The shallow part of the section is noisy, which may make interpretation difficult. After a conventional f-x deconvolution, a much cleaner section is obtained (Figure 6(a)). Some coherent signal is left in the noise section shown in Figure 6(d), especially around 0.75s and the 350th trace. After local signal-and-noise orthogonalization, we obtained a clean denoised section but with strong amplitude in some regions and almost no loss of useful component in the noise section (Figure 6(d)). In this example, the vertical and lateral smoothing radii are both 2 samples. Unlike the synthetic data test and the previous field data test, for this field data test, the clean signal is unknown, so the SNR measurement cannot be applied in this case. The two corresponding similarity maps are shown in Figure 7. The local similarity after applying the proposed approach is nearly zero along the whole seismic section.

The third field data example is borrowed from (20). The noisy data is shown in Figure 9(a). The initial signal and noise models are prepared using $f-x$ regularized nonstationary autoregression (RNAR). Following the approach of (20), the initially denoised data are shown in Figure 9(a). The corresponding noise section is shown in Figure 9(c). The general denoising approach shows good performance considering the clean denoised image and random removed noise. However, we can find small useful signals from the noise section, as indicated by the frame boxes in Figure 9(c). The denosied image using the proposed approach is shown in Figure 9(b). After using local orthogonalization to retrieve useful signal, the coherent signal disappears in the final noise section (Figure 9(d)), as indicated by the frame boxes. The zoomed sections corresponding to frame boxes A and B in Figure 9 are shown in Figure 10. In this example, we chose the vertical and lateral smoothing radius as 5 samples. Before using the proposed approach, the local similarity map between the denoised data and removed noise (Figure 11(a)) showed some high-similarity anomalies. After applying the proposed approach, the similarity map is nearly zero everywhere across the whole section.

CONCLUSIONS

Conventional methods for noise attenuation may cause some leakage of signal energy in the noise section as a result of incorrect parameter selection or inadequacy of denoising assumptions. We have shown that it is possible to retrieve the leakage energy by applying a weighting operator to the initial signal and adding the retrieved signal energy to the initially denoised data to obtain the final denoised data. In order to design an optimized weighting operator, we introduce a new local attribute, called local orthogonalization weight (LOW). LOW can be obtained by solving a least-square minimization problem using shaping regularization with a smoothness

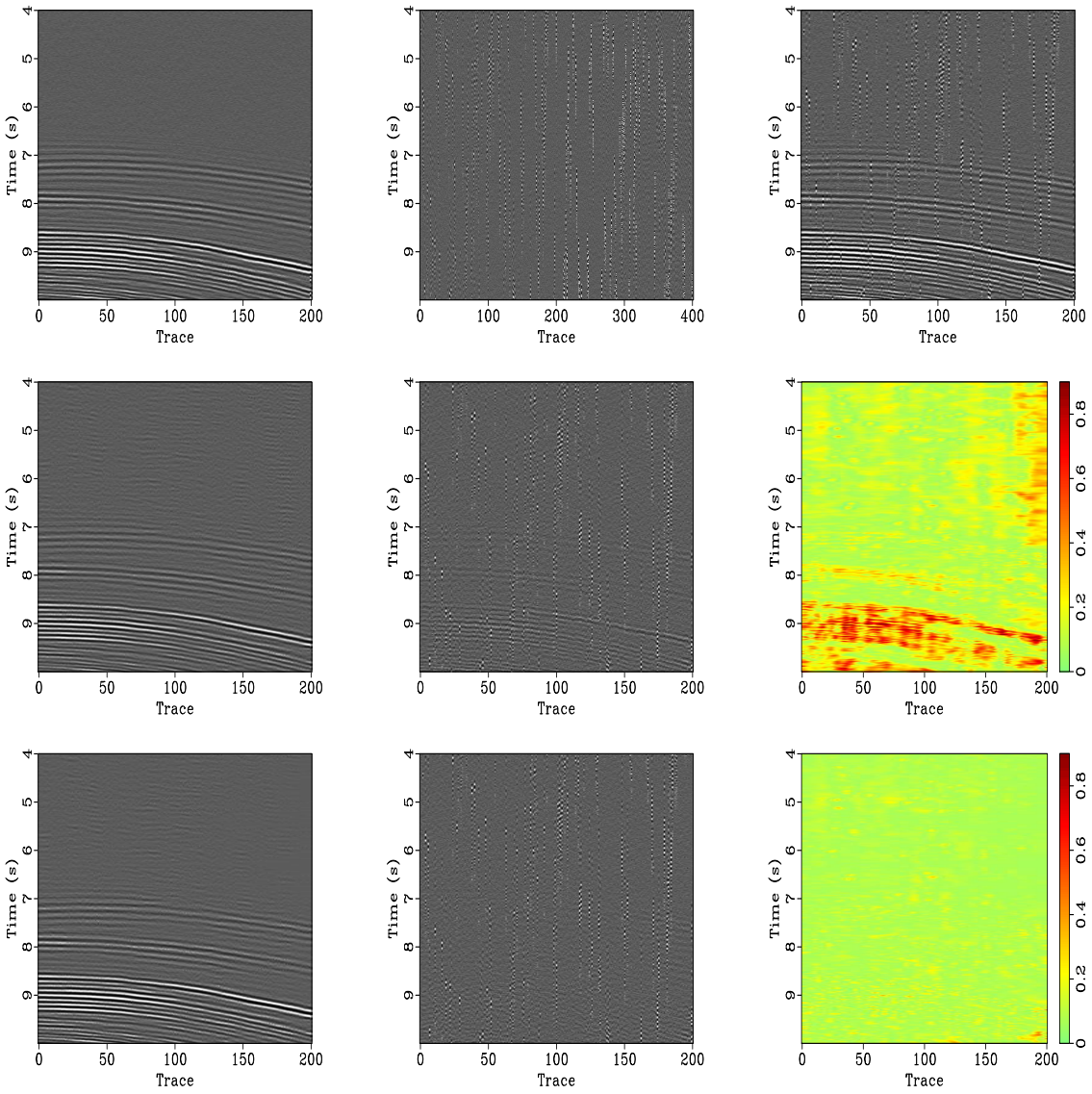


Figure 4: Comparison of using the local-orthogonalization-based random noise attenuation approach for deblending test before and after. (a) Unblended section (not clean, also containing random noise). (b) Blending noise section. (c) Blended section. (d) Deblended result using $f - x$ deconvolution. (e) Noise section using $f - x$ deconvolution. (f) Local similarity map between (d) and (e). (g) Deblended result using the proposed method. (h) Noise section using the proposed method. (i) Local similarity map between (g) and (h).

ortho/orthofair fairunblended2,fair-noise,fairblended2,fairdeblended2fx2,fairdeblended2dif-fx2,fairdif-s

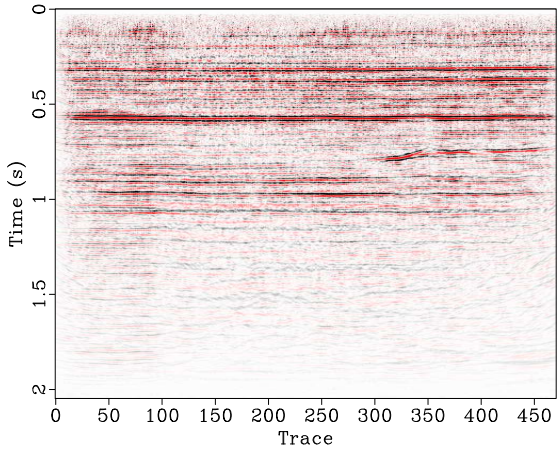


Figure 5: Noisy post-stack land field data (13). ortho/orthonine pp

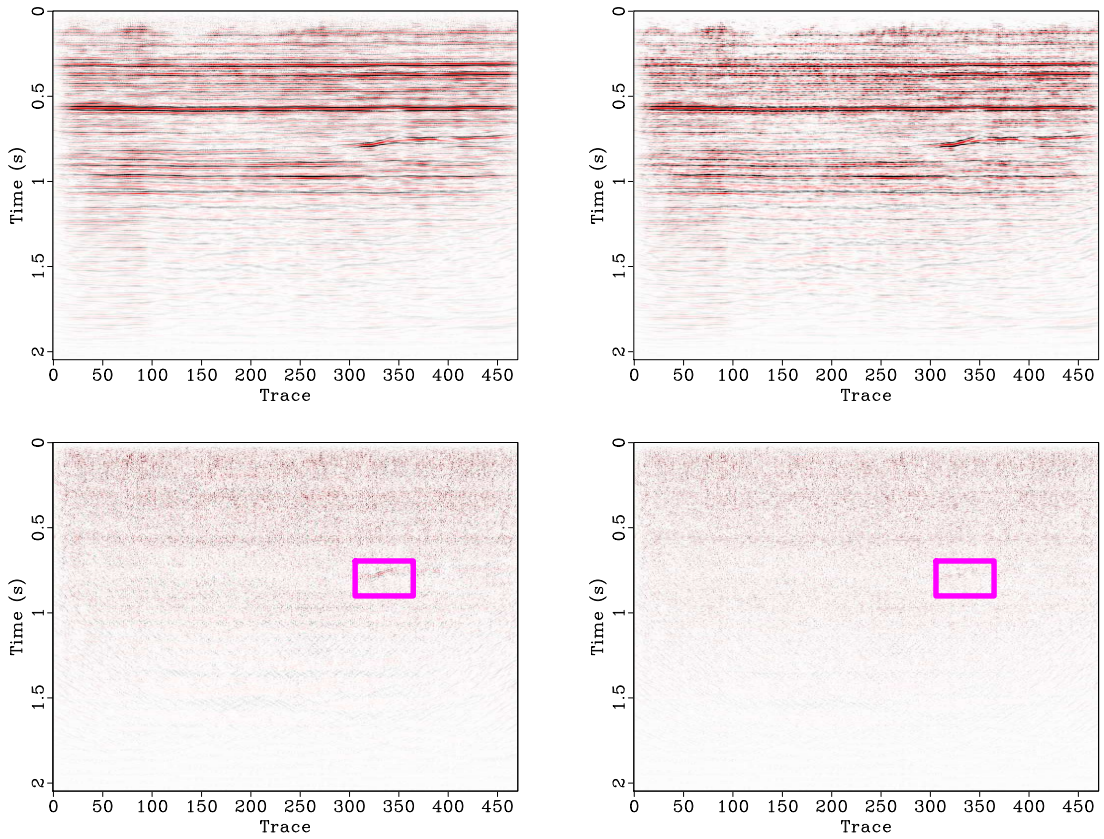


Figure 6: (a) Denoised section using $f - x$ deconvolution. (b) Denoised section using the proposed approach. (c) Noise section using $f - x$ deconvolution. (d) Noise section using the proposed approach. ortho/orthonine pp-fx,pp-ortho,ppdiff-fx0,ppdiff-ortho0

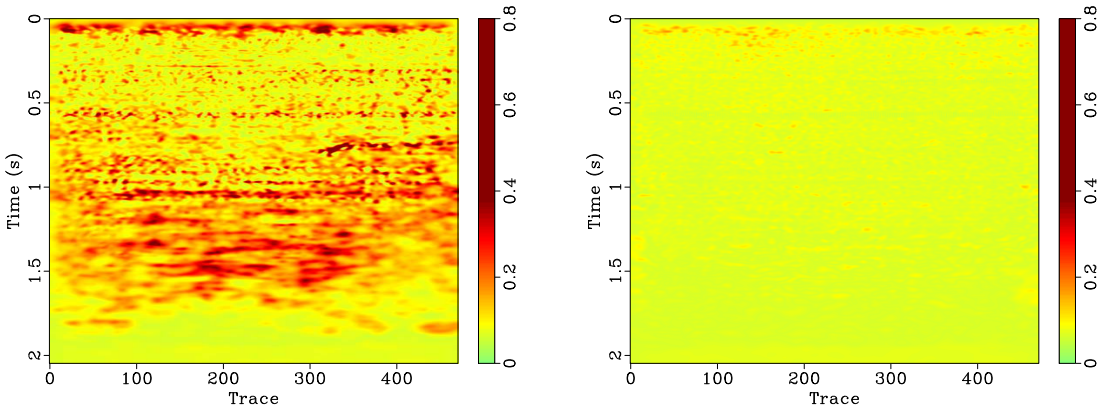


Figure 7: Comparison of similarity between noise section and denoised section with and without using the proposed random noise attenuation approach. (a) Similarity map using $f - x$ deconvolution. (b) Similarity map using the proposed method. ortho/orthonine pp-simi,pp-simi-ortho

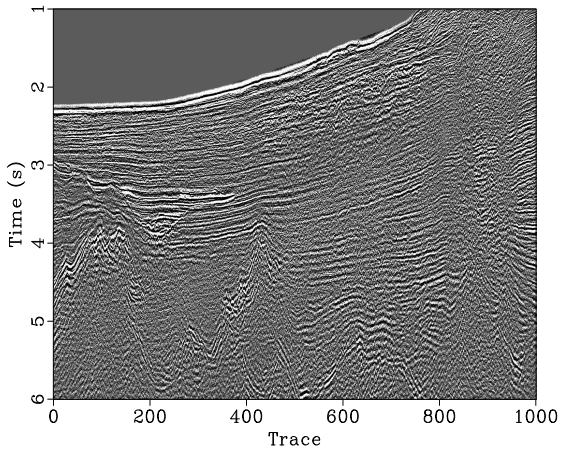


Figure 8: Noisy post-stack field data (20). ortho/orthorna data

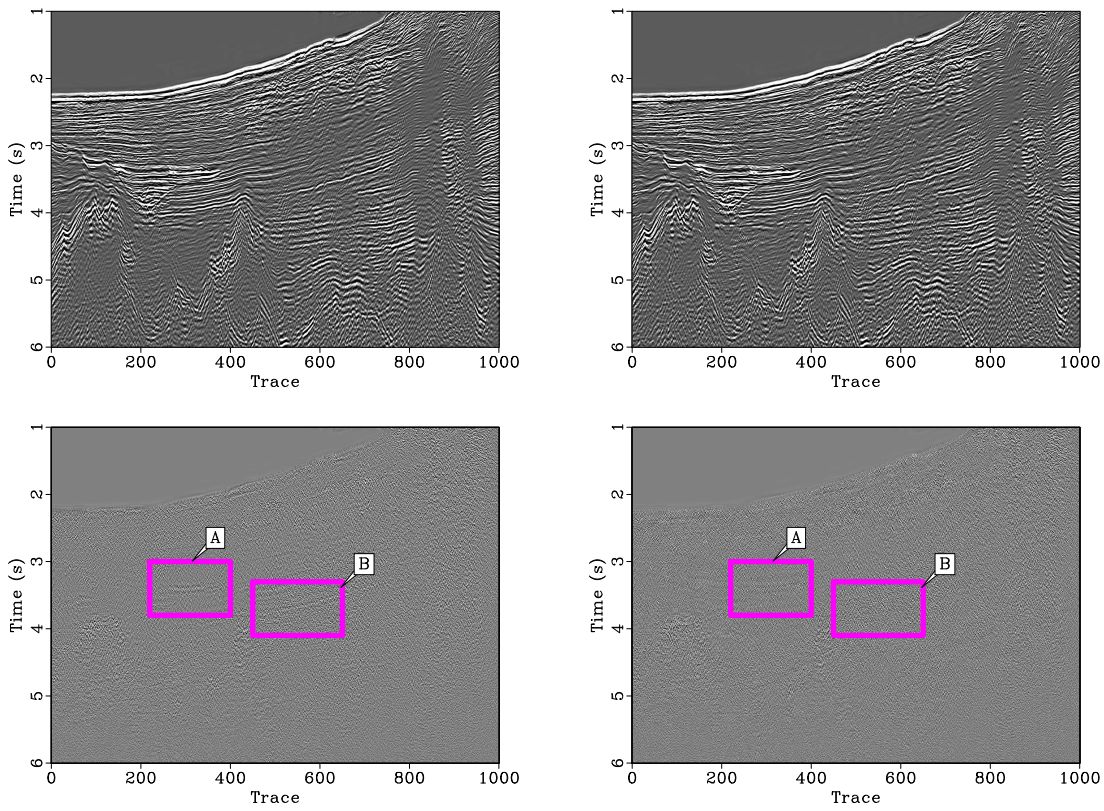


Figure 9: (a) Denoised section using $f-x$ RNAR. (b) Denoised section using the proposed approach. (c) Noise section using $f-x$ RNAR. (d) Noise section using the proposed approach. ortho/orthorna npre,rna-ortho,nprediff0,rnadiff-ortho0

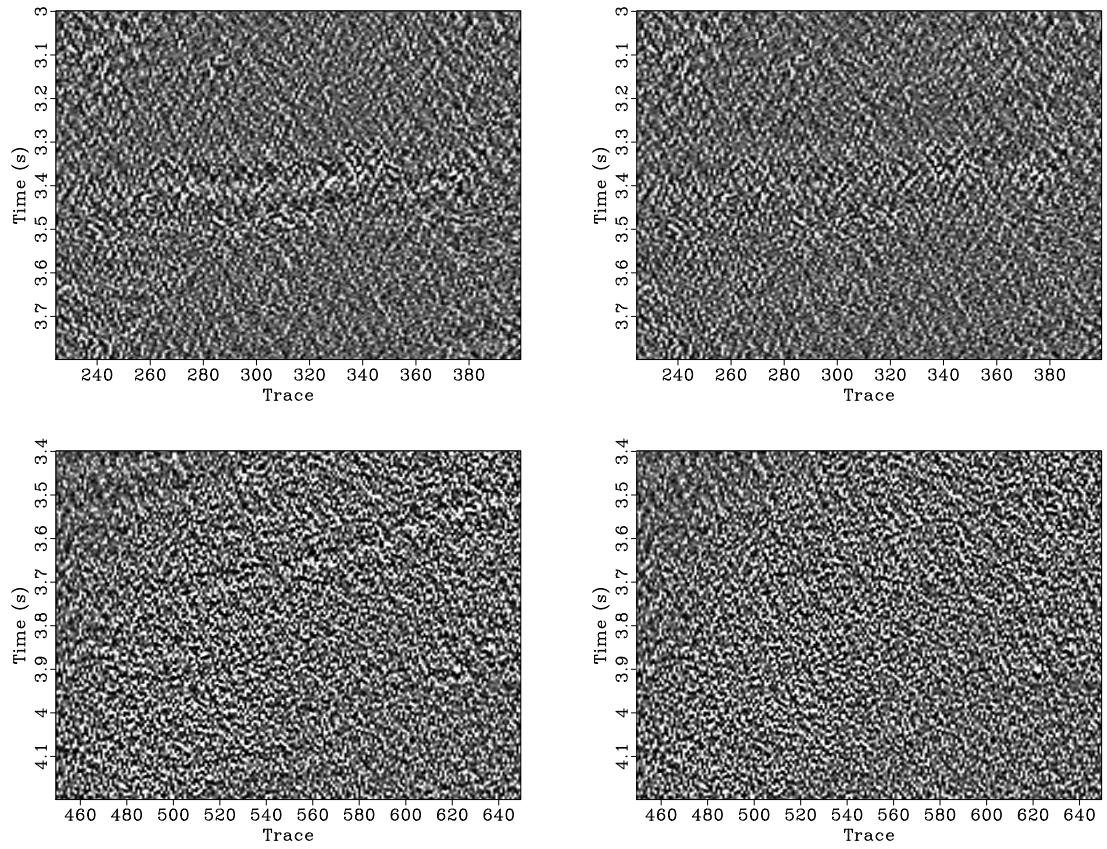


Figure 10: (a) Zoomed section using f-x RNAR (frame box A shown in Figure 9). (b) Zoomed section using the proposed approach (frame box A shown in Figure 9). (c) Zoomed section using f-x RNAR (frame box B shown in Figure 9). (d) Zoomed section using the proposed approach (frame box B shown in Figure 9). ortho/orthorna zoom-rnadif-a,zoom-orthodif-a,zoom-rnadif-b,zoom-orthodif-b

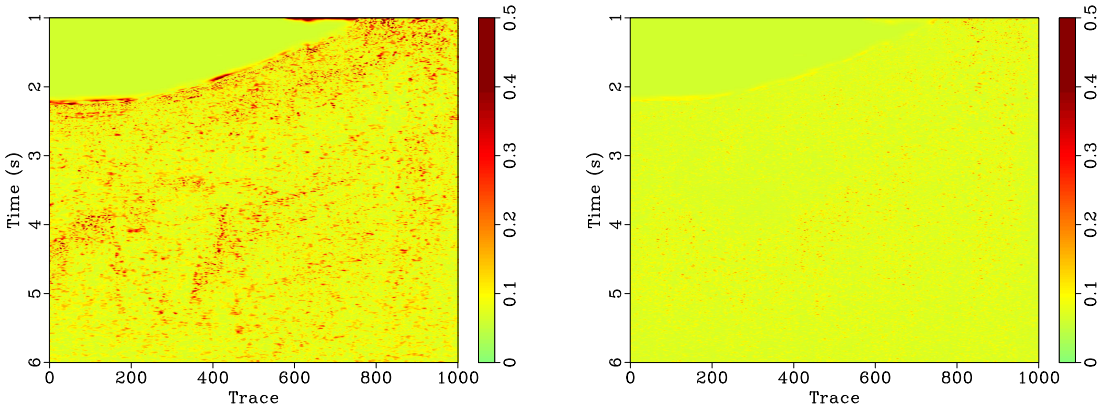


Figure 11: Comparison of similarity between noise section and denoised section with and without using the proposed random noise attenuation approach. (a) Similarity map using $f - x$ RNAR. (b) Similarity map using the proposed method. ortho/orthorna rna-simi,rna-simi-ortho

constraint. The proposed denoising approach corresponds to locally orthogonalizing signal and noise. Once the initial signal and noise models are given, the proposed approach can help retrieve the residual signal in the noise section. The local orthogonalization approach to random noise attenuation is applicable to processing of blended simultaneous-source seismic data, in which the preservation of useful signal is particularly important. Although the examples were 2-D, the method applies equally well in 3-D and in any other number of dimensions.

ACKNOWLEDGMENTS

We thank FairfieldNodal for the data used in the first field data example. We also thank Luke Decker for helpful discussions, Bill Harlan and two anonymous reviewers for constructive suggestions that made the manuscript better. The paper was prepared in the Madagascar open-source software environment (16).

APPENDIX A: SIGNAL-AND-NOISE ORTHOGONALIZATION

As shown schematically in Figure A-1, the initially estimated signal and noise are denoted by $\mathbf{s}_0$ and $\mathbf{n}_0$, respectively. By projecting $\mathbf{n}_0$ to the direction of $\mathbf{s}_0$, we can get the projection $w\mathbf{s}_0$. The other component of $\mathbf{n}_0$ is the final estimated noise, as shown in equation 3. The final estimated signal is thus the summation of the initially estimated signal $\mathbf{s}_0$ and the projection component $w\mathbf{s}_0$.

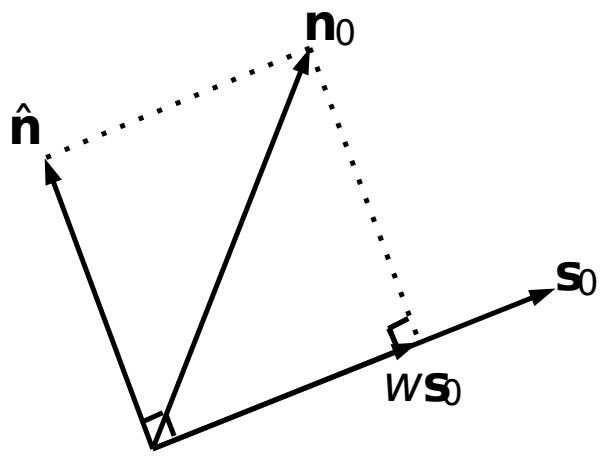


Figure A-1: Demonstration of signal-and-noise orthogonalization. ortho/. demo

When $w = \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0}$, the following equation holds:

$$\begin{aligned} \hat{\mathbf{n}}^T \hat{\mathbf{s}} &= (\mathbf{n}_0 - w\mathbf{s}_0)^T (\mathbf{s}_0 + w\mathbf{s}_0) = \mathbf{n}_0^T \mathbf{s}_0 - w\mathbf{s}_0^T \mathbf{s}_0 + w\mathbf{n}_0^T \mathbf{s}_0 - w^2 \mathbf{s}_0^T \mathbf{s}_0 \\ &= \mathbf{n}_0^T \mathbf{s}_0 - \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0} \mathbf{s}_0^T \mathbf{s}_0 + \frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0} \mathbf{n}_0^T \mathbf{s}_0 - \left(\frac{\mathbf{n}_0^T \mathbf{s}_0}{\mathbf{s}_0^T \mathbf{s}_0} \right)^2 \mathbf{s}_0^T \mathbf{s}_0 = 0 \end{aligned} \tag{A-1}$$

Here, $\hat{\mathbf{s}}$ and $\hat{\mathbf{n}}$ denote the final estimated signal and noise, respectively, and appear orthogonal to each other. This orthogonalization approach is also known as Gram-Schmidt orthogonalization (18). Note that w defined above can be obtained by solving the least-squares optimization problem:

$$\min_w \|w\mathbf{s}_0 - \mathbf{n}_0\|_2^2, \tag{A-2}$$

which we extend to equation 7 in the main text.

APPENDIX B: REVIEW OF LOCAL SIMILARITY

(10) defined local similarity between vectors $\mathbf{a}$ and $\mathbf{b}$ as:

$$\mathbf{c} = \sqrt{\mathbf{c}_1^T \mathbf{c}_2} \tag{B-1}$$

where $\mathbf{c}_1$ and $\mathbf{c}_2$ come from two least-squares minimization problems:

$$\mathbf{c}_1 = \arg\min_{\mathbf{c}_1} \|\mathbf{a} - \mathbf{B}\mathbf{c}_1\|_2^2, \tag{B-2}$$

$$\mathbf{c}_2 = \arg\min_{\mathbf{c}_2} \|\mathbf{b} - \mathbf{A}\mathbf{c}_2\|_2^2, \tag{B-3}$$

where $\mathbf{A}$ is a diagonal operator composed from the elements of $\mathbf{a}$: $\mathbf{A} = diag(\mathbf{a})$ and $\mathbf{B}$ is a diagonal operator composed from the elements of $\mathbf{b}$: $\mathbf{B} = diag(\mathbf{b})$. Least-squares problems B-2 and B-3 can be solved with the help of shaping regularization with a smoothness constraint:

$$\mathbf{c}_1 = [\lambda_1^2\mathbf{I} + \mathcal{T}(\mathbf{B}^T\mathbf{B} - \lambda_1^2\mathbf{I})]^{-1}\mathcal{T}\mathbf{B}^T\mathbf{a}, \tag{B-4}$$

$$\mathbf{c}_2 = [\lambda_2^2\mathbf{I} + \mathcal{T}(\mathbf{A}^T\mathbf{A} - \lambda_2^2\mathbf{I})]^{-1}\mathcal{T}\mathbf{A}^T\mathbf{b}, \tag{B-5}$$

where $\mathcal{T}$ is a smoothing operator, and λ_1 and λ_2 are two parameters controlling the physical dimensionality and enabling fast convergence when inversion is implemented iteratively. These two parameters can be chosen as $\lambda_1 = \|\mathbf{B}^T\mathbf{B}\|_2$ and $\lambda_2 = \|\mathbf{A}^T\mathbf{A}\|_2$ (10). The definition of $\mathbf{c}_1$ and $\mathbf{c}_2$ are equivalent to definition of LOW in this paper.

REFERENCES

[1] Abma, R. L., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: 72nd Annual International Conference and Exhibition, EAGE, Extended Abstracts.

[2] Beasley, C. J., R. E. Chambers, and Z. Jiang, 1998, A new look at simultaneous sources: 68th Annual International Meeting, SEG, Expanded Abstracts, 133–135.

[3] Bekara, M., and M. van der Baan, 2009, Random and coherent noise attenuation by empirical mode decomposition: Geophysics, **74**, V89–V98.

[4] Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.

[5] Canales, L., 1984, Random noise reduction: SEG expanded abstracts: 54th Annual international meeting, 525–527.

[6] Chen, Y., 2014, Deblending using a space-varying median filter: Exploration Geophysics, doi:http://dx.doi.org/10.1071/EG14051.

[7] Chen, Y., S. Fomel, and J. Hu, 2014a, Iterative deblending of simultaneous-source seismic data using seislet-domain shaping regularization: Geophysics, **79**, V179–V189.

[8] Chen, Y., and J. Ma, 2014, Random noise attenuation by *f-x* empirical mode decomposition predictive filtering: Geophysics, **79**, V81–V91.

[9] Chen, Y., J. Yuan, Z. Jin, K. Chen, and L. Zhang, 2014b, Deblending using normal moveout and median filtering in common-midpoint gathers: Journal of Geophysics and Engineering, **11**, 045012.

[10] Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.

[11] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[12] ———, 2009, Adaptive multiple subtraction using regularized nonstationary regression: Geophysics, **74**, V25–V33.
[13] ———, 2013, Seismic data decomposition into spectral components using regularized nonstationary autoregression: Geophysics, **78**, O69–O76.
[14] Fomel, S., M. Backus, K. Fouad, B. Hardage, and G. Winters, 2005, A multistep approach to multicomponent seismic image registration with application to a West Texas carbonate reservoir study: 75th Annual International Meeting, SEG, Expanded Abstracts, 1018–1021.
[15] Fomel, S., and L. Jin, 2009, Time-lapse image registration using the local similarity attribute: Geophysics, **74**, A7–A11.
[16] Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013, Madagascar open-source software project: Journal of Open Research Software, **1**, e8.
[17] Fomel, S., and M. van der Baan, 2014, Local skewness attribute as a seismic phase detector: Interpretation, **2**, SA49–SA56.
[18] Hazewinkel, M., 2001, Encyclopedia of mathematics: Springer.
[19] Huo, S., Y. Luo, and P. G. Kelamis, 2012, Simultaneous sources separation via multidirectional vector-median filtering: Geophysics, **77**, V123–V131.
[20] Liu, G., X. Chen, J. Du, and K. Wu, 2012, Random noise attenuation using *f-x* regularized nonstationary autoregression: Geophysics, **77**, V61–V69.
[21] Liu, G., S. Fomel, and X. Chen, 2011a, Stacking angle-domain common-image gathers for normalization of illumination: Geophysical Prospecting, **59**, 244–255.
[22] ———, 2011b, Time-frequency analysis of seismic data using local attributes: Geophysics, **76**, P23–P34.
[23] Liu, G., S. Fomel, L. Jin, and X. Chen, 2009, Stacking seismic data using local correlation: Geophysics, **74**, V43–V48.
[24] Liu, Y., and S. Fomel, 2012, Seismic data analysis using local time-frequency decomposition: Geophyscial Prospecting, **60**, 1–10.
[25] Liu, Y., S. Fomel, and G. Liu, 2010, Nonlinear structure-enhancing filtering using plane-wave prediction: Geophysical Prospecting, **58**, 415–427.
[26] Soubaras, R., 1995, Prestack random and impulsive noise attenuation by f-x projection filtering: 65th Annual International Meeting, SEG, Expanded Abstracts, 711–714.
[27] Zhang, R., X. Song, S. Fomel, M. K. Sen, and S. Srinivasan, 2013, Time-lapse seismic data registration and inversion for CO_2 sequestration study at Cranfield: Geophysics, **78**, B329–B338.

Post-stack velocity analysis by separation and imaging of seismic diffractions

Sergey Fomel[*]*, Evgeny Landa*[†]*, and M. Turhan Taner*[‡*]

ABSTRACT

Small geological features manifest themselves in seismic data in the form of diffracted waves, which are fundamentally different from seismic reflections. Using two field data examples and one synthetic example, we demonstrate the possibility of separating seismic diffractions in the data and imaging them with optimally chosen migration velocities. Our criterion for separating reflection and diffraction events is the smoothness and continuity of local event slopes that correspond to reflection events. For optimal focusing, we develop the local varimax measure. The objectives of this work are velocity analysis implemented in the post-stack domain and high-resolution imaging of small-scale heterogeneities. Our examples demonstrate the effectiveness of the proposed method for high-resolution imaging of such geological features as faults, channels, and salt boundaries.

INTRODUCTION

Diffracted and reflected seismic waves are fundamentally different physical phenomena (14). Most seismic data processing is tuned to imaging and enhancing reflected waves, which carry most of the information about subsurface. The value of diffracted waves, however, should not be underestimated (13). When seismic exploration focuses on identifying small subsurface features (such as faults, fractures, channels, and rough edges of salt bodies) or small changes in seismic reflectivity (such as those caused by fluid presence or fluid flow during reservoir production), it is diffracted waves that contain the most valuable information.

In this paper, we develop an integrated approach for extracting and imaging of diffracted events. We start with stacked or zero-offset data as input and produce time-migrated images with separated and optimally focused diffracted waves as output. The output of our processing flow can be compared to coherence cubes (1; 18). While the coherence cube algorithm tries to enhance incoherent features, such as faults, in the migrated image domain, we perform the separation in unmigrated data, where these features appear in the form of diffracted waves.

[*]**e-mail:** sergey.fomel@beg.utexas.edu

We also introduce diffraction-event focusing as a criterion for migration velocity analysis, as opposed to the usual "flat-gather" criterion used in seismic imaging. Focusing analysis is applicable not only to multi-coverage prestack data but also to post-stack or single-coverage data.

The idea of extracting information from seismic diffractions is not new. (11) used forward modeling and local slant stacks for estimating velocities from diffractions; (15) used common-diffraction-point sections for imaging of diffraction energy and detecting local heterogeneities; (22) simulated diffraction responses for enhancing velocity analysis. (20) incorporated diffraction imaging in wave-equation migration velocity analysis.

The novelty of our approach is in integration of two essential steps:

1. Separating diffracted and reflected events in the data space,

2. Focusing analysis for automatic detection of migration velocities optimal for imaging diffractions.

We explain both steps and illustrate their application with field and synthetic datasets.

SEPARATING DIFFRACTIONS

The underlying assumption that we employ for separating diffracted and reflected events is that, in a stacked data volume, background reflections correspond to strong coherent events with continuously variable slopes. Removing those events reveals other coherent information, often in the form of seismic diffractions. We propose to identify and remove reflection events with the method of plane-wave destruction (2; 5). Plane-wave destruction estimates continuously variable local slopes of dominant seismic events by forming a prediction of each data trace from its neighboring traces with optimally compact non-stationary filters that follow seismic energy along the estimated slopes. Minimizing the prediction residual while constraining the local slopes to vary smoothly provides an optimization objective function analogous to differential semblance (23). Iterative optimization of the objective function generates a field of local slopes. The prediction residual then contains all events, including seismic diffractions, that do not follow the dominant slope pattern. An analogous idea, but with implementation based on prediction-error filters, was previously discussed by (3) and (21). Although separation of reflection and diffraction energy can never be exact, our method serves the practical purpose of enhancing the wave response of small subsurface discontinuities.

IMAGING DIFFRACTIONS

How can one detect the spatially-variable velocity necessary for focusing of different diffraction events? A good measure of focusing is the *varimax norm* used by (25) for minimum-entropy deconvolution and by (17) for zero-phase correction. The varimax norm is defined as

$$\phi = \frac{N \sum_{i=1}^{N} s_i^4}{\left(\sum_{i=1}^{N} s_i^2 \right)^2} \ , \tag{1}$$

where s_i are seismic signal amplitudes inside a window of size N. Varimax is simply related to kurtosis of zero-mean signals.

Rather than working with data windows, we turn focusing into a continuously variable attribute using the technique of *local attributes* (8). Noting that the correlation coefficient of two sequences a_i and b_i is defined as

$$c[a,b] = \frac{\sum_{i=1}^{N} a_i \, b_i}{\sqrt{\sum_{i=1}^{N} a_i^2 \sum_{i=1}^{N} b_i^2}} \tag{2}$$

and the correlation of a_i with a constant is

$$c[a,1] = \frac{\sum_{i=1}^{N} a_i}{\sqrt{N \sum_{i=1}^{N} a_i^2}} \ , \tag{3}$$

one can interpret the varimax measure in equation 1 as the inverse of the squared correlation coefficient between s_i^2 and a constant: $\phi = 1/c[s^2,1]^2$. Well-focused signals have low correlation with a constant and correspondingly high varimax.

Going further toward a continuously variable focusing attribute, notice that the squared correlation coefficient can be represented as the product of two quantities $c[s^2,1]^2 = p\,q$, where

$$p = \frac{\sum_{i=1}^{N} s_i^2}{N} \ , \quad q = \frac{\sum_{i=1}^{N} s_i^2}{\sum_{i=1}^{N} s_i^4} \ . \tag{4}$$

Furthermore, p is the solution of the least-squares minimization problem

$$\min_p \sum_{i=1}^{N} \left(s_i^2 - p\right)^2 \ , \tag{5}$$

and q is the solution of the least-squares minimization

$$\min_q \sum_{i=1}^{N} \left(1 - q\, s_i^2\right)^2 \ . \tag{6}$$

This allows us to define a continuously variable attribute ϕ_i by using continuously variable quantities p_i and q_i, which are defined as solutions of regularized optimization problems

$$\min_{p_i} \left(\sum_{i=1}^{N} \left(s_i^2 - p_i\right)^2 + R\left[p_i\right] \right) \ , \tag{7}$$

$$\min_{q_i} \left(\sum_{i=1}^{N} \left(1 - q_i\, s_i^2\right)^2 + R\left[q_i\right] \right) \ , \tag{8}$$

where R is a regularization operator designed to avoid trivial solutions by enforcing a desired behavior (such as smoothness). Shaping regularization (9) provides a particularly convenient method for enforcing smoothing in an iterative optimization scheme.

We apply the local focusing measure to obtain migration-velocity panels for every point in the image. First, we follow the procedure outlined in the previous section to replace a stacked or zero-offset section with a section containing only separated diffractions. Next, we migrate the data multiple times with different migration velocities. This is accomplished by *velocity continuation* (6), a method that performs time-migration velocity analysis by continuing seismic images in velocity with the process also called "image waves" (12). The velocity continuation theory (7) shows that one can accomplish time migration with a set of different velocities by making differential steps in velocity similarly to the method of cascaded migrations (16) but described and implemented as a continuous process. While comparable in theory to an ensemble of Stolt migrations (10; 19), velocity continuation has the advantage of working directly in the image domain. It is implemented with an efficient and robust algorithm based on the Fast Fourier Transform.

Finally, we compute ϕ_i for every sample point in each of the migrated images. Thus, N in equations 4 and 5 refers to the total number of sample points in an image. The output is *focusing image gathers* (FIGs), exemplified in Figure 1. A FIG is analogous to a conventional migration-velocity analysis panel and suitable for picking migration velocities. The main difference is that the velocity information is obtained from analysis of diffraction focusing as opposed to semblance of flattened image gathers used in prestack analysis.

EXAMPLES

Three different examples illustrate applications of our method to imaging of geological faults and irregular salt boundaries.

Fault detection

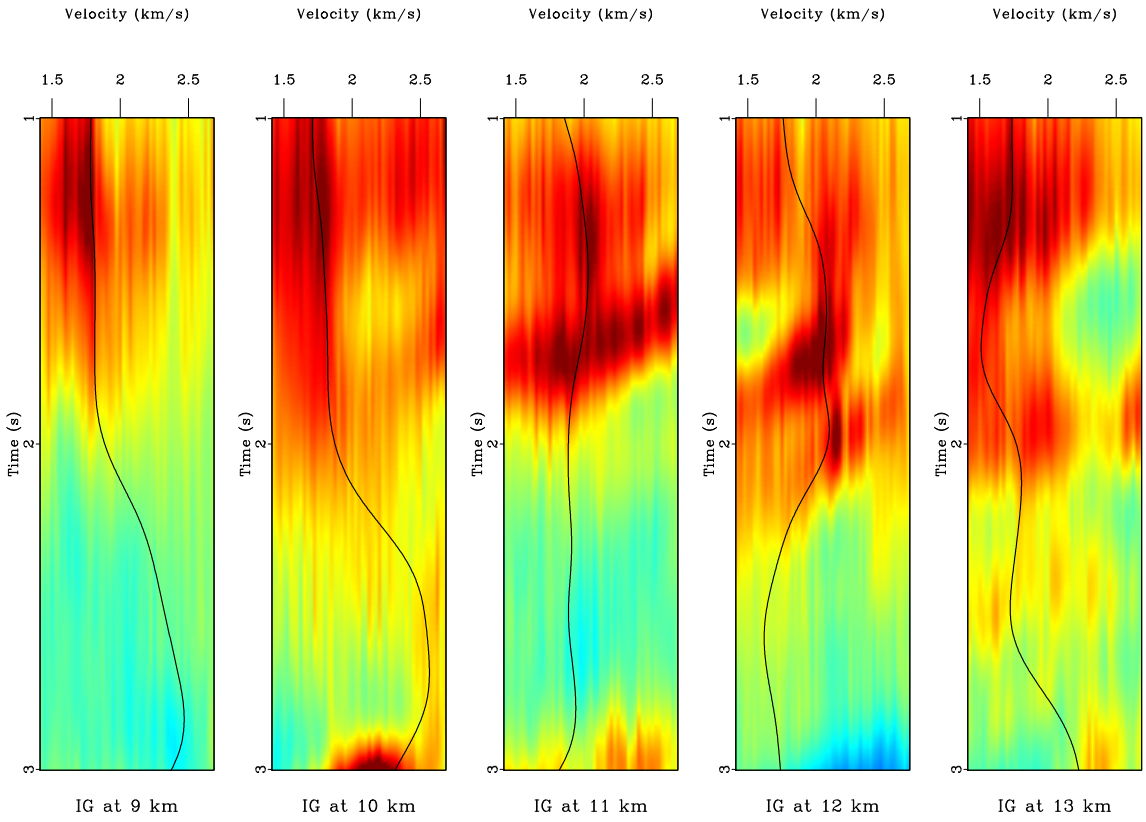


Figure 1: Focusing image gathers (FIG) for post-stack migration velocity analysis by diffraction focusing. Red colors indicate strong focusing. Superimposed black curves are slices of the picked migration velocity shown in Figure 3(b). diffr/fault panel

The data for our first example are shown in Figure 2(a), which displays a stacked section of a vintage Gulf of Mexico dataset (4). Diffractions caused by irregular fault boundaries are preserved in the stack thanks to dip moveout processing but are hardly visible underneath strong reflection responses. Figure 2(b) shows the dominant slope of reflection events estimated by the plane-wave destruction method. Numerous diffractions were separated from reflections by plane-wave destruction and are shown in Figure 3(a).

Figure 3(b) shows the migration velocity picked from focusing common-image gathers (FIGs). Example FIGs are shown in Figure 1. Figure 4(a) is the image of diffracted events, which collapse to collectively form fault surfaces. Figure 4(b) is the

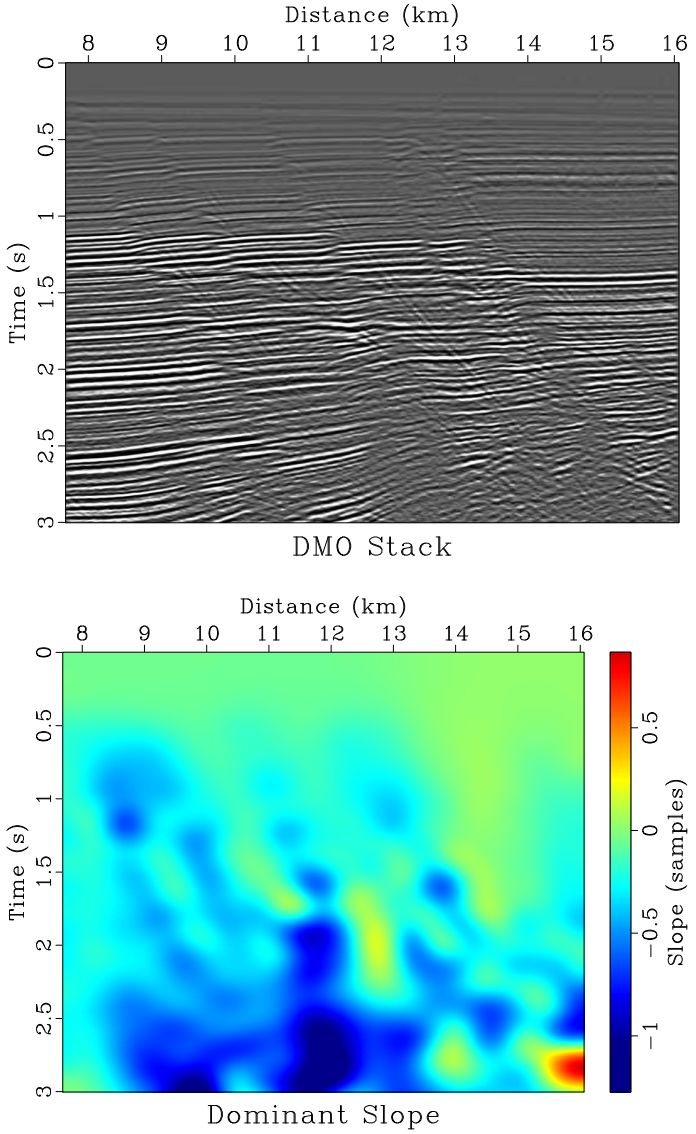


Figure 2: First test example. (a) Stacked section from a Gulf of Mexico dataset. (b) Local slopes estimated by plane-wave destruction. diffr/fault bei-stk2,bei-dip

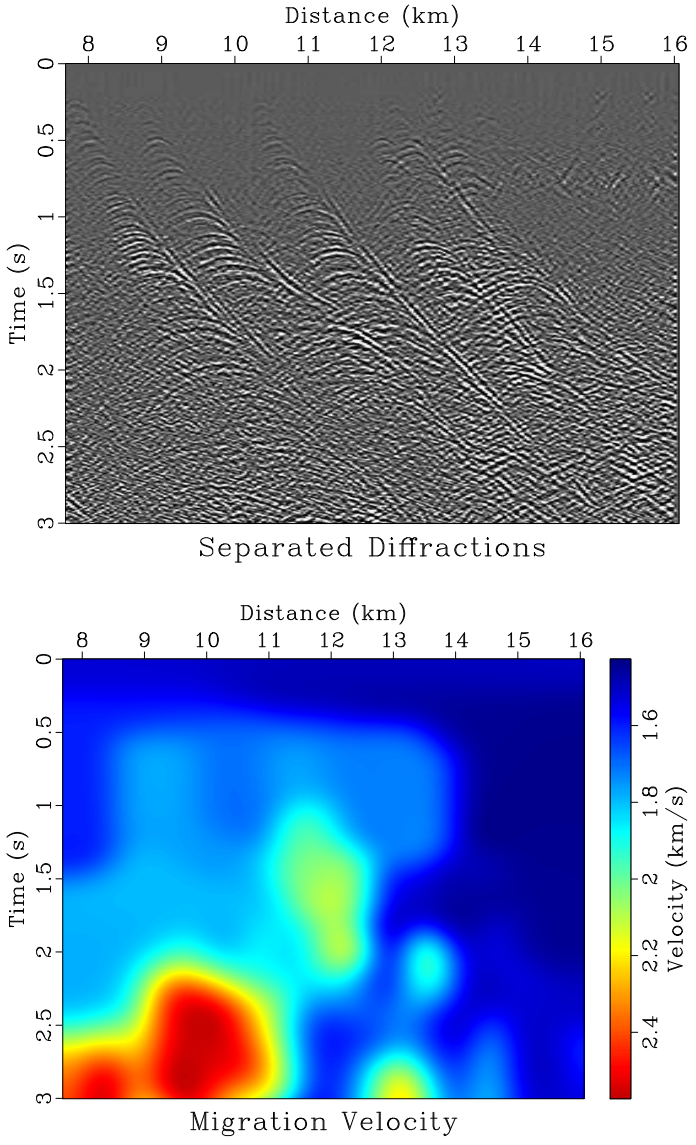


Figure 3: Diffraction separation. (a) Diffraction events separated from data in Figure 2(a). (b) Migration velocity picked from local varimax scans after velocity continuation of diffractions. diffr/fault bei-pwd,bei-pik

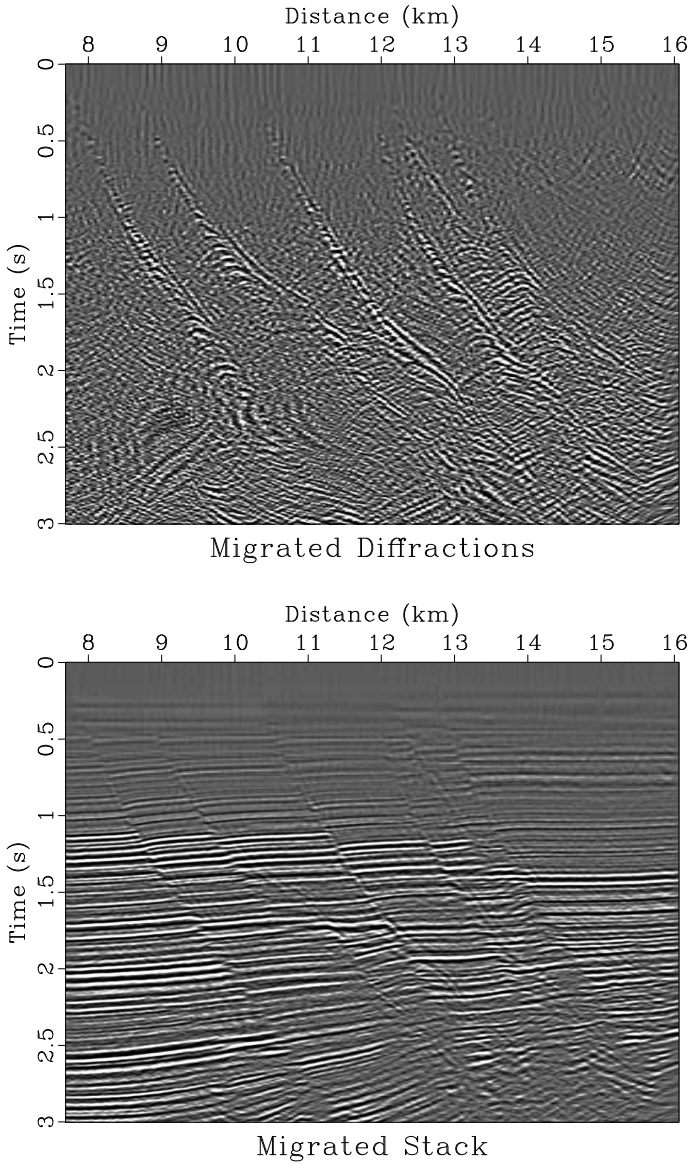


Figure 4: Migrated images. (a) Migrated diffractions from Figure 3(a). (b) Initial data from Figure 2(a) migrated with velocity estimated by diffraction imaging. diffr/fault bei-slc,bei-slc2

image obtained by migrating the original stack with velocities estimated from diffraction focusing analysis. In this final image, fault surfaces align with discontinuities in seismic reflectors. The image compares favorably with images of the same dataset from the conventional processing shown by (4).

Salt detection

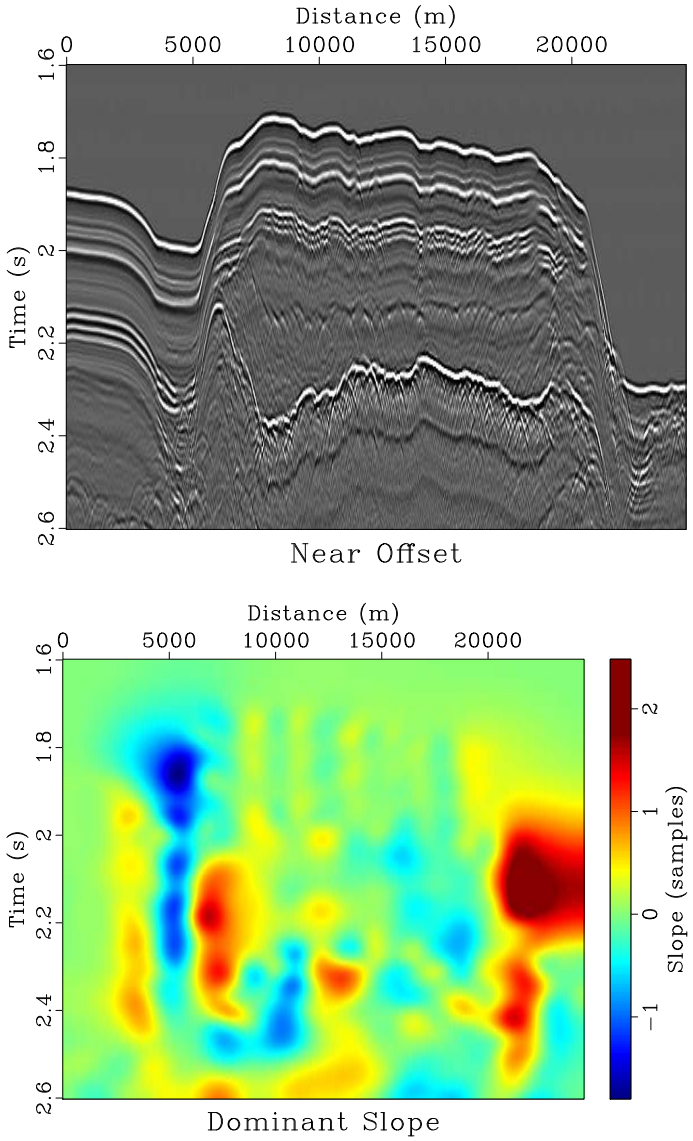


Figure 5: Second test example. (a) Near-offset section from a Gulf of Mexico dataset. (b) Local slopes estimated by plane-wave destruction. diffr/gom gom,gom-dip

Figure 5(a) shows another example, also from the Gulf of Mexico. We used the nearest-offset section for diffraction analysis. Plane-wave destruction estimates dominant slopes of continuous reflection events [Figure 5(b)] and reveals numerous

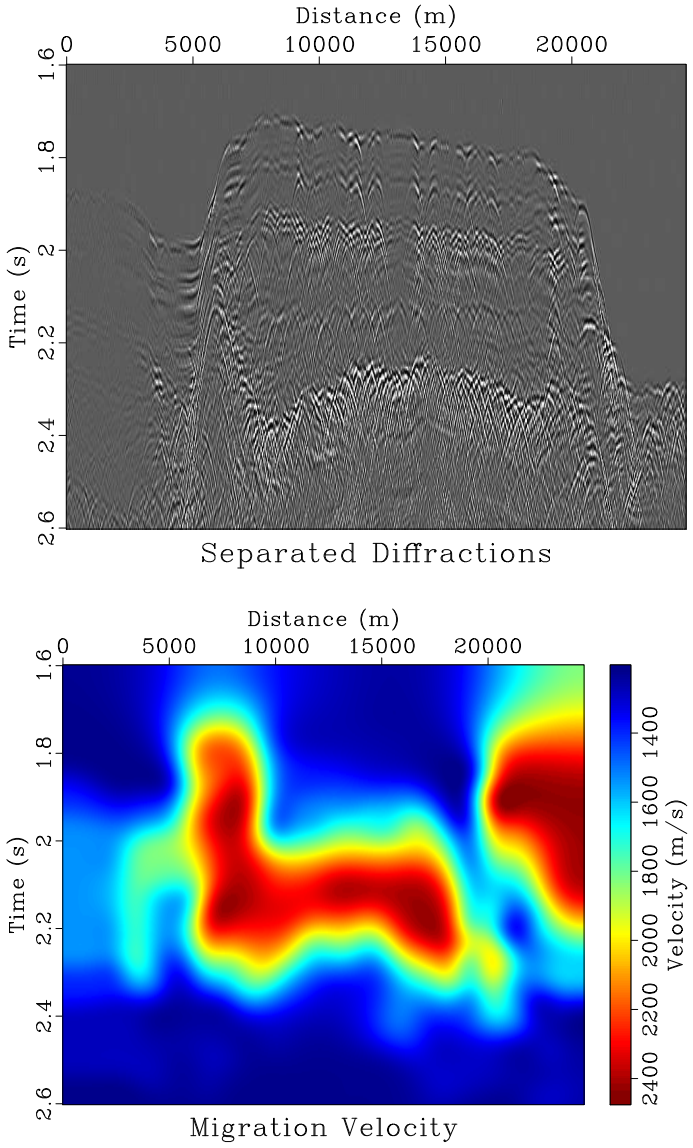


Figure 6: Diffraction separation. (a) Diffraction events separated from data in Figure 5(a). (b) Migration velocity picked from local varimax scans after velocity continuation of diffractions. diffr/gom gom-pwd,gom-pik

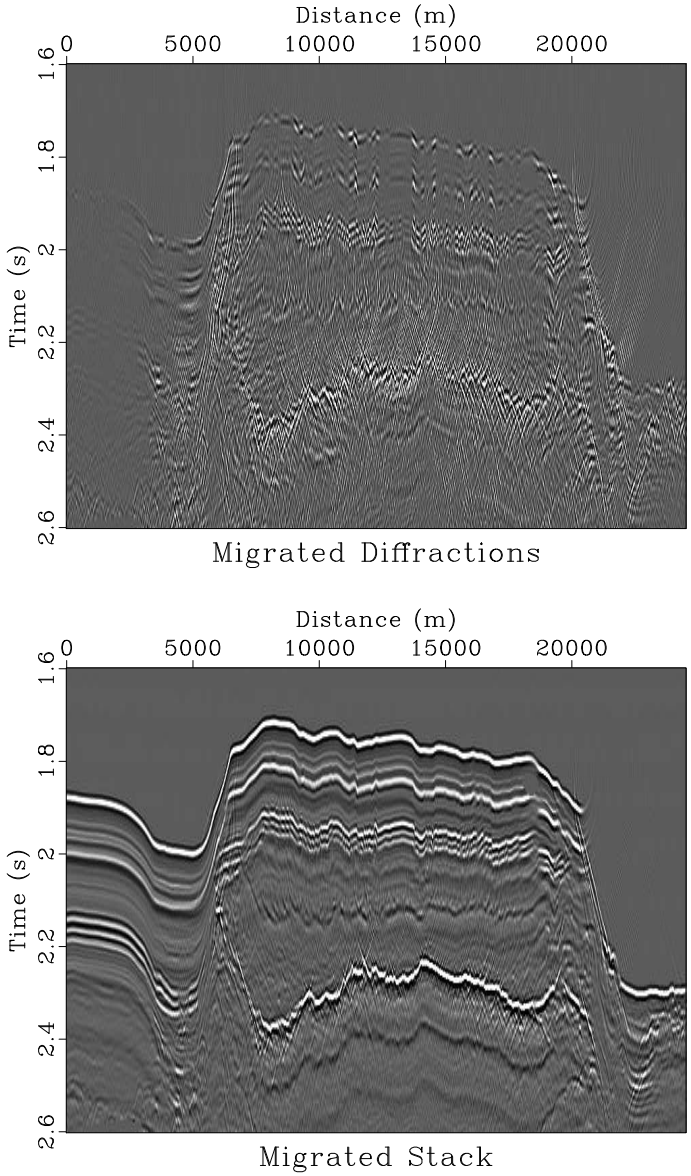


Figure 7: Migrated images. (a) Migrated diffractions from Figure 6(a). (b) Initial data from Figure 5(a) migrated with velocity estimated by diffraction imaging. diffr/gom gom-slc,gom-slc2

diffractions generated by rough edges of a salt body [Figure 6(a)]. We used shaping regularization (9) with the smoothing radius of 40-by-10 samples to constrain the slope-estimation process. Focusing analysis generates a time migration velocity [Figure 6(b)] suitable for collapsing diffractions [Figure 7(a)]. Both sharp edges of the salt body and continuous specular reflections appear in the final image [Figure 7(b)]. Inevitably, prestack depth migration (as opposed to time migration) is required to properly position the salt boundary in depth. Time migration, however, provides a reasonable first-order approximation computed at a small fraction of the cost.

Channel detection

The third example is a 3-D synthetic dataset. The velocity model was designed to simulate a complex sand channel geometry in a deep-water clastic reservoir (Figure 8(a)). Including an overburden with stochastically generated velocity fluctuations on top of the reservoir model, we generated zero-offset data shown in Figure 8(b). The data contain reflections from continuous parts of the model and numerous diffractions generated by the channel edges. Separating diffractions using in-line plane-wave destruction (Figure 9), we compare depth-migrated images of the original data and of the separated diffractions (Figure 10). The fine details of the stacked channel geometry are revealed by diffraction imaging.

CONCLUSIONS

We have developed a method of efficient migration velocity analysis based on separation and imaging of seismic diffractions. The efficiency follows from the fact that the proposed analysis is applied in the post-stack domain as opposed to the conventional prestack velocity analysis. We used continuity of dominant reflections in the zero-offset or stacked sections as a criterion for separating reflections from diffractions. We then imaged separated diffractions using local focusing analysis for picking optimal migration velocities. A prestack extension of our approach was presented by (24).

ACKNOWLEDGMENTS

We thank TOTAL for partially supporting this work. The field datasets used in this study were released by Western Geophysical, and the synthetic reservoir model was created by Jim Jennings at the Bureau of Economic Geology in collaboration with Chevron. We are grateful to Ken Larner for a thorough and helpful review.

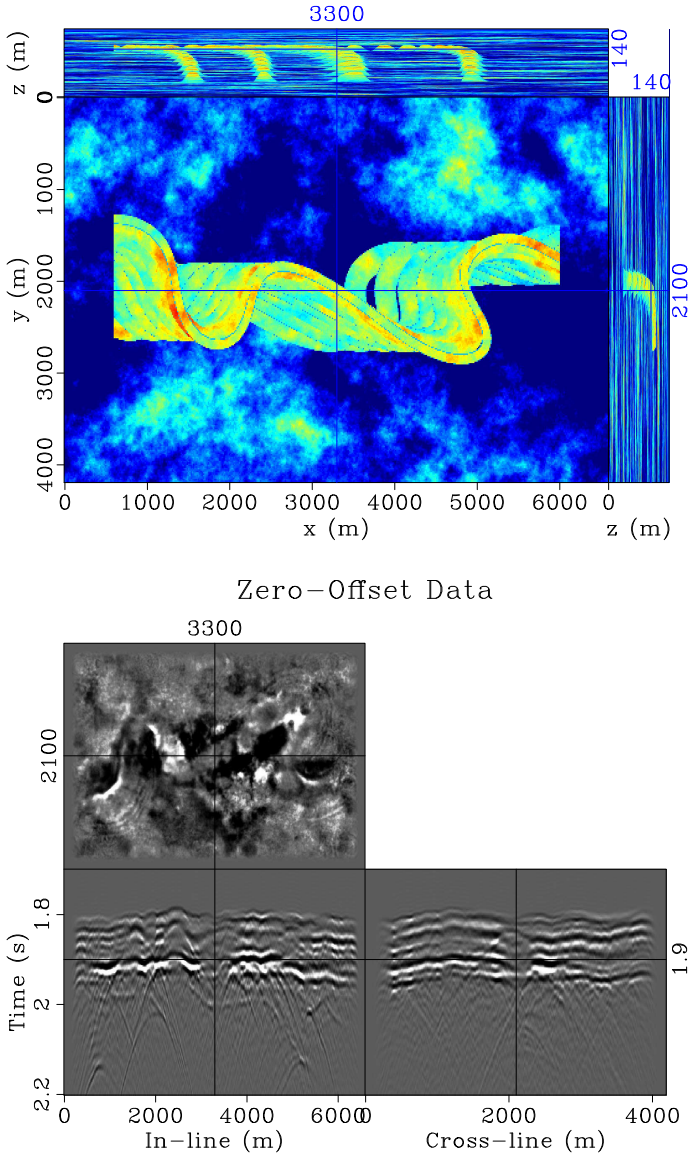


Figure 8: 3-D synthetic test. (a) Synthetic velocity model for a channelized reservoir. (b) Modeled zero-offset data. diffr/chan3d zovel,zovol

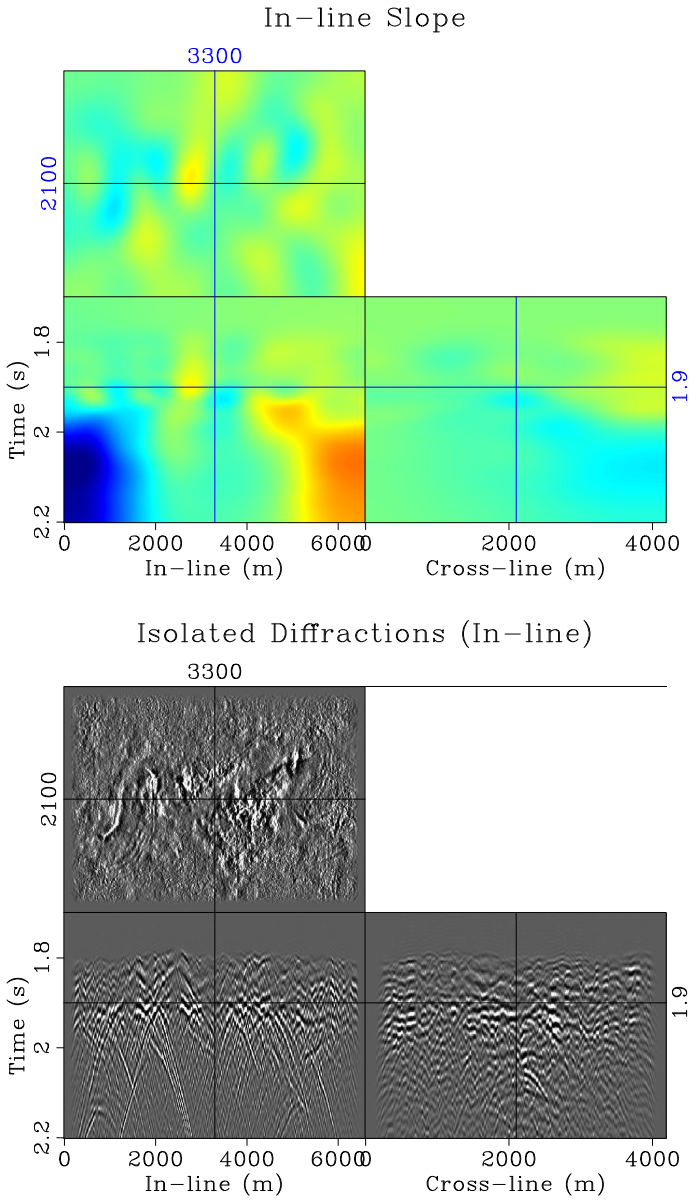


Figure 9: Diffraction separation for the 3-D synthetic test from Figure 8. (a) Dominant in-line slope estimated by plane-wave destruction. (b) Diffractions separated from the data. diffr/chan3d xdip,xpwd

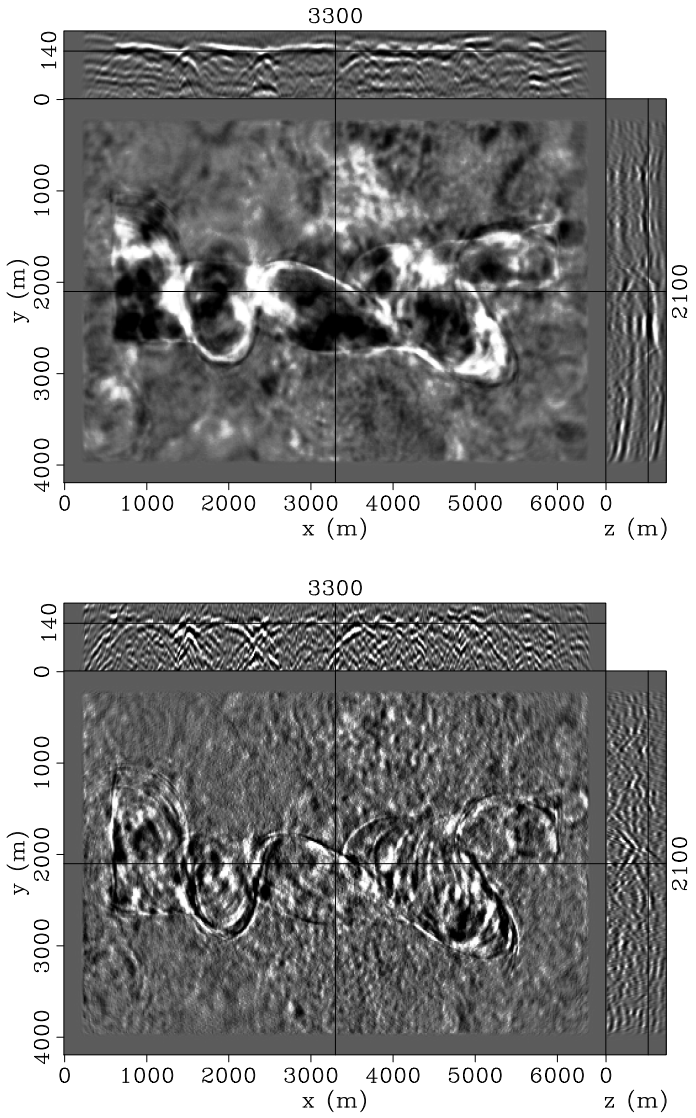


Figure 10: Depth migration of the 3-D synthetic test data. (a) Migrated data. (b) Migrated diffractions. diffr/chan3d zomig,pwmig

REFERENCES

[1] Bahorich, M., and S. Farmer, 1995, 3-D seismic discontinuity for faults and stratigraphic features: The coherence cube: The Leading Edge, **14**, 1053–1058. (Discussion with reply by author in TLE-15-3-172).

[2] Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.

[3] ———, 1994, Applications of two- and three-dimensional filtering: 64th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1572–1575.

[4] ———, 2005, Basic Earth Imaging: Stanford Exploration Project.

[5] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

[6] ———, 2003a, Time-migration velocity analysis by velocity continuation: Geophysics, **68**, 1662–1672.

[7] ———, 2003b, Velocity continuation and the anatomy of residual prestack time migration: Geophysics, **68**, 1650–1661.

[8] ———, 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.

[9] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[10] Fowler, P., 1984, Velocity independent imaging of seismic reflectors: 54th Ann. Internat. Mtg, Soc. of Expl. Geophys., Session:S1.8.

[11] Harlan, W. S., J. F. Claerbout, and F. Rocca, 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, 1869–1880.

[12] Hubral, P., M. Tygel, and J. Schleicher, 1996, Seismic image waves: Geophysical Journal International, **125**, 431–442.

[13] Khaidukov, V., E. Landa, and T. Moser, 2004, Diffraction imaging by focusing-defocusing: An outlook on seismic superresolution: Geophysics, **69**, 1478–1490.

[14] Klem-Musatov, K., 1994, Theory of Seismic Diffractions: Soc. of Expl. Geophys. (Edited by Hron F. Fronta and Larry Lines).

[15] Landa, E., and S. Keydar, 1998, Seismic monitoring of diffraction images for detection of local heterogeneities: Geophysics, **63**, 1093–1100.

[16] Larner, K., and C. Beasley, 1987, Cascaded migrations - Improving the accuracy of finite-difference migration: Geophysics, **52**, 618–643. (Errata in GEO-52-8-1165).

[17] Levy, S., and D. W. Oldenburg, 1987, Automatic phase correction of common-midpoint stacked data: Geophysics, **52**, 51–59.

[18] Marfurt, K. J., R. L. Kirlin, S. L. Farmer, and M. S. Bahorich, 1998, 3-D seismic attributes using a semblance-based coherency algorithm: Geophysics, **63**, 1150–1165.

[19] Mikulich, W., and D. Hale, 1992, Steep-dip v(z) imaging from an ensemble of Stolt-like migrations: Geophysics, **57**, 51–59.

[20] Sava, P. C., B. Biondi, and J. Etgen, 2005, Wave-equation migration velocity analysis by focusing diffractions and reflections: Geophysics, **70**, U19–U27.

[21] Schwab, M., J. Claerbout, and C. Holden, 1996, Revealing geological discontinuities by plane reflector suppression: 66th Ann. Internat. Mtg, Soc. of Expl. Geophys., 302–305.

[22] Soellner, W., and W. Yang, 2002, Diffraction response simulation: A 3D velocity inversion tool: 72nd Ann. Internat. Mtg, Soc. of Expl. Geophys., 2293–2296.
[23] Symes, W. W., and J. J. Carazzone, 1991, Velocity inversion by differential semblance optimization: Geophysics, **56**, 654–663.
[24] Taner, M. T., S. Fomel, and E. Landa, 2006, Prestack separation of seismic diffractions using plane-wave decomposition: 76th Ann. Internat. Mtg, Soc. of Expl. Geophys., 2401–2404.
[25] Wiggins, R., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21–25.

Stacking seismic data using local correlation

Guochang Liu[*†], *Sergey Fomel*[†], *Long Jin*[‡], *Xiaohong Chen*[**]

ABSTRACT

Stacking plays an important role in improving signal-to-noise ratio and imaging quality of seismic data. However, for low-fold-coverage seismic profiles, the result of conventional stacking is not always satisfactory. To address this problem, we have developed a method of stacking in which we use local correlation as a weight for stacking common-midpoint gathers after NMO processing or common-image-point gathers after prestack migration. Application of the method to synthetic and field data showed that stacking using local correlation can be more effective in suppressing random noise and artifacts than other stacking methods.

INTRODUCTION

Stacking as one of the three crucial techniques (deconvolution, stacking, and migration) plays an important role in improving signal-to-noise ratio (S/R) in seismic data processing (22). Conventional stacking, which is performed by averaging an NMO-corrected data set or migrated data set, is optimal only when noise components in all traces are uncorrelated, normally distributed, stationary, and of equal magnitude (14; 15). Therefore, different stacking technologies have been proposed, along with improvements in optimizing weights of seismic traces.

(18) proposes using an S/N-based weighted stack to further minimize noise. Using cross-correlation of seismic traces and normalized cross-correlation processing, (4) proposes preserved frequency stacking. (19) proposes optimum weighted stack for multiple suppression, with weight determined by solving a set of optimization equations. (15) propose a stack-and-denoise method called SAD, which exploits the structure of seismic signals to obtain an enhanced stack. (23) present a high-order correlative weighted stacking technique on the basis of wavelet transformation and high-order statistics. By estimating the probability distribution of noise, (21) applies a maximum-likelihood estimator to stacking. To eliminate artifacts in angle-domain common-image gathers (CIGs) caused by sparsely sampled wavefields, (20) presents a selective stacking approach that applies local smoothing of envelope function to achieve the weighting function. (17) proposes smart stacking, which is based on optimizing seismic amplitudes of the stacked signal by excluding harmful samples from the stack and applying larger weight to the central part of the sample population.

[*]**e-mail:** sergey.fomel@beg.utexas.edu

The global correlation coefficient can measure the similarity of two signals, but it is not a local attribute. Not only does the sliding-window global-correlation approach need many parameters to be specified, but this approach cannot smoothly characterize thin layers well. (8) uses shaping regularization, which controls locality and smoothness to define local correlation (7). Local correlation is applied to multicomponent seismic image registration (9; 7) and time-lapse image registration (10).

In this paper,we present a new stacking method using local correlation. This method applies time-dependent smooth weights (which are taken as local correlation coefficients between reference traces and prestack traces), stacks the common-midpoint (CMP) gather, and effectively discards parts of the data that least contribute to stacked reflection signals. Using synthetic and field data examples, we show that, compared with other stacking methods, this method can greatly improve the S/N and suppress artifacts.

METHODOLOGY

Review of local correlation

The global uncentered correlation coefficient between two discrete signals $\mathbf{a}_i$ and $\mathbf{b}_i$ can be defined as the functional

$$\gamma = \frac{\sum_{i=1}^{N} \mathbf{a}_i \mathbf{b}_i}{\sqrt{\sum_{i=1}^{N} \mathbf{a}_i^2 \sum_{i=1}^{N} \mathbf{b}_i^2}}, \tag{1}$$

where N is the length of a signal. The global correlation in equation 1 supplies only one number for the whole signal. For measuring the similarity between two signals locally, one can define the sliding-window correlation coefficient

$$\gamma_w(t) = \frac{\sum_{i=t-w/2}^{t+w/2} \mathbf{a}_i \mathbf{b}_i}{\sqrt{\sum_{i=t-w/2}^{t+w/2} \mathbf{a}_i^2 \sum_{i=t-w/2}^{t+w/2} \mathbf{b}_i^2}}, \tag{2}$$

where w is window length.

(7) proposes the local correlation attribute that identifies local changes in signal similarity in a more elegant way. In a linear algebra notation, the correlation coefficient in equation 1 can be represented as a product of two least-squares inverses γ_1

and γ_2:

$$\gamma^2 = \gamma_1\gamma_2 \ , \tag{3}$$

$$\gamma_1 = \arg\min_{\gamma_1} \parallel \mathbf{b} - \gamma_1\mathbf{a} \parallel^2 = (\mathbf{a}^T\mathbf{a})^{-1}(\mathbf{a}^T\mathbf{b}) \ , \tag{4}$$

$$\gamma_2 = \arg\min_{\gamma_2} \parallel \mathbf{a} - \gamma_1\mathbf{b} \parallel^2 = (\mathbf{b}^T\mathbf{b})^{-1}(\mathbf{b}^T\mathbf{a}) \ , \tag{5}$$

where $\mathbf{a}$ and $\mathbf{b}$ are vector notions for $\mathbf{a}_i$ and $\mathbf{b}_i$. Let $\mathbf{A}$ and $\mathbf{B}$ be two diagonal operators composed of the elements of a and b. Localizing equations 4 and 5 amounts to adding regularization to inversion. Using shaping regularization (8), scalars γ_1 and γ_2 turn into vectors $\mathbf{c}_1$ and $\mathbf{c}_2$, defined as

$$\mathbf{c}_1 = [\lambda^2\mathbf{I} + \mathbf{S}(\mathbf{A}^T\mathbf{A} - \lambda^2\mathbf{I})]^{-1}\mathbf{S}\mathbf{A}^T\mathbf{b} \ , \tag{6}$$

$$\mathbf{c}_2 = [\lambda^2\mathbf{I} + \mathbf{S}(\mathbf{B}^T\mathbf{B} - \lambda^2\mathbf{I})]^{-1}\mathbf{S}\mathbf{B}^T\mathbf{a} \ , \tag{7}$$

where λ scaling controls relative scaling of operators $\mathbf{A}$ and $\mathbf{B}$ and where $\mathbf{S}$ is a shaping operator such as Gaussian smoothing with an adjustable radius. The component-wise product of vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ defines the local correlation measure. Local correlation is a measure of the similarity between two signals. An iterative, conjugate-gradient inversion for computing the inverse operators can be applied in equations 6 and 7. Interestingly, the output of the first iteration is equivalent to the algorithm of fast local cross-correlation proposed by (13).

Stacking using local correlation

The problem of combining a collection of seismic traces into a single trace is commonly referred to as stacking in seismic data processing. This process is used to attenuate random noise and simultaneously amplify the coherent signal in the gather. Typically, the desired stacked trace is estimated by averaging traces from the CMP gather (14):

$$\bar{a}_j(t) = \frac{1}{N}\sum_{i=1}^{N} a_{i,j}(t), \qquad j = (1, 2, 3, \ldots, M) \ , \tag{8}$$

where N is the fold of the stack and $a_{i,j}(t)$ is the sample value on trace i from the jth CMP gather at two-way time t. Such a technique provides the optimal unbiased estimate of $\bar{a_j}(t)$. (18) proposes weighted stacking of seismic data:

$$\bar{a}_j(t) = \frac{1}{\sum_{i=1}^{N} w_{i,j}}\sum_{i=1}^{N} w_{i,j}\cdot a_{i,j}(t), \qquad j = (1, 2, 3, \ldots, M) \ , \tag{9}$$

where $w_{i,j}$ denotes the weight of the ith trace from the jth CMP gather, which is determined by noise variances $w_{i,j} = 1/\sigma^2_{i,j}$. However, it is difficult to estimate noise variances reliably in practice. (15) use an iterative algorithm called "leave me out"

(LMO) to estimate noise variances from data. The desired signal is assumed to be flat with constant amplitude across all the traces within a gather in the LMO method.

For using time-dependent smooth weights in the stacking process, we choose the local correlation coefficient from the previous section as weights to stack seismic data. We find that local correlation better characterizes local similarity between prestack and reference traces than the sliding-window method.

Consider the two noisy signals with Gaussian random noise but different noise levels in Figure 1c and 1d. The signals are derived from adding noise on convolution of the Ricker wavelet (Figure 1a) with synthetic reflectivity (Figure 1b). The distribution of noise in (Figure 1c) is $N(\mu, \sigma) = N(0, 10^{-6})$, where μ and σ are mean and variance of noise, respectively. The distribution of noise in (Figure 1d) is $N(\mu, \sigma) = N(0, 0.07)$. The sliding-window correlation and local correlation between Figure 1c and Figure 1d are shown in Figure 1e and Figure 1f, respectively. Note that local correlation coefficients (Figure 1f) are smooth and better distinguish the thin layer, represented by the first two reflectivities in Figure 1b. In application to stacking, the prestack trace is analogous to Figure 1d with larger variance noise, and the reference trace is analogous to Figure 1c with smaller variance noise.

To implement seismic data stacking using local correlation, we first apply the equal-weight stack to the NMO-corrected CMP gather to obtain the reference trace. Then we compute the local correlation coefficients between the reference trace and the NMO-corrected CMP gather and apply soft thresholding (6) to all local correlation coefficients. Finally, we apply the weighted stack to the CMP gather using local correlation to get the final stacked trace. Mathematically, stacking using the local correlation approach modifies equation 9 to

$$\bar{a}_j(t) = \frac{1}{K_j H_j(t)} \sum_{j=1}^{N} w_{i,j} \cdot a_{i,j}(t), \qquad j = (1, 2, 3, \ldots, M) \ , \tag{10}$$

$$w_{i,j}(t) = \begin{cases} \eta_{i,j}(t) - \varepsilon, & \eta_{i,j} > \varepsilon \\ 0, & \eta_{i,j} \le \varepsilon \end{cases} \ , \tag{11}$$

where ε is the threshold value, $K_j = \sum_{t=0}^{t} \sum_{i=0}^{N} w_{i,j}(t)$ is the sum of weights of the jth CMP gather, $H_j(t)$ is the number of samples with $w_{i,j} \cdot a_{i,j}(t) \neq 0$ for a given two-way time, and $\eta_{i,j}(t)$ is the local correlation between the ith prestack trace from jth gather and the reference trace. The local correlation $\eta_{i,j}(t)$ can be computed using equations 6 and 7. The reference trace is derived from averaging all the traces of one CMP gather. Here we assume that the equal-weight stacked trace is close to the desired trace. Because the weights are a function of two-way traveltime and offset, recovery scalar K_j has the same value for the same CMP gather. Meanwhile, the samples with $w_{i,j}(t) \cdot a_{i,j} \neq 0$ at a given two-way time are assumed to be full noise or null value such as muting parts; we therefore use $H_j(t)$ to scale the stacked trace.

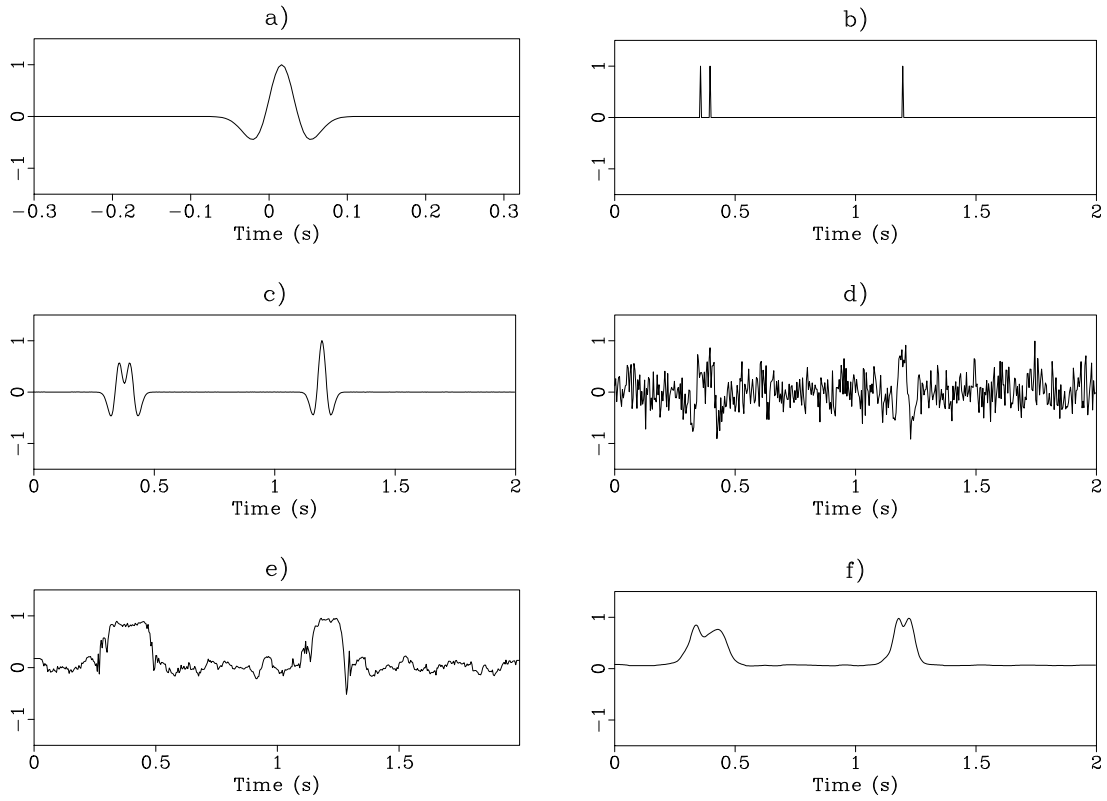


Figure 1: (a) Zero-phase Ricker wavelet. (b) Reflection coefficient. (c) Noisy signal with $N(\mu, \sigma) = N(0, 10^{-6})$. (d) Noisy signal with $N(\mu, \sigma) = N(0, 0.07)$. (e) Sliding-window correlation. (f) Local correlation. simistack/similarity ricker,ref,signal1,signal2,wcorr,simi

Changes occurring between equation 9 and equations 10 and 11 result from time-dependent smooth weights for the stack and application of thresholding to the weights. All local correlation coefficients below a specified threshold are discarded, and the rest, with values above the threshold, are included. We thus stack only those parts of prestack data whose similarity to the reference trace is comparatively large. Equations 10 and 11 implicitly estimate the noise variance by analyzing local cross-correlations between prestack trace and the reference trace. This operation can be understood as a nonlinear filter that enhances the coherency of events. We perform this operation for all gathers using this method to improve the stack profile.

When applied after angle-domain migration, normalization provided by soft thresholding is analogous to true-amplitude illumination compensation from the so-called Beylkin determinant (1; 3). Local correlation normalizes the image by the number of strongly illuminated angles in angle-domain CIGs.

In the following, we discuss the distinctions between seismic stacking using local correlation and other methods. Our method creates time-dependent smooth weights without depending on sliding windows, as compared to other weighted stacking methods such as statistically optimal stacking (18; 15) and the semblance method (22). In contrast to smart stacking, proposed by (17) and based on sign difference between sample point and the alpha-trimmed mean to remove frequency distortions, our method applies waveform similarity between prestack trace and mean to make the stacking selective.

EXAMPLES

To illustrate the proposed method using synthetic and field data, we apply our approach to three examples. The first example is a simple case involving a fivefold prestack gather (Figure 2a) with a timeshifted-upward trace, which might be distortion by poor static correction. The peak of the signal in this gather is one.We add Gaussian random noise with distribution $N(\mu, \sigma) = N(0, 0.01)$ on the five traces. The result of an equal-weight stack is shown in Figure 2c. The upside wing in Figure 2c is distorted because of the first time-shift trace. Then we use three weighted stacking methods to stack the five traces.

Figure 2d and Figure 2e illustrates results of smart stacking (17) and LMO-based weighted stacking in which the weights are computed by the LMO method (18; 15). Figure 2f shows the result of stacking using local correlation with weights (Figure 2b) determined by the similarity between the prestack trace (Figure 2a) and the reference (Figure 2c). Because the waveform in the first trace in Figure 2a is most likely noise or artifact, it is reasonable that the weight in the stack procedure is lower. Use of local correlation as weights of prestack traces lets us select those portions, which are more similar to the reference trace to contribute to the stack.

Comparing the three methods, one can find that smart stacking and LMO-based weighted stacking can remove upside wing distortion cleanly, but stacking using local

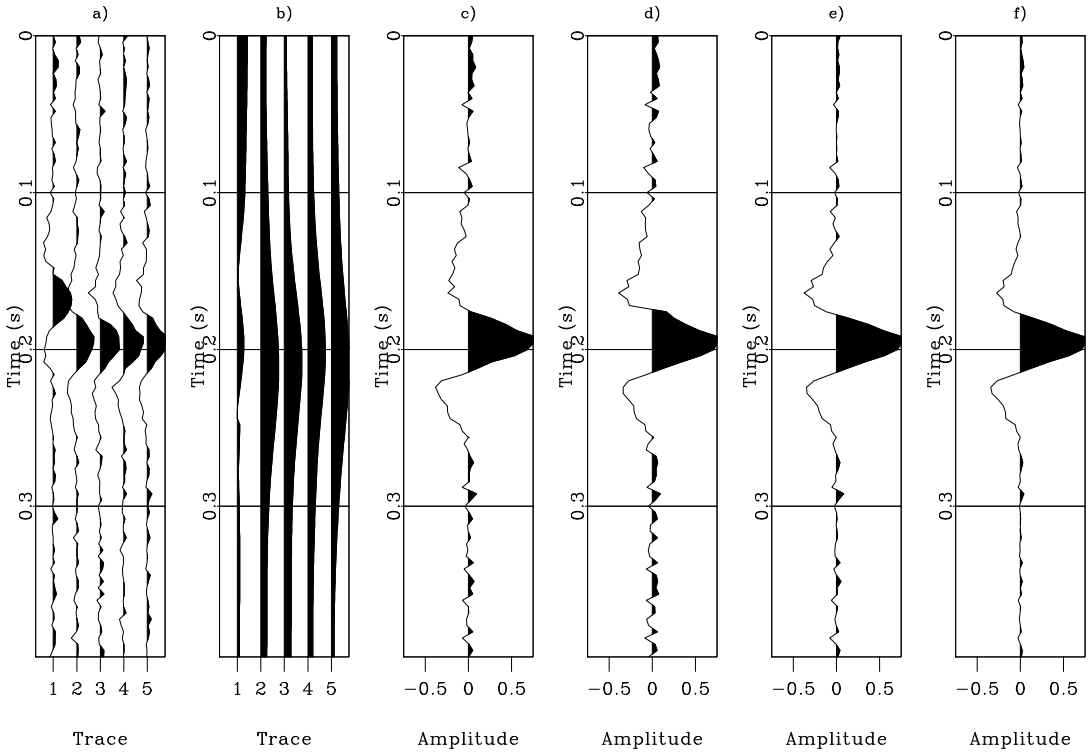


Figure 2: Simple stacking test with fivefold gather. (a) Prestack gather. (b)Weights used in local-correlation weighted stacking. (c) Conventional equal-weight stacking method (S/N=8.4). (d) Smart stacking method (S/N=9.2). (d) LMO-based weighted method (S/N=10.2). (f) Local-correlation weighted stacking (S/N=13.5). simistack/simple compare

correlation removes more random noise than the other two methods and meanwhile corrects upside wing distortion.

To judge the effect of denoising quantitatively between different methods, we apply equal-weight stacking on the last four traces without any noise to get the exact desired stacked trace $d_j(t)$, which can be regarded as a signal trace. The S/N of the jth CMP can therefore be estimated as

$$\mathrm{S/N}_j = 10\log_{10}\left(\frac{\sum_t [d_j^2(t)]}{\sum_t [d_j(t)-\bar{a}_j(t)]^2}\right) , \tag{12}$$

where $\bar{a}_j(t)$ is the stacked trace from different stacking methods. In the first simple example Figure 2, the S/N of equal-weight stacking is 8.4 dB and the other three weighted methods are, respectively, 9.2, 10.2, and 13.5 dB. Stacking using local correlation can improve S/N greatly.

The second example is a 2D synthetic model that includes four reflectors. Synthetic data are generated with Kirchhoff modeling. The peak of the data set is one and Gaussian random noise with distribution $N(\mu,\sigma) = N(0, 0.05)$ is added. We show the results of stacking one CMP gather (Figure 3a) by three methods in Figure 3c-e. Compared to other methods, our method is the most effective in denoising.

The stacked profile of all CMP gathers is shown in Figure 4. We use equation 12 to compute the S/N of the stacked profile (Figure 4). The S/Ns of three methods are 7.1, 9.6, 10.9 dB, respectively. Noise is attenuated more effectively in the stacking result using local correlation (Figure 4c).

The third example involves a historic 2D line from the Gulf of Mexico (5). The stacked sections, using three different methods, are shown in Figure 5. Figure 6 shows the local correlation cube between prestack and reference traces. Similar cubes have been used in multicomponent seismic image registration (9; 7) and time-lapse image registration (10). For synthetic data, the exact desired stacked section can be calculated by stacking prestack traces without any noise. But for field data, the S/N is difficult to estimate using equation 12. We therefore use singular value decomposition (SVD) (2) to evaluate different stacking methods. The SVD of stacked section matrix gives

$$\mathbf{M} = \mathbf{U}\Sigma\mathbf{V}^T . \tag{13}$$

The diagonal elements σ_r of Σ are the singular values of $\mathbf{M}$. The S/N can be estimated as (11; 16; 12)

$$\mathrm{S/N} = 10\log_{10}\left(\frac{\sigma_1^2 - \frac{1}{R-1}\sum_{r=2}^{R}\sigma_r^2}{\frac{1}{R-1}\sum_{r=2}^{R}\sigma_r^2}\right) , \tag{14}$$

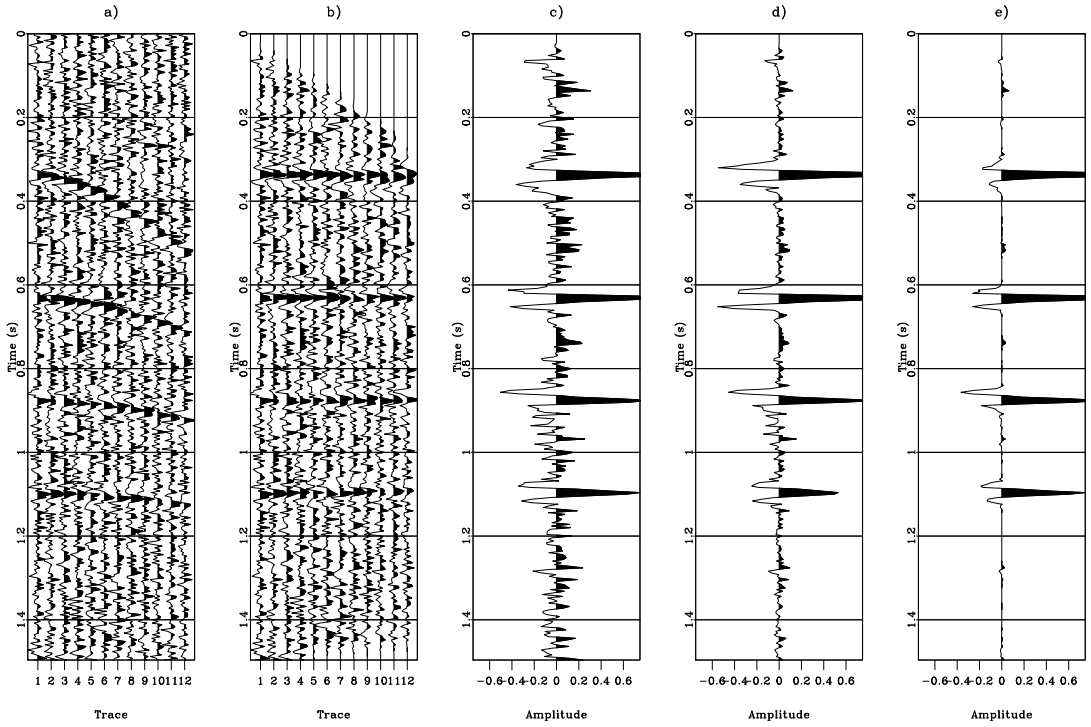


Figure 3: (a) One CMP gather from synthetic data set. (b) NMO-corrected gather. (c) Result of conventional equal-weight stacking. (d) Result of LMO-based weighted stacking. (e) Result of local-correlation weighted stacking. simistack/flat4 onestack1

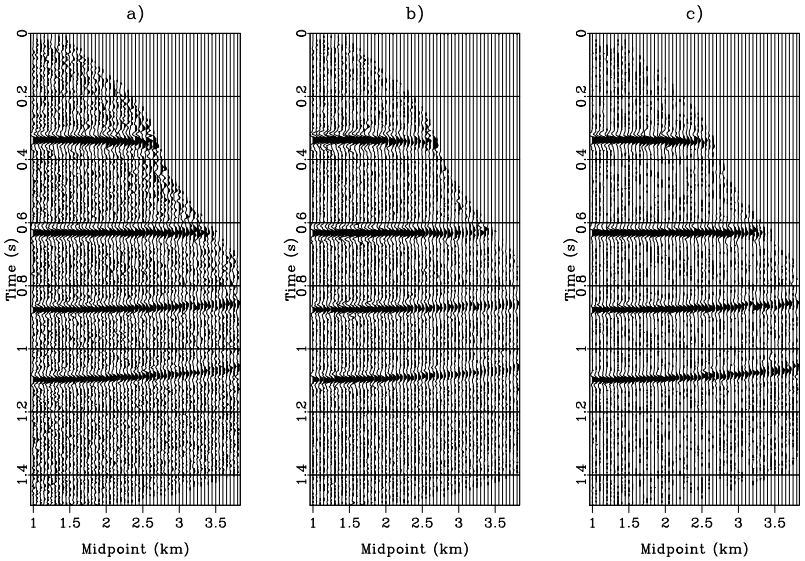


Figure 4: Comparison among three stacking methods including all synthetic CMP gathers. (a) Conventional equal-weight stacking (S/N=7.1). (b) LMO-based weighted stacking (S/N=9.6). (c) Local-correlation weighted stacking (S/N=10.9). simistack/flat4 stackss

where R is the number of all singular values. The S/Ns of stacked sections resulting from three stacking methods are, respectively, 27.4, 29.2, and 33.9 dB. Comparing Figure 5a-c, we can find also that random noise is attenuated and coherent reflections are enhanced better using local correlation (e.g., 0.5–1.5-s range).

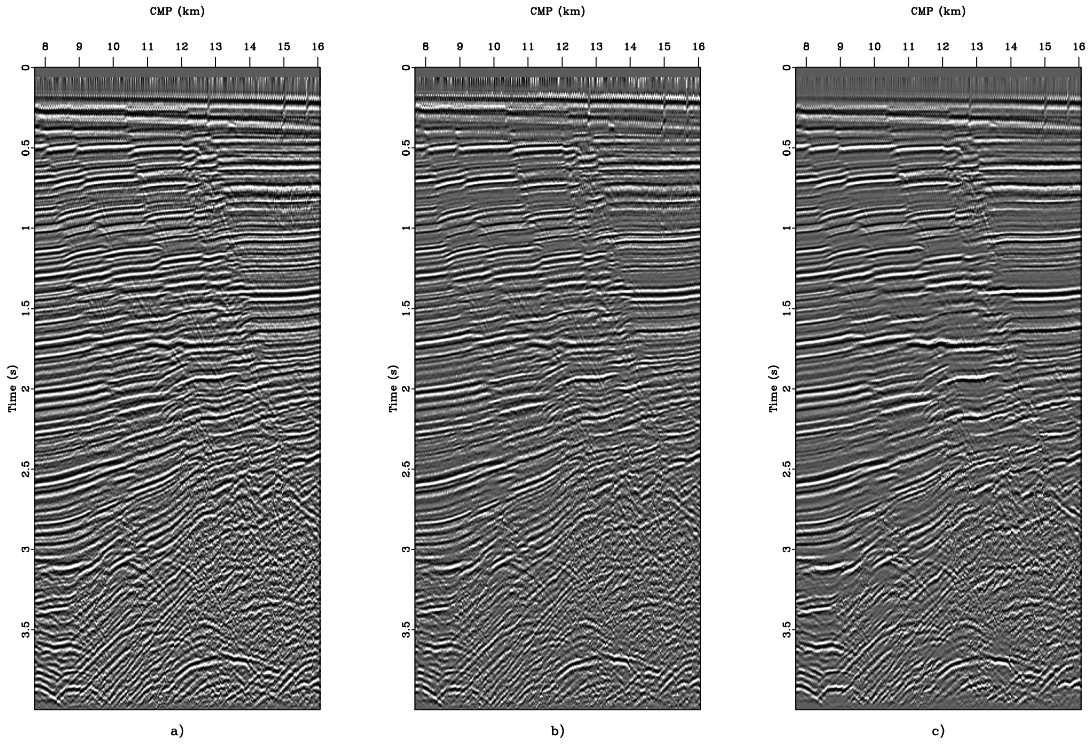


Figure 5: Results of (a) conventional equal-weight stacking (S/N=27.4), (b) LMO-based weighted stacking (S/N=29.2) and (c) local-correlation weighted stacking (S/N=33.9). simistack/bei field

CONCLUSION

We have developed a new method of stacking NMO-corrected or migrated seismic data using local correlation. We substitute local correlation for the weight value in statistically weighted stacking and then use soft thresholding to make the stacking selective. Because weights are derived from the input data, our method can be regarded as a nonlinear filter. Synthetic and field data examples confirm that our approach can be significantly more effective than other weighted stacking methods in improving S/N and suppressing distortions resulting from prestack processing. Seismic stacking using local correlation can give a poor result if the quality of the reference trace is very poor. Because the coherency enhancement from local correlation is not based on physics, one should use this approach with caution when aiming to preserve physically meaningful amplitudes.

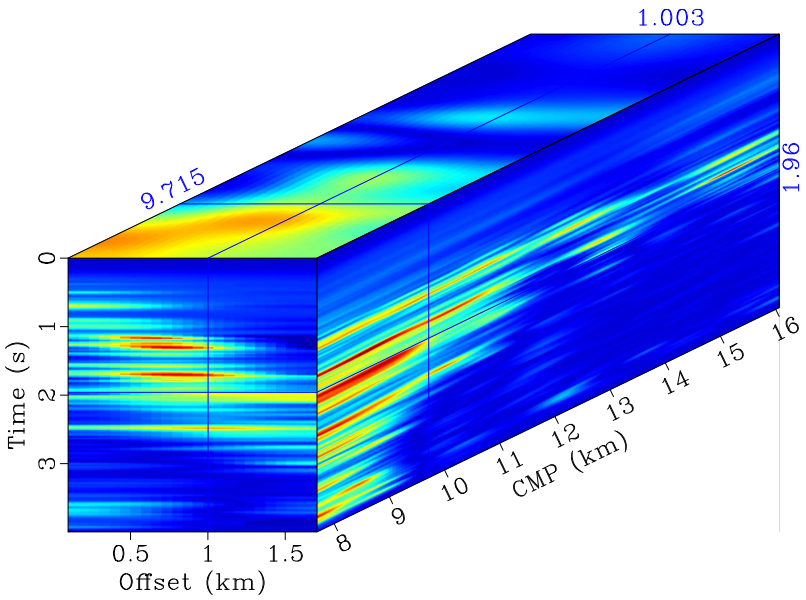


Figure 6: Local correlation cube of the field-data example. simistack/bei weight

ACKNOWLEDGMENTS

We would like to thank Yang Liu and Jingye Li for inspiring discussions and three reviewers for helpful suggestions. G. Liu would like to thank the China Scholarship Council (grant 2007U44003) for partial support of this work. X. Chen's research was supported partially by the National Basic Research Program of China (973 program, grant 2007CB209606) and by the National High Technology Research and Development Program of China (863 program, grant 2006AA09A102-09). Publication was authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Albertin, U., H. Jaramillo, D. Yingst, R. Bloor, W. Chang, C. Beasley, and E. Mobley, 1999, Aspects of true amplitude migration: 69th Annual International Meeting, SEG, Expanded Abstracts, 1358–1361.

[2] Andrews, H. C., and C. L. Patterson, 1976, Singular value decomposition and digital image processing: IEEE Transactions on Acoustics, Speech, and Signal Processing, **24**, 26–53.

[3] Audebert, F. L., and P. Froidevaux, 2005, Regularization of illumination in angle domains – a key to true amplitude migration: The Leading Edge, **24**, 643–654.

[4] Chang, X. W., Y. J. Guang, C. M. Zhang, and G. Q. Xu, 1996, Stacking based on correlation saving frequency: OilGeophysics Prospecting (in Chinese), **31**, 110–113.

[5] Claerbout, J. F., 2005, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

[6] Donoho, D. L., 1995, De-noising by soft-thresholding: IEEE Transactions on Information Theory, **41**, 613–627.
[7] Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.
[8] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.
[9] Fomel, S., M. Backus, K. Fouad, B. Hardage, and G. Winters, 2005, Amultistep approach to multicomponent seismic image registration with application to a West texas carbonate reservoir study: 75th Annual International Meeting, SEG, Expanded Abstracts, 1018–1021.
[10] Fomel, S., and L. Jin, 2007, Time-lapse image registration using the local similarity attribute: 77th Annual International Meeting, SEG, Expanded Abstracts, 2979–2983.
[11] Freire, S. L. M., and T. J. Ulrych, 1988, Application of singular value decomposition to vertical seismic profiling: Geophysics, **53**, 778–785.
[12] Grion, S., and A. Mazzotti, 1998, Stacking weights determination by means of svd and crosscorrelation: 68th Annual International Meeting, SEG, Expanded Abstracts, 1135–1138.
[13] Hale, D., 2006, Fast local crosscorrelation of images: 76th Annual International Meeting, SEG, Expanded Abstracts, 3160–3163.
[14] Mayne, W. H., 1962, Common reflection point horizontal data stacking techniques: Geophysics, **27**, 927–938.
[15] Neelamani, R., T. A. Dickens, and M. Deffenbaugh, 2006, Stack-and-denoise: A new method to stack seismic datasets: 76th Annual International Meeting, SEG, Expanded Abstracts, 2827–2831.
[16] Peterson, B. J., and D. R. DeGroat, 1988, Nearly optimal, SVD-based stacking of seismic data: Maple Press.
[17] Rashed, M. A., 2008, Smart stacking: A new CMP stacking technique for seismic data: The Leading Edge, **27**, 462–467.
[18] Robinson, J. C., 1970, Statistically optimal stacking of seismic data: Geophysics, **35**, 436–446.
[19] Schoenberger, M., 1996, Optimum weighted stack for multiple suppression: Geophysics, **61**, 891–901.
[20] Tang, Y., 2007, Selective stacking in the reflection-angle and azimuth domain: 77th Annual International Meeting, SEG, Expanded Abstracts, 2320–2324.
[21] Trickett, S., 2007, Maximum-likelihood-estimation stacking: 77th Annual International Meeting, SEG, Expanded Abstracts, 2640–2643.
[22] Yilmaz, O., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: SEG.
[23] Zhang, S. Z., and Y. X. Xu, 2006, Higher-order correlative stacking for seismic data in the wavelet domain: Chinese Journal of Geophysics (in Chinese), **49**, 554–560.

Local skewness attribute as a seismic phase detector

Sergey Fomel[*] *and Mirko van der Baan*[†*]

ABSTRACT

We propose a novel seismic attribute, local skewness, as an indicator of localized phase of seismic signals. The proposed attribute appears to have a higher dynamical range and a better stability than the previously used local kurtosis. Synthetic and real data examples demonstrate the effectiveness of local skewness in detecting and correcting time-varying, locally-observed phase of seismic signals.

INTRODUCTION

Wavelet phase is an important characteristic of seismic signals. Physical causal systems exhibit minimum-phase behavior (20). For purposes of seismic interpretation, it is convenient to deal with zero-phase wavelets with minimum or maximum amplitudes centered at the horizons of interest because it leads to the highest resolution as well as more accurate estimates of both reflection times and spacings (21). Zero-phase correction is therefore a routine procedure applied to seismic images before they are passed to the interpreter (1).

It is important to make a distinction between phases of *propagating* and *locally observed* wavelets (25; 24). The former is the physical wavelet that propagates through the Earth, thereby sampling the local geology. It is subject to geometric spreading, attenuation, and concomitant dispersion. The latter is the wavelet as observed at a certain point in space and time. Its immediate shape results from its interaction (convolution) with the reflectivity of the Earth and the current shape of the propagating wavelet. For instance, a thin layer with a positive change in seismic impedance has opposite polarities of seismic reflectivities at the top and the bottom interface, which make it act like a derivative filter and generate a wavelet with the locally observed phase subjected to a 90° rotation (31). In the absence of well log information, it is usually difficult to separate unambiguously the locally observed phase from the phase of the propagating wavelet. Nevertheless, measuring the local phase can provide a useful attribute for analyzing seismic data (23; 10; 24; 30).

(14) proposed a method of phase detection based on maximization of the varimax norm or kurtosis as an objective measure of zero-phaseness. By rotating the phase and

[*]**e-mail:** sergey.fomel@beg.utexas.edu

measuring the kurtosis of seismic signals, one can detect the phase rotations necessary for zero-phase correction (22). (23) applied local kurtosis, a smoothly nonstationary measure (9), and demonstrated its advantages in measuring phase variations as compared with kurtosis measurements in sliding windows. Local kurtosis is an example of a local attribute (5) defined by utilizing regularized least-squares inversion.

In this paper, we revisit the problem of phase estimation and propose a novel attribute, *local skewness*, as a phase detector. Analogous to local kurtosis, local skewness is defined using local similarity measurements via regularized least squares. This attribute is maximized when the locally observed phase is close to zero. Advantages of the new attribute are a higher dynamical range and a better stability, which make it suitable for picking phase corrections. Using synthetic and field-data examples, we demonstrate properties and applications of the proposed attribute.

LOCALIZED PHASE ESTIMATION

Our goal is to estimate the time-variant, localized phase from seismic data. What objective measure can indicate that a certain signal has a zero phase? One classic measure is the varimax norm or kurtosis (29; 14; 28). Varimax is defined as

$$\phi[\mathbf{s}] = \frac{N \sum_{n=1}^{N} s_n^4}{\left(\sum_{n=1}^{N} s_n^2 \right)^2} \, , \tag{1}$$

where $\mathbf{s} = \{s_1, s_2, \ldots, s_N\}$ represents a vector of seismic amplitudes inside a window of size N. Varimax is simply related to kurtosis of zero-mean signals.

The statistical rationale behind the Wiggins algorithm and its variants is that convolution of any filter with a time series that is white with respect to all statistical orders makes the outcome more Gaussian. The optimum deconvolution filter is therefore one that ensures the deconvolution output is maximally non-Gaussian (4). The constant-phase assumption made by (14) and (28) reduces the number of free parameters to one, thereby stabilizing performance compared with the Wiggins method. Wavelets derived in seismic-to-well ties often have a near-constant phase, thus justifying this assumption.

Noticing that the correlation coefficient of two sequences a_n and b_n is defined as

$$\gamma[\mathbf{a}, \mathbf{b}] = \frac{\sum_{n=1}^{N} a_n \, b_n}{\sqrt{\sum_{n=1}^{N} a_n^2 \sum_{n=1}^{N} b_n^2}} \tag{2}$$

and the correlation of a_n with a constant is

$$\gamma[\mathbf{a}, \mathbf{1}] = \frac{\sum_{n=1}^{N} a_n}{\sqrt{N \sum_{n=1}^{N} a_n^2}} , \tag{3}$$

(9) interpreted the kurtosis measure in equation 1 as the inverse of the squared correlation coefficient between s_n^2 and a constant, $\phi[\mathbf{s}] = 1/\gamma^2[\mathbf{s}^2, \mathbf{1}]$. Well-focused or zero-phase signals exhibit low correlation with a constant and correspondingly higher kurtosis (Figure 1). This provides an alternative interpretation to the goal of making the deconvolution outcome maximally non-Gaussian for desired phase estimation. Note that equation 10 is usually applied to zero-mean sequences **a** and **b**. This is neglected in the derivation of expression 3.

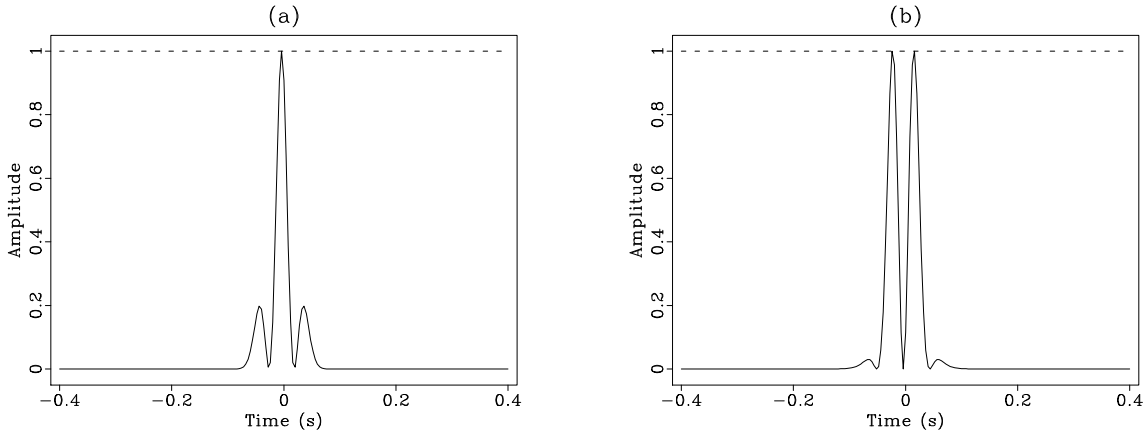


Figure 1: (a) Squared 0°-phase Ricker wavelet compared with a constant. (b) Squared 90°-phase Ricker wavelet compared with a constant. The 90°-phase signal has a higher correlation with a constant and correspondingly a lower kurtosis. phase/signal kur

In this paper, we suggest a different measure, skewness, for measuring the apparent phase of seismic signals. Skewness of a sequence s_n is defined as (2)

$$\kappa[\mathbf{s}] = \frac{\frac{1}{N}\sum_{n=1}^{N} s_n^3}{\left(\frac{1}{N}\sum_{n=1}^{N} s_n^2\right)^{3/2}} . \tag{4}$$

In statistics, skewness is used for measuring asymmetry of probability distributions.

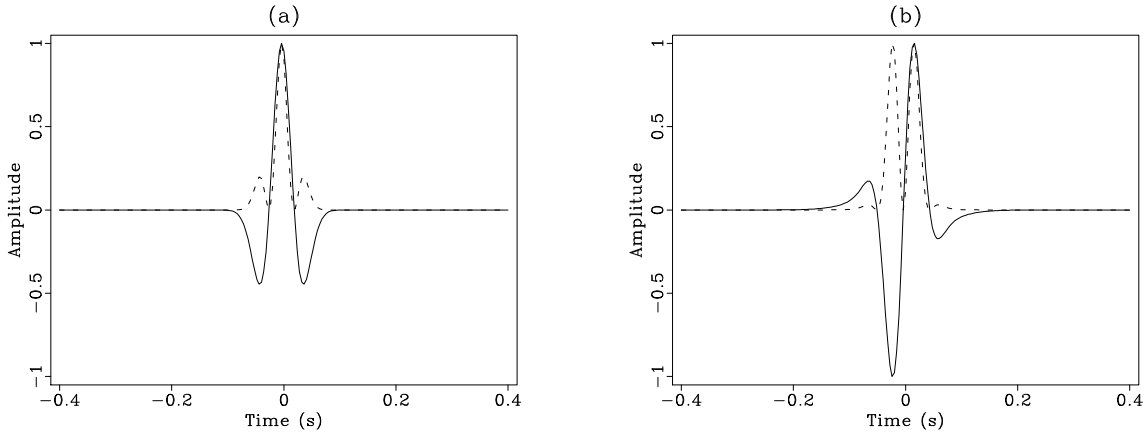


Figure 2: (a) 0°-phase Ricker wavelet compared with its square. (b) 90°-phase Ricker wavelet compared with its square. The 0°-phase has a stronger correlation with its square and correspondingly a higher skewness. phase/signal sqr

Simple algebraic manipulations show that skewness squared can be represented as

$$\kappa^2[\mathbf{s}] = \frac{\left(\sum_{n=1}^{N} s_n^2 \cdot s_n\right)^2}{\sum_{n=1}^{N} s_n^4 \sum_{n=1}^{N} s_n^2} \frac{\sum_{n=1}^{N} s_n^4 \sum_{n=1}^{N} 1^2}{\left(\sum_{n=1}^{N} s_n^2\right)^2} = \frac{\gamma^2[\mathbf{s}^2, \mathbf{s}]}{\gamma^2[\mathbf{s}^2, \mathbf{1}]} = \phi[\mathbf{s}]\, \gamma^2[\mathbf{s}^2, \mathbf{s}] \; . \tag{5}$$

In other words, squared skewness is equal to the kurtosis measure modulated by the squared correlation coefficient between the signal and its square. Zero-phase signals tend to exhibit higher correlation with the square and correspondingly higher skewness (Figure 2). Following experiments with synthetic and field data, we find it advantageous to use sometimes the inverse skewness

$$\frac{1}{\kappa^2[\mathbf{s}]} = \frac{\gamma^2[\mathbf{s}^2, \mathbf{1}]}{\gamma^2[\mathbf{s}^2, \mathbf{s}]} \; . \tag{6}$$

Unlike kurtosis which measures non-Gaussianity, skewness is related to asymmetry. Whereas convolution of two non-Gaussian sequences makes the outcome more Gaussian, convolution of two asymmetric series becomes more symmetric. Both phenomena are a consequence of the central limit theorem. A zero-phase wavelet is more compact than a nonzero phase one (21), and therefore also more asymmetric. Skewness-based criteria can thus detect the appropriate wavelet phase by applying a series of constant phase rotations to the data and then evaluating the angle that produces the most skewed distribution.

The two measures do not necessarily agree with one another, which is illustrated in Figures 3 and 4. For an isolated positive spike convolved with a compact zero-phase

wavelet, the two measures agree in the picking of the zero-phase result as having both a high kurtosis and a high skewness (Figure 3). For a slightly more complex case of a double positive spike convolved with the same wavelet (Figure 4), the two measures disagree: kurtosis picks a signal rotated by 90° whereas skewness picks the original signal. Note that, in both examples, skewness exhibits a significantly higher dynamical range, which makes it more suitable for picking optimal phase rotations.

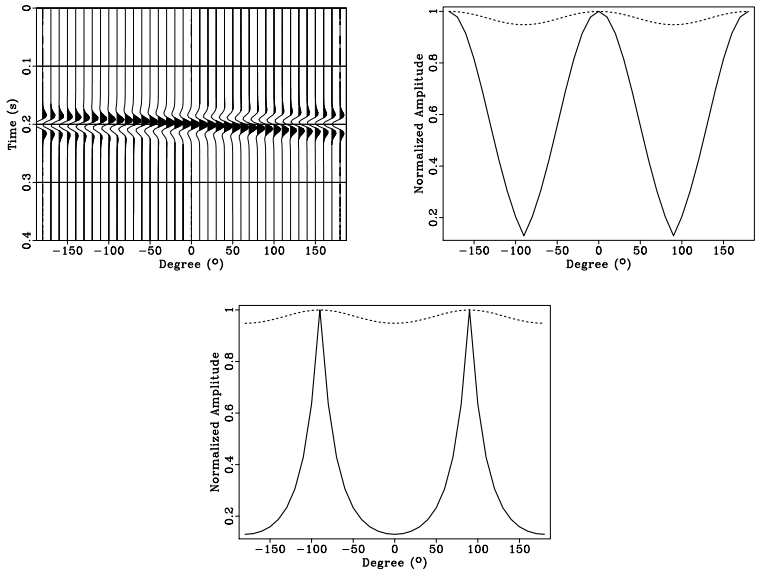


Figure 3: (a) Ricker wavelet rotated through different phases. (b) Skewness (solid line) and kurtosis (dashed line) as functions of the phase rotation angle. (c) Inverse skewness (solid line) and inverse kurtosis (dashed line) as functions of the phase rotation angle. The two measures agree in picking the signal at 0° and 180°. Note the higher dynamical range of skewness. phase/zero ricker-all,ricker-sq-corr,ricker-sq-corr-inv

DEFINING SKEWNESS AS A LOCAL ATTRIBUTE

The method of local attributes (5) is a technique for extending stationary or instantaneous attributes to smoothly varying or nonstationary attributes by employing a regularized least-squares formulation. In particular, the scalar correlation coefficient γ in equation 10 is replaced with a vector, $\mathbf{c}$, defined as a componentwise product of vectors $\mathbf{c}_1$ and $\mathbf{c}_2$, where

$$\mathbf{c}_1 = \left[\lambda^2 \mathbf{I} + \mathbf{S}\left(\mathbf{A}^T\mathbf{A} - \lambda^2\mathbf{I}\right)\right]^{-1} \mathbf{S}\,\mathbf{A}^T\mathbf{b}\,, \tag{7}$$

$$\mathbf{c}_2 = \left[\lambda^2 \mathbf{I} + \mathbf{S}\left(\mathbf{B}^T\mathbf{B} - \lambda^2\mathbf{I}\right)\right]^{-1} \mathbf{S}\,\mathbf{B}^T\mathbf{a}\,. \tag{8}$$

In equations 7-8, $\mathbf{a}$ and $\mathbf{b}$ are vectors composed of a_n and b_n, respectively; $\mathbf{A}$ and $\mathbf{B}$ are diagonal matrices composed of the same elements; and $\mathbf{S}$ is a smoothing operator.

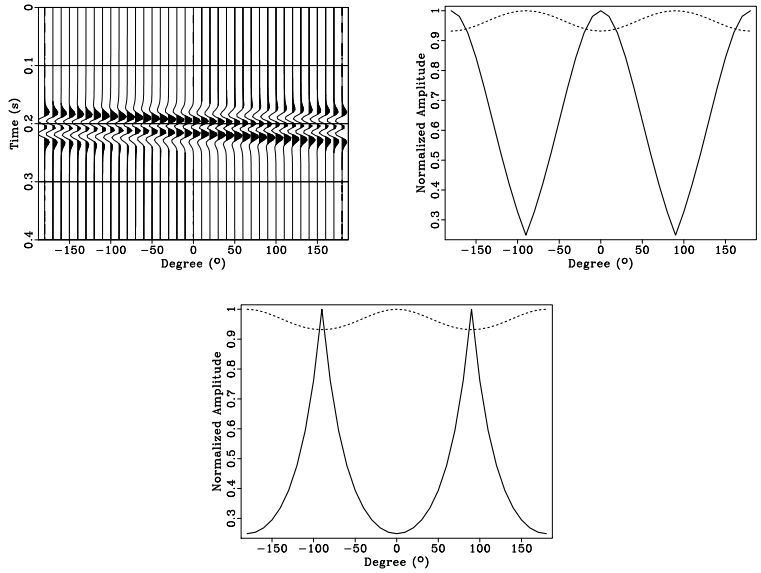


Figure 4: (a) Ricker wavelet convolved with a double impulse and rotated through different phases. (b) Skewness (solid line) and kurtosis (dashed line) as functions of the phase rotation angle. (c) Inverse skewness (solid line) and inverse kurtosis (dashed line) as functions of the phase rotation angle. The two measures disagree by 90° in picking the optimal phase. The skewness attribute picks a better focused signal. phase/zero ricker2-all,ricker2-sq-corr,ricker2-sq-corr-inv

We use triangle smoothing (3) controlled by specifying the smoothing radius, which can be different in vertical and horizontal directions.

Regularized inversion appearing in equations 7 and 8 is justified in the method of shaping regularization (6). The corresponding local similarity attribute has been used previously to align multicomponent and time-lapse images (5; 8; 13; 32), to detect focusing of diffractions (9), to enhance stacking (15; 17), to create time-frequency distributions (16), and to perform zero-phase correction with local kurtosis (23). In this paper, we apply it to zero-phasing seismic data using local skewness.

We illustrate the proposed zero-phase correction procedure in Figure 5. The input synthetic trace contains a set of Ricker wavelets with a gradually variable phase (Figure 5(a)). We start with a number of phase rotations with different angles, each time computing the local skewness. The result of this step is displayed in Figure 5(b) and shows a clear high-similarity trend. After picking the trend, adding 90° to it, and performing the corresponding nonstationary trace rotation, we end up with the phase-corrected trace, shown in Figure 5(c). All the original phase rotations are clearly detected and removed. The radius of the regularization smoothing in this example was 100 samples or 0.4 s.

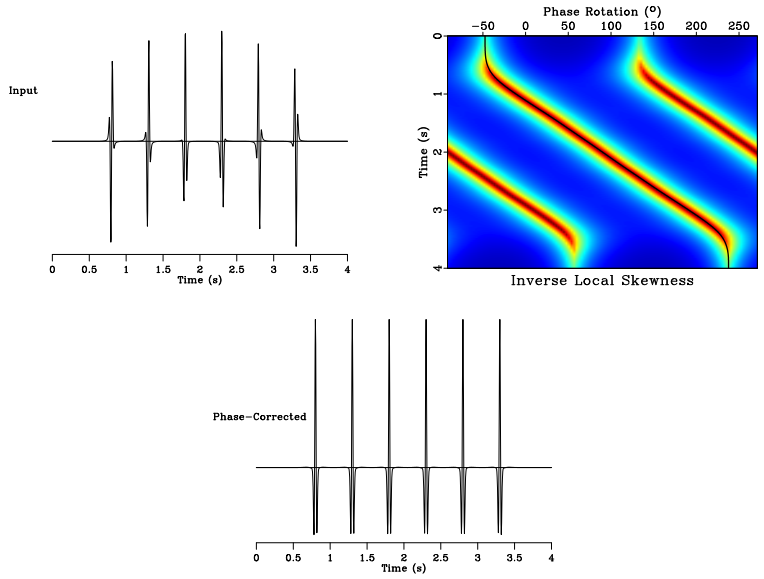


Figure 5: (a) Input synthetic trace with variable-phase events. (b) Inverse local skewness as a function of the phase rotation angle. Red colors correspond to high inverse similarity. (c) Synthetic trace after non-stationary rotation to zero phase using picked phase. phase/nonstat trace,tskew,trace0

APPLICATION EXAMPLE

.

The input dataset for our field-data example is the Boonsville dataset from the Fort Worth Basin in North-Central Texas, USA (11; 12). The formations of interest are the Ellenburger Carbonates, the Barnett Shale and the Bend Conglomerates (see Figure 6 for a stratigraphic column). The Ellenburger Carbonates are of Ordovician age. Their karstification due to post-Ellenburger carbonate dissolution and subsequent cavern collapses has created sags in the overlying formations, affecting sedimentation patterns and structures in the overlying Barnett Shales and Bend Conglomerates. The collapse features look like vertical chimneys with roughly circular cross-sections, extending up to 600-760 m above the Ellenburger Carbonates (11), sometimes even reaching into the Strawn Group above the Bend Conglomerates.

The Barnett Shales are of Mississippian age. They are the target of much current exploitation in Texas as these are tight-shale reservoirs (19). Zones with karst-induced cavern collapses form a drilling and completion hazard for the mainly horizontal drilling programs in these tight-shale reservoirs and must be mapped. They may affect local fracture densities and thus permeabilities and reservoir drainage positively but can also lead to fluid barriers due to reservoir compartmentalization.

The shallower clastic Bend Conglomerates are of Middle Pennsylvanian (Atokan) age. The formation has a thickness of 300-360 m in this area with depths between

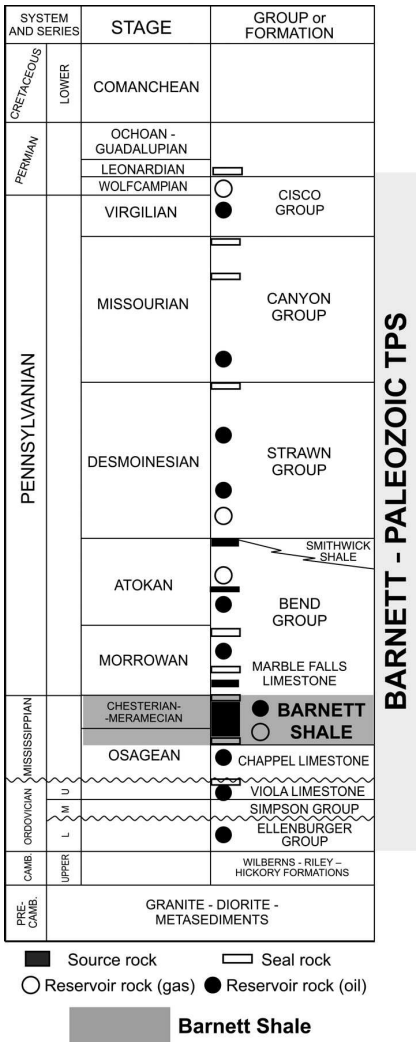


Figure 6: Stratigraphic column of the Fort Worth Basin where the Boonsville dataset is located, after (19). Karstification in the Ellenburger Carbonates has caused local sags in the overlying Barnett and Bend Conglomerate formations, creating reservoir compartmentalization. phase/. barnett-strat

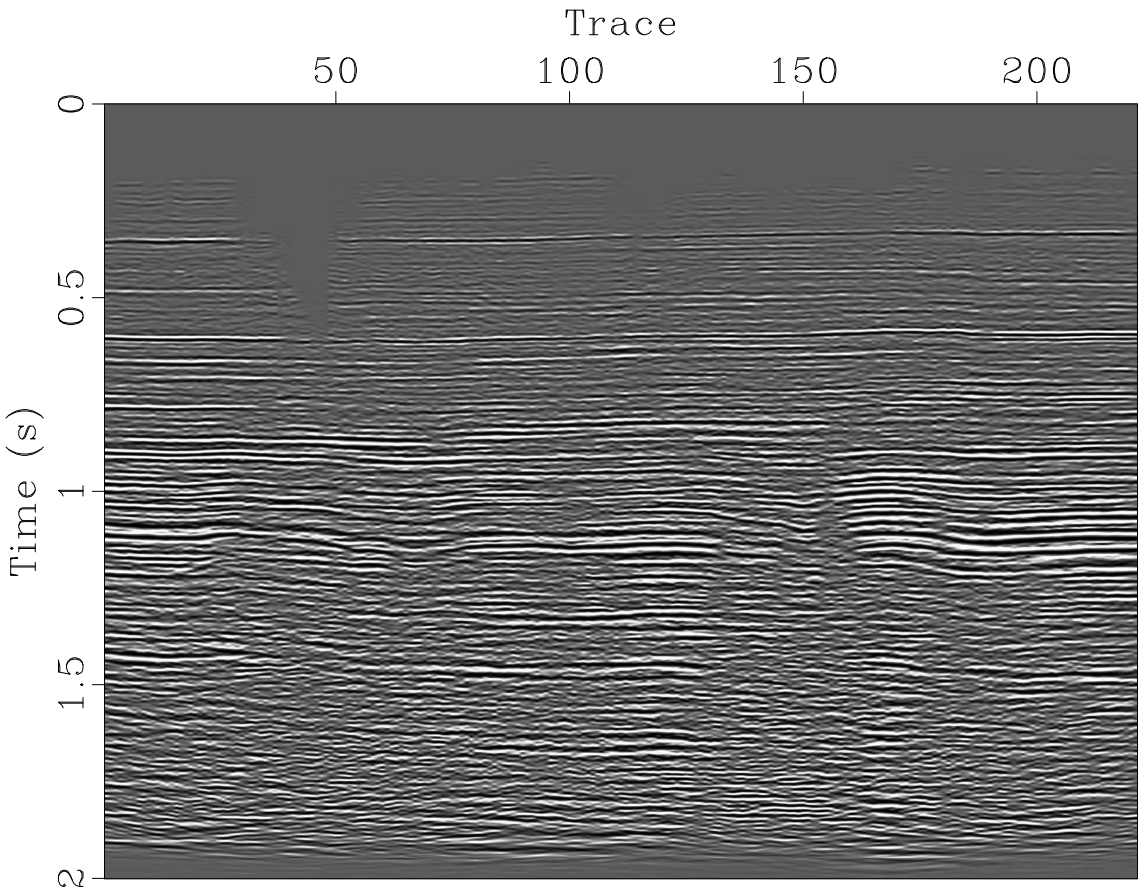


Figure 7: Input data: a section of the Boonsville dataset phase/boon boon

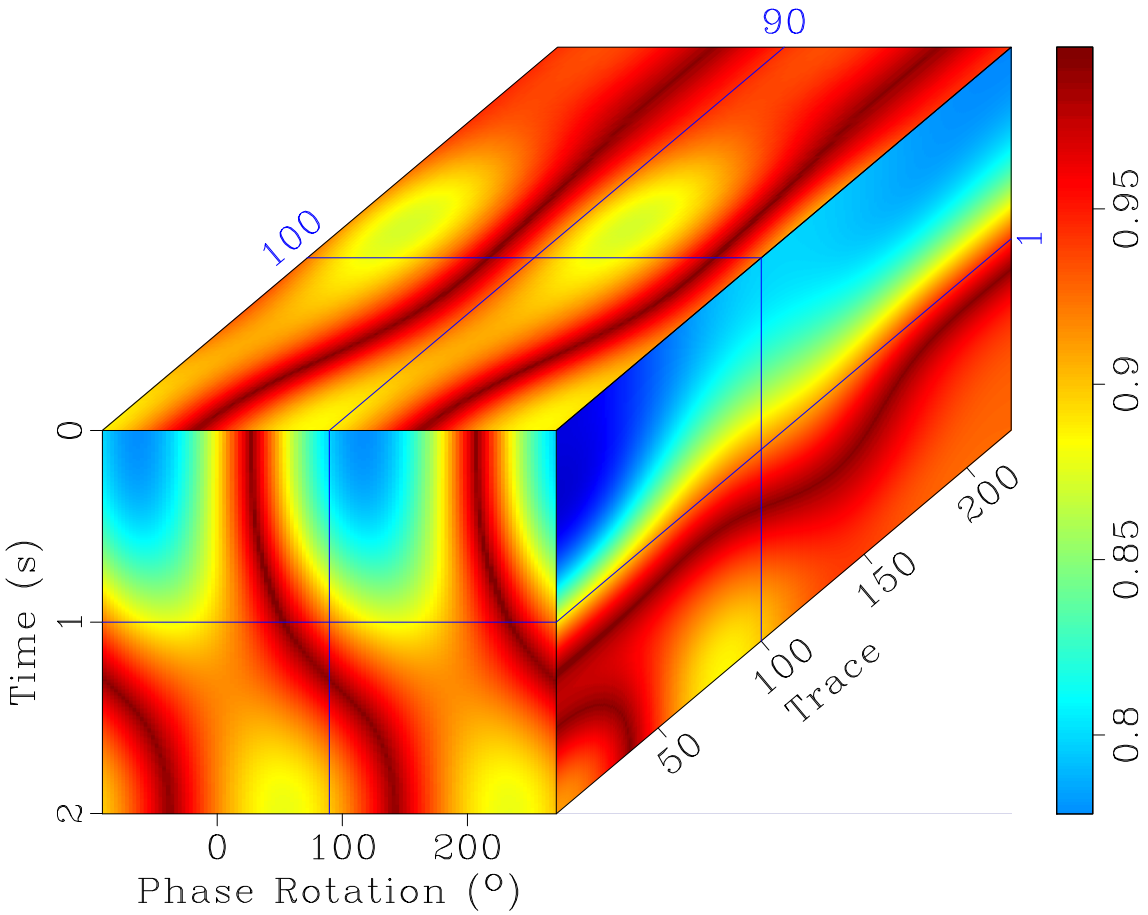


Figure 8: Inverse local skewness as a function of the phase rotation angle, with application to the section from Figure 7. Red colors correspond to high inverse similarity. phase/boon bfoc

1370 to 1830 m. It was targeted throughout the 1980s and 1990s as it contains several gas and oil-bearing reservoirs in a stacked fashion. (11; 12) describe how the karstification has greatly impacted the system tracts and sedimentation patterns in the Bend formation which were characterized by low accommodation space. Resulting reservoir compartmentalization is a significant challenge for this formation and has also affected the reflector character. Reflection near the base of this formation display both reflector weakening and sometimes even polarity reversals in areas depressed due to local sagging. Acquisition and processing of this dataset are described by (11). A stacked section is shown in Figure 7. A zero-phase correction has been applied to the data but has left regions with variable localized phase.

Our processing sequence is similar to the one used in Figure 5. First, we apply a number of phase rotations with different angles and compute local skewness for each rotation. The regularization lengths in this examples were 500 samples or 0.5 s in time and 50 traces in space. The result is displayed in Figure 8. Next, we apply automatic picking with the algorithm described by (7) to extract the nonstationary phase rotation that maximizes the local skewness. Finally, the phase correction is applied to the data, with the result displayed in Figure 9. A zoomed-in comparison shows the effects of non-stationary phase correction: rotating major seismic events to zero degrees and improving their continuity. These effects can be useful both for improving structural interpretation and for improved matching of seismic data and well logs.

We applied the local phase detection to the 3-D volume in a window centered on the target horizon (Figure 10). The estimated local phase variation along the target horizon is shown in Figure 11. Comparing the amplitude before and after phase correction (Figure 12), we observe a noticeable improvement in event continuity. Once the processing and interpretation are done on the zero-phase-corrected volume, it is easy to restore the original phase by applying the inverse phase rotation.

DISCUSSION

The data example as well as the previous case studies by (25) underline how analysis of the local phase can be used as a complementary attribute to spectral decomposition to highlight variations in wavelet character. There are two complimentary applications, namely analysis of the propagating and locally observed wavelets. The targeted wavelet type is determined by the chosen regularization length: the propagating wavelet is estimated by using long temporal regularization lengths, and the locally observed one from shorter lengths. The underlying assumption is that, for long regularization lengths, variations in the local geology are averaged out, revealing only the propagating wavelet. In this paper, we used relatively short regularization lengths as the aim is to highlight changes in the local reflection character.

Well-log analyses have demonstrated that the Earth's reflectivity series is non-Gaussian (27) and, to first order, white (26). In addition, impedances tend to in-

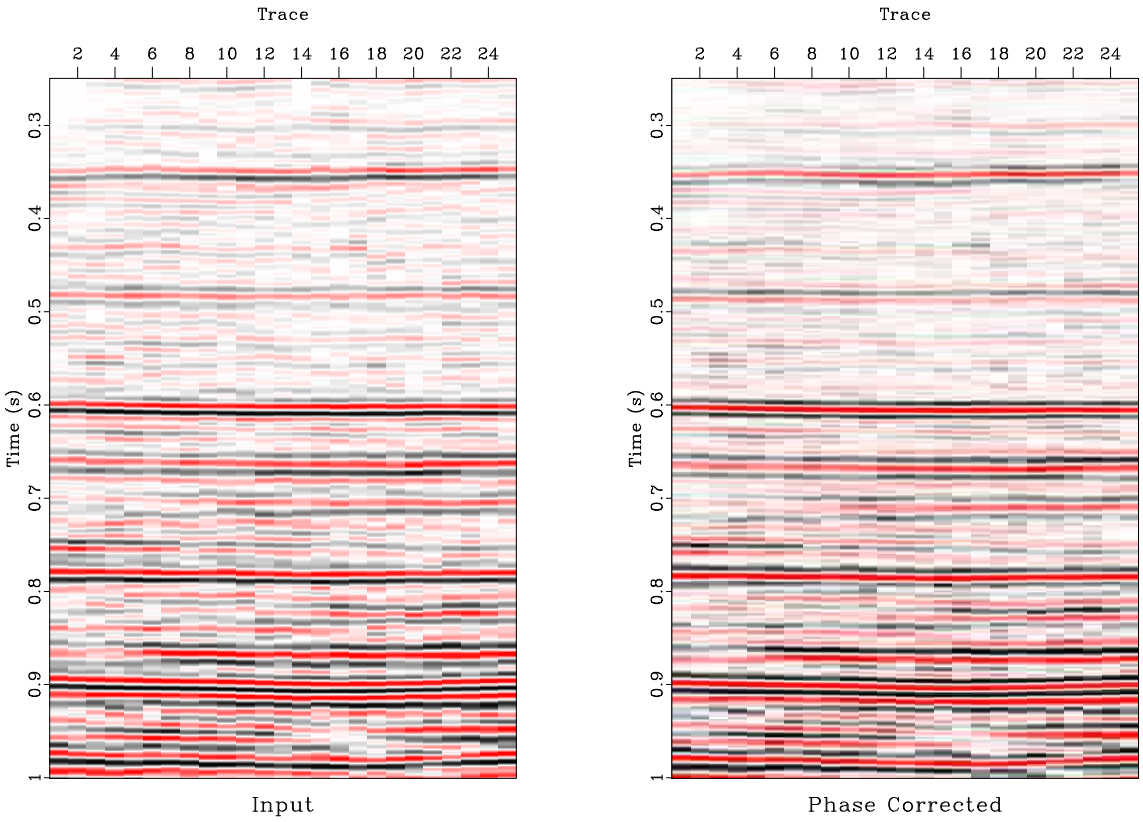


Figure 9: Zoomed-in comparison of the data before phase correction (a) and after phase correction (b). Nonstationary phase correction helps in identifying significant horizons and increasing their resolution in time. phase/boon boon-win

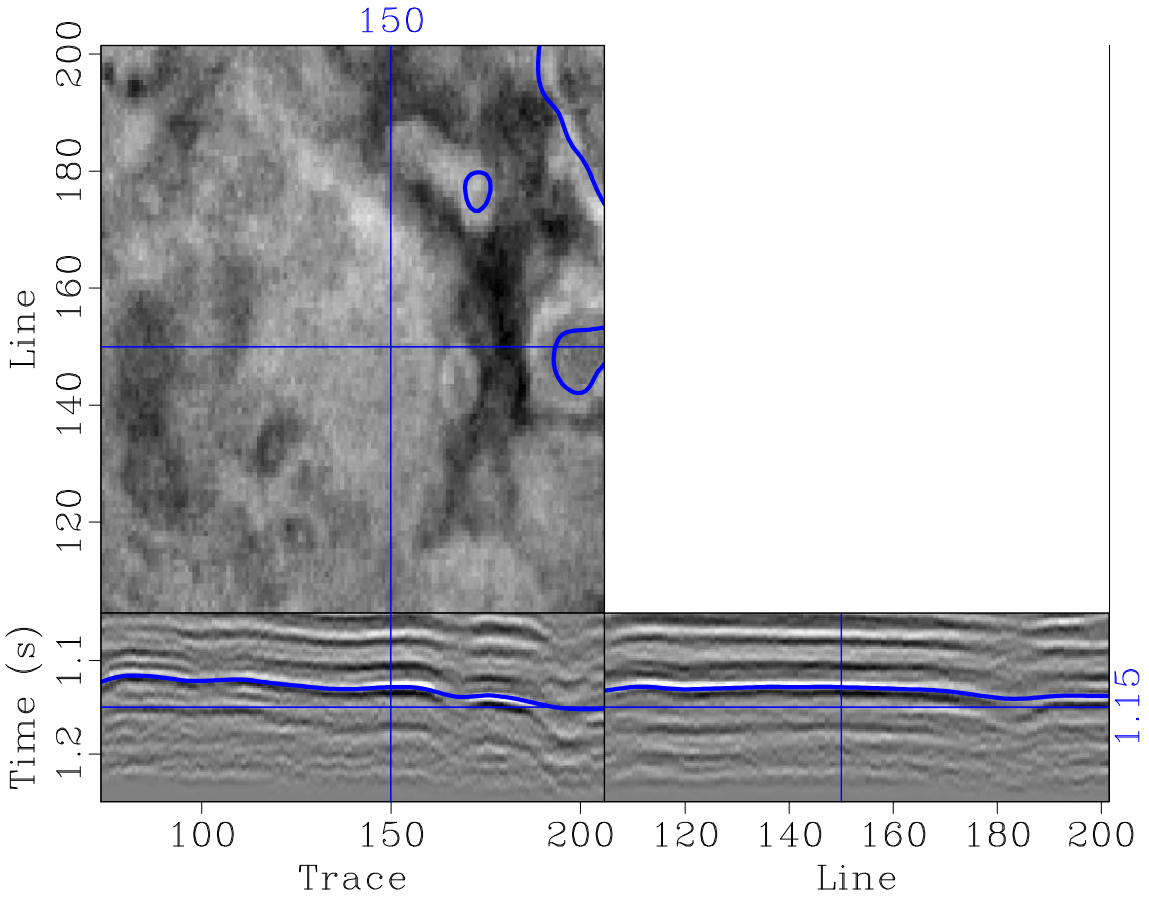


Figure 10: Boonsville dataset windowed around the target horizon. phase/boon3 boonm

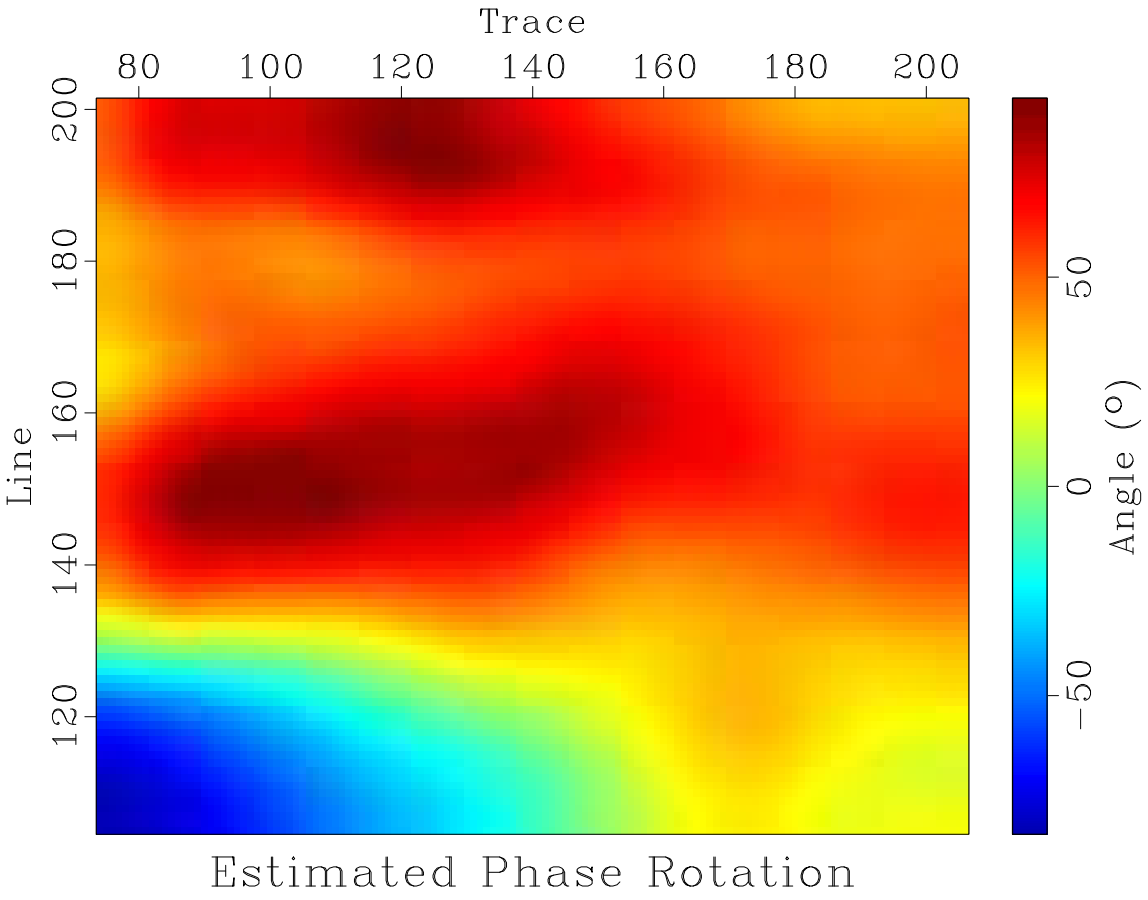


Figure 11: Local phase variation along the target horizon estimated from the data shown in Figure 10. phase/boon3 mpik

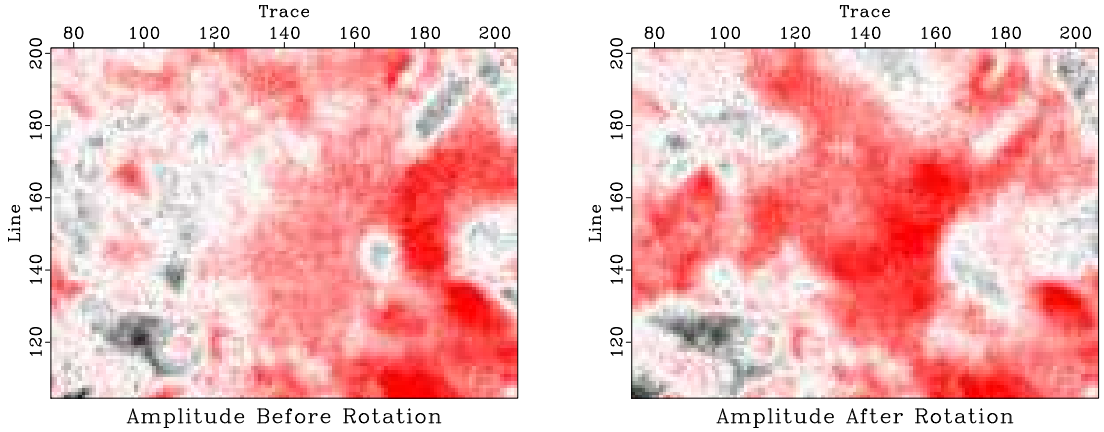


Figure 12: Amplitude along the target horizon before and after phase rotation. phase/boon3 ampl,ampr

crease with depth, hence positive reflection coefficients are slightly more likely than negative one, producing an asymmetric reflectivity distribution. Statistically, the skewness-based criterion assumes that the Earth's reflectivity series are white and asymmetric. This is in contrast to kurtosis used previously by (23), which assumes a non-Gaussian reflectivity series. Both the non-Gaussianity and asymmetry assumptions seem warranted but may fail if the local reflectivity series becomes respectively purely Gaussian or symmetric.

The local skewness attribute has the advantage over kurtosis because of its higher dynamic range, which facilitates picking. Variance is the second statistical order, skewness is the third one, and kurtosis is related to the fourth order. Estimation variances increases with the order of a moment (18). In other words, less samples are needed to estimate skewness with the same accuracy as kurtosis. We hypothesize that this contributes to the higher dynamic range of the skewness criterion.

CONCLUSIONS

We have presented a novel approach to nonstationary identification of apparent (locally observed) phase. Our approach is based on a new attribute, local skewness. In synthetic and field-data examples, local skewness exhibits a tendency to pick focused signals and a higher dynamical range than the previously used local kurtosis. Its computation involves a local similarity between the input signal and its square. Practical applications of using local skewness for zero-phase correction of seismic signals should be combined with well-log analysis in order to better separate the locally-observed phase from the propagating-wavelet phase.

ACKNOWLEDGMENTS

We thank Milo Backus, Guochang Liu, Mike Perz, and Hongliu Zeng for stimulating discussions.

REFERENCES

[1] Brown, A. R., 1999, Interpretation of three-dimensional seismic data: AAPG and SEG.
[2] Bulmer, M., 1979, Principles of statistics: Dover Publications.
[3] Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.
[4] Donoho, D., 1981, On minimum entropy deconvolution, *in* Applied time series analysis II: Academic Press, 565–608.
[5] Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.
[6] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.
[7] ———, 2009, Velocity analysis using AB semblance: Geophysical Prospecting, **57**, 311–321.
[8] Fomel, S., and L. Jin, 2009, Time-lapse image registration using the local similarity attribute: Geophysics, **74**, A7–A11.
[9] Fomel, S., E. Landa, and M. T. Taner, 2007, Post-stack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.
[10] Fomel, S., and M. van der Baan, 2010, Local similarity with the envelope as a seismic phase detector, *in* 80th Ann. Internat. Mtg: Soc. of Expl. Geophys., 1555–1559.
[11] Hardage, B. A., D. L. Carr, D. E. Lancaster, J. L. Simmons Jr, R. Y. Elphick, V. M. Pendleton, and R. A. Johns, 1996a, 3-D seismic evidence of the effects of carbonate karst collapse on overlying clastic stratigraphy and reservoir compartmentalization: Geophysics, **61**, 1336–1350.
[12] Hardage, B. A., D. L. Carr, D. E. Lancaster, J. L. Simmons Jr, D. S. Hamilton, R. Y. Elphick, K. L. Oliver, and R. A. Johns, 1996b, 3-D seismic imaging and seismic attribute analysis of genetic sequences deposited in low-accommodation conditions: Geophysics, **61**, 1351–1362.
[13] Kazemeini, S. H., C. Juhlin, and S. Fomel, 2010, Monitoring CO_2 response on surface seismic data; a rock physics and seismic modeling feasibility study at the CO_2 sequestration site, Ketzin, Germany: Journal of Applied Geophysics, **71**, 109–124.
[14] Levy, S., and D. W. Oldenburg, 1987, Automatic phase correction of common-midpoint stacked data: Geophysics, **52**, 51–59.
[15] Liu, G., S. Fomel, and X. Chen, 2011a, Stacking angle-domain common-image gathers for normalization of illumination: Geophysical Prospecting, **59**, 244–255.
[16] ———, 2011b, Time-frequency analysis of seismic data using local attributes: Geophysics, **76**, P23–P34.
[17] Liu, G., S. Fomel, L. Jin, and X. Chen, 2009, Stacking seismic data using local correlation: Geophysics, **74**, V43–V48.
[18] Mendel, J. M., 1991, Tutorial on higher-order statistics (spectra) in signal processing and system theory: Proc. IEEE, **79**, 278–305.
[19] Pollastro, R. M., D. M. Jarvie, R. J. Hill, and C. W. Adams, 2007, Geologic framework of the Mississipian Barnett Shale, Barnett-Paleozoic total petroleum

system, Bend-arch–Fort Worth Basin, Texas: AAPG Bulletin, **91 (4)**, 405–436.
[20] Robinson, E. A., and S. Treitel, 2000, Geophysical signal analysis: Society Of Exploration Geophysicists.
[21] Schoenberger, M., 1974, Resolution comparison of minimum-phase and zero-phase signals: Geophysics, **39**, 826–833.
[22] Van der Baan, M., 2008, Time-varying wavelet estimation and deconvolution using kurtosis maximization: Geophysics, **73**, V11–V18.
[23] Van der Baan, M., and S. Fomel, 2009, Nonstationary phase estimation using regularized local kurtosis maximization: Geophysics, **74**, A75–A80.
[24] Van der Baan, M., S. Fomel, and M. Perz, 2010a, Nonstationary phase estimation: A tool for seismic interpretation?: The Leading Edge, **29**, 1020–1026.
[25] Van der Baan, M., M. Perz, and S. Fomel, 2010b, Nonstationary phase estimation for analysis of wavelet character, *in* 72nd Conference and Exhibition: EAGE, D020.
[26] Walden, A. T., and J. W. J. Hosken, 1985, An investigation of the spectral properties of primary reflection coefficients: Geophys. Prospect., **33**, 400–435.
[27] ———, 1986, The nature of the non-gaussianity of primary reflection coefficients and its significance for deconvolution: Geophys. Prospect, **34**, 1038–1066.
[28] White, R. E., 1988, Maximum kurtosis phase correction: Geophysical Journal, **95**, 371–389.
[29] Wiggins, R., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21–35.
[30] Xu, Y., P. Thore, and S. Duchenne, 2012, The reliability of the kurtosis-based wavelet estimation, *in* SEG Technical Program Expanded Abstracts 2012: Society of Exploration Geophysicists.
[31] Zeng, H., and M. M. Backus, 2005, Interpretive advantages of 90°-phase wavelets. Part 1: Modeling: Geophysics, **70**, C7–C15.
[32] Zhang, R., X. Song, S. Fomel, M. K. Sen, and S. Srinivasan, 2013, Time-lapse seismic data registration and inversion for CO_2, sequestration study at Cranfield: Geophysics, **78**, B329–B338.

A fast algorithm for 3D azimuthally anisotropic velocity scan

Jingwei Hu, Sergey Fomel, and Lexing Ying[*]

ABSTRACT

Conventional velocity scan can be computationally expensive for large-size seismic data, particularly when the presence of anisotropy requires multiparameter estimation. We introduce a fast algorithm for 3D azimuthally anisotropic velocity scan, which is a generalization of the previously proposed 2D butterfly algorithm for hyperbolic Radon transform. To compute the semblance in a two-parameter residual moveout domain, the numerical complexity of our algorithm is roughly $O(N^3 \log N)$ as opposed to $O(N^5)$ of the straightforward velocity scan, with N being representative of the number of points in either dimension of data space or parameter space. We provide both synthetic and field-data examples to illustrate the efficiency and accuracy of the algorithm.

INTRODUCTION

Multiazimuth seismic data reveal the Earth's seismic response along different azimuthal directions. Detecting and measuring the anisotropy in such data can be useful for characterizing fractures or stress in the subsurface (14). When apparent azimuthal anisotropy is present, conventional single-parameter isotropic velocity scan and normal moveout are usually inadequate. To further flatten the events, a residual anisotropic moveout is necessary. This, however, brings several difficulties in the implementation. First, the computational cost increases dramatically compared with the single-parameter case. If we assume for simplicity that there are N sample points in every dimension of data and model (parameter) domains, then the numerical complexity of a two-parameter velocity scan will be at least $O(N^5)$; i.e., summing over $O(N^2)$ data points for each of $O(N^3)$ values (time + two parameters). Second, the simultaneous automatic picking from a high-dimensional semblance volume also poses a challenge (1; 11; 2; 13).

In this work, we attempt to solve a fundamental problem related to the first difficulty. Specifically, we introduce a fast algorithm to speed up the velocity-scan process. The stacking procedure involved in computing the semblance can be regarded as a generalized Radon transform (3). Following our previous work on the hyperbolic Radon transform (9; 10), we formulate the time-domain summation as a discrete

[*]**e-mail:** sergey.fomel@beg.utexas.edu

oscillatory integral in the frequency domain, and apply the 3D version of the FIO (Fourier integral operator) butterfly algorithm (5). As a result, complexity of the velocity scan reduces to roughly $O(N^3 \log N)$, where N is representative of the number of points in either dimension of data space or model space. An alternative approach was developed by (4), but may not be applicable for noisy data.

THEORY

As explained by (8), a pure-mode (P or S) reflection event in an effectively azimuthally anisotropic medium can be described by

$$t = \sqrt{\tau^2 + W_{11}x^2 + W_{22}y^2 + 2W_{12}xy}, \tag{1}$$

where t is two-way CMP traveltime, τ is two-way zero-offset traveltime, (x, y) is the full source-receiver offset in surface survey coordinates, and

$$\mathbf{W} = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix} \tag{2}$$

is the slowness matrix. Equation 1 follows from a truncated 2D Taylor expansion. Geometrically, it represents a curved surface that is hyperbolic in cross section and elliptic in map view.

Ideally, one can perform a semblance scan (12) over the three parameters W_{11}, W_{22}, and W_{12} simultaneously to estimate the velocity and perform NMO correction. However, this approach, if not impossible, is extremely expensive for large-size seismic data. Furthermore, since these parameters are not orthogonal, the semblance plots might appear to be extended and ambiguous, hence presenting difficulties for picking (7).

(6) proposed a stable way of detecting azimuthal anisotropy using an orthogonal parametrization of the moveout function, which is based on an equivalent reformulation of equation 1,

$$t = \sqrt{\tau^2 + W_{\text{avg}}(x^2 + y^2) + W_{\cos}(x^2 - y^2) + 2W_{\sin}xy}. \tag{3}$$

The cosine and sine dependent slownesses $W_{\cos}$ and $W_{\sin}$ are usually much smaller than the averaged slowness W_{avg}. Therefore, a possible workflow for anisotropic velocity analysis and NMO correction can proceed in three steps:

1. Perform an isotropic velocity scan to estimate W_{avg} and flatten seismic events.
2. Perform a residual anisotropic moveout to account for $W_{\cos}$ and $W_{\sin}$ dependent terms simultaneously.
3. Convert orthogonal parameters to more intuitive anisot-ropy parameters. For instance, the normal moveout velocity at azimuth α can be recovered by

$$V_{\text{nmo}}^{-2}(\alpha) = W_{\text{avg}} + W_{\cos}\cos 2\alpha + W_{\sin}\sin 2\alpha. \tag{4}$$

In this procedure, the first two steps require a velocity-scan process. In fact, because x and y are symmetric in $W_{\text{avg}}(x^2+y^2)$, the single-parameter isotropic scan involved in the first step can be handled efficiently by a 2D butterfly algorithm, as discussed in our previous work (9; 10). Our goal in this paper is to speed up the more expensive, two-parameter velocity scan in the second step.

To be specific, what we need for residual moveout is to compute a semblance as follows (assuming that the W_{avg} part has been moved out from the previous step):

$$S(\tau, W_{\cos}, W_{\sin}) = \frac{\left(\sum_{x,y} d(t(x,y;\tau,W_{\cos},W_{\sin}),x,y)\right)^2}{N_x N_y \sum_{x,y} d^2(t(x,y;\tau,W_{\cos},W_{\sin}),x,y)}, \tag{5}$$

where $d(t,x,y)$ is a 3D CMP dataset after isotropic moveout and

$$t(x,y;\tau,W_{\cos},W_{\sin}) = \sqrt{\tau^2 + W_{\cos}(x^2-y^2) + 2W_{\sin}xy}. \tag{6}$$

Basic formulation

The right-hand side of equation 5 is a quotient of two (discrete) generalized Radon transforms (3). They can be expressed in a unified way as (to simplify the notation, we write $p = W_{\cos}$, $q = W_{\sin}$ in this and next subsections)

$$(Rg)(\tau,p,q) = \sum_{x,y} g(\sqrt{\tau^2 + p(x^2-y^2) + 2qxy}, x, y), \tag{7}$$

where g is d or some composite function of d.

To construct the fast algorithm, we first rewrite equation 7 in the frequency domain as

$$(Rg)(\tau,p,q) = \sum_{f,x,y} e^{2\pi i f\sqrt{\tau^2+p(x^2-y^2)+2qxy}} \hat{g}(f,x,y), \tag{8}$$

where f is frequency and $\hat{g}(f,x,y)$ is the Fourier transform of $g(t,x,y)$ in time. We next perform a linear transformation to map all discrete points in (f,x,y) and (τ,p,q) domains to points in the unit cube $[0,1]^3$; i.e., a point $(f,x,y) \in [f_{\min},f_{\max}] \times [x_{\min},x_{\max}] \times [y_{\min},y_{\max}]$ is mapped to $\mathbf{k} = (k_1,k_2,k_3) \in [0,1]\times[0,1]\times[0,1] = K$ via

$$\begin{aligned} f &= (f_{\max} - f_{\min})k_1 + f_{\min}, \\ x &= (x_{\max} - x_{\min})k_2 + x_{\min}, \\ y &= (y_{\max} - y_{\min})k_3 + y_{\min}; \end{aligned}$$

a point $(\tau,p,q) \in [\tau_{\min},\tau_{\max}]\times[p_{\min},p_{\max}]\times[q_{\min},q_{\max}]$ is mapped to $\mathbf{x} = (x_1,x_2,x_3) \in [0,1]\times[0,1]\times[0,1] = X$ via

$$\begin{aligned} \tau &= (\tau_{\max} - \tau_{\min})x_1 + \tau_{\min}, \\ p &= (p_{\max} - p_{\min})x_2 + p_{\min}, \\ q &= (q_{\max} - q_{\min})x_3 + q_{\min}. \end{aligned}$$

If we define a phase function $\Phi(\mathbf{x}, \mathbf{k})$ as

$$\Phi(\mathbf{x}, \mathbf{k}) = f\sqrt{\tau^2 + p(x^2 - y^2) + 2qxy}, \tag{9}$$

then equation 8 can be recast as

$$(Rg)(\mathbf{x}) = \sum_{\mathbf{k}\in K} e^{2\pi i\Phi(\mathbf{x},\mathbf{k})}\hat{g}(\mathbf{k}), \quad \mathbf{x} \in X \tag{10}$$

(throughout the paper, K and X are used to denote either sets of discrete points or the cubic domains containing them).

Fast 3D butterfly algorithm

Equation 10 is the discretized form of a 3D oscillatory integral of the type

$$u(\mathbf{x}) = \int_K e^{2\pi i\Phi(\mathbf{x},\mathbf{k})} v(\mathbf{k})\, d\mathbf{k}, \quad \mathbf{x} \in X, \tag{11}$$

whose fast evaluation can be realized by a butterfly algorithm (5).

The overall structure of the 3D butterfly algorithm basically follows its 2D analogue. The idea is to partition the computational domains X and K recursively into a pair of *octrees*, T_X and T_K, ending at level $l = \log N$ (see Figure 1(d) for an illustration). Here N is chosen as an integer power of two, which is on the order of the maximum of $|\Phi(\mathbf{x}, \mathbf{k})|$ for all possible $\mathbf{x}$ and $\mathbf{k}$ (so it is mainly determined by the range of variables (f, x, y) and (τ, p, q)). A crucial property of this structure is that at arbitrary level l, the side lengths $w(A)$ of a box A in T_X and $w(B)$ of a box B in T_K always satisfy $w(A)w(B) = 1/N$. Then when $\mathbf{x}$, $\mathbf{k}$ restricted in A and B respectively, one can construct a low-rank, separated expansion for the kernel function $e^{2\pi i\Phi(\mathbf{x},\mathbf{k})}$ (via a Chebyshev interpolation):

$$\left| e^{2\pi i\Phi(\mathbf{x},\mathbf{k})} - \sum_{r=1}^{r_\epsilon} \alpha_r^{AB}(\mathbf{x})\beta_r^{AB}(\mathbf{k}) \right| < \epsilon. \tag{12}$$

By embedding this approximation in the above tree structure and traversing T_X from top to bottom, T_K from bottom to top, we arrive at a fast algorithm running in complexity $O(N^3 \log N)$ (there are N^3 pairs of boxes (A, B) on every level, and there are $\log N$ levels in total). Detailed description of the algorithm can be found in (10), where the difference between 2D and 3D formulations should be clear from the context.

Considering the initial Fourier transform for preparing data in (f, x, y) domain, the overall complexity of our algorithm is roughly $O(N_x N_y N_t \log N_t) + O(C_1(r_\epsilon)(N_f N_x N_y + N_\tau N_p N_q)) + O(C_2(r_\epsilon)N^3 \log N)$ (r_ϵ terms are due to low-rank approximations, and $C_2(r_\epsilon)$ is bigger than $C_1(r_\epsilon)$).

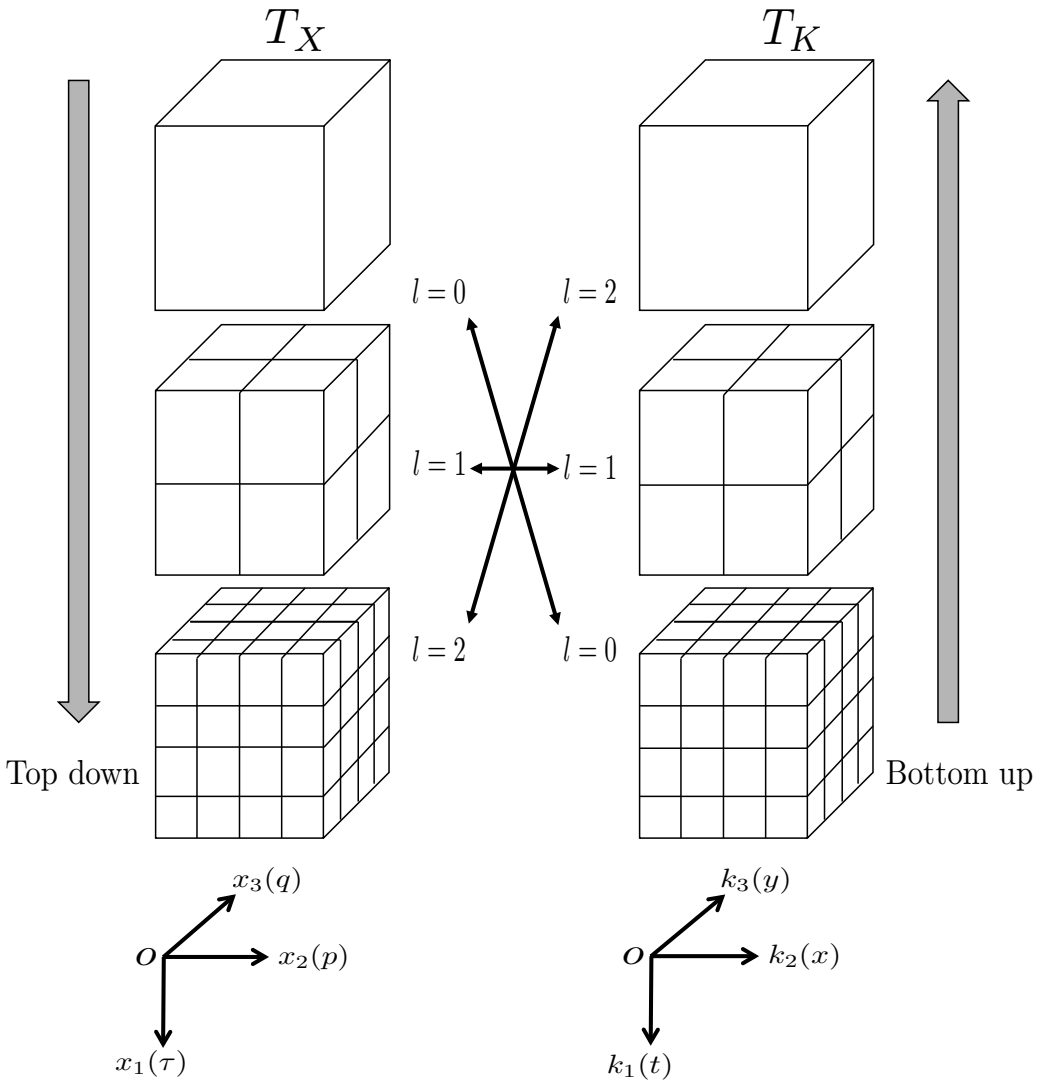


Figure 1: Butterfly tree structure for the special case of $N = 4$. scan3/. cube

event	τ	W_{avg}	$W_{\cos}$	$W_{\sin}$
1	0.7	0.3	0	0
2	1.8	0.29	0.021	0.021
3	2.6	0.25	-0.01	-0.017
4	3.4	0.15	0	0.02

Table 1: Parameters used to generate the seismic events in Figure 2.

By comparison, the conventional straightforward velocity scan requires at least $O(N_\tau N_p N_q N_x N_y)$ computations, which may quickly become a bottleneck as the problem size increases. Yet the efficiency of our algorithm is controlled mainly by $O(N^3 \log N)$ with an ϵ-dependent constant, where N is determined by the degree of oscillations in the kernel $e^{2\pi i \Phi(\mathbf{x},\mathbf{k})}$. Generally speaking, N depends on the maximum frequency and offset in the dataset, and the range of parameters in the model space. In practice, N can often be chosen smaller than the grid size.

The significance of above analysis for the fast algorithm lies in the fact that the input and output data sizes $N_t N_x N_y$ and $N_\tau N_p N_q$ have little impact on the final computational cost; a dense sampling therefore becomes affordable.

NUMERICAL EXAMPLES

Example 1

We first consider a simple 3D synthetic CMP gather consisting of four isolated events, each with a different degree of azimuthal anisotropy (Figure 2(a)). The moveout parameters τ, W_{avg}, $W_{\cos}$, and $W_{\sin}$ used to generate the events are specified in Table 1. Figure 1 shows the data after isotropic NMO using the exact W_{avg}. Except for the first flattened isotropic event, the other three events clearly require an additional moveout. The computed semblance by the fast algorithm is shown in Figure 3, where manually picked parameters coincide well with exact values. Besides accuracy, what is remarkable is that, even for this moderate-sized problem ($N_t = N_\tau = 1000$, $N_x = N_y = N_{W_{\cos}} = N_{W_{\sin}} = 100$), CPU time of the butterfly algorithm (for a single Radon transform) is about 139 s, whereas the direct velocity scan takes 4681 s.

Example 2

We now further investigate the properties of the fast algorithm using a more realistic 3D synthetic CMP gather (Figure 9(a)). The semblance plot computed by the fast algorithm is shown in Figure 5. Figure 6(a) is the isotropically NMO corrected data. After residual moveout using picked velocities from the semblance, curved events are flattened to the right position (Figure 6(b)).

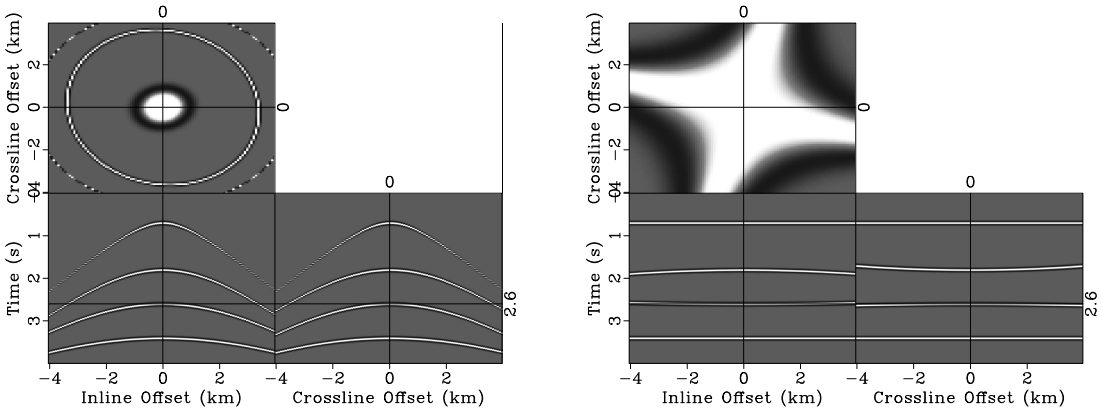


Figure 2: 3D synthetic CMP gather (a) before and (b) after isotropic NMO. $N_t = 1000$, $N_x = N_y = 100$. $\Delta t = 0.004$ s, $\Delta x = \Delta y = 80$ m. scan3/fourevents cmp-before,cmp

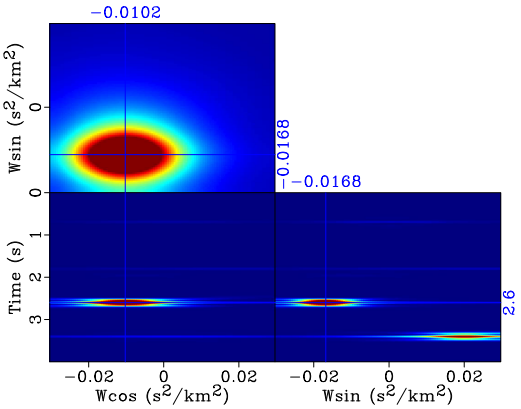


Figure 3: Semblance plot (event 3) computed by the fast algorithm. $N_\tau = 1000$, $N_{W_{\cos}} = N_{W_{\sin}} = 100$. scan3/fourevents NMOsemb

$N_{W_{\cos}} \times N_{W_{\sin}}$	Direct velocity scan	Fast butterfly algorithm	Speedup factor
10×10	1847 s	145 s	**12.7**
20×20	7394 s	146 s	**50.6**
100×100	$\sim$ 184700 s	159 s	$\sim$ **1162**
200×200	$\sim$ 738800 s	196 s	$\sim$ **3769**

Table 2: CPU time of direct velocity scan and fast butterfly algorithm for different $N_{W_{\cos}}$ and $N_{W_{\sin}}$ applied to the synthetic data in Figure 9(a).

We next fix $N_t = N_\tau = 1000$, $N_x = N_y = 400$ and compare CPU time of the fast algorithm and the direct velocity scan for different $N_{W_{\cos}}$ and $N_{W_{\sin}}$ (Table 2). When $N_{W_{\cos}}$ and $N_{W_{\sin}}$ increase by a factor of 2, computation time of the direct velocity scan increases nearly by a factor of 4, which is consistent with our previous discussion on numerical complexity. On the other hand, CPU time of the fast algorithm is not affected much by the size of output sampling, again confirming our expectations.

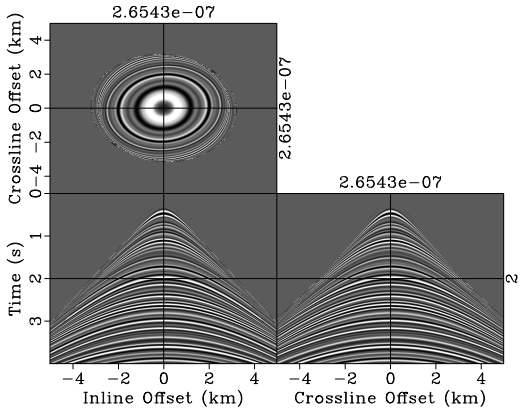


Figure 4: 3D synthetic CMP gather. $N_t = 1000$, $N_x = N_y = 400$. $\Delta t = 0.004$ s, $\Delta x = \Delta y = 25$ m. scan3/noncwavg-1 data

Example 3

Finally we consider a field-data example. A subset of the McElroy dataset from West Texas was formed in a supergather (Figure 7). This dataset was studied by (4), in which they proposed a velocity-independent moveout correction to avoid costly velocity scan. With the fast algorithm, we are now able to compute the semblance efficiently: only 45 s for a single Radon transform when $N_t = N_\tau = 400$, $N_x = N_y = 297$, and $N_{W_{\cos}} = N_{W_{\sin}} = 200$; direct computation at this sampling would take approximately 30 hours.

Although the original data have been isotropically NMO corrected, the time slice

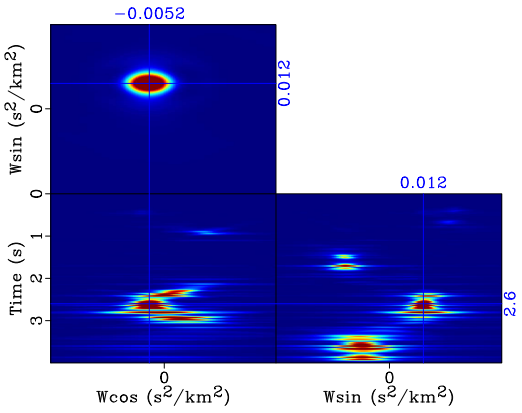


Figure 5: Semblance plot computed by the fast algorithm. $N_\tau = 1000$, $N_{W_{\cos}} = N_{W_{\sin}} = 200$. scan3/noncwavg-1 NMOsemb-1

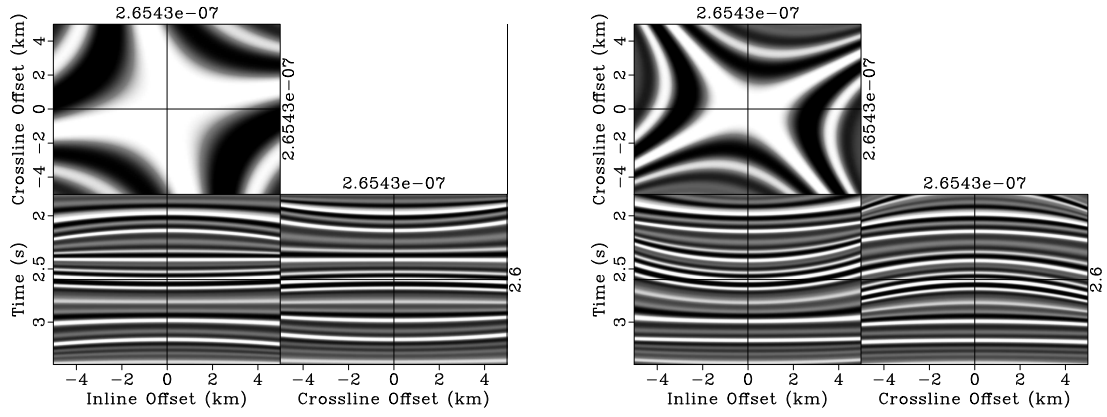


Figure 6: Synthetic gather (a) before and (b) after residual moveout using picked velocities from the semblance scan. scan3/noncwavg-1 NMOdatacut,RMOdatacut

still shows a subtle directional trend to flatness (Figure 7). From the semblance plot (Figure 8), we observe some nonzero values of anisotropic parameters.

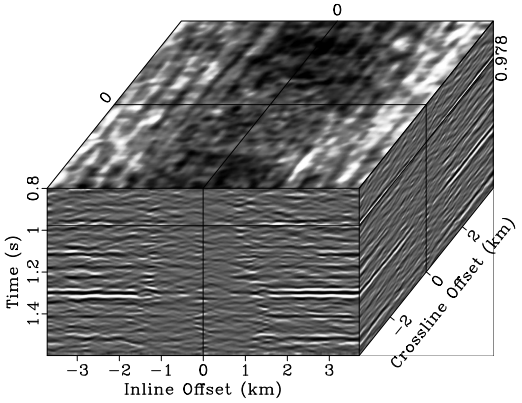


Figure 7: An isotropically NMO-corrected supergather from the McElroy dataset, West Texas. $N_t = 400$, $N_x = N_y = 297$. $\Delta t = 0.002$ s, $\Delta x = \Delta y = 25$ m. scan3/realdata super

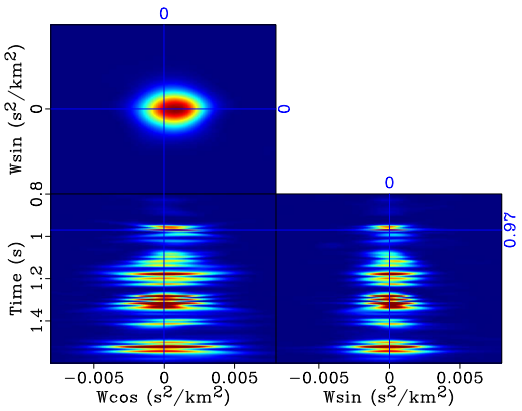


Figure 8: Semblance plot computed by the fast algorithm. $N_\tau = 400$, $N_{W_{\cos}} = N_{W_{\sin}} = 200$. scan3/realdata semb

CONCLUSIONS

We have introduced a fast algorithm for 3D, azimuthally aniso-tropic velocity scan. Our synthetic and field-data experiments show that the method can be orders of magnitude faster than the conventional velocity scan, especially for large-size data and dense parameter sampling.

To illustrate the idea, equation 3 was used throughout the paper. Applicability of the butterfly algorithm is not limited to this form of the moveout function, as long as the transform can be written in the form of equation 10.

ACKNOWLEDGMENTS

We thank Chevron for providing the field data. We thank KAUST and sponsors of the Texas Consortium for Computational Seismology (TCCS) for financial support.

REFERENCES

[1] Adler, F., and S. Brandwood, 1999, Robust estimation of dense 3D stacking velocities from automated picking: 69th Annual International Meeting, SEG, Expanded Abstracts.

[2] Arnaud, J., D. Rappin, J.-P. Dunand, and V. Curinier, 2004, High density picking for accurate velocity and anisotropy determination: 74th Annual International Meeting, SEG, Expanded Abstracts.

[3] Beylkin, G., 1984, The inversion problem and applications of the generalized Radon transform: Communications on Pure and Applied Mathematics, **37**, 579–599.

[4] Burnett, W., and S. Fomel, 2009, 3D velocity-independent elliptically anisotropic moveout correction: Geophysics, **74**, WB129–WB136.

[5] Candès, E., L. Demanet, and L. Ying, 2009, A fast butterfly algorithm for the computation of Fourier integral operators: Multiscale Modeling and Simulation, **7**, 1727–1750.

[6] Davidson, M., H. Swan, S. Sil, J. Howell, R. Olson, and C. Zhou, 2011, A robust workflow for detecting azimuthal anisotropy: 81st Annual International Meeting, SEG, Expanded Abstracts, 259–263.

[7] Fowler, P. J., A. Jackson, and B. Hootman, 2006, Orthogonal parameters for anisotropic velocity analysis: 76th Annual International Meeting, SEG, Expanded Abstracts, 169–173.

[8] Grechka, V., and I. Tsvankin, 1998, 3-D description of normal moveout in anisotropic inhomogeneous media: Geophysics, **63**, 1079–1092.

[9] Hu, J., S. Fomel, L. Demanet, and L. Ying, 2012, A fast butterfly algorithm for the hyperbolic Radon transform: 82nd Annual International Meeting, SEG, Expanded Abstracts.

[10] ———, 2013, A fast butterfly algorithm for generalized Radon transforms: Geophysics, accepted.

[11] Siliqi, R., D. L. Meur, F. Gamar, L. Smith, J. P. Toure, and P. Herrmann, 2003, High-density moveout parameter fields V and Eta. Part I: Simultaneous automatic picking: 73rd Annual International Meeting, SEG, Expanded Abstracts.

[12] Taner, M. T., and F. Koehler, 1969, Velocity spectra – Digital computer derivation and applications of velocity functions: Geophysics, **34**, 859–881.

[13] Tao, Y., M. Davidson, H. Swan, S. Fomel, J. Malloy, J. Howell, S. Chiu, and R. Olson, 2012, Constrained simultaneous automatic picking for VVAZ analysis: 82nd Annual International Meeting, SEG, Expanded Abstracts.

[14] Tsvankin, I., and V. Grechka, 2011, Seismology of azimuthally anisotropic media and seismic fracture characterization: Society of Exploration Geophysicists.

Stacking angle-domain common-image gathers for normalization of illumination

Guochang Liu, China University of Petroleum-Beijing and The University of Texas at Austin; Sergey Fomel, The University of Texas at Austin; Xiaohong Chen, China University of Petroleum-Beijing **

ABSTRACT

Unequal illumination of the subsurface highly impacts the quality of seismic imaging. Different image points of the media have different folds of reflection-angle illumination, which can be caused by irregular acquisition or by wave propagation in complex media. To address this problem, we present a method of stacking angle-domain common-image gathers (ADCIGs), in which we use local similarity with soft thresholding to decide the folds of local illumination. Normalization by local similarity regularizes local illumination of reflection angles for each image point of the subsurface model. This approach can restore good fidelity of amplitude by selective stacking in the image space, whatever the cause of acquisition or propagation irregularities. We use two synthetic examples to demonstrate that our method can normalize migration amplitudes and effectively suppress migration artifacts.

INTRODUCTION

The image of geology produced by Kirchhoff migration or wave-equation migration often suffers from artifacts, especially in the case of irregular acquisition or complex wave propagation (25; 32). Artifacts sometimes bury seismic reflectors, especially in subsalt areas, and make interpretation and model building difficult. Developments in acquisition technology (14; 20) have provided richer coverage and better illumination of the subsurface; however, areas with poor illumination still exist in complex geological environments, such as subsalt zones.

(24; 17) used preconditioned inversion in the reflection angle domain to improve imaging in complex media. (36; 13) proposed focusing on dip and AVA compensation in common-image gathers to suppress artifacts and improve signal-noise ratio in Kirchhoff migration. To compensate irregular illumination at image points, (15) presented illumination-angle compensation in Kirchhoff migration, in which inclusion of an amplitude-weighting term can dramatically improve the migrated image. (6; 2) proposed a method of regularization of illumination in the multiangle

***e-mail:** sergey.fomel@beg.utexas.edu

domain for Kirchhoff migration, in which a hit-count technique is applied to suppress migration artifacts and balance migration amplitudes. (25) presented interactive dip-gather stacking for attenuating artifacts of Kirchhoff migration. (32) analytically demonstrated artifacts in angle-domain common-image gathers (ADCIGs) by sparsely-sampled wavefields and proposed selective stacking on the basis of local smoothing of the envelope function in shot-profile wave-equation migration. (19) proposed the MAZ-stack method for weighting signal in areas of poor illumination for multiazimuth seismic data, in which weights of stacking were chosen as binary (zero or one).

Local similarity is a local attribute measured between two signals (9) that has been applied to multicomponent seismic image registration (11; 9) and time-lapse image registration (12). In an earlier work (18), we also applied local similarity as a weight for stacking common-midpoint gathers in order to improve signal-noise ratio of seismic data.

In this paper, we present another application of local similarity in stacking ADCIGs in order to normalize illumination. Our method applies local similarity between initial image and ADCIGs using a soft threshold to normalize stacking. It can attenuate migration artifacts and restore migration amplitudes. This method can be regarded as true-amplitude illumination compensation arising from the so-called Beylkin determinant (3; 1; 2).

ANGLE-DOMAIN COMMON-IMAGE GATHERS

ADCIGs can be produced both by Kirchhoff methods (36; 13) and wave-equation methods (23; 21; 35; 27; 5). In wave-equation migration methods, angle gathers can be produced either in the data space during imaging or in the image space after imaging (8). These methods can be also applied in source-receiver migration (27), shot-profile migration (26), and reverse-time migration (4; 37).

To extract ADCIGs, in wave equation migration we first compute subsurface offset-domain common-image gathers (ODCIGs) using crosscorrelation imaging condition in space and time, followed by extraction at zero time (29):

$$I(\mathbf{m},\mathbf{h}) = \sum_{\omega} U_s(\mathbf{m}-\mathbf{h},\omega)U_r^*(\mathbf{m}+\mathbf{h},\omega)\ . \tag{1}$$

Here, $I(\mathbf{m},\mathbf{h})$ is the subsurface-offset dependent image, $\mathbf{m}=[m_x,m_y,m_z]$ is a vector of locations of image points, $\mathbf{h}=[h_x,h_y,h_z]$ is a vector of local source-receiver separation in the image space, and $*$ means the conjugate.

In two dimensions, angle-domain common-image gathers $I(\mathbf{m},\theta)$ can be obtained by a simple slant-stack operation on ODCIGs $I(\mathbf{m},\mathbf{h})$ after migration (31; 33; 27). For three dimensions, we can extract ADCIGs after applying imaging condition by transforming local offset gathers in the depth domain (8; 28), in which local structural

dips need to be estimated. In examples herein, we extract ADCIGs in the image space after applying imaging conditions and test our proposed method on a 2D case.

MEASUREMENT OF LOCAL SIMILARITY

(9) defined the local similarity attribute using shaping regularization (10). The global correlation coefficient between two discrete signals a_t and b_t is defined as

$$\gamma = \frac{\sum_{i=1}^{N} \mathbf{a}_t \mathbf{b}_t}{\sqrt{\sum_{i=1}^{N} \mathbf{a}_t{}^2 \sum_{i=1}^{N} \mathbf{b}_t{}^2}} \ . \tag{2}$$

To locally measure correlation between two signals according to the definition of local similarity (9; 12), local similarity γ_t can be represented as the product of two least-squares inverses:

$$\gamma_t{}^2 = p_t q_t \ , \tag{3}$$

$$p_t = \arg \min_{p_t} \left(\sum_t \left(a_t - p_t b_t\right)^2 + R\left[p_t\right] \right) \ , \tag{4}$$

$$q_t = \arg \min_{q_t} \left(\sum_t \left(a_t - q_t b_t\right)^2 + R\left[q_t\right] \right) \ , \tag{5}$$

where R is a regularization operator designed to constrain the solution in a desired behavior, such as smoothness. Shaping regularization (10) can conveniently be applied in solving inverse problems (4) and (5) iteratively. The local similarity can smoothly measure the correlation between two signals locally. It is an estimate of waveform similarity of two seismic signals.

STACKING USING LOCAL SIMILARITY FOR NORMALIZATION OF ILLUMINATION

Once we obtain ADCIGs by migration, the final image of subsurface can be obtained by stacking over the angle axis in ADCIGs. Assuming reflectivity is angle independent, we can write equal-weight stacking as

$$\widehat{I}(\mathbf{m}) = \frac{1}{N} \sum_{\theta=1}^{N} \mathbf{I}(\mathbf{m}, \theta) \ . \tag{6}$$

where N is the number of samples in the reflection angle. To get rid of the artifacts caused by poor sampling of the sources or receiver wavefield or complex structure, (32) suggested stacking only angles with good illumination and dense sampling and applied selective stacking using envelopes as weights of stacking.

In this paper, we use local similarity with soft thresholding as weights to normalize the stacking:

$$\widetilde{I}(\mathbf{m}) = \frac{1}{KN}\sum_{\theta=1}^{N} w(\mathbf{m},\theta)\mathbf{I}(\mathbf{m},\theta)\ , \tag{7}$$

$$w(\mathbf{m},\theta) = \begin{cases} \gamma(\mathbf{m},\theta)-\alpha & \text{if } s(\mathbf{m},\theta) > \alpha \\ 0 & \text{if } s(\mathbf{m},\theta) \le \alpha \end{cases}\ . \tag{8}$$

Here, $\gamma(\mathbf{m},\theta)$ is local similarity between initial imagecomputed $\widehat{I}$ by equation refeq:e6 and ADCIGs $I(\mathbf{m},\theta)$. Local similarity can be computed by equations 3 through refeq:e5. Equation 8 is soft thresholding (7), which can define the fold of local illumination for each image point, $\mathbf{m}$. Normalizationis K the sum of weights. Normalization N in equation 7, the number of samples with $w(\mathbf{m},\theta) \neq 0$, can restore migration amplitude by discarding some samples of angle, which are likely artifacts or noise. We assume that the discarded samples in ADCIGs are entirely artifacts or noise, probably caused by irregular acquisition or complex propagation in the complex subsurface model.

In the following, we use a simple example to illustrate our method. Figure 1a is a simple synthetic ADCIG with 30 samples of reflection angle, and we add Gaussian random noise to the ADCIG. Assume that the reflection coefficients are equal and angle independent. Note that the third and fourth reflectors are not entirely illuminated. The image created by directly stacking this ADCIG is shown in Figure 1d. Third and fourth reflectors are not restored in equal-weight stacking because of poor illumination. The local similarity between the direct equal-weight stacking image (Figure 1c) and ADCIG (Figure 1a) computed by equations 3 through 5 is shown in Figure 1b. Note that the local similarity can approximate local illumination of the reflection angle. (34) presented a method of computing illumination distribution as a function of reflection angle on the basis of one-way wave equation. We apply soft thresholding in local similarity to select angles with good illumination to contribute to the stacking. And then we use the number of samples of reflection angles with good illumination to normalize the image. In comparing our method with the equal-weight stacking method, note that the third and the fourth reflection coefficients are balanced well and the signal-noise ratio is also improved by our method (Figure 1e). Amplitudes in Figure 1e are almost the same as those of the ideal image (Figure 1c).

In Kirchhoff integral migration, Beylkin's theory (3) shows how to take into account irregularities of illumination caused by acquisition irregularities or wave propagation in complex media (2). The Beylkin determinant estimates the local smooth illumination density function by computation of traveltime between surface locations and subsurface image point. In this paper, the local fold of illumination in the denominator of equation 7 is estimated by local similarity with soft thresholding. Because this estimation of the fold of local illumination is computed after imaging, the method can be thought of as postprocessing after migration.

When reflection coefficients are angle dependent (known as amplitude versus angle [AVA]), the proposed method can be regarded as an average effect of all angle-

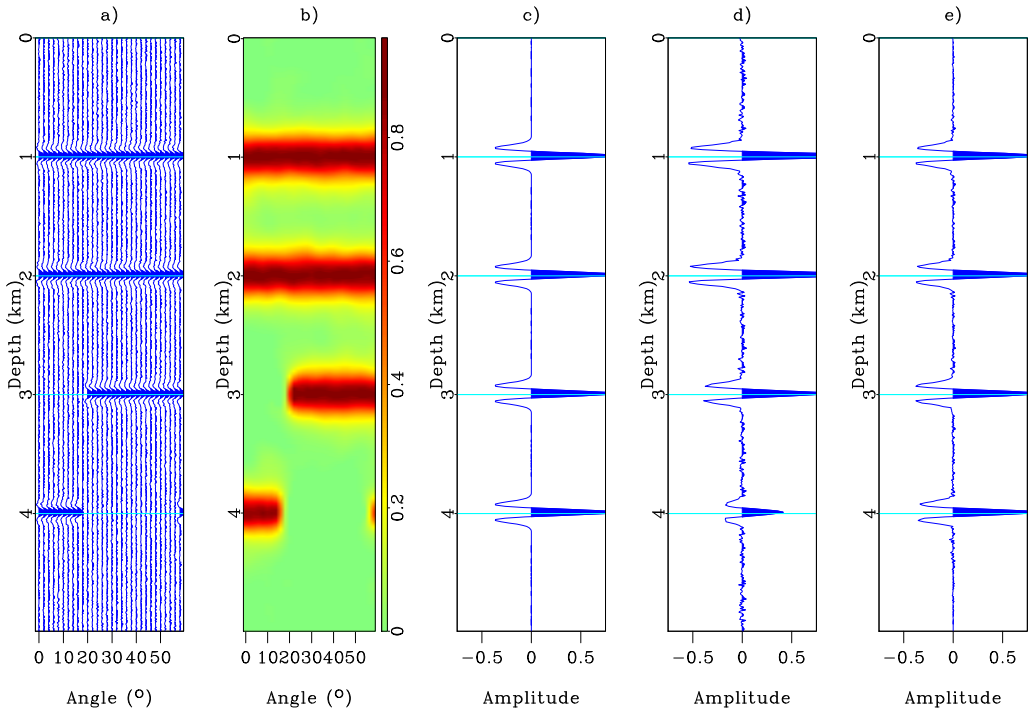


Figure 1: A simple example. (a) Synthetic ADCIGs with different folds; (b) local similarity; (c) ideal reflectivity; (d) initial equal-weight stacking; (e) stacking with normalization using local similarity. stackadcig/simple compare1

dependent reflection coefficients. It can enhance the image and restore relative amplitudes. When considering the AVA effect in stacking processing, one should use the accurate AVA inversion to obtain the reflection coefficient of the same reflection angle for all image points (17), such as zero-reflection-angle reflectivity, or retrieve the map of relative slowness perturbations from multishot seismic data (16).

EXAMPLES

Two synthetic examples illustrate the method of stacking ADCIGs for illumination normalization.

The first example involves a simple constant-velocity model with three reflectors. The reflecting interfaces are caused by density contrasts, which are chosen in such a way that all reflection coefficients for all reflectors and reflection angles are identical. A similar model was applied by (38) and (30). The synthetic dataset is generated using Kirchhoff modeling. It includes 25 shots between 3,000 and 5,000 m, and range of offset is -3,000 to 3,000 m. PSPI wave equation migration is performed for generating ADCIGs (Figure 2(a)). The three reflectors can be found to have different folds caused by acquisition geometry. The parts that are not illuminated in ADCIGs do not all have zero values; some are occupied by noise or artifacts. Therefore, appropriate normalization when stacking all reflection coefficients of different reflection angles together must be ensured. The local similarity cube (Figure 2(b)) can provide normalization for stacking. By comparing equal-weight stacking and our method (Figure 3), we find that image amplitudes are improved, especially on the sides of the model, which are not entirely illuminated because of acquisition.

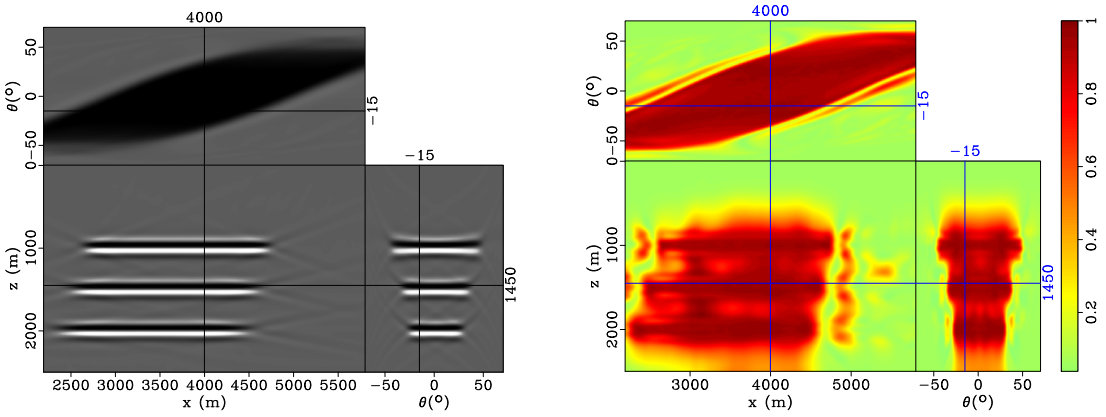


Figure 2: (a) ADCIGs for the simple synthetic example; (b) local similarity cube. stackadcig/flat ppangtr-h,similtr

The second example involves the Sigsbee2A model (22). This dataset is known to have illumination problems below the salt. Figure 4 shows the ADCIGs of Sigsbee2A model. Figures 5(a) and 6(a), respectively, show images of 0 ° and 20 ° reflection

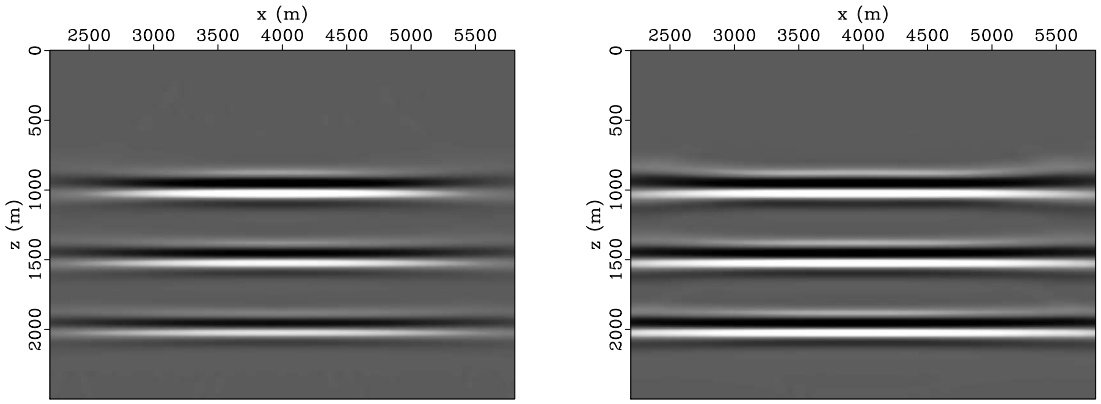


Figure 3: (a) Equal-weight stack; (b) stack normalized by local similarity. stackadcig/flat stack,simistack

angles. Figures 5(b) and 6(b), respectively, show the local similarity cube of 0 ° and 20 ° reflection angles. Comparing local similarity cubes of different reflection angles, we can find that folds of local illumination are not equal for all image points. The folds are closely uniform in areas with simple structures, but differences can be found in complex media, such as in subsalt areas (Figures 5(b) and 6(b)). Figure 7 shows the result of equal-weight stacking, and Figure 8 shows the result of our method. With normalization, reflection coefficients are restored, such as below the salt (Figure 8). Our method's image beneath the salt is clearer than that of the conventional stacking method, and amplitude of the base line, with strong reflectivity at the bottom of the Sigsbee2A model, is more balanced. Note that fault and diffraction points below the salt become clearer after normalization using local similarity.

CONCLUSIONS

We have presented a method for stacking angle-domain common-image gathers for illumination normalization in which we use a local similarity cube between initial image and ADCIGs as the approximate fold of local illumination for each image point of the media. Normalization by local fold can restore migration amplitude and attenuate artifacts possibly caused by irregular acquisition geometry or by wave propagation in complex media. Our method is easy to implement and can be seen as a postprocessing step after migration. Two synthetic examples show that this method can be used to enhance migrated images and to restore migration amplitudes.

ACKNOWLEDGMENTS

The first author would like to thank Long Jin for inspiring discussions and the China Scholarship Council (grant 2007U44003) for partial support of this work. We would

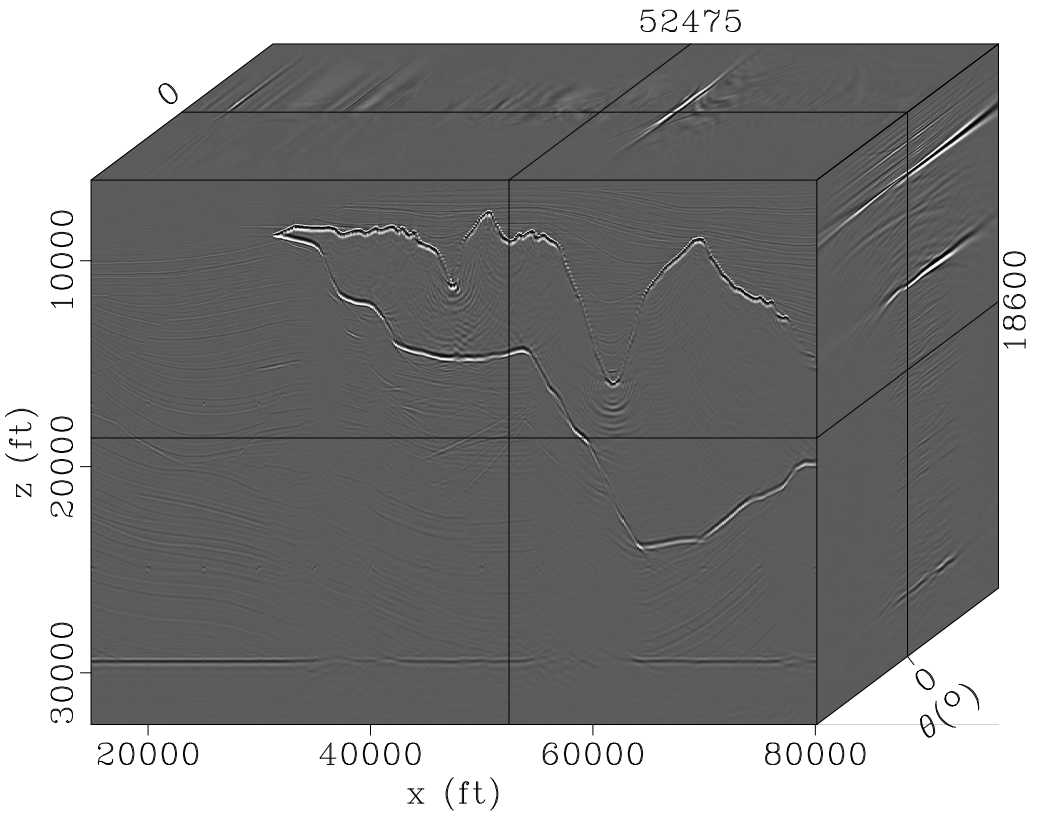


Figure 4: ADCIGs of the Sigsbee2A Model. stackadcig/sigsbee angs

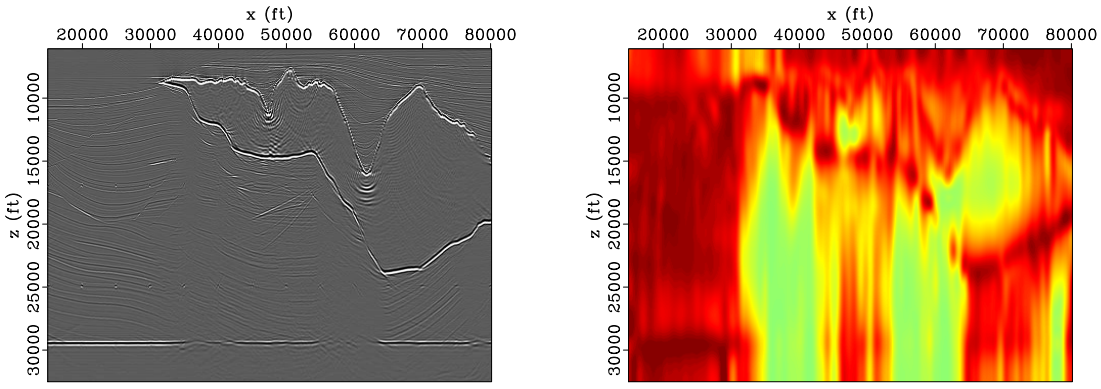


Figure 5: 0° angle image and (a) its local similarity (b) of the Sigsbee2A model. stackadcig/sigsbee ang1,simil1

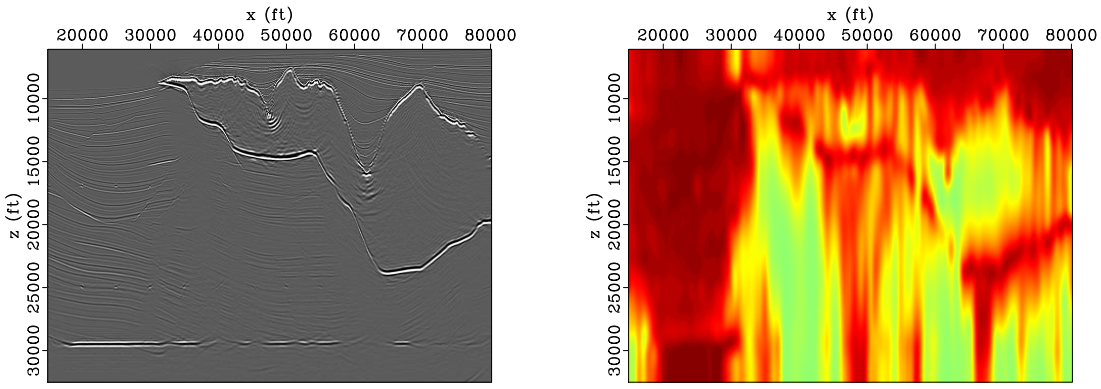


Figure 6: 20° angle image and (a) its local similarity (b) of the Sigsbee2A model. stackadcig/sigsbee ang2,simil2

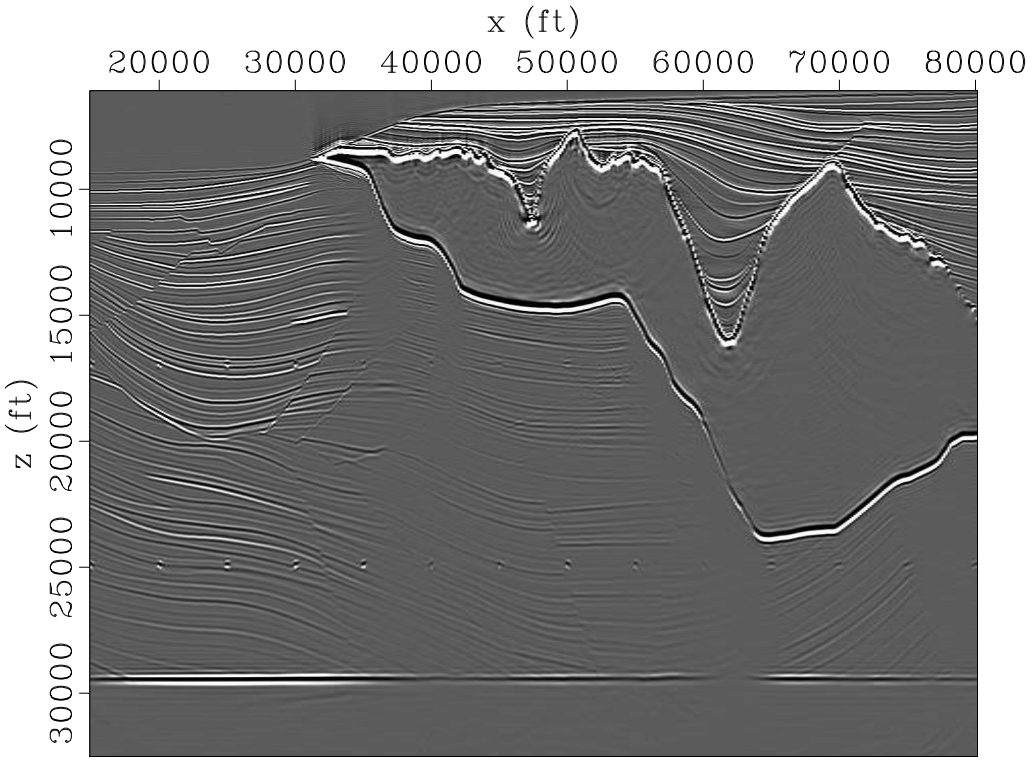


Figure 7: The image of Sigsbee2A using initial equal-weight stacking. stackadcig/sigsbee sigstack

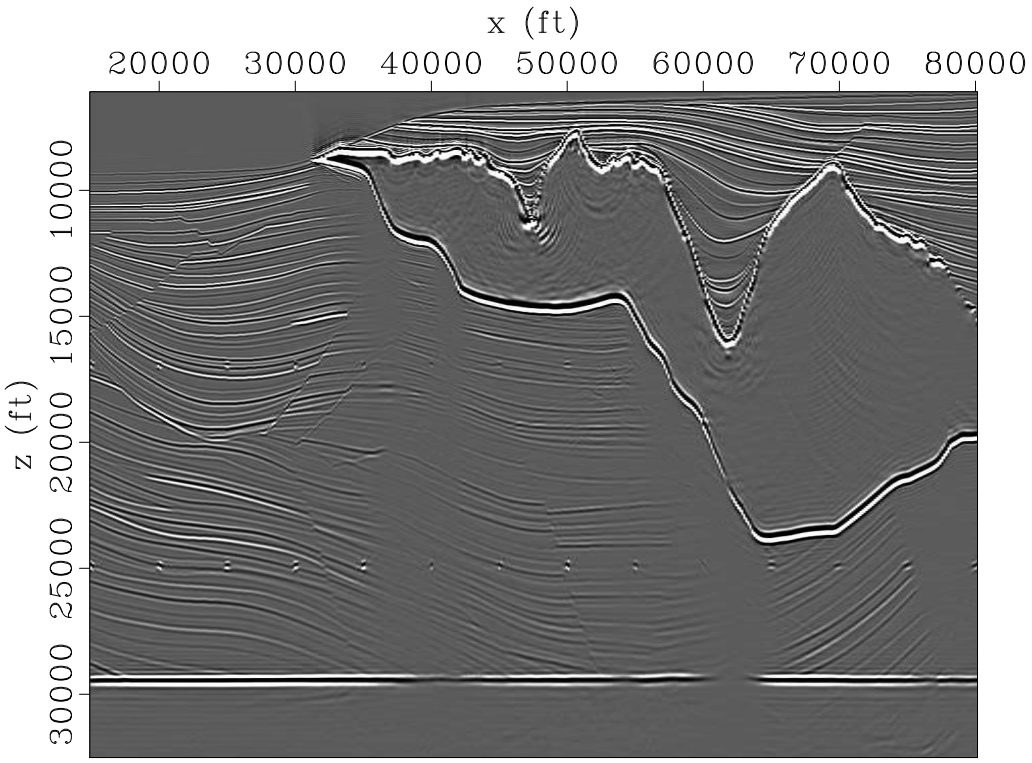


Figure 8: The image of Sigsbee2A using stacking with normalization by local similarity. stackadcig/sigsbee sigsimistack

like to thank Paul Sava for help with wave-equation migration code.

REFERENCES

[1] Albertin, U., H. Jaramillo, D. Yingst, R. Bloor, W. Chang, C. Beasley, and E. Mobley, 1999, Aspects of true amplitude migration: 69th Annual International Meeting, Soc. of Expl. Geophys., 1358–1361.

[2] Audebert, F., L. Nicoletis, and P. Froidevaux, 2005, Regularization of illumination in angle domains–a key to true amplitude migration: The Leading Edge, **24**, 643–654.

[3] Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inverse of a causal generalized radon transform: Journal of Mathematics and Physics, **26**, 99–108.

[4] Biondi, B., and G. Shan, 2002, Prestack imaging of overturned reflections by reverse time migration: 72th Annual International Meeting, Soc. of Expl. Geophys., 1284–1287.

[5] Biondi, B., and W. Symes, 2004, Angle-domain common image gathers for migration velocity analysis by wavefield-continuation imaging: Geophysics, **69**, 1283–1298.

[6] Bloor, R., U. Albertin, H. Jaramillo, and D. Yingst, 1999, Equalized prestack depth migration: 69th Annual International Meeting, Soc. of Expl. Geophys., 1362–1365.

[7] Donoho, D. L., 1995, De-noising by soft-thresholding: IEEE Transaction on Information Theory, **41**, 613–627.

[8] Fomel, S., 2004, Theory of 3-d angle gathers in wave-equation imaging: 74th Annual International Meeting, Soc. of Expl. Geophys., 1053–1058.

[9] ———, 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.

[10] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[11] Fomel, S., M. Backus, K. Fouad, B. Hardage, and G. Winters, 2005, A multistep approach to multicomponent seismic image registration with application to a west texas carbonate reservoir study: 75th Annual International Meeting, Soc. of Expl. Geophys., 1018–1021.

[12] Fomel, S., and J. Long, 2009, Time-lapse image registration using the local similarity attribute: Geophysics, **74**, A7–A11.

[13] Hoop, S. B.-D. M. V. D., and B. Ursin, 2003, Focusing in dip and ava compensation on scattering-angle/azimuth common image gathers: Geophysics, **68**, 232–254.

[14] Howard, M. S., and N. Moldoveanu, 2006, Marine survey design for rich-azimuth seismic using surface streamers: 76th Annual International Meeting, Soc. of Expl. Geophys., 2905–2909.

[15] Kessinger, W., 2004, Illumination angle compensation in kirchhoff depth migration: 76th Annual International Meeting, Soc. of Expl. Geophys., 1017–1021.

[16] Kiyashchenko, D., R. E. Plessix, B. Kashtan, and V. Troyan, 2007, A modi-

fied imaging principle for true-amplitude wave-equation migration: Geophysical Journal International, **168**, 1093–1104.
[17] Kuhl, H., and M. D. Sacchi, 2003, Least-square wave-equation migration for avp/ava inversion: Geophysics, **68**, 262–273.
[18] Liu, G., S. Fomel, L. Jin, and X. Chen, 2009, Seismic data stacking using local correlation: Geophysics, **74**, V43–V48.
[19] Manning, T., C. Page, S. A. Hall, J. Keggin, B. Barley, W. Rietveld, E. Fromyr, and R. V. Borselen, 2008, Leveraging the value of multi-azimuth (maz) seismic through maz-stack: 70th conference and Exhibition, EAGE, G013.
[20] Michell, S., E. Shoshitaishvili, D. Chergotis, J. Sharp, and J. Etgen, 2006, Wide azimuth streamer imaging of mad dog: Have we solved the subsalt imaging problem?: 76th Annual International Meeting, Soc. of Expl. Geophys., 2905–2909.
[21] Mosher, C., and D. Foster, 2000, Common angle imaging conditions for prestack depth migration: 70th Annual International Meeting, Soc. of Expl. Geophys., 830–833.
[22] Paffenholz, J., 2000, Sigsbee2 synthetic subsalt dataset: image quality as function of migration algorithm and velocity model error: 71th Annual International Meeting, Soc. of Expl. Geophys., Workshop W–5.
[23] Prucha, M., B. Biondi, and W. Symes, 1999, Angle-domain common-image gathers by wave-equation migration: 69th Annual International Meeting, Soc. of Expl. Geophys., 824–827.
[24] Prucha, M. L., R. G. Clapp, and B. Biondi, 2000, Seismic image regularization in the reflection angle domain: SEP-103, 109–119.
[25] Qin, F., B. Wang, P. Zhang, and F. Audebert, 2005, Kirchhoff presdm interactive dip-gather stacking and dip illumination panel generation: 75th Annual International Meeting, Soc. of Expl. Geophys., 1882–1885.
[26] Rickett, J. E., and P. C. Sava, 2002, Offset and angle-domain common image-point gathers for shot-profile migration: Geophysics, **67**, 883–889.
[27] Sava, P. C., and S. Fomel, 2003, Angle-domain common-angles by wavefield continuation methods: Geophysics, **68**, 1065–1074.
[28] ———, 2005, Coordinate-independent angle-gathers for wave equation migration: 75th Annual International Meeting, Soc. of Expl. Geophys., 2052–2055.
[29] ———, 2006, Time-shift imaging condition in seismic migration: Geophysics, **71**, 209–217.
[30] Schleicher, J., J. C. Costa, and A. Novais, 2008, A comparison of imaging condition for wave-equation shot-profile migration: Geophysics, **73**, 219–227.
[31] Stolt, R. H., and A. B. Weglein, 1985, Migration and inversion of seismic data: Geophysics, **50**, 2458–3472.
[32] Tang, Y., 2007, Selective stacking in the reflection-angle and azimuth domain: 77th Annual International Meeting, Soc. of Expl. Geophys., 2320–2324.
[33] Weglein, A. B., and R. H. Stolt, 1999, Migration-inversion revisited: The Leading Edge, **18**, 950–957.
[34] Xie, X., S. Jin, and R. Wu, 2006, Wave-equation-based seismic illumination analysis: Geophysics, **71**, 169–177.
[35] Xie, X., and R. Wu, 2002, Extracting angle domain information from migrated

wavefield: 72th Annual International Meeting, Soc. of Expl. Geophys., 1360–1363.
[36] Xu, S., H. Chauris, G. Lambare, and M. Noble, 2001, Common-angle migration: A strategy for imaging complex media: Geophysics, **66**, 1877–1894.
[37] Zhang, Y., J. Sun, and S. Gray, 2007, Reverse-time migration: amplitude and implementation issues: 77th Annual International Meeting, Soc. of Expl. Geophys., 2147–2149.
[38] Zhang, Y., G. Zhang, and N. Bleistein, 2003, Theory of true-amplitude one-way wave equations and true amplitude common-shot migration: Geophysics, **70**, E1–E10.

Automated spectral recomposition with application in stratigraphic interpretation

Yihua Cai[*†], *Sergey Fomel*[†], *and Hongliu Zeng*[†*]

ABSTRACT

Analyzing seismic attributes in the frequency domain is helpful for reservoir characterization. To analyze the reservoir interval of interest in detail, it is important to capture the seismic response at each frequency subset. Spectral recomposition can be used to extract significant components from the seismic spectrum. We propose a separable nonlinear least-squares algorithm for spectral recomposition, which estimates both linear and nonlinear parts automatically in separate steps. Our approach is applied to estimate fundamental signal parameters, peak frequencies and amplitudes, with which the seismic spectrum can be reconstructed. Automated spectral recomposition helps us visualize frequency-dependent geological features on both cross sections and time slices by extracting significant frequency components. Spectral recomposition can also indicate how frequency contents attenuate with time.

INTRODUCTION

Frequency-domain seismic attributes can be useful in stratigraphic and hydrocarbon reservoir characterization (3; 10). If the seismic response can be captured at each frequency subset, the reservoir interval of interest can then be scrutinized in greater detail. Spectral decomposition is a technique that was developed at Amoco in the 1990's (16). Various time-frequency analysis methods have been employed for frequency decomposition since then. (7) applied short-time Fourier transform, which unfortunately suffers from a time-frequency resolution limit (4). (12) and (5) applied spectral decomposition in the time domain by decomposing the input seismogram into constituent wavelets and then summing the Fourier spectra of individual wavelets. This approach experiences difficulties when the frequency range is large, because it relies on the accuracy of wavelet decomposition, whose residuals commonly introduce bias into "frequency gathers" (5). (11) and (12) implemented spectral decomposition by time-frequency analysis using an iterative inversion framework with the help of local attributes.

(19) defined *frequency recomposition* in seismic forward modeling as an estimation of components of the seismic spectrum. They showed how to make forward seismic

[*]**e-mail:** sergey.fomel@beg.utexas.edu

models by recomposing single-frequency models into a multi-frequency model. Different from all previous approaches, spectral recomposition models and reconstructs the seismic spectrum instead of decomposing it. However, (19) manually picked component frequencies and amplitudes, which may not be accurate and highly depends on personal experience. In this paper, we propose spectral recomposition using separable nonlinear least-squares estimation (9), which simultaneously and automatically estimates both linear and nonlinear parts of the Ricker wavelet spectrum. It provides an accurate and direct estimation of amplitudes and peak frequencies of various Ricker wavelets.

A problem is separable if the model can be represented as a linear combination of functions that have a nonlinear parametric dependence. A separable least-squares estimation fits frequencies and amplitudes with large variations, and provides computing confidence, as well as prediction and calibration intervals. The Gauss-Newton algorithm, a method of minimizing the residual sum of squares, is effective both when residuals are small and when measurement errors are additive and the data set is large (15). An analogous method was used previously by (2) for fitting von Kármán distributions, and by (13) for fitting a single Ricker wavelet.

We represent a seismic spectrum as the sum of different Ricker components and use the Gauss-Newton method to fit it with a sum of Ricker wavelet spectra so as to estimate the peak frequency and amplitude of each component. On a field data example from the Gulf of Mexico, we show that automatic spectral recomposition can improve seismic stratigraphic interpretation and help in seismic attribute studies.

THEORY

We represent a seismic spectrum as the sum of different Ricker components (19):

$$d(f) \approx \sum_{i=1}^{n} a_i \psi_i(m_i, f) \ , \tag{1}$$

where $d(f)$ is the spectrum of a seismic trace, and a_i and m_i are the amplitude and peak frequency of the i-th Ricker spectrum component, given as

$$R(f) = a\,\psi(m, f) = a\,\frac{f^2}{m^2}\,exp(-\frac{f^2}{m^2}) \ . \tag{2}$$

Thus, the model is a linear combination of Ricker wavelet spectra, which has nonlinear functions and depends on multiple parameters. To estimate the Ricker wavelet spectra, we need both $\mathbf{a} = \{a_1, a_2, ..., a_n\}$ and $\mathbf{m} = \{m_1, m_2, ..., m_n\}$ coefficients. The estimation error is

$$r_j = d(f_j) - \sum_{i=1}^{n} a_i(m_i)\psi_i(m_i, f_j) \ . \tag{3}$$

The optimal least-squares estimation requires

$$\min_{\mathbf{a},\mathbf{m}} \|\mathbf{r}(\mathbf{a}, \mathbf{m})\|_2^2 \ . \tag{4}$$

The goal of separable nonlinear least-squares estimation (1) is to find a global minimizer of the sum of squares of nonlinear functions. The separability aspect comes from solving linear and nonlinear parts separately (17). The algorithm we use in this paper is known as the variable projection algorithm (9). It provides solutions for **a** and **m** by exploring the fact that **r** depends on **a** linearly.

NUMERICAL METHOD

Assuming nonlinear parameters of **m**, linear parameters of **a** can be obtained by solving the linear least-squares problem,

$$\mathbf{a} = \boldsymbol{\psi}(\mathbf{m})^{\dagger}\mathbf{d} \ , \tag{5}$$

where $\boldsymbol{\psi}(\mathbf{m})$ is the matrix composed of $\psi_i(m_i, f_j)$ and $\boldsymbol{\psi}(\mathbf{m})^{\dagger}$ is the Moore-Penrose generalized inverse of the $\boldsymbol{\psi}(\mathbf{m})$ matrix. Replacing this **a** in the original function, the minimization problem takes the form

$$\min_{\mathbf{m}} \left\| (\boldsymbol{I} - \boldsymbol{\psi}(\mathbf{m})\boldsymbol{\psi}(\mathbf{m})^{\dagger})\mathbf{d} \right\|_2^2 \ , \tag{6}$$

where the linear parameters have been eliminated (9). We use the Gauss-Newton method (1) to linearize the problem as follows:

$$\begin{aligned} d(f_i) &\approx \sum_i R_j(m_i, f_j) + \sum_i \frac{\partial R_j}{\partial m_i}\Delta m_i \\ &\approx \sum_i a_i \, \boldsymbol{\psi}(m_i, f_j) + \sum_i [a_i' \, \boldsymbol{\psi}(m_i, f_j) + a_i \, \boldsymbol{\psi}'(m_i, f_j)]\Delta m_i. \end{aligned} \tag{7}$$

Starting with initial values of m_i, we are able to solve for a_i and a_i' using equation 5. Then we solve for the model increment $\Delta\, m_i$. After a number of iterations, summation of $\Delta\, m_i$ converges to the estimated value. The Gauss-Newton method is efficient. In most cases, approximately 20 iterations provide an acceptable convergence. Fitting more component frequencies helps minimize the residual. Geological factors can help the user decide how many components to include in the model. In addition, providing good initial values helps the algorithm avoid being trapped in a local minimum.

APPLICATION

In this section, we show how spectral recomposition can be used in interpretation. First, we test automated spectral recomposition using synthetic and field data. We then use automated spectral recomposition to simulate and estimate how frequency components attenuate in the subsurface. Finally, we perform thin-bed stratigraphic interpretation using data from Starfak and Tiger Shoal fields of offshore Louisiana.

Synthetic and field data test

To test our method, we generated a wavelet composed of three Ricker wavelets with peak frequencies of 10, 20, and 50 Hz (Figure 1.) Applying a separable nonlinear least-squares estimation, we estimated peak frequencies at 9.999, 19.999, and 49.995 Hz. The residual sum of squares equals approximately 10^{-7} after 50 iterations. The estimation fits the wavelet spectrum accurately, as shown in Figure 1. Generally speaking, the more terms we use to fit the data, the better fitting result we may obtain. However, it is geological reasoning, not statistical factors that determines how many terms should be used in the model. A real data example, shows that spectral recomposition works well for field data (Figure 2). In this case, the estimated peak frequencies are approximately 15 Hz, 31 Hz, and 43 Hz.

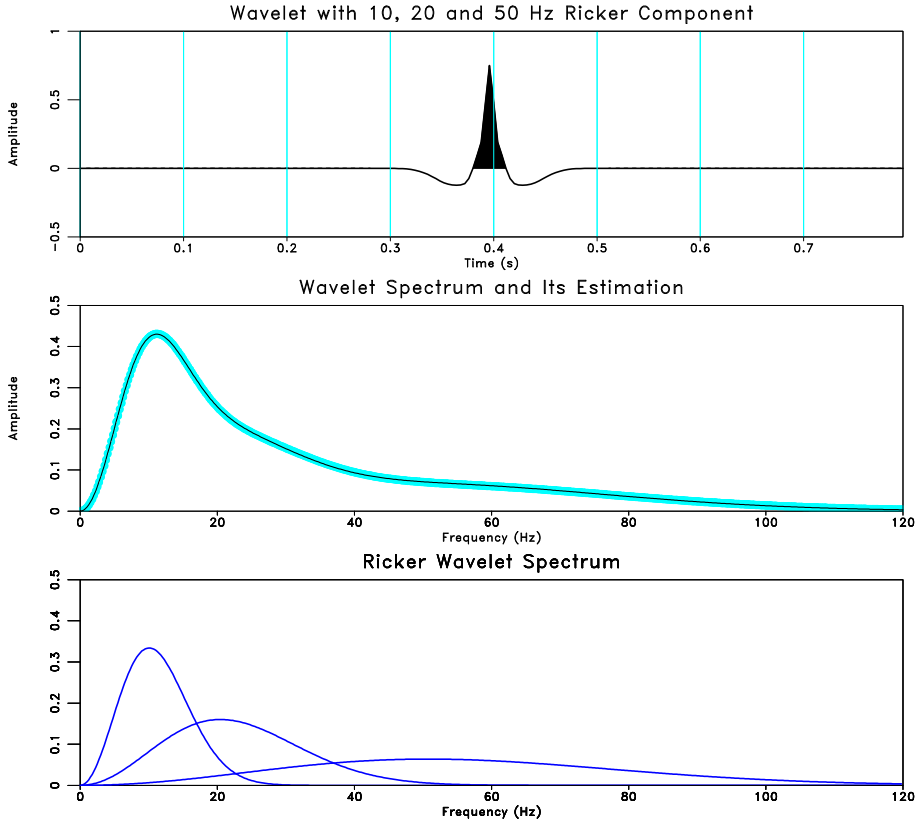


Figure 1: (a) A wavelet composed of 10, 20 and 50 Hz Ricker components. (b) The estimated wavelet spectrum components are plotted individually. The estimated peak frequencies of these components are 9.999, 19.999 and 49.995 Hz. (c) We computed and estimated the spectrum of the wavelet in (a). Spectral recomposition result (in green) fits the spectrum (in black) very well. specrecom/ricker rk

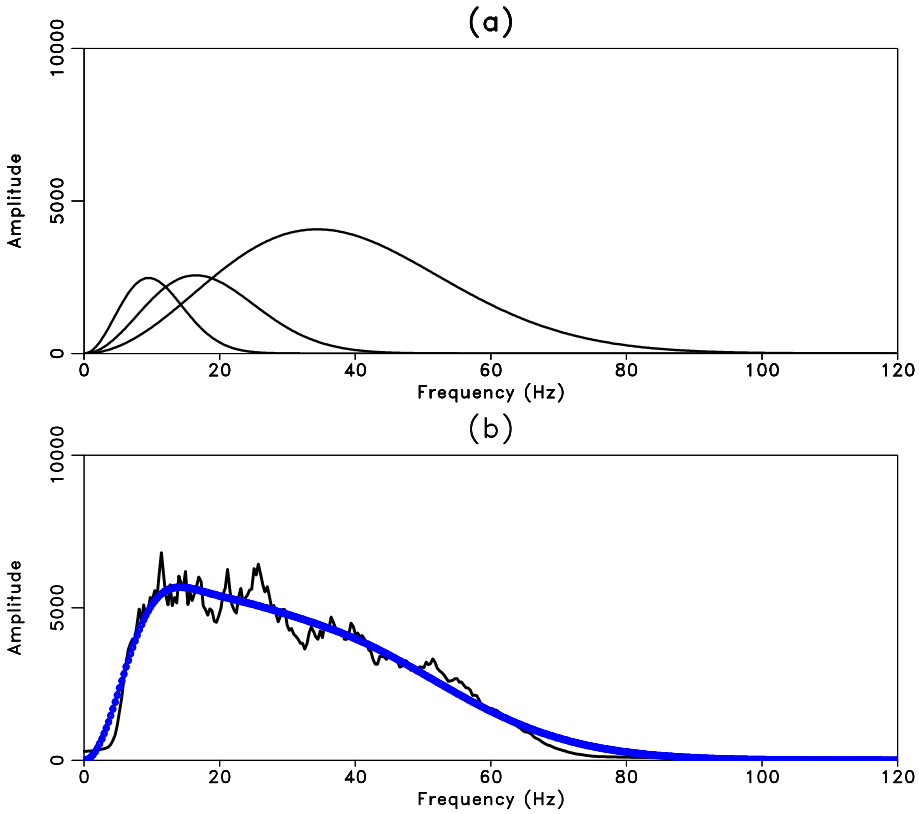


Figure 2: (a) Spectra of ricker wavelet components. Each spectral component is well separated, probably indicating different geological factors. (b) Spectral recomposition result (in blue) fits the seismic spectrum (in black) well. specrecom/gulf recomp

Frequency attenuation estimation

Spectral recomposition can be used to indicate how various frequency components attenuate in the subsurface. Different from spectral decomposition, spectral recomposition extracts significant components from seismic data by modeling the spectrum, i.e. fitting spectrum data with summation of Ricker components. With the help of local time-frequency analysis (11), we re-composed the spectra of a seismic trace at each time depth. Figure 3 shows recomposition examples at four different time depths. Figure 4 compares our spectral recomposition result with the result of time-frequency analysis. At approximately 1 second, one of the components has a peak frequency as high as 65 Hz. It attenuates gradually to about 8 Hz in 4 seconds. Spectral decomposition indicates information of whole frequency spectrum at each time depth, but provides us no extracted key components. It can be easily found that the spectral recomposition result is consistent with time-frequency analysis result. Thus, spectral recomposition provides information related to any specific layer in which the user might be interested and builds a deeper understanding of seismic attenuation in the subsurface.

Stratigraphic interpretation

Setting the bandwidth for the frequency component ensures an accurate interpretation. Carrying valuable geological information, the main lobe of a wavelet has a finite duration, which means that it covers a certain range in the frequency domain. However, interpreters tend to pick a single frequency with bandpass filtering, which hardly covers the corresponding seismic information because a single frequency appears as a spike ignoring all other frequency content. Spectral recomposition helps in computation of frequency bandwidth of the main lobe. One can compute $t_d = \sqrt{6}/(f_m\,\pi)$ as in Figure 5, where t_d is the dominant wavelength and f_m is the peak frequency. The first zero crossing is computed as $t_r = t_d/\sqrt{3}$ (18). Having estimated the dominant wavelength, we can set the corresponding bandwidth in the frequency domain.

Spectral recomposition extracts significant components from seismic data. Picking the peak frequencies and setting their bandwidths produce results, as shown in Figures 6(b) and 6(c). The same cross section is displayed with a manually picked peak frequency in Figure 6(a). Note that Figures 6(b) and 6(c) reveal more geologic features. Reflections and seismic events are displayed better in Figure 6(b) and 6(c); for example, a fault system clearly shows up in both Figure 6(b) and 6(c), but not in Figure 6(a). Deeper events are also well displayed in Figure 6(c), but not in Figure 6(a).

The concept behind spectral recomposition is that the seismic response of the geologic unit has a characteristic expression in the frequency domain that is indicative of its significant components (19). Hence, spectral recomposition can be used in time-slice or stratal-slice interpretation. We start by picking a depositional surface (geologic-time surface) so that any seismic attribute extracted on such a surface could

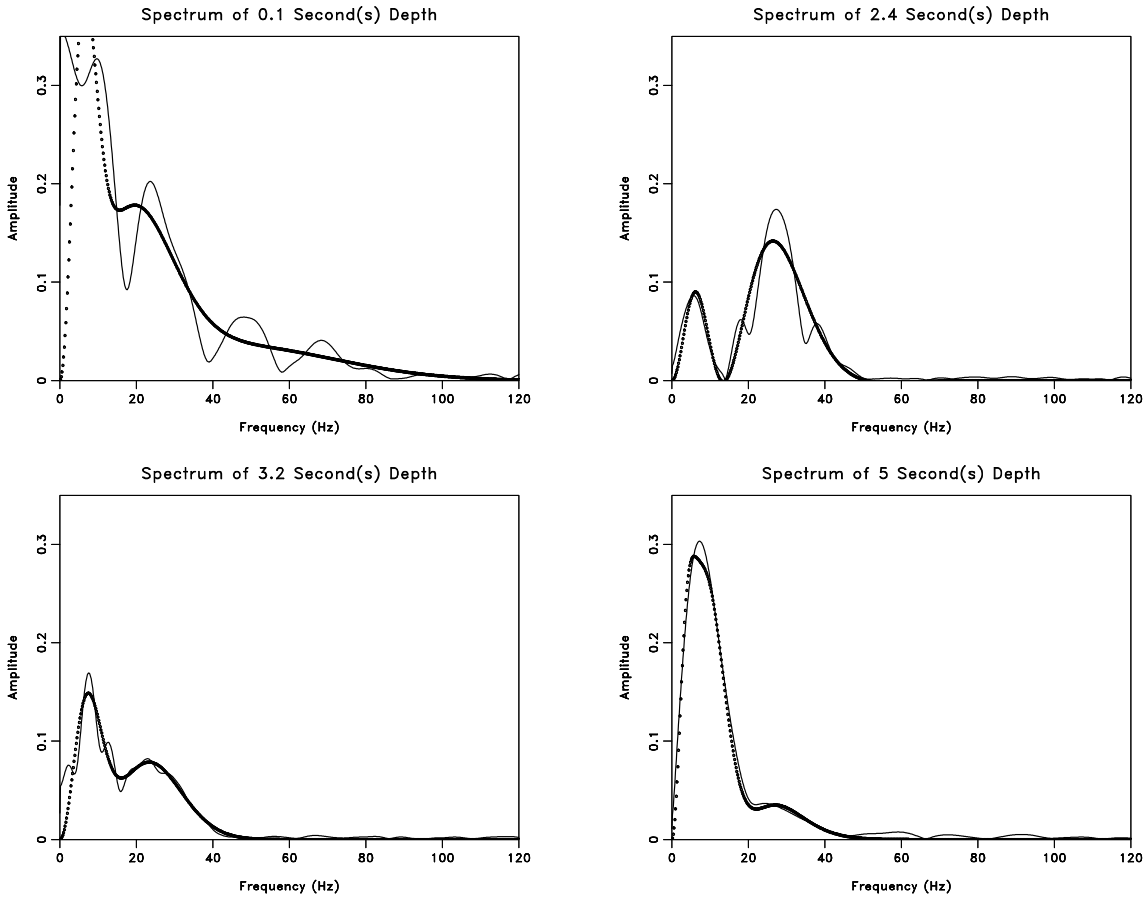


Figure 3: Four time depths, 0, 2, 3 and 5 second have been plotted in (a), (b), (c) and (d) to show spectral recomposition reconstructs Ricker component at different depths. Field data results have been plotted in solid black lines. The spectral recomposition results have been plotted in dotted red lines. specrecom/tf tf2

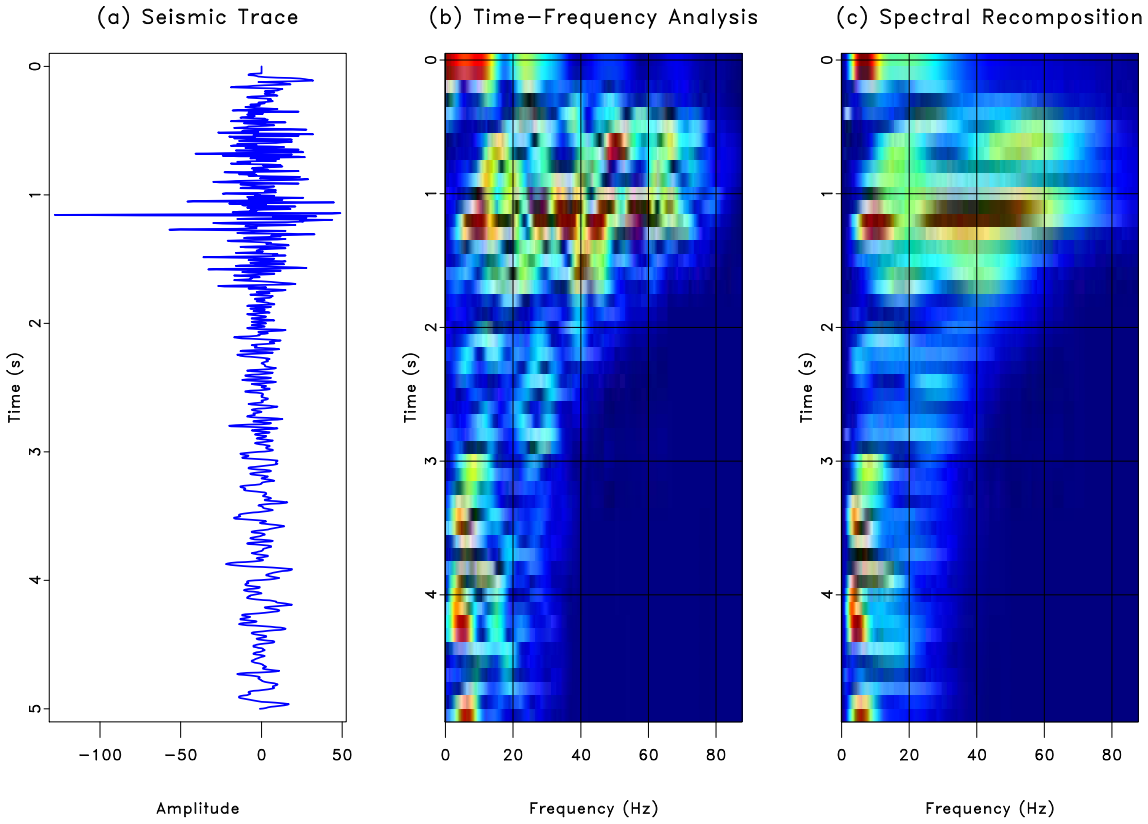


Figure 4: (a) One single seismic trace; (b) Time-frequency analysis result of the trace in (a); (c) Automated spectral recomposition reconstructs Ricker components at each time depth. The reconstructed result is quite consistent with time-frequency analysis in (b). specrecom/tf tf

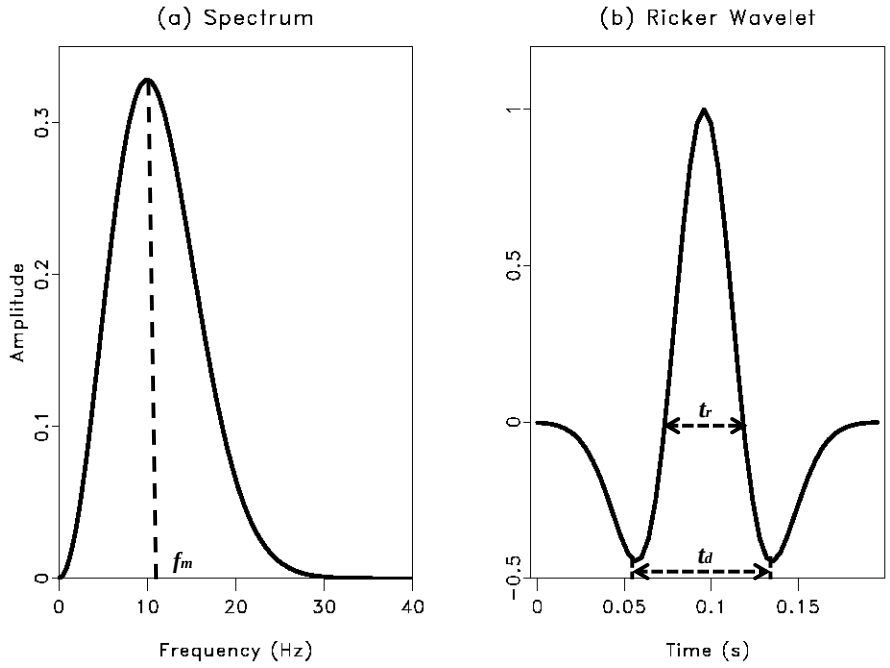


Figure 5: After extracting peak frequency f_m, we compute the dominant wavelength and bandpass filter. (a) Spectrum with peak frequency f_m. (b) Ricker wavelet with dominant wavelength t_d and first zero crossing t_r. specrecom/. rkband

represent a genetic depositional unit. Such a seismic surface display was termed a stratal slice (21; 22). To showcase the "recomposition" technique developed here, we used data from Starfak and Tiger Shoal fields of offshore Louisiana, a 135-mile^2 3-D survey area. The study area lies along the western periphery of the ancestral Mississippi River depocenter (14), most recently designated the central Mississippi sediment-dispersal axis by (8).

(24) constructed and interpreted stratal slices by using special tools to restore and refine multiple seismic slices in the wire-line-log context in both the traveltime domain and the relative geologic-time domain. Their approach includes conditioning seismic data to log lithology by 90^o phasing to achieve better well log integration, imaging, and interpretation of the sequential, planoform geomorphology of the depositional systems (20). Our goal is to use spectral recomposition to further improve the image of seismic depositional facies; e.g. a high-frequency component helps indicate thin stratigraphic layers, while a low-frequency component helps identify thick stratigraphic layers. Depositional facies of different thickness can be differentiated using spectral recomposition. After extracting significant spectral components and setting band pass filters, we plotted the stratal slices of Starfak and Tiger Shoal field. Figure 7(a) is an upper Miocene-age horizon extracted by (23). Spectral recomposition shows that the component with 22 Hz peak frequency is a significant one, thus we have Figure 7(b). As can be seen, Figure 7(b) improves imaging of the architec-

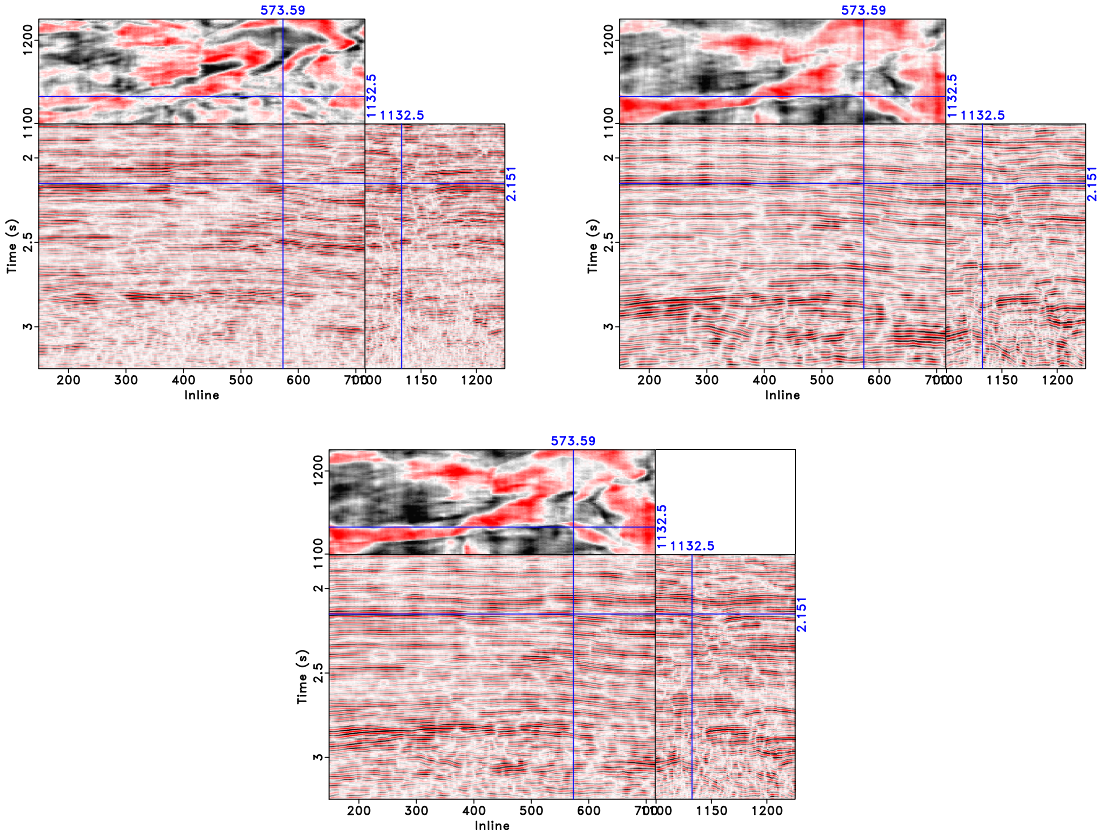


Figure 6: (a) Seismic data from Starfak and Tiger Shoal field displayed with manually picked frequency bandpass filter. Using spectral recomposition, we display the same seismic data with two significant components in (b) and (c). A fault system between inline 800 to 900 clearly stands out in (b). Deep events below 2.4s are well displayed in (c). However, neither the fault system nor the deep events are well displayed in (a). specrecom/gulf sgr330,sgr317,sgr321

tural elements within the depositional system. The thin, narrow channel sandstones in relict delta are more clearly imaged in Figure 7(b) than in Figure 7(a). The shelf systems can be easily identified in Figure 7(b) but not in Figure 7(a). Figure 7(b) is also less affected by acquisition footprint compared with Figure 7(a). More than that, spectral recomposition works more efficiently. Both true time and thickness information has been extracted in the process. No prior time or depth-thickness estimate is required for automatic spectral recomposition.

After extracting significant signal components, the result of spectral recomposition naturally admits a Red-Green-Blue (RGB) color-blending plot. RGB color-blending plot takes advantage of the fact that human eyes can discriminate a large numbers of colors. One can plot horizontal slices of different frequency content in different colors, either red, green or blue, and combine them into one figure (6). Picking frequency subsets automatically with solid geological and geophysical reasoning, spectral recomposition helps standardize the workflow of the RGB color-blending plot. As we have seen in Figure 3, spectral recomposition extracts dominant components at each time depth. Given a time slice or stratal slice, we extract the significant frequency components, compute the duration of the main lope of the wavelet from the area of interest, and set the appropriate bandwidth in the frequency domain. Plotting three different frequency components in RGB colors and combining them together, we produce RGB color-blending plots in Figure 8. With a sense of relative sandstone thickness distribution, RGB color blending of spectral recomposition brings useful information together and helps reveal geological features of interest.

CONCLUSION

Automated spectral recomposition using separable nonlinear least squares represents the seismic spectrum as a sum of Ricker components and efficiently estimates their peak frequencies and amplitudes. With the seismic spectrum reconstructed from component frequencies, spectral recomposition can be used in seismic interpretation. We adopted Ricker wavelet in the analysis because of its popularity. Other wavelets may also be used with corresponding estimation numerical strategy.

Applying spectral recomposition, we have been able to better visualize seismic images in both cross sections and stratal slices. This technique has improved the interpreter's ability to image the various elements of the depositional system in extracted stratal slices. Spectral recomposition can also be used in forward modeling, in studying how different frequency components attenuate in the subsurface, newand in estimating thin-bed thicknesses using tuning frequencies. It provides a robust and phase-independent approach to seismic thickness estimation. Compared with conventional methods involving adjacent peaks and troughs picking, spectral recomposition requires only peak-frequency and amplitude estimation.

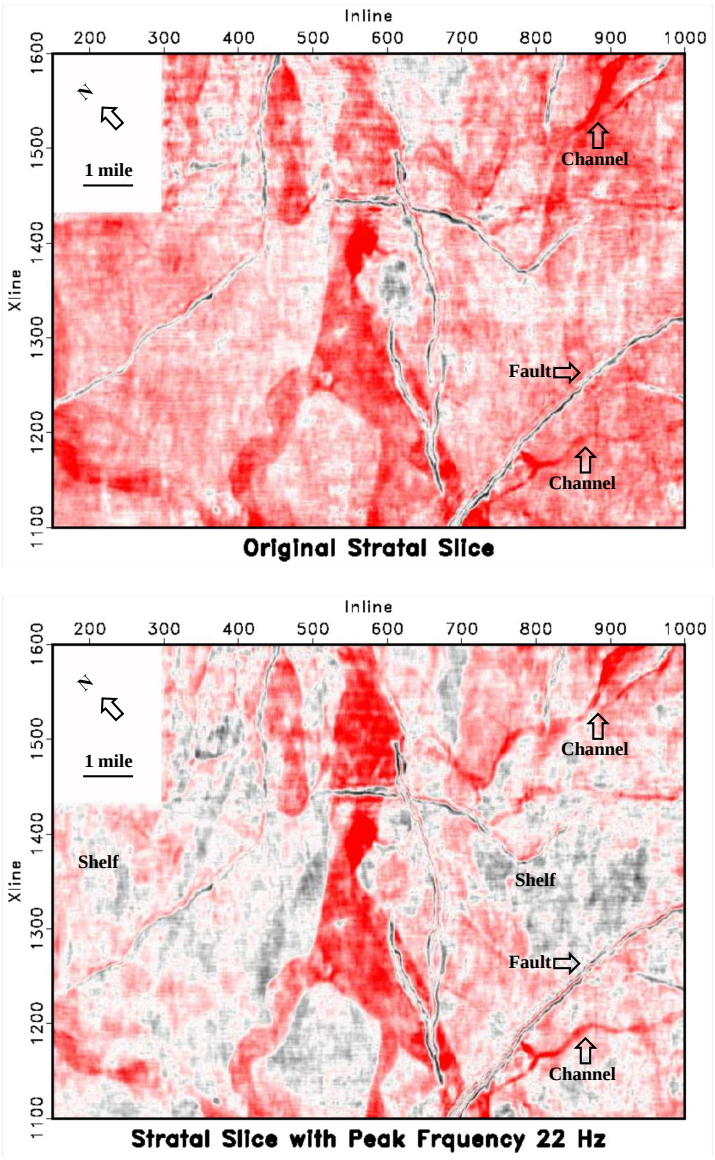


Figure 7: Use of spectrum recomposition to improve thin-bed stratigraphic interpretation in Miocene, Starfak-Tiger Shoal fields, offshore Louisiana. (a) A stratal slice made by (24). (b) Same stratal slice extracted from 22 Hz component of seismic data. specrecom/. slice0,slice1

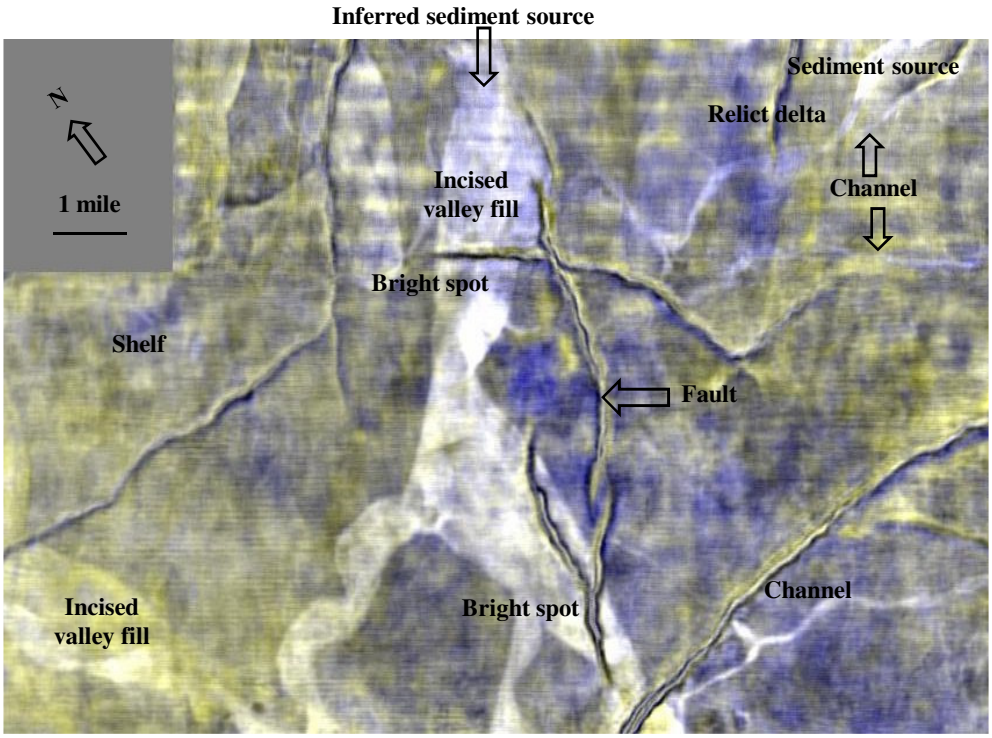


Figure 8: Components of 15, 22, and 29 Hz of Starfak-Tiger Shoal stratal slice picked and plotted using color blending, which provides more detailed information. White bright colors indicate thicker sands; white dim colors refer to thinner sands; blue and dark colors indicate shale. Interpreters can easily identify different depositional facies in plot. specrecom/. rgb

ACKNOWLEDGEMENT

We thank Chevron for the seismic data and the Madagascar development community for developing the software used in this paper. Publication is authorized by the Director, Bureau of Economic Geology.

REFERENCES

[1] Björck, Å., 1996, Numerical Methods for Least Squares Problems: Society for Industrial and Applied Mathematics.

[2] Browaeys, T. J., and S. Fomel, 2009, Fractal heterogeneities in sonic logs and low-frequency scattering attenuation: Geophysics, **74**, WA77–WA92.

[3] Castagna, J. P., S. Sun, and R. W. Siegfried, 2003, Instantaneous spectral analysis: Detection of low-frequency shadows associated with hydrocarbons: The Leading Edge, **22**, 120–127.

[4] Chakraborty, A., and D. Okaya, 1995, Frequency-time decomposition of seismic data using wavelet-based methods: Geophysics, **60**, 1906–1916.

[5] Chen, G., G. Matteucci, B. Fahmy, and C. Finn, 2008, Spectral-decomposition response to reservoir fluids from a deepwater West Africa reservoir: Geophysics, **73**, C23–C30.

[6] Chopra, S., and K. Marfurt, 2007, Seismic attributes for prospect identification and reservoir characterization: Society of Exploration Geophysicists.

[7] Dilay, A., and J. Eastwood, 1995, Spectral analysis applied to seismic monitoring of thermal recovering: The Leading Edge, **14**, 1117–1122.

[8] Galloway, W. E., P. E. Ganey-Curry, X. Li, and R. T. Buffler, 2000, Cenozoic depositional history of the Gulf of Mexico Basin: AAPG Bulletin, **84**, 1743–1774.

[9] Golub, G. H., and V. Pereyra, 1973, The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate: Numerical Analysis, **10**, 413–432.

[10] Li, Y., X. Zheng, and Y. Zhang, 2011, High-frequency anomalies in carbonate reservoir characterization using spectral decomposition: Geophysics, **76**, V47–V57.

[11] Liu, G., S. Fomel, and X. Chen, 2011, Time-frequency analysis of seismic data using local attributes: Geophysics, **76**, P23–P34.

[12] Liu, J., 2006, Spectral Decomposition and Its Application in Mapping Stratigraphy and Hydrocarbons: PhD thesis, Univeristy of Houston.

[13] Liu, Y., and S. Fomel, 2010, Local time-frequency transform and its application to ground-roll noise attenuation, *in* 80th Annual International Meeting: Society of Exploration Geophysicists, 3711–3716.

[14] McGookey, D. P., 1975, Gulf coast cenozoic sediments and structures: An excellent example of extra-continental sedimentation: Gulf Coast Association of Geological Societies Transactions, **25**, 104–120.

[15] Osborne, M. R., 2007, Separable least squares, variable projection, and the Gauss-Newton algorithm: Electronic Transactions on Numerical Analysis, **28**, 1–15.

[16] Partyka, G. J., J. Gridley, and J. Lopez, 1999, Interpretational applications of spectral decomposition in reservior characterization: The Leading Edge, **18**, 353–360.

[17] Scolnik, H. D., 1972, On the solution of non-linear least squares problems: IFIP Congress, IFIP Congress, 1258–1265.

[18] Sheriff, R., 2002, Encyclopedic Dictionary of Applied Geophysics: Soc. Expl. Geophys.

[19] Tomasso, M., R. Bouroullec, and D. R. Pyles, 2010, The use of spectral recomposition in tailored forward seismic modeling of outcrop analogs: AAPG Bulletin, **94**, 457–474.

[20] Zeng, H., and M. M. Backus, 2005, Interpretive advantages of 90°-phase wavelets. Part 1: Modeling: Geophysics, **70**, C7–C15.

[21] Zeng, H., M. M. Backus, K. T. Barrow, and N. Tyler, 1998a, Stratal slicing, part I: Realistic 3-D seismic model: Geophysics, **63**, 502–513.

[22] Zeng, H., S. C. Henry, and J. P. Riola, 1998b, Stratal slicing, part II: Real 3-D seismic data: Geophysics, **63**, 514–522.

[23] Zeng, H., and T. F. Hentz, 2004, High-frequency sequence stratigraphy from seismic sedimentology: Applied to Miocene, Vermilion Block 50, Tiger Shoal area, offshore Louisiana: AAPG Bulletin, **88**, 153–174.

[24] Zeng, H., T. F. Hentz, and L. W. Wood, 2001, Stratal slicing of Miocene-Pliocene sediments in Vermilion Block 50-Tiger Shoal area, offshore Louisiana: the Leading Edge, **20**, 408–418.

Nonlinear structure-enhancing filtering using plane-wave prediction

Yang Liu[*†], *Sergey Fomel*[†], *Guochang Liu*[‡*]

ABSTRACT

Attenuation of random noise and enhancement of structural continuity can significantly improve the quality of seismic interpretation. We present a new technique, which aims at reducing random noise while protecting structural information. The technique is based on combining structure prediction with either similarity-mean filtering or lower-upper-middle (LUM) filtering. We use structure prediction to form a structural prediction of seismic traces from neighboring traces. We apply a nonlinear similarity-mean filter or an LUM filter to select best samples from different predictions. In comparison with other common filters, such as mean or median, the additional parameters of the nonlinear filters allow us to better control the balance between eliminating random noise and protecting structural information. Numerical tests using synthetic and field data show the effectiveness of the proposed structure-enhancing filters.

INTRODUCTION

Extracting structural information is the most important goal of seismic interpretation. A number of approaches have been proposed to preserve and enhance structural information. (13) applied nonlinear filters, such as median, trimmed mean, and adaptive Gaussian, over planar surfaces parallel to the structural dip. (4) and (11) applied structure-oriented filtering based on anisotropic diffusion. (9) suggested the method of plane-wave construction. (1) developed 3-D edge-preserving smoothing (EPS) operators to reduce noise without blurring sharp discontinuities. (24) used a mix of finite-impulse-response (FIR) and infinite-impulse-response (IIR) filters, followed by f-x filtering, to perform structure-preserving seismic processing. (26) introduced a frequency-dependent, structurally conformable filter. Random noise attenuation and structure protection are always ambivalent problems in seismic data processing. The main challenge in designing effective structure-enhancing filters is controlling the balance between noise attenuation and signal preservation.

Mean filter (stacking) plays an important role in improving signal-to-noise ratio in seismic data processing (27). Conventional stacking is a simple and effective method of denoising, but it is optimal only when noise has a Gaussian distribution (18). Therefore, alternative stacking methods were proposed. (21) proposed a

[*]**e-mail:** sergey.fomel@beg.utexas.edu

signal-to-noise-ratio-based weighted stack to further minimize noise. (25) developed optimum-weighted stacking. To eliminate artifacts in angle-domain common-image gathers, (22) presented a selective-stacking approach, which applied local smoothing of envelope function to achieve the weighting function. (20) proposed "smart stacking," which is based on optimizing seismic amplitudes of the stacked signal by excluding harmful samples from the stack and applying more weight to the center of the samples population. (16) introduced time-dependent smooth weights to design a similarity-mean stack, which is a nonstationary, nonlinear operator.

Lower-upper-middle (LUM) filters (12) are a class of rank-order-based filters that generalize the concept of the median filter. They combine the characteristics of two subclasses, LUM smoothers and LUM sharpeners, and find a variety of signal and image-processing applications. A special LUM smoother, probably still the most widely used, is a simple median of the points within a window. The median filter is a well-known method that can effectively suppress spike-like noise in nonstationary signal processing. Median filters find important applications in seismic data analysis. (19) incorporated noise reduction using median filter into standard Kirchhoff time migration. (28) used a hyperbolic median filter to suppress multiples. (15; 17) implemented random-noise attenuation using 2-D multistage median filter and nonstationary time-varying median filter. However, in many cases the median filter introduces too much smoothing. LUM sharpeners have excellent detail preserving characteristics because of their ability to pass small details with minimal distortion. For a given window size and shape, the parameters of LUM filter can be adjusted independently to yield a wide range of characteristics.

In this paper, we combine structure prediction (8) with either nonlinear similarity-mean filtering or lower-upper-middle (LUM) filtering to define a structure-enhancing filter. Structure prediction generates structure-conforming predictions of seismic events from neighboring traces, whereas a similarity-mean filtering or an LUM filtering reduces several predictions to optimal output values. We test the proposed method on synthetic and field data examples to demonstrate its ability to enhance structural details of seismic images.

THEORY

Structure prediction

The method of plane-wave destruction (2) uses a local plane-wave model for characterizing the structure of seismic data. It finds numerous applications in seismic imaging and data processing (5). Letting a seismic section, $\mathbf{s}$, be a collection of traces, $\mathbf{s} = [\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_N]^T$, the plane-wave destruction operation can be defined in a linear operator notation as

$$\mathbf{d} = \mathbf{D}(\sigma)\,\mathbf{s}\ , \tag{1}$$

where $\mathbf{d}$ is the destruction residual and $\mathbf{D}$ is the nonstationary plane-wave de-

struction operator defined as follows

$$\begin{bmatrix} \mathbf{d}_1 \\ \mathbf{d}_2 \\ \mathbf{d}_3 \\ \cdots \\ \mathbf{d}_N \end{bmatrix} = \begin{bmatrix} \mathbf{I} & 0 & 0 & \cdots & 0 \\ -\mathbf{P}_{1,2}(\sigma_1) & \mathbf{I} & 0 & \cdots & 0 \\ 0 & -\mathbf{P}_{2,3}(\sigma_2) & \mathbf{I} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -\mathbf{P}_{N-1,N}(\sigma_{N-1}) & \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{s}_1 \\ \mathbf{s}_2 \\ \mathbf{s}_3 \\ \cdots \\ \mathbf{s}_N \end{bmatrix} , \quad (2)$$

where $\mathbf{I}$ stands for the identity operator, σ_i is local dip pattern, and $\mathbf{P}_{i,j}(\sigma_i)$ is an operator for prediction of trace j from trace i according to the dip pattern σ_i. A trace is predicted by shifting it according to the local seismic event slopes. Minimizing the destruction residual $\mathbf{d}$ provides a method of estimating the local slopes σ (5).

The least-squares minimization of $\mathbf{d}$ is achieved by using iterative conjugate-gradient (CG) method and smooth regularization. Local dip at a fault position cannot be accurately estimated, its value will depend on the initial slope estimate and regularization.

Prediction of a trace from a distant neighbor can be accomplished by simple recursion, i.e., predicting trace k from trace 1 is simply

$$\mathbf{P}_{1,k} = \mathbf{P}_{k-1,k} \cdots \mathbf{P}_{2,3}\, \mathbf{P}_{1,2} \; . \quad (3)$$

(8) applied plane-wave prediction to predictive painting of seismic images. In this paper, we use a similar construction to recursively predict a trace from its neighbors.

An example is shown in Figure 1a. The input data is borrowed from (3): a synthetic seismic image containing dipping beds, an uncomformity, and a fault. Figure 1b shows the same image with Gaussian noise added. Figure 1c shows local slopes measured from the noisy image by plane-wave destruction. The estimated slope field correctly depicts the constant slope in the top part of the image and the sinusoidal variation of slopes in the bottom. In the next step, we predict every trace from its neighbor traces according to the local slope, as described by (8). We chose a total of 14 prediction steps (7 from the left and 7 from the right), which, with the addition of the original section, generated a data volume (Figure 1d). The prediction axis corresponds to index k in equation 1. The volume is flat along the prediction direction, which confirms the ability of plane-wave destruction to follow the local structure. However, it still contains some discontinuous information because of the faults. In the next step, we apply nonlinear structure-enhancing filtering to process the data along the prediction direction.

Choices of nonlinear filters

We have tried two different choices to select best samples from prediction data volume: Gaussian similarity-mean filter and lower-upper-middle (LUM) filter.

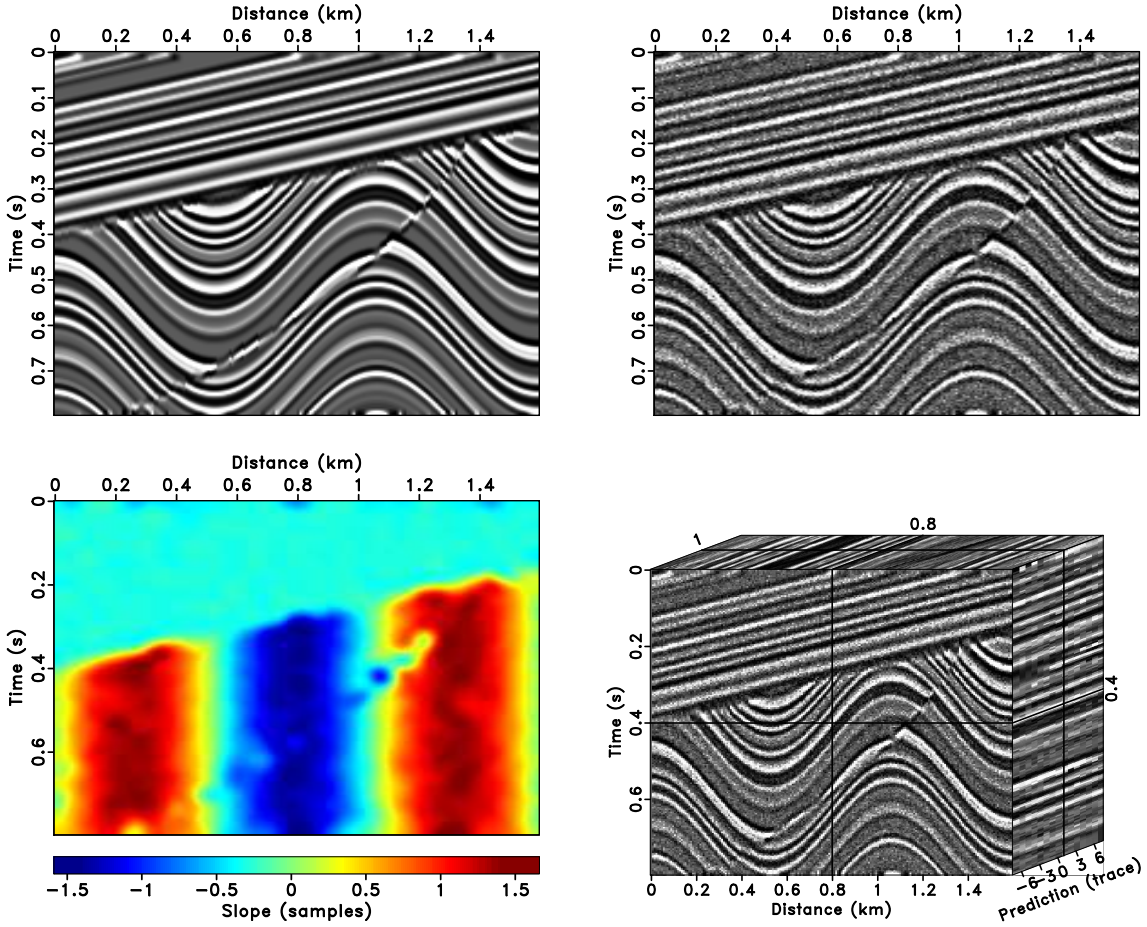


Figure 1: Noise-free synthetic image (a), noisy image (b), local slopes estimated from Figure 1b (c), and predictive data volume, where every trace is supplemented with predictions from its neighbors (d). strfilter/spray sigmoid1,gnoise1,ndip1,cube

Gaussian similarity-mean filter

The similarity-mean filter is a nonlinear filter that uses local correlation coefficients as desired weight coefficients (16). It is described in Appendix A. We chose the shaping operator with smoothing radius of 10 samples in time to calculate local similarity coefficients between the predicted data at each prediction step and original data (reference data). Figure 2a displays similarity-weight coefficients from local correlation. The elements with the shortest prediction distance get largest weights because they provide the most accurate prediction and therefore are most similar to the original image. We used zero-value boundary conditions for the prediction, so the predicted amplitudes from the left most and the right most sides are zero. This results in the similarity coefficients on the corners of the weight cube to be zero. In the weighted mean filter, large weight coefficients get selected when the similarity is strong between processed data and reference data. We introduce additionally Gaussian weights to localize the smoothing characteristics of the filter

$$w_i = e^{-h_i^2/h_r^2} \, , \tag{4}$$

where h_i is the distance to trace i and h_r is the reference parameter that controls the shape of the weight function. This construction is analogous to bilateral or non-local filtering (23; 10). Figure 2b shows the product of Gaussian weights and similarity weights. The prediction data volumes only with similarity weights applied and with Gaussian similarity weights applied are shown in Figure 2c and 2d respectively. After applying Gaussian and similarity weights, we stack the data in Figure 2d along the prediction direction. The result is shown in Figure 3b.

For comparison, we used the standard mean filter to process the prediction data volume (Figure 1d) along the prediction direction. The result is shown in Figure 3a and corresponds, in this case, to simple box smoothing along the local image structure. The standard mean filter simply stacks all information along the prediction direction. It enhances structural continuity but smears information across the fault.

For further discussion, we show the difference between the noisy image (Figure 1b) and structure-enhancing results with the standard mean filter and the Gaussian similarity-mean filter (Figure 3a and 3b). We kept the same scale of magnitude and plotting clips as that of the input image. From Figure 4a and 4b, the coherent events are well protected by the two methods because structure prediction can exactly predict coherent information. However fault information is destroyed by the mean filter (Figure 4a), while the similarity-mean filter provides a result where fault information is protected well, whereas random noise is attenuated (Figure 4b).

Lower-upper-middle filter

The lower-upper-middle (LUM) filter is a nonlinear filter that is simple to define and yet effective for noise attenuation in non-stationary signal processing (12). It has two

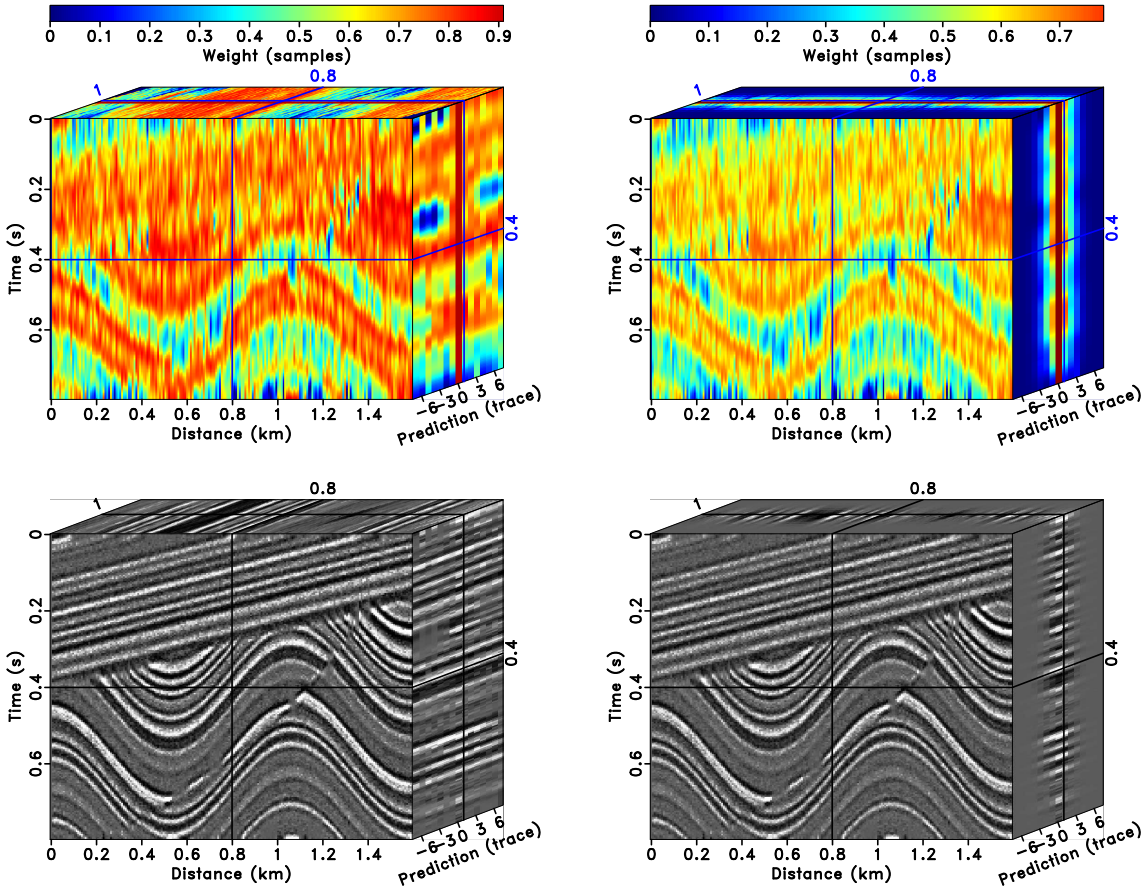


Figure 2: Similarity weights (a), product of Gaussian weights and similarity weights (b), the data only with similarity weights applied (c), and the data with Gaussian similarity weights applied (d). strfilter/spray weight01,weightcube1,pwdatacube,wdatacube

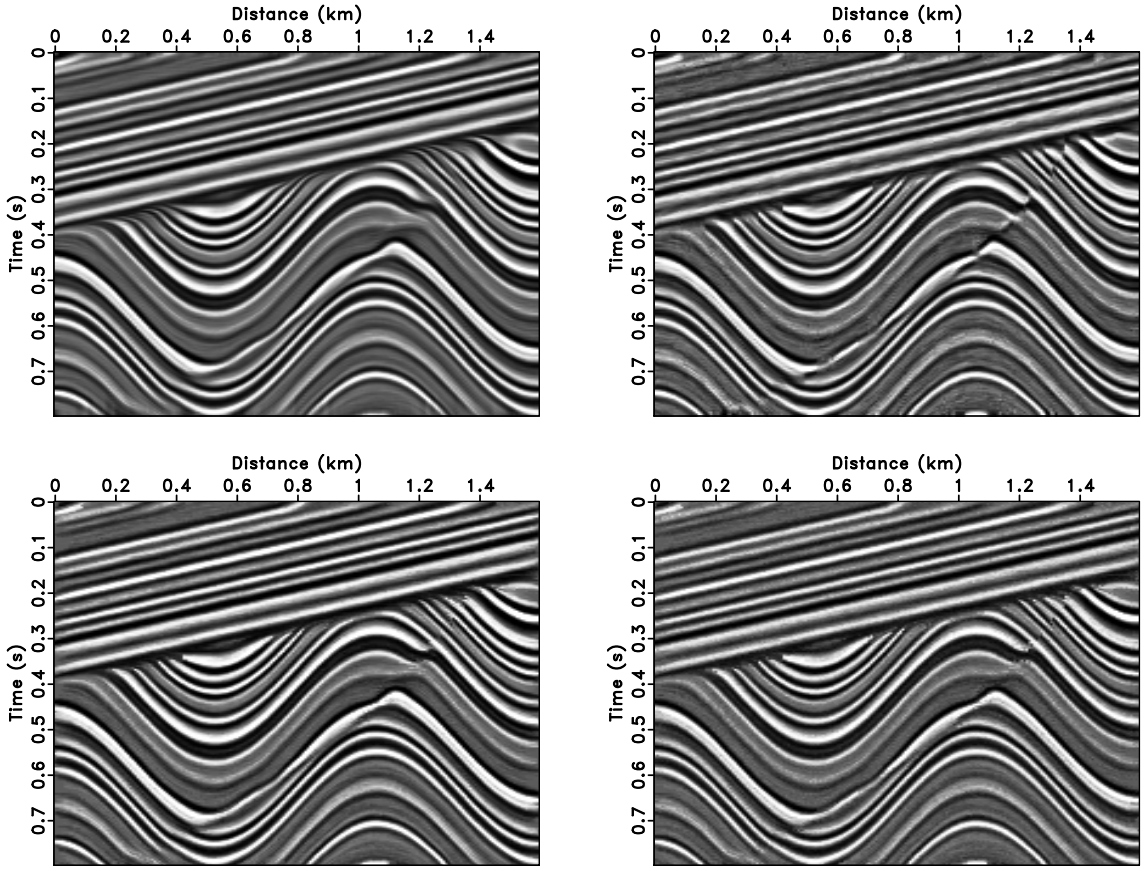


Figure 3: Structure-enhancing results using different methods. Standard mean filtering (a), similarity-mean filtering (b), standard median filtering with filter-window length 15 (c), and lower-upper-middle (LUM) filtering with parameters $k = l = 7$ (d). strfilter/spray mean1,gsimilarstack1,mf,median1

parameters, one for smoothing and the other for sharpening. A general class of LUM filters includes LUM smoothers and LUM sharpeners as special cases (Appendix B).

By manipulating the parameters k (for smoothing) and l (for sharpening), the lower-upper-middle (LUM) filter takes on a variety of characteristics. We found that $k = l = (N-1)/2$ works well for our synthetic and field data examples. In the synthetic example, special smoothing and sharpening parameters, $k = l = 7$, were chosen to balance the ability between noise attenuation and fault protection. After applying the LUM filter on the synthetic noisy image, we obtain the image shown in Figure 3d. Comparing with Figure 3b, the LUM filter displays the similar-quality result as Gaussian similarity-mean filter. However, the LUM filter is somewhat easier to control than the similarity-mean filter.

The standard median filter with filter-window length 15 is compared to the lower-upper-middle (LUM) filter. After applying the median filter on the prediction direction of Figure 1d, the result is shown in Figure 3c. When comparing with the mean filter (Figure 3a), the median filter has a better fault-protection ability but weaker noise-attenuation result. However, it still makes edges of some faults ambiguous. The LUM filter uses smoothing and sharpening parameters to limit the smoothing characteristics of the standard median filter. Therefore, it strikes a reasonable balance between structure enhancement and fault protection. Figure 4c and 4d show the difference between the noisy image (Figure 1b) and structure-enhancing results with the standard median filter (Figure 3c) and the LUM filter (Figure 3d). While coherent events can be preserved by either of the two filters, the median filter has a better result of fault protection than the mean filter (Figure 4c) and the LUM filter can further reduce the fault damage of the standard median filter (Figure 4d).

The key steps of our method are illustrated schematically in Figure 5.

FIELD DATA EXAMPLES

For the first field-data test, we use a time-migrated seismic image from a historic Gulf of Mexico dataset (3). The input is shown in Figure 6a. After estimating the field of local slopes (Figure 6b), we apply structure prediction to calculate the structural data volume (Figure 7a), which effectively flattens the data. As in the synthetic example, the length of the prediction axis is 15. Next, we apply two different filtering approaches. The first choice is the similarity-mean filter. We display the product of Gaussian weights and similarity weights in Figure 7b. After applying the similarity-mean filter, we can get a structure-enhanced image (Figure 8a): the fault structures are protected well while noise is removed. Our second choice is an lower-upper-middle (LUM) filter with parameters $k = l = 7$. The enhanced result is shown in Figure 8b. Similarly to the similarity-mean filter, the LUM filter highlights locally continuous reflectors while preserving the geometry of faults. Figure 9 shows the difference sections, in which the structure-enhancing results using two filtering methods (Figure 8) are subtracted from the original data (Figure 6a). After processing, the coherent events

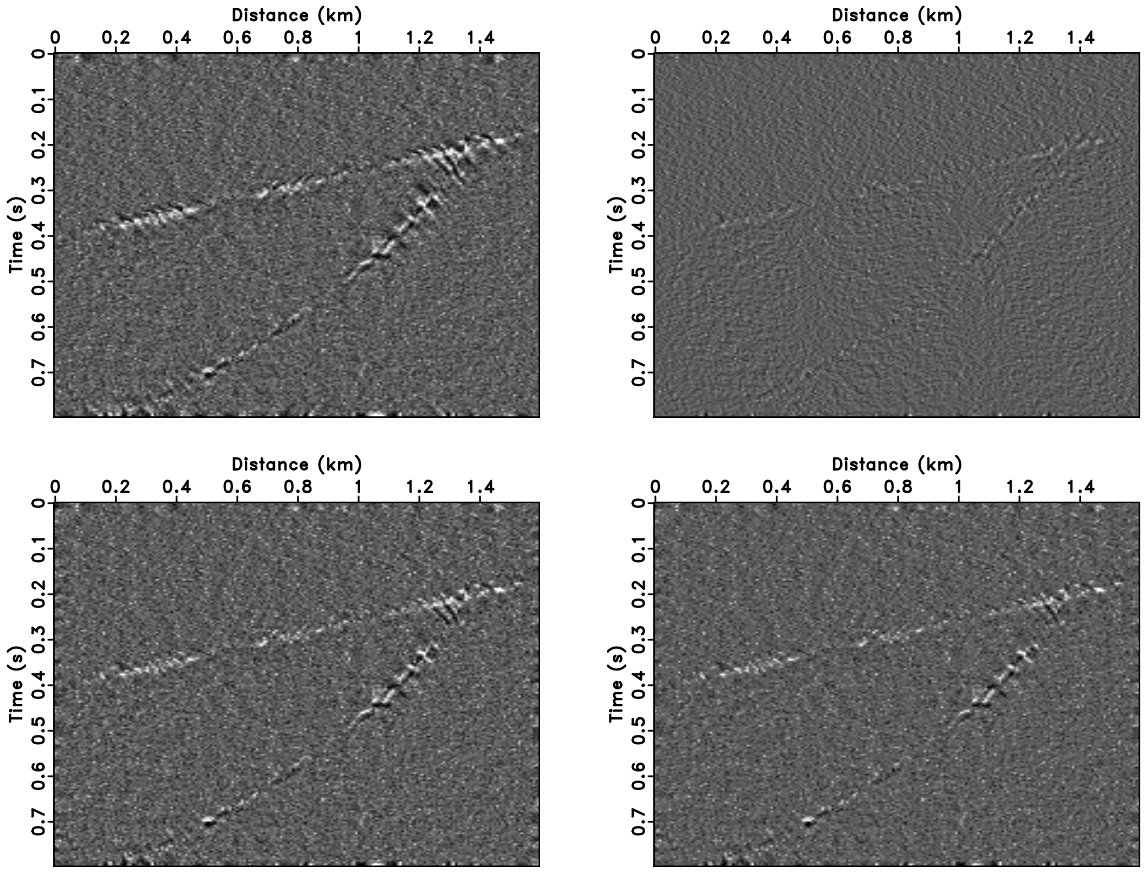


Figure 4: Difference between Figure 1b and structure-enhancing results (Figure 3). Standard mean filtering (a), similarity-mean filtering (b), standard median filtering (c), and lower-upper-middle (LUM) filtering (d). strfilter/spray mdif1,gdif1,mdif,ldif1

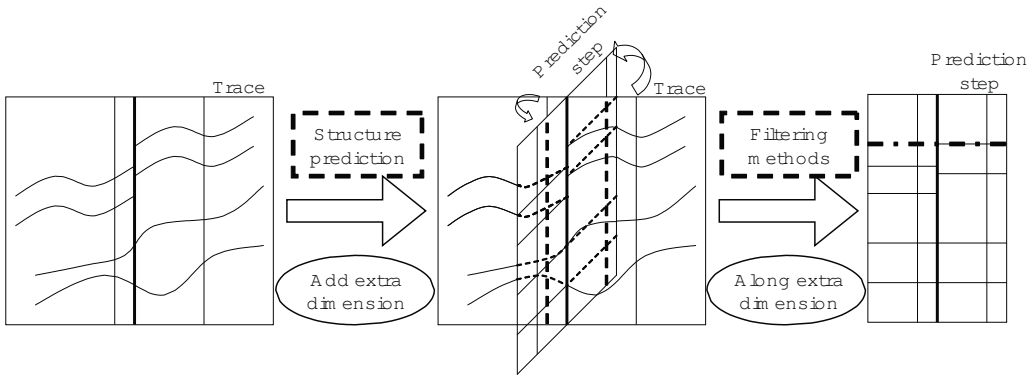


Figure 5: Schematic illustration of the proposed workflow. strfilter/XFig strat

are preserved by both methods. The LUM filter also removes some short events with crossing dips.

An extension of the method to 3-D is straightforward and follows the 3-D predictive painting construction of (8). We use a 3-D field data volume from the Gulf of Mexico to further test our methods (Figure 10). The dataset includes faults having an upside-down cone shape and some amount of random noise. Although variations in amplitude along reflectors can be stratigraphically significant, our goal is to enhance structurally continuous events. We chose a total of 24 prediction traces (2 prediction-step distances around the reference trace), which, with the addition of the original trace, generated a 4-D data volume. The corresponding inline and crossline dips were measured automatically from the image using plane-wave destruction (Figure 11 and 12). Dip sections display different tendencies in inline and crossline directions. Figure 13 and 14 show results of nonlinear structure-enhancing filtering. When processing along the prediction direction, both of our nonlinear filters work well for noise reduction and fault protection. Figure 15 and 16 show the difference sections between the original data (Figure 10) and the structure-enhancing results using two filtering methods (Figure 13 and 14). Similar to the results in the synthetic and 2-D field examples, both methods can protect useful information well.

CONCLUSION

We have introduced a new technique that combines structure prediction and signal-enhancing filters: similarity-mean filter and lower-upper-middle (LUM) filter, for attenuating random noise while enhancing structural information. The technique is generic. One can replace the proposed filters with other prediction and signal-enhancing filter methods. Experiments with synthetic and field data show that nonlinear structure-enhancing filters can both eliminate random noise and preserve fault information in seismic images. Our method makes an implicit assumption that data is made of long, coherent events and random noise. It may need a further improvement to deal with areas where either crossing dips or no coherent events are present.

ACKNOWLEDGMENTS

We thank Hongliu Zeng for useful discussions. We thank assistant editor Tijmen Jan Moser, associate editor Peter Hanssen, and two anonymous reviewers for their constructive comments and suggestions. We thank Texaco for providing the 3-D Gulf of Mexico dataset, and BGP Americas for partial financial support of this work. Publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

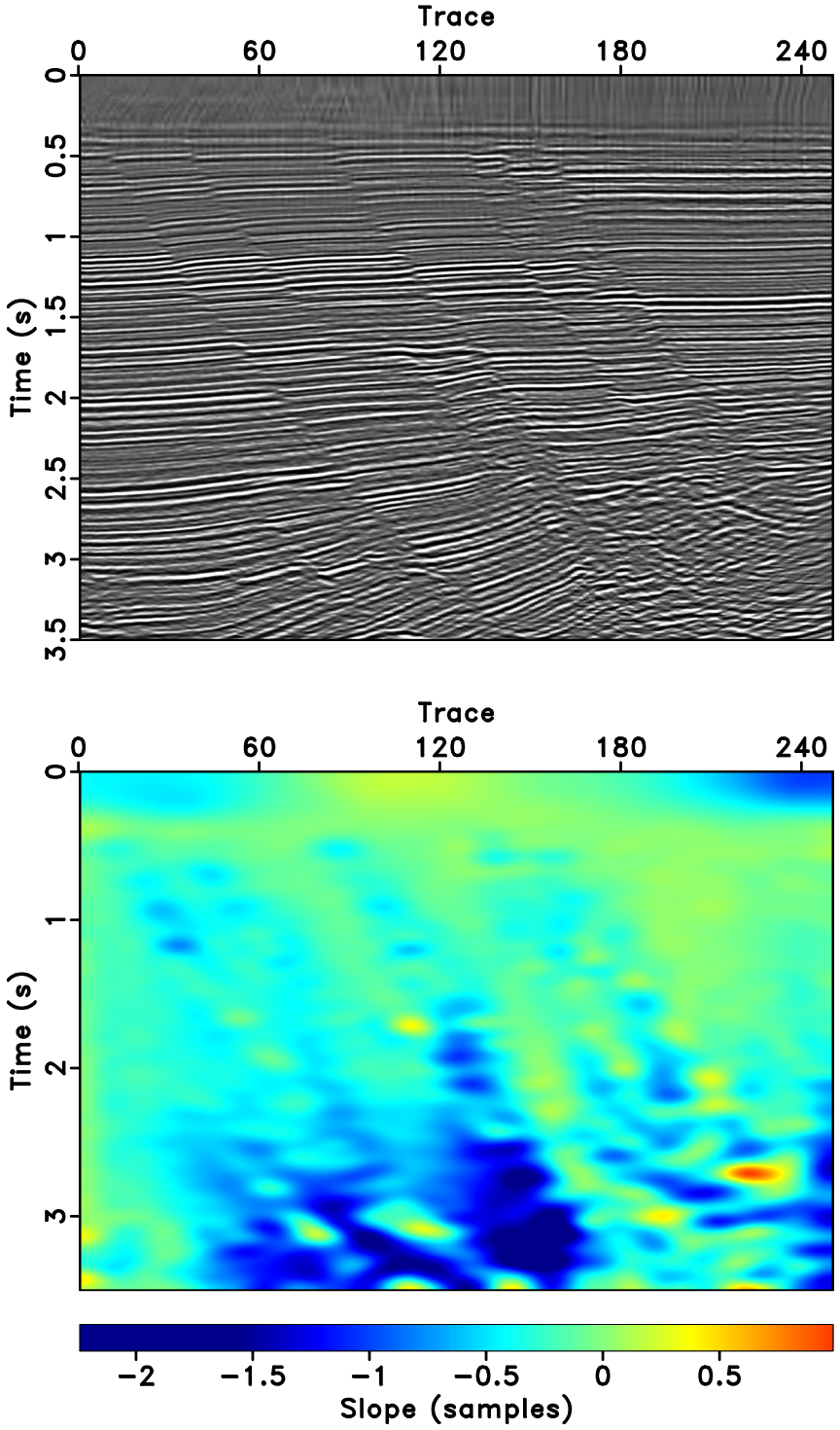


Figure 6: 2-D field data (a) and local slopes (b). strfilter/bei bei,rdip

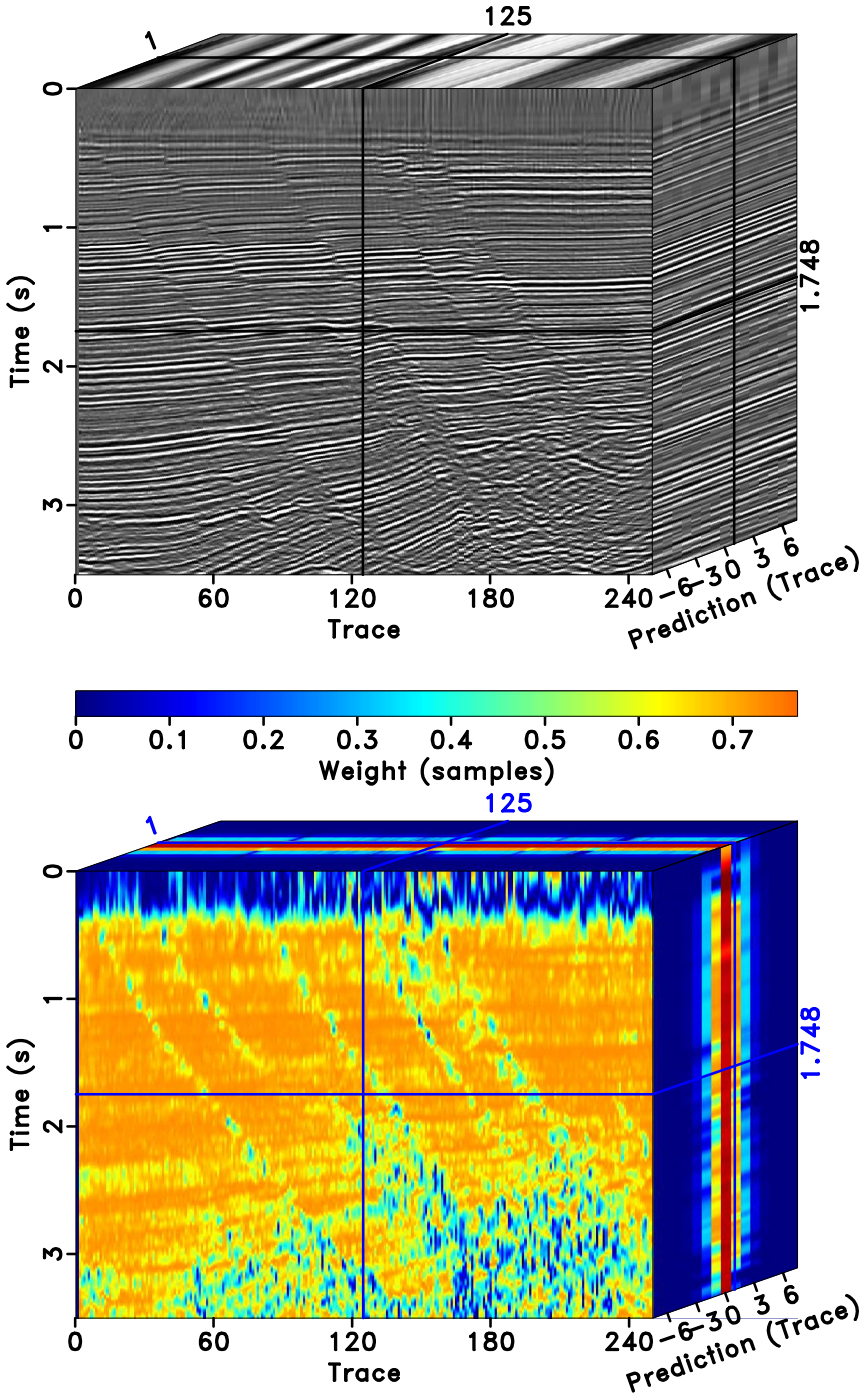


Figure 7: Predictive data volume (a) and product of Gaussian weights and similarity weights (b). strfilter/bei rcube,rweight

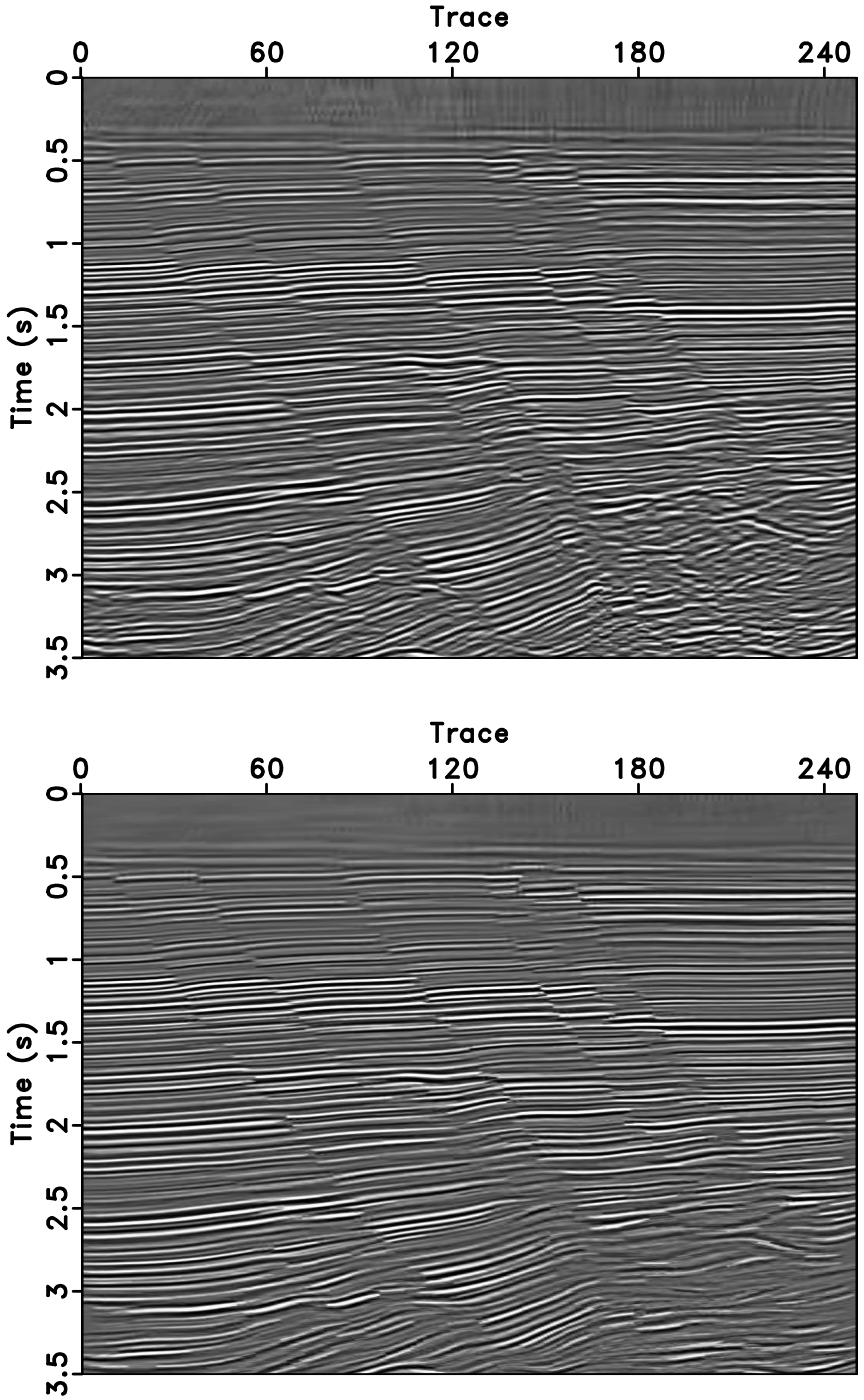


Figure 8: Structure-enhancing results for image from Figure 6a using similarity-mean filtering (a) and lower-upper-middle (LUM) filtering (b). strfilter/bei rgsimilarstack,rmedian

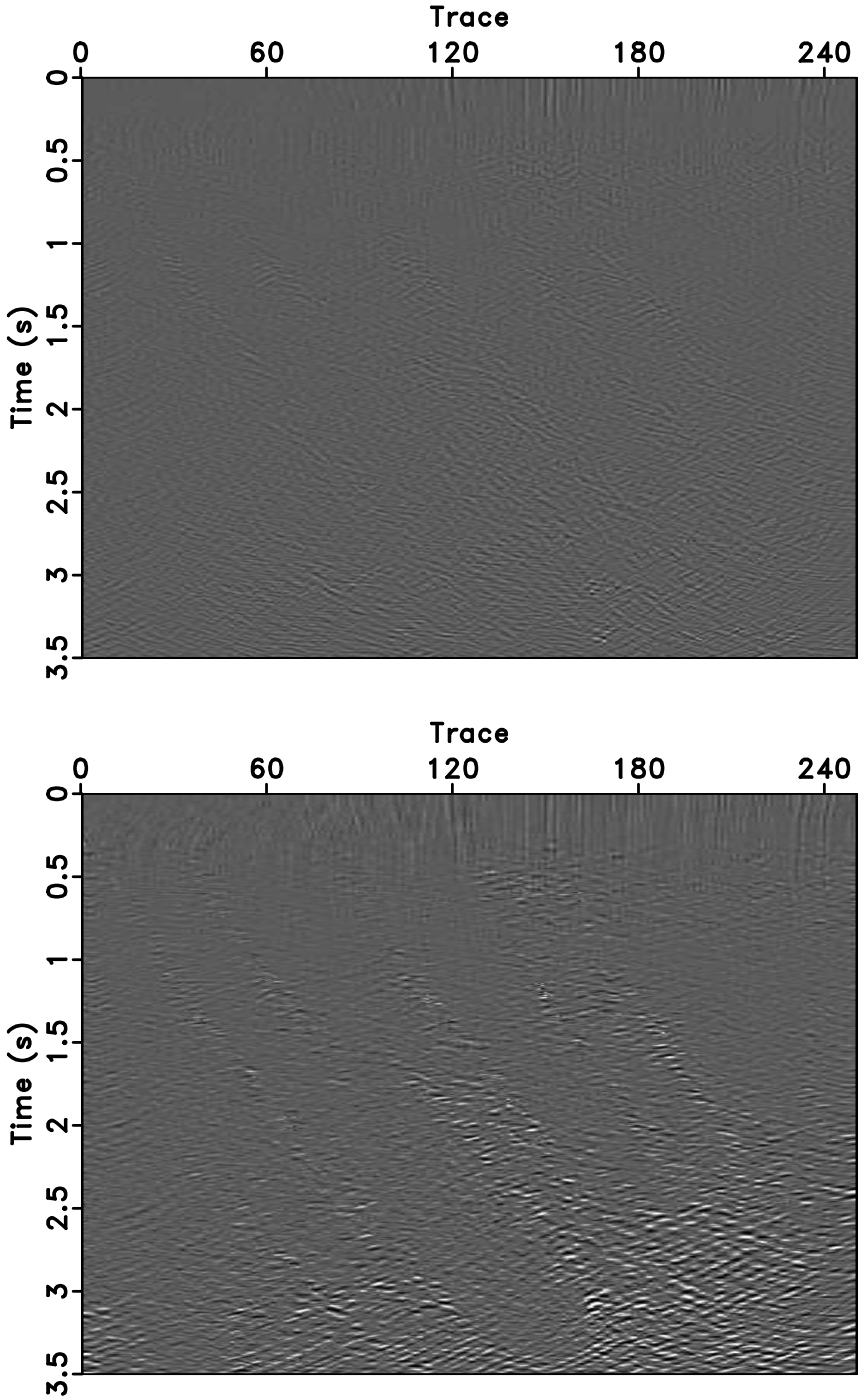


Figure 9: Difference between Figure 6a and structure-enhancing results (Figure 8). Similarity-mean filtering (a) and lower-upper-middle (LUM) filtering (b). strfilter/bei rgdif,rldif

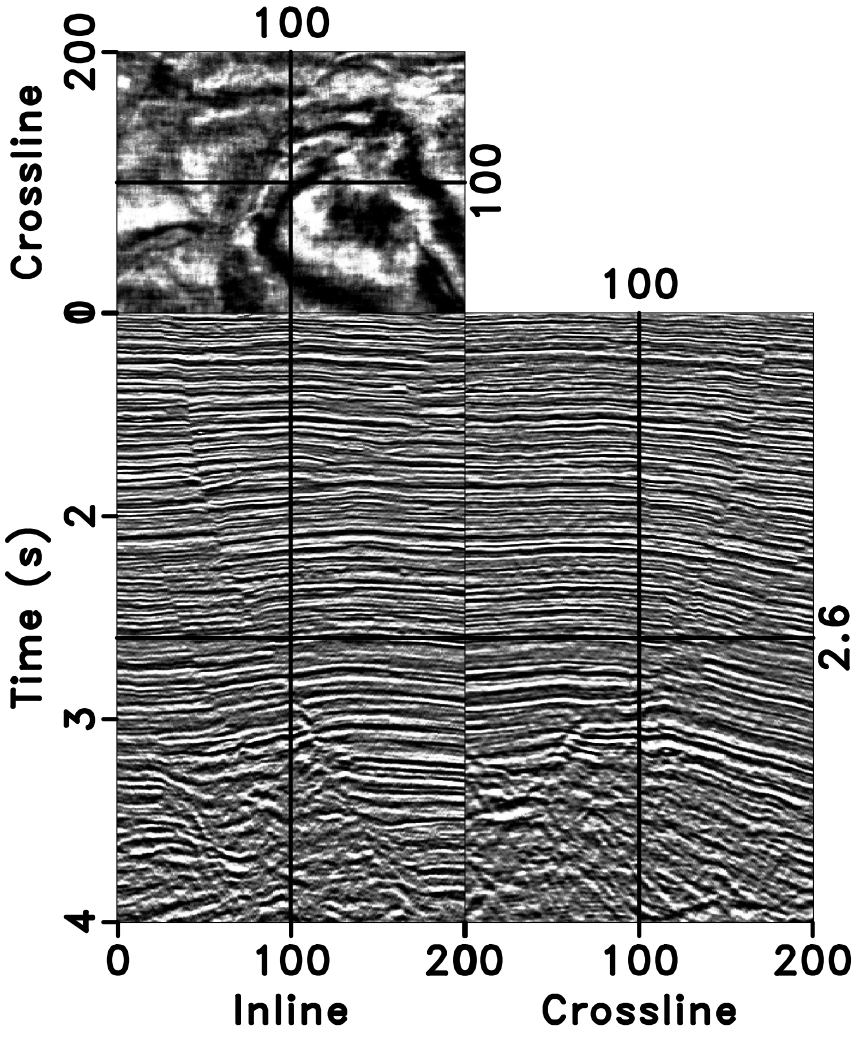


Figure 10: 3-D image from Gulf of Mexico. strfilter/hongliu sgr-ori

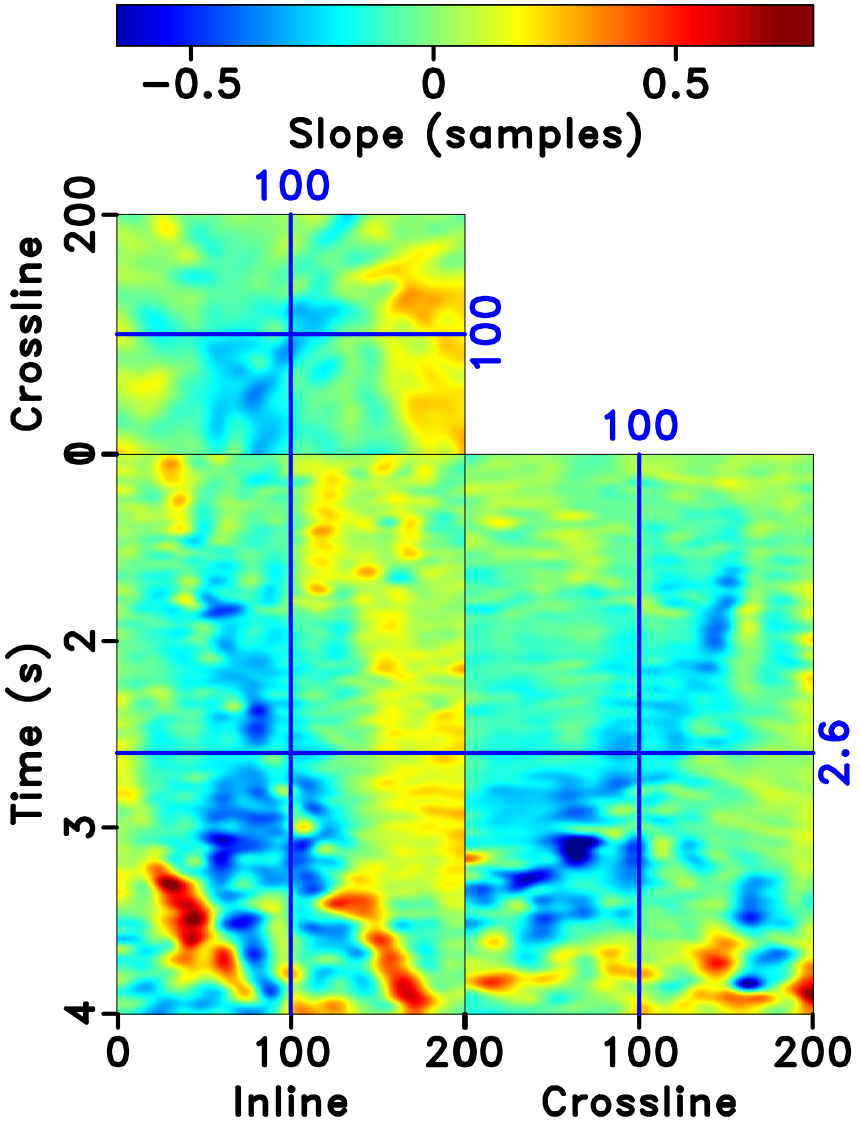


Figure 11: Local inline dip for the 3-D image from Figure 10. strfilter/hongliu sgr-dip1

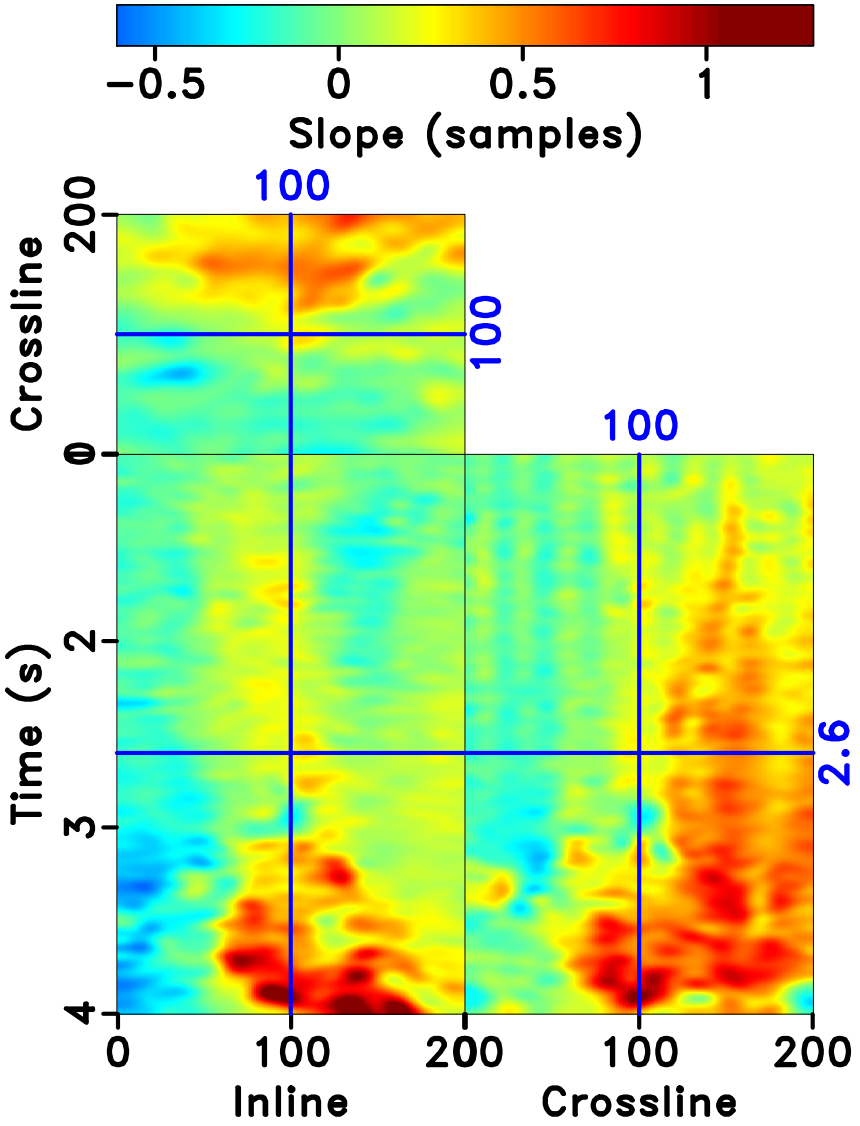


Figure 12: Local crossline dip for the 3-D image from Figure 10. strfilter/hongliu sgr-dip2

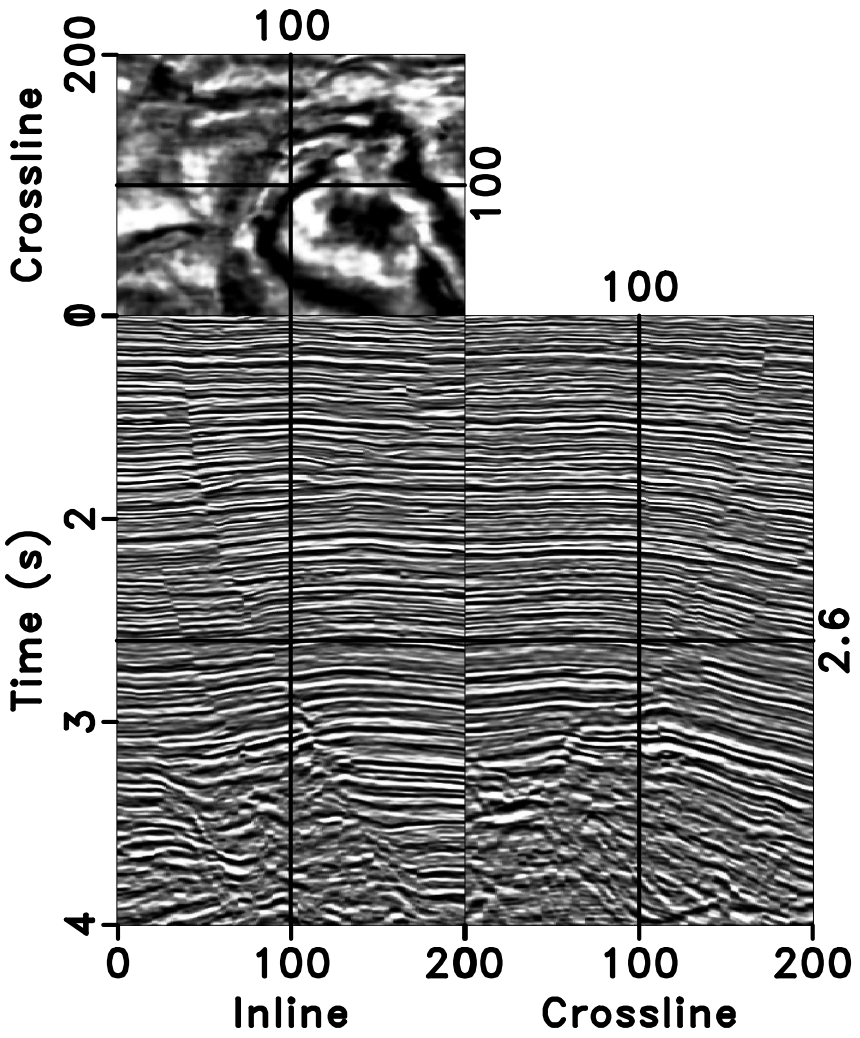


Figure 13: Structure-enhancing result using similarity-mean filtering. Compare with Figure 10. strfilter/hongliu sgr-gsstack

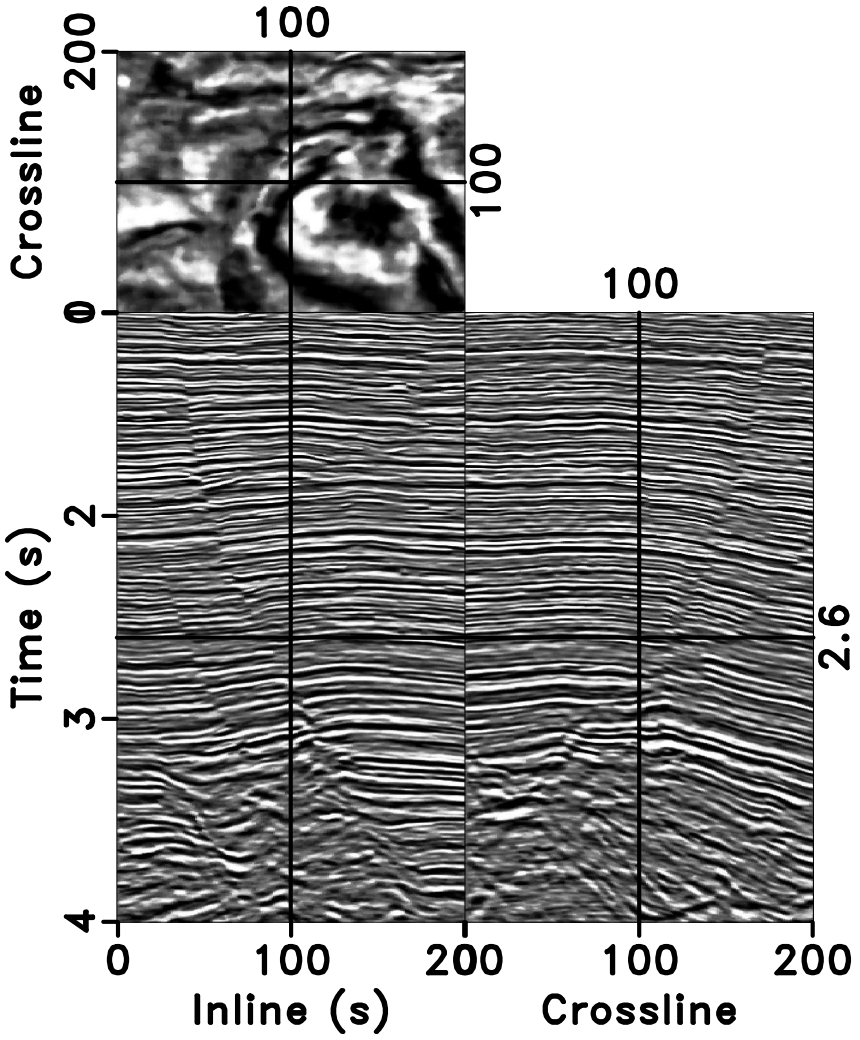


Figure 14: Structure-enhancing result using lower-upper-middle (LUM) filtering. Compare with Figure 10. strfilter/hongliu sgr-lum

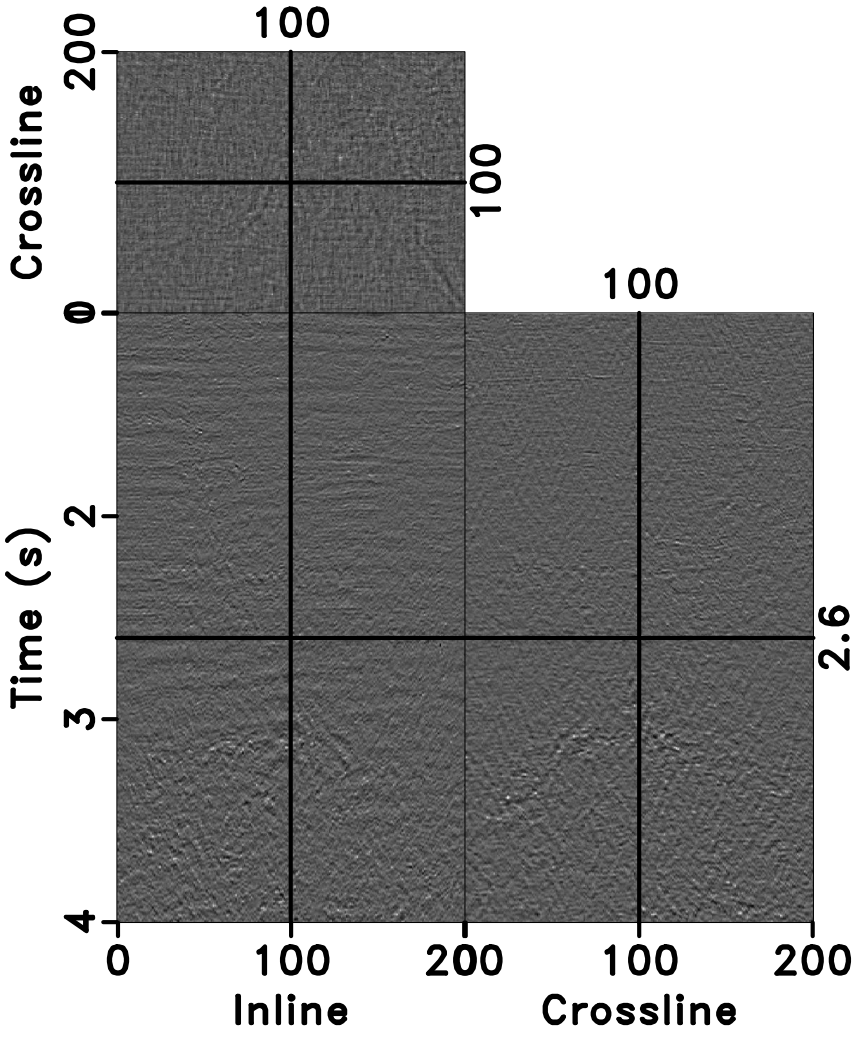


Figure 15: Difference between Figure 10 and Figure 13. strfilter/hongliu sgr-gdif

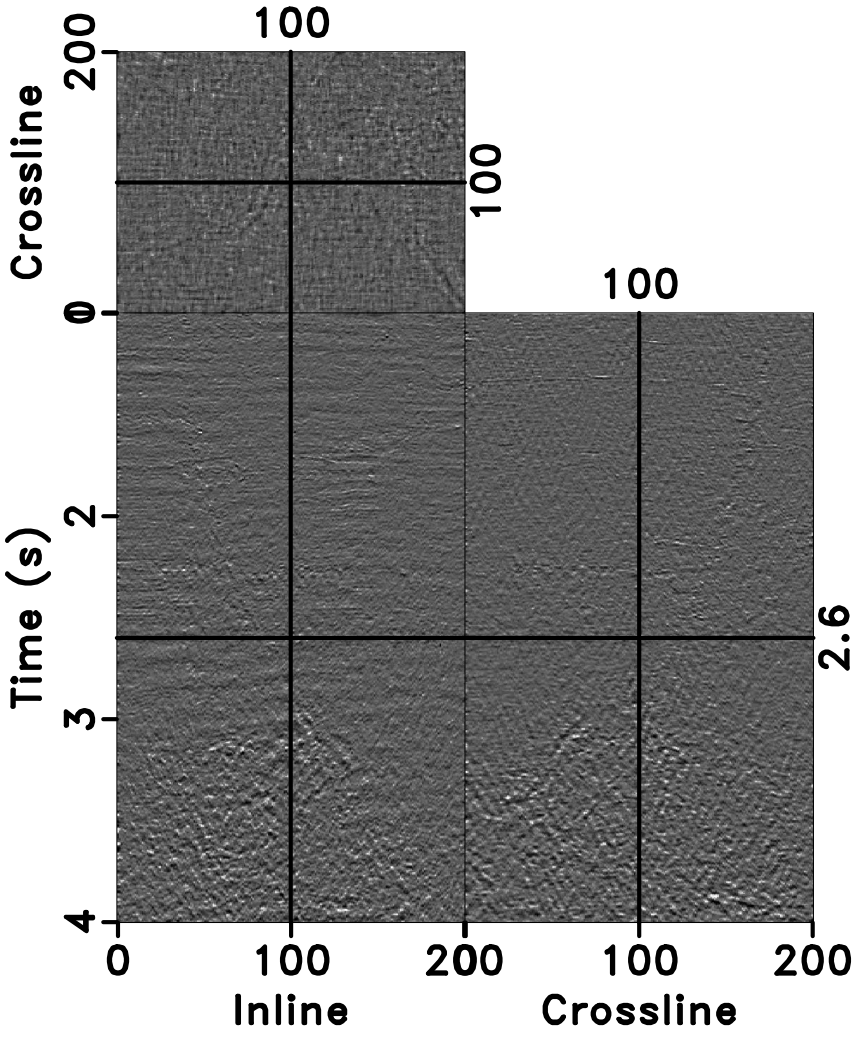


Figure 16: Difference between Figure 10 and Figure 14. strfilter/hongliu sgr-ldif

APPENDIX A: SIMILARITY-MEAN FILTER

(6) defined local similarity as follows. The global correlation coefficient between two different signals $a(t)$ and $b(t)$ is the functional

$$\gamma = \frac{\langle a(t), b(t)\rangle}{\sqrt{\langle a(t), a(t)\rangle \langle b(t), b(t)\rangle}} \ , \tag{A-1}$$

where $\langle x(t), y(t)\rangle$ denotes the dot product between two signals

$$\langle x(t), y(t)\rangle = \int x(t)y(t)dt \ . \tag{A-2}$$

In a linear algebra notation, the squared correlation coefficient γ from equation A-1 can be represented as a product of two least-squares inverses

$$\gamma^2 = \gamma_1 \gamma_2 \ , \tag{A-3}$$

$$\gamma_1 = (\mathbf{a}^T\mathbf{a})^{-1}(\mathbf{a}^T\mathbf{b}) \ , \tag{A-4}$$

$$\gamma_2 = (\mathbf{b}^T\mathbf{b})^{-1}(\mathbf{b}^T\mathbf{a}) \ , \tag{A-5}$$

where $\mathbf{a}$ is a vector notation for $a(t)$, $\mathbf{b}$ is a vector notation for $b(t)$, and $\mathbf{x}^T\mathbf{y}$ denotes the dot product operation defined in equation A-2. Let $\mathbf{A}$ be a diagonal operator composed of the elements of $\mathbf{a}$ and $\mathbf{B}$ be a diagonal operator composed of the elements of $\mathbf{b}$. Localizing equations A-4 and A-5 amounts to adding regularization to inversion. Scalars γ_1 and γ_2 turn into vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ defined, using shaping regularization (7)

$$\mathbf{c}_1 = [\lambda^2\mathbf{I} + \mathbf{S}(\mathbf{A}^T\mathbf{A} - \lambda^2\mathbf{I})]^{-1}\mathbf{S}\mathbf{A}^T\mathbf{b} \ , \tag{A-6}$$

$$\mathbf{c}_2 = [\lambda^2\mathbf{I} + \mathbf{S}(\mathbf{B}^T\mathbf{B} - \lambda^2\mathbf{I})]^{-1}\mathbf{S}\mathbf{B}^T\mathbf{a} \ , \tag{A-7}$$

where λ scaling controls the relative scaling of operators $\mathbf{A}$ and $\mathbf{B}$. Finally, the componentwise product of vectors $\mathbf{c}_1$ and $\mathbf{c}_2$ defines the local similarity measure.

For using time-dependent smooth weights in the stacking process, the local similarity amplitude can be chosen as a weight for stacking seismic data. We thus stack only those parts of the predicted data whose similarity to the reference one is comparatively large (16).

APPENDEX B: LOWER-UPPER-MIDDLE FILTER

In this appendix, we review lower-upper-middle (LUM) filters introduced by (12). Consider a window function containing a set of N samples centered about the sample $x^\star$. We assume N to be odd. This set of observations will be denoted by $\{x_1, x_2, \cdots, x_N\}$. The rank-ordered set can be written as

$$x_{(1)} \le x_{(2)} \le \cdots \le x_{(N)} \, . \tag{B-1}$$

The estimate of the center sample will be denoted $y^\star$.

Lower-upper-middle smoother

Lower-upper-middle (LUM) smoother is equivalent to center-weighted medians (14). The output of the LUM smoother with parameter k is given by

$$y^\star = med\{x_{(k)}, x^\star, x_{(N-k+1)}\}\,, \tag{B-2}$$

where $1 \le k \le (N+1)/2$.

Thus, the output of the lower-upper-middle (LUM) smoother is $x_{(k)}$ if $x^\star < x_{(k)}$. If $x^\star > x_{(N-k+1)}$, then the output of the LUM smoother is $x_{(N-k+1)}$. Otherwise the output of the LUM smoother is simply $x^\star$.

Lower-upper-middle sharpener

We can define a value centered between the lower- and upper-order statistics, $x_{(l)}$ and $x_{(N-l+1)}$. This midpoint or average, denoted t_l, is given by

$$t_l = (x_{(l)} + x_{(N-l+1)})/2\,. \tag{B-3}$$

Then, the output of the lower-upper-middle (LUM) sharpener with parameter l is given by

$$y^\star = \begin{cases} x_{(l)}, & \text{if } x_{(l)} < x^\star \le t_l \\ x_{(N-l+1)}, & \text{if } t_l < x^\star < x_{(N-l+1)} \\ x^\star & \text{otherwise} \end{cases}\,. \tag{B-4}$$

Thus, if $x_{(l)} < x^\star < x_{(N-l+1)}$, then $x^\star$ is shifted outward to $x_{(l)}$ or $x_{(N-l+1)}$ according to which is closest to $x^\star$. Otherwise the sample $x^\star$ is unmodified. By changing the parameter l, various levels of sharpening can be achieved. In the case where $l = (N+1)/2$, no sharpening occurs and the lower-upper-middle (LUM) sharpener is simply an identity filter. In the case where $l = 1$, a maximum amount of sharpening is achieved since $x^\star$ is being shifted to one of the extreme-order statistics $x_{(1)}$ or $x_{(N)}$.

Lower-upper-middle filter

To obtain an enhancing filter that is robust and can reject outliers, the philosophies of the lower-upper-middle (LUM) smoother and lower-upper-middle (LUM) sharpener must be combined. This leads us to the general lower-upper-middle (LUM) filter. A direct definition is as follows:

$$y^\star = \begin{cases} x_{(k)}, & \text{if } x^\star < x_{(k)} \\ x_{(l)}, & \text{if } x_{(l)} < x^\star < t_l \\ x_{(N-l+1)}, & \text{if } t_l < x^\star < x_{(N-l+1)} \\ x_{(N-k+1)}, & \text{if } x_{(N-l+1)} < x^\star \\ x^\star, & \text{otherwise} \end{cases}\,. \tag{B-5}$$

where t_l is the midpoint between $x_{(l)}$ and $x_{(N-l+1)}$ defined in equation B-3, and $1 \le k \le l \le (N+1)/2$.

REFERENCES

[1] AlBinHassan, N. M., Y. Luo, and M. N. Al-Faraj, 2006, 3-D edge-preserving smoothing and applications: Geophysics, **71**, P5–P11.
[2] Claerbout, J. F., 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.
[3] ———, 2008, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.
[4] Fehmers, G. C., and C. F. W. Höcker, 2003, Fast structural interpretation with structure-oriented filtering: Geophysics, **68**, 1286–1293.
[5] Fomel, S., 2002, Appliations of plane-wave destruction filters: Geophysics, **67**, 1946–1960.
[6] ———, 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.
[7] ———, 2007b, Shaping regularization in geophysical estimation problems: Geophysics, **72**, R29–R36.
[8] ———, 2008, Predictive painting of 3-D seismic volumes: 78th Annual International Meeting, SEG, Expanded Abstracts, 864–868.
[9] Fomel, S., and A. Guitton, 2006, Regularizing seismic inverse problems by model reparameterization using plane-wave construction: Geophysics, **71**, A43–A47.
[10] Gilboa, G., and S. Osher, 2008, Nonlocal operators with applications to image processing: Multiscale Model & Simulation, **7**, 1005–1028.
[11] Hale, D., 2009, Structure-oriented smoothing and semblance: CWP Report 635, Colorado School of Mines, 261–270.
[12] Hardie, R. C., and C. G. Boncelet, 1993, LUM filters: A class of Rank-Order-Based filters for smoothing and sharpening: IEEE Transactions on Signal Processing, **41**, 1061–1076.
[13] Hoeber, H., S. Brandwood, and D. Whitcombe, 2006, Structurally consistent filtering: Presented at the 68th Annual International Conference and Exhibition, EAGE, Expanded Abstracts.
[14] Justusson, B. I., 1981, Median filtering: Statistical properties: Two-Dimensional Digital Signal Processing, II: Transforms and Median Filters, Springer, 161–196.
[15] Liu, C., Y. Liu, B. Yang, D. Wang, and J. Sun, 2006, A 2D multistage median filter to reduce random seismic noise: Geophysics, **71**, V105–V110.
[16] Liu, G., S. Fomel, L. Jin, and X. Chen, 2009a, Stacking seismic data using local correlation: Geophysics, **74**, V43–V48.
[17] Liu, Y., C. Liu, and D. Wang, 2009b, A 1-D time-varying median filter for seismic random, spike-like noise elimination: Geophysics, **74**, V17–V24.
[18] Mayne, W. H., 1962, Common reflection point horizontal data stacking techniques: Geophysics, **27**, 927–938.
[19] Mi, Y., and G. F. Margrave, 2000, Median filtering in Kirchhoff migration for noisy data: 70th Annual International Meeting, SEG, Expanded Abstracts, 822–825.
[20] Rashed, M. A., 2008, Smart stacking: A new cmp stacking technique for seismic data: The Leading Edge, **27**, 462–467.
[21] Robinson, J. C., 1970, Statistically optimal stacking of seismic data: Geophysics,

35, 436–446.

[22] Tang, Y., 2007, Selective stacking in the reflection-angle and azimuth domain: 77th Annual International Meeting, SEG, Expanded Abstracts, 2320–2324.

[23] Tomasi, C., and R. Manduchi, 1998, Bilateral filtering for gray and color images: Proceedings of IEEE International Conference on Computer Vision, IEEE, 836–846.

[24] Traonmilin, Y., and P. Herrmann, 2008, Structurally consistent f-x filtering: 78th Annual International Meeting, SEG, Expanded Abstracts, 2642–2646.

[25] Tyapkin, Y., and B. Ursin, 2005, Optimum stacking of seismic records with irregular noise: Journal of Geophysics and Engineering, **2**, 177–187.

[26] Whitcombe, D. N., L. Hodgson, H. Hoeber, and Z. Yu, 2008, Frequency dependent, structurally conformable filtering: 78th Annual International Meeting, SEG, Expanded Abstracts, 2617–2621.

[27] Yilmaz, O., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: SEG.

[28] Zhang, R., and T. J. Ulrych, 2003, Multiple suppression based on the migration operator and a hyperbolic median filter: 73th Annual International Meeting, SEG, Expanded Abstracts, 1949–1952.

A parallel sweeping preconditioner for heterogeneous 3D Helmholtz equations

Jack Poulson[*], *Björn Engquist* [†], *Siwei Li* [‡], *and Lexing Ying* [§*]

ABSTRACT

A parallelization of a sweeping preconditioner for 3D Helmholtz equations without large cavities is introduced and benchmarked for several challenging velocity models. The setup and application costs of the sequential preconditioner are shown to be $O(\gamma^2 N^{4/3})$ and $O(\gamma N \log N)$, where $\gamma(\omega)$ denotes the modestly frequency-dependent number of grid points per Perfectly Matched Layer. Several computational and memory improvements are introduced relative to using black-box sparse-direct solvers for the auxiliary problems, and competitive runtimes and iteration counts are reported for high-frequency problems distributed over thousands of cores. Two open-source packages are released along with this paper: *Parallel Sweeping Preconditioner (PSP)* and the underlying distributed multifrontal solver, *Clique*.

INTRODUCTION

While definite elliptic partial differential equations can be efficiently solved by a wide variety of techniques (e.g., multigrid, ILU, or structured matrix factorizations), indefinite elliptic equations tend to be more challenging. This paper is concerned with three-dimensional heterogeneous Helmholtz equations of the form,

$$\mathcal{A}u \equiv \left[-\Delta - \frac{\omega^2}{c^2(x)}\right] u(x) = f(x), \tag{1}$$

where $c(x)$ is the spatially varying wave speed, and $u(x)e^{-i\omega t}$ is the time-harmonic response to an acoustic wave equation with forcing function $f(x)e^{-i\omega t}$. It is important to recognize that $-\Delta$ is positive-definite and that its combination with the negative-definite $-\frac{\omega^2}{c^2}$ term results in an indefinite system.

Before discussing the overall asymptotic complexity of solution techniques, it is helpful to first motivate why high frequency problems require large numbers of degrees of freedom: Given the wave speed bounds $c_{\min} \leq c(x) \leq c_{\max}$, we can define the minimum wavelength as $\lambda_{\min} = 2\pi c_{\min}/\omega$. In order to resolve oscillations in the solution using piecewise polynomial basis functions, e.g., with finite-difference and finite-element methods, it is necessary to increase the number of degrees of freedom

[*]**e-mail:** sergey.fomel@beg.utexas.edu

in each direction at least linearly with the number of wavelengths spanned by the domain. In order to combat pollution effects (4), which are closely related to phase errors in the discrete solution, one must use asymptotically more than a constant number of grid points per wavelength with standard discretization schemes. Nevertheless, it is common practice to resolve the domain to as few as ten points per wavelength. In any case, piecewise polynomial discretizations require $\Omega(\omega^d)$ degrees of freedom in d dimensions.

Until recently, doubling the frequency of Eq. (1) not only increased the size of the linear system by at least a factor of 2^d, it also doubled the number of iterations required for convergence with standard preconditioned Krylov methods (6; 17; 18). Thus, denoting the number of degrees of freedom in a three-dimensional finite-element or finite-difference discretization as $N = \Omega(\omega^3)$, every linear solve required $\Omega(\omega^4)$ work with traditional iterative techniques. Engquist and Ying recently introduced two classes of *sweeping* preconditioners for Helmholtz equations without large cavities (14; 15): Both approaches approximate a block LDL^T factorization of the Helmholtz operator in block tridiagonal form in a manner which exploits a radiation boundary condition. The first approach performs a block tridiagonal factorization algorithm in $\mathcal{H}$-matrix arithmetic (25; 22), while the second approach approximates the Schur complements of the factorization using auxiliary problems with artificial radiation boundary conditions. Though the $\mathcal{H}$-matrix sweeping preconditioner has theoretical support for two-dimensional problems (14; 28), there is not yet justification for three-dimensional problems.

This paper therefore focuses on the second approach, which relies on multifrontal factorizations (27; 34; 12; 21) of the approximate auxiliary problems in order to achieve an $O(\gamma^2 N^{4/3})$ setup cost and an $O(\gamma N \log N)$ application cost, where $\gamma(\omega)$ denotes the number of grid points used for each Perfectly Matched Layer (PML) (29). While the sweeping preconditioner is competitive with existing techniques even for a single right-hand side, its main advantage is for problems with large numbers of right-hand sides, as the preconditioner appears to converge in $O(1)$ iterations for problems without large cavities. Thus, after setting up the preconditioner, typically only $O(\gamma N \log N)$ work is required for each solution.

Moving PML sweeping preconditioner

The focus of this paper is on parallelization of the second class of sweeping preconditioners mentioned above, which makes use of auxiliary problems with artificial radiation boundary conditions in order to approximate the Schur complements that arise during block LDL^T factorization. The approach is referred to as a *moving PML* preconditioner since the introductory paper represented the artificial radiation boundary conditions using PML.

One interpretation of radiation conditions is that they allow for the analysis of a finite portion of an infinite domain, as their purpose is to enforce the condition that

waves propagate outwards and not reflect at the boundary of the truncated domain. This concept is crucial to understanding the Schur complement approximations that take place within the moving PML sweeping preconditioner which is reintroduced in this paper for the sake of completeness.

For the sake of simplicity, we will describe the preconditioner in the context of a second-order finite-difference discretization over the unit cube, with PML used to approximate a radiation boundary condition on the $x_3 = 0$ face and homogeneous Dirichlet boundary conditions applied on all other boundaries (an x_1x_3 cross-section is shown in Fig. 1). If the domain is discretized into an $n \times n \times n$ grid, then ordering the vertices in the grid such that vertex (i_1, i_2, i_3) is assigned index $i_1 + i_2 n + i_3 n^2$ results in a block tridiagonal system of equations, say

$$\begin{pmatrix} A_{0,0} & A_{1,0}^T & & & \\ A_{1,0} & A_{1,1} & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & A_{n-1,n-2}^T \\ & & & A_{n-1,n-2} & A_{n-1,n-1} \end{pmatrix} \begin{pmatrix} u_0 \\ u_1 \\ \vdots \\ u_{n-2} \\ u_{n-1} \end{pmatrix} = \begin{pmatrix} f_0 \\ f_1 \\ \vdots \\ f_{n-2} \\ f_{n-1} \end{pmatrix}, \tag{2}$$

where $A_{i,j}$ propagates sources from the i'th x_1x_2 plane into the j'th x_1x_2 plane, and the overall linear system is complex symmetric (*not* Hermitian) due to the imaginary terms introduced by the PML (15).

If we were to ignore the sparsity within each block, then the following naïve factorization and solve algorithms would be appropriate for a direct solver, where the application of S_i^{-1} implicitly makes use of the factorization of S_i.

Algorithm 0.0.1: Naïve block tridiagonal LDL^T factorization. $O(n^7) = O(N^{7/3})$ work is required.

$S_0 := A_{0,0}$
Factor S_0
for $i = 0, ..., n-2$ **do**
  $S_{i+1} := A_{i+1,i+1} - A_{i+1,i} S_i^{-1} A_{i+1,i}^T$
  Factor S_{i+1}

While the computational complexities of Algs. 0.0.1 and 0.0.2 are significantly higher than multifrontal algorithms (27; 12; 21), which have $O(N^2)$ factorization complexity and $O(N^{4/3})$ solve complexity for regular three-dimensional meshes, they are the conceptual starting points of the sweeping preconditioner.[†]

Suppose that we paused Alg. 0.0.1 after computing the i'th Schur complement, S_i, where the i'th x_1x_2 plane is in the interior of the domain (i.e., it is not in a PML

[†]In fact, they are the starting points of *both* classes of sweeping preconditioners. The $\mathcal{H}$-matrix approach essentially executes these algorithms with $\mathcal{H}$-matrix arithmetic.

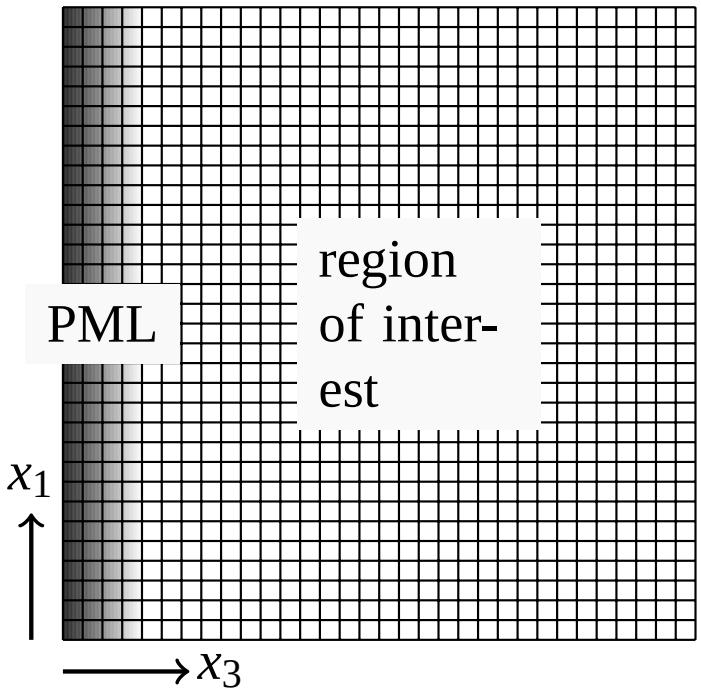


Figure 1: An $x_1 x_3$ cross section of a cube with PML on its $x_3 = 0$ face. The domain is shaded in a manner which loosely corresponds to its extension into the complex plane. psp/Tikz plane-with-single-pml

Algorithm 0.0.2: Naïve block LDL^T solve. $O(n^5) = O(N^{5/3})$ work is required.

```
// Apply L^{-1}
for i = 0, ..., n-2 do
    u_{i+1} := u_{i+1} - A_{i+1,i}(S_i^{-1} u_i)
// Apply D^{-1}
for i = 0, ..., n-1 do
    u_i := S_i^{-1} u_i
// Apply L^{-T}
for i = n-2, ..., 0 do
    u_i := u_i - S_i^{-1}(A_{i+1,i}^T u_{i+1})
```

region). Due to the ordering imposed on the degrees of freedom of the discretization, the first i Schur complements are equivalent to those that would have been produced from applying Alg. 0.0.1 to an auxiliary problem formed by placing a homogeneous Dirichlet boundary condition on the $(i+1)$'th x_1x_2 plane and ignoring all of the successive planes (see the left half of Fig. 2). While this observation does not immediately yield any computational savings, it does allow for a qualitative description of the inverse of the i'th Schur complement, S_i^{-1}: it is the restriction of the half-space Green's function of the exact auxiliary problem onto the i'th x_1x_2 plane (recall that PML can be interpreted as approximating the effect of a domain extending to infinity).

The main approximation made in the sweeping preconditioner can now be succinctly described: since S_i^{-1} is a restricted half-space Green's function which incorporates the velocity field of the first i planes, it is natural to approximate it with another restricted half-space Green's function which only takes into account the *local* velocity field, i.e., by moving the PML next to the i'th plane (see the right half of Fig. 2).

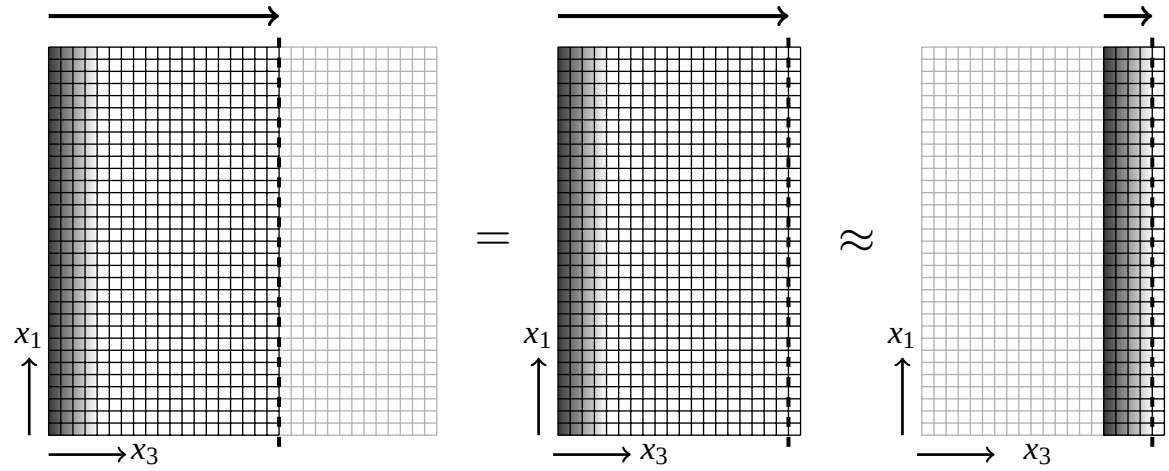


Figure 2: (Left) A depiction of the portion of the domain involved in the computation of the Schur complement of an x_1x_2 plane (marked with the dashed line) with respect to all of the planes to its left during execution of Alg. 0.0.1. (Middle) An equivalent auxiliary problem which generates the same Schur complement; the original domain is truncated just to the right of the marked plane and a homogeneous Dirichlet boundary condition is placed on the cut. (Right) A local auxiliary problem for generating an approximation to the relevant Schur complement; the radiation boundary condition of the exact auxiliary problem is moved next to the marked plane. psp/Tikz auxiliary

If we use $\gamma(\omega)$ to denote the number of grid points of PML as a function of the frequency, ω, then it is important to recognize that the depth of the approximate auxiliary problems in the x_3 direction is $\Omega(\gamma)$.[‡] If the depth of the approximate

[‡]In all of the experiments in this paper, $\gamma(\omega)$ was either 5 or 6, and the subdomain depth never exceeded 10.

auxiliary problems was $O(1)$, then combining nested dissection with the multifrontal method over the regular $n \times n \times O(1)$ mesh would only require $O(n^3)$ work and $O(n^2 \log n)$ storage (21). However, if the PML size is a slowly growing function of frequency, then applying 2D nested dissection to the *quasi-2D* domain requires $O(\gamma^3 n^3)$ work and $O(\gamma^2 n^2 \log n)$ storage, as the number of degrees of freedom in each front increases by a factor of γ and dense factorizations have cubic complexity.

Let us denote the quasi-2D discretization of the local auxiliary problem for the i'th plane as H_i, and its corresponding approximation to the Schur complement S_i as $\tilde{S}_i$. Since $\tilde{S}_i$ is essentially a dense matrix, it is advantageous to come up with an abstract scheme for applying $\tilde{S}_i^{-1}$: Assuming that H_i was ordered in a manner consistent with the $(i_1, i_2, i_3) \mapsto i_1 + i_2 n + i_3 n^2$ global ordering, the degrees of freedom corresponding to the i'th plane come last and we find that

$$H_i^{-1} = \begin{pmatrix} \star & \star \\ \star & \tilde{S}_i^{-1} \end{pmatrix}, \tag{3}$$

where the irrelevant portions of the inverse have been marked with a $\star$. Then, trivially,

$$H_i^{-1} \begin{pmatrix} 0 \\ u_i \end{pmatrix} = \begin{pmatrix} \star & \star \\ \star & \tilde{S}_i^{-1} \end{pmatrix} \begin{pmatrix} 0 \\ u_i \end{pmatrix} = \begin{pmatrix} \star \\ \tilde{S}_i^{-1} u_i \end{pmatrix}, \tag{4}$$

which implies a method for quickly computing $\tilde{S}_i^{-1} u_i$ given a factorization of H_i:

Algorithm 0.0.3: Application of $\tilde{S}_i^{-1}$ to u_i given a multifrontal factorization of H_i. $O(\gamma^2 n^2 \log n)$ work is required.

Form $\hat{u}_i$ as the extension of u_i by zero over the artificial PML
Form $\hat{v}_i := H_i^{-1} \hat{u}_i$
Extract $\tilde{S}_i^{-1} u_i$ from the relevant entries of $\hat{v}_i$

From now on, we write T_i to refer to the application of the (approximate) inverse of the Schur complement for the i'th plane.

There are two more points to discuss before defining the full serial algorithm. The first is that, rather than performing an approximate LDL^T factorization using a discretization of $\mathcal{A}$, the preconditioner is instead built from a discretization of an *artificially damped* version of the Helmholtz operator, say

$$\mathcal{J} \equiv \left[-\Delta - \frac{(\omega + i\alpha)^2}{c^2(x)} \right], \tag{5}$$

where $\alpha \approx 2\pi$ is responsible for the artificial damping. This is in contrast to shifted Laplacian preconditioners (5; 16), where α is typically $O(\omega)$ (18), and our motivation is to avoid introducing large long-range dispersion error by damping the long range interactions in the preconditioner. Just as A refers to the discretization of the original Helmholtz operator, $\mathcal{A}$, we will use J to refer to the discretization of the artificially damped operator, $\mathcal{J}$.

The second point is much easier to motivate: since the artificial PML in each approximate auxiliary problem is of depth $\gamma(\omega)$, processing a single plane at a time would require processing $O(n)$ subdomains with $O(\gamma^3 n^3)$ work each. Clearly, processing $O(\gamma)$ planes at once has the same asymptotic complexity as processing a single plane, and so combining $O(\gamma)$ planes reduces the setup cost from $O(\gamma^3 N^{4/3})$ to $O(\gamma^2 N^{4/3})$. Similarly, the memory usage becomes $O(\gamma N \log N)$ instead of $O(\gamma^2 N \log N)$.[§] If we refer to these sets of $O(\gamma)$ contiguous planes as *panels*, which we label from 0 to $m-1$, where $m = O(n/\gamma)$, and correspondingly redefine $\{u_i\}$, $\{f_i\}$, $\{S_i\}$, $\{T_i\}$, and $\{H_i\}$, we have the following algorithm for setting up an approximate block LDL^T factorization of the discrete artificially damped Helmholtz operator:

Algorithm 0.0.4: Setup phase of the sweeping preconditioner. $O(\gamma^2 N^{4/3})$ work is required.

$S_0 := J_{0,0}$
Factor S_0
for $i = 1, ..., m-1$ **do**
  Form H_i by prefixing PML to $J_{i,i}$
  Factor H_i

Once the preconditioner is set up, it can be applied using a straightforward modification of Alg. 0.0.2 which avoids redundant solves by combining the application of L^{-1} and D^{-1}:

Algorithm 0.0.5: Application of the sweeping preconditioner. $O(\gamma N \log N)$ work is required.

`// Apply` L^{-1} `and` D^{-1}
for $i = 0, ..., m-2$ **do**
  $u_i := T_i u_i$
  $u_{i+1} := u_{i+1} - J_{i+1,i} u_i$
$u_{m-1} := T_{m-1} u_{m-1}$
`// Apply` L^{-T}
for $i = m-2, ..., 0$ **do**
  $u_i := u_i - T_i (J_{i+1,i}^T u_{i+1})$

Given that the preconditioner can be set up with $O(\gamma^2 N^{4/3})$ work, and applied with $O(\gamma N \log N)$ work, it provides a near-linear scheme for solving Helmholtz equations if only $O(1)$ iterations are required for convergence. The experiments in this paper seem to indicate that, as long as the velocity model does not include a large cavity, the sweeping preconditioner indeed only requires $O(1)$ iterations.

[§]Note that increasing the number of planes per panel provides a mechanism for interpolating between the sweeping preconditioner and a full multifrontal factorization.

Though this paper is focused on the parallel solution of Helmholtz equations, which are the time-harmonic form of acoustic wave equations, Tsuji et al. have shown that the moving PML sweeping preconditioner is equally effective for time-harmonic Maxwell's equations (38; 39), and we believe that the same will hold true for time-harmonic linear elasticity. The rest of the paper will be presented in the context of the Helmholtz equation, but we emphasize that the parallelization should carry over to more general wave equations in a conceptually trivial way.

Related work

A domain decomposition variant of the sweeping preconditioner was recently introduced (37) which results in fast convergence rates, albeit at the expense of requiring PML padding on both sides of each subdomain. Recalling our previous analysis with respect to the PML size, γ, the memory usage of the preconditioner scales linearly with the PML size, while the setup cost scales quadratically. Thus, the domain decomposition approach should be expected to use twice as much memory and require four times as much work for the setup phase. On the other hand, doubling the subdomain sizes allows for more parallelism in both the setup and solve phases, and less sweeps seem to be required.

Another closely related method is the *Analytic ILU factorization* (19). Like the sweeping preconditioner, it uses local approximations of the Schur complements of the block LDL^T factorization of the Helmholtz matrix represented in block tridiagonal form. There are two crucial differences between the two methods:

- Roughly speaking, AILU can be viewed as using Absorbing Boundary Conditions (ABC's) (13) instead of PML when forming approximate subdomain auxiliary problems. While ABC's result in strictly 2D local subproblems, versus the *quasi-2D* subdomain problems which result from using PML, they are well-known to be less effective approximations of the Sommerfeld radiation condition (and thus the local Schur complement approximations are less effective). The sweeping preconditioner handles its non-trivial subdomain factorizations via a multifrontal algorithm.

- Rather than preconditioning with an approximate LDL^T factorization of the original Helmholtz operator, the sweeping preconditioner sets up an approximate factorization of a *slightly damped* Helmholtz operator in order to mitigate the dispersion error which would result from the Schur complement approximations.

These two improvements are responsible for transitioning from the $O(\omega)$ iterations required with AILU to the $O(1)$ iterations needed with the sweeping preconditioner (for problems without large cavities).

Two other iterative methods warrant mentioning: the two-grid shifted-Laplacian approach of (8) and the multilevel-ILU approach of (6). Though both require $O(\omega)$ iterations for convergence, they have very modest memory requirements. In particular, (8) demonstrates that, with a properly tuned two-grid approach, large-scale heterogeneous 3D problems can be solved with impressive timings.

There has also been a recent effort to extend the fast-direct methods presented in (43) from definite elliptic problems into the realm of low-to-moderate frequency time-harmonic wave equations (40; 41). In particular, their experiments (see Table 3 of (40)) suggest an asymptotic complexity of roughly $O(N^{1.8})$, which is a noticeable improvement over the $O(N^2)$ complexity of traditional 3D sparse-direct methods.

PARALLEL SWEEPING PRECONDITIONER

The setup and application stages of the sweeping preconditioner (Algs. 0.0.4 and 0.0.5) essentially consist of $m = O(n/\gamma)$ multifrontal factorizations and solves, respectively. The most important detail is that *the subdomain factorizations can be performed in parallel, while the subdomain solves must happen sequentially.*[¶] When we also consider that each subdomain factorization requires $O(\gamma^3 n^3)$ work, while subdomain solves only require $O(\gamma^2 n^2 \log n)$ work, we see that, relative to the subdomain factorizations, subdomain solves must extract a factor of $m = O(n/\gamma)$ more parallelism from a factor of $O(\gamma n/\log n)$ less operations. We thus have a strong hint that, unless the subdomain solves are carefully handled, they will be the limiting factor in the scalability of the sweeping preconditioner.

Parallel multifrontal algorithms

While a large number of techniques exist for parallelizing multifrontal factorizations and triangular solves, we focus on parallelizations which combine subtree-to-subteam (20) mappings of processes to the elimination tree (34) that also make use of two-dimensional distributions of the frontal matrices (35).[‖] More specifically, we make use of supernodal (3) elimination trees defined through nested dissection (see Figs. 3 and 4), which have been shown to result in highly scalable factorizations (24; 23) and moderately scalable triangular solutions (26).

Roughly speaking, the analysis in (26) shows that, if p_F processes are used in the multifrontal factorization of our quasi-2D subdomain problems, then we must have $\gamma n = \Omega(p_F^{1/2})$ in order to maintain constant efficiency as p_F is increased; similarly, if p_S processes are used in the multifrontal triangular solves for a subdomain, then we must

[¶]While it is tempting to try to expose more parallelism with techniques like cyclic reduction (which is a special case of a multifrontal algorithm), their straightforward application destroys the Schur complement properties that we exploit for our fast algorithm.

[‖]Cf. (1), which advocates for only distributing the root frontal matrix two-dimensionally and using a one-dimensional distribution for all other fronts.

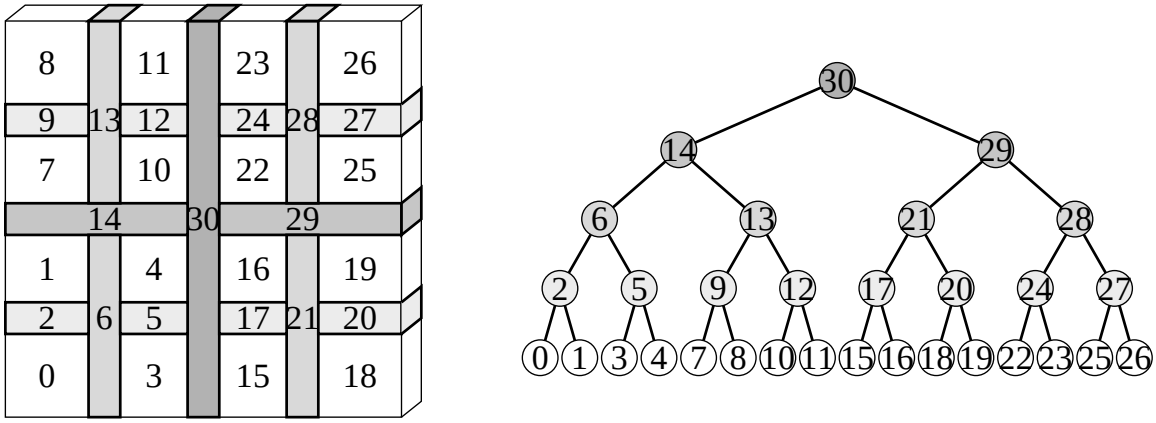


Figure 3: A separator-based supernodal elimination tree (right) over a quasi-2D subdomain (left). psp/Tikz sep-tree

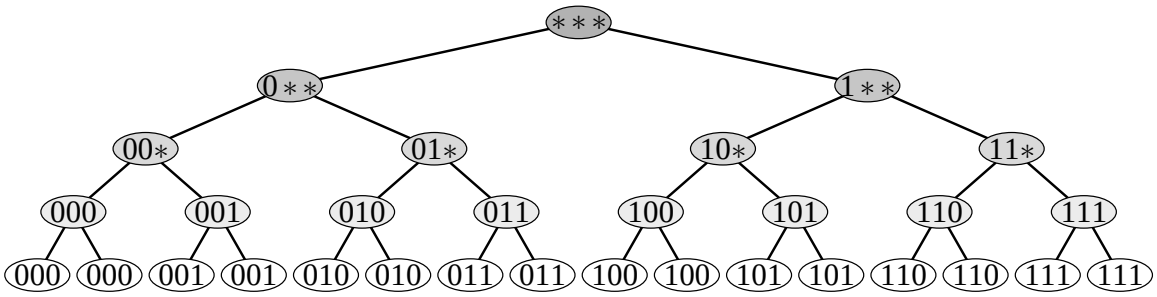


Figure 4: Overlay of the process ranks (in binary) of the owning subteams of each supernode from the elimination tree in Fig. 3 when the tree is assigned to eight processes using a subtree-to-subteam mapping; a '*' is used to denote both 0 and 1, so that '00∗' represents processes 0 and 1, '01∗' represents processes 2 and 3, and '∗ ∗ ∗' represents all eight processes. psp/Tikz subteam

have $\gamma n \approx \Omega(p_S)$ (where we use $\approx$ to denote that the equality holds within logarithmic factors). Since we can simultaneously factor the $m = O(n/\gamma)$ subdomain matrices, we denote the total number of processes as p and set $p_S = p$ and $p_F = O(p/m)$; then the subdomain factorizations only require that $n^3 = \Omega(p/\gamma)$, while the subdomain solves have the much stronger constraint that $n \approx \Omega(p/\gamma)$. This last constraint should be considered unacceptable, as we have the conflicting requirement that $n^3 \approx O(p/\gamma)$ in order to store the factorizations in memory. It is therefore advantageous to consider more scalable alternatives to standard multifrontal triangular solves, even if they require additional computation.

Selective inversion

The lackluster scalability of dense triangular solves is well known and a scheme known as *selective inversion* was introduced in (32) and implemented in (31) specifically to avoid the issue; the approach is characterized by directly inverting every distributed dense triangular matrix which would have been solved against in a normal multifrontal triangular solve. Thus, after performing selective inversion, every parallel dense triangular solve can be translated into a parallel dense triangular matrix-vector multiply.

Suppose that we have paused a multifrontal LDL^T factorization just before processing a particular front, F, which corresponds to some supernode, $\mathcal{S}$. Then all of the fronts for the descendants of $\mathcal{S}$ have already been handled, and F can be partitioned as

$$F = \begin{pmatrix} F_{TL} & \star \\ F_{BL} & F_{BR} \end{pmatrix}, \tag{6}$$

where F_{TL} holds the Schur complement of supernode $\mathcal{S}$ with respect to all of its descendants, F_{BL} represents the coupling of $\mathcal{S}$ and its descendants to $\mathcal{S}$'s ancestors, and F_{BR} holds the Schur complement updates from the descendants of $\mathcal{S}$ for the ancestors of $\mathcal{S}$. Using hats to denote input states, e.g., $\hat{F}_{TL}$ to denote the input state of F_{TL}, the first step in processing the frontal matrix F is to overwrite F_{TL} with its LDL^T factorization, which is to say that $\hat{F}_{TL}$ is overwritten with the strictly lower portion of a unit lower triangular matrix L_F and a diagonal matrix D_F such that $\hat{F}_{TL} = L_F D_F L_F^T$.

The partial factorization of F can then be completed via the following steps:

1. Solve against L_F^T to form $F_{BL} := F_{BL} L_F^{-T}$.
2. Form the temporary copy $Z_{BL} := F_{BL}$.
3. Finalize the coupling matrix as $F_{BL} := F_{BL} D_F^{-1}$.
4. Finalize the update matrix as $F_{BR} := F_{BR} - \hat{F}_{BL} \hat{F}_{TL}^{-1} \hat{F}_{BL}^T = F_{BR} - Z_{BL} F_{BL}^T$.

After adding F_{BR} onto the parent frontal matrix, only F_{TL} and F_{BL} are needed in order to perform a multifrontal solve. For instance, applying L^{-1} to some vector x proceeds up the elimination tree (starting from the leaves) in a manner similar to the factorization; after handling all of the work for the descendants of some supernode $\mathcal{S}$, only a few dense linear algebra operations with $\mathcal{S}$'s corresponding frontal matrix, say F, are required. Denoting the portion of x corresponding to the degrees of freedom in supernode $\mathcal{S}$ by $x_{\mathcal{S}}$, we must perform:

1. $x_{\mathcal{S}} := L_F^{-1} x_{\mathcal{S}}$
2. $x_U \equiv -F_{BL} x_{\mathcal{S}}$
3. Add x_U onto the entries of x corresponding to the parent supernode.

The key insight of selective inversion is that, *if we invert each distributed dense unit lower triangular matrix L_F in place, all of the parallel dense triangular solves in a multifrontal triangular solve are replaced by parallel dense matrix-vector multiplies.* It is also observed in (32) that the work required for the selective inversion is typically only a modest percentage of the work required for the multifrontal factorization, and that the overhead of the selective inversion is easily recouped if there are several right-hand sides to solve against.

Since each application of the sweeping preconditioner requires two multifrontal solves for each of the $m = O(n/\gamma)$ subdomains, which are relatively small and likely distributed over a large number of processes, selective inversion will be shown to yield a very large performance improvement. We also note that, while it is widely believed that direct inversion is numerically unstable, in (11) Druinsky and Toledo provide a review of (apparently obscure) results dating back to Wilkinson (in (42)) which show that $x := \text{inv}(A) * b$ is as accurate as a backwards stable solve if reasonable assumptions are met on the accuracy of $\text{inv}(A)$. Since $\text{inv}(A) * b$ is argued to be more accurate when the columns of $\text{inv}(A)$ have been computed with a backwards-stable solver, and both $\text{inv}(F_{TL})$ and $\text{inv}(F_{TL}^T)$ must be applied after selective inversion, it might be worthwhile to modify selective inversion to compute and store two different inverses of each F_{TL}: one by columns and one by rows.

Global vector distributions

The goal of this subsection is to determine an appropriate scheme for distributing global vectors, i.e., ones representing a function over the entire domain (as opposed to only over a panel). And while the factorizations themselves may have occurred on subteams of $O(p/m)$ processes each, in order to make use of all available processes for every subdomain solve, at this point we assume that each auxiliary problem's frontal tree has been selectively inverted and is distributed using a subtree-to-subteam

$$\begin{pmatrix} 0 & 2 & 4 & 0 & 2 & 4 & 0 \\ 1 & 3 & 5 & 1 & 3 & 5 & 1 \\ 0 & 2 & 4 & 0 & 2 & 4 & 0 \\ 1 & 3 & 5 & 1 & 3 & 5 & 1 \\ 0 & 2 & 4 & 0 & 2 & 4 & 0 \\ 1 & 3 & 5 & 1 & 3 & 5 & 1 \\ 0 & 2 & 4 & 0 & 2 & 4 & 0 \end{pmatrix} \qquad \begin{array}{ccccc} 0 & - & 2 & - & 4 \\ | & & | & & | \\ 1 & - & 3 & - & 5 \end{array}$$

Figure 5: Overlay of the owning process ranks of an 7×7 matrix distributed over a 2×3 process grid in the $[M_C, M_R]$ distribution, where M_C assigns row i to process row $i \bmod 2$, and M_R assigns column j to process column $i \bmod 3$ (left). The process grid is shown on the right.

mapping (recall Fig. 4) over the entire set of p processes.**

Since a subtree-to-subteam mapping will assign each supernode of an auxiliary problem to a team of processes, and each panel of the original 3D domain is by construction a subset of the domain of an auxiliary problem, it is straightforward to extend the supernodal subteam assignments to the full domain. We should then be able to distribute global vectors so that no communication is required for readying panel subvectors for subdomain solves (via extension by zero for interior panels, and trivially for the first panel). Since our nested dissection process does not partition in the shallow dimension of quasi-2D subdomains (see Fig. 3), extending a vector defined over a panel of the original domain onto the PML-padded auxiliary domain simply requires individually extending each supernodal subvector by zero in the x_3 direction.

Consider an element-wise two-dimensional cyclic distribution (30) of a frontal matrix F over q processes using an $r \times c$ process grid, where r and c are $O(\sqrt{q})$. Then the (i, j) entry will be stored by the process in the $(i \bmod r, j \bmod c)$ position in the process grid. Using the notation from (30), this distributed front would be denoted as $F[M_C, M_R]$, while its top-left quadrant would be referred to as $F_{TL}[M_C, M_R]$ (see Fig. 5 for a depiction of an $[M_C, M_R]$ matrix distribution).

Loosely speaking, $F[X, Y]$ means that each column of F is distributed using the scheme denoted by X, and each row is distributed using the scheme denoted by Y. For the element-wise two-dimensional distribution used for F, $[M_C, M_R]$, we have that the columns of F are distributed like Matrix Columns (M_C), and the rows of F are distributed like Matrix Rows (M_R). While this notation may seem vapid when only working with a single distributed matrix, it is useful when working with products of distributed matrices. For instance, if a '$\star$' is used to represent rows/columns being redundantly stored (i.e., not distributed), then the result of every process multiplying

**In cases where the available solve parallelism has been exhausted but the problem cannot be solved on less processes due to memory constraints, it would be preferable to solve with subdomains stored on subsets of processes.

$$\begin{pmatrix} 0 \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 \\ 2 \\ 4 \\ 1 \\ 3 \\ 5 \\ 0 \end{pmatrix}$$

Figure 6: Overlay of the owning process ranks of a vector of height 7 distributed over a 2×3 process grid in the $[V_C, \star]$ vector distribution (left) and the $[V_R, \star]$ vector distribution (right).

$$\begin{pmatrix} \{0,2,4\} \\ \{1,3,5\} \\ \{0,2,4\} \\ \{1,3,5\} \\ \{0,2,4\} \\ \{1,3,5\} \\ \{0,2,4\} \end{pmatrix}, \qquad \begin{pmatrix} \{0,1\} \\ \{2,3\} \\ \{4,5\} \\ \{0,1\} \\ \{2,3\} \\ \{4,5\} \\ \{0,1\} \end{pmatrix}$$

Figure 7: Overlay of the owning process ranks of a vector of height 7 distributed over a 2×3 process grid in the $[M_C, \star]$ distribution (left) and the $[M_R, \star]$ distribution (right).

its local submatrix of $A[X, \star]$ with its local submatrix of $B[\star, Y]$ forms a distributed matrix $C[X, Y] = (AB)[X, Y] = A[X, \star]\, B[\star, Y]$, where the last expression refers to the local multiplication process.

We can now decide on a distribution for each supernodal subvector, say $x_{\mathcal{S}}$, based on the criteria that it should be fast to form $F_{TL} x_{\mathcal{S}}$ and $F_{TL}^T x_{\mathcal{S}}$ in the same distribution as $x_{\mathcal{S}}$, given that F_{TL} is distributed as $F_{TL}[M_C, M_R]$. Suppose that we define a Column-major Vector distribution (V_C) of $x_{\mathcal{S}}$, say $x_{\mathcal{S}}[V_C, \star]$, to mean that entry i is owned by process $i \bmod q$, which corresponds to position $(i \bmod r, \lfloor i/r \rfloor \bmod c)$ in the process grid (if the grid is constructed with a column-major ordering of the process ranks; see the left side of Fig. 6). Then a call to `MPI_Allgather` (10) within each row of the process grid would allow for each process to collect all of the data necessary to form $x_{\mathcal{S}}[M_C, \star]$, as for any process row index $s \in \{0, 1, ..., r-1\}$,

$$\{i \in \mathbb{N}_0 : i \bmod r = s\} = \bigcup_{t=0}^{c-1} \{i \in \mathbb{N}_0 : i \bmod q = s + tr\}. \tag{7}$$

See the left side of Fig. 7 for an example of an $[M_C, \star]$ distribution of a 7×3 matrix.

Similarly, if $x_{\mathcal{S}}$ was distributed with a Row-major Vector distribution (V_R), as shown on the right side of Fig. 6, say $x_{\mathcal{S}}[V_R, \star]$, so that entry i is owned by the process in position $(\lfloor i/c \rfloor \bmod r, i \bmod c)$ of the process grid, then a call to `MPI_Allgather`

within each column of the process grid would provide each process with the data necessary to form $x_{\mathcal{S}}[M_R, \star]$. Under reasonable assumptions, both of these redistributions can be shown to have per-process communication volume lower bounds of $\Omega(n/\sqrt{p})$ (if F_{TL} is $n \times n$) and latency lower bounds of $\Omega(\log_2(\sqrt{p}))$ (9). We also note that translating between $x_{\mathcal{S}}[V_C, \star]$ and $x_{\mathcal{S}}[V_R, \star]$ simply requires permuting which process owns each local subvector, so the communication volume would be $O(n/p)$, while the latency cost is $O(1)$.

We have thus described efficient techniques for redistributing $x_{\mathcal{S}}[V_C, \star]$ to both the $x_{\mathcal{S}}[M_R, \star]$ and $x_{\mathcal{S}}[M_C, \star]$ distributions, which are the first steps for our parallel algorithms for forming $F_{TL}x_{\mathcal{S}}$ and $F_{TL}^T x_{\mathcal{S}}$, respectively: Denoting the distributed result of each process in process column $t \in \{0, 1, ..., c-1\}$ multiplying its local submatrix of $F_{TL}[M_C, M_R]$ by its local subvector of $x_{\mathcal{S}}[M_R, \star]$ as $z^{(t)}[M_C, \star]$, it holds that $(F_{TL}x_{\mathcal{S}})[M_C, \star] = \sum_{t=0}^{c-1} z^{(t)}[M_C, \star]$. Since Eq. (7) also implies that each process's local data from a $[V_C, \star]$ distribution is a subset of its local data from a $[M_C, \star]$ distribution, a simultaneous summation and scattering of $\{z^{(t)}[M_C, \star]\}_{t=0}^{c-1}$ within process rows, perhaps via `MPI_Reduce_scatter` or `MPI_Reduce_scatter_block`, yields the desired result, $(F_{TL}x_{\mathcal{S}})[V_C, \star]$. An analogous process with $(F_{TL}[M_C, M_R])^T = F_{TL}^T[M_R, M_C]$ and $x_{\mathcal{S}}[M_C, \star]$ yields $(F_{TL}^T x_{\mathcal{S}})[V_R, \star]$, which can then be cheaply permuted to form $(F_{TL}^T x_{\mathcal{S}})[V_C, \star]$. Both calls to `MPI_Reduce_scatter_block` can be shown to have the same communication lower bounds as the previously discussed `MPI_Allgather` calls (9).

As discussed at the beginning of this section, defining the distribution of each supernodal subvector specifies a distribution for a global vector, say $[\mathcal{G}, \star]$. While the $[V_C, \star]$ distribution shown in the left half of Fig. 6 simply assigns entry i of a supernodal subvector $x_{\mathcal{S}}$, distributed over q processes, to process $i \bmod q$, we can instead choose an alignment parameter, σ, where $0 \leq \sigma < q$, and assign entry i to process $(i + \sigma) \bmod q$. If we simply set $\sigma = 0$ for every supernode, as the discussion at the beginning of this subsection implied, then at most $O(\gamma n)$ processes will store data for the root separator supernodes of a global vector, as each root separator only has $O(\gamma n)$ degrees of freedom by construction. However, there are $m = O(n/\gamma)$ root separators, so we can easily allow for up to $O(n^2)$ processes to share the storage of a global vector if the alignments are carefully chosen. It is important to notice that the top-left quadrants of the frontal matrices for the root separators can each be distributed over $O(\gamma^2 n^2)$ processes, so $O(\gamma^2 n^2)$ processes can actively participate in the corresponding triangular matrix-vector multiplications.

Parallel preconditioned GMRES(k)

Since, by hypothesis, only $O(1)$ iterations of GMRES(k) will be required for convergence with the sweeping preconditioner, a cursory inspection of Algorithm 0.0.5 reveal that most of the work in a preconditioned iterative method, such as GMRES(k), will be performed in the multifrontal solves during the preconditioner application, but a modest portion will also be spent in sparse matrix-vector multiplication with the

discrete Helmholtz operator, A, and the off-diagonal blocks of the discrete artificially damped Helmholtz operator, J. It is thus important to parallelize the sparse matrix-vector multiplies, but it is not crucial that the scheme be optimal, and so we simply distribute A and J in the same manner as vectors, i.e., with the $[\mathcal{G}, \star]$ distribution derived from the auxiliary problems' frontal distributions.

For performance reasons, it is beneficial to solve as many right-hand sides simultaneously as possible: both the communication latency and the costs of loading the local data from frontal and sparse matrices from main memory can be amortized over all of the right-hand sides. Another idea is to extend the so-called `trsm` algorithm for triangular solves with many right-hand sides (i.e., more right-hand sides than processes), which is well-known in the field of dense linear algebra (30), into the realm of sparse-direct solvers via the dense frontal triangular solves. This approach was not pursued in this paper due to the modest storage space available on Lonestar and is left for future work. Another performance improvement might come from exploiting block variants of GMRES (36), which can potentially lower the number of required iterations.

Clique

In order to implement the previously discussed techniques for scalable multifrontal factorizations and solves (via selective inversion), an open-source distributed multifrontal solver named Clique was built on top of Elemental (30), a library for distributed-memory dense linear algebra. In addition to being designed to support the techniques we discussed above: selective inversion, subtree-to-submesh mappings, and two-dimensional frontal matrix distributions, it was also written with a strong emphasis on *memory scalability*. This is because the sweeping preconditioner requires large numbers of factorizations of relatively small sparse matrices, and so it is crucial that the per-process memory usage for each subdomain factorization decreases inversely with the total number of processes.

We note that Clique was designed specifically to provide a memory-scalable multifrontal implementation for our parallel sweeping preconditioner, and so there is not yet support for pivoting. We plan to add a pivoted LU factorization in the near future.

Parallel Sweeping Preconditioner (PSP)

Given the discussion in Section , it is most convenient to describe our prototype implementation of a parallel sweeping preconditioner based upon its deviations from our proposed approach. The primary difference is that there is not yet support for simultaneously factoring the subdomain auxiliary problems and then redistributing each frontal tree to the entire set of processes. This will certainly lead to large improvements in the scalability of the setup phase, but it is left for future work.

EXPERIMENTAL RESULTS

Our experiments were performed on the Texas Advanced Computing Center (TACC) machine, Lonestar, which is comprised of 1,888 compute nodes, each equipped with two hex-core 3.33 GHz processors and 24 GB of memory, which are connected with QDR InfiniBand using a fat-tree topology. Our tests launched eight MPI processes per node in order to provide each MPI process with 3 GB of memory.

Our experiments took place over five different 3D velocity models:

- A uniform background with a high-contrast barrier. The domain is the unit cube and the wave speed is 1 except in $[0,1] \times [0.25, 0.3] \times [0, 0.75]$, where it is 10^{10}.
- A wedge problem over the unit cube, where the wave speed is set to 2 if $Z \leq 0.4 + 0.1x_2$, 1.5 if otherwise $Z \leq 0.8 - 0.2x_2$, and 3 in all other cases.
- A two-layer model defined over the unit cube, where $c = 4$ if $x_2 < 0.5$, and $c = 1$ otherwise.
- A waveguide over the unit cube: $c(\mathbf{x}) = 1.25(1 - 0.4e^{-32(|x_1-0.5|^2+|x_2-0.5|^2)})$.
- The SEG/EAGE Overthrust model (2), see Fig. 9.

In all of the following experiments, the shortest wavelength of each model is resolved with roughly ten grid points and the performance of the preconditioner is tested using the following four forcing functions:

- a single *shot* centered at $\mathbf{x}_0$, $f_0(\mathbf{x}) = ne^{-10n\|\mathbf{x}-\mathbf{x}_0\|^2}$,
- three shots, $f_1(\mathbf{x}) = \sum_{i=0}^{2} ne^{-10n\|\mathbf{x}-\mathbf{x}_i\|^2}$,
- a Gaussian beam centered at $\mathbf{x}_2$ in direction $\mathbf{d}$, $f_2(\mathbf{x}) = e^{i\omega\mathbf{x}\cdot\mathbf{d}}e^{-4\omega\|\mathbf{x}-\mathbf{x}_2\|^2}$, and
- a plane wave in direction $\mathbf{d}$, $f_3(\mathbf{x}) = e^{i\omega\mathbf{x}\cdot\mathbf{d}}$,

where $\mathbf{x}_0 = (0.5, 0.5, 0.1)$, $\mathbf{x}_1 = (0.25, 0.25, 0.1)$, $\mathbf{x}_2 = (0.75, 0.75, 0.5)$, and $\mathbf{d} = (1, 1, -1)/\sqrt{3}$. Note that, in the case of the Overthrust model, these source locations should be interpreted proportionally (e.g., an x_3 value of 0.1 means a depth which is 10% of the total depth of the model). Due to the oscillatory nature of the solution, we simply choose the zero vector as our initial guess in all experiments.

The first experiment was meant to test the convergence rate of the sweeping preconditioner over domains spanning 50 wavelengths in each direction (resolved to ten points per wavelength) and each test made use of 256 nodes of Lonestar. During the course of the tests, it was noticed that a significant amount of care must be taken when setting the amplitude of the derivative of the PML takeoff function (i.e., the

	velocity model			
	barrier	wedge	two-layers	waveguide
Hz	50	75	50	37.5
PML amplitude	3.0	4.0	4.0	2.0
iterations	28	49	48	52

Table 1: The number of iterations required for convergence for four model problems (with four forcing functions per model). The grid sizes were 500^3 and roughly 50 wavelengths were spanned in each direction. Five grid points were used for all PML discretizations, four planes were processed per panel, and the damping factors were all set to 7.

"C" variable in Eq. (2.1) from (15)). For the sake of brevity, hereafter we refer to this variable as the *PML amplitude*. A modest search was performed in order to find a near-optimal value for the PML amplitude for each problem. These values are reported in Table 1 along with the number of iterations required for the relative residuals for all four forcing functions to reduce to less than 10^{-5}.

It was also observed that, at least for the waveguide problem, the convergence rate would be significantly improved if 6 grid points of PML were used instead of 5. In particular, the 52 iterations reported in Table 1 reduce to 27 if the PML size is increased by one. Interestingly, the same number of iterations are required for convergence of the waveguide problem at half the frequency (and half the resolution) with five grid points of PML. Thus, it appears that the optimal PML size is a slowly growing function of the frequency.[††] We also note that, though it was not intentional, each of the these first four velocity models is invariant in one or more direction, and so it would be straightforward to sweep in a direction such that only $O(1)$ panel factorizations would need to be performed, effectively reducing the complexity of the setup phase to $O(\gamma^2 N)$.

The last experiment was meant to simultaneously test the convergence rates and scalability of the new sweeping preconditioner using the Overthrust velocity model (see Fig. 9) at various frequencies, and the results are reported in Table 2. It is important to notice that this is not a typical weak scaling test, as the number of grid points per process was fixed, *not* the local memory usage or computational load, which both grow superlinearly with respect to the total number of degrees of freedom. Nevertheless, including the setup phase, it took less than three minutes to solve the 3.175 Hz problem with four right-hand sides with 128 cores, and just under seven and a half minutes to solve the corresponding 8 Hz problem using 2048 cores. Also, while it is by no means the main message of this paper, the timings without making use of selective inversion are also reported in parentheses, as the technique is not widely implemented.[‡‡]

[††] A similar observation is also made in (37).

[‡‡] Other than Clique, the only other implementation appears to be in DSCPACK (31).

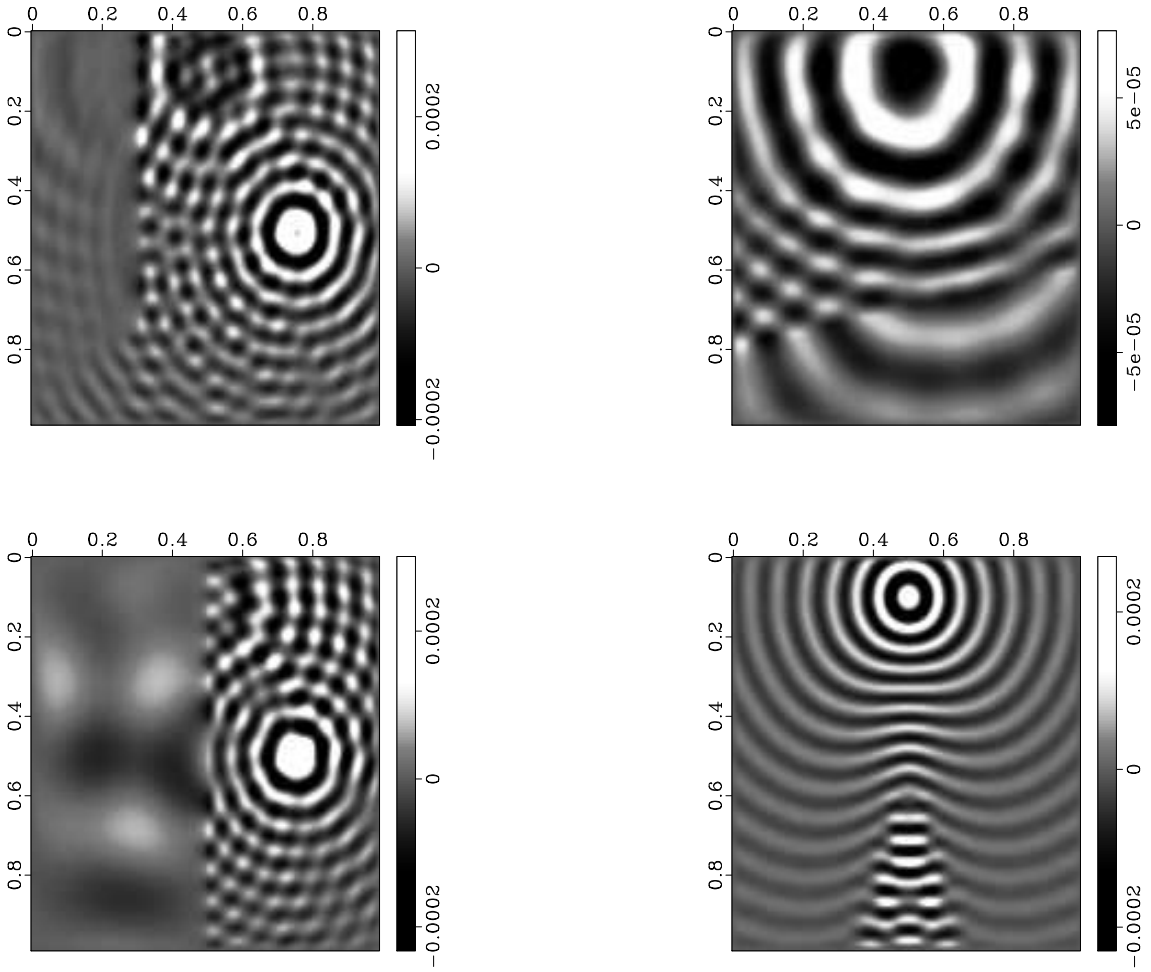


Figure 8: A single x_2x_3 plane from each of the four analytical velocity models over a 500^3 grid with the smallest wavelength resolved with ten grid points. (Top-left) the three-shot solution for the barrier model near $x_1 = 0.7$, (bottom-left) the three-shot solution for the two-layer model near $x_1 = 0.7$, (top-right) the single-shot solution for the wedge model near $x_1 = 0.7$, and (bottom-right) the single-shot solution for the waveguide model near $x_1 = 0.55$. psp/models solution-models

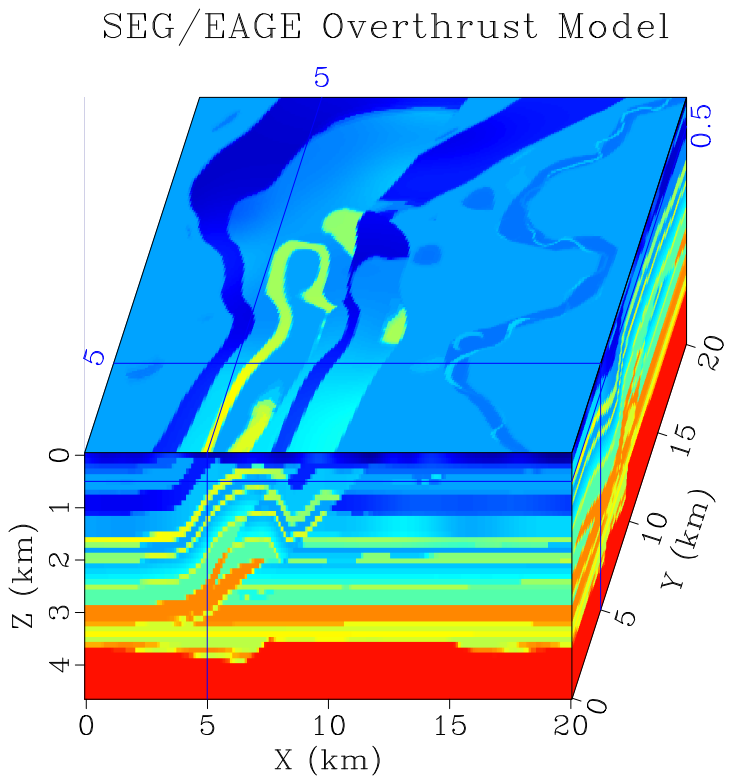


Figure 9: Three cross-sections of the SEG/EAGE Overthrust velocity model, which represents an artificial 20 km × 20 km × 4.65 km domain, containing an *overthrust* fault, using samples every 25 m. The result is an 801 × 801 × 187 grid of wave speeds varying discontinuously between 2.179 km/sec (blue) and 6.000 km/sec (red). psp/overthrust overthrust

	number of processes				
	128	256	512	1024	2048
Hz	3.175	4	5.04	6.35	8
grid	$319^2 \times 75$	$401^2 \times 94$	$505^2 \times 118$	$635^2 \times 145$	$801^2 \times 187$
setup (sec)	48.40 (46.23)	66.33 (63.41)	92.95 (86.90)	130.4 (120.6)	193.0 (175.2)
apply (sec/rhs)	0.468 (1.07)	0.550 (1.28)	0.645 (2.40)	0.700 (3.33)	0.880 (6.13)
3 digits (iter's)	42	44	42	39	40
4 digits (iter's)	54	57	57	58	58
5 digits (iter's)	63	69	70	68	72

Table 2: Convergence rates and timings on TACC's Lonestar for the SEG/EAGE Overthrust model, where timings in parentheses do not make use of selective inversion. All cases used a double-precision second-order stencil with five grid spacings for all PML (with an amplitude of 7.5), and a damping parameter of 2.25π. The preconditioner was configured with four planes per panel and eight processes per node, and the 'apply' timings are with respect to a single application of the preconditioner to four right-hand sides.

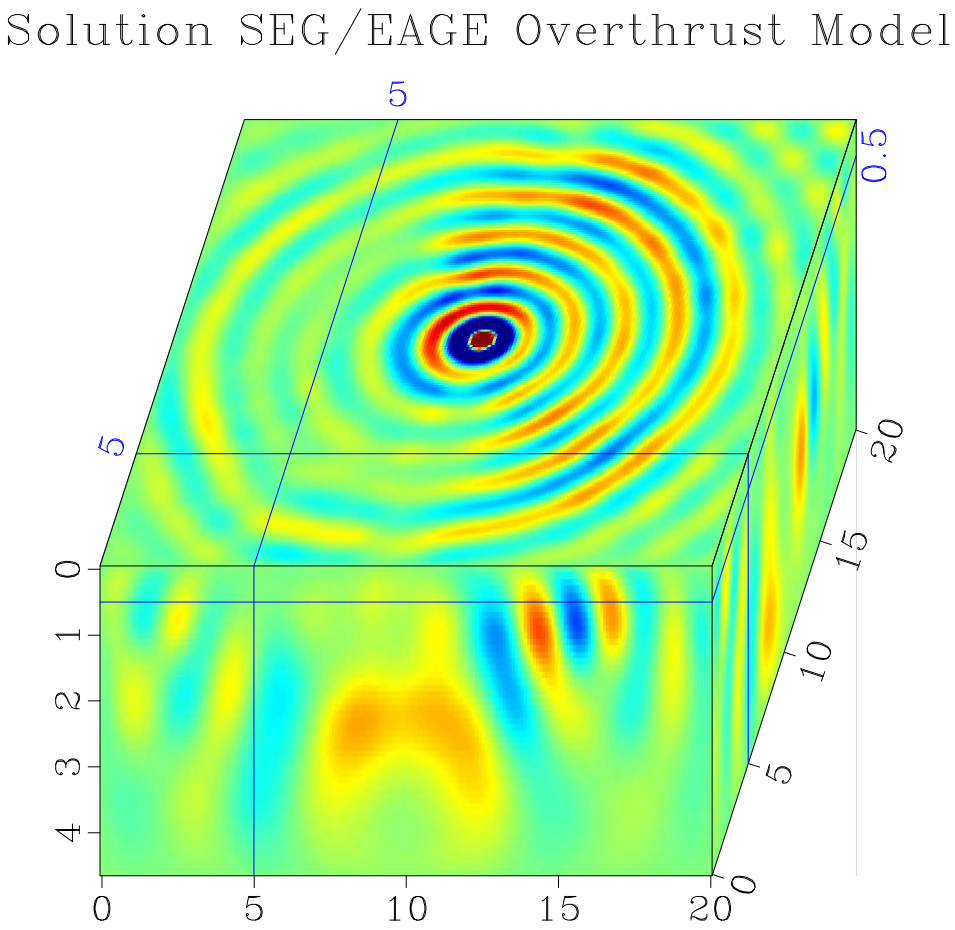


Figure 10: Three planes from a solution with the Overthrust model with a single localized shot at the center of the x_1x_2 plane at a depth of 456 m: (top) a x_2x_3 plane near $x_1 = 14$ km, (middle) an x_1x_3 plane near $x_2 = 14$ km, and (bottom) an x_1x_2 plane near $x_3 = 0.9$ km. psp/overthrust solution-overthrust

CONCLUSIONS

A parallelization of the *moving PML* sweeping preconditioner has been presented which has allowed us to efficiently solve 3D Helmholtz equations in parallel with essentially $O(1)$ iterations, with the only observed frequency-dependence arising from a moderate growth in the PML size with increasing frequency. This size of the PML, $\gamma(\omega)$ was explained to result in a linear growth in the memory requirements of the preconditioner and a quadratic growth in the setup cost. Results were then presented for a variety of models, one of which had a velocity field which varied by ten orders of magnitude, and convergence was shown to be essentially independent of frequency for the challenging Overthrust model.

Also, despite the requirement that each panel must be solved against one at a time when applying the preconditioner, a custom approach was introduced and implemented which eliminates most of the communication associated with performing traditional black-box sparse-direct factorizations and solves over each subdomain. These implementations are now released as part of the open-source packages Clique and Parallel Sweeping Preconditioner (PSP). There are at least five important directions for future work:

- developing a heuristic for tailoring the PML profile to the velocity field,
- extending the preconditioner to more general discretizations and time-harmonic wave equations,
- finding a fast preconditioner for problems with large cavities (perhaps through more general local auxiliary problems),
- testing the performance improvements resulting from simultaneously factoring the subdomain problems and then redistributing the frontal trees, as well as a `trsm` approach to solving many right-hand sides, and
- carefully studying the spectrum of the preconditioned operator for various classes of velocity models.

AVAILABILITY

The distributed dense linear algebra library, Elemental, is available under the New BSD License at `http://code.google.com/p/elemental`. The distributed multifrontal solver, Clique, is available under the GPLv3 at `http://github.com/poulson/Clique`. The Parallel Sweeping Preconditioner (PSP) code is available under the GPLv3 at `http://github.com/poulson/PSP`.

ACKNOWLEDGMENTS

The authors acknowledge TACC for usage of their computing resources and thank Bill Barth for suggesting the `tacc_affinity` option and Tommy Minyard for helping with several large runs. We also thank Anshul Gupta for WSMP details and the referees for their insightful comments.

REFERENCES

[1] P. R. Amestoy, I. S. Duff, J. Koster, and J.-Y. L'Excellent, *A fully asynchronous multifrontal solver using distributed dynamic scheduling*, SIAM J. Matrix Anal., 23 (2001), no. 1, pp. 15–41.

[2] F. Aminzadeh, J. Brac, and T. Kunz, *3-D Salt and Overthrust Models*, SEG/EAGE 3-D Modeling Series 1, Society of Exploration Geophysicists, Tulsa, OK, 1997.

[3] C. Ashcraft, R. Grimes, J. Lewis, B. Peyton, and H. Simon, *Progress in sparse matrix methods for large sparse linear systems on vector supercomputers*, Internat. J. Supercomputer Applications, 1 (1987), pp. 10–30.

[4] I. M. Babuška and S. A. Sauter, *Is the pollution effect of the FEM avoidable for the Helmholtz equation considering high wave numbers?*, SIAM Review, 42 (2000), no. 3, pp. 451–484.

[5] A. Bayliss, C. Goldstein, and E. Turkel, *An iterative method for the Helmholtz equation*, J. Comput. Phys., 49 (1983), pp. 443–457.

[6] M. Bollhoefer, M. Grote, and O. Schenk, *Algebraic multilevel preconditioner for the Helmholtz equation in heterogeneous media*, SIAM J. Sci. Comp., 31 (2009), pp. 3781–3805.

[7] M. Bollhoefer and Y. Saad, *Multilevel preconditioners constructed from inverse-based ILUs*, SIAM J. Sci. Comp., 27 (2006), pp. 1627–1650.

[8] H. Calandra, S. Gratton, X. Pinel, and X. Vasseur, *An improved two-grid preconditioner for the solution of three-dimensional Helmholtz problems in heterogeneous media*, CERFACS, Toulouse, France, Technical Report, 2012, TR/PA/12/2. Available at: `http://www.cerfacs.fr/algor/reports/2012/TR_PA_12_2.pdf`.

[9] E. Chan, M. Heimlich, A. Purkayastha, and R. A. van de Geijn, *Collective communication: theory, practice, and experience*, Concurrency and Computation: Practice and Experience, 19 (2007), no. 13, pp. 1749–1783.

[10] J. J. Dongarra and D. W. Walker, *MPI: A standard message passing interface*, Supercomputer, 12 (1996), no. 1, pp. 56–68.

[11] A. Druinsky and S. Toledo, *How accurate is* inv$(A) * b?$, CoRR, abs/1201.6035 (2012), 9 pages. Available at: `http://arxiv.org/abs/1201.6035`.

[12] I. S. Duff and J. K. Reid, *The multifrontal solution of indefinite sparse symmetric linear equations*, ACM Trans. Math. Software, 9 (1983), pp. 302–325.

[13] B. Engquist and A. Majda, *Absorbing Boundary Conditions for the numerical simulation of waves*, Mathematics of Computation, 31 (1977), pp. 629–651.

[14] B. Engquist and L. Ying, *Sweeping preconditioner for the Helmholtz equation: hierarchical matrix representation*, Commun. on Pure and App. Math., 64 (2011), pp. 697–735.

[15] B. Engquist and L. Ying, *Sweeping preconditioner for the Helmholtz equation: moving perfectly matched layers*, SIAM J. Multiscale Modeling and Simulation, 9 (2011), pp. 686–710.

[16] Y. Erlangga, C. Vuik, and C. Oosterlee, *On a class of preconditioners for solving the Helmholtz equation*, Applied Numer. Math., 50 (2004), pp. 409–425.

[17] Y. Erlangga, *Advances in iterative methods and preconditioners for the Helmholtz equation*, Archives Comput. Methods in Engin., 15 (2008), pp. 37-66.

[18] O. G. Ernst and M. J. Gander, *Why it is difficult to solve Helmholtz problems with classical iterative methods*, in Numerical Analysis of Multiscale Problems, I. Graham, T. Hou, O. Lakkis, and R. Scheichl, eds., Springer-Verlag, New York, NY, 2011, pp. 325–363.

[19] M. J. Gander and F. Nataf, *AILU for Helmholtz problems: a new preconditioner based on the analytic parabolic factorization*, in J. Comput. Acoustics, 9 (2001), pp. 1499–1506.

[20] A. George, J. W. H. Liu, and E. Ng, *Communication reduction in parallel sparse cholesky factorization on a hypercube*, in Hypercube Multiprocessors, M. T. Heath, ed., SIAM, Philadelphia, PA, 1987, pp. 576–586.

[21] A. George, *Nested dissection of a regular finite element mesh*, SIAM J. Numer. Anal., 10 (1973), pp. 345–363.

[22] L. Grasedyck and W. Hackbusch, *Construction and arithmetics of $\mathcal{H}$-matrices*, Computing, 70 (2003), no. 4, pp. 295–334.

[23] A. Gupta, S. Koric, and T. George, *Sparse matrix factorization on massively parallel computers*, Proc. of Conf. on High Perf. Comp. Networking, Storage, and Anal. (SC '09), ACM, New York, NY, 2009. Article 1, 12 pages. Available at: `http://doi.acm.org/10.1145/1654059.1654061`.

[24] A. Gupta, G. Karypis, and V. Kumar, *A highly scalable parallel algorithm for sparse matrix factorization*, IEEE Trans. Parallel and Dist. Systems, 8 (1997), no. 5, pp. 502–520.

[25] W. Hackbusch, *A sparse matrix arithmetic based on $\mathcal{H}$-matrices. I. Introduction to $\mathcal{H}$-matrices*, Computing, 62 (1999), no. 2, pp. 89–108.

[26] M. Joshi, A. Gupta, G. Karypis, and V. Kumar, *A high-performance two dimensional scalable parallel algorithm for solving sparse triangular systems*, Proc. of Internat. Conf. on High Perf. Comp. (HiPC), (1997), pp. 137–143.

[27] J. W. H. Liu, *The multifrontal method for sparse matrix solution: theory and practice*, SIAM Rev., 34 (1992), no. 1, pp. 82–109.

[28] P.-G. Martinsson and V. Rokhlin, *A fast direct solver for scattering problems involving elongated structures*, J. Comput. Phys., 221 (2007), no. 1, pp. 288–302.

[29] S. G. Johnson, *Notes on perfectly matched layers (PMLs)*, Massachusetts Institute of Technology, Technical Report, 2007; updated 2010. Available at: `http://www-math.mit.edu/~stevenj/18.369/pml.pdf`.

[30] J. Poulson, B. Marker, R. A. van de Geijn, J. R. Hammond, and N. A. Romero, *Elemental: a new framework for distributed memory dense matrix computations*, ACM Trans. Math. Software, Note: to appear.

[31] P. Raghavan, *Domain-Separator Codes for the parallel solution of sparse linear systems*, The Pennsylvania State University, University Park, PA, Technical Report, 2002, CSE-02-004.

[32] P. Raghavan, *Efficient parallel sparse triangular solution using selective inversion*, Parallel Processing Letters, 8 (1998), no. 1, pp. 29–40.

[33] Y. Saad and M. H. Schultz, GMRES*: A generalized minimal residual method for solving nonsymmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.

[34] R. Schreiber, *A new implementation of sparse Gaussian elimination*, ACM Trans. Math. Software, 8 (1982), no. 3, pp. 256–276.

[35] R. Schreiber, *Scalability of sparse direct solvers*, in Graph Theory and Sparse Matrix Computation, A. George, J. R. Gilbert, and J. W. H. Liu, eds., Springer-Verlag, New York, NY, 1993, pp. 191–209.

[36] V. Simoncini and E. Gallopoulos, *Convergence properties of block GMRES and matrix polynomials*, Linear Algebra and its Applications, 247 (1996), pp. 97–119.

[37] C. Stolk, *A rapidly converging domain decomposition method for the Helmholtz equation*, CoRR, abs/1208.3956 (2012), 14 pages. Available at: `http://arxiv.org/abs/1208.3956`.

[38] P. Tsuji, B. Engquist, and L. Ying, *A sweeping preconditioner for time-harmonic Maxwell's equations with finite elements*, J. Comp. Phys, Note: to appear.

[39] P. Tsuji and L. Ying, *A sweeping preconditioner for Yee's finite difference approximation of time-harmonic Maxwell's Equations*, J. Frontiers of Math. China, Note: to appear.

[40] S. Wang, M. V. de Hoop, and J. Xia, *On 3D modeling of seismic wave propagation via a structured parallel multifrontal direct Helmholtz solver*, Geophysical Prospecting, 59 (2011), pp. 857–873.

[41] S. Wang, X. S. Li, J. Xia, Y. Situ, and M. V. de Hoop, *Efficient scalable algorithms for hierarchically semiseparable matrices*, Submitted to SIAM J. Sci. Comput., 2011. Available at: `http://www.math.purdue.edu/~xiaj/work/parhss.pdf`.

[42] J. H. Wilkinson, *Rounding Errors in Algebraic Processes*. Prentice-Hall, Englewood Cliffs, N. J., 1963.

[43] J. Xia, S. Chandrasekaran, M. Gu, and X. Li, *Superfast multifrontal method for large structured linear systems of equations*, SIAM J. Matrix Anal. Appl., 31 (2009), no. 3, pp. 1382–1411.

Modeling of pseudo-acoustic P-waves in orthorhombic media with a lowrank approximation

*Xiaolei Song and Tariq Alkhalifah**

ABSTRACT

Wavefield extrapolation in pseudo-acoustic orthorhombic anisotropic media suffers from wave-mode coupling and stability limitations in the parameter range. We use the dispersion relation for scalar wave propagation in pseudo-acoustic orthorhombic media to model acoustic wavefields. The wavenumber-domain application of the Laplacian operator allows us to propagate the P-waves exclusively, without imposing any conditions on the parameter range of stability. It also allows us to avoid dispersion artifacts commonly associated with evaluating the Laplacian operator in space domain using practical finite difference stencils. To handle the corresponding space-wavenumber mixed-domain operator, we apply the lowrank approximation approach. Considering the number of parameters necessary to describe orthorhombic anisotropy, the lowrank approach yields space-wavenumber decomposition of the extrapolator operator that is dependent on space location regardless of the parameters, a feature necessary for orthorhombic anisotropy. Numerical experiments show that the proposed wavefield extrapolator is accurate and practically free of dispersion. Furthermore, there is no coupling of qSv and qP waves, because we use the analytical dispersion solution corresponding to the P-wave.

INTRODUCTION

Nowadays, a growing number of seismic modeling and imaging techniques are being developed to handle wave propagation in transversely isotropic media (TI). Such anisotropic phenomena are typical in sedimentary rocks, in which the process of lithification usually produces identifiable layering. In anisotropic media, the velocity is no longer described by a single parameter. Equations for anisotropic wave propagation are more complicated, even for simple cases. Although exact expressions for phase velocities in VTI media involve four independent parameters, It has been observed that only three parameters influence wave propagation and are of interest to surface seismic processing (4). Different approximations have been developed to simplify anisotropic equations, such as the weak-anisotropy approximation (34), elliptical approximations (27; 9), acoustic approximations (1; 2) and anelliptic approximations (10; 29; 19). Tectonic movement of the crust may rotate the rocks and tilt the

*e-mail: sergey.fomel@beg.utexas.edu

natural vertical orientation of the symmetry axis (VTI), causing a tilted TI (TTI) anisotropy. In addition, tectonic stresses may also fracture rocks, inducing another TI with a symmetry axis parallel to the stress direction and usually normal to the sedimentation-based TI. The combination of these effects can be represented by an orthorhombic model with three mutually orthogonal planes of mirror symmetry; the P-waves in each symmetry plane can be described kinematically as an independent TI model. Realization of the importance of orthorhombic models mainly comes from observation of azimuthal velocity variations in flat-layered rocks, which may indicate valuable fracture properties of reservoirs (37).

Wavefields in anisotropic media are well described by the anisotropic elastic-wave equation. However, in practice, we often have little information about shear waves and prefer to deal with scalar wavefields, especially for conventional imaging of subsurface structure. (2) derived an acoustic scalar wave equation for VTI media by careful reparametrization followed by setting the shear velocity along the symmetry axis to zero, which provided accurate kinematics for the conventional elastic wavefield. Later on, (3) followed the same approach and introduced an acoustic wave equation of the sixth order in axis-aligned orthorhombic media. (23) presented coupled systems of partial differential equations for pseudo-acoustic wave propagation in orthorhombic media by extending their previous work in TI media (22). (41) extended self-adjoint differential operators in TTI media (12; 43) to orthorhombic media.

Pseudo-acoustic P-wave modeling with coupled equations may have shear-wave numerical artifacts in the simulated wavefield (26; 40; 13). Those artifacts as well as sharp changes in symmetry axis tilting may introduce severe numerical dispersion and instability in modeling. (38) proposed to reduce the instability by making $\epsilon = \delta$ in regions with rapid tilt changes. (18) suggested that including a finite shear-wave velocity enhances the stability when solving the coupled equations. These methods can alleviate the instability problem; however, they may alter the wave propagation kinematics or still leave shear-wave components in the P-wave simulation. Shear-wave artifacts can be removed from the P-wavefield in the phase-shift extrapolation method because the P- and S-wave solutions lie in a different part of the wavenumber spectrum (5). A number of spectral methods are proposed to provide solutions which can completely avoid the shear-wave artifacts (16; 30; 21; 31; 8; 39; 24).

In this paper, we adopt a dispersion relation for orthorhombic anisotropic media (3) and introduce a mixed-domain acoustic wave extrapolator for time marching in orthorhombic media. We use the lowrank approximation (20; 21) to handle this mixed-domain operator. We demonstrate by numerical examples that our method is kinematically accurate. Furthermore, there is no coupling of quasi-P and quasi-SV waves in the wavefield and no constraints on Thomsen's parameters required for stability.

THEORY

Acoustic Wave Extrapolation

The acoustic wave equation is widely used in forward seismic modeling and reverse-time migration (6; 17):

$$\frac{\partial^2 p}{\partial t^2} = v(\mathbf{x})^2 \, \nabla^2 p \; , \tag{1}$$

where $p(\mathbf{x}, t)$ is the seismic pressure wavefield and $v(\mathbf{x})$ is the wave propagation velocity.

Assuming the model is homogeneous $v(\mathbf{x}) \equiv v_0$, after a Fourier transform in space, we get the following explicit expression in the wavenumber domain:

$$\frac{d^2 \hat{p}}{dt^2} = -v_0^2 |\mathbf{k}|^2 \hat{p} \; , \tag{2}$$

where

$$\hat{p}(\mathbf{k}, t) = \int_{-\infty}^{+\infty} p(\mathbf{x}, t) e^{i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x} \; . \tag{3}$$

Equation 2 has the following analytical solution:

$$\hat{p}(\mathbf{k}, t + \Delta t) = e^{\pm i |\mathbf{k}| v_0 \Delta t} \hat{p}(\mathbf{k}, t) \; , \tag{4}$$

which leads to the well-known second-order time-marching scheme (15; 33):

$$\begin{aligned} p(\mathbf{x}, t + \Delta t) + p(\mathbf{x}, t - \Delta t) = \\ 2 \int_{-\infty}^{+\infty} \hat{p}(\mathbf{k}, t) \cos(|\mathbf{k}| v_0 \Delta t) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{k} \; . \end{aligned} \tag{5}$$

Equation 5 provides a very accurate and efficient solution in the case of a constant-velocity medium with the aid of FFTs. When the seismic wave velocity varies in the medium, equation 5 turns into a reasonable approximation by replacing v_0 with $v(\mathbf{x})$, and taking small time steps, Δt. However, FFTs can no longer be applied directly to evaluate the inverse Fourier transform, because a space-wavenumber mixed-domain term appears in the integral operation:

$$W(\mathbf{x}, \mathbf{k}) = \cos(|\mathbf{k}| v(\mathbf{x}) \Delta t). \tag{6}$$

As a result, a straightforward numerical implementation of wave extrapolation in a variable velocity medium with mixed-domain matrix 6 will increase the cost from $O(N_x log N_x)$ to $O(N_x^2)$, the original cost for the homogeneous case, in which N_x is the total size of the three-dimensional space grid. A number of numerical methods (16; 28; 42; 11; 20; 21; 30; 32; 31) have been proposed to overcome this mixed-domain

problem.

In the case of orthorhombic acoustic modeling, we derive a new phase operator $\phi(\mathbf{x}, \mathbf{k})$ to replace $|\mathbf{k}|v(\mathbf{x})$ of the isotropic model. We describe the details in the next section.

Dispersion Relation for Orthorhombic Anisotropic Media

In transversely isotropic (TI) media, the model is fully characterized by five elastic parameters and density. In orthorhombic media, nine elastic parameters and density are needed to describe the elastic model. The stiffness tensor c_{ijkl} for an orthorhombic model can be represented, using the compressed two-index Voigt notation, as follows:

$$\mathbf{C} = \begin{bmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{22} & c_{23} & 0 & 0 & 0 \\ c_{13} & c_{23} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{bmatrix}. \tag{7}$$

Instead of strictly adhering to the orthorhombic media used by (35; 36), (3) slightly changed the notations and used the following nine parameters determined from the above stiffness tensor:

$$\begin{aligned} v_v &= \sqrt{\frac{c_{33}}{\rho}} \\ v_{s1} &= \sqrt{\frac{c_{55}}{\rho}} \\ v_{s2} &= \sqrt{\frac{c_{44}}{\rho}} \\ v_{s3} &= \sqrt{\frac{c_{66}}{\rho}} \\ v_1 &= \sqrt{\frac{c_{13}(c_{13}+2c_{55})+c_{33}c_{55}}{\rho(c_{33}-c_{55})}} \\ v_2 &= \sqrt{\frac{c_{23}(c_{23}+2c_{44})+c_{33}c_{44}}{\rho(c_{33}-c_{44})}} \\ \eta_1 &= \frac{c_{11}(c_{33}-c_{55})}{2c_{13}(c_{13}+2c_{55})+2c_{33}c_{55}} - \frac{1}{2} \\ \eta_2 &= \frac{c_{22}(c_{33}-c_{44})}{2c_{23}(c_{23}+2c_{44})+2c_{33}c_{44}} - \frac{1}{2} \\ \delta &= \frac{(c_{12}+c_{66})^2-(c_{11}-c_{66})^2}{2c_{11}(c_{11}-c_{66})}, \end{aligned} \tag{8}$$

where v_v is P-wave vertical phase velocity, v_{s1} and v_{s2} are S-wave vertical phase velocity polarized in the $[x_2, x_3]$ and $[x_1, x_3]$ planes, v_{s3} is S-wave horizontal phase

velocity polarized in the $[x_1, x_3]$ but propagating in the x_1 direction, v_1 and v_2 are NMO P-wave velocities for horizontal reflectors in the $[x_1, x_3]$ and $[x_2, x_3]$ planes, and η_1, η_2, and δ are anisotropic parameters in the $[x_1, x_3]$, $[x_2, x_3]$, and $[x_1, x_2]$ planes.

The Christoffel equation in 3D anisotropic media takes the following general form (7):

$$\Gamma_{ik}(x_s, \mathbf{p}) = a_{ijkl}(x_s) p_j p_l - \delta_{ik}, \tag{9}$$

with $p_j = \frac{\partial \tau}{\partial x_j}$ and $a_{ijkl} = \frac{c_{ijkl}}{\rho}$, where p_j are components of the phase vector $\mathbf{p}$, τ is travel-time along the ray, ρ is density, $x_s, s = 1, 2, 3$ are Cartesian coordinates for position along the ray, and δ_{ik} is the Kronecker delta function.

(1) pointed out that setting the S-wave velocity to zero does not compromise accuracy in traveltime computations for TI media. This conclusion can be applied to orthorhombic media as well (35). (3) showed that the kinematics of wave propagation is well described by acoustic approximation.

In orthorhombic media, the Christoffel equation 9 reduces to the following form if v_{s1}, v_{s2}, and v_{s3} are set to zero:

$$\begin{bmatrix} p_1^2 v_1^2 \xi_1 - 1 & \gamma p_1 p_2 v_1^2 \xi_1 & p_1 p_3 v_1 v_v \\ \gamma p_1 p_2 v_1^2 \xi_1 & p_2^2 v_2^2 \xi_2 - 1 & p_2 p_3 v_2 v_v \\ p_1 p_3 v_1 v_v & p_2 p_3 v_2 v_v & p_3^2 v_v^2 - 1 \end{bmatrix}, \tag{10}$$

where $\gamma = \sqrt{1 + 2\delta}$, $\xi_1 = 1 + 2\eta_1$ and $\xi_2 = 1 + 2\eta_2$.

We evaluate the determinant of matrix C-12 and set it to zero. After replacing p_1 with $\frac{k_x}{\phi}$, p_2 with $\frac{k_y}{\phi}$, and p_3 with $\frac{k_z}{\phi}$, we obtain a cubic polynomial in ϕ^2 as follows:

$$\begin{aligned} &-\phi^6 + \phi^4 \left(2v_1^2\eta_1 k_x^2 + v_1^2 k_x^2 + 2v_2^2\eta_2 k_y^2 + v_2^2 k_y^2 + v_v^2 k_z^2\right) \\ &+\phi^2 (v_1^4\gamma^2\xi_1^2 k_x^2 k_y^2 - v_2^2 v_1^2 \xi_1 \xi_2 k_x^2 k_y^2 - 2v_v^2 v_1^2 \eta_1 k_x^2 k_z^2 \\ &-2v_2^2 v_v^2 \eta_2 k_y^2 k_z^2) - v_1^4 v_v^2 \gamma^2 \xi_1^2 k_x^2 k_y^2 k_z^2 + 2v_1^3 v_2 v_v^2 \gamma \xi_1 k_x^2 k_y^2 k_z^2 \\ &-v_1^2 v_2^2 v_v^2 \left(1 - 4\eta_1\eta_2\right) k_x^2 k_y^2 k_z^2 = 0 \ . \end{aligned} \tag{11}$$

One of the roots of the cubic polynomial corresponds to P-waves in acoustic media and is given by the following expression:

$$\phi^2 = \frac{1}{6} \left| -2^{2/3} d - \frac{2\sqrt[3]{2}\,(a^2 + 3b)}{d} + 2a \right| , \tag{12}$$

where

$$a = 2v_1^2\eta_1 k_x^2 + v_1^2 k_x^2 + 2v_2^2\eta_2 k_y^2 + v_2^2 k_y^2 + v_v^2 k_z^2,$$

$$b = v_1^4 k_x^2 k_y^2 \left(2\gamma\eta_1 + \gamma\right)^2 - v_2^2 v_1^2 \left(2\eta_1 + 1\right)\left(2\eta_2 + 1\right) k_x^2 k_y^2$$

$$-2v_v^2 v_1^2 \eta_1 k_x^2 k_z^2 - 2v_2^2 v_v^2 \eta_2 k_y^2 k_z^2,$$

$$c = v_v^2 v_1^4 \left(-k_x^2\right) k_y^2 k_z^2 \left(2\gamma\eta_1 + \gamma\right)^2 + 2v_2 v_v^2 v_1^3 \gamma \left(2\eta_1 + 1\right) k_x^2 k_y^2 k_z^2$$

$$-v_2^2 v_v^2 v_1^2 \left(1 - 4\eta_1\eta_2\right) k_x^2 k_y^2 k_z^2,$$

$$d = \sqrt[3]{-2a^3 + 3\left(e - 9c\right) - 9ab},$$

$$e = \sqrt{|-3b^2\left(a^2 + 4b\right) + 6ac\left(2a^2 + 9b\right) + 81c^2|}.$$

This root reduces to the isotropic P-wave solution when we set $v_1 = v_2 = v_3 = v$, $\eta_1 = \eta_2 = 0$, and $\gamma = 1$, in which ϕ in expression 12 is then given by $|\mathbf{k}|v$, which is the same dispersion relation in isotropic media as that is shown in equation 6. In VTI media: $v_1 = v_2 = v$, $\eta_1 = \eta_2 = \eta$, and $\gamma = 1$, ϕ in expression 12 reduces to

$$\phi(\mathbf{x}, \mathbf{k}) = \sqrt{\frac{1}{2}(v_h^2\, k_h^2 + v_v^2\, k_z^2) + \frac{1}{2}\sqrt{(v_h^2\, k_h^2 + v_v^2\, k_z^2)^2 - \frac{8\eta}{1+2\eta} v_h^2 v_v^2\, k_h^2\, k_z^2}}\,, \tag{13}$$

where $v_h = v\sqrt{1+2\eta}$ is the P-wave phase velocity in the symmetry plane, and $k_h = \sqrt{k_x^2 + k_y^2}$. Expression 13 is the same as the dispersion relation for VTI media (1; 2; 19).

Tilted Orthorhombic Anisotropy

Tectonic movement of the crust may rotate the rocks and tilt the plane containing the vertical cracks, causing a tilted anisotropy. In the case of tilted orthorhombic media, k_x, k_y, and k_z need to be replaced by $\hat{k}_x$, $\hat{k}_y$, and $\hat{k}_z$, which are spatial wavenumbers evaluated in a rotated coordinate system aligned with the vectors normal to the orthorhombic symmetry planes:

$$\begin{aligned} \hat{k}_x &= k_x \cos\phi + k_y \sin\phi \\ \hat{k}_y &= -k_x \sin\phi \cos\theta + k_y \cos\phi \cos\theta + k_z \sin\theta \\ \hat{k}_z &= k_x \sin\phi \sin\theta - k_y \cos\phi \sin\theta + k_z \cos\theta\ , \end{aligned} \tag{14}$$

where θ is the dip angle measured with respect to vertical and ϕ is the azimuth angle, which is the angle between the original X-coordinate and the rotated one. The original vertical axis has the direction of $\{\sin\theta\sin\phi, -\sin\theta\cos\phi, \cos\theta\}$. For a more general rotation, one needs three angles to describe the transformation (41).

LOWRANK APPROXIMATION

For orthorhombic media, the mixed-domain phase operator, ϕ, is given by equation 12. Considering inhomogeneous media, we choose lowrank approximation (20; 21) to implement the mixed-domain operator.

(20; 21) showed that mixed-domain matrix $W(\mathbf{x},\mathbf{k}) = \cos(\phi(\mathbf{x},\mathbf{k})\Delta t)$, which appears in wavefield extrapolation, can be decomposed using a separable representation:

$$W(\mathbf{x},\mathbf{k}) \approx \sum_{m=1}^{M}\sum_{n=1}^{N} W(\mathbf{x},\mathbf{k}_m)a_{mn}W(\mathbf{x}_n,\mathbf{k}). \tag{15}$$

$W(\mathbf{x},\mathbf{k}_m)$ is a submatrix of $W(\mathbf{x},\mathbf{k})$ that consists of a few columns associated with $\mathbf{k}_m$, $W(\mathbf{x}_n,\mathbf{k})$ is another submatrix that contains some rows associated with $\mathbf{x}_n$, and a_{mn} stands for the coefficients. The construction of the separated form 15 follows the method of (14). The main observation is that the columns of $W(\mathbf{x},\mathbf{k}_m)$ are able to span the column space of the original matrix and that the rows of $W(\mathbf{x}_n,\mathbf{k})$ can span the row space as well as possible.

In the case of smooth models, the mixed-domain operator can be decomposed by a low-rank approximation. In models with serious roughness and randomness, the time step may be restricted to small values or otherwise; the rank will end up high. As a result, the computational cost maybe high.

To perform a linear-time lowrank decompositon as proposed by (21), we first need to restrict the mixed-domain $\mathbf{W}$ to n randomly selected rows. In practice, n can be scaled as $O(r\log N_x)$ and r is the numerical rank of $\mathbf{W}$. Then, we perform pivoted QR algorithm (25) to find the corresponding columns for $W(\mathbf{x},\mathbf{k}_m)$. To find the rows for $W(\mathbf{x}_n,\mathbf{k})$, we apply the pivoted QR algorithm to $\mathbf{W}^*$.

Representation 15 speeds up the computation of $p(\mathbf{x},t+\Delta t)$ because

$$\begin{aligned} p(\mathbf{x},t+\Delta t) + p(\mathbf{x},t-\Delta t) &= 2\int e^{-i\mathbf{x}\cdot\mathbf{k}} W(\mathbf{x},\mathbf{k})\hat{p}(\mathbf{k},t)d\mathbf{k} \\ &\approx 2\sum_{m=1}^{M} W(\mathbf{x},\mathbf{k}_m)\left(\sum_{n=1}^{N} a_{mn}\left(\int e^{-i\mathbf{x}\cdot\mathbf{k}} W(\mathbf{x}_n,\mathbf{k})\hat{p}(\mathbf{k},t)d\mathbf{k}\right)\right). \end{aligned} \tag{16}$$

Evaluation of the last formula requires N inverse FFTs. Correspondingly, with lowrank approximation, the cost can be reduced to $O(NN_x\log N_x)$, where N_x is the model size and N is a small number, related to the rank of the above decomposition and it is automatically calculated for some given error level (10^{-5}) with a pre-determined Δt.

Figure 1(a)-1(c) shows an orthorhombic model with smoothly varying velocity – v_1: 1500–3088 m/s, v_2: 1500–3686 m/s, v_v: 1500–3474 m/s, $\eta_1 = 0.3$, $\eta_2 = 0.1$, and

$\gamma = 1.03$. The time step $\Delta t = 4ms$. Figure 2 display error of lowrank decomposition for $\cos(\phi\Delta t)$ at the location (-1.925 km, -1.925 km, -1.925 km) with relatively high velocity values, $v_1 = 2.257$ km/s, $v_2 = 2.534$ km/s, $v_v = 2.438$ km/s. One can find the error level is around 10^{-5}. Figure 3 display error of lowrank decomposition for $\cos(\phi\Delta t)$ at the location (0.575 km, 0.575 km, 0.575 km) with relatively low velocity values, $v_1 = 1.544$ km/s, $v_2 = 1.561$ km/s, $v_v = 1.554$ km/s. One can find the error is also well controlled.

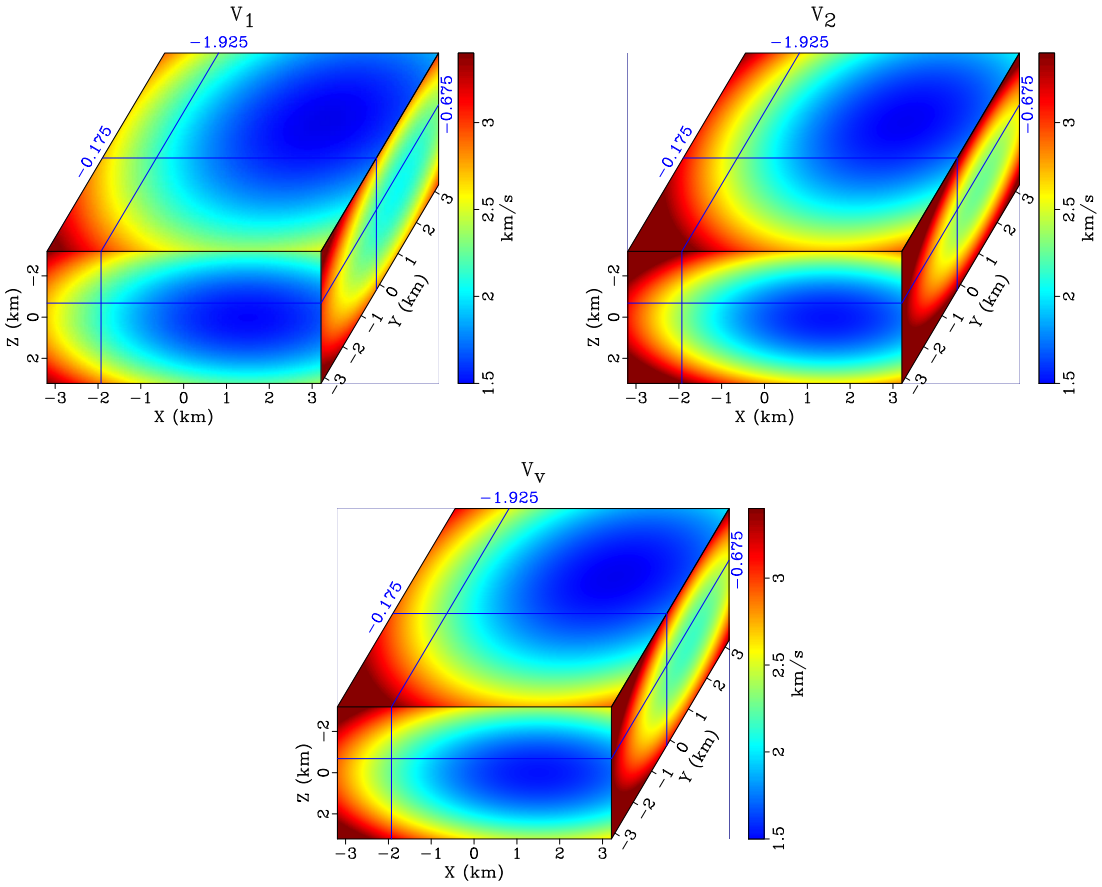


Figure 1: An orthorhombic model with smoothly varying velocity: (a) v_1: 1500–3088 m/s; (b) v_2: 1500–3686 m/s; (c) v_v: 1500–3474 m/s. orthorhombic/tiltn velxfig,velyfig,velzfig

We propose using the above lowrank approximation algorithm to handle mixed-domain operator ϕ in equation 12 for wave extrapolation in orthorhombic media.

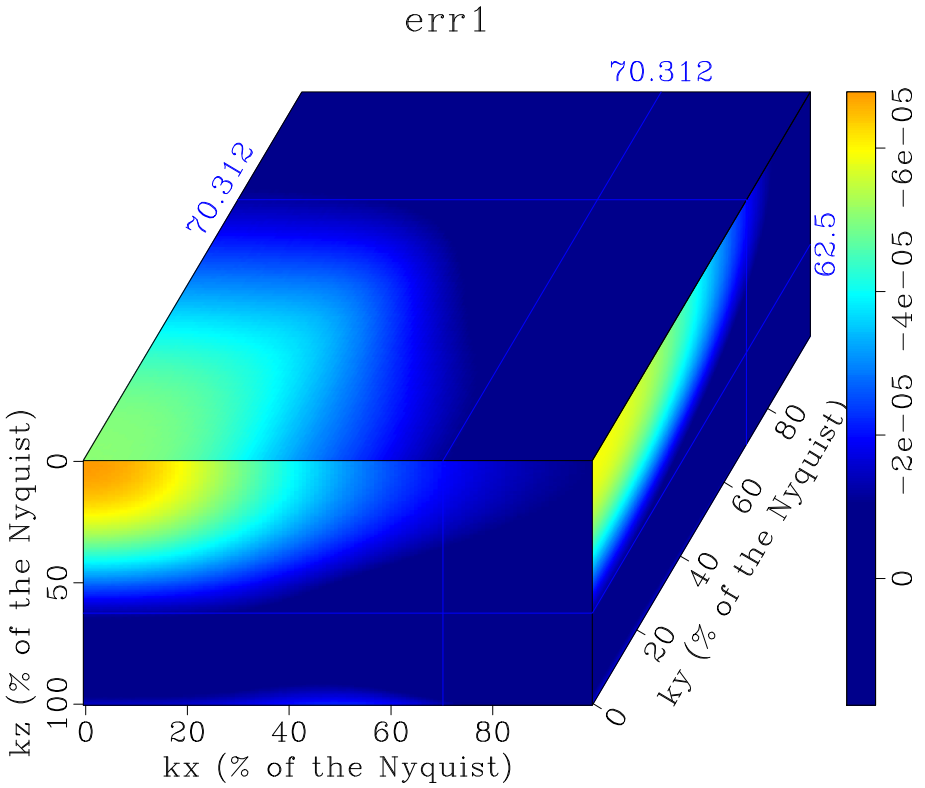


Figure 2: Error plot for the lowrank approximation for $\cos(\phi\Delta t)$ at the location (-1.925 km, -1.925 km, -1.925 km) with relatively high velocity values, $v_1 = 2.257$ km/s, $v_2 = 2.534$ km/s, $v_v = 2.438$ km/s. orthorhombic/tiltn errfig1

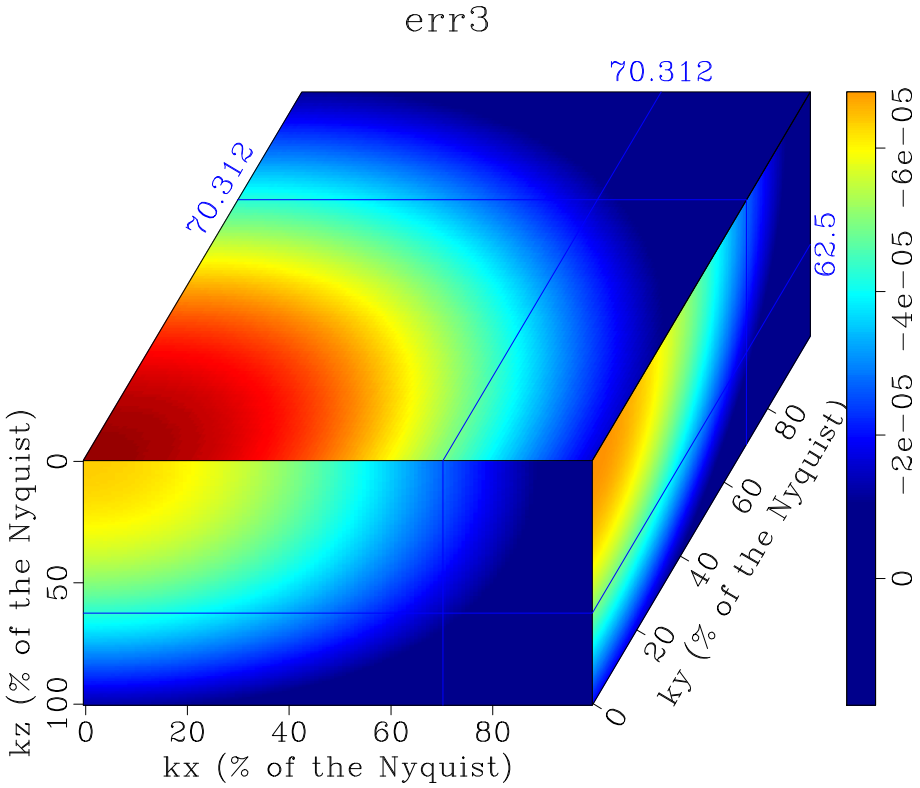


Figure 3: Error plot for the lowrank approximation for $\cos(\phi\Delta t)$ at the location (0.575 km, 0.575 km, 0.575 km) with relatively low velocity values, $v_1 = 1.544$ km/s, $v_2 = 1.561$ km/s, $v_v = 1.554$ km/s. orthorhombic/tiltn errfig3

NUMERICAL EXAMPLES

Figure 4(a)–4(c) shows wavefield snapshots (depth, inline, and crossline) in a vertical orthorhombic medium with constant parameters: $v_v = 2km/s$, $v_1 = 2.1km/s$, $v_2 = 2.05km/s$, $\eta_1 = 0.3$, $\eta_2 = 0.1$, and $\gamma = 1$. The time-step size is 1 ms and the space grid sizes in three directions are all 25 m. As the model is homogeneous, the rank is 1 for the lowrank decomposition. The depth slice is anelliptical, whereas the inline and crossline display different diamond shapes, indicating different VTI properties. In Figures 4(a)–4(c), red dashed lines are calculated using ray tracing. Note that the red dashed lines match the wavefront from the lowrank method very well.

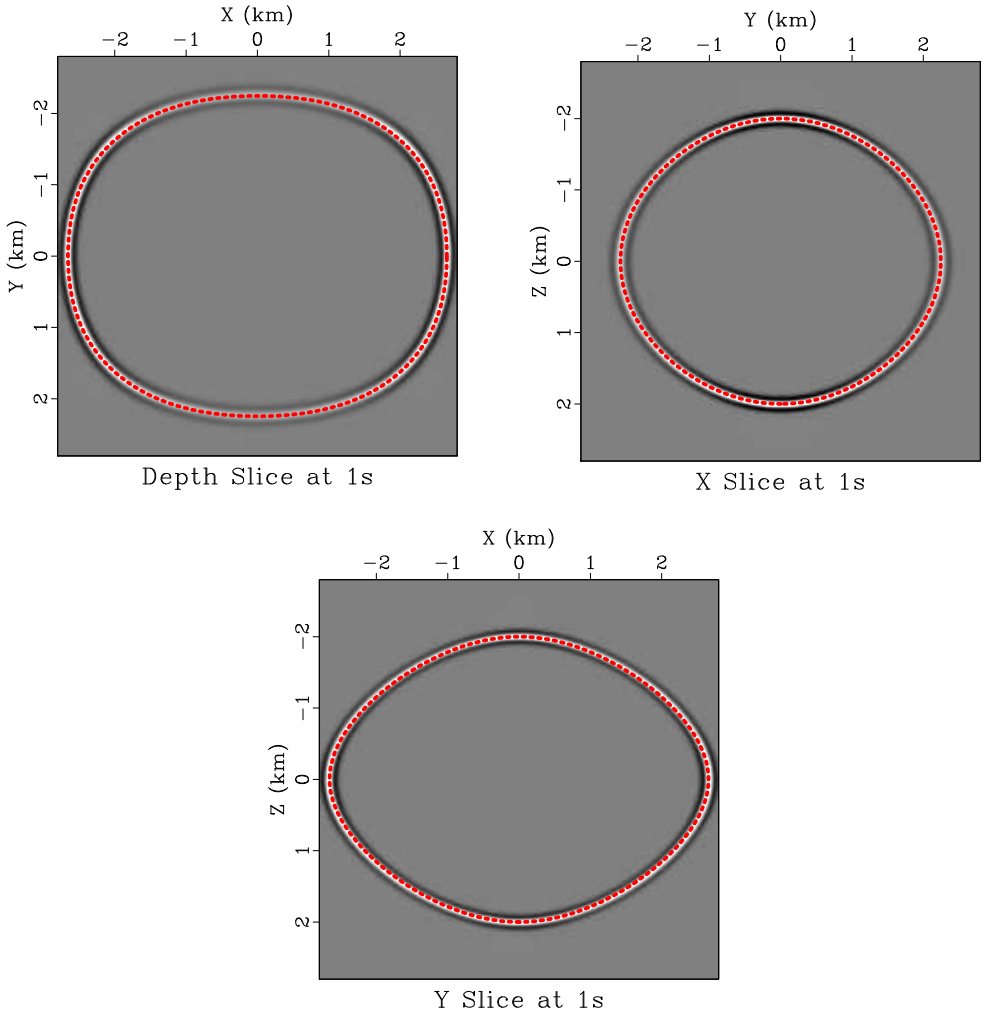


Figure 4: Three slices of the wavefield snapshot based on the dispersion relation 12 at 1 second in a vertical orthorhombic medium: (a) Depth Slice; (b) Inline Slice; (c) Crossline Slice. Also plotted are red curves representing the wavefront at that time calculated using raytracing. orthorhombic/orth wavexy,waveyz,wavexz

To show that the lowrank approximation method can handle rough velocity models, we use a two-layer velocity model with high velocity contrast. The first layer has lower velocity parameters: $v_v = 1.5km/s$, $v_1 = 1.6km/s$, $v_2 = 1.7km/s$, while the values in the other layer are much higher: $v_v = 3.5km/s$, $v_1 = 4.1km/s$, $v_2 = 4.2km/s$. And we use the same anisotropic parameters for both layers: $\eta_1 = 0.3$, $\eta_2 = 0.1$, and $\gamma = 1$. For this test, we use a time step size of 1 ms and a space grid size of 25 m. The rank is 2 calculated by the lowrank decomposition within an error level of 10^{-5}. Figure 5(a) displays the depth slice above the reflector at 0.6 second. Note the snapshot shows the reflection from the velocity contrast. Figure 5(b) and 5(c) show the inline and crossline slices, which indicate strong anisotropy in the medium.

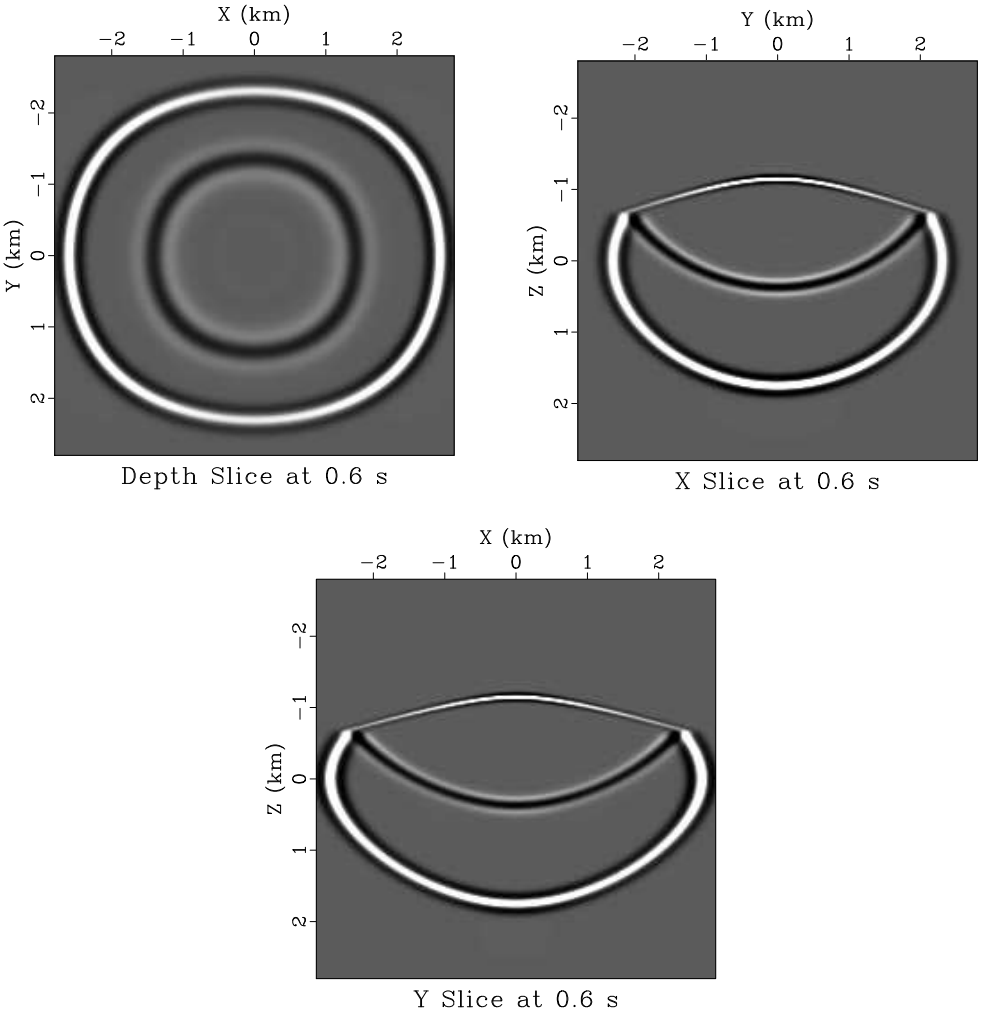


Figure 5: Three slices of the wavefield snapshot by the dispersion relation 12 at 0.6 second in a 2-layer vertical orthorhombic model (high velocity contrast): (a) Depth Slice; (b) Inline Slice; (c) Crossline Slice. orthorhombic/test wavexyt,waveyzt,wavexzt

Our next example is wavefield snapshots in an orthorhombic model with smoothly varying velocity, shown in Figure 1(a)-1(c): v_1: 1500–3088 m/s, v_2: 1500–3686 m/s,

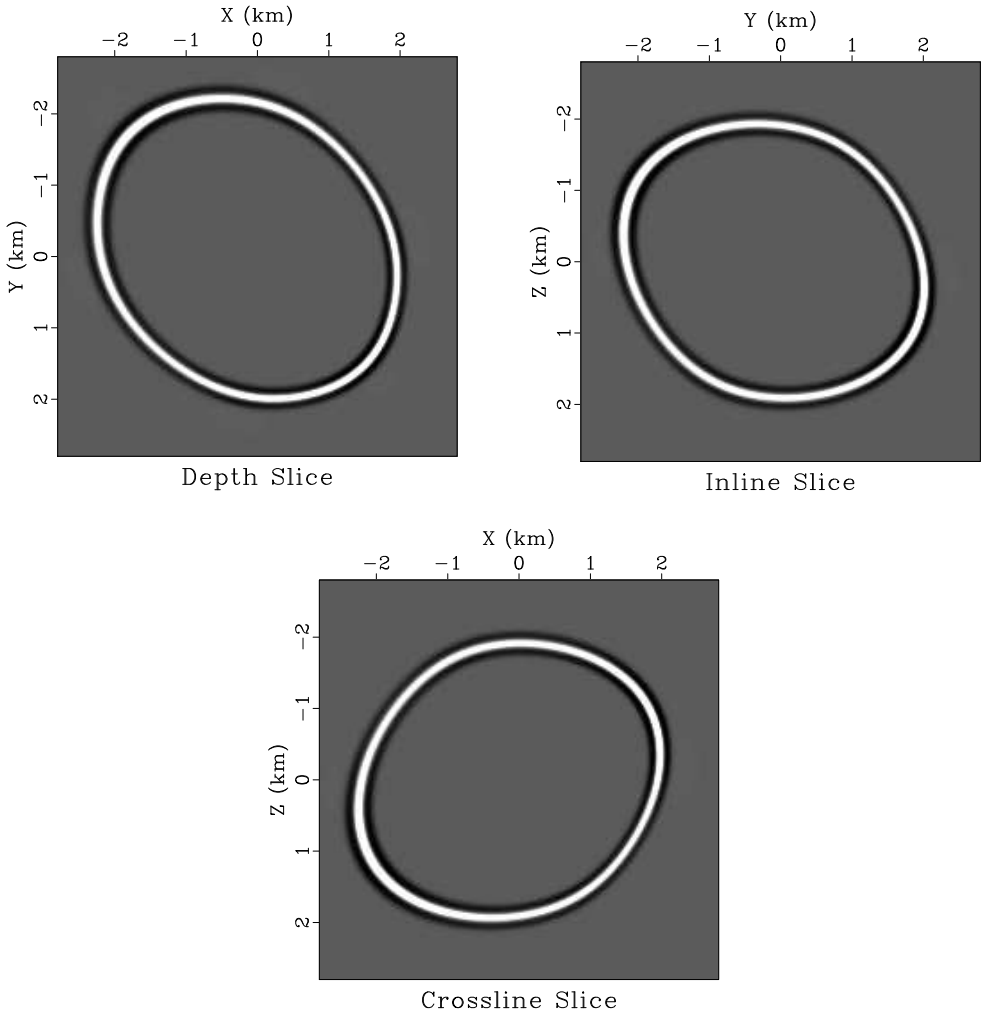


Figure 6: Wavefield snapshots based on the dispersion relation 12 in an rotated and tilted orthorhombic medium ($\theta = \phi = 45\,^\circ$) with variable velocity shown in Figure 1(a)-1(c): (a) Depth Slice; (b) Inline Slice; (c) Crossline Slice orthorhombic/tiltn snapxy4,snapyz4,snapxz4

v_v: 1500–3474 m/s, $\eta_1 = 0.3$, $\eta_2 = 0.1$, and $\gamma = 1.03$. The time-step size is 4 ms. We also rotate the model ($\theta = \phi = 45\,^\circ$). Figure 6(a)–6(c) shows corresponding wavefield snapshots by the dispersion relation 12 in depth, inline, and crossline slices through the central source location. The inline section (Figure 6(b)) displays the strongest anisotropic property, because η_1 is as large as 0.3. Note that the snapshots are free of dispersion and that there is no coupling of qSV and qP waves in the middle. Lowrank parameters were $M = 7$ and $N = 7$. Therefore, the cost is 7 FFTs at each time step.

Table 1 displays rank N required for maintaining an error level of 10^{-5} with different time step size Δt. From table 1, one could find for this smooth model, $\Delta t = 4$ ms and $N = 7$ is the optimal choice for cost consideration. For models with very wide range of parameters and rather complicated structures, the resulting rank may be high, because more space locations and wavenumbers are required to properly represent the original mixed-domain matrix. In order to reduce the computational cost, one may consider Lowrank Finite differences proposed by (31), which is a space-domain finite-difference scheme in which the coefficients of the Laplacian finite-difference stencil is derived from the lowrank approximation.

Δt (ms)	0.5	1	2	3	4	5
$RankN$	5	5	7	7	7	12

Table 1: Rank N calculated from the lowrank approximation of the propagation matrix for a 2D smooth orthorhombic model with different time step size Δt for a given error level 10^{-5}.

CONCLUSIONS

We derive and adopt a dispersion relation for acoustic orthorhombic media so as to model seismic wavefields in such media. To handle the space-wavenumber mixed-domain operator, we apply the lowrank approximation to reduce computational cost. Numerical experiments show that the proposed wavefield extrapolator is accurate. There is no coupling of qSV and qP in the wavefield snapshots because we use the dispersion relation. In addition, our approach yields practically dispersion-free wavefields, and it is also free of stability limitations on media parameters.

ACKNOWLEDGMENTS

We thank Sergey Fomel and Lexing Ying for their help with the lowrank algorithm, and Mirko van der Baan, Richard Bale, Paul J. Fowler, Samuel Gray, and one anonymous reviewer for helpful reviews. We also thank KAUST (King Abdullah University

of Science and Technology) and TCCS (Texas Consortium for Computational Seismology) for financial support. This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Alkhalifah, T., 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.

[2] ———, 2000, An acoustic wave equation for anisotropic media: Geophysics, **65**, 1239–1250.

[3] ———, 2003, An acoustic wave equation for orthorhombic anisotropy: Geophysics, **68**, 1169–1172.

[4] Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

[5] Bale, R. A., 2007, Phase-shift migration and the anisotropic acoustic wave equation: 69th Annual EAGE Meeting, EAGE, Expanded Abstracts, C021.

[6] Bednar, J., 2005, A brief history of seismic migration: Geophysics, **70**, 3MJ–20MJ.

[7] Chapman, C., 2004, Fundamentals of seismic wave propagation: Cambridge University Press.

[8] Chu, C., and P. L. Stoffa, 2011, Application of normalized pseudo-Laplacian to elastic wave modeling on staggered grids: Geophysics, **76**, T113–T121.

[9] Dellinger, J., and F. Muir, 1988, Imaging reflections in elliptically anisotropic media: Geophysics, **53**, 1616–1618.

[10] Dellinger, J., F. Muir, and M. Karrenbach, 1993, Anelliptic approximations for TI media: Jour. Seis. Expl., **2**, 23–40.

[11] Du, X., R. P. Fletcher, and P. J. Fowler, 2010, Pure P-wave propagators versus pseudo-acoustic propagators for RTM in VTI media: 72nd Annual EAGE Meeting, EAGE, Expanded Abstracts, C013.

[12] Duveneck, E., and P. M. Bakker, 2011, Stable P-wave modeling for reverse-time migration in tilted TI media: Geophysics, **76**, S65–S75.

[13] Duveneck, E., P. Milcik, P. M. Bakker, and C. Perkins, 2008, Acoustic VTI wave equations and their application for anisotropic reverse-time migration: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2186–2190.

[14] Engquist, B., and L. Ying, 2009, A fast directional algorithm for high frequency acoustic scattering in two dimensions: Commun. Math. Sci., **7**, 327–345.

[15] Etgen, J., 1989, Accurate wave equation modeling, *in* SEP-60: Stanford Exploration Project, 131–148.

[16] Etgen, J., and S. Brandsberg-Dahl, 2009, The pseudo-analytical method: application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation: 79nd Annual International Meeting, SEG, Expanded Abstracts, 2552–2556.

[17] Etgen, J., S. H. Gray, and Y. Zhang, 2009, An overview of depth imaging in exploration geophysics: Geophysics, **74**, WCA5–WCA17.

[18] Fletcher, R. P., X. Du, and P. J. Fowler, 2009, Reverse time migration in tilted transversely isotropic (TTI) media: Geophysics, **74**, WCA179–WCA187.

[19] Fomel, S., 2004, On anelliptic approximations for qP velocities in VTI media: Geophysical Prospecting, **52**, 247–259.
[20] Fomel, S., L. Ying, and X. Song, 2010, Seismic wave extrapolation using a lowrank symbol approximation: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3092–3096.
[21] ———, 2012, Seismic wave extrapolation using lowrank symbol approximation: Geophysical Prospecting, doi: 10.1111/j.1365–2478.2012.01064.x.
[22] Fowler, P. J., X. Du, and R. P. Fletcher, 2010, Coupled equations for reverse time migration in transversely isotropic media: Geophysics, **75**, S11–S22.
[23] Fowler, P. J., and R. King, 2011, Modeling and reverse time migration of orthorhombic pseudo-acoustic P-waves: 81th Ann. Internat. Mtg., Soc. Expl. Geophys., 190–195.
[24] Fowler, P. J., and C. Lapilli, 2012, Generalized pseudospectral methods for modeling and reverse-time migration in orthorhombic media: 74th Annual EAGE Meeting, EAGE, Expanded Abstracts, AO22.
[25] Golub, G. H., and C. F. Van Loan, 1996, Matrix computations: John Hopkins.
[26] Grechka, V., L. Zhang, and J. W. Rector, 2004, Shear waves in acoustic anisotropic media: Geophysics, **69**, 576–582.
[27] Helbig, K., 1983, Elliptical anisotropyits significance and meaning: Geophysics, **48**, 825–832.
[28] Liu, F., S. A. Morton, S. Jiang, L. Ni, and J. P. Leveille, 2009, Decoupled wave equations for P and SV waves in an acoustic VTI media: 79th Ann. Internat. Mtg., Soc. Expl. Geophys., 2844–2848.
[29] Muir, F., 1985, A practical anisotropic system: SEP-44, 55–58.
[30] Song, X., and S. Fomel, 2011, Fourier finite-difference wave propagation: Geophysics, **76**, T123–T129.
[31] Song, X., S. Fomel, and L. Ying, 2013, Lowrank finite-differences and lowrank Fourier finite-differences for seismic wave extrapolation in the acoustic approximation: Geophysical Journal International, doi: 10.1093/gji/ggt017.
[32] Song, X., S. Fomel, L. Ying, and T. Ding, 2011, Lowrank finite-differences for wave extrapolation: 81th Ann. Internat. Mtg., Soc. Expl. Geophys., 3372–3376.
[33] Soubaras, R., and Y. Zhang, 2008, Two-step explicit marching method for reverse time migration: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2272–2276.
[34] Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966.
[35] Tsvankin, I., 1997, Anisotropic parameters and P-wave velocity for orthorhombic media: Geophysics, **62**, 1292–1309.
[36] ———, 2005, Seismic signature and analysis of reflection data in anisotropic media: Elsevier Science.
[37] Tsvankin, I., and V. Grechka, 2011, Seismology of azimuthally anisotropic media and seismic fracture characterization: Society of Exploration Geophysicists.
[38] Yoon, K., S. Suh, J. Ji, J. Cai, and B. Wang, 2010, Stability and speedup issues in TTI RTM implementation: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3221–3225.
[39] Zhan, G., R. C. Pestana, and P. L. Stoffa, 2012, Decoupled equations for reverse time migration in tilted transversely isotropic media: Geophysics, **77**, T37–T45.

[40] Zhang, H., G. Zhang, and Y. Zhang, 2009, Removing s-wave noise in TTI reverse time migration: 79th Ann. Internat. Mtg., Soc. Expl. Geophys., 2849–2853.
[41] Zhang, H., and Y. Zhang, 2011, Reverse time migration in vertical and tilted orthorhombic media: 81th Ann. Internat. Mtg., Soc. Expl. Geophys., 185–189.
[42] Zhang, Y., and G. Zhang, 2009, One-step extrapolation method for reverse time migration: Geophysics, **74**, A29–A33.
[43] Zhang, Y., H. Zhang, and G. Zhang, 2011, A stable TTI reverse time migration and its implementation: Geophysics, **76**, WA3–WA9.

A robust approach to time-to-depth conversion and interval velocity estimation from time migration in the presence of lateral velocity variations

*Siwei Li and Sergey Fomel**

ABSTRACT

The problem of conversion from time-migration velocity to an interval velocity in depth in the presence of lateral velocity variations can be reduced to solving a system of partial differential equations. In this paper, we formulate the problem as a nonlinear least-squares optimization for seismic interval velocity and seek its solution iteratively. The input for inversion is the Dix velocity which also serves as an initial guess. The inversion gradually updates the interval velocity in order to account for lateral velocity variations that are neglected in the Dix inversion. The algorithm has a moderate cost thanks to regularization that speeds up convergence while ensuring a smooth output. The proposed method should be numerically robust compared to the previous approaches, which amount to extrapolation in depth monotonically. For a successful time-to-depth conversion, image-ray caustics should be either nonexistent or excluded from the computational domain. The resulting velocity can be used in subsequent depth-imaging model building. Both synthetic and field data examples demonstrate the applicability of the proposed approach.

INTRODUCTION

Time-domain seismic processing has been a popular and effective tool in areas with mild lateral velocity variations (35; 28; 1). Time migration produces images in the time coordinate as opposed to the usual depth coordinate. The time coordinate, along with time-migration velocities, is determined during time migration by optimizing image qualities. However, it is highly sensitive to lateral velocity changes (6; 3). Therefore, the time-migrated image is usually distorted (20; 7; 8; 9; 21). For this reason, in many cases time migrations are inadequate for accurate geological interpretation of subsurface structures. On the other hand, time-migration velocities can be conveniently and efficiently estimated either by repeated migrations (36) or by velocity continuation (15). Depth migration can handle general media and output

*e-mail: sergey.fomel@beg.utexas.edu

images in regular Cartesian depth coordinates. But it requires an accurate interval velocity model construction, which is in practice both challenging and time-consuming. An iterative process of tomographic updates is often employed (4; 31), where a good initial interval velocity model is crucial for achieving the global minimum. It is thus of great interest to convert the time-migration velocity to the depth domain in order to unravel inherent distortions in time-domain images and to provide a reasonable starting model for building depth-imaging velocities.

The relationship between time and depth coordinates was first explained by (20) through the concept of *image rays*. An interval velocity model can be converted to one in the time domain by tracing image rays that dive into earth with slowness vector normal to the surface. The time coordinates are then defined by the traveltime along image rays and its surface location (22; 28). However this process is not trivially revertible and it does not reveal directly the connection between time-migration velocity and interval velocity. According to (12), in a layered medium where $v = v(z)$, the image rays run straightly downward and the time-migration velocities are the root-mean-square (RMS) velocities appearing in a truncated Taylor approximation for traveltimes. The Dix inversion formula is exact in a $v(z)$ medium, where the conversion between time- and depth-domain attributes is theoretically straightforward.

Even a mild lateral velocity variation can cause image rays to bend and invalidate Dix inversion. (7; 8) studied and established theoretical relations between the time-migration velocity and the seismic velocity in depth for general media using the paraxial ray-tracing theory. They showed that the conventional Dix velocity is equal to the ratio of the interval velocity and the geometrical spreading of image rays. This is consistent with Dix formula because when $v = v(z)$ the geometrical spreading equals to one everywhere. In order to carry out the time-to-depth conversion in the presence of lateral velocity variations, one can solve a nonlinear partial differential equation (PDE) of elliptic type with boundary conditions on the surface (9). The problem is mathematically ill-posed. (9) adopt a two-step numerical procedure for the time-to-depth conversion. The first step is a Lax-Friedrichs-like finite-difference method or a spectral Chebyshev method to solve for geometrical spreading in the time coordinate. The next step is a Dijkstra-like solver motivated by the fast marching method (29) for velocity conversion and coordinate mapping. In this approach, it is crucial to yield the development of low harmonics and damp the high harmonics during the first stage. (21) discussed an extension of the time-to-depth conversion problem along 2-D profiles in 3-D. An essential part of the algorithm is the time-stepping (integration) along image-rays. The robustness of these methods may not be satisfactory in practice because of stability issues that arise from the ill-posed nature of the problem.

In this paper, we start with the theoretical relations derived by (7; 8) but cast the original problem in a nonlinear iterative optimization framework. This idea is motivated by the observation that, for arbitrary depth velocity model, two of the PDEs can be always satisfied, while the remaining one associated with image-ray geometrical spreading can be rewritten in the form of a cost functional that indicates

errors in interval velocity. A key benefit in our formulation is that each iteration will give a global update of the whole velocity model. In contrast, the previous approaches only consider information locally at each time step. Another advantage is that we are able to include regularization in the optimization framework in order to deal with the ill-posedness issue.

The paper is organized as follows. We first show theoretical derivations of all necessary components involved in the proposed approach. Next, we develop a numerical implementation for 2-D time-to-depth conversion. Finally, we test the algorithm on synthetic and field data examples. We conclude the paper by giving some discussions on the proposed method.

THEORY

For completeness, we will first review some basic concepts related to time-to-depth conversion. Then we show the theoretical derivation for an optimization formulation of the problem and a corresponding inversion procedure. For simplicity, we restrict our analysis below to the 2-D case.

Connection between time- and depth-domain attributes

In Figure 1, we illustrate image rays in 2-D and a *forward* mapping from depth coordinate (z, x) to time coordinate (t_0, x_0) (20). Throughout this paper t_0 stands for one-way time, while time migration usually produces images in two-way time (35), i.e. $(2t_0, x_0)$. Under the assumption of no caustics, for each subsurface location (z, x) we consider the image ray from (z, x) to the surface, where it emerges at point $(0, x_0)$, with slowness vector normal to the surface. Here x_0 is the location of the image ray at the earth surface and is a scalar. t_0 is the traveltime along this image ray between (z, x) and $(0, x_0)$. The forward mapping $t_0(z, x)$ and $x_0(z, x)$ can be done with a knowledge of interval velocity $v(z, x)$. A unique *inverse* mapping $z(t_0, x_0)$ and $x(t_0, x_0)$ also exists that enables us to directly map the time-migrated image to depth.

The counterpart for $v(z, x)$ in the time-domain is the time-migration velocity $v_m(t_0, x_0)$, which is commonly estimated in prestack Kirchhoff time migration (35; 15). In a $v(z)$ medium, v_m corresponds to the root-mean-square (RMS) velocity:

$$v_m(t_0) = \sqrt{\frac{1}{t_0}\int_0^{t_0} v^2(z(t))\,dt}\ . \tag{1}$$

A time-to-depth velocity conversion can be done by first applying the Dix inversion formula (12), which is theoretically exact in this case:

$$v_d(t) = \sqrt{\frac{d}{dt_0}(t_0 v_m^2(t_0))}\ , \tag{2}$$

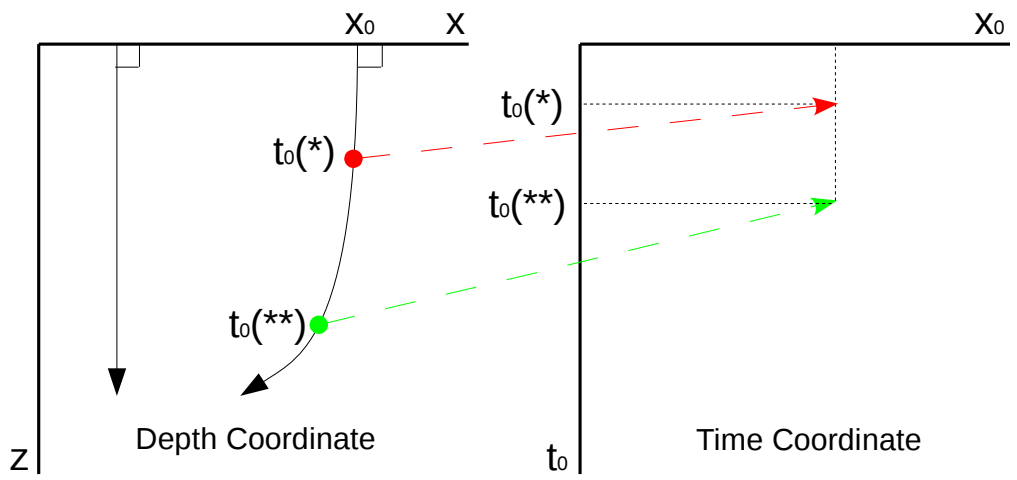


Figure 1: Image ray in (left) the depth-domain can be traced with a source at location x_0 with slowness vector normal to the earth surface. Each depth coordinate (z, x) along this image ray is then mapped into (right) the time coordinate (t_0, x_0) by using its corresponding traveltime t_0 and source location x_0. time2depth2/. imageray

followed by performing a simple conversion from $v_d(t)$ to $v(z)$ according to $z = \frac{1}{2}\int_0^t v_d(t)dt$.

In equations 1 and 2, there is no dependency on x_0 or x, because image rays are all vertical. For an arbitrary medium, image rays will bend as they travel through the medium (22). Therefore, in general, v_m is a function of both t_0 and x_0 and no longer satisfies the simple expression 1, which limits the applicability of the Dix formula. (7) proved that the seismic velocity and the Dix velocity in this case are connected through geometrical spreading Q of image rays:

$$v_d(t_0, x_0) \equiv \sqrt{\frac{\partial}{\partial t_0}(t_0 v_m^2(t_0, x_0))} = \frac{v(z(t_0, x_0), x(t_0, x_0))}{Q(t_0, x_0)} \ . \tag{3}$$

In equation 3, the generalized Dix velocity is defined by (7) in a way similar to equation 2 with a change from d/dt_0 to a partial differentiation with respect to t_0. The quantity Q is related to $x_0(z, x)$ using its definition (26; 7), as follows:

$$|\nabla x_0|^2 = \frac{1}{Q^2} \ . \tag{4}$$

Combining equations 3 and 4 results in

$$|\nabla x_0(z, x)|^2 = \frac{v_d^2(t_0(z, x), x_0(z, x))}{v^2(z, x)} \ . \tag{5}$$

The traveltimes along image rays obey the eikonal equation (20; 10), thus

$$|\nabla t_0(z,x)|^2 = \frac{1}{v^2(z,x)} \ . \tag{6}$$

Finally, since x_0 remains constant along each image ray, there is an orthogonality condition between gradients of t_0 and x_0 (7):

$$\nabla t_0(z,x) \cdot \nabla x_0(z,x) = 0 \ . \tag{7}$$

Equations 5, 6 and 7 form a system of nonlinear PDEs for $t_0(z,x)$ and $x_0(z,x)$. The input is $v_d(t_0,x_0)$, estimated from $v_m(t_0,x_0)$ by equation 3. Solving a boundary-value problem for the PDEs should provide $v(z,x)$, as well as $t_0(z,x)$ and $x_0(z,x)$. Because seismic acquisitions are limited to the earth surface, we can only use boundary conditions at the surface. For a rectangular Cartesian domain with $z=0$ being the surface, the boundary conditions are

$$\left\{ \begin{array}{rcl} t_0(0,x) &=& 0 \ , \\ x_0(0,x) &=& x \ . \end{array} \right. \tag{8}$$

In Appendix A, we show that the time-to-depth conversion is an ill-posed problem because it requires solving a Cauchy-type problem for an elliptic PDE. The missing boundary conditions on sides of the computational domain other than those in equation 8 can induce numerical instability when extrapolating in t_0 (or, equivalently, z). Instead, we consider an alternative formulation of the problem in the following section.

The optimization formulation

Given boundary conditions 8, equation 6 describes the traveltime t_0 of a velocity model with a plane-wave source at the surface. For a given t_0, equation 7 is a first-order linear PDE on x_0 and thus computation of x_0 is straightforward. Our idea is inspired by a natural logic: if the resulting x_0 does not satisfy equation 5, we need to modify v in a way such that the misfit decreases, and repeat the process until convergence.

Mathematically, we define a cost function $f(z,x)$ based on equation 5:

$$f(z,x) = \nabla x_0 \cdot \nabla x_0 - v_d^2 \, w \ , \tag{9}$$

where for convenience we use slowness-squared $w(z,x) = v^{-2}(z,x)$ instead of v. Note that f is dimensionless. The discretized form of equation 9 reads

$$\mathbf{f} = (\nabla \mathbf{x_0} \cdot \nabla) \, \mathbf{x_0} - \mathrm{diag}(\mathbf{v_d} \star \mathbf{v_d}) \, \mathbf{w} \equiv \mathbf{L}_{x_0} \, \mathbf{x_0} - \mathrm{diag}(\mathbf{v_d} \star \mathbf{v_d}) \, \mathbf{w} \ . \tag{10}$$

In equation 10, $\mathbf{f}, \mathbf{x_0}, \mathbf{v_d}$ and $\mathbf{w}$ are all column vectors after discretizing the computational domain (z,x). For example, $\mathbf{x_0}$ is the discretized column vector of $x_0(z,x)$. The

vector $\mathbf{v_d}$ may require interpolation because it is in (t_0, x_0) while the discretization is in (z, x). The interpolation can be done after forward mapping from (z, x) to (t_0, x_0) at current velocity model. We denote an operator which is a matrix $\mathbf{L}_{x_0} = \nabla \mathbf{x_0} \cdot \nabla$. The other operator diag() expands a vector into a diagonal matrix. Finally, the symbol $\star$ stands for an element-wise vector-vector multiplication.

As is common in many optimization problems, we seek to minimize the least-squares norm of $\mathbf{f}$:

$$E[\mathbf{w}] = \frac{1}{2}\|\mathbf{f}\|^2 = \frac{1}{2}\mathbf{f}^T\mathbf{f} \ , \tag{11}$$

where the superscript T stands for transpose. The Gauss-Newton method in optimization requires linearizing the cost function in equation 10:

$$\frac{\partial f}{\partial w} = 2\left(\nabla x_0 \cdot \nabla\right)\frac{\partial x_0}{\partial w} - 2\, v_d\, w\, \frac{\partial v_d}{\partial w} - v_d^2 \ . \tag{12}$$

The Fréchet derivative matrix $\mathbf{J}$ required by inversion is the discretized form of equation 12, i.e., $\mathbf{J} = \partial\mathbf{f}/\partial\mathbf{w}$. In Appendix B we find that $\mathbf{J}$ is a cascade and summation of several parts. An update $\delta\mathbf{w}$ at current $\mathbf{w}$ is found by solving the following normal equation arising from the Gauss-Newton approach (5):

$$\delta\mathbf{w} = \left[\mathbf{J}^T\mathbf{J}\right]^{-1}\mathbf{J}^T(-\mathbf{f}) \ . \tag{13}$$

Equations 5 and 17 together suggest a nonlinear inversion procedure for solving the original system of PDEs 5, 6 and 7. The inversion is analogous to traveltime tomography but with more complexity. The cost 9 can be interpreted as difference between modeled and observed geometrical spreadings. However, both of them depend on the model v, while in traveltime tomography the observed arrival times are independent of v. The forward modeling in our case involves two steps, which construct a curvilinear coordinate system that is sensitive to lateral velocity variations. On the other hand, the forward modeling in traveltime tomography consists of only one step. Last but not least, unlike traveltime tomography, we have observations everywhere in the computational domain, except for areas excluded due to instabilities of the numerical implementation, as we will discuss later.

Before introducing a numerical implementation, we would like to point out several important facts and assumptions that make a successful time-to-depth conversion possible by the proposed method:

- Caustics must be excluded from the computational domain. In regions where caustics develop, the gradient ∇x_0 goes to infinity and the cost function is not well-defined. For all numerical examples in this paper, we do not encounter this issue. In the Discussion section, we provide a possible strategy to cope with this limitation.

- According to derivations in Appendix B, the calculation of $\delta\mathbf{w}$ depends on values of $\partial\mathbf{v_d}/\partial\mathbf{t_0}$ and $\partial\mathbf{v_d}/\partial\mathbf{x_0}$. Thus the input v_d should be differentiable.

This requirement can be satisfied during v_m estimation by using regularization (15).

- Similarly to all nonlinear inversions, the proposed method requires a prior model that is sufficiently close to desired model at the global minimum $E = 0$. Meanwhile, to guarantee stability and a smooth output, some form of regularization should be imposed during inversion (13; 37).
- Our formulation does not change the ill-posed nature of the original problem. One assumption is that condition 8 describes all *in-flow* domain boundaries of t_0 and x_0. In other words, the image rays are only allowed to be either parallel to or exiting (*out-flow*) all other boundaries of the computational domain except the surface.

For the prior model, we adopt the Dix-inverted model. In other words, we seek to refine the interval velocity given by equation 2 by taking the geometrical spreading of image rays into consideration according to equation 3.

NUMERICAL IMPLEMENTATION

Below we outline the steps involved in computing one linearization update:

1. Given current $v(z, x)$, solve equation 6 for $t_0(z, x)$ with $t_0(0, x) = 0$;
2. Given $t_0(z, x)$ and $x_0(0, x) = x$, solve equation 7 for $x_0(z, x)$;
3. Given $t_0(z, x)$ and $x_0(z, x)$, interpolate $v_d(z, x)$ from $v_d(t_0, x_0)$ and compute $f(z, x)$ based on equation 9;
4. Assemble linear operator B-5 and solve equation 17 for $\delta w(z, x)$.

First, we apply the fast-marching method (29; 30) to solve the eikonal equation 6 by initializing a plane-wave source at $z = 0$. Computation for x_0 can be incorporated into t_0 by adopting the upwind finite-differences of t_0 for equation 7. In Figure 2, consider a currently updated grid point C during forward modeling of t_0. If it has only one upwind neighbor A that is inside the wave-front, $t_0(C) > t_0(A)$, then the image ray must be aligned with grid segment AC and therefore $x_0(C) = x_0(A)$. We refer to this scenario as one-sided. If C has two upwind neighbors A and B, $t_0(C) > \max\{t_0(A), t_0(B)\}$, and they are both inside the wave-front, then the image ray must intersect the simplex ABC from an angle. In this case, we compute $x_0(C)$ from

$$\frac{(t_0(C) - t_0(A))(x_0(C) - x_0(A))}{h_z^2} + \frac{(t_0(C) - t_0(B))(x_0(C) - x_0(B))}{h_x^2} = 0 \, . \qquad (14)$$

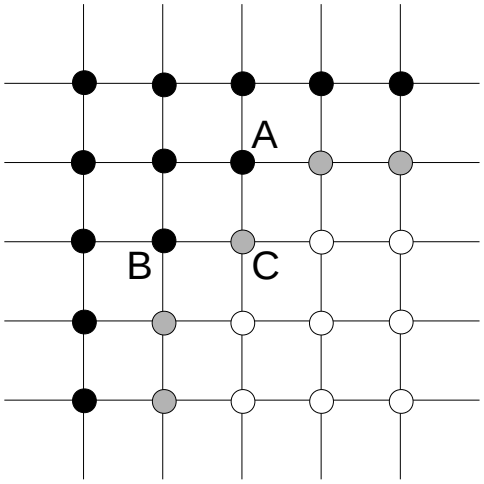


Figure 2: A modified fast-marching method for forward modeling. Black dots represent region that has been swept by the wave-front, gray dots are the expanding wave-front and grid points being updated, and white dots are region yet to be reached. time2depth2/. fmm

Because $\mathbf{x_0}$ at certain grid points is calculated by one-sided scenario, $\mathbf{L}_{x_0}$ there contains all zeros. Consequently, an evaluation of the cost $\mathbf{f}$ at these locations with $\mathbf{L}_{x_0}\,\mathbf{x_0}$ becomes inaccurate. We exclude these regions from $\mathbf{f}$ and expect inversion to fill them.

Next, we apply simple bilinear interpolation for $v_d(z,x)$ and estimate $\delta\mathbf{w}$ by solving equation 17 using shaping regularization (16). We use a triangular smoother with adjustable size as the shaping operator. We find in numerical tests that shaping significantly improves convergence speed compared to that of the traditional Tikhonov regularization (32) with gradient operators. We also observe that without regularization the model update can be undesirably oscillatory. We believe this phenomenon is related to the ill-posedness of the PDEs.

Finally, we reduce computational cost by adopting the method of conjugate gradients (19) and an efficient implementation of $\mathbf{J}$, as well as its adjoint, according to the equations derived in Appendix B. For this purpose, we choose the upwind finite-difference scheme (17; 24) based on $\mathbf{t_0}$ for both $\mathbf{L}_{t_0}$ and $\mathbf{L}_{x_0}$. As shown by (25), applying $\mathbf{J}$ and its transpose involves only sparse triangularized matrix-vector multiplications and is therefore inexpensive. For example, at each grid point $\mathbf{L}_{t_0}^{-1}$ relies on only its upwind neighbors. The computational complexity of $\mathbf{J}$ and $\mathbf{J}^T$ is $O(N)$, where N is the total number of grid points.

EXAMPLES

Constant velocity gradient model

In a medium with a constant velocity gradient

$$v(z,x) = v_0 + g_z z + g_x x \ , \tag{15}$$

image rays are segments of circles parallel to each other, and all attributes involved in time-to-depth conversion have analytical solutions. In Appendix C, we derive the following relationships:

$$x_0(z,x) = x + \frac{\sqrt{(v_0+g_x x)^2 + g_x^2 z^2} - (v_0 + g_x x)}{g_x} \ , \tag{16}$$

$$t_0(z,x) = \frac{1}{g}\text{arccosh}\left[\frac{g^2\left(\sqrt{(v_0+g_x x)^2 + g_x^2 z^2} + g_z z\right) - v g_z^2}{v g_x^2}\right] \ , \tag{17}$$

where $g = \sqrt{g_z^2 + g_x^2}$. The migration velocity v_m and Dix velocity v_d take the following expressions:

$$v_m(t_0,x_0) = \frac{(v_0+g_x x_0)^2}{t_0\left(g\coth(g t_0) - g_z\right)} \ , \tag{18}$$

$$v_d(t_0,x_0) = \frac{(v_0+g_x x_0) g}{g\cosh(g t_0) - g_z \sinh(g t_0)} \ . \tag{19}$$

It is easy to verify from equations 16 and 17 that $|\nabla x_0| = 1$, $|\nabla t_0| = 1/v$ and $\nabla x_0 \cdot \nabla t_0 = 0$. Because there is no geometrical spreading of image rays in this case, the Dix velocity will be equal to the interval velocity according to equation 3. However, a Dix-inverted model will still be distorted if $g_x \neq 0$ because of the lateral shift of image rays.

Figures 3 and 4 show a velocity model with $v(z,x) = 1.5 + 0.75z + 0.5x$ km/s and the corresponding analytical $v_d(t_0,x_0)$, $x_0(z,x)$ and $t_0(z,x)$. Clearly, the right domain boundary is of in-flow type that violates our assumption. To address this challenge, we include Dix velocity in regions beyond the original left and right boundaries during inversion, but mask out the cost in these regions. It means that the time-to-depth conversion is performed in a sub-domain of time-domain attributes, such that information on the in-flow boundary is available. Afterwards, we apply Dix inversion to the expanded model and use the result as the prior model. We use the exact Dix velocity in equation 19 for evaluating the right-hand side of 5. Then, in total three linearization updates are carried out, which decreases E to relative 0.6%. The radiuses of triangular smoother in shaping are 8 m vertically and 30 m horizontally (8 m $\times$ 30 m). At last, we cut the computational domain back to its original size. Figures 5 and 6 compare the cost and model misfit before and after inversion.

We also synthesize data with Kirchhoff modeling (18) for several horizontal reflectors using the exact model, and examine the subsurface scattering-angle common-image-gathers from Kirchhoff prestack depth migration (34) as an evidence of interval velocity improvements. In Figure 7, the shallower events do not improve significantly because the image rays have not yet bent considerably. Deeper events become noticeably flatter after applying the proposed method.

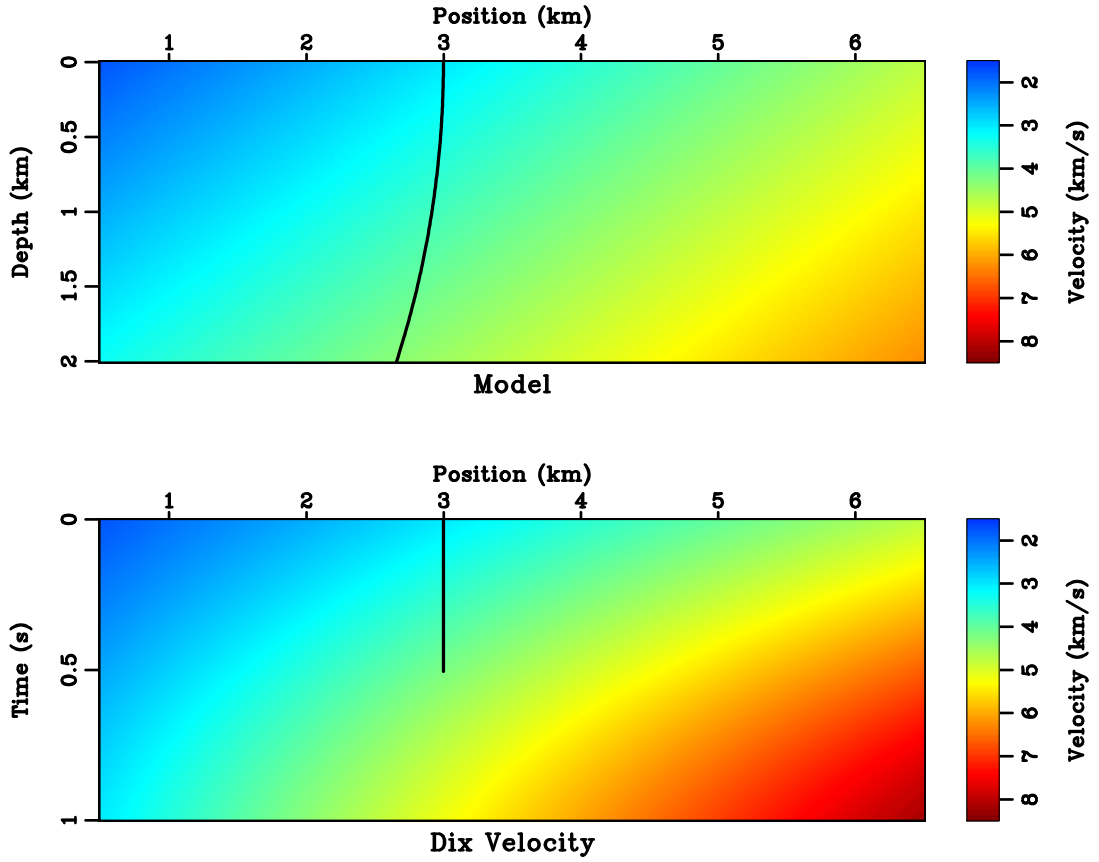


Figure 3: (Top) a constant velocity gradient model and (bottom) the analytical Dix velocity v_d. A curved image ray is mapped to the time domain as a straight line. time2depth2/vgrad vgrad

Constant horizontal slowness-squared gradient model

Another medium that provides analytical time-to-depth conversion formulas is

$$w(z,x) = w_0 - 2q_x x \ , \tag{20}$$

i.e., the slowness-squared changes linearly in the horizontal direction. In Appendix D, we show that in this case the coordinate mapping follows

$$x_0(z,x) = \frac{2w_0 x + q_x z^2}{w_0 + 2q_x x + \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}} \ , \tag{21}$$

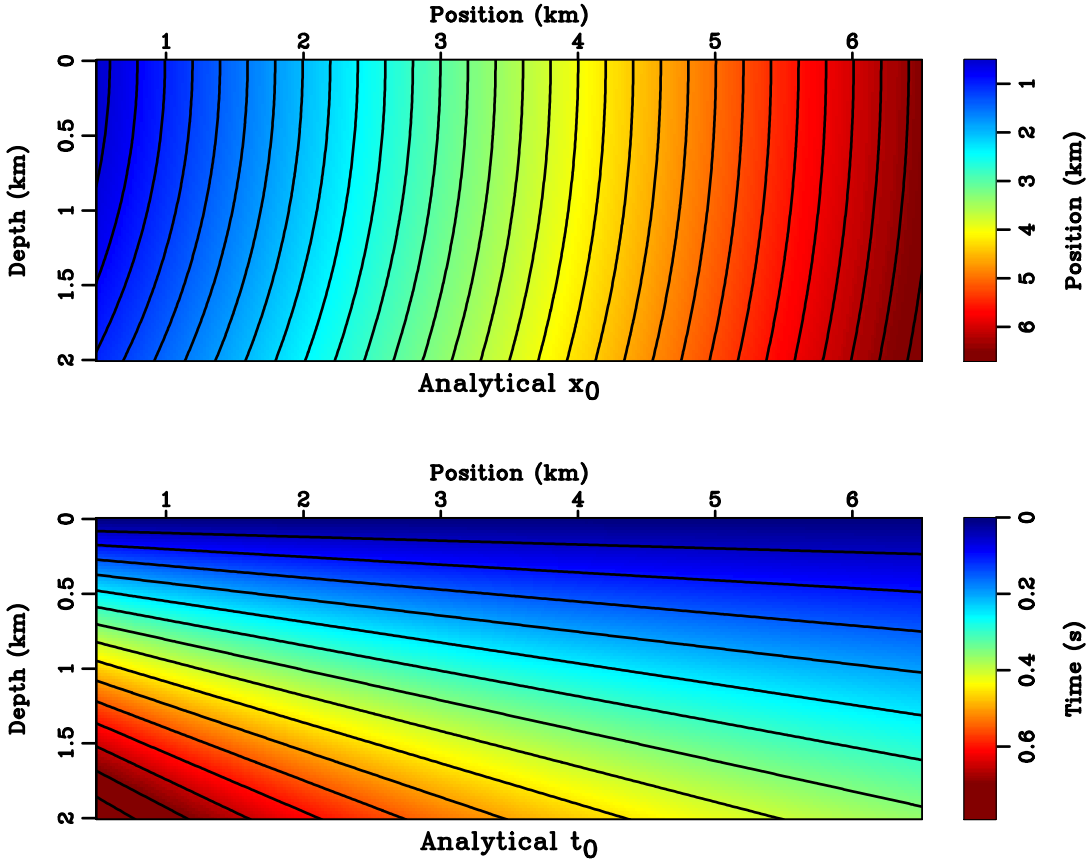


Figure 4: Analytical values of (top) t_0 and (bottom) x_0 of the model in Figure 3. Both figures are overlaid with contour lines that, according to equation 7, are perpendicular to each other. Each contour line of x_0 is an image ray, while the contours of t_0 illustrate the propagation of a plane-wave. time2depth2/vgrad analy

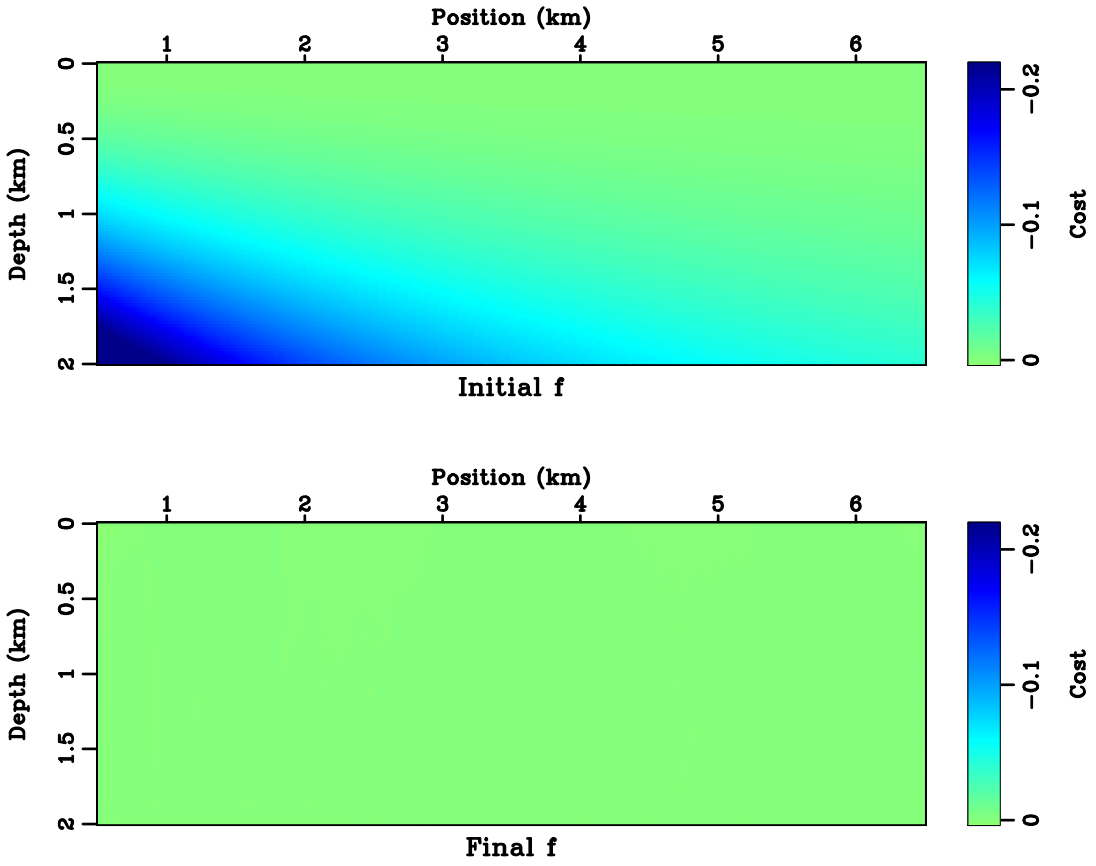


Figure 5: The cost defined by equation 9 (top) before and (bottom) after inversion. The least-squares norm of cost E is decreased from 9.855 to 0.057. time2depth2/vgrad cost

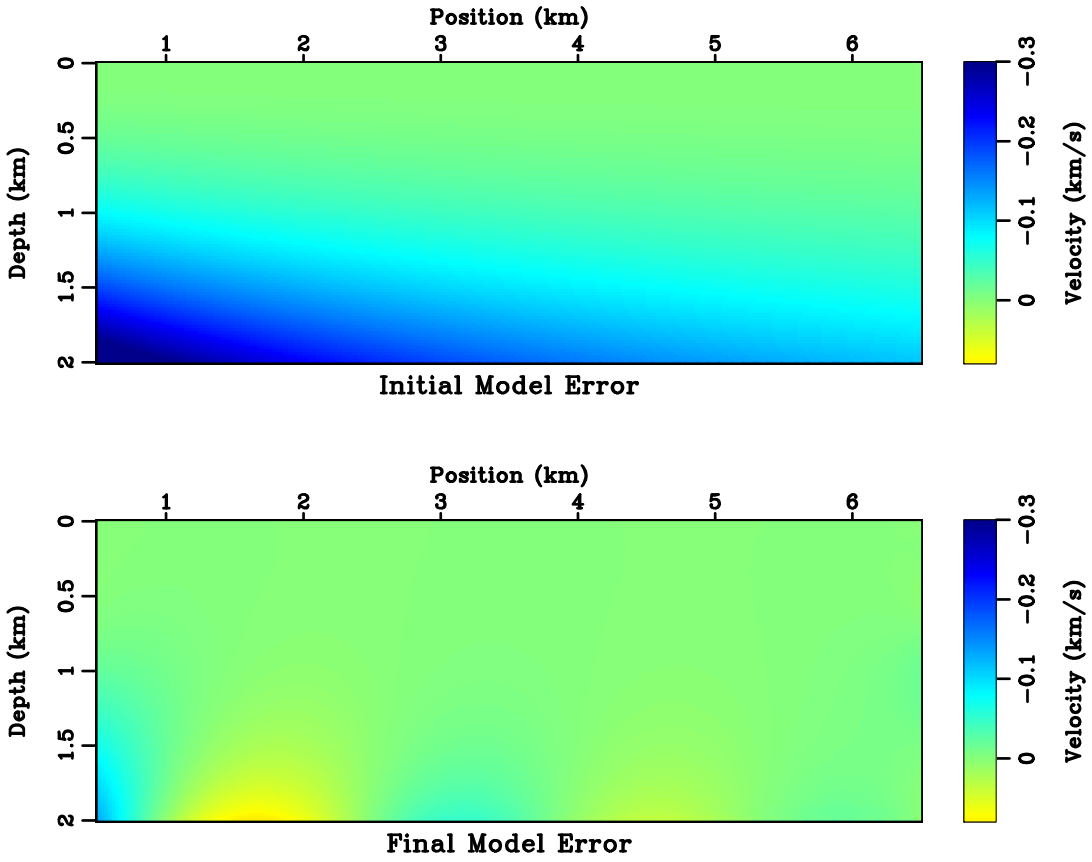


Figure 6: The difference between exact model and (top) initial model and (bottom) inverted model. The least-squares norm of model misfit is decreased from 15.6 km^2/s^2 to 2.7 km^2/s^2. time2depth2/vgrad error

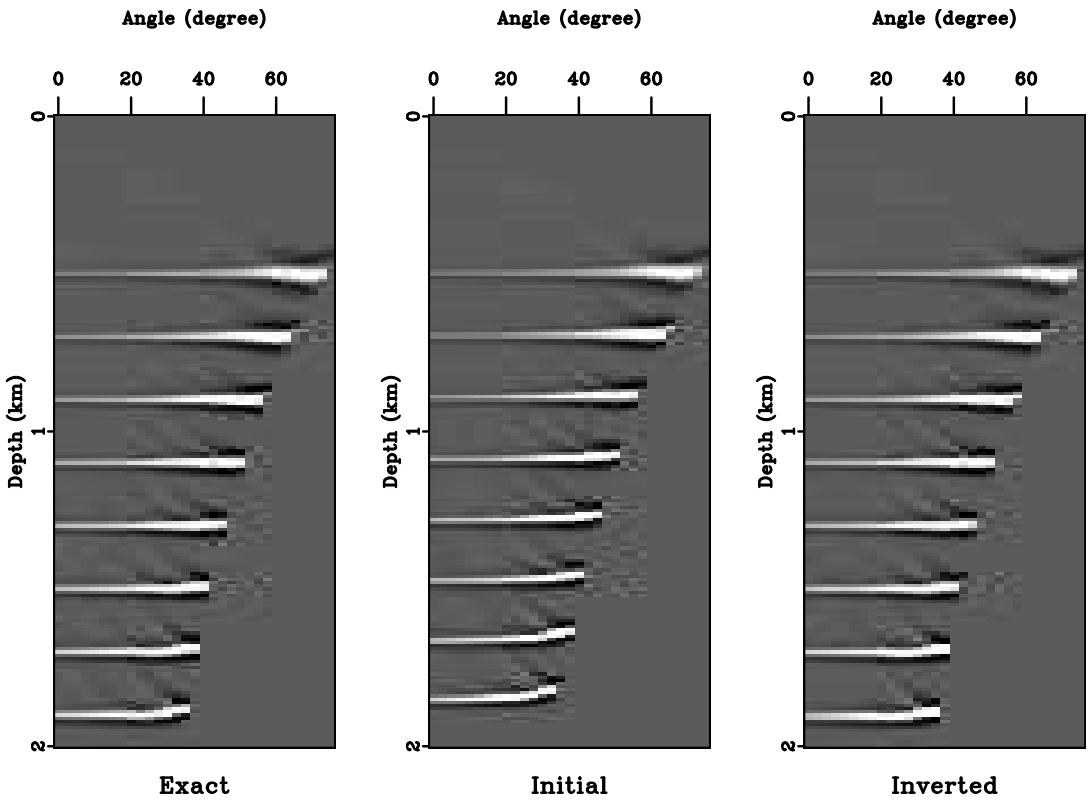


Figure 7: Comparison of the subsurface scattering-angle common-image-gathers at $x = 1.5$ km of (left) exact model, (middle) prior model and (right) inverted model. time2depth2/vgrad cigv

$$t_0(z,x) = \frac{\sqrt{2}z\left[2w_0 - 4q_x x + \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}\right]}{3\left[w_0 - 2q_x x + \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}\right]^{\frac{1}{2}}} . \tag{22}$$

The analytical expression for $v_d(t_0, x_0)$ is more complex and can be found in equations D-13 and D-14.

In Figure 8 we illustrate $x_0(z,x)$ and $t_0(z,x)$ in the model $w(z,x) = 1 - 0.052x$ s^2/km^2. To deal with the in-flow boundary issue, we apply the method described previously for the constant velocity gradient example. Unlike equation 16, equation 21 indicates varying geometrical spreadings in the domain. Figure 9 shows the corresponding analytical $Q^2(z,x)$ and $v_d(t_0, x_0)$. The geometrical spreading is significant at the lower-right corner of the domain, which translates to the cost at approximately the same location in Figure 10. We use analytical Dix velocity as the input in the inversion. Starting from the Dix-inverted model and after three linearization updates, E decreases to relative 0.0045%. The size of triangular smoother is 8 m $\times$ 20 m. The model misfit, as demonstrated in Figure 11, is also improved.

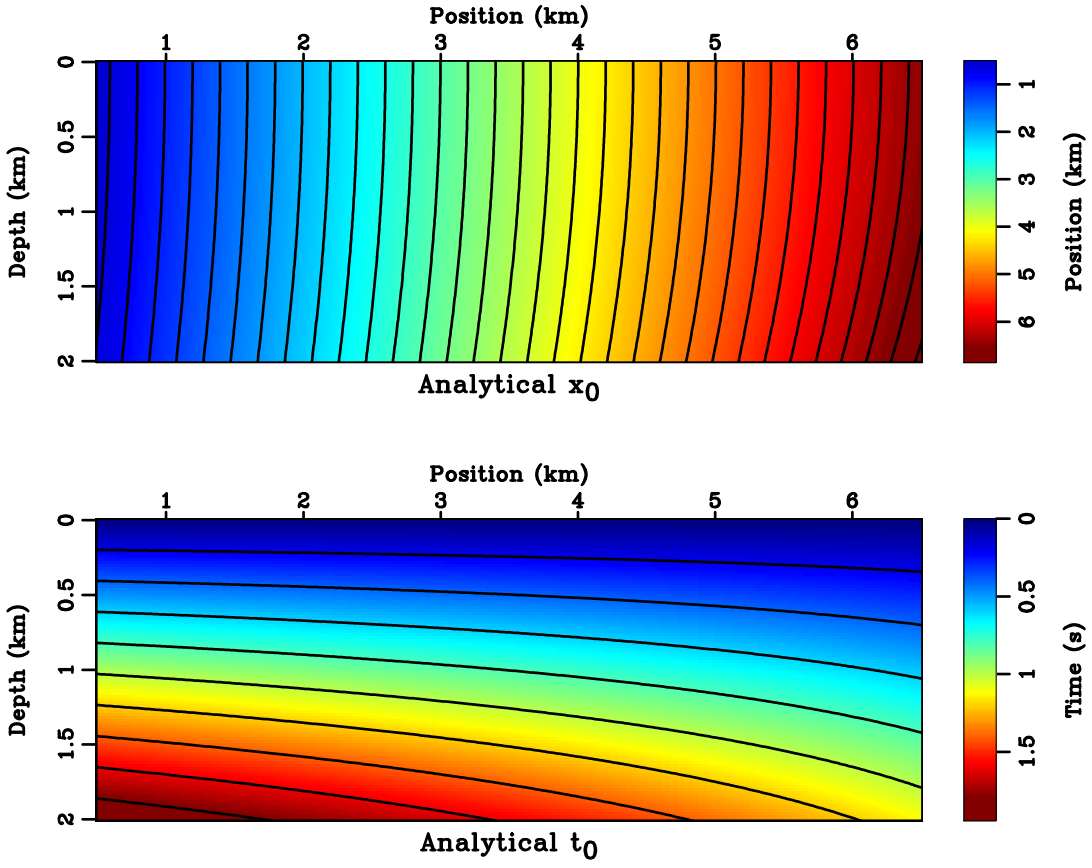


Figure 8: Analytical values of (top) t_0 and (bottom) x_0, overlaid with contour lines. time2depth2/hs2grad hs2analy

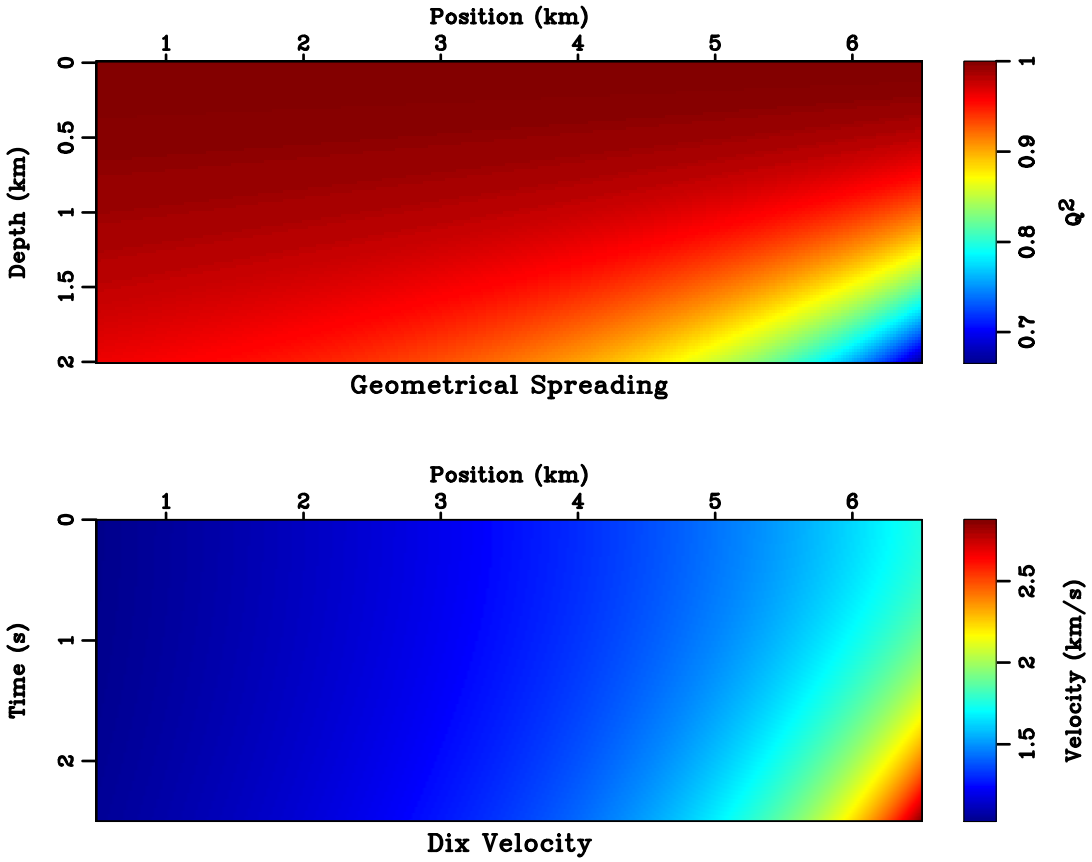


Figure 9: The (top) geometrical spreading and (bottom) Dix velocity associated with the model used in Figure 8. time2depth2/hs2grad hs2grad

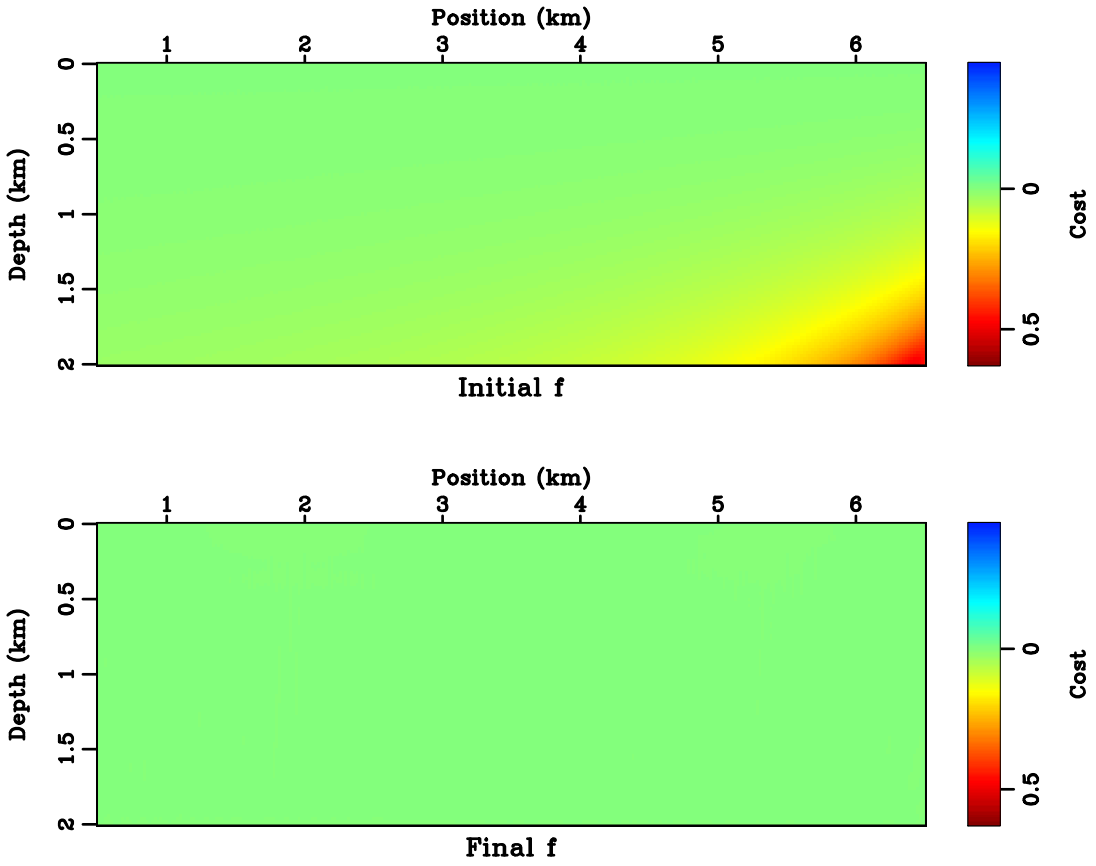


Figure 10: The cost (top) before and (bottom) after inversion. The least-squares norm of cost E is decreased from 10.431 to 0.047. time2depth2/hs2grad hs2cost

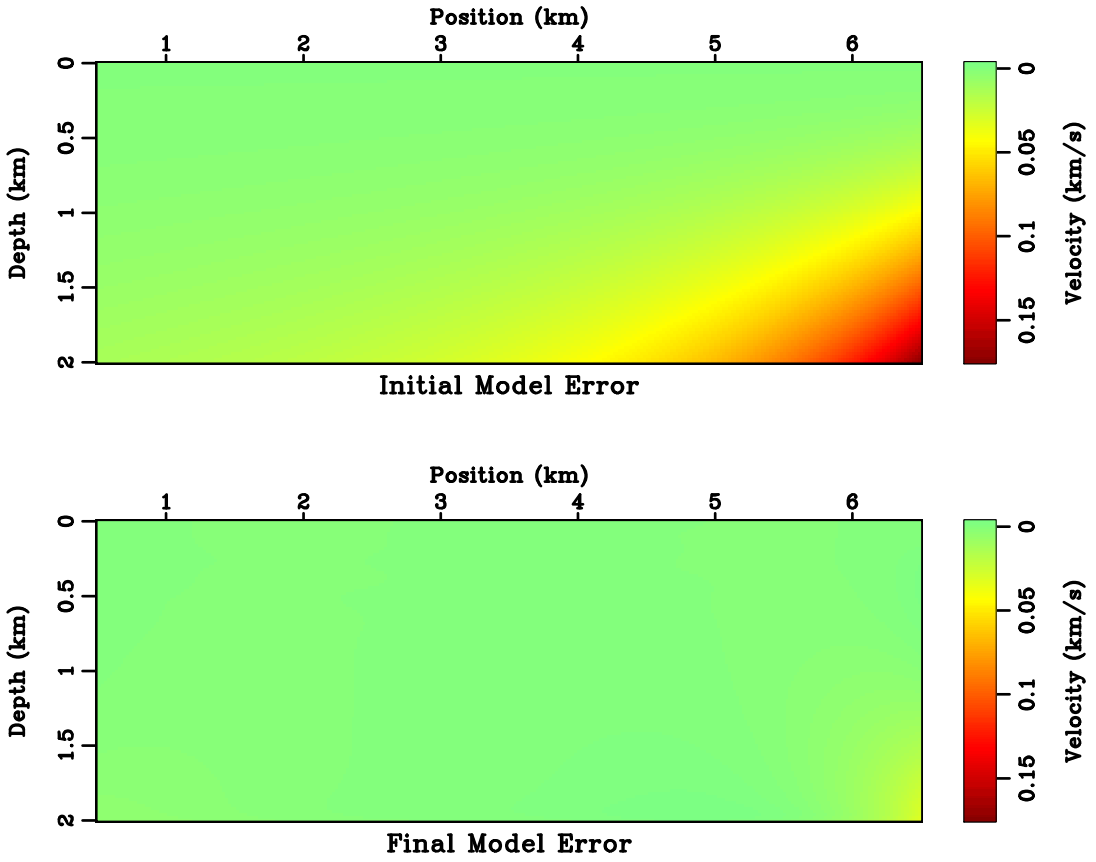


Figure 11: The difference between exact model and (top) initial model and (bottom) inverted model. The least-squares norm of model misfit is decreased from 5.0 km^2/s^2 to 0.5 km^2/s^2. time2depth2/hs2grad hs2error

Spiral model

Figure 12 shows a synthetic model borrowed from (8). The Dix inversion recovers the shallow part of the model but deteriorates quickly as geometrical spreading of image rays grows in the deeper section.

As a simple verification for the linearization process, we add a small positive velocity perturbation at location $(1, 3)$ km to the synthetic model. Comparisons between the exact and linearly predicted attributes are illustrated in Figures 13, 14 and 15. In accordance with forward modeling, where we solve firstly t_0, then x_0, and finally f, the linearization (see Appendix B) is carried out following the same sequence. First, Figure 13 justifies our upwind finite-differences implementation of the linearized eikonal equation. The positive perturbation in v in Figure 12 causes t_0 to decrease in a narrow downwind region. Next, the area affected by the perturbation in Figure 14 is wider than that in Figure 13. It also has both positive and negative amplitudes. These phenomenon are physical because image rays should bend in opposite directions in response to the perturbation. Finally, effects in cost f in Figure 15 show alternating polarities and are broader in width compared to that of dt_0 and dx_0. They indicate a complicated dependency of f on w. Note the good agreements in both shape and magnitude between exact and linearly predicted quantities in all three steps.

Because there is no analytical formula for Dix velocity in this model, we compute v_d by tracing image rays numerically in the exact model v. Also, based on Figure 12, there is no in-flow boundary other than $z = 0$. Therefore, we do not need to extend the domain as in the preceding examples. We use the Dix-inverted model as the prior model and run the inversion. It turns out that the first linearization update is sufficient for achieving the desired global minimum as shown in Figure 16.

Field data example

The field data shown in Figure 17 is from a section of Gulf of Mexico dataset (11). We estimate v_m using the method of velocity continuation (15) and convert it to v_d. Similar to the spiral model, no domain extension is needed. In Figure 18, the Dix-inverted prior model highly resembles the Dix velocity, because the Dix formula only scales the vertical axis from time to depth regardless of horizontal v_d variations. Figure 19 compares the cost before and after five linearization updates, with a 125 m $\times$ 335 m triangular smoother. In Figure 20, the l_2 norm of the cost, E, has a rapid decrease to relative 1.3%. Figure 21 illustrates the inverted model and interval velocity update.

Next, we map the time-migrated image to depth using t_0 and x_0 generated during inversion. Spline interpolation (27) is used during the coordinate mapping. We also migrate the prestack data by Kirchhoff depth migration (23) (PSDM). Figure 22 compares the time-mapped image and PSDM image of the inverted model. A good agreement between these two images justifies that time-to-depth conversion has effec-

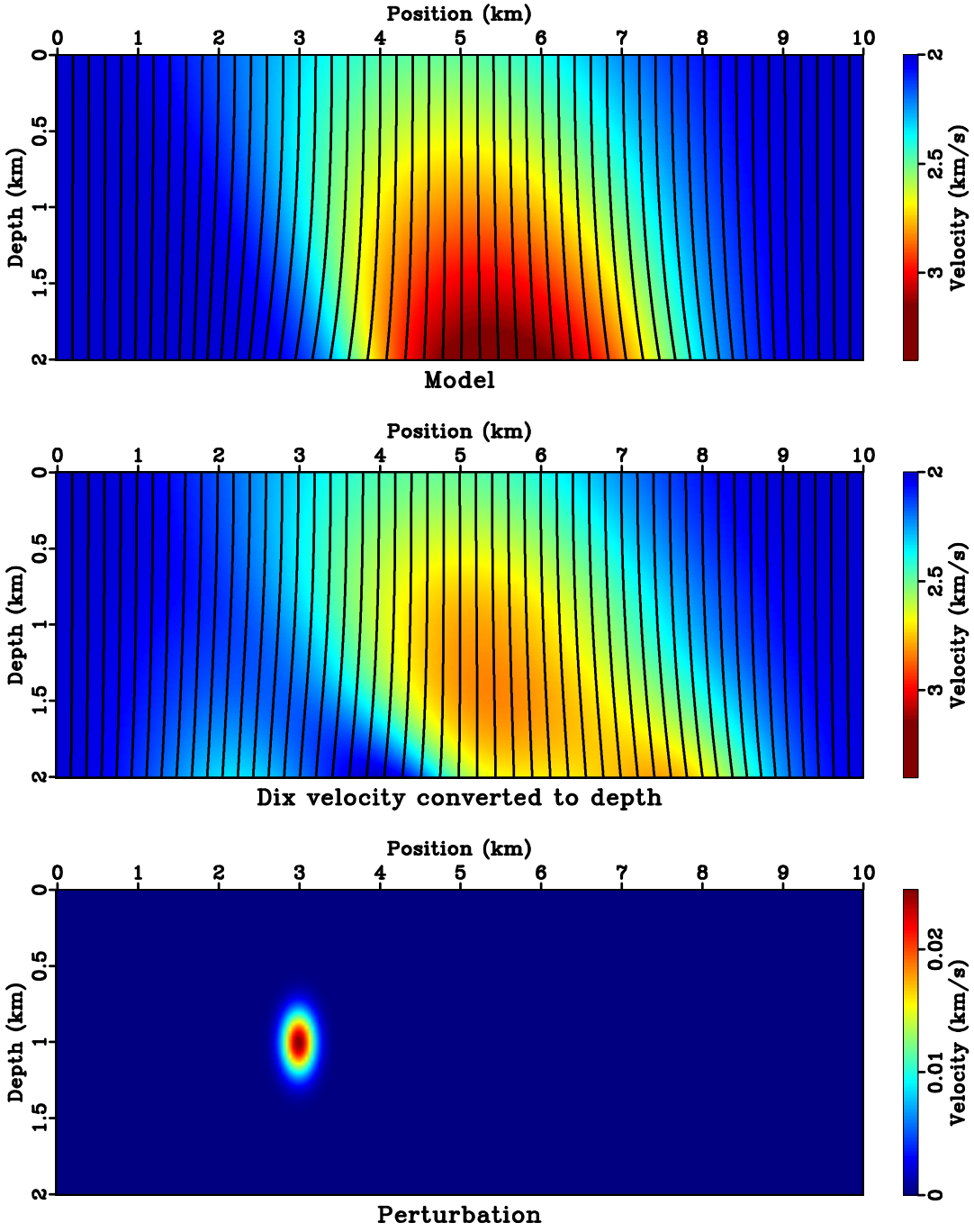


Figure 12: (Top) a synthetic model and (middle) Dix velocity converted to depth. Both overlaid with image rays. (Bottom) the model perturbation for testing linearization. time2depth2/synth vz0

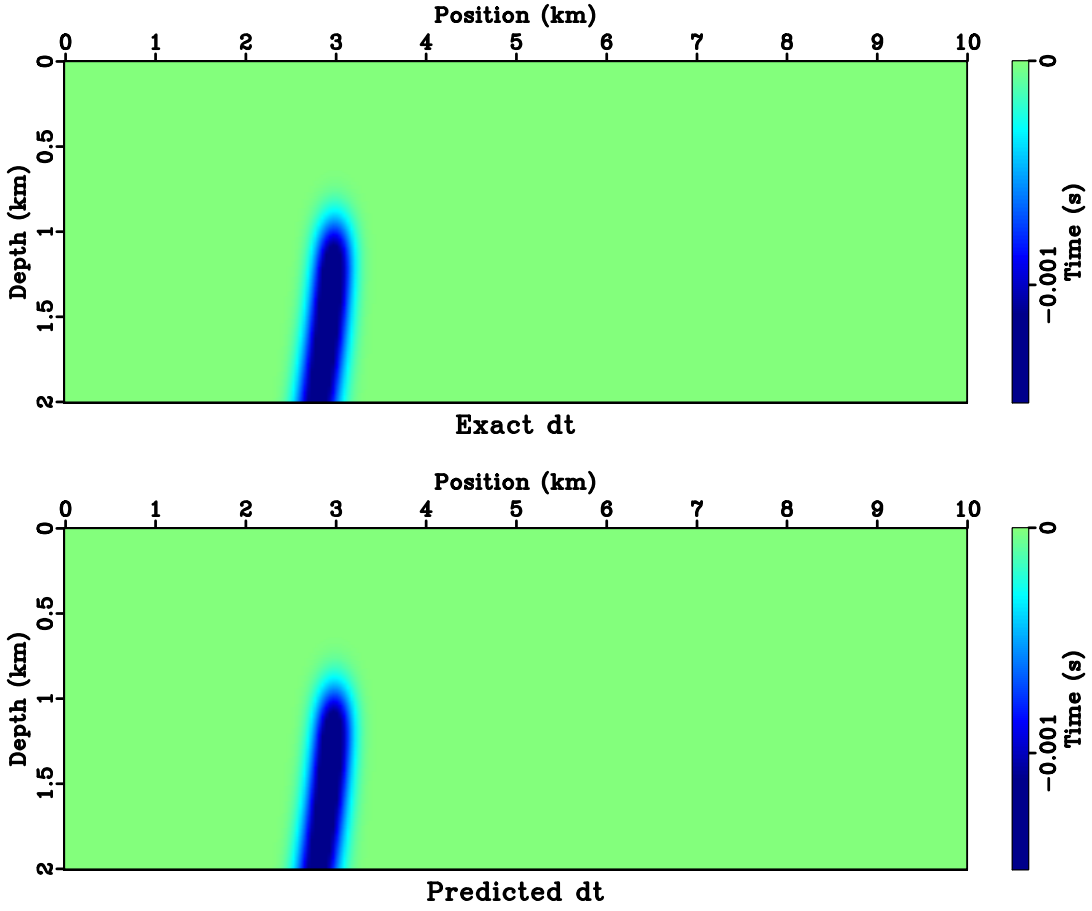


Figure 13: (Top) exact dt_0 and (bottom) linearly predicted dt_0 by equation B-4. time2depth2/synth pdt

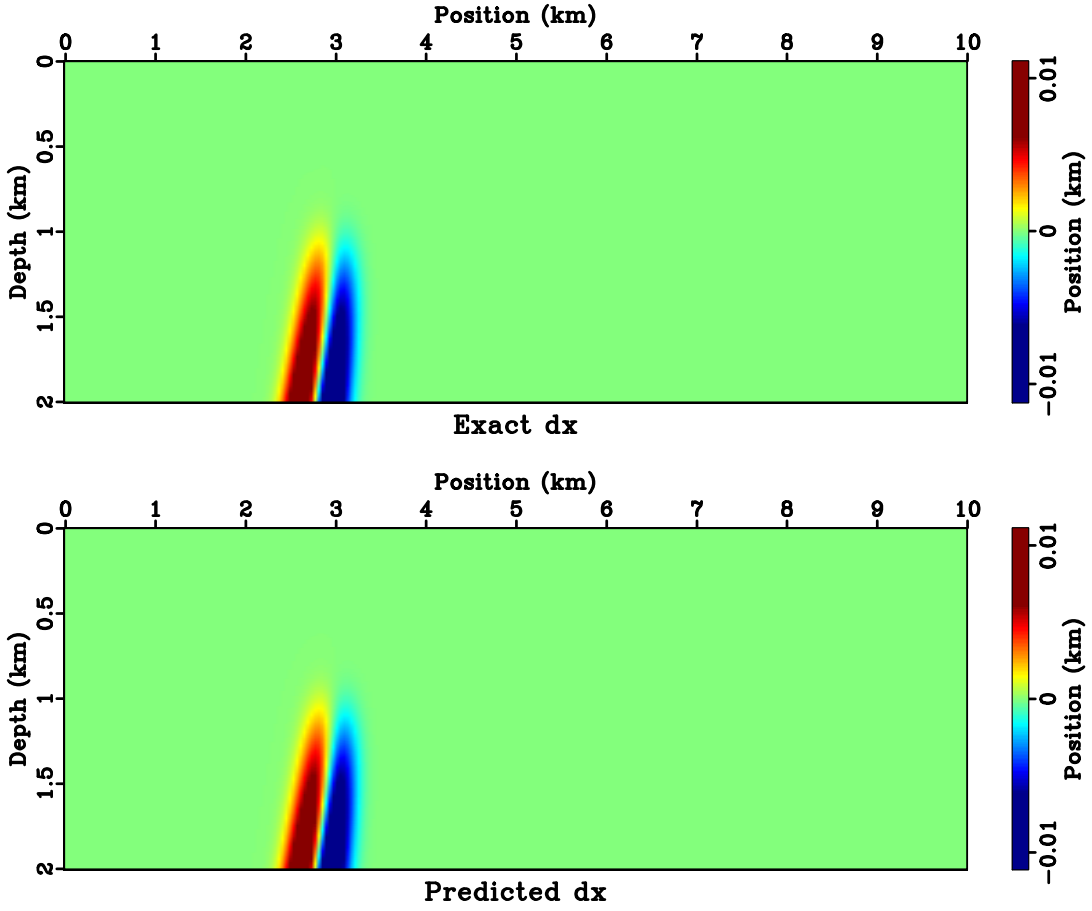


Figure 14: (Top) exact dx_0 and (bottom) linearly predicted dx_0 by equation B-3. time2depth2/synth pdx

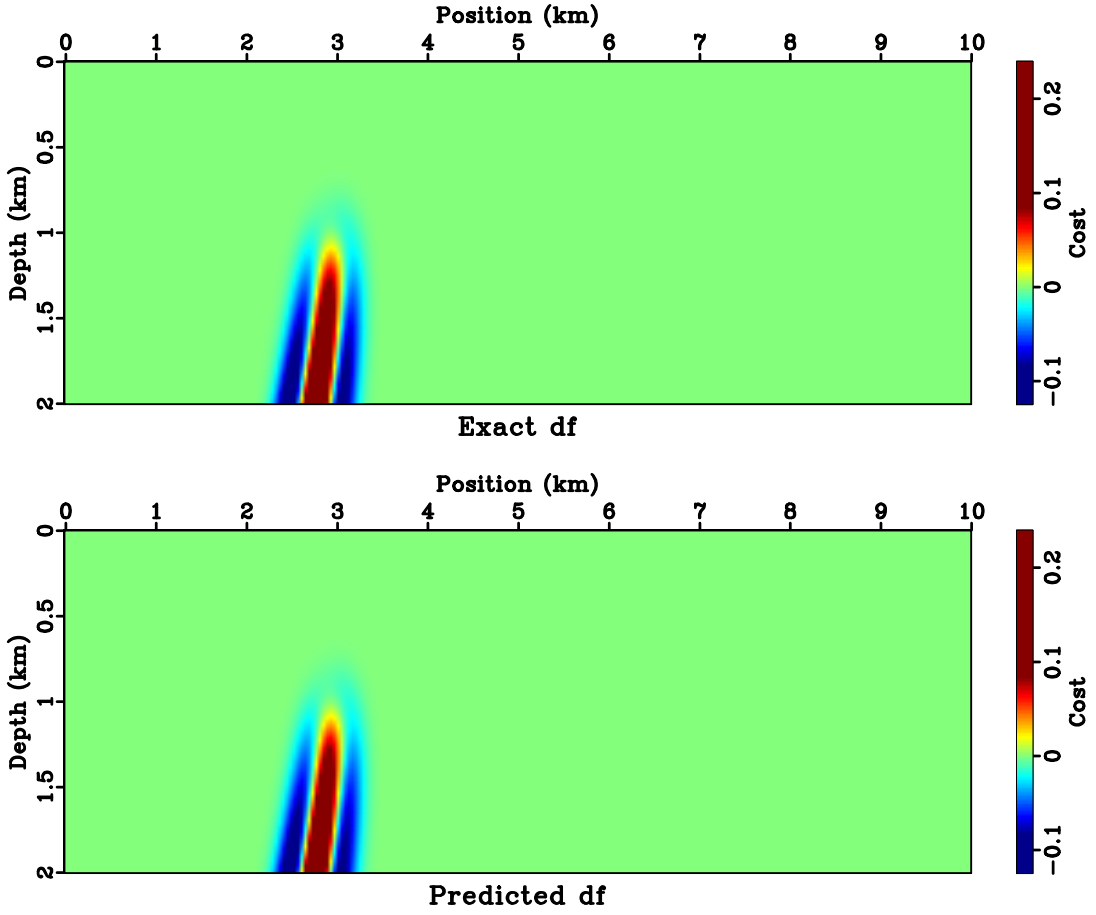


Figure 15: (Top) exact *df* and (bottom) linearly predicted *df* by equation B-5. time2depth2/synth diffcost

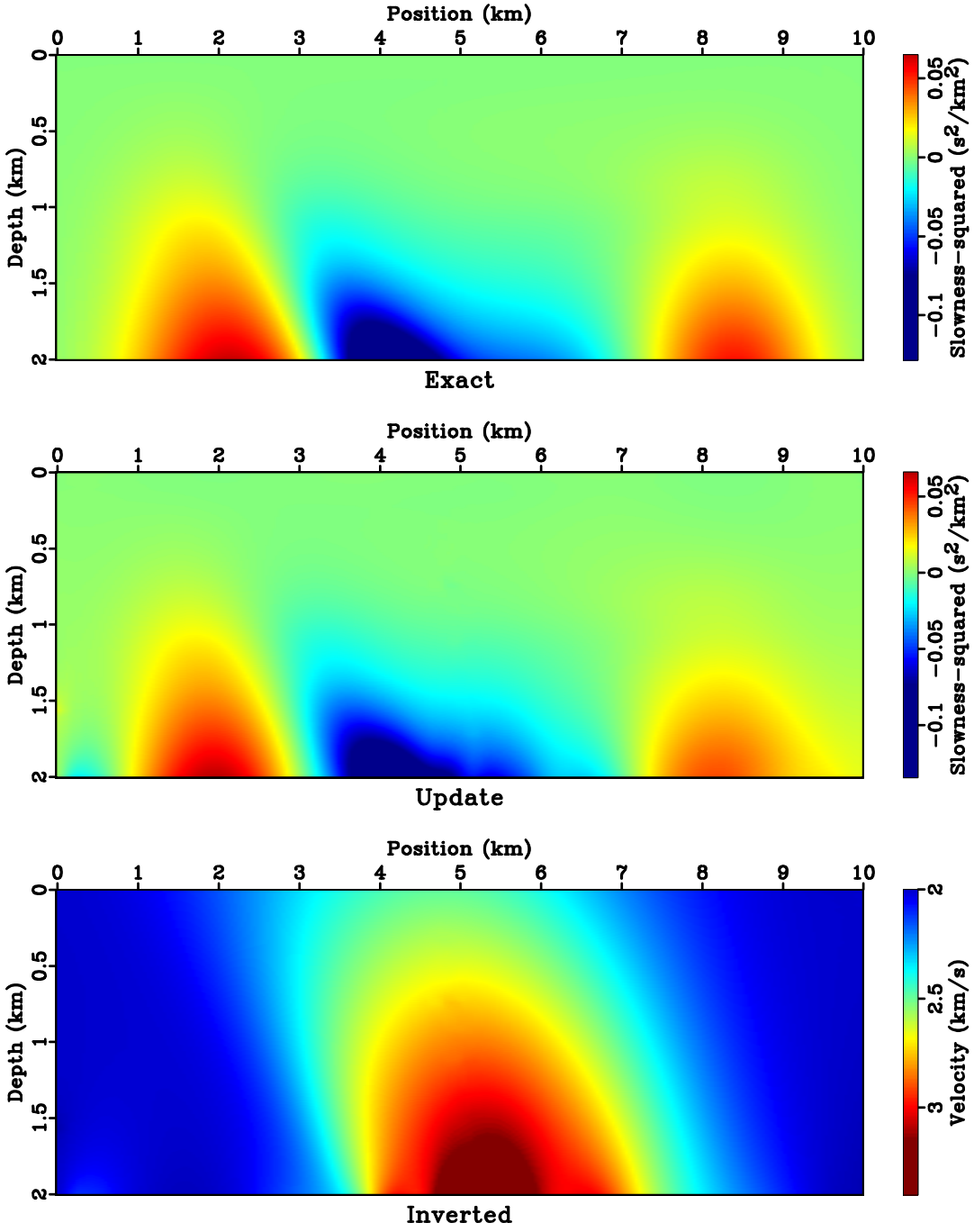


Figure 16: (Top) the exact δw and (middle) the computed δw of the first linearization step. (Bottom) the inverted interval velocity model. Compare with Figure 12. time2depth2/synth bgrad

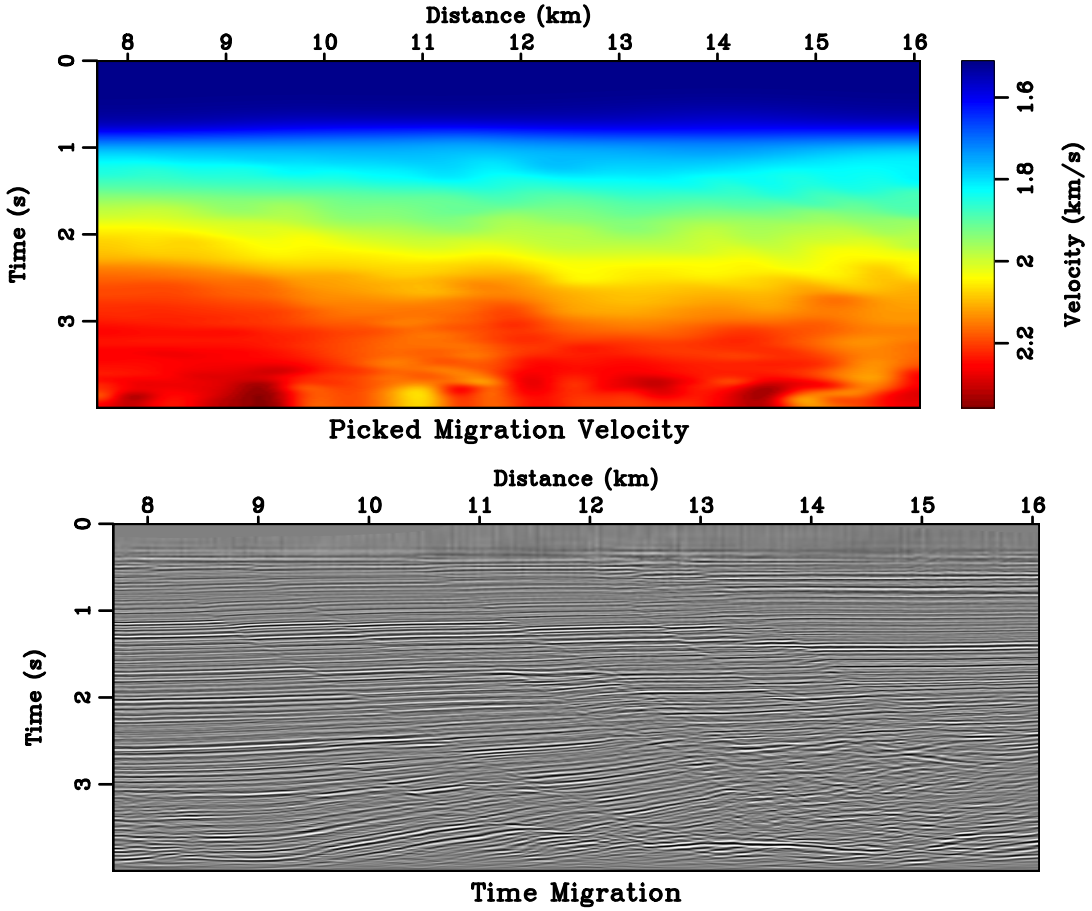


Figure 17: (Top) the estimated time-migration velocity of a section of Gulf of Mexico dataset and (bottom) the corresponding time-migrated image. time2depth2/beinew vdix

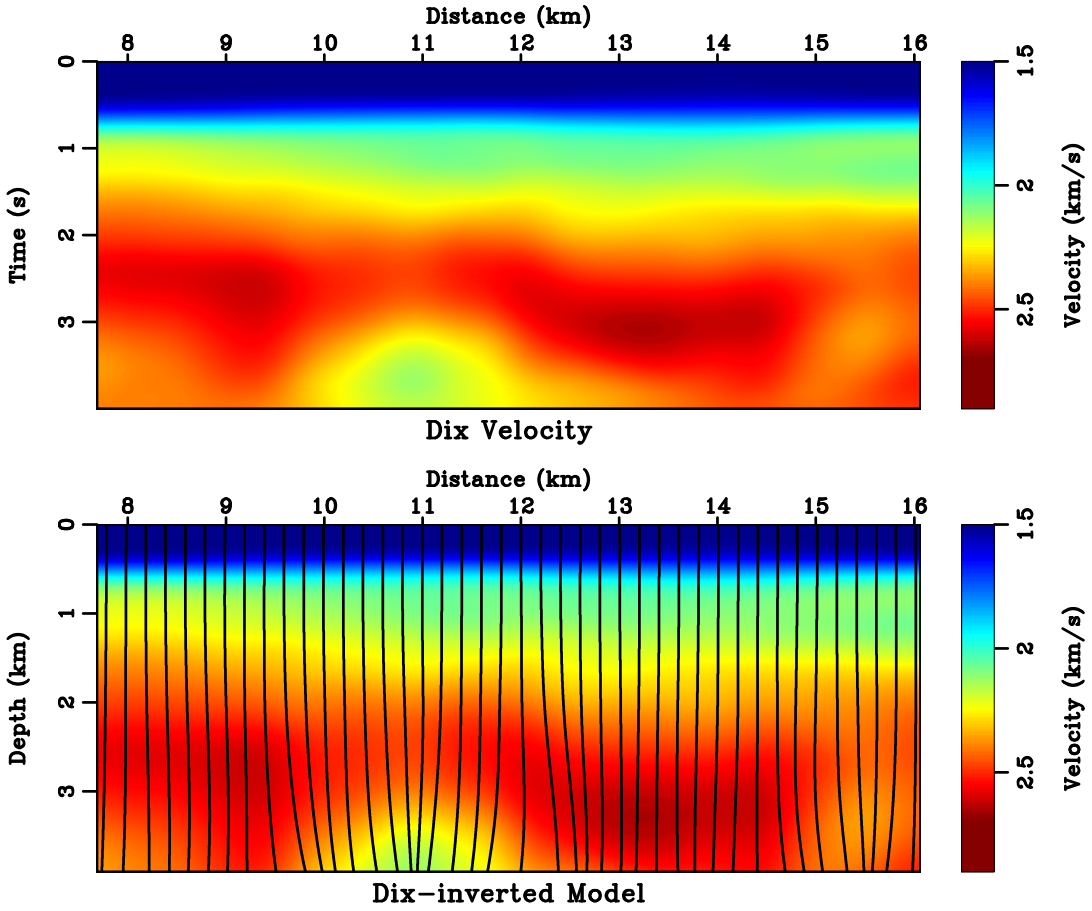


Figure 18: (Top) the Dix velocity converted from v_m in Figure 17 and (bottom) the Dix-inverted prior model for inversion, overlaid with image rays. time2depth2/beinew init

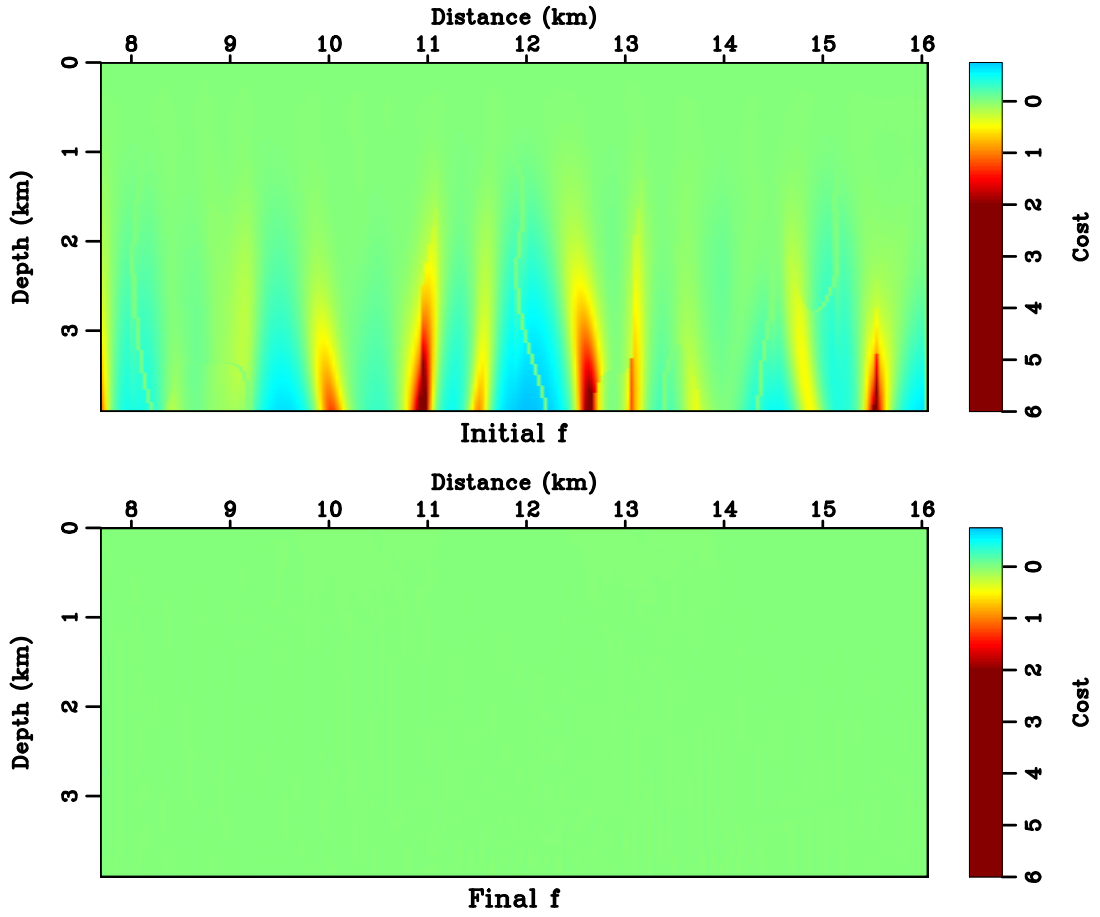


Figure 19: The cost of (top) prior model ($E = 140.25$) and (bottom) inverted model ($E = 1.81$). time2depth2/beinew inv

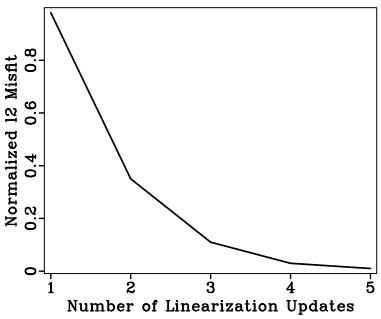


Figure 20: Convergence history of the proposed optimization-based time-to-depth conversion. time2depth2/beinew hist

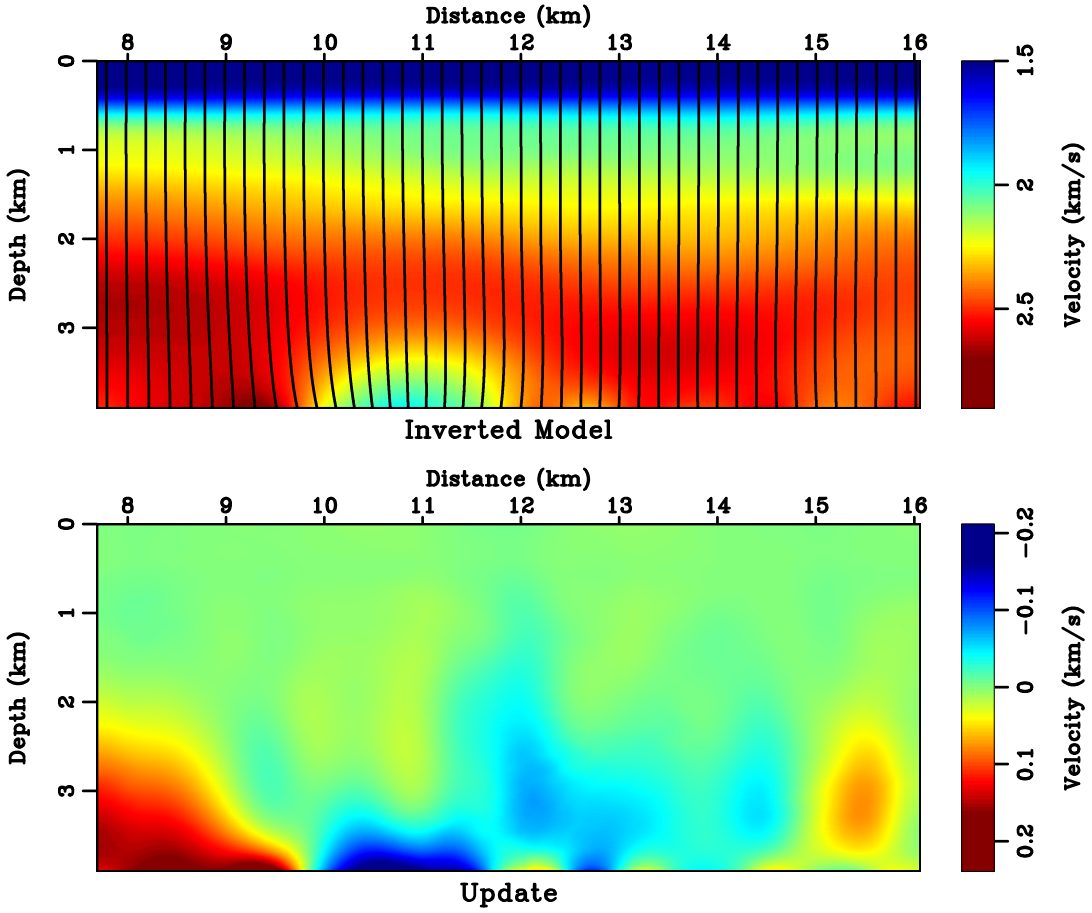


Figure 21: (Top) the inverted model, overlaid with image rays, and (bottom) its difference from the prior model in Figure 18. time2depth2/beinew dinv

tively unravelled the distorted time coordinate. Figure 23 compares PSDM images of the prior and inverted models. The velocity update in Figure 21 results in not only changes in structural dips (for example at $(3, 12)$ km) but also improved reflector continuity (for example at $(3.7, 11)$ km). Moreover, the Kirchhoff migration outputs surface offset common-image gathers. We choose two midpoint locations, $x = 11$ km and $x = 12$ km, and show their common-image gathers in Figure 24. In deeper sections, flat dashed lines are overlaid as references for the flatness of gathers. The two common-image gathers of prior model appear curved in opposite directions. After time-to-depth conversion, both gathers get flattened across the whole offset range, verifying a correct velocity update.

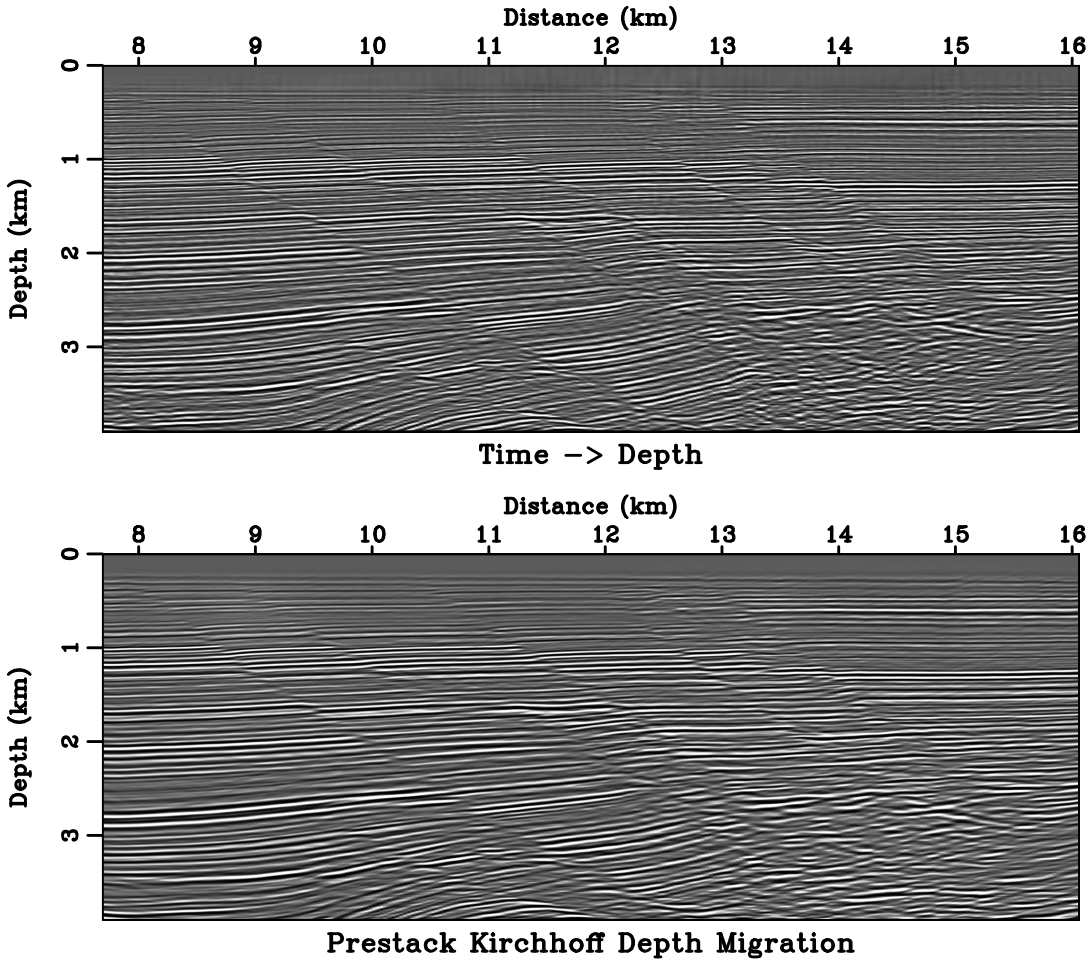


Figure 22: (Top) the time-migrated image in Figure 17 is mapped to depth using products of the time-to-depth conversion. (Bottom) PSDM image using inverted model in Figure 21. time2depth2/beinew dmig

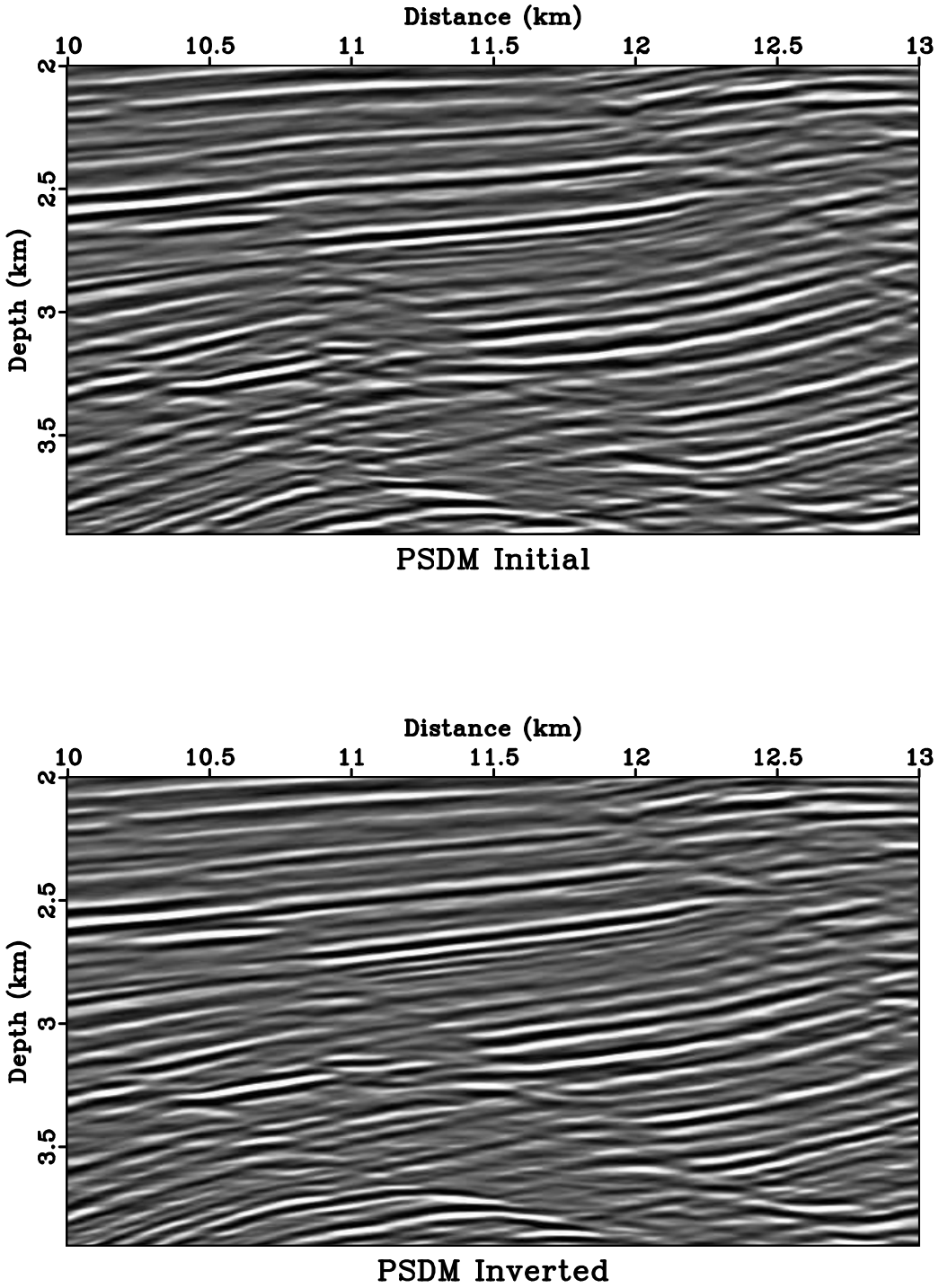


Figure 23: PSDM images of (top) the prior model and (bottom) the inverted model. Both images are plotted for the same central deep part. time2depth2/beinew ddmig0

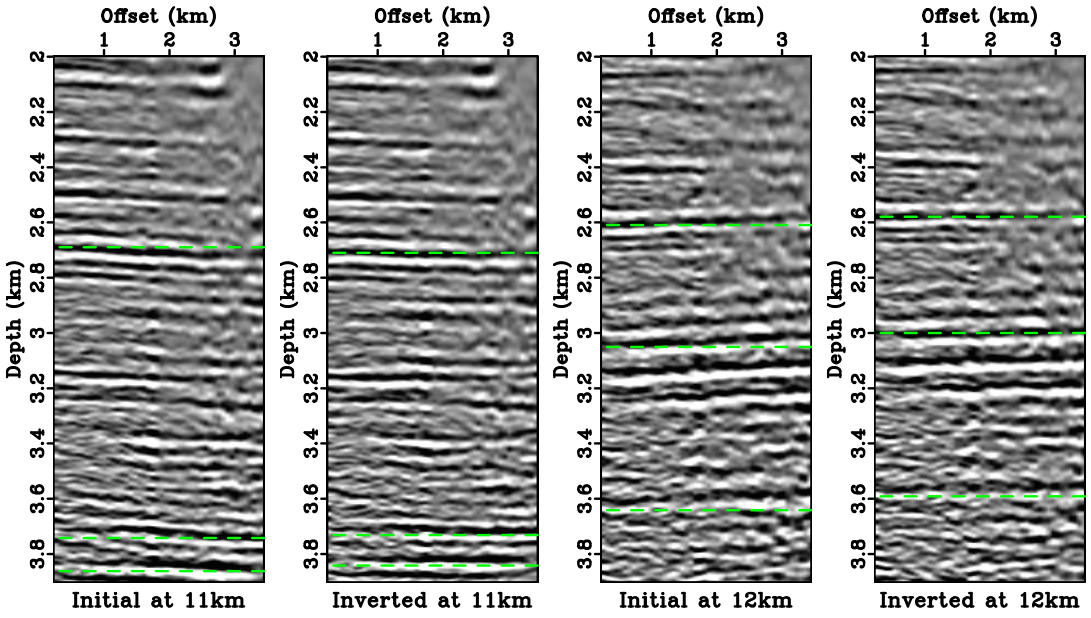


Figure 24: The surface offset common-image gathers of prior and inverted models. time2depth2/beinew cig0

DISCUSSION

Note that in our method the inversion is applied directly for slowness-squared $\mathbf{w}$ (and thus velocity), and the update $\delta\mathbf{w}$ will incorporate dependency of velocities throughout the domain, as physically honored by image rays. Previous methods (7; 8; 21) that rely on time-stepping in t_0 account for such dependencies only locally. Moreover, regularization provides a convenient and effective way of handling ill-posedness of the underlying PDEs. For these reasons, the proposed method should be numerically robust compared to the previous approaches.

A 3-D extension of the proposed method is straightforward. Instead of a scalar x_0 we need to handle both inline and crossline coordinates $\mathbf{x_0} = (x_0, y_0)$. Consequently, the geometrical spreading in 3-D becomes a matrix $\mathbf{Q}$, whose determinant is related to the generalized Dix velocity $v_d(t_0, \mathbf{x_0})$ and interval velocity $v(z, \mathbf{x})$ (7). Starting from this relationship, an iterative time-to-depth conversion can be established by following procedures similar to the ones described in this paper.

The main limitation of our approach is the failure of underlying theory at caustics, which in turn limits either the depth or the extent of lateral velocity variation of the model. A possible solution is to divide the original domain into several depth intervals, then apply velocity estimation and redatuming from one interval to another recursively (2; 23). By doing so, image-ray crossing may not develop within each interval thanks to a limited depth.

Another issue is the handling of in-flow boundaries other than the earth surface.

In the synthetic examples, we avoided this problem by limiting the interval velocity model within image ray coverage. Alternatively, we could pad the input v_d laterally. The padding should simply replicate the original boundaries. By doing so, image rays at the new left and right boundaries must run straightly downward as the media there are of $v(z)$ type. Although we could not expect the inversion to fix the in-flow boundary, the resulting errors should be local around that boundary.

Finally, in our approach time-migration velocity not only determines the prior model but also drives the inversion. Consequently, errors in time-migration velocity have a direct influence on the accuracy of inverted model. Note that usually the picking of time-migration velocity is carried out with certain smoothing. Combined with regularization in the time-to-depth conversion, the resulting interval velocity model may contain limited fine-scale features and high velocity contrasts. In this regard, we suggest our method as an efficient estimation of an initial guess for subsequent depth-imaging velocity model refinements.

CONCLUSIONS

We have introduced a novel nonlinear inversion formulation for time-to-depth velocity conversion and coordinate mapping, which involves iterative least-squares optimization constrained by a system of partial differential equations. We turn one of the governing equations that relates seismic velocity to geometrical spreading of image rays into a cost function, and linearize the other two equations: the eikonal equation and the orthogonality condition. Regularization provides extra constraints during inversion. The proposed method appears to be robust and the inversion in practice converges fast.

ACKNOWLEDGMENTS

We are grateful to Tijmen Jan Moser, Alexey Stovas, Anton Duchkov, and two anonymous reviewers for their valuable comments and suggestions. We thank BGP International for partial financial support of this research. We also thank Alexander Klokov, Alexander Vladimirsky and Changlong Wang for helpful discussions.

APPENDIX A: ILL-POSEDNESS OF THE TIME-TO-DEPTH CONVERSION PROBLEM

Let us consider the problem of solving for $z(t_0, x_0)$ and $x(t_0, x_0)$ instead of $t_0(z, x)$ and $x_0(z, x)$ by recasting equations 5, 6 and 7 in the time coordinates:

$$\left(\frac{\partial x}{\partial x_0}\right)^2 + \left(\frac{\partial z}{\partial x_0}\right)^2 = \frac{v^2(z(t_0, x_0), x(t_0, x_0))}{v_d^2(t_0, x_0)} \,, \tag{A-1}$$

$$\left(\frac{\partial x}{\partial t_0}\right)^2 + \left(\frac{\partial z}{\partial t_0}\right)^2 = v^2(z(t_0, x_0), x(t_0, x_0)) \ , \tag{A-2}$$

$$\frac{\partial x}{\partial x_0}\frac{\partial x}{\partial t_0} + \frac{\partial z}{\partial x_0}\frac{\partial z}{\partial t_0} = 0 \ . \tag{A-3}$$

The corresponding boundary conditions are

$$\begin{cases} z(0, x_0) &= 0 \ , \\ x(0, x_0) &= x_0 \ . \end{cases} \tag{A-4}$$

From equation A-3

$$\frac{\partial x}{\partial t_0} = -\frac{\partial z}{\partial x_0}\frac{\partial z}{\partial t_0} \Big/ \frac{\partial x}{\partial x_0} \ . \tag{A-5}$$

Substituting equation A-5 into equation A-2 and combining with equation A-1 produces

$$\frac{\partial z}{\partial t_0} = v_d \frac{\partial x}{\partial x_0} \ ; \tag{A-6}$$

$$\frac{\partial x}{\partial t_0} = -v_d \frac{\partial z}{\partial x_0} \ , \tag{A-7}$$

where we assume that both $\partial z/\partial t_0$ and $\partial x/\partial x_0$ remain positive and there is no caustics.

On the first glance, equations A-6 and A-7 seem suitable for numerically extrapolating $x(t_0, x_0)$ and $z(t_0, x_0)$ in t_0 direction using the boundary conditions A-4. After such an extrapolation, one would be able to reconstruct $v(t_0, x_0)$ from equation A-2 and thus solve the original problem. However, by further decoupling the system using the equivalence of the second-order mixed derivatives, we discover that the underlying PDEs are elliptic. For instance, applying $\partial/\partial x_0$ to both sides of equation A-7 results in

$$\frac{\partial^2 x}{\partial t_0 \partial x_0} = -\frac{\partial}{\partial x_0}\left(v_d \frac{\partial z}{\partial x_0}\right) \ . \tag{A-8}$$

Meanwhile, dividing by v_d and applying $\partial/\partial t_0$ to both sides of equation A-6 leads to

$$\frac{\partial}{\partial t_0}\left(\frac{1}{v_d}\frac{\partial z}{\partial t_0}\right) = \frac{\partial^2 x}{\partial x_0 \partial t_0} \ . \tag{A-9}$$

Comparing equations A-8 and A-9, we find

$$\frac{\partial}{\partial x_0}\left(v_d \frac{\partial z}{\partial x_0}\right) + \frac{\partial}{\partial t_0}\left(\frac{1}{v_d}\frac{\partial z}{\partial t_0}\right) = 0 \ . \tag{A-10}$$

Analogously,

$$\frac{\partial}{\partial x_0}\left(v_d \frac{\partial x}{\partial x_0}\right) + \frac{\partial}{\partial t_0}\left(\frac{1}{v_d}\frac{\partial x}{\partial t_0}\right) = 0 \ . \tag{A-11}$$

Solving elliptic equations A-10 and A-11 with the Cauchy-type boundary conditions A-4 is an ill-posed problem (14). A different formulation, leading to a non-linear elliptic PDE, was previously discussed by (9).

APPENDIX B: THE FRÉCHET DERIVATIVE OPERATOR

We continue the derivation of the Fréchet derivative matrix from equation 12. Using the same notations as introduced in equation 10, after discretization equation 12 becomes

$$\mathbf{J} \equiv \frac{\partial \mathbf{f}}{\partial \mathbf{w}} = 2\,\mathbf{L}_{x_0}\,\frac{\partial \mathbf{x_0}}{\partial \mathbf{w}} - 2\,\mathrm{diag}(\mathbf{v_d} \star \mathbf{w})\,\frac{\partial \mathbf{v_d}}{\partial \mathbf{w}} - \mathrm{diag}(\mathbf{v_d} \star \mathbf{v_d})\ . \tag{B-1}$$

Because v_d is in time-domain (t_0, x_0), we need to apply the chain-rule for its derivative with respect to w, i.e.,

$$\frac{\partial \mathbf{v_d}}{\partial \mathbf{w}} = \frac{\partial \mathbf{v_d}}{\partial \mathbf{t_0}}\,\frac{\partial \mathbf{t_0}}{\partial \mathbf{w}} + \frac{\partial \mathbf{v_d}}{\partial \mathbf{x_0}}\,\frac{\partial \mathbf{x_0}}{\partial \mathbf{w}}\ , \tag{B-2}$$

where $\mathbf{t_0}$ is the discretized column vector of $t_0(z, x)$. According to equation 7 and after denoting another matrix operator $\mathbf{L}_{t_0} = \nabla \mathbf{t_0} \cdot \nabla$, we find

$$\frac{\partial \mathbf{x_0}}{\partial \mathbf{w}} = -(\nabla \mathbf{t_0} \cdot \nabla)^{-1}(\nabla \mathbf{x_0} \cdot \nabla)\,\frac{\partial \mathbf{t_0}}{\partial \mathbf{w}} \equiv -\mathbf{L}_{t_0}^{-1}\mathbf{L}_{x_0}\,\frac{\partial \mathbf{t_0}}{\partial \mathbf{w}}\ . \tag{B-3}$$

Another differentiation of equation 6 leads to

$$\frac{\partial \mathbf{t_0}}{\partial \mathbf{w}} = \frac{1}{2}(\nabla \mathbf{t_0} \cdot \nabla)^{-1} \equiv \frac{1}{2}\mathbf{L}_{t_0}^{-1}\ . \tag{B-4}$$

Finally, by inserting equations B-2 through B-4 into B-1, we complete the derivation of the Fréchet derivative matrix:

$$\begin{aligned} \mathbf{J} \;&=\; -\mathbf{L}_{x_0}\mathbf{L}_{t_0}^{-1}\mathbf{L}_{x_0}\mathbf{L}_{t_0}^{-1} - \mathrm{diag}(\mathbf{v_d} \star \mathbf{w})\frac{\partial \mathbf{v_d}}{\partial \mathbf{t_0}}\mathbf{L}_{t_0}^{-1} \\ &+\; \mathrm{diag}(\mathbf{v_d} \star \mathbf{w})\frac{\partial \mathbf{v_d}}{\partial \mathbf{x_0}}\mathbf{L}_{t_0}^{-1}\mathbf{L}_{x_0}\mathbf{L}_{t_0}^{-1} - \mathrm{diag}(\mathbf{v_d} \star \mathbf{v_d})\ . \end{aligned} \tag{B-5}$$

APPENDIX C: ANALYTICAL EXPRESSIONS FOR THE CONSTANT VELOCITY GRADIENT MEDIUM

In order to derive the time-to-depth conversion analytically, we first trace image rays in the depth coordinate for $z(t_0, x_0)$ and $x(t_0, x_0)$. Then we carry out a direct inversion to find $t_0(z, x)$ and $x_0(z, x)$. The Dix velocity can be obtained at last following equations 3 and 4.

Continuing from equation 15, we write the velocity in a coordinate relative to the image ray

$$v(z, x) = v_0 + g_z z + g_x x = \tilde{v}_0 + \mathbf{g} \cdot (\mathbf{x} - \mathbf{x_0})\ , \tag{C-1}$$

where $\mathbf{g} = [g_z, g_x]^T$ and $\tilde{v}_0 = v_0 + g_x x_0$. At the starting point, image ray satisfies

$$\begin{cases} \mathbf{x_0} &= [0, x_0]^T \ , \\ \mathbf{p_0} &= [\tilde{v}_0^{-1}, 0]^T \ , \\ t_0 &= t \ . \end{cases} \tag{C-2}$$

Here we denote ray parameter $\mathbf{p} = \nabla t$ and $\mathbf{p_0}$ is the ray parameter at source. The Hamiltonian for ray tracing reads $H(\mathbf{x}, \mathbf{p}) = \mathbf{p} \cdot \mathbf{p} - v^{-2} \equiv 0$. The corresponding ray tracing system is (33):

$$\begin{cases} d\mathbf{x}/d\xi &= \mathbf{p}v^3 \ , \\ d\mathbf{p}/d\xi &= -\nabla v \ , \\ dt/d\xi &= \nabla t \cdot d\mathbf{x}/d\xi = \mathbf{p} \cdot \mathbf{p}v^3 = v \ . \end{cases} \tag{C-3}$$

Equation C-1 indicates $\nabla v = \mathbf{g}$, which means $d\mathbf{p}/d\xi$ can be integrated analytically and provides

$$\mathbf{p} = \mathbf{p_0} - \mathbf{g}\xi \ . \tag{C-4}$$

From the eikonal equation and considering $\mathbf{p_0} \cdot \mathbf{p_0} = \tilde{v}_0^{-2}$ and $g = |\mathbf{g}| = \sqrt{g_z^2 + g_x^2}$, we have

$$v = \frac{1}{\sqrt{\mathbf{p} \cdot \mathbf{p}}} = \left(\tilde{v}_0^{-2} - 2\mathbf{p_0} \cdot \mathbf{g}\xi + g^2\xi^2\right)^{-\frac{1}{2}} \ . \tag{C-5}$$

Integrating equation C-5 over ξ gives

$$t = \frac{1}{g}\text{arccosh}\left(1 + \frac{g^2\xi^2}{\tilde{v}_0^{-2} + v^{-1}\tilde{v}_0^{-1} - \mathbf{p_0} \cdot \mathbf{g}\xi}\right) \ . \tag{C-6}$$

Meanwhile, combining equations C-1 and C-3, we find $d\mathbf{p}/d\xi \cdot (\mathbf{x} - \mathbf{x_0}) + d\mathbf{x}/d\xi \cdot \mathbf{p} = \tilde{v}_0$, i.e., $\mathbf{p} \cdot (\mathbf{x} - \mathbf{x_0}) = \tilde{v}_0\xi$. Suppose

$$\mathbf{x} - \mathbf{x_0} = \alpha\mathbf{p_0} + \beta\mathbf{g} \tag{C-7}$$

then

$$\begin{cases} \mathbf{g} \cdot (\mathbf{x} - \mathbf{x_0}) &= \alpha\mathbf{p_0} \cdot \mathbf{g} + \beta g^2 = v - \tilde{v}_0 \ , \\ \mathbf{p_0} \cdot (\mathbf{x} - \mathbf{x_0}) &= \alpha\tilde{v}_0^{-2} + \beta\mathbf{p_0} \cdot \mathbf{g} = \tilde{v}_0\xi + (v - \tilde{v}_0)\xi = v\xi \ . \end{cases} \tag{C-8}$$

Solving equation C-8 provides $\alpha(\xi, v)$ and $\beta(\xi, v)$, which after substituting into equation C-7 leads to

$$\mathbf{x} = \mathbf{x_0} + \frac{(v - \tilde{v}_0)\left[\mathbf{g} - (\mathbf{p_0} \cdot \mathbf{g})\tilde{v}_0^2\mathbf{p_0}\right] + v\tilde{v}_0^2\left[g^2\mathbf{p_0} - (\mathbf{p_0} \cdot \mathbf{g})\mathbf{g}\right]\xi}{g^2 - (\mathbf{p_0} \cdot \mathbf{g})^2\tilde{v}_0^2} \ . \tag{C-9}$$

Note equation C-2 states $\mathbf{p_0} \cdot \mathbf{g} = g_z\tilde{v}_0^{-1}$ and thus equations C-4, C-6 and C-9 can be further simplified.

To connect depth- and time-domain attributes, we first invert equation C-6 such that ξ is expressed by t_0 and x_0

$$\xi(t_0, x_0) = \frac{g_z(1 - \cosh(|gt_0|)) + g\sinh(gt_0)}{g^2\tilde{v}_0^2} \ . \tag{C-10}$$

Next, we insert equations C-5 and C-10 into C-9 in order to change its parameterization from (ξ, v) to (t_0, x_0). The result is written for the z and x components of $\mathbf{x}$ separately, as follows:

$$x(t_0,x_0) = x_0 + \frac{\tilde{v}_0 g_x (1-\cosh(gt_0))}{g(g\cosh(gt_0) - g_z \sinh(gt_0))} \; , \tag{C-11}$$

$$z(t_0,x_0) = \frac{\tilde{v}_0 \left[g_z(1-\cosh(gt_0)) + g\sinh(gt_0)\right]}{g(g\cosh(gt_0) - g_z \sinh(gt_0))} \; . \tag{C-12}$$

Inverting equations C-11 and C-12 results in

$$x_0(z,x) = x + \frac{\sqrt{(v_0+g_x x)^2 + g_x^2 z^2} - (v_0 + g_x x)}{g_x} \; , \tag{C-13}$$

$$t_0(z,x) = \frac{1}{g}\text{arccosh}\left\{\frac{g^2\left[\sqrt{(v_0+g_x x)^2 + g_x^2 z^2} + g_z z\right] - v g_z^2}{v g_x^2}\right\} \; . \tag{C-14}$$

In the last step, we derive the analytical formula for the Dix velocity. Note that from equation C-13 $|\nabla x_0|^2 = 1$, i.e., there is no geometrical spreading. The image rays are circles parallel to each other. Therefore according to equation 3 $v_d = v$ and is found by combining equations C-5 and C-10

$$v_d(t_0,x_0) = \frac{(v_0 + g_x x_0) g}{g\cosh(gt_0) - g_z \sinh(gt_0)} \; . \tag{C-15}$$

The time-migration velocity v_m, on the other hand, is

$$v_m(t_0,x_0) = \frac{(v_0 + g_x x_0)^2}{t_0(g\coth(gt_0) - g_z)} \; . \tag{C-16}$$

APPENDIX D: ANALYTICAL EXPRESSIONS FOR THE CONSTANT HORIZONTAL SLOWNESS-SQUARED GRADIENT MEDIUM

In this appendix, we study the following medium:

$$w(z,x) = w_0 - 2q_x x = w_0 - 2\mathbf{q}\cdot\mathbf{x} \; , \tag{D-1}$$

where $\mathbf{q} = [0, q_x]^T$.

Similarly to the constant velocity gradient medium, it is convenient to write down the ray-tracing system in the form (33)

$$\left\{\begin{array}{lcl} d\mathbf{x}/d\sigma &=& \mathbf{p} \; , \\ d\mathbf{p}/d\sigma &=& \nabla w/2 \; , \\ dt/d\sigma &=& \nabla t \cdot d\mathbf{x}/d\sigma = \mathbf{p}\cdot\mathbf{p} \; . \end{array}\right. \tag{D-2}$$

Given equation D-1, $\nabla w = -2\mathbf{q}$ and thus $d\mathbf{p}/d\sigma = -\mathbf{q}$. After integration over σ, equation D-2 becomes

$$\left\{ \begin{array}{lcl} \mathbf{x} & = & \mathbf{x_0} + \mathbf{p_0}\sigma - \mathbf{q}\sigma^2/2 \ , \\ \mathbf{p} & = & \mathbf{p_0} - \mathbf{q}\sigma \ , \\ t & = & |\mathbf{p_0}|^2\sigma - \mathbf{p_0}\cdot\mathbf{q}\sigma^2 + |\mathbf{q}|^2\sigma^3/3 \ . \end{array} \right. \tag{D-3}$$

For a particular image ray

$$\left\{ \begin{array}{lcl} \mathbf{x_0} & = & [0, x_0]^T \ , \\ \mathbf{p_0} & = & [\sqrt{w_0 - 2q_x x_0}, 0]^T \ , \\ t_0 & = & t \ , \end{array} \right. \tag{D-4}$$

the equation for $\mathbf{x}$ in D-3 simplifies to

$$\left\{ \begin{array}{l} x = x_0 - q_x\sigma^2/2 \ , \\ z = \sigma\sqrt{w_0 - 2q_x x_0} \ . \end{array} \right. \tag{D-5}$$

Solving equation D-5 for σ as a function of z and x

$$\sigma(z,x) = \left[\frac{(w_0 - 2q_x x) - \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}}{2q_x^2} \right]^{\frac{1}{2}} \ . \tag{D-6}$$

Combining equations D-3 through D-6, we find

$$\begin{array}{rcl} x_0(z,x) & = & x + \dfrac{1}{2}q_x\sigma^2 \\ & = & \dfrac{2w_0 x + q_x z^2}{w_0 + 2q_x x + \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}} \ , \end{array} \tag{D-7}$$

$$\begin{array}{rcl} t_0(z,x) & = & (w_0 - 2q_x x_0)\sigma + \dfrac{1}{3}q_x^2\sigma^3 \\ & = & \dfrac{\sqrt{2}z\left[2w_0 - 4q_x x + \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}\right]}{3\left[w_0 - 2q_x x + \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}\right]^{\frac{1}{2}}} \ . \end{array} \tag{D-8}$$

According to equation 4, D-7 can give rise to the geometrical spreading:

$$Q^2(z,x) = \frac{2[(w_0 - 2q_x x)^2 - 4q_x^2 z^2]}{(w_0 - 2q_x x)\left[w_0 - 2q_x x + \sqrt{(w_0 - 2q_x x)^2 - 4q_x^2 z^2}\right]} \ . \tag{D-9}$$

It is more convenient to express equations D-1 and D-9 in σ and x_0 instead of directly in t_0 and x_0:

$$w(\sigma, x_0) = w_0 - 2q_x x_0 + q_x^2\sigma^2 \ , \tag{D-10}$$

$$Q^2(\sigma, x_0) = 1 + q_x^2\sigma^2 \left(\frac{1}{w_0 - 2q_x x_0} - \frac{4}{w_0 - 2q_x x_0 + q_x^2\sigma^2} \right) , \tag{D-11}$$

where we must resolve $\sigma = \sigma(t_0, x_0)$. This is done by revisiting equation D-8. For given t_0 and x_0, σ is the root of a depressed cubic function of the following form:

$$\sigma^3 + \frac{3(w_0 - 2q_x x_0)}{q_x^2}\sigma - \frac{3}{q_x^2}t_0 = 0 . \tag{D-12}$$

After some algebraic manipulations, we find

$$\begin{aligned} \sigma(t_0, x_0) &= \left[\frac{3t_0 + \sqrt{9t_0^2 + 4(w_0 - 2q_x x_0)^3/q_x^2}}{2q_x^2} \right]^{\frac{1}{3}} \\ &- \frac{w_0 - 2q_x x_0}{q_x} \left[\frac{2}{3q_x t_0 + \sqrt{9q_x^2 t_0^2 + 4(w_0 - 2q_x x_0)^3}} \right]^{\frac{1}{3}} . \end{aligned} \tag{D-13}$$

Finally, inserting equations D-10 and D-11 into equation 3 results in the Dix velocity:

$$v_d(t_0, x_0) = \frac{\sqrt{w_0 - 2q_x x_0}}{w_0 - 2q_x x_0 - q_x^2\sigma^2} . \tag{D-14}$$

REFERENCES

[1] Bartel, D., M. Busby, J. Nealon, and J. Zaske, 2006, Time to depth conversion and uncertainty assessment using average velocity modeling: 66th Annual International Meetings, SEG Expanded Abstracts, 2166–2170.

[2] Bevc, D., 1997, Imaging complex structures with semirecursive Kirchhoff migration: Geophysics, **62**, 577–588.

[3] Bevc, D., J. L. Black, and G. Palacharla, 1995, Plumes: Response of time migration to lateral velocity variation: Geophysics, **60**, 1118–1127.

[4] Bishop, T. N., K. P. Bube, R. T. Cutler, R. T. Langan, P. L. Love, J. R. Resnick, R. T. Shuey, D. A. Spindler, and H. W. Wyld, 1985, Tomographic determination of velocity and depth in laterally varying media: Geophysics, **50**, 903–923.

[5] Björck, A., 1996, Numerical methods for least squares problems: SIAM.

[6] Black, J. L., and M. A. Brzostowski, 1994, Systematics of time migration errors: Geophysics, **59**, 1419–1434.

[7] Cameron, M., S. Fomel, and J. Sethian, 2007, Seismic velocity estimation from time migration: Inverse Problems, **23**, 1329–1369.

[8] ———, 2008, Time-to-depth conversion and seismic velocity estimation using time-migration velocity: Geophysics, **73**, no. 5, VE205–VE210.

[9] ———, 2009, Analysis and algorithms for a regularized Cauchy problem arising from a nonlinear elliptic PDE for seismic velocity estimation: Journal of Computational Physics, **228**, 7388–7411.

[10] Chapman, C., 2004, Fundamentals of seismic wave propagation: Cambridge University Press.
[11] Claerbout, J. F., 1996, Basic earth imaging: http://sepwww.stanford.edu/sep/prof/bei10.2008.pdf.
[12] Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68–86.
[13] Engl, H. W., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.
[14] Evans, L. C., 2010, Partial differential equations: American Mathematical Society.
[15] Fomel, S., 2003, Time-migration velocity analysis by velocity continuation: Geophysics, **68**, 1662–1672.
[16] ———, 2007, Shaping regularization in geophysical estimation problems: Geophysics, **72**, no. 2, R29–R36.
[17] Franklin, J. B., and J. M. Harris, 2001, A fast marching level set method for monotonically advancing fronts: Journal of Computational Acoustics, **9**, 1095–1109.
[18] Haddon, R. A. W., and P. W. Buchen, 1981, Use of Kirchhoff's formula for body wave calculations in the earth: Geophysical Journal of the Royal Astronomical Society, **67**, 587–598.
[19] Hestenes, M. R., and E. Stiefel, 1952, Method of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**, 409–436.
[20] Hubral, P., 1977, Time migration—some ray theoretical aspects: Geophysical Prospecting, **25**, 738–745.
[21] Iversen, E., and M. Tygel, 2008, Image-ray tracing for joint 3D seismic velocity estimation and time-to-depth conversion: Geophysics, **73**, S99–S114.
[22] Larner, K. L., L. Hatton, B. S. Gibson, and I. S. Hsu, 1981, Depth migration of imaged time sections: Geophysics, **46**, 734–750.
[23] Li, S., and S. Fomel, 2013, Kirchhoff migration using eikonal-based computation of traveltime source-derivatives: Geophysics, **78**, no. 4, S211–S219.
[24] Li, S., S. Fomel, and A. Vladimirsky, 2011, Improving wave-equation fidelity of Gaussian beams by solving the complex eikonal equation: 71st Annual International Meetings, SEG Expanded Abstracts, 3829–3834.
[25] Li, S., A. Vladimirsky, and S. Fomel, 2013, First-break traveltime tomography with the double-square-root eikonal equation: Geophysics, **78**, no. 6, U89–U101.
[26] Popov, M. M., 2002, Ray theory and Gaussian beam method for geophysicists: EDUFBA.
[27] Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2007, Numerical recipes: The art of scientific computing: Cambridge University Press.
[28] Robein, E., 2003, Velocities, time-imaging and depth-imaging in reflection seismics: EAGE.
[29] Sethian, J., 1999, Fast marching methods: SIAM Review, **41**, 199–235.
[30] Sethian, J. A., and A. M. Popovici, 1999, 3-D imaging using higher order fast marching traveltimes: Geophysics, **64**, 516–523.

[31] Stork, C., and R. W. Clayton, 1992, Using constraints to address the instabilities of automated prestack velocity analysis: Geophysics, **57**, 404–419.
[32] Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Sovet Math. Dokl., 1035–1038.
[33] Červený, V., 2001, Seismic ray theory: Cambridge University Press.
[34] Xu, S., H. Chauris, G. Lambaré, and M. Noble, 2001, Common-angle migration: A strategy for imaging complex media: Geophysics, **66**, 1877–1894.
[35] Yilmaz, O., 2001, Seismic data analysis: SEG.
[36] Yilmaz, O., I. Tanir, and C. Gregory, 2001, A unified 3-d seismic workflow: Geophysics, **66**, 1699–1713.
[37] Zhdanov, M., 2002, Geophysical inverse theory and regularization problems: Elsevier Science.

Predictive painting of 3-D seismic volumes

*Sergey Fomel**

ABSTRACT

Structural information is the most important content of seismic images. I introduce a numerical algorithm for spreading information in 3-D volumes according to the local structure of seismic events. The algorithm consists of two steps. First, local spatially-variable inline and crossline slopes of seismic events are estimated by the plane-wave-destruction method. Next, a seed trace is inserted in the volume, and the information contained in that trace is spread inside the volume, thus automatically "painting" the data space. Immediate applications of this technique include automatic horizon picking and flattening in applications to both prestack and post-stack seismic data analysis. Synthetic and field data tests demonstrate the effectiveness of predictive painting.

INTRODUCTION

Structural information is the most important content of seismic images. One way to characterize structure is to assign a dominant local slope attribute to all elements in a volume. (4) proposed the method of plane-wave destruction for detecting local slopes of seismic events. Closely related ideas were developed in the differential semblance optimization framework (21; 17). Plane-wave destruction finds many important applications in seismic data analysis, including data regularization, noise attenuation, and velocity-independent imaging (10; 12; 13; 3).

The main principle of plane-wave destruction is prediction: each seismic trace gets predicted from its neighbors that are shifted along the event slopes, and the prediction error gets minimized to estimate optimal slopes. In this paper, I propose to extract the prediction operator from the plane-wave destruction process and to use it for recursive spreading of information inside the seismic volume. I call this spreading *predictive painting*.

One particular kind of information that becomes meaningful when spread in a volume is *relative geologic age*, in the terminology of (20): seismic layers arranged according to the relative age of sedimentation. Once relative geological age is established everywhere in the volume, it is possible to flatten seismic images by extracting *stratal slices* (22) without manual picking of horizons. Even though flattened seismic horizons do not necessarily correspond to equivalent true geologic ages, flattening improves the interpreter's ability to understand and quantify the structural architecture

*e-mail: sergey.fomel@beg.utexas.edu

of sedimentary layers (23). The idea of using local shifts for automatic picking was introduced by (1) and (18). (19) presents an alternative approach involving instantaneous phase unwrapping. Analogous techniques are implemented by (7; 2).

Flattening and automatic picking of horizons are important not only for final structural interpretation but also for prestack imaging and data analysis and for extracting prestack amplitude attributes. The idea of prestack gather flattening using local cross-correlations was developed previously by a number of authors (16; 9; 14; 15).

The predictive painting method, introduced in this paper, provides a new approach for extracting and applying structural patterns, with superior computational performance. The advantages of the proposed method include both conceptual simplicity and computational efficiency. In the next sections, I describe the basic algorithm for automatic painting and demonstrate its performance with synthetic and field data examples.

DESTRUCTION AND PREDICTION OF PLANE WAVES

Plane-wave destruction originates from a local plane-wave model for characterizing seismic data (10). The mathematical basis is the local plane differential equation

$$\frac{\partial P}{\partial x} + \sigma \frac{\partial P}{\partial t} = 0 \;, \tag{1}$$

where $P(t,x)$ is the wave field and σ is the local slope, which may also depend on t and x (4). In the case of a constant slope, equation A-1 has the simple general solution

$$P(t,x) = f(t - \sigma x) \;, \tag{2}$$

where $f(t)$ is an arbitrary waveform. Equation A-2 is nothing more than a mathematical description of a plane wave. Assuming that the slope $\sigma(t,x)$ varies in time and space, one can design a local operator to propagate each trace to its neighbors.

Let $\mathbf{s}$ represent a seismic section as a collection of traces: $\mathbf{s} = [\mathbf{s}_1\ \mathbf{s}_2\ \ldots\ \mathbf{s}_N]^T$, where $\mathbf{s}_k$ corresponds to $P(t,x_k)$ for $k = 1, 2, \ldots$ A plane-wave destruction operator (10) effectively predicts each trace from its neighbor and subtracts the prediction from the original trace. In the linear operator notation, the plane-wave destruction operation can be defined as

$$\mathbf{r} = \mathbf{D}\,\mathbf{s} \;, \tag{3}$$

where $\mathbf{r}$ is the destruction residual, and $\mathbf{D}$ is the destruction operator defined as

$$\mathbf{D} = \begin{bmatrix} \mathbf{I} & 0 & 0 & \cdots & 0 \\ -\mathbf{P}_{1,2} & \mathbf{I} & 0 & \cdots & 0 \\ 0 & -\mathbf{P}_{2,3} & \mathbf{I} & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & -\mathbf{P}_{N-1,N} & \mathbf{I} \end{bmatrix} \;, \tag{4}$$

where $\mathbf{I}$ stands for the identity operator, and $\mathbf{P}_{i,j}$ describes prediction of trace j from trace i. Prediction of a trace consists of shifting the original trace along dominant event slopes. The prediction operator is a numerical solution of equation A-1 for local plane wave propagation in the x direction. The dominant slopes are estimated by minimizing the prediction residual $\mathbf{r}$ using regularized least-squares optimization. I employ shaping regularization (11) for controlling the smoothness of the estimated slope fields. In the 3-D case, a pair of inline and crossline slopes, $\sigma_x(t,x,y)$ and $\sigma_y(t,x,y)$, and a pair of destruction operators, $\mathbf{D}_x$ and $\mathbf{D}_y$, are required to characterize the 3-D structure. Each prediction in 3-D occurs in either inline or crossline direction and thus conforms to equation 4. However, as explained below in the discussion of Dijkstra's algorithm, it is possible to arrange all 3-D traces in a sequence for further processing.

Prediction of a trace from a distant neighbor can be accomplished by simple recursion. Predicting trace k from trace 1 is simply

$$\mathbf{P}_{1,k} = \mathbf{P}_{k-1,k} \cdots \mathbf{P}_{2,3}\,\mathbf{P}_{1,2}\ . \tag{5}$$

If $\mathbf{s}_r$ is a reference trace, then the prediction of trace $\mathbf{s}_k$ is $\mathbf{P}_{r,k}\,\mathbf{s}_r$. I call the recursive operator $\mathbf{P}_{r,k}$ *predictive painting*. Once the elementary prediction operators in equation 4 are determined by plane-wave destruction, predictive painting can spread information from a given trace to its neighbors recursively by following the local structure of seismic events. The next section illustrates the painting concept using 2-D examples.

PREDICTIVE PAINTING IN 2-D

The input for the first example is borrowed from (5) and shown in Figure 2(a). It is a synthetic seismic image containing dipping beds, an uncomformity, and a fault. Figure 1(b) shows local dips measured by the plane-wave destruction algorithm. The slope estimate correctly depicts the constant dip in the top part of the image and the sinusoidal variation of the dip in the bottom. Figure 2(a) shows the output of predictive painting: I assign the reference trace, selected in the middle of the image, with several horizon picks, which are then automatically spread into the image space by using prediction operators from equation 1. Figure 2(b) shows another kind of painting: This time, the reference trace contains simply the time values along this trace. When spread by predictive painting, it turns into the *relative geologic age* attribute, as defined by (20). Relative age indicates how much a given trace is shifted with respect to the reference trace. Unshifting each trace accomplishes automatic flattening. The result is shown in Figure 3(a). Most of the horizons are successfully flattened, although the algorithm fails to "heal" some of events across the fault because of significant structural discontinuities. In cases like that, the geological insight of the interpreter is invaluable and cannot be easily replaced by an automatic algorithm.

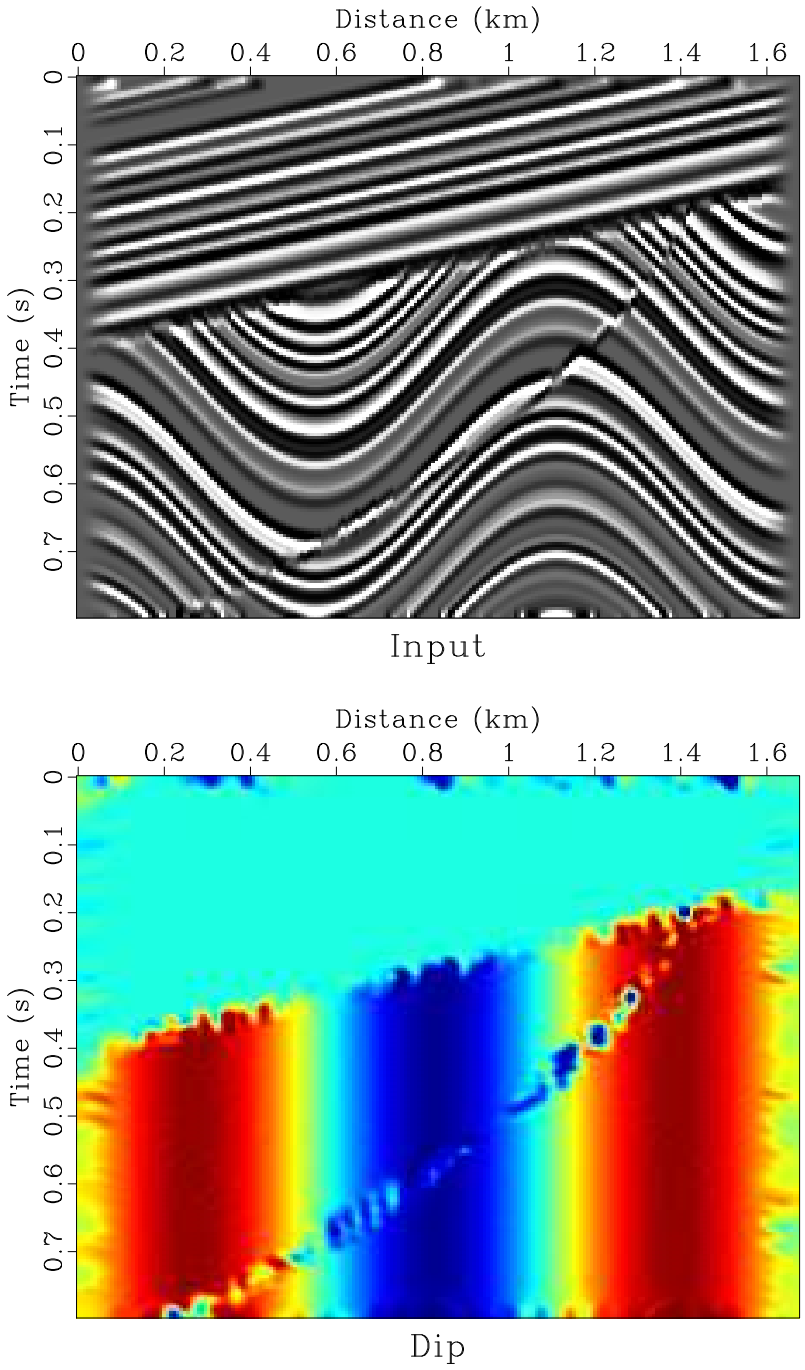


Figure 1: (a) Synthetic image from (5). (b) Local dip estimate. flat/flat sigmoid,sdip

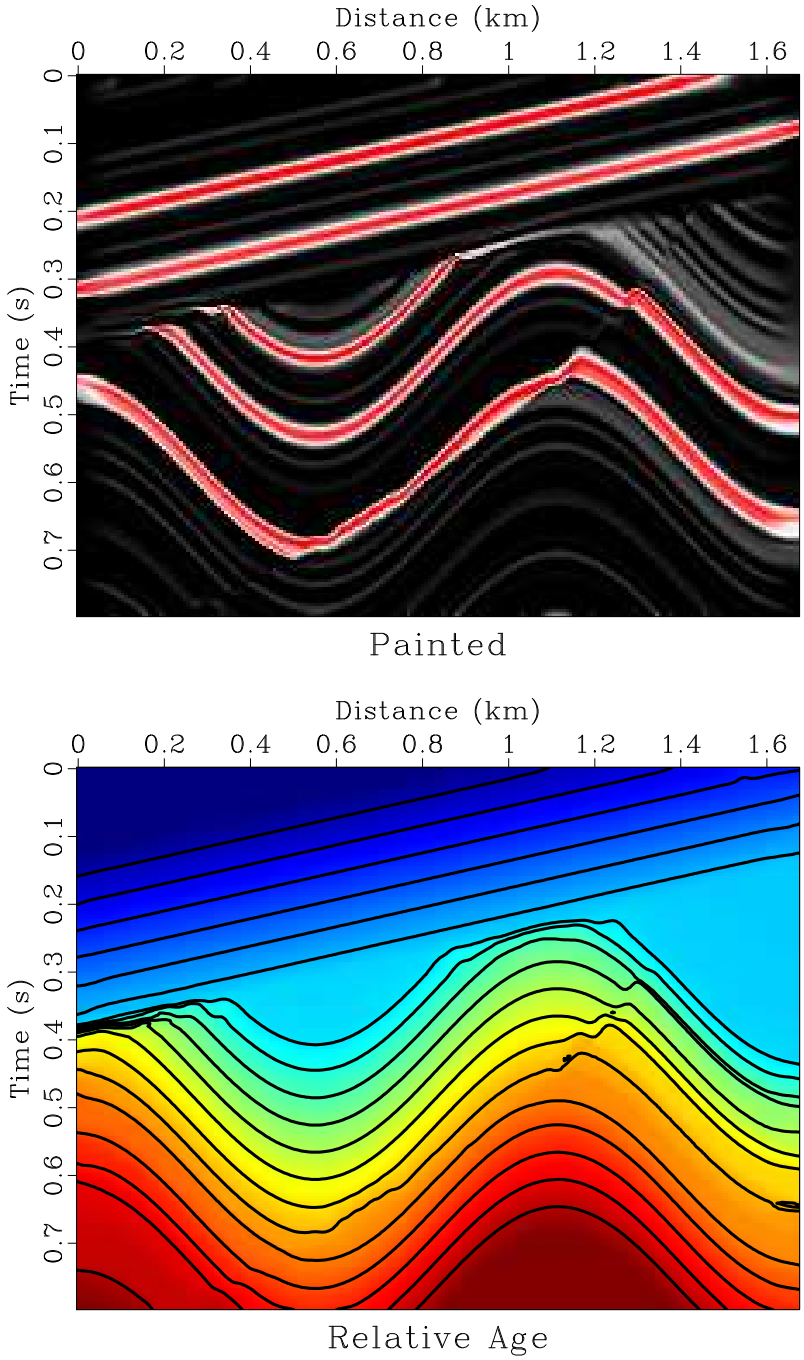


Figure 2: Predictive painting using synthetic image from Figure 2(a). (a) Painted horizons. (b) Painted relative age. The reference trace is in the center of the image. flat/flat spaint,spick

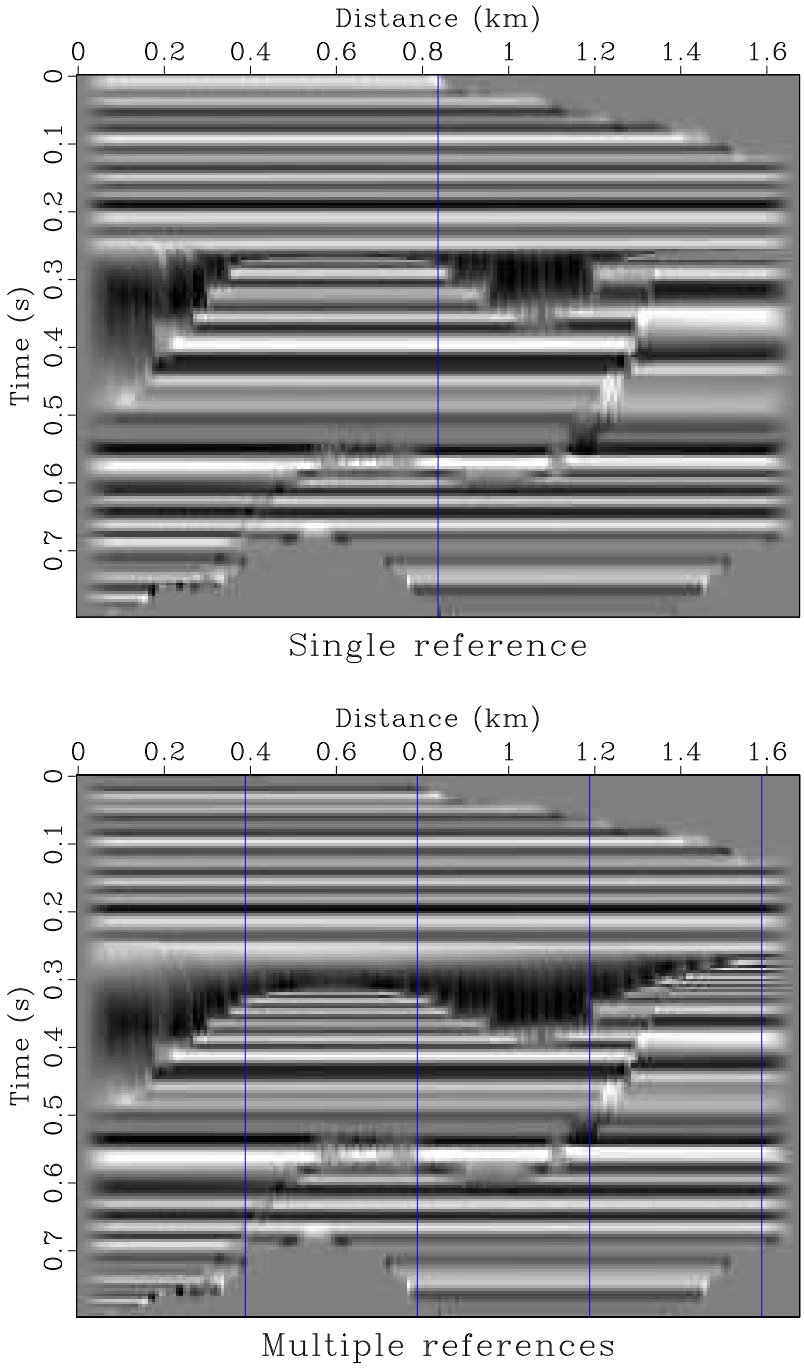


Figure 3: Synthetic image from Figure 2(a) flattened using a single reference trace (a) and multiple references (b). The reference traces are marked by vertical lines. flat/flat sflat,flat

One reference trace is not necessarily structurally connected to all events in the volume. By using multiple references and averaging the relative age among all of them, one can obtain a more accurate extraction of stratal slice information. The result of using multiple references, shown in Figure 3(b), contains more detailed information about horizons that are not structurally visible from the single reference trace.

As mentioned in the introduction, flattening and automatic picking of horizons are useful not only for post-stack structural interpretation but also for prestack data analysis. Figure 4 shows an application of 2-D predictive painting to a marine CMP (common midpoint) gather. After the field of local slopes has been found (Figure 4b), predictive painting is applied to mark individual events (Figure 4c) or to flatten them (Figure 4d), which effectively accomplishes moveout correction. This processing is automatic and does not require manual picking or any prior assumption on the moveout shape. After extracting the moveout information, the moveout parameters can be estimated by least-squares fitting, as described by (3). It is important to note that the gather flattening method is prone to errors in the presence of crossing events, such as multiple reflections, since only the dominant slopes of the strongest events are going to be picked up by the slope estimation procedure.

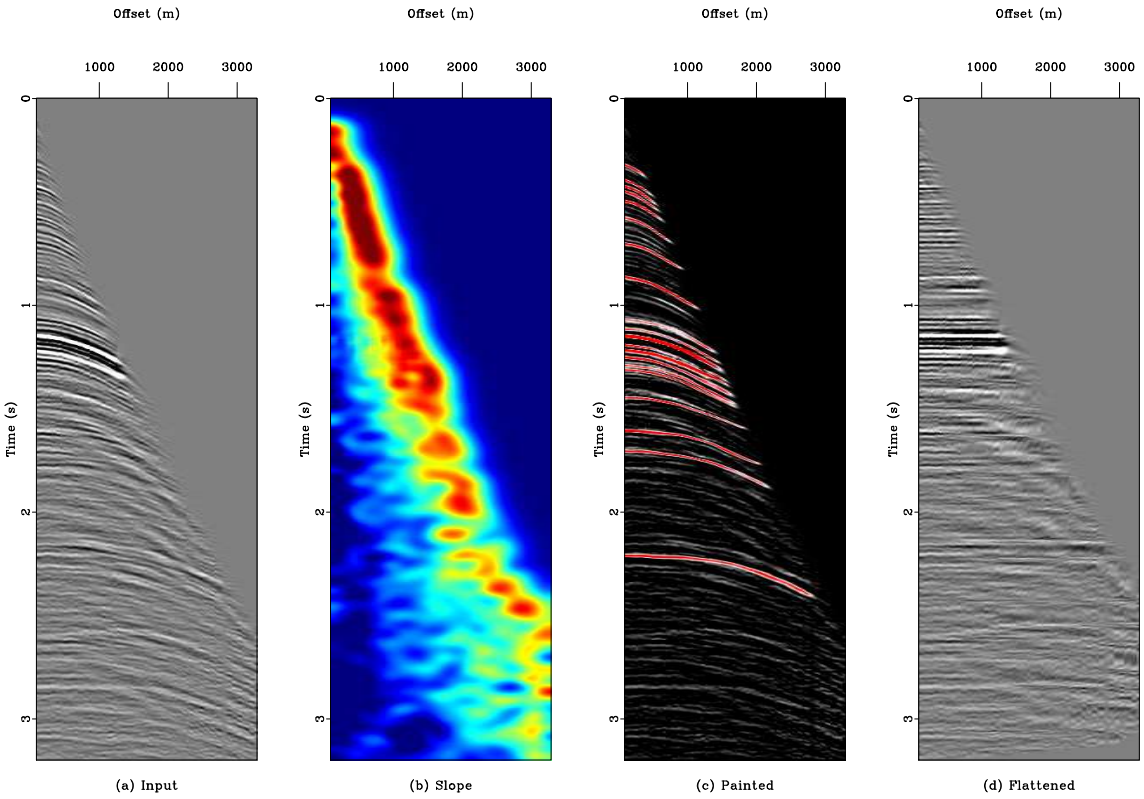


Figure 4: A marine CMP gather (a), estimated local slopes (b), events marked by predictive painting (c), and flattening (d). flat/flatelf pgath

PREDICTIVE PAINTING IN 3-D

The challenge of predictive flattening in 3-D is in selecting a recursive path that the reference trace should follow to paint its neighbors. For choosing this path, I adopt a version of Dijkstra's shortest path algorithm (8; 6). Dijkstra's algorithm finds the path between two nodes in a network of nodes, where there is a cost associated in connecting each node with its neighbors. In our case, the nodes are seismic traces in a 3-D cube. I use the semblance between neighboring traces as a cost function. Dijkstra's algorithm finds the shortest (minimum-cost) path by effectively arranging all nodes in a sequence from low to high cost and evaluating each new node using the information from previous nodes. I run the shortest-path algorithm starting from the reference trace and paint other traces in a recursive sequence using the information from previously painted traces. Using semblance as a cost function helps avoiding 3-D misties by forcing the shortest path to go around possible fault areas.

The 3-D data test is reproduced from (18). It uses a portion of a depth-migrated 3-D image with structural folding and angular unconformities (Figure 5(a)). Inline and crossline dips are measured automatically from the image using plane-wave destruction (Figures 6(a) and 6(b)). Figure 7(a) shows painting of individual strong horizons in the volume. Figure 5(b) displays automatic flattening using predictive painting of the relative geologic age. Figure 7(b) shows some of the corresponding equal-relative-age horizons displayed on top of the original image. Predictive painting is able to correctly identify the most significant three-dimensional structural features.

CONCLUSIONS

I have introduced predictive painting, a numerical algorithm for automatic spreading of information in 3-D seismic volumes according to the local structure of seismic events. The structure is extracted using plane-wave destruction, which operates by predicting each trace in the volume from its inline and crossline neighbors. In the second step, prediction operators are used to spread information inside the volume. This procedure is automatic and computationally fast because it requires only a small fixed number of operations per each trace or data sample. Synthetic and field data tests demonstrate the effectiveness of predictive painting in accomplishing such tasks as automatic flattening and horizon picking.

Further research should concentrate on combining automatic tools with interactive interpretation to allow the information extracted from seismic data to be integrated with the geological insight of the interpreter.

ACKNOWLEDGMENTS

I would like to thank William Burnett, Dave Hale, Yang Liu, Jesse Lomask, James Rickett, and Hongliu Zeng for inspiring discussions. I am especially grateful to late

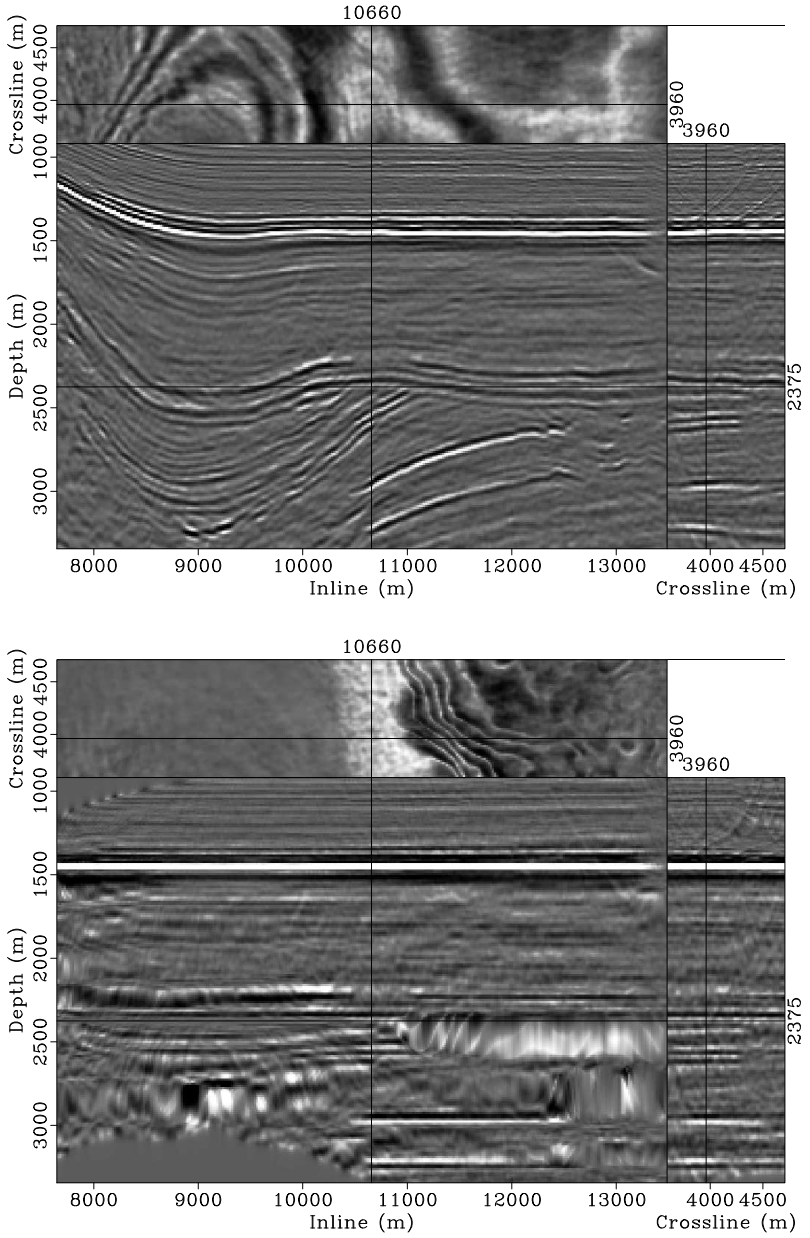


Figure 5: A North Sea image from (18) (a) and its predictive flattening (b). flat/comaz win,wflat

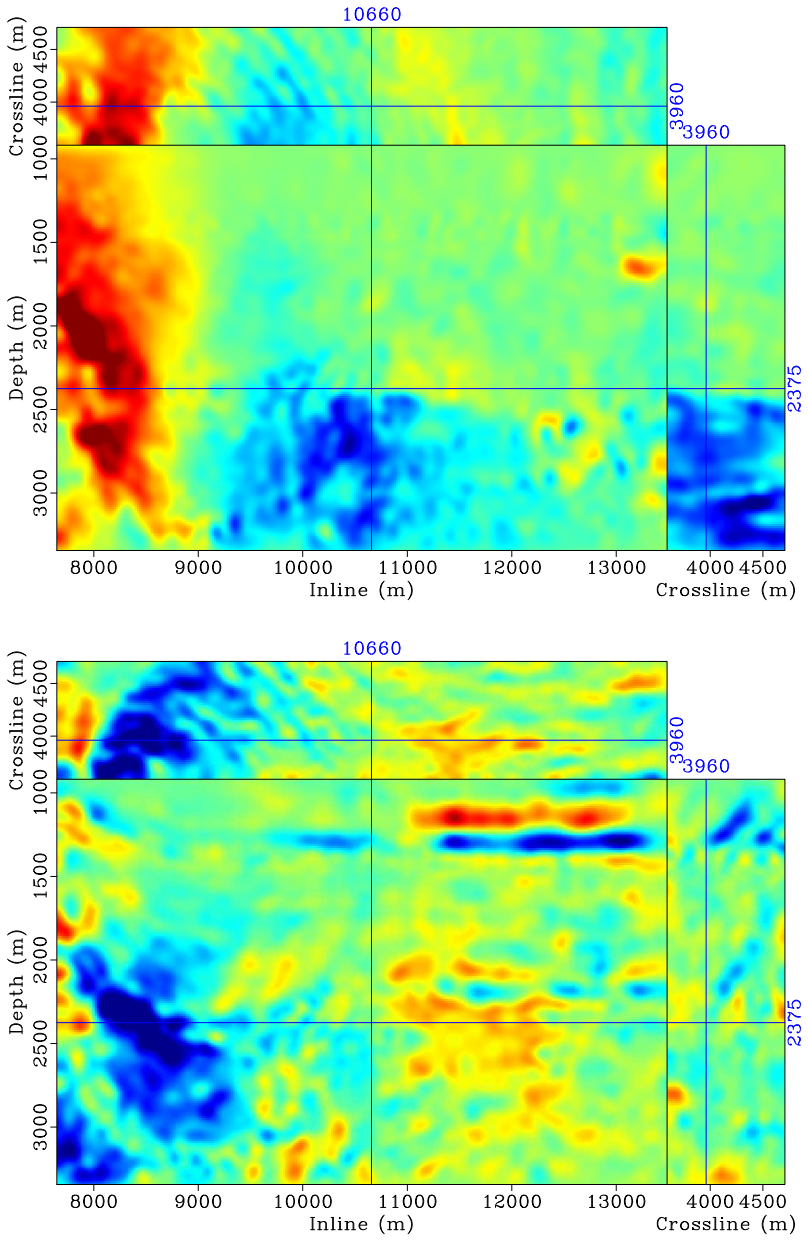


Figure 6: Inline (a) and crossline (b) slopes in the North Sea Image estimated by plane-wave destruction. Blue colors indicate negative slope; red colors, positive slope. flat/comaz wdip1,wdip2

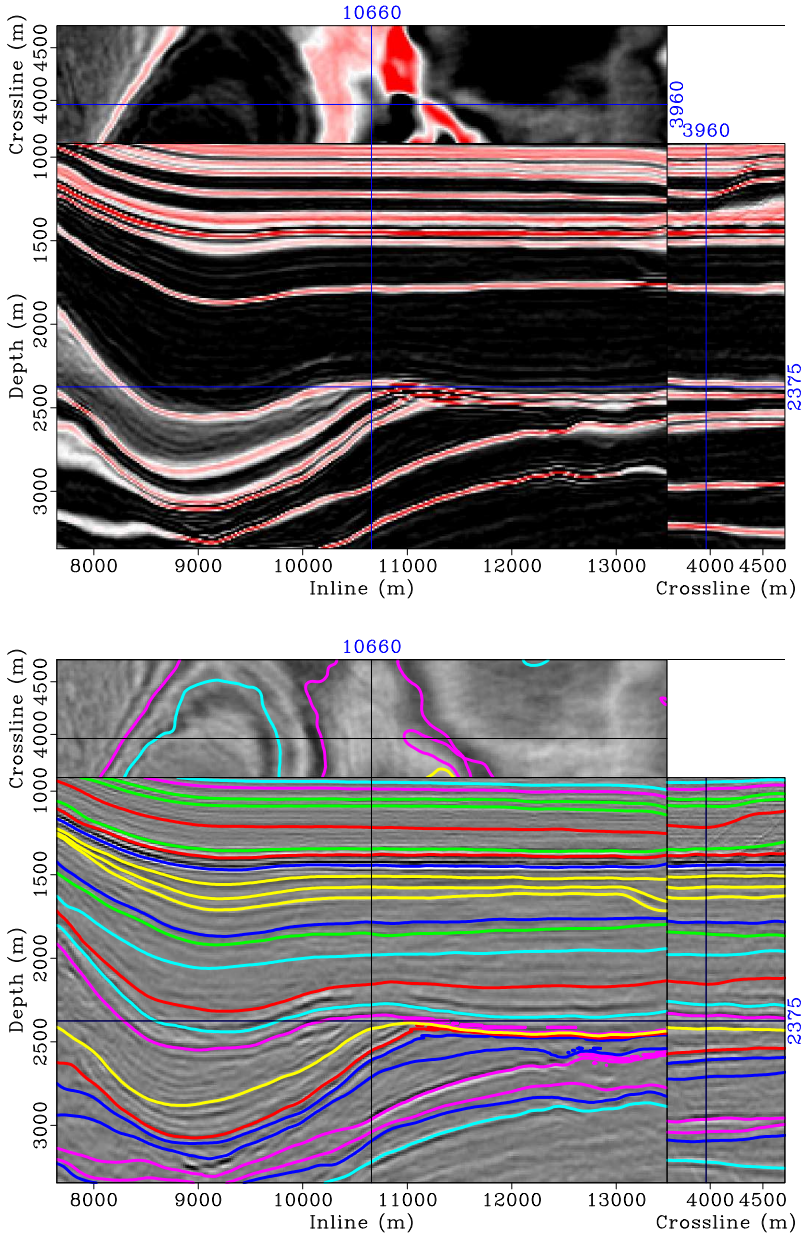


Figure 7: Predictive painting (a) and automatic picking (b) of major horizons in the North Sea Image. flat/comaz wpaint,wcont

Tury Taner for his encouragement and support of this work.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Bienati, N., and U. Spagnolini, 2001, Multidimensional wavefront estimation from differential delays: IEEE Trans. on Geoscience and Remote Sensing, **39**, 655–664.

[2] Bruin, G. D., N. Hemstra, and A. Pouwel, 2007, Stratigraphic surfaces in the depositional and chronostratigraphic (Wheeler-transformed) domain: The Leading Edge, **26**, 883–886.

[3] Burnett, W., and S. Fomel, 2009, Moveout analysis by time-warping, *in* 79th Ann. Internat. Mtg.: Soc. of Expl. Geophys.

[4] Claerbout, J. F., 1992, Earth soundings analysis: Processing versus inversion: Blackwell Scientific Publications.

[5] ———, 2006, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.

[6] Cormen, T. H., C. E. Leiserson, R. L. Rivest, and C. Stein, 2001, Introduction to algorithms: MIT Press and McGraw-Hill.

[7] de Groot, P., G. de Bruin, and K. McBeath, 2006, Opendtect SSIS - Sequence Stratigraphic Interpretation System: Drilling & Exploration World, **15**, 31–34.

[8] Dijkstra, E. W., 1959, A note on two problems in connexion with graphs: Numerische Mathematik, **1**, 269271.

[9] Duveneck, E., and B. Traub, 2006, Automatic moveout correction by local event correlations on coherency-enhanced gathers, *in* 76th Ann. Internat. Mtg.: Soc. of Expl. Geophys., 3036–3040.

[10] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

[11] ———, 2007a, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[12] ———, 2007b, Velocity-independent time-domain seismic imaging using local event slopes: Geophysics, **72**, S139–S147.

[13] Fomel, S., E. Landa, and M. T. Taner, 2007, Post-stack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.

[14] Gulunay, N., F. Gamar, H. Hoeber, M. Dyce, C. MacKenzie, and D. Whitcombe, 2007a, Robust residual gather flattening, *in* 77th Ann. Internat. Mtg.: Soc. of Expl. Geophys., 229–233.

[15] Gulunay, N., M. Magesan, and H. H. Roende, 2007b, Gather flattening: The Leading Edge, **26**, 1538–1543.

[16] Hinkley, D., G. W. Bear, and C. Dawson, 2004, Prestack gather flattening for AVO, *in* 74th Ann. Internat. Mtg.: Soc. of Expl. Geophys., 2329–2329.

[17] Kim, S., and W. W. Symes, 1998, Smooth detectors of linear phase: Inverse Problems, **14**, 101–112.

[18] Lomask, J., A. Guitton, S. Fomel, J. Claerbout, and A. Valenciano, 2006, Flattening without picking: Geophysics, **71**, 13–20.
[19] Stark, T., 2003, Unwrapping instantaneous phase to generate a relative geologic time volume: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 5019–5022.
[20] ———, 2004, Relative geologic time (age) volume: Relating every seismic sample to a geologically reasonable horizon: The Leading Edge, **23**, 928–932.
[21] Symes, W. W., 1994, The plane wave detection problem: Inverse Problems, **10**, 1361–1391.
[22] Zeng, H., M. M. Backus, K. T. Barrow, and N. Tyler, 1998a, Stratal slicing, Part I: Realistic 3-D seismic model: Geophysics, **63**, 502–513.
[23] Zeng, H., S. C. Henry, and J. P. Riola, 1998b, Stratal slicing, Part II: Real 3-D seismic data: Geophysics, **63**, 514–522.

Seislet transform and seislet frame

*Sergey Fomel and Yang Liu**

ABSTRACT

We introduce a digital wavelet-like transform, which is tailored specifically for representing seismic data. The transform provides a multiscale orthogonal basis with basis functions aligned along seismic events in the input data. It is defined with the help of the wavelet lifting scheme combined with local plane-wave destruction. In the 1-D case, the seislet transform is designed to follow locally sinusoidal components. In the 2-D case, it is designed to follow local plane wave components with smoothly variable slopes. If more than one component is present, the transform turns into an overcomplete representation or a tight frame. In these terms, the classic digital wavelet transform is simply a seislet transform for a zero frequency (in 1-D) or zero slope (in 2-D).

The main objective of the new transform is an effective seismic data compression for designing efficient data analysis algorithms. Traditional signal processing tasks such as noise attenuation and trace interpolation become simply defined in the seislet domain. When applied in the offset direction on common midpoint or common image point gathers, the seislet transform finds an additional application in optimal stacking of seismic records.

INTRODUCTION

Wavelet transforms have found many applications in science and engineering (28), including geophysics (20; 11; 40; 26). The power of wavelet transforms, in comparison with the classic Fourier transform, lies in their ability to represent non-stationary signals. As a result, wavelets can provide a compact basis for non-stationary data decomposition. Having a compact basis is useful both for data compression and for designing efficient numerical algorithms.

A number of wavelet-like transforms that explore directional characteristics of images have been proposed in the image analysis literature (41). Among those transforms are bandelets (32), contourlets (12), curvelets (33), directionlets (39), shearlets (21), etc. Unlike isotropic wavelets, directional transforms attempt to design basis functions so that they appear elongated anisotropically along 2-D curves or 3-D surfaces, which might be characteristic for an image. Therefore, these transforms achieve better accuracy and better data compression in representing non-stationary images with curved edges. Curvelets are particularly appropriate for seismic data because

*e-mail: sergey.fomel@beg.utexas.edu

they provide a provably optimal decomposition of wave-propagation operators (4). Application of the curvelet transform to seismic imaging and seismic data analysis has been an area of active research (24; 14; 5; 23).

Although the wavelet theory originated in seismic data analysis (30), none of the known wavelet-like transforms were designed specifically for seismic data. Even though some of the transforms are applicable for representing seismic data, their original design was motivated by different kinds of data, such as piecewise-smooth images. In this paper, we investigate the possibility of designing a transform tailored specifically for seismic data. In analogy with the previous naming games, we call such a transform *the seislet transform* (17).

In seismic data analysis, it is common to represent signals as sums of sinusoids (in 1-D) or plane waves (in 2-D) with the help of the digital Fourier transform (DFT). Certain methods for seismic data regularization, such as the anti-leakage Fourier transform (42), the Fourier reconstruction method (44; 45; 46), or POCS (1) rely on the ability to represent signals sparsely in the transform domain. The digital wavelet transform (DWT) is often preferred to the Fourier transform for characterizing digital images, because of its ability to localize events in both time and frequency domains (25; 28). However, DWT may not be optimal for describing data that consist of individual sinusoids or plane waves. It is for those kinds of data that the seislet transform attempts to achieve an optimally compact representation.

The approach taken in this paper follows the general recipe for digital wavelet transform construction known as the *lifting scheme* (34). The lifting scheme provides a convenient and efficient construction for digital wavelet transforms of different kinds. The key ingredients of this scheme are a prediction operator and an update operator defined at different digital scales. The goal of the prediction operator is to predict regular parts of the input data so that they could be subtracted from the analysis. The goal of the update operator is to carry essential parts of the input data to the next analysis scale. Conventional wavelet transforms use prediction and update operators designed for characterizing locally smooth images. In this paper, we show how designing prediction and update tailored for seismic data can improve the effectiveness of the transform in seismic applications. In 1-D, our prediction and update operators focus on predicting sinusoidal signals with chosen frequencies. In 2-D and 3-D, we use predictions along locally dominant event slopes found by the method of plane-wave destruction (7; 16). One can extend the idea of the seislet transform further by changing the definition of prediction and update operators in the lifting scheme (27).

The seislet transform decomposes a seismic image into an orthogonal basis which is analogous to the wavelet basis but aligned along dominant seismic events. In 1-D, the classic wavelet transform is equivalent to the seislet transform with a zero frequency. In 2-D, the wavelet transform in the horizontal direction is equivalent to the seislet transform with a zero slope. When more than one frequency or more than one slope field are employed for analysis, the seislet transform turns into an overcomplete representation or a tight frame.

The paper is organized as follows. We start by reviewing the digital wavelet transform and the lifting scheme. Next, we modify the lifting scheme to define 1-D and 2-D seislet transforms. Finally, we generalize the transform construction to a frame. We illustrate applications of both the seislet transform and the seislet frame with synthetic and field data examples.

LIFTING SCHEME FOR WAVELET TRANSFORMS

In order to define the new transform, we follow the general recipe for digital wavelet transforms provided by (35). In the most general terms, the lifting scheme (34) can be defined as follows:

1. Organize the input data as a sequence of records.
2. Break the data into even and odd components $\mathbf{e}$ and $\mathbf{o}$.
3. Find a residual difference $\mathbf{r}$ between the odd component and its prediction from the even component:
$$\mathbf{r} = \mathbf{o} - \mathbf{P}[\mathbf{e}] \ , \tag{1}$$
where $\mathbf{P}$ is a *prediction* operator.
4. Find a coarse approximation $\mathbf{c}$ of the data by updating the even component
$$\mathbf{c} = \mathbf{e} + \mathbf{U}[\mathbf{r}] \ , \tag{2}$$
where $\mathbf{U}$ is an *update* operator.
5. The coarse approximation $\mathbf{c}$ becomes the new data, and the sequence of steps is repeated at the next scale level.

A digital wavelet transform consists of data approximation at the coarsest level and residuals from all the levels. The key in designing an effective transform is making sure that the prediction operator $\mathbf{P}$ leaves small residuals while the update operator $\mathbf{U}$ preserves essential features of the original data while promoting them to the next scale level. For example, one can obtain the classic Haar wavelet by defining the prediction operator as a simple shift from the nearest sample:

$$\mathbf{P}[\mathbf{e}]_k = \mathbf{e}_k \ , \tag{3}$$

with the update operator designed to preserve the DC (zero frequency) component of the signal. Alternatively, the $(2,2)$ Cohen-Daubechies-Feauveau biorthogonal wavelets (9) are constructed by making the prediction operator a linear interpolation between two neighboring samples,

$$\mathbf{P}[\mathbf{e}]_k = \left(\mathbf{e}_{k-1} + \mathbf{e}_k\right)/2 \ , \tag{4}$$

and by constructing the update operator to preserve the running average of the signal (35), as follows:

$$\mathbf{U}[\mathbf{r}]_k = (\mathbf{r}_{k-1} + \mathbf{r}_k)/4 \ . \tag{5}$$

The digital wavelet transform is an efficient operation. Assuming that the prediction and update operation take a constant cost per record, the number of operations at the finest scale is proportional to the total number of records N, the next scale computation takes $O(N/2)$, etc. so that the total number of operations is proportional to $N + N/2 + N/4 + \ldots + 2 = 2\,(N-1)$, which is smaller than the $O(N \log N)$ cost of the Fast Fourier Transform.

The digital wavelet transform is also easily invertible. Reversing the lifting scheme operations provides the inverse transform algorithm, as follows:

1. Start with the coarsest scale data representation $\mathbf{c}$ and the coarsest scale residual $\mathbf{r}$.
2. Reconstruct the even component $\mathbf{e}$ by reversing the operation in equation 2, as follows:
$$\mathbf{e} = \mathbf{c} - \mathbf{U}[\mathbf{r}] \ , \tag{6}$$
3. Reconstruct the odd component $\mathbf{o}$ by reversing the operation in equation 10, as follows:
$$\mathbf{o} = \mathbf{r} + \mathbf{P}[\mathbf{e}] \ , \tag{7}$$
4. Combine the odd and even components to generate the data at the previous scale level and repeat the sequence of steps.

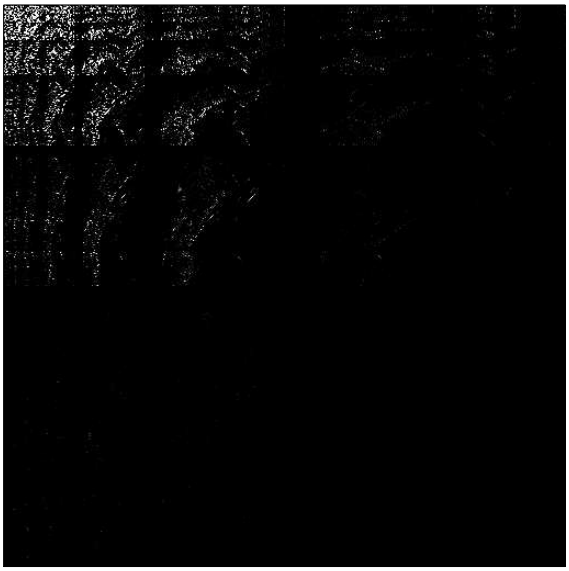
Figure 1 shows a classic benchmark image from the image analysis literature and its digital wavelet transform using 2-D biorthogonal wavelets. Thanks to the general smoothness of the "Lena" image, the residual differences from equation 2 (stored as wavelet coefficients at different scales) have a small dynamic range, which enables an effective compression of the image in the transform domain. Wavelet compression algorithms are widely used in practice for compression of natural images. As for compression of seismic data, the classic DWT may not be optimal, because it does not take into account the specific nature of seismic data patterns. In the next section, we turn the wavelet transform into the seislet transform, which is tailored for representing seismic data.

FROM WAVELETS TO SEISLETS

We adopt the general idea of the lifting scheme to transforming seismic data. The key idea of the seislet transform (17) is recognizing that

- seismic data can be organized as collections of traces or records;

Lena



2−D Wavelet Transform

Figure 1: Benchmark "Lena" image from image analysis literature (a) and its 2-D digital wavelet transform using bi-orthogonal wavelets (b). seislet/lena lena,linear2

- prediction of one seismic trace or record from the other and update of records on the next scale should follow features characteristic for seismic data.

1-D seislet transform

The prediction and update operators employed in the lifting scheme can be understood as digital filters. In the Z-transform notation, the Haar prediction filter from equation 9 is

$$P(Z) = Z \tag{8}$$

(shifting each sample by one), and the linear interpolation filter from equation 2 is

$$P(Z) = 1/2\,(1/Z + Z)\ . \tag{9}$$

These predictions are appropriate for smooth signals but may not be optimal for a sinusoidal signal. In comparison, the prediction

$$P(Z) = Z/Z_0\ , \tag{10}$$

where $Z_0 = e^{i\,\omega_0 \Delta t}$, perfectly characterizes a sinusoid with ω_0 circular frequency sampled on a Δt grid. In other words, if a constant signal ($\omega_0 = 0$) is perfectly predicted by shifting each trace to its neighbor, a sinusoidal signal ($\omega_0 \neq 0$) requires the shift to be modulated by an appropriate frequency. Likewise, the linear interpolation in equation 9 needs to be replaced by a filter tuned to a particular frequency in order to predict a sinusoidal signal with that frequency perfectly:

$$P(Z) = 1/2\,(Z_0/Z + Z/Z_0)\ . \tag{11}$$

The analysis easily extends to higher-order filters.

2-D seislet transform

If we view seismic data as collections of traces, we can predict one trace from the other by following local slopes of seismic events. Such a prediction is a key operation in the method of plane-wave destruction (16). In fact, it is the minimization of prediction error that provides a criterion for estimating local slopes (7). For completeness, we include a review of plane-wave destruction in the appendix.

The prediction and update operators for a simple seislet transform are defined by modifying the biorthogonal wavelet construction in equations 2-5 as follows:

$$\mathbf{P}[\mathbf{e}]_k = \left(\mathbf{S}_k^{(+)}[\mathbf{e}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{e}_k]\right)/2 \tag{12}$$

$$\mathbf{U}[\mathbf{r}]_k = \left(\mathbf{S}_k^{(+)}[\mathbf{r}_{k-1}] + \mathbf{S}_k^{(-)}[\mathbf{r}_k]\right)/4\ , \tag{13}$$

where $\mathbf{S}_k^{(+)}$ and $\mathbf{S}_k^{(-)}$ are operators that predict a trace from its left and right neighbors correspondingly by shifting seismic events according to their local slopes. The predictions need to operate at different scales, which, in this case, mean different separation distances between the traces. Equations 12-13, in combination with the forward and inverse lifting schemes 10-2 and 6-7, provide a complete definition of the 2-D seislet transform.

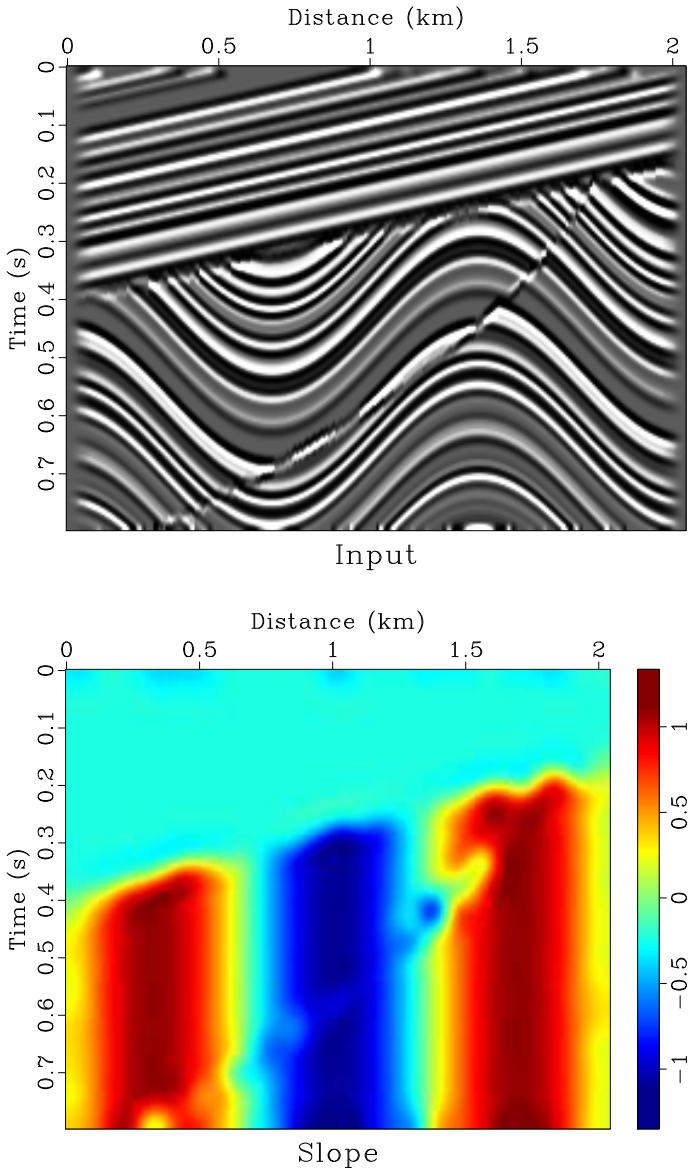


Figure 2: Synthetic seismic image (a) and local slopes estimated by plane-wave destruction (b). seislet/seis sigmoid,sigmoiddip

Figure 2(a) shows a synthetic seismic image from (8). After estimating local slopes from the image by plane-wave destruction (Figure 2(b)), we applied the 2-D seislet transform described above. The transform is shown in Figure 3(b) and should be

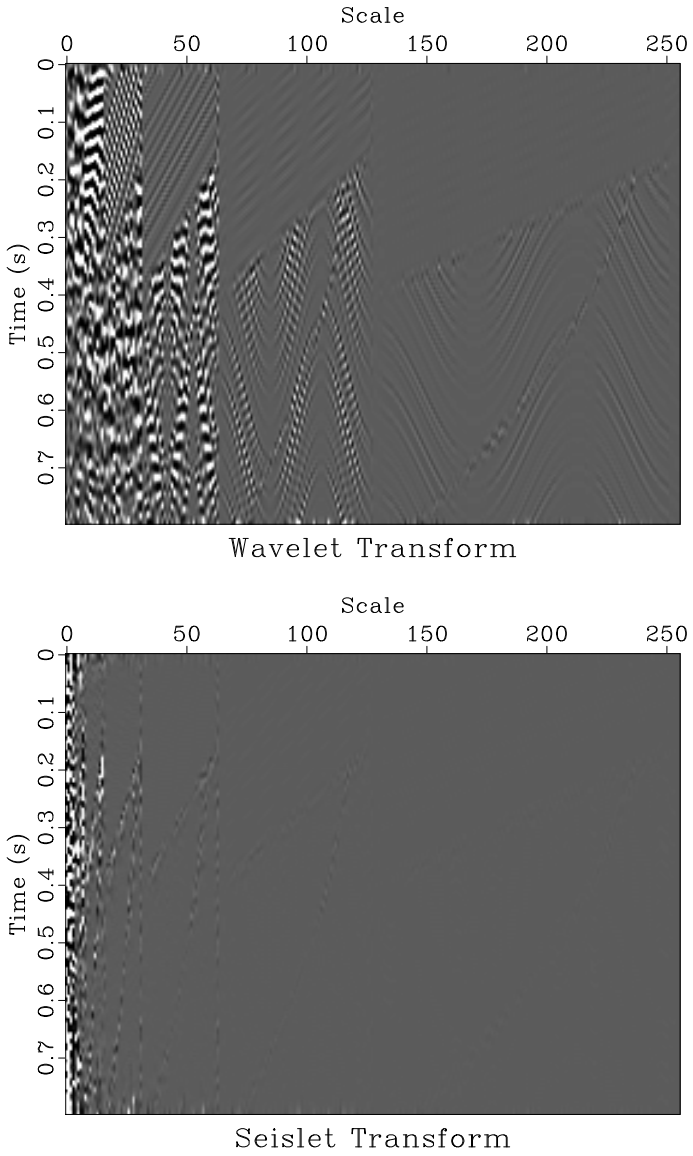


Figure 3: Wavelet transform (a) and seislet transform (b) of the synthetic image from Figure 2. seislet/seis sigmoidwvlt,sigmoidseis

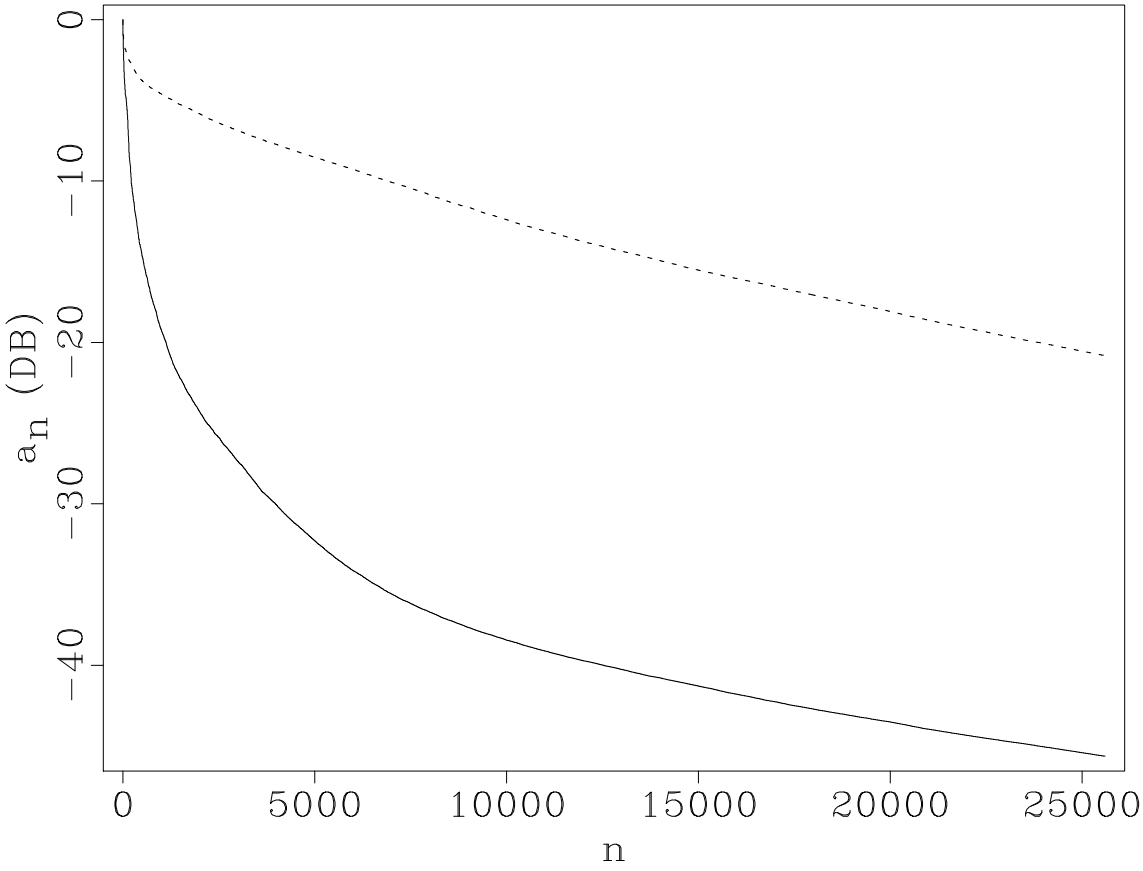


Figure 4: Transform coefficients sorted from large to small, normalized, and plotted on a decibel scale. Solid line: seislet transform. Dashed line: wavelet transform. seislet/seis sigmoidcoef

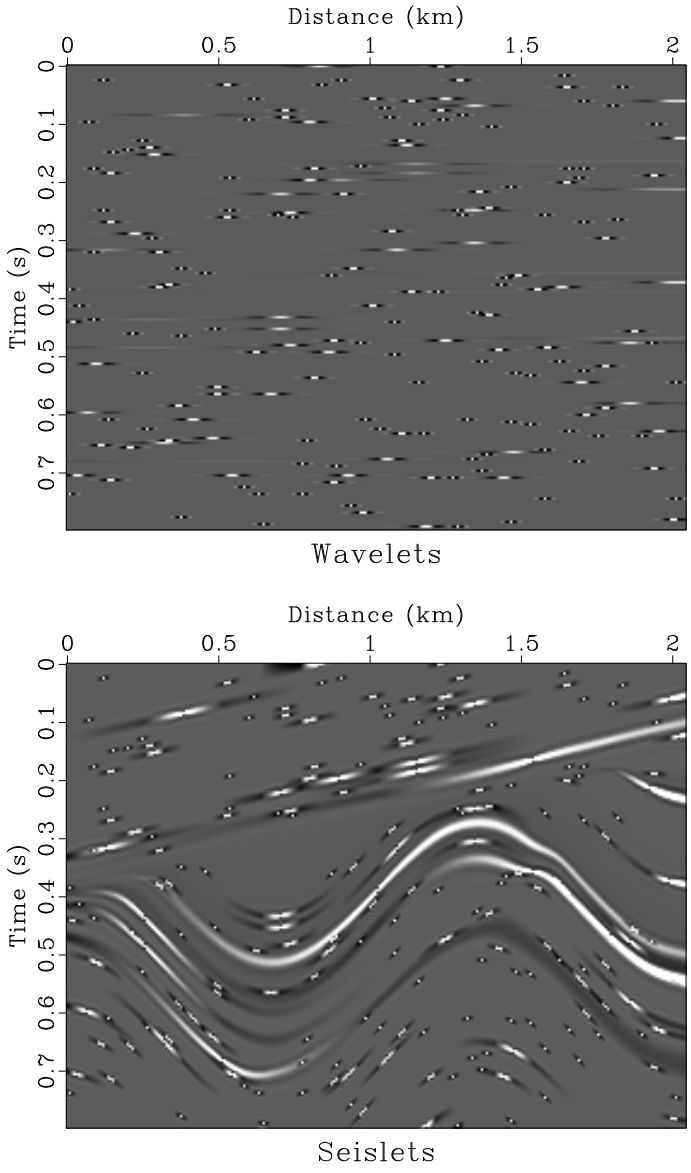


Figure 5: Randomly selected representative basis functions for wavelet transform (a) and seislet transform (b). seislet/seis sigmoidimpw,sigmoidimps

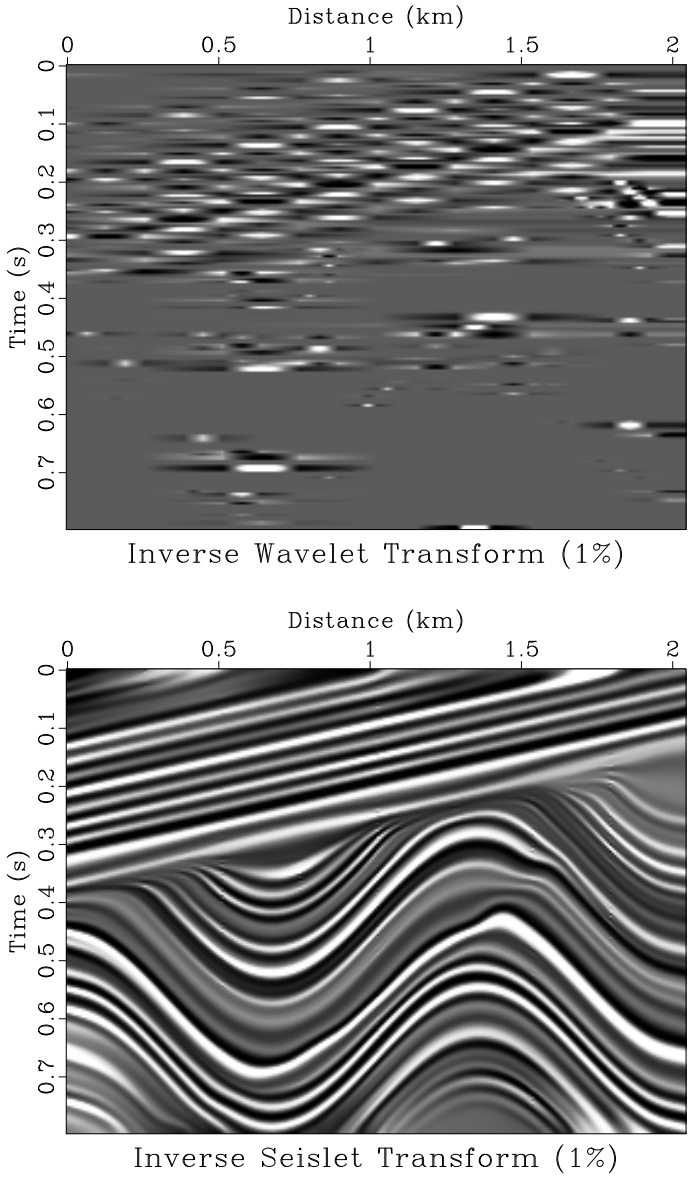


Figure 6: Synthetic image reconstruction using only 1% of significant coefficients (a) by inverse wavelet transform (b) by inverse seislet transform. Compare with Figure 2(a). seislet/seis sigmoidwrec1,sigmoidsrec1

compared with the corresponding wavelet transform in Figure 3(a). Apart from the fault and unconformity regions, where the image is not predictable by continuous local slopes, the 2-D seislet transform coefficients are small, which enables an effective compression. In contrast, the wavelet transform has small residual coefficients at fine scales but develops large coefficients at coarser scales. Figure 4 shows a comparison between the decay of coefficients (sorted from large to small) between the wavelet transform and the seislet transform. A significantly faster decay of the seislet coefficients is evident.

Effectively, the wavelet transform in this case is equivalent to the 2-D seislet transform with the erroneous zero slope. Figure 5 shows example basis functions for the wavelet and 2-D seislet transform used in this example. If using only a small number of the most significant coefficients, the wavelet transform fails to reconstruct the most important features of the original image, while the 2-D seislet transform achieves an excellent reconstruction (Figure 6). We use the method of soft thresholding (13) for selecting the most significant coefficients.

Example applications of 2-D seislet transform

In this section, we discuss some example applications of the 2-D seislet transform.

Denoising and trace interpolation

Figure 7(a) shows a common-midpoint gather from a North Sea dataset. Plane-wave destruction estimates local slopes shown in Figure 7(b) and enables the 2-D seislet transform shown in Figure 8(a). Small dynamic range of seislet coefficients implies a good compression ratio. Figure 8(b) shows data reconstruction using only 5% of the significant seislet coefficients.

If we choose the significant coefficients at the coarse scale and zero out difference coefficients at the finer scales, the inverse transform will effectively remove incoherent noise from the gather (Figure 8). Thus, denoising is a naturally defined operation in the 2-D seislet domain (Figure 9).

If we extend the seislet transform domain and interpolate the smooth local slope to a finer grid, the inverse seislet transform will accomplish trace interpolation of the input gather (Figure 10). We extend not simply with zeros but with small random noise to account for the fact that realistic noise is unpredictable and therefore exists on different scale levels. In this example, the number of traces is increased by four. Thus, trace interpolation also turns out to be a natural operation when viewed from the 2-D seislet domain.

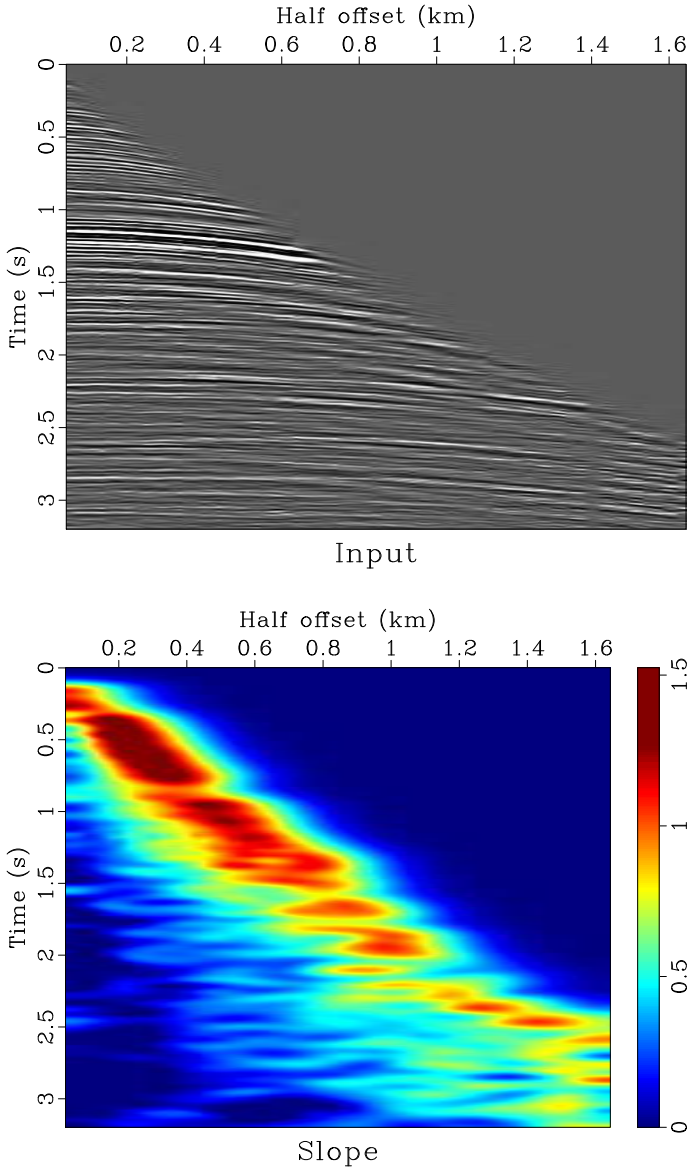


Figure 7: Common-midpoint gather (a) and local slopes estimated by plane-wave destruction (b). seislet/gath gath,gathdip

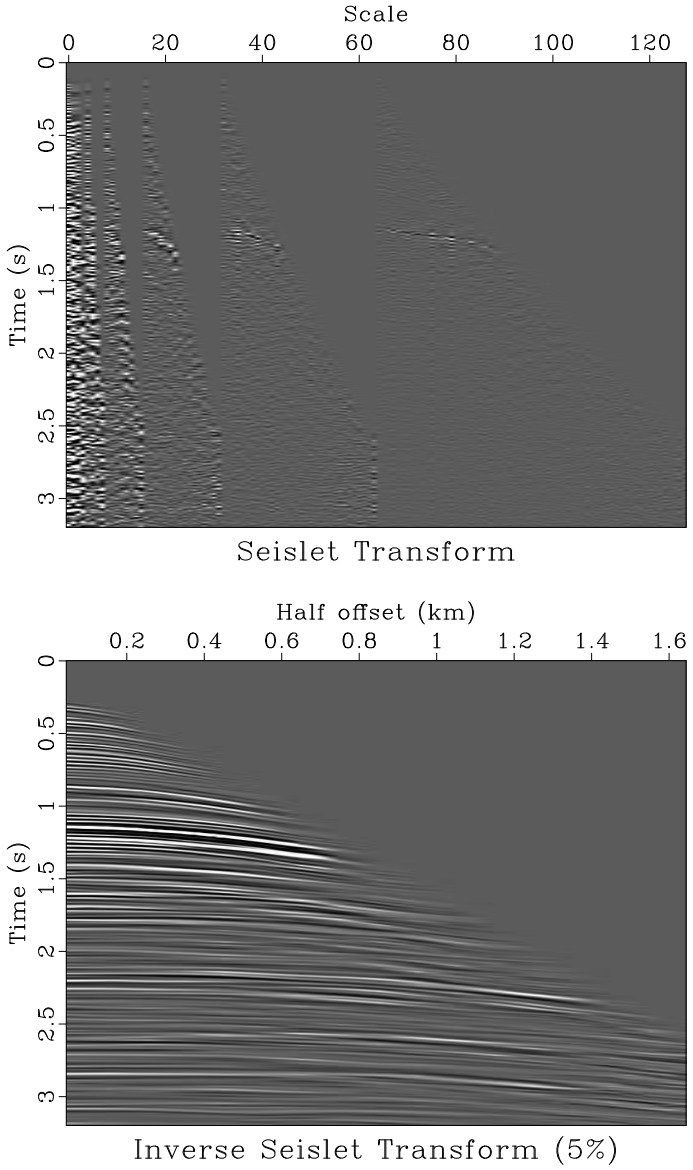


Figure 8: Seislet transform of the input gather (a) and (a) and data reconstruction using only 5% of significant seislet coefficients (b). seislet/gath gathseis,gathsrec5

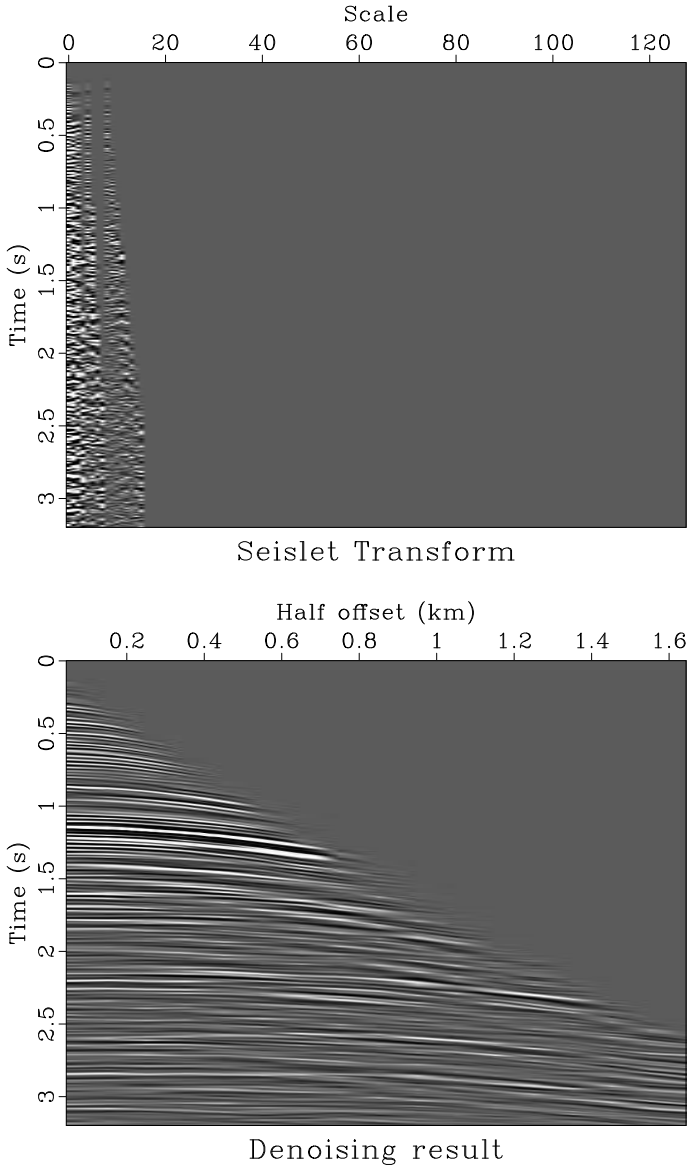


Figure 9: Zeroing seislet difference coefficients at fine scales (a) enables effective denoising of the reconstructed data (b). seislet/gath sign,gathslet16

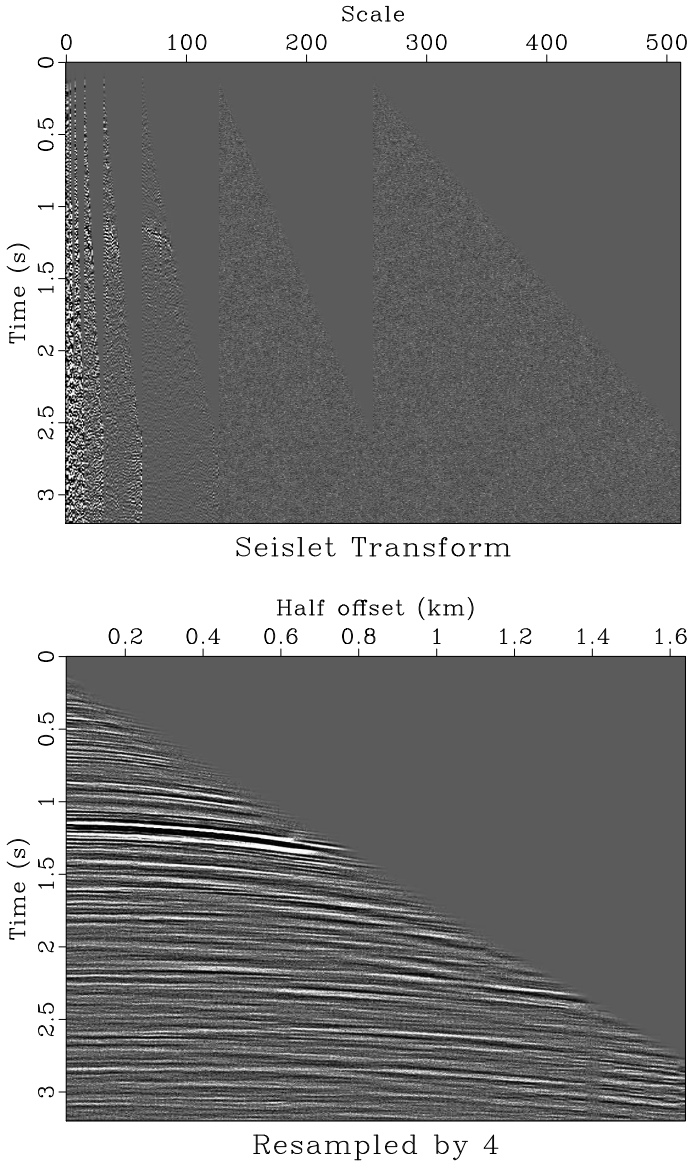


Figure 10: Extending seislet transform with random noise (a) enables trace interpolation in the reconstructed data. The interpolated section (b) has 4 times more traces than the original shown in Figure 7(a). seislet/gath seis4,gath4

Seislet stack

The seislet transform acquires a special meaning when applied in the offset direction on common midpoint or common image point gathers. According to the lifting construction, the zero-order seislet coefficient is nothing more than seismic stack computed in a recursive manner by successive partial stacking of neighboring traces. As a consequence, seislet stack avoids the problem of "NMO stretch" associated with usual stacking (22) as well as the problem of nonhyperbolic moveouts (19). All other gather attributes including multiple reflections and amplitude variation with offset appear in the higher order seislet coefficients. Figure 11 shows a comparison between the conventional normal moveout stack and the seislet stack. The higher resolution of the seislet stack is clearly visible. Figure 12 compares the common-midpoint gather after conventional normal moveout correction and an effective seislet moveout computed by separating contributions from individual traces to the seislet stack.

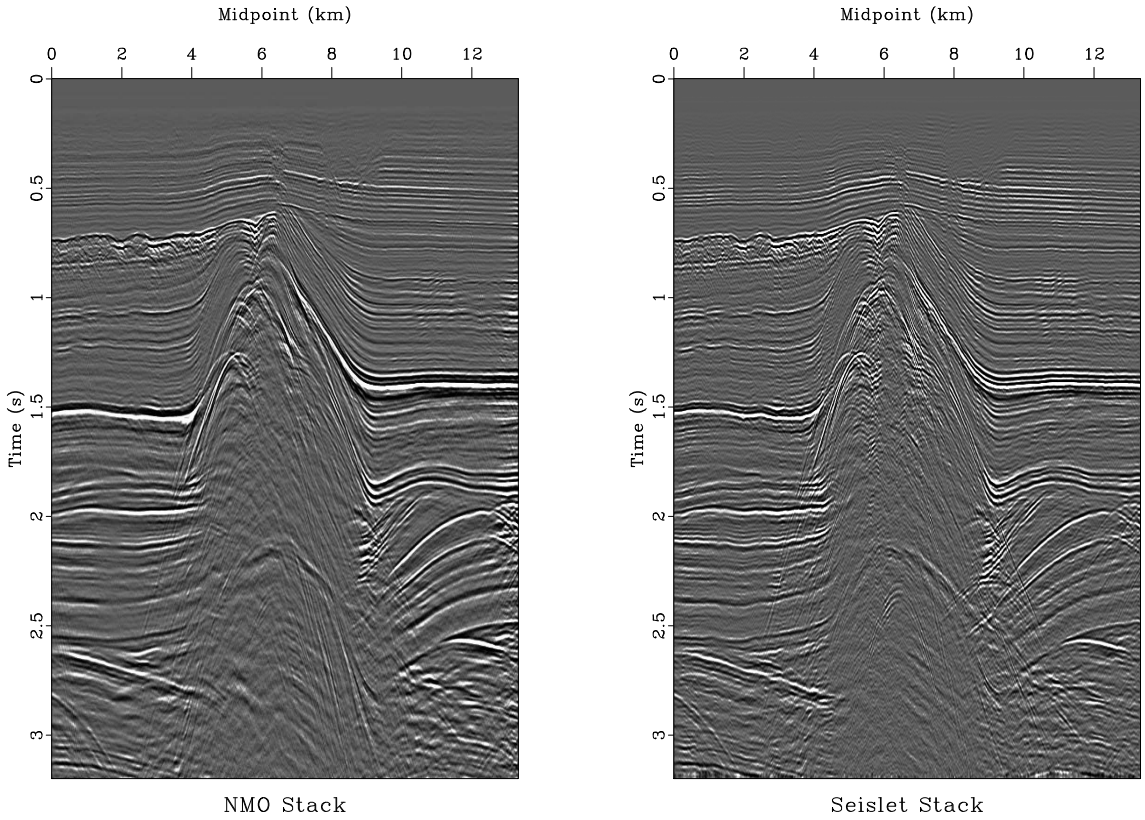


Figure 11: Left: conventional normal-moveout stack. Right: seislet stack. seislet/gath sstack

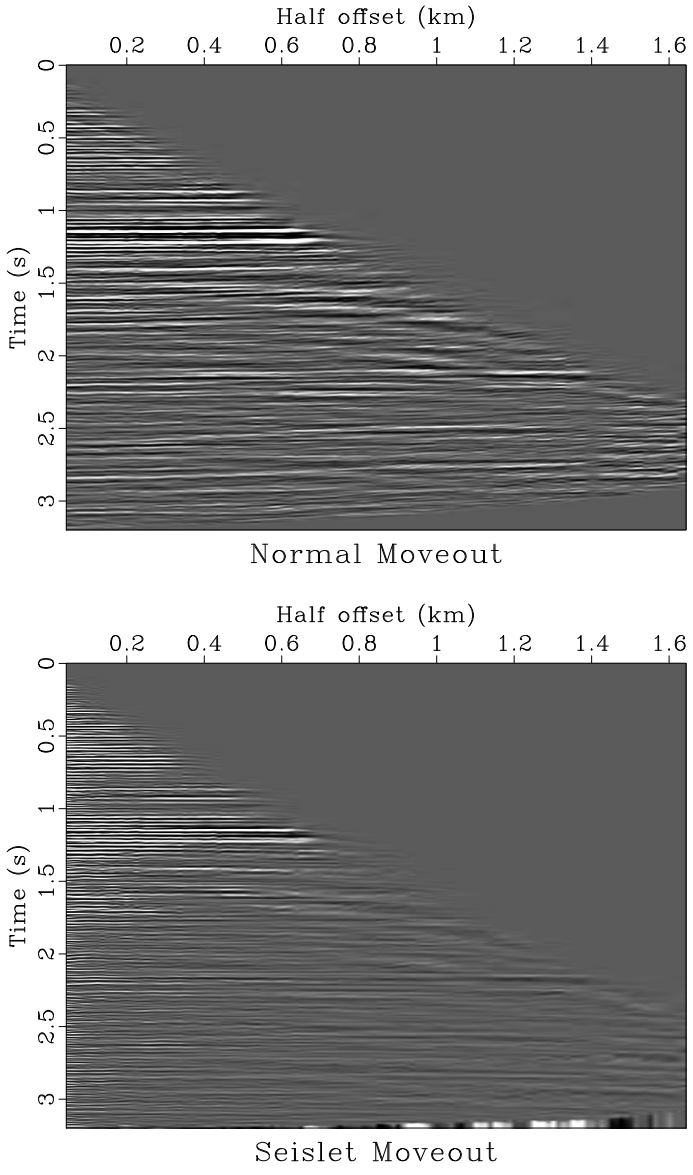


Figure 12: Input gather after normal moveout correction (a) and effective seislet moveout (b). seislet/gath nmo,snmo

FROM TRANSFORM TO FRAME

The 1-D and 2-D transforms, defined in the previous sections, are appropriate for analyzing signals, which have a single dominant sinusoid or plane-wave component. In practice, it is common to analyze signals composed of multiple sinusoids (in 1-D) or plane waves (in 2-D). If a range of frequencies or plane-wave slopes is chosen, and the appropriate transform is constructed for each of them, all the transform domains taken together will constitute an overcomplete representation or a frame (28).

Mathematically, if $\mathbf{F}_n$ is the orthonormal seislet transform for n-th frequency or plane wave, then, for any data vector $\mathbf{d}$,

$$\sum_{n=1}^{N} \|\mathbf{F}_n\,\mathbf{d}\|^2 = \sum_{n=1}^{N} \mathbf{d}^T\,\mathbf{F}_n^T\,\mathbf{F}_n\,\mathbf{d} = \sum_{n=1}^{N} \|\mathbf{d}\|^2 = N\,\|\mathbf{d}\|^2\;, \tag{14}$$

which means that all transforms taken together constitute a *tight frame* with constant N.

For example, in the 1-D case, one can find appropriate frequencies by autoregressive spectral analysis (2; 29). We define the algorithm for the *1-D seislet frame* as follows:

1. Select a range of coefficients $Z_1, Z_2, \ldots, Z_k$. When using autoregressive spectral analysis, these coefficients are simply the roots of the prediction-error filter. Alternatively, they can be defined from an appropriate range of frequencies $\omega_1, \omega_2, \ldots, \omega_k$.
2. For each of the coefficients, perform the 1-D seislet transform.

Because of its over-completeness, a frame representation for a given signal is not unique. In order to assure that different frequency components do not leak into other parts of the frame, it is advantageous to employ sparseness-promoting inversion. We adopt a nonlinear shaping regularization scheme (18), analogous to the sparse inversion method of (10), and define sparse decomposition as an iterative process

$$\widehat{\mathbf{f}}_{k+1} = \mathbf{S}[\mathbf{F}\,\mathbf{d} + (\mathbf{I} - \mathbf{F}\,\mathbf{F}')\,\widehat{\mathbf{f}}_k]\;, \tag{15}$$

$$\mathbf{f}_{k+1} = \mathbf{f}_k + \mathbf{F}\,\mathbf{d} - \mathbf{F}\,\mathbf{F}'\widehat{\mathbf{f}}_{k+1}\;, \tag{16}$$

where $\mathbf{f}_k$ is the seislet frame at k-th iteration, $\widehat{\mathbf{f}}_k$ is an auxiliary quantity, $\mathbf{d}$ is input data, $\mathbf{I}$ is the identity operator, $\mathbf{F}$ and $\mathbf{F}'$ are frame construction and deconstruction operators

$$\begin{aligned} \mathbf{F} &\equiv \begin{bmatrix} \mathbf{F}_1 & \mathbf{F}_2 & \cdots & \mathbf{F}_k \end{bmatrix}^T\;, \\ \mathbf{F}' &\equiv \begin{bmatrix} \mathbf{F}_1^{-1} & \mathbf{F}_2^{-1} & \cdots & \mathbf{F}_k^{-1} \end{bmatrix}\;, \end{aligned}$$

where $\mathbf{F}_j$ is the seislet transform for an individual frequency, and $\mathbf{S}$ is a nonlinear shaping operator, such as soft thresholding (13). The iteration 15-16 starts with

$\mathbf{f}_0 = \mathbf{0}$ and $\widehat{\mathbf{f}}_0 = \mathbf{F}\,\mathbf{d}$ and is related to the linearized Bregman iteration (31; 43). We find that a small number of iterations is usually sufficient for convergence and achieving both model sparseness and data recovery.

1-D data analysis with 1-D seislet frame

We use a simple synthetic test to verify the compression effectiveness of 1-D seislet frame. A test signal mixing two sinusoids with different frequencies and some random noise is displayed in Figure 13(a). We use a prediction-error filter to detect the signal frequencies and to design the corresponding seislet frame. The result is shown in Figure 13(b). The 1-D seislet frame algorithm with shaping regularization compresses the sinusoidal signal into two nearly perfect impulses with some dispersive random noise. For comparison, we also apply DFT and DWT to transform the signal (Figures 13(d) and 13(c)). In the Fourier transform domain, the signal appears as two impulses corresponding to the chosen frequency components. The resolution is not perfect because of spectral leakage caused by non-periodic input data. In the wavelet domain, the transform coefficients are not compressed well. For further comparison, we plot the coefficients in the three different transform domains, sorted from large to small, on a decibel scale (Figure 14). The significantly faster rate of coefficient decay shows the superiority of the 1-D seislet frame in compressing sinusoidal signals.

2-D data analysis with 1-D seislet frame

To analyze 2-D data, one can apply 1-D seislet frame in the distance direction after the Fourier transform in time (the F-X domain). In this case, different frame frequencies correspond to different plane-wave slopes (3). We use a simple plane-wave synthetic model to verify this observation (Figure 3(a)). The F-X plane is shown in Figure 15(b). We find a prediction-error-filter (PEF) in each frequency slice and detect its roots to select appropriate spatial frequencies. We use Burg's algorithm for PEF estimation (2; 6) and an eigenvalue-based algorithm for root finding (15). The seislet coefficients and the corresponding recovered plane-wave components are shown in Figure 16. Similarly to the 1-D example, information from different plane-waves gets strongly compressed in the transform domain.

2-D data analysis with 2-D seislet frame

To show an example of 2-D data analysis with 2-D seislet frames, we use the CMP gather from Figure 7(a). We try two different choices for selecting a set of dip fields for the frame construction.

First, we define dip fields by scanning different constant dips (Figure 17(a)). In this case, the zero-scale coefficients out of the 2-D seislet frame correspond to the

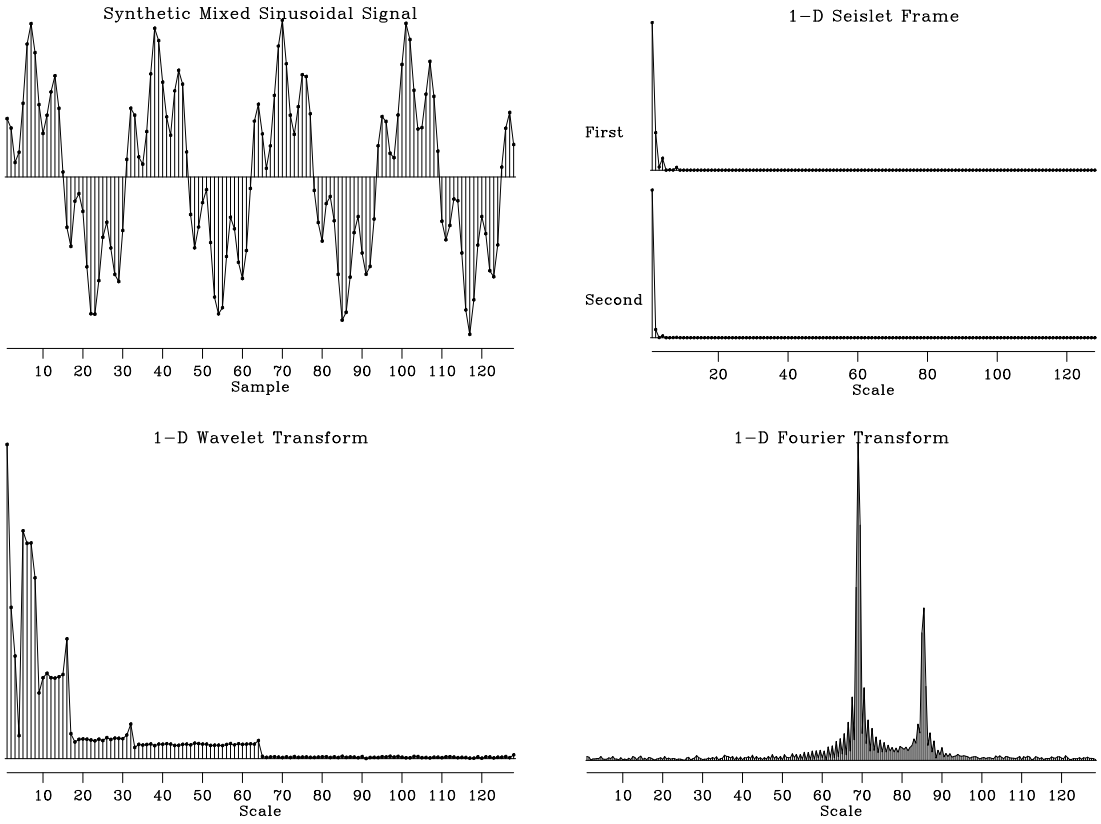


Figure 13: Mixed sinusoidal signal (a), 1-D seislet frame (b), 1-D wavelet transform (c), and 1-D Fourier transform (d). seislet/sin2 tatrace,taft,tdwt,fourier

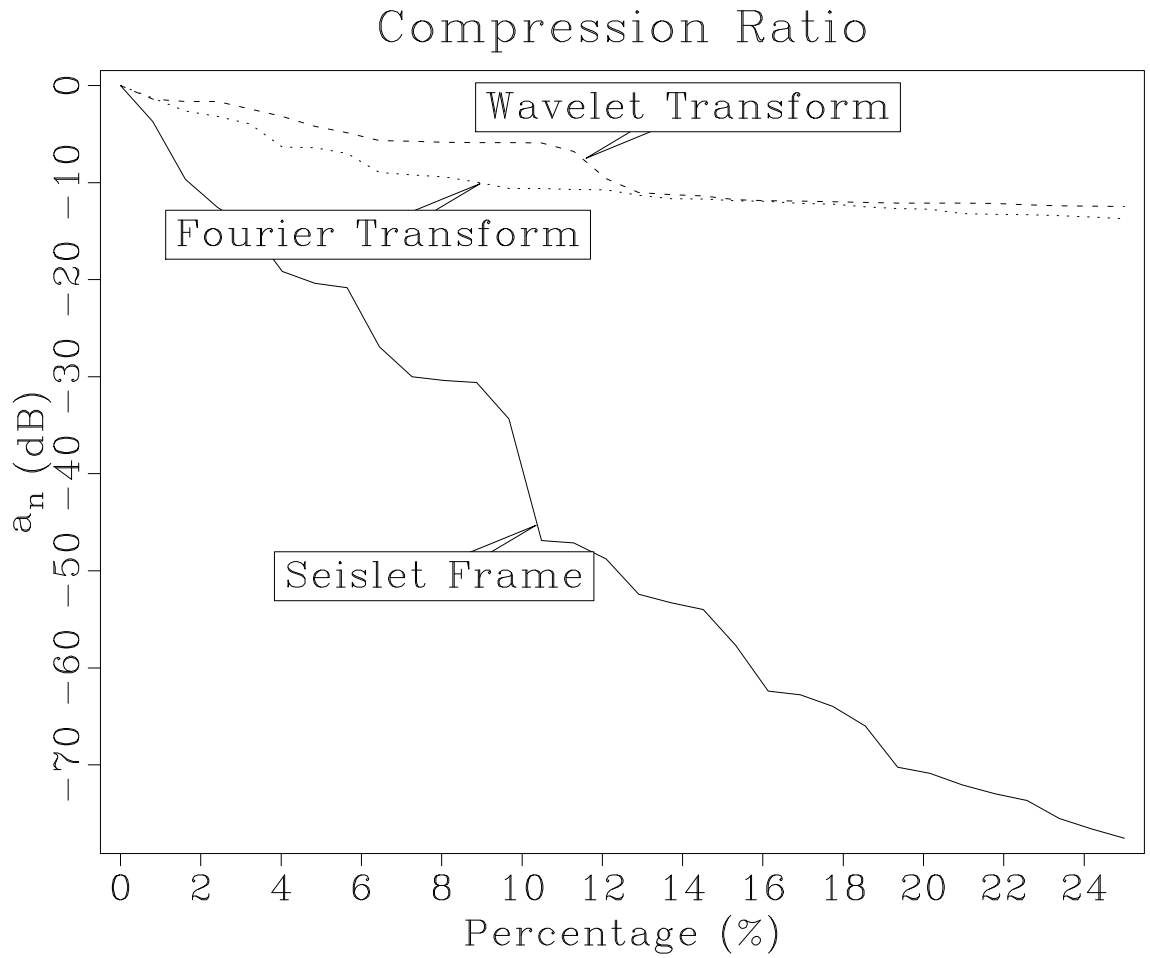


Figure 14: Compression comparison between digital Fourier transform, digital wavelet transform, and 1-D seislet frame. Transform coefficients are sorted from large to small, normalized, and plotted on a decibel scale. seislet/sin2 tlog

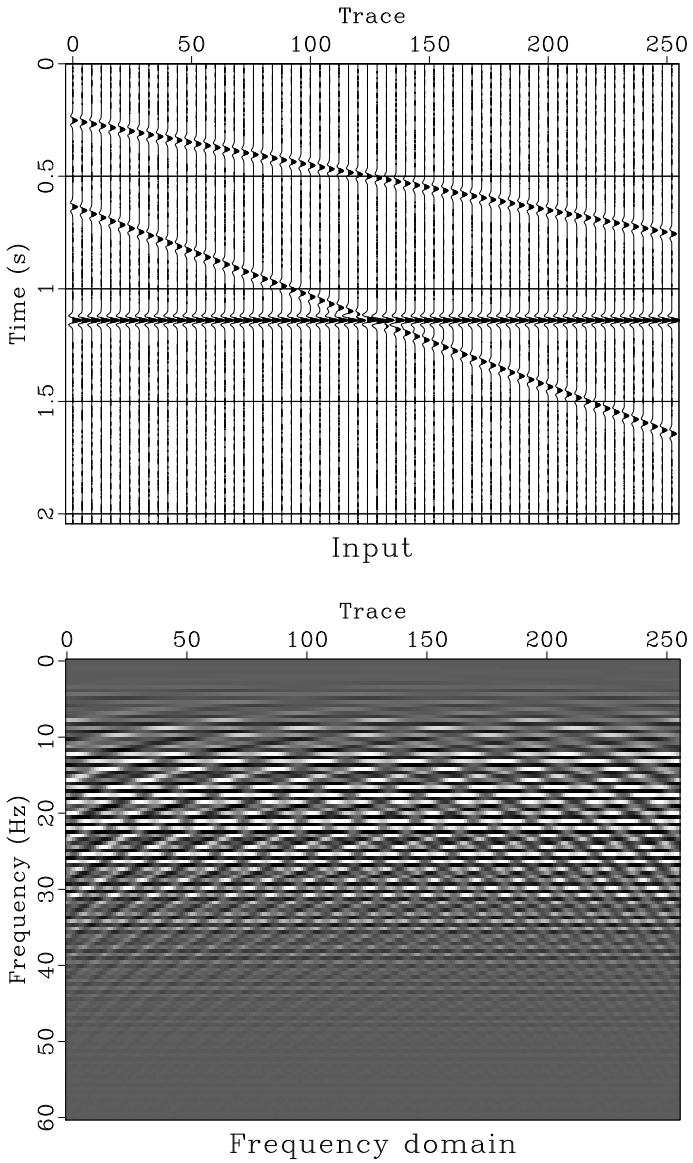


Figure 15: Synthetic plane-wave data (a) and corresponding Fourier transform along the time direction (b). seislet/plane plane,fft

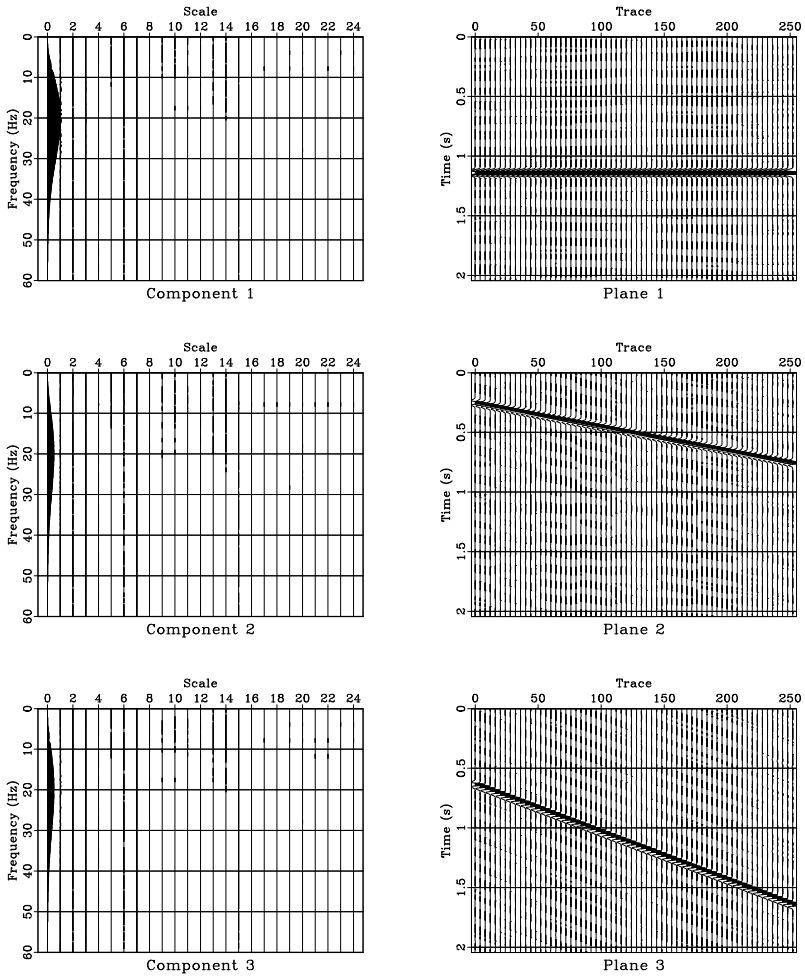


Figure 16: Seislet coefficients (left) and corresponding recovered plane-wave components (right) for three different parts of the 1-D seislet frame in the F-X domain. seislet/plane plane1,plane2,plane3

slant-stack (Radon transform) gather (Figure 18(a)). Figure 19(a) shows randomly selected example frame functions for the 2-D seislet frame using constant dips

Our second choice is a set of dip fields defined by the hyperbolic shape of seismic events on the CMP gather:

$$t(x) = \sqrt{t_0^2 + \frac{x^2}{v^2}}\ , \tag{17}$$

where $t(x)$ is traveltime for reflection at offset x, t_0 is the zero-offset traveltime, and v is the root-mean-square velocity. For a range of constant velocities, the direct relationship between dip and velocity is given by

$$p = \frac{dt}{dx} = \frac{x}{v^2 t}\ . \tag{18}$$

The dip field $p(x, t, v)$ is shown in Figure 17(b). Analogously to the case of constant dips, the frame coefficients at the zero scale correspond to the hyperbolic Radon transform (37), with the primary and multiple reflections distributed in different velocity ranges (Figure 18(b)). Figure 19(b) shows randomly selected frame functions for the 2-D seislet frame with varying dip fields defined by a range of constant velocities.

DISCUSSION

Data size	1-D FFT	1-D DWT	1-D seislet	2-D seislet
1024×1024	0.06	0.03	0.17	1.03
512×512	0.02	0.01	0.04	0.22

Table 1: CPU times (in seconds) for different transforms.

How efficient are the proposed algorithms? The CPU times, in our implementation, are shown in Table 1. They confirm that, while the seislet transform and frame can be more expensive than FFT or DWT, they are still comfortably efficient in practice. In applications of the 2-D seislet transform, the main cost may not be in the transform itself but in iterative estimation of the slope fields. In practical large-scale applications, it is advantageous to break the input data in parts and process them in parallel.

How effective are the seislet transform and frame in compressing seismic data? In the case of the 2-D seislet transform that requires a slope field, it appears that one would need to store this field in addition to the compressed data. However, since we force the estimated slopes to be smooth, the slope field can be easily compressed with one of the classic compression algorithms. Consider the example in Figure 2. Suppose that we apply lossy compression and require 99% of the energy to be preserved. The seislet transform compression ratio in this case is less that 1% while the corresponding wavelet transform ratio is 26%. Applied to the smooth slope field

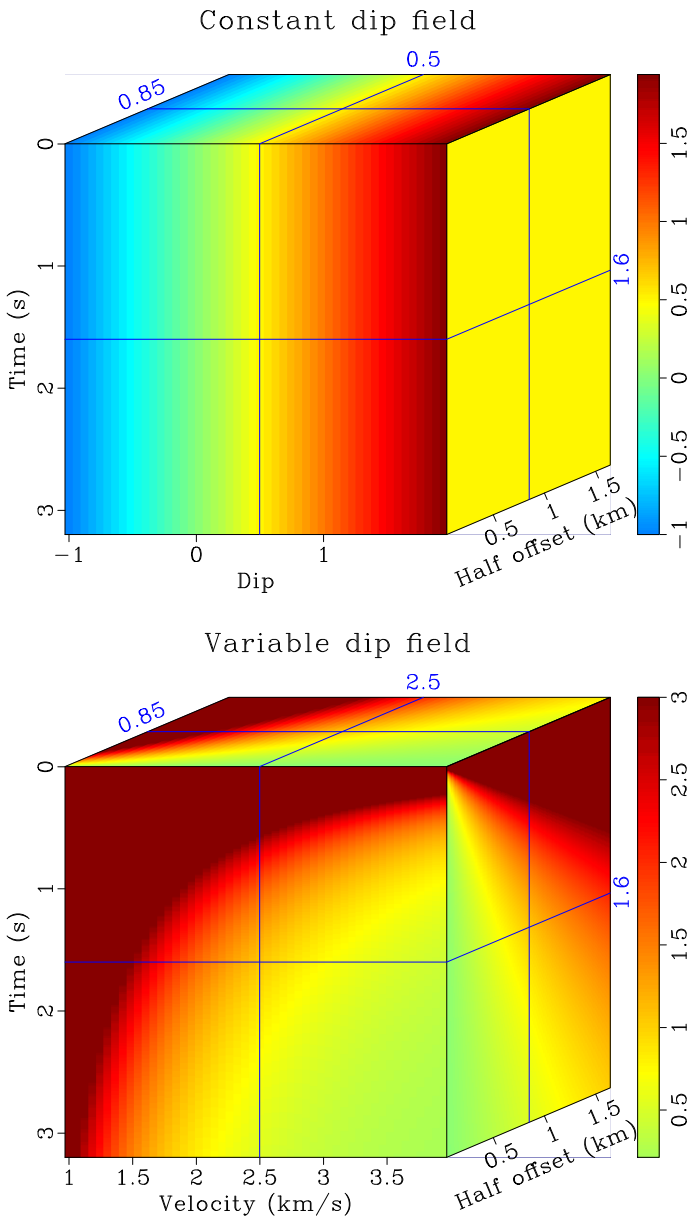


Figure 17: Constant dip field (a) and time and space varying dip field (b). seislet/diplet cdips,rrdips

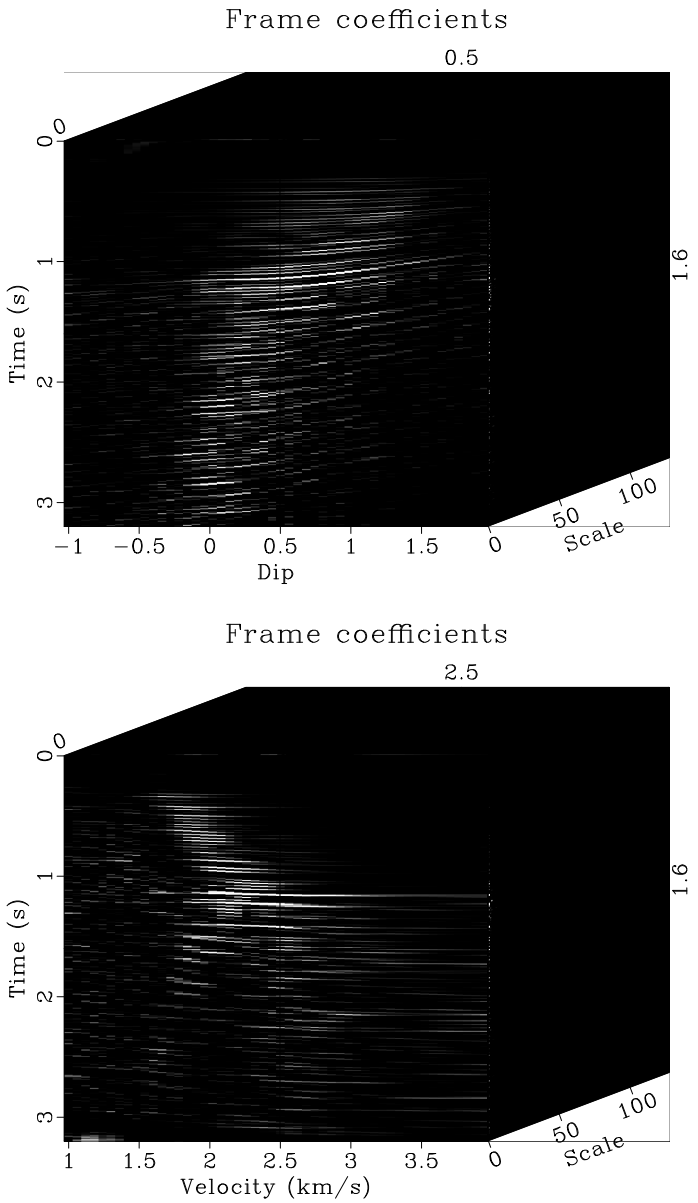


Figure 18: 2-D seislet frame coefficients with constant dip field (a) and with varying dip field (b). seislet/diplet cdiplet,rrdiplet

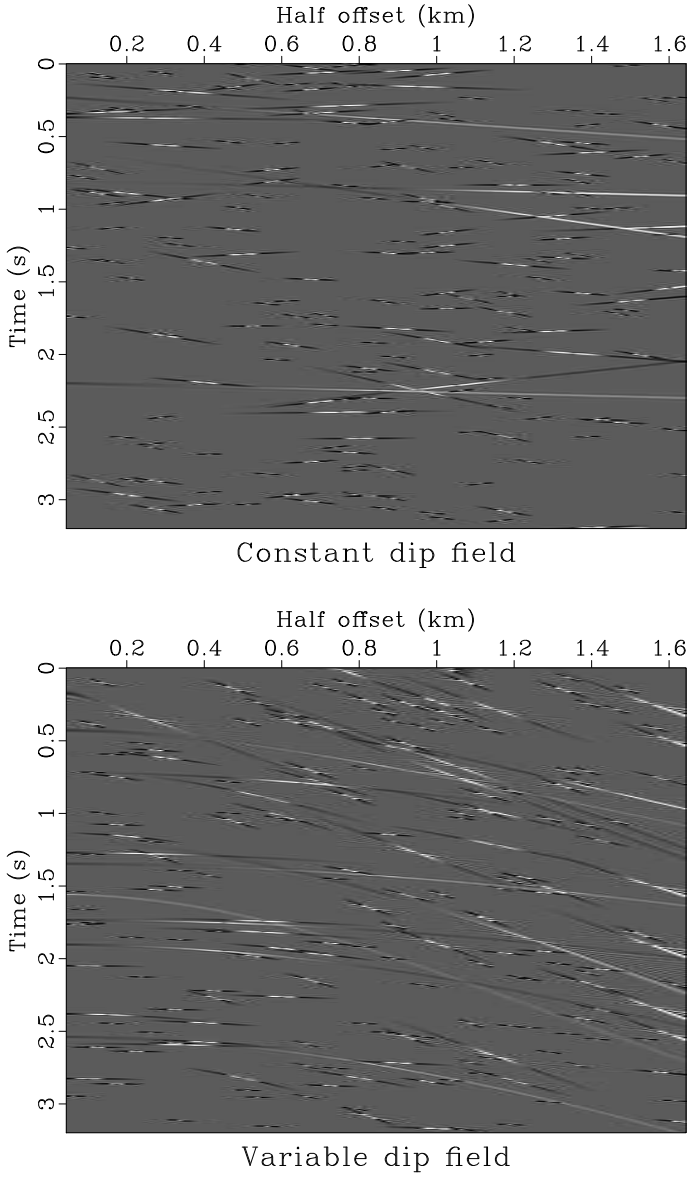


Figure 19: Randomly selected representative frame functions for 2-D seislet frame with constant dip field (a) and varying dip field (b). seislet/diplet cdipimps,rrdipimps

from Figure 2(b), the wavelet transform compresses it to about 0.1%. This example shows that compressing seismic data with the seislet transform and the corresponding slope field with the wavelet transform can be significantly more effective that trying to compress seismic data with the wavelet transform.

CONCLUSIONS

We have introduced a new digital transform named *seislet transform* because of its ability to characterize and compress seismic data in the manner similar to that of digital wavelet transforms. We define the seislet transform by combining the wavelet lifting scheme with local plane-wave destruction. In 1-D, the seislet transform follows sinusoidal components. In 2-D, it follows locally plane events. When more than one sinusoid or more than one local slope are applied for the analysis, the transform turns into an overcomplete representation or a frame. The seislet transform and seislet frame can achieve a better compression ratio than either the digital Fourier transform (DFT) or the digital wavelet transform (DWT).

The seislet transform provides a convenient orthogonal basis with the basis functions spanning different scales analogously to those of the digital wavelet transform but aligned along the dominant seismic events. Traditional signal analysis operations such as denoising and trace interpolation become simply defined in the seislet domain and allow for efficient algorithms. Seismic stacking also has a simple meaning of the zeroth-order seislet coefficient computed in an optimally efficient manner by recursive partial stacking and thus avoiding the usual problems with wavelet stretch and nonhyperbolic moveouts.

ACKNOWLEDGMENTS

We thank BGP Americas for a partial financial support of this work. The first author is grateful to Huub Douma for inspiring discussions and for suggesting the name “seislet”. This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

REVIEW OF PLANE-WAVE DESTRUCTION

This appendix reviews the basic theory of plane-wave destruction (16).

Following the physical model of local plane waves, we define the mathematical basis of plane-wave destruction filters via the local plane differential equation (7)

$$\frac{\partial P}{\partial x} + \sigma \frac{\partial P}{\partial t} = 0 \;, \tag{A-1}$$

where $P(t,x)$ is the wave field, and σ is the local slope, which may also depend on t and x. In the case of a constant slope, equation A-1 has the simple general solution

$$P(t,x) = f(t - \sigma x) \ , \tag{A-2}$$

where $f(t)$ is an arbitrary waveform. Equation A-2 is nothing more than a mathematical description of a plane wave.

If we assume that the slope σ does not depend on t, we can transform equation A-1 to the frequency domain, where it takes the form of the ordinary differential equation

$$\frac{d\hat{P}}{dx} + i\omega\,\sigma\,\hat{P} = 0 \tag{A-3}$$

and has the general solution

$$\hat{P}(x) = \hat{P}(0)\,e^{i\omega\,\sigma x} \ , \tag{A-4}$$

where $\hat{P}$ is the Fourier transform of P. The complex exponential term in equation A-4 simply represents a shift of a t-trace according to the slope σ and the trace separation x.

In the frequency domain, the operator for transforming the trace $x-1$ to the neighboring trace x is a multiplication by $e^{i\omega\,\sigma}$. In other words, a plane wave can be perfectly predicted by a two-term prediction-error filter in the F-X domain:

$$a_0\,\hat{P}(x) + a_1\,\hat{P}(x-1) = 0 \ , \tag{A-5}$$

where $a_0 = 1$ and $a_1 = -e^{i\omega\,\sigma}$. The goal of predicting several plane waves can be accomplished by cascading several two-term filters. In fact, any F-X prediction-error filter represented in the Z-transform notation as

$$A(Z_x) = 1 + a_1 Z_x + a_2 Z_x^2 + \cdots + a_N Z_x^N \tag{A-6}$$

can be factored into a product of two-term filters:

$$A(Z_x) = \left(1 - \frac{Z_x}{Z_1}\right)\left(1 - \frac{Z_x}{Z_2}\right)\cdots\left(1 - \frac{Z_x}{Z_N}\right) \ , \tag{A-7}$$

where $Z_1, Z_2, \ldots, Z_N$ are the zeroes of polynomial A-6. According to equation A-5, the phase of each zero corresponds to the slope of a local plane wave multiplied by the frequency. Zeroes that are not on the unit circle carry an additional amplitude gain not included in equation A-3.

In order to incorporate time-varying slopes, we need to return to the time domain and look for an appropriate analog of the phase-shift operator A-4 and the plane-prediction filter A-5. An important property of plane-wave propagation across different traces is that the total energy of the propagating wave stays invariant throughout the process: the energy of the wave at one trace is completely transmitted to the next

trace. This property is assured in the frequency-domain solution A-4 by the fact that the spectrum of the complex exponential $e^{i\omega\sigma}$ is equal to one. In the time domain, we can reach an equivalent effect by using an all-pass digital filter. In the Z-transform notation, convolution with an all-pass filter takes the form

$$\hat{P}_{x+1}(Z_t) = \hat{P}_x(Z_t)\frac{B(Z_t)}{B(1/Z_t)} \ , \tag{A-8}$$

where $\hat{P}_x(Z_t)$ denotes the Z-transform of the corresponding trace, and the ratio $B(Z_t)/B(1/Z_t)$ is an all-pass digital filter approximating the time-shift operator $e^{i\omega\sigma}$. In finite-difference terms, equation A-8 represents an implicit finite-difference scheme for solving equation A-1 with the initial conditions at a constant x. The coefficients of filter $B(Z_t)$ can be determined, for example, by fitting the filter frequency response at low frequencies to the response of the phase-shift operator. This leads to a version of Thiran's maximally-flat all-pass fractional-delay filters (36; 38).

Taking both dimensions into consideration, equation A-8 transforms to the prediction equation analogous to A-5 with the 2-D prediction filter

$$A(Z_t, Z_x) = 1 - Z_x\frac{B(Z_t)}{B(1/Z_t)} \ . \tag{A-9}$$

In order to characterize several plane waves, we can cascade several filters of the form A-9 in a manner similar to that of equation A-7. A modified version of the filter $A(Z_t, Z_x)$, namely the filter

$$C(Z_t, Z_x) = A(Z_t, Z_x)B(1/Z_t) = B(1/Z_t) - Z_xB(Z_t) \ , \tag{A-10}$$

avoids the need for polynomial division. In case of a 3-point filter $B(Z_t)$, the 2-D filter A-10 has exactly six coefficients. It consists of two columns, each column having three coefficients and the second column being a reversed copy of the first one.

REFERENCES

[1] Abma, R., and N. Kabir, 2006, 3D interpolation of irregular data with a POCS algorithm: Geophysics, **71**, E91–E97.

[2] Burg, J. P., 1975, Maximum entropy spectral analysis: PhD thesis, Stanford University.

[3] Canales, L. L., 1984, Random noise reduction: 54th Annual International Meeting, SEG, Expanded Abstracts, Session: S10.1.

[4] Candés, E. J., and L. Demanet, 2005, The curvelet representation of wave propagators is optimally sparse: Communications on Pure and Applied Mathematics, **58**, 1472–1528.

[5] Chauris, H., and T. Nguyen, 2008, Seismic demigration/migration in the curvelet domain: Geophysics, S35–S46.

[6] Claerbout, J. F., 1976, Fundamentals of geophysical data processing: Blackwell Scientific Publications.

[7] ———, 1992, Earth Soundings Analysis: Processing Versus Inversion: Blackwell Scientific Publications.
[8] ———, 2008, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.
[9] Cohen, A., I. Daubechies, and J. Feauveau, 1992, Biorthogonal bases of compactly supported wavelets: Communications on Pure and Applied Mathematics, **45**, 485–560.
[10] Daubechies, I., M. Defrise, and C. D. Mol, 2004, An iterative thresholding algorithm for linear inverse problems with a sparsity constraint: Communications on Pure and Applied Mathematics, **LVII**, 1413–1457.
[11] Dessing, F. J., 1997, A wavelet transform approach to seismic processing: PhD thesis, Delft University of Technology.
[12] Do, M. N., and M. Vetterli, 2005, The contourlet transform: an efficient directional multiresolution image representation: IEEE Transactions on Image Processing, **14**, 2091–2106.
[13] Donoho, D. L., 1995, De-noising by soft-thresholding: IEEE Trans. on Inform. Theory, **41**, 613–627.
[14] Douma, H., and M. V. de Hoop, 2007, Leading-order seismic imaging using curvelets: Geophysics, **72**, S231–S248.
[15] Edelman, A., and H. Murakami, 1995, Polynomial roots from companion matrix eigenvalues: Mathematics of Computation, **64**, 763–776.
[16] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.
[17] ———, 2006, Towards the seislet transform, *in* 76th Ann. Internat. Mtg: Soc. of Expl. Geophys., 2847–2850.
[18] ———, 2008, Nonlinear shaping regularization in geophysical inverse problems: 78th Annual International Meeting, SEG, Expanded Abstracts, 2046–2051.
[19] Fomel, S., and V. Grechka, 2001, Nonhyperbolic reflection moveout of P waves. An overview and comparison of reasons: Technical Report CWP-372, Colorado School of Mines.
[20] Foster, D. J., C. C. Mosher, and S. Hassanzadeh, 1994, Wavelet transform methods for geophysical applications: 64th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1465–1468.
[21] Guo, K., and D. Labate, 2007, Optimally sparse multidimensional representation using shearlets: SIAM Journal on Mathematical Analysis, **39**, 298–318.
[22] Haldorsen, J. B. U., and P. A. Farmer, 1989, Resolution and NMO-stretch: Imaging by stacking: Geophysical Prospecting, **37**, 479–492.
[23] Herrmann, F. J., and G. Hennenfent, 2008, Non-parametric seismic data recovery with curvelet frames: Geophysical Journal International, **173**, 233–248.
[24] Herrmann, F. J., D. Wang, G. Hennenfent, and P. Moghaddam, 2007, Curvelet-based seismic data processing: a multiscale and nonlinear approach: Geophysics, **73**, A1–A5.
[25] Jensen, A., and A. la Cour-Harbo, 2001, Ripples in mathematics: the discrete wavelet transform: Springer.
[26] Kazemeini, S. H., C. Juhlin, K. Zinck-Jorgensen, and B. Norden, 2009, Appli-

cation of the continuous wavelet transform on seismic data for mapping of channel deposits and gas detection at the CO2 site, Ketzin, Germany: Geophysical Prospecting, **57**, 111–123.

[27] Liu, Y., and S. Fomel, 2009, Oc-seislet: seislet transform construction with differential offset continuation, *in* 79th Ann. Internat. Mtg: Soc. of Expl. Geophys., 3228–3232.

[28] Mallat, S., 2009, A wavelet tour of signal processing: The sparse way: Academic Press.

[29] Marple, S. L., 1987, Digital spectral analysis with applications: Prentice Hall.

[30] Morlet, J., 1981, Sampling theory and wave propagation: 51st Ann. Internat. Mtg, Soc. of Expl. Geophys., Session:S15.1.

[31] Osher, S., M. Burger, D. Goldfarb, J. Xu, and W. Yin, 2005, An iterative regularization method for total variation-based image restoration: Multiscale Modeling & Simulation, **4**, 460–489.

[32] Pennec, E. L., and S. Mallat, 2005, Sparse geometrical image representation with bandelets: IEEE Transactions on Image Processing, **14**, 423–438.

[33] Starck, J. L., E. J. Candés, and D. L. Donoho, 2002, The curvelet transform for image denoising: IEEE Transactions on Image Processing, **11**, 670–684.

[34] Sweldens, W., 1995, The lifting scheme: A new philosophy in biorthogonal wavelet constructions: Wavelet Applications in Signal and Image Processing III, Proc. SPIE 2569, 68–79.

[35] Sweldens, W., and P. Schröder, 1996, Building your own wavelets at home, *in* Wavelets in Computer Graphics: ACM SIGGRAPH Course notes, 15–87.

[36] Thiran, J.-P., 1971, Recursive digital filters with maximally flat group delay: IEEE Transactions on Circuit Theory, **18**, 659–664.

[37] Thorson, J. R., and J. F. Claerbout, 1985, Velocity stack and slant stack stochastic inversion: Geophysics, **50**, 2727–2741.

[38] Välimäki, V., and T. T. Laakso, 2001, Fractional delay filters – design and applications, *in* Nonuniform sampling: theory and practice: Kluwer Academic/Plenum Publishers, 835–895.

[39] Velisavljevic, V., 2005, Directionlets: anisotropic multi-directional representation with separable filtering: PhD thesis, Ecole Polytechnique Fédérale de Lausanne.

[40] Wapenaar, K., R. Ghose, G. Toxopeus, and J. Fokkema, 2005, The wavelet transform as a tool for geophysical data integration: Integrated Computer-Aided Engineering, **12**, 5–23.

[41] Welland, G., ed., 2003, Beyond wavelets: Academic Press.

[42] Xu, S., Y. Zhang, D. Pham, and G. Lambaré, 2005, Antileakage Fourier transform for seismic data regularization: Geophysics, **70**, V87–V95.

[43] Yin, W., S. Osher, D. Goldfarb, and J. Darbon, 2008, Bregman interative algorithoms for l_1-minimization with application to compressed sensing: SIAM J. Imaging Sciences, **1**, 143–168.

[44] Zwartjes, P., and A. Gisolf, 2006, Fourier reconstruction of marine-streamer data in four spatial coordinates: **71**, V171–V186.

[45] ———, 2007, Fourier reconstruction with sparse inversion: Geophysical Prospecting, **55**, 199–221.

[46] Zwartjes, P., and M. Sacchi, 2007, Fourier reconstruction of nonuniformly sampled, aliased seismic data: **72**, V21–V32.

3D velocity-independent elliptically-anisotropic moveout correction

*William Burnett and Sergey Fomel**

ABSTRACT

Azimuthal anisotropy or lateral velocity variations cause azimuthal variations in moveout velocity which can lead to seismic image degradation if not properly handled. In cases where apparent azimuthally anisotropic moveout is present, a single picked velocity is inadequate to flatten an event on a 3D CMP gather. Conventional velocity analysis techniques require a significant amount of time and effort, especially in areas where apparent anisotropy is observed. We propose a velocity-independent imaging approach to perform an elliptically anisotropic moveout correction in 3D. The velocity-independent approach relies on volumetric local traveltime slopes rather than aggregate velocities, and therefore provides an azimuthally flexible description of traveltime geometries throughout the gather. We derive theoretical expressions for extracting the moveout slowness matrix and the angle between the symmetry and acquisition axes as volumetric local attributes. A practical inversion scheme to extract the same parameters is also developed. These parameters are used to solve for moveout slowness as a function of azimuth. Tests on a synthetic CMP gather show accurate results for the automatic moveout correction and the inversion scheme. A field data example from West Texas illustrates the application of the automatic moveout correction as a residual moveout.

INTRODUCTION

Common geological occurrences such as dipping interfaces, lateral velocity variations, or HTI media can lead to real or apparent azimuthal anisotropy, in which case the P-wave moveout velocity becomes elliptically dependent on azimuth (4). The symmetry axes of apparent azimuthal anisotropy often correspond to geologically meaningful parameters such as the strike and dip directions of the reflector (8),the directions of tectonic stress (12), the preferred orientation of vertical fractures (1), or any combination of these factors. Failure to account for azimuthal velocity variations often leads to stack degradation, improper time-to-depth conversion, inaccurate AVO/AVOA, and overall poorer image results (16). (11) and (14) demonstrate that migration algorithms which can handle azimuthal variations in the velocity model can visibly improve seismic imaging results.

*e-mail: sergey.fomel@beg.utexas.edu

Conventional manual velocity analysis procedures take up a significant part of the time needed to process seismic data. Even with semi-automated picking tools, this phase of a typical processing flow alone may take weeks or even months for modern 3D data sets. Accurate automated traveltime picking algorithms are the main tools for modern velocity analysis, and have greatly reduced the time and manual work required to hand-pick velocities (13). However, these tools still require significant manual inspection and editing for quality control.

The conventional production processing flow does not include picking azimuthally-dependent velocities, but two approaches are commonly used to handle and characterize azimuthal variations in velocity. The first, and historically more popular approach, is to sort CMP gathers into azimuth sectors, and then perform isotropic velocity analysis, processing, and migration on each sector. The individual moveout parameters from all sectors are plotted together, and then fit with a sinusoid to characterize the principal moveout directions and the percentage of anisotropy. (6) describe another approach, where NMO is first performed with a smooth global velocity model. If apparent anisotropy is detected, trace-to-trace traveltime shifts are estimated automatically, and the traveltime surface is fit with an ellipse characterized by the moveout slowness matrix $\mathbf{W}$. The second approach has become more popular in production because of its robustness, and in a case-study comparing the two, (9) provides an example where the non-sectoring approach yielded a more reliable azimuthal velocity model.

The concept of velocity-independent imaging (10) is attractive because it can be very efficient when compared the time and manual work required to hand-pick velocities (3). The underlying strategy of velocity-independent imaging relies on measuring traveltime slopes throughout the data set rather than hyperbolic traveltimes or velocities themselves (17). (2) demonstrates that plane-wave destruction filters provide an automated and effective way to measure local slopes in a seismic volume. Measured slopes can then be used to automate any common time-domain imaging step (3). Previous work concerning automatic moveout corrections does not extend to the 3D case. In doing so here, we demonstrate that the azimuthal flexibility of automatic moveout correction in 3D is especially useful in the presence of real or apparent azimuthal anisotropy.

Rather than using a single picked velocity profile to apply the NMO correction, using the local slopes of a given 3D reflection event allows the event to be flattened regardless of azimuthal variations in NMO velocity. In practice, these slopes can be measured automatically throughout the volume, so no traveltime surfaces need to be picked. The velocity-independent approach can still be used to extract moveout or interval velocities throughout the data set as data attributes (3). Our method also suggests that, by measuring local curvatures throughout the seismic data volume, the orientation of the symmetry axes can automatically be estimated with respect to the acquisition coordinates. We present theoretical expressions for azimuthal anisotropy moveout parameters as volumetric attributes, and demonstrate a practical inversion scheme for the same parameters using the velocity-independent approach. Synthetic

and field examples are used to validate our proposed method and show the variety of potential applications.

THEORY

Following (4), the elliptical NMO equation in 3D can be written with the help of a truncated Taylor series expansion as,

$$t_0 = \sqrt{t^2 - (W_x x^2 + W_y y^2 + 2W_{xy} xy)}, \tag{1}$$

where t is the event arrival time, t_0 is the moveout corrected time, and x and y are the components of full offset in the x and y survey directions, respectively. W_x and W_y are the conventionally-measured moveout slownesses squared values (along the same survey coordinates). Equation 1 describes NMO with elliptical velocity, where the third parameter, W_{xy}, arises from observing the ellipse from rotated coordinates. In practice, one can perform elliptically anisotropic NMO using conventional velocity picking in the inline and crossline directions, but the principal directions of moveout must also be estimated. Rewriting equation 1 as a matrix-vector multiplication between the offset vector and the slowness matrix $\mathbf{W}$ allows one to solve for the angle between the acquisition coordinates and the medium symmetry axes, denoted as α here. This can be done either by finding the eigenvectors of the system (4), or by using geometric arguments and well-known relations between the formulas for a rotated ellipse and its unrotated equivalent (15). Using the latter approach gives an expression for α in terms of conventional slowness parameters,

$$\alpha = \frac{1}{2} \tan^{-1} \left(\frac{2W_{xy}}{W_x - W_y} \right). \tag{2}$$

In this expression, α is the angle from a survey axis measured counter-clockwise toward the nearest symmetry axis. If W_x is equal to W_y, then the arc-tangent argument goes to infinity, corresponding to $\alpha = 45°$. Although equation 2 is a straightforward way of finding the coordinate rotation angle, finding the eigenvalues and eigenvectors allows one to resolve between the fast and slow principal moveout directions. The eigenvalues λ_1 and λ_2 of the slowness matrix,

$$\mathbf{W} = \begin{pmatrix} W_x & W_{xy} \\ W_{xy} & W_y \end{pmatrix}, \tag{3}$$

can be found following (4). Rewritten here in our notation, they demonstrate how the eigenvalues,

$$\lambda_{1,2} = \frac{1}{2} \left[W_x + W_y \pm \sqrt{(W_x - W_y)^2 + 4W_{xy}^2} \right], \tag{4}$$

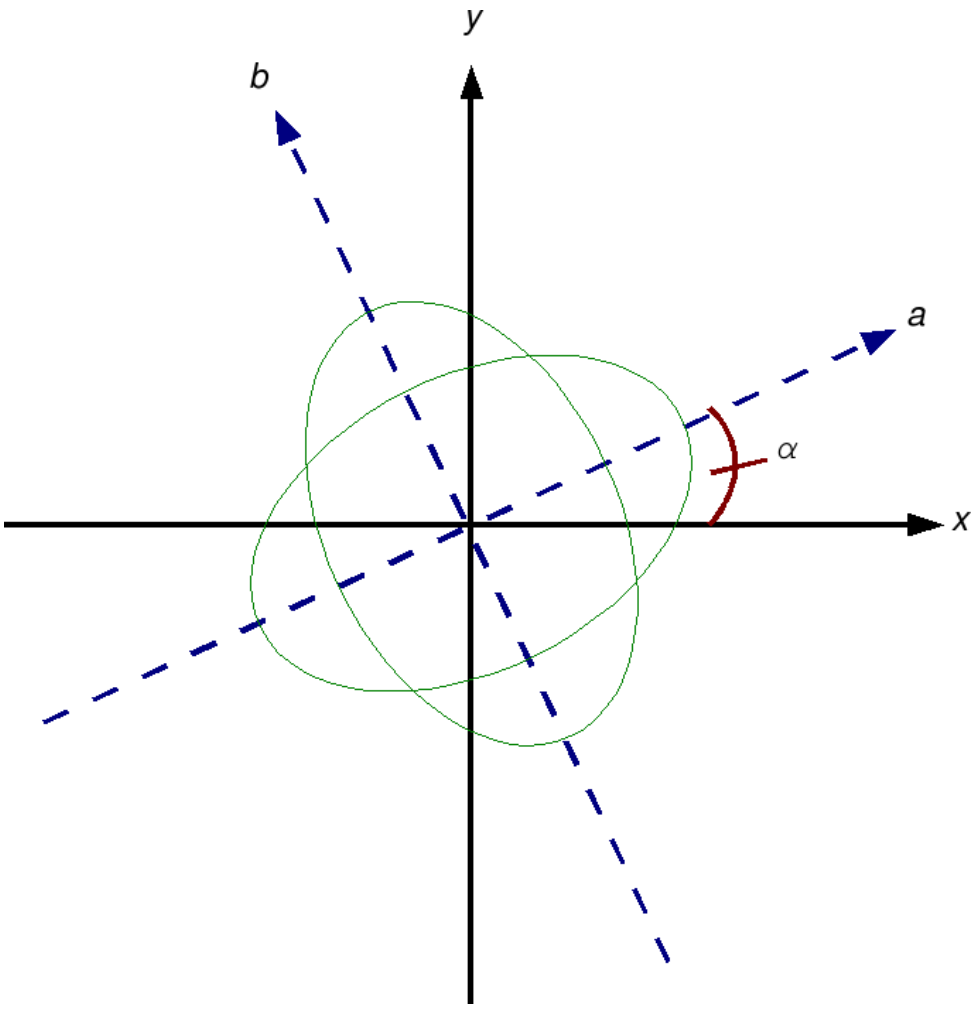


Figure 1: Elliptic NMO velocity depends on the natural symmetry axes (a-b), not the acquisition coordinates (x-y). The angle α describes the counter-clockwise rotation between the two coordinate frames. nmo3/. HTINMO

can be used together with α to solve for the NMO slowness S as a function of source-receiver azimuth θ:

$$S^2_{nmo}(\theta) = \lambda_1 \cos^2(\theta - \alpha) + \lambda_2 \sin^2(\theta - \alpha). \tag{5}$$

Since equation 1 describes the predicted hyperbolic traveltime curve on a CMP gather for a variety of cases where either real or apparent elliptical anisotropy may be present, the local slope of an event on an inline or crossline CMP gather can be related to the conventional moveout slowness parameters by taking the derivative of 1 with respect to x and y. Ignoring higher order terms and assuming the parameters vary slowly along x and y, gives a first-order approximation of how the measured slopes relate to conventional moveout parameters:

$$p_x(t,x,y) = \frac{\partial t}{\partial x} = \frac{W_x x + W_{xy} y}{t}, \tag{6}$$

$$p_y(t,x,y) = \frac{\partial t}{\partial y} = \frac{W_y y + W_{xy} x}{t}. \tag{7}$$

By substitution back into 1, we arrive at the velocity-independent expression for 3D elliptical moveout in terms of local slopes:

$$t_0 = \sqrt{t^2 - t\,(p_x x + p_y y)}. \tag{8}$$

Notice that only two parameters (p_x and p_y) must be measured to completely predict the NMO corrected time. More importantly, these parameters can be measured automatically using a local slope estimation algorithm, such as plane-wave destruction (2). Equation 8 is a 3D extension for the 2D equation from (10).

Automated processes allow one to save time spent on a project, but it may seem that the insight and information gained during a more interactive conventional processing flow would be lost. A significant part of production velocity analysis involves picking or examining the velocity model directly, which provides an early and intuitive link between the seismic data and the subsurface geology. The velocity model and anisotropy information are themselves invaluable sources of geologic information. They also control the positioning of events in the final image, so an ability to extract these parameters is desirable.

The relation between local slopes and moveout velocity has been documented for the 2D case (10; 17; 3). In the 3D case where apparent azimuthal anisotropy is present, at least three conventional slowness or velocity-like values are needed to characterize moveout (W_x, W_y, and W_{xy} from equation 1). Although equation 8 suggests they are not necessary for moveout in terms of local slopes, these values can be used to characterize anisotropy, and may also be useful for other subsequent processing. Simply rearranging equations 3 and 7 gives expressions for W_x and W_y:

$$W_x = \frac{tp_x - W_{xy}y}{x}, \tag{9}$$

and,

$$W_y = \frac{tp_y - W_{xy}x}{y}. \tag{10}$$

Both of these parameters require an estimate of W_{xy}. A first-order approximation of W_{xy} can be found by differentiating equation 3 with respect to y or equation 7 with respect to x:

$$W_{xy} = t\frac{\partial p_x}{\partial y} + p_x p_y = t\frac{\partial p_y}{\partial x} + p_x p_y = tp_{xy} + p_x p_y. \tag{11}$$

Since slopes are measured as a local attribute, the inline and crossline local slopes comprise data volumes with the same dimensions and coordinates as the input CMP. Applying a 1D derivative filter to these volumes allows one to obtain either mixed-derivative in equation 11, and solve for the apparent anisotropy angle α, using equation 1. This angle can also be expressed in terms of local slopes. Combining equations 9, 10, and 11 yields,

$$W_x - W_y = \frac{t}{xy}[yp_x - xp_y + (p_{xy} + (x^2 - y^2)\frac{p_x p_y}{t})]. \tag{12}$$

Now everything needed to express α independently of velocity is found in equations 11 and 12. Combining them with equation 2 gives,

$$\alpha(t,x,y) = \frac{1}{2}\tan^{-1}\left(\frac{2xy(tp_{xy} + p_x p_y)}{t(yp_x - xp_y) + (tp_{xy} + p_x p_y)(x^2 - y^2)}\right). \tag{13}$$

Implementing equation A-8 creates an attribute for each input data sample describing the counter-clockwise azimuthal angle between the symmetry coordinates and the acquisition coordinates. Applying NMO to this attribute volume yields $\alpha(t_0,x,y)$, which should then theoretically be constant at each time-slice if the moveout were exactly described by equation 1.

Finding local estimates of slowness and anisotropy parameters using equations 9-A-8 remains at this point only an interesting theoretical idea. A more robust and practical approach to extracting velocity and anisotropy parameters is to exploit the shear number of volumetric slope measurements made to perform the velocity-independent NMO correction. For a given CMP with dimensions $(n_x \times n_y \times n_t)$, the NMO correction applies a shift of time-squared,

$$t_\Delta(t_0,x,y) = t^2(x,y) - t_0^2(x,y), \tag{14}$$

which can be automatically computed for every output coordinate using equation (8) and stored as another volume of the same dimensions. Once NMO is applied, the time axis of the CMP gather represents t_0, so the slowness matrix $\mathbf{W}$ and α should each be constant for a given time value. Each time-slice from either the data or one of the attribute volumes can be viewed as an $(n_x \times n_y)$ matrix, which can be re-indexed into a vector of length $(n_x \times n_y)$. If the x and y indexes from the time-slice are i and j respectively, then the value from position (i, j) in the matrix is mapped to the $k = i + jn_x$ position in the vector. Using this notation, a highly overdetermined problem follows from writing equation 1 as a matrix-vector multiplication:

$$\mathbf{t}_\Delta = \mathbf{X}\mathbf{w}. \tag{15}$$

where the k^{th} element of $\mathbf{t}_\Delta$ is,

$$t_{\Delta k} = t_\Delta(t_0, x_i, y_j), \tag{16}$$

the k^{th} row of $\mathbf{X}$ is given by the vector,

$$\mathbf{x}_k = \begin{pmatrix} x_i^2 & y_j^2 & 2x_i y_j \end{pmatrix}, \tag{17}$$

and

$$\mathbf{w} = \begin{pmatrix} W_x \\ W_y \\ W_{xy} \end{pmatrix}. \tag{18}$$

Linear system 15 has $(n_x \times n_y)$ equations with only three unknowns. By solving 15 for each time-slice in the output CMP, we construct the slowness matrix $\mathbf{W}(t_0)$, and use it with equations A-8, 3, and 5 to extract the coordinate rotation angle $\alpha(t_0)$ and the NMO slowness as a function of azimuth, $S(t_0, \theta)$.

EXAMPLES

We provide two examples to illustrate the performance of our approach. In the first example, we consider a simple 3D synthetic CMP gather (Figure 2(a)) with four events, each with a different degree of apparent azimuthal anisotropy. The synthetic CMP in Figure 2(a) was created by first specifying the moveout slowness matrix, $\mathbf{W}$ for each event. Each of the four events was modeled individually by applying inverse 3D NMO to a flat reflection based on equation 1. The exact parameters used to model the four events are specified in Table 1. The four events were then added together into a single CMP gather with a small amount of random noise (10% of the signal amplitude). The result of this approach differs from real cases in that the

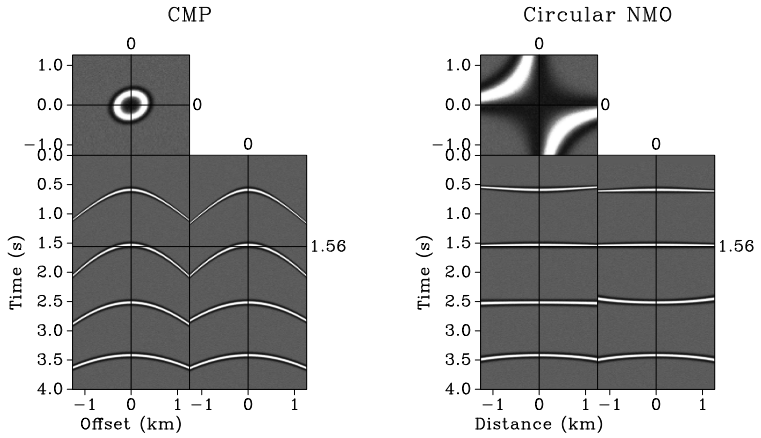


Figure 2: (a.) A synthetic 3D CMP gather with four events of varying apparent elliptical anisotropy. The three panels in the display show a time-slice view (upper square panel), a crossline view (central panel), and an inline view (right panel) of the same volume. (b.) An isotropic NMO correction using a picked velocity function appropriate for flattening certain events. At best, isotropic NMO can flatten either the inline or crossline directions well, but there is no single velocity function that will flatten both. nmo3/azimuthtest cmp3d,nmo063d

traveltime surface for each of the events is completely independent from overlying events. However, this approach allows us to specify the exact moveout slownesses of each the events without additional work.

Event Moveout Parameters					
Event	t_0 (s)	$W_x(s^2/km^2)$	$W_y(s^2/km^2)$	$W_{xy}(s^2/km^2)$	$\alpha(^\circ)$
A	0.59	0.14	0.16	-0.01	14.0
B	1.53	0.30	0.30	-0.04	-44.3
C	2.51	0.32	0.26	-0.03	-21.8
D	3.41	0.24	0.25	-0.005	26.57

Table 1: Moveout parameters used for events in Figure 2(a).

Conventional velocity semblance scans may yield multiple peaks for the same event when apparent azimuthal anisotropy is present. One must interpret the correct velocity in these areas, which can lead to inconsistent results between different processing geophysicists, since the cause of multiple semblance panel peaks can be ambiguous (multiples, noise, anisotropy, etc.). Resolving this ambiguity properly can add even more time and work to the production approach. Figure 2(b) shows a possible result of picking a single velocity profile which flattens certain events in either the inline or crossline view. For a root-mean-square velocity model, if an intermediate value is chosen for each event, both directions are flattened poorly. Event B has a particularly difficult scenario for the production approach; the symmetry axes are nearly 45° from

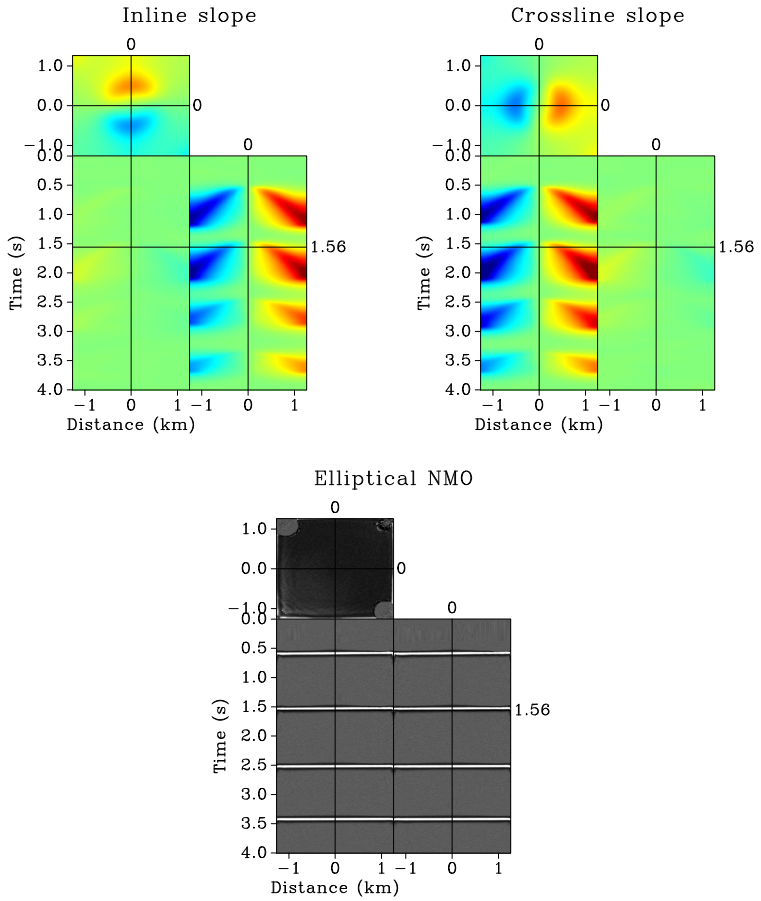


Figure 3: The (a.) inline and (b.) crossline slopes of the CMP gather from Figure 2(a). (c.) These slopes are used with equation 8 to automatically perform the proposed elliptically anisotropic moveout correction. All four events are flattened perfectly where the slopes are not aliased. nmo3/azimuthtest pxsmooth3d,pysmooth3d,PNMO3d

the acquisition axes, which makes the apparent moveout velocities along the x and y axes practically equal. A production velocity analysis is likely to not even detect the anisotropy in this case, because viewed from the acquisition axes, event B appears isotropic. The time-slice panel of Figure 2(b) reveals the poor performance of the isotropic NMO correction along other source-receiver azimuths of event B.

By measuring the local slopes of an input CMP gather as a volumetric attribute, the geometry of each traveltime surface is captured, even away from the x and y acquisition axes. Figures 3(a) and 3(b) show sections of the automatically measured inline and crossline slope volumes for the CMP from Figure 2(a). The time-slice views clarify that the slopes are measured in the x and y directions throughout the volume, not just along the x and y zero-offset axes. Comparison of the slope volumes with Figure 2(a) also shows that there are clearly non-zero slopes in areas without data. No initial slope fields were used to get these measurements, but the results of the plane-wave destruction filter application were regularized with a smoothing constraint by using shaping regularization (3). This constraint is enforced to help ensure that the moveout correction varies both spatially and temporally in a stable fashion. In Figure 3(c), the velocity-independent elliptically anisotropic moveout correction is applied using these slopes, and all of the events are flattened well in both directions. The time-slice view of Figure 3(c) now shows the overall superior performance of the velocity-independent correction, but also reveals its limitations. As events become steep relative to the trace spacing, local slope measurements can be aliased. Towards the corners of the example gather, the slopes become too steep to be measured reliably, and in these areas, the automatic moveout correction performs poorly. In field data, crossline trace spacing is often much coarser than inline spacing, which may lead to similar aliasing problems. However, the effects of aliasing can often be mitigated with a few simple extra steps. By first applying a constant velocity isotropic NMO correction to the data before measuring slopes, the events will be flatter and less likely to have aliased slopes at far offsets. An inverse NMO correction using the constant velocity can then be applied to the measured slopes. The constant velocity can then be converted to p_x and p_y components and added to the slope measurements to obtain the unaliased slope fields of the input CMP gather.

Another perspective of the same test is shown in Figures 4(a) and 4(b). The traces from the synthetic CMP gather have been binned into offset and azimuth coordinates to display the familiar sinusoidal signature of azimuthal traveltime variations. The various squared traveltime shifts (t_Δ values) applied by the automatic NMO correction were computed during implementation using equation 8, and then stored as a volumetric attribute. For each of the events, time-slices of this volume are shown in Figures 5(a)-5(d). The elliptical variation is clearly displayed for events A, B, and C, but the subtle variation for event D makes it difficult to detect the apparent anisotropy. Each time-slice from the same volume was re-indexed into a vector following the scheme described in the previous section, and then fed into a least-squares solver for equation 15 to yield $\mathbf{W}(t_0)$. The results of this inversion are displayed in Figures 6(a)-6(b). At the times of the events, all three parameters have been extracted accurately. The shaping regularization used to create smooth slope

fields also leads to similar smoothness in the **W** estimates. Because of the random noise in the synthetic data, slope measurements away from events are also random. The best-fit surface through these random slopes tends to be a flat plane, which is characterized on a CMP gather by zero slowness, causing the W_x and W_y estimates to tend toward zero between events. It is important to note that the values shown in Figures 6(a)-6(d) each rely on the accuracy of the NMO shifts computed at the corresponding value of t_0. Only the sparse times of this synthetic CMP gather with data have meaningful slope estimates and therefore meaningful **W** and α estimates.

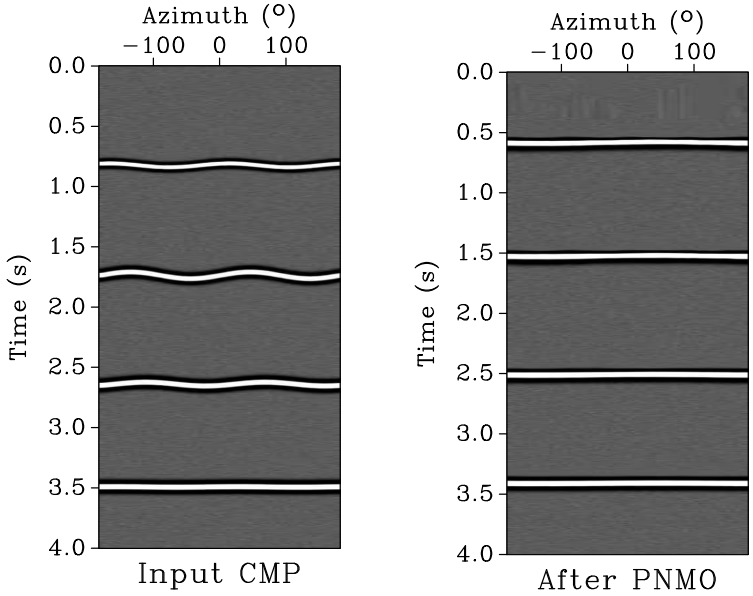


Figure 4: (a.) A common offset (0.75 km) display of CMP from Figure 2(a) with azimuth on the horizontal axis. (b.) The same traces after the automatic moveout correction. All events are shifted up to their appropriate t_0 and flattened. nmo3/azimuthtest oacmp,oaPNMO

The extracted moveout slowness matrices $\mathbf{W}(t_0)$ are used with equation 2 to estimate $\alpha(t_0)$. The results of estimating $\alpha(t_0)$ are displayed in Figure 6(d), and the values at the times of each event are accurate. We conclude this example by solving for NMO slowness-squared as using $\alpha(t_0)$ and the corresponding eigenvalues at each t_0 with equation 5. These results, shown in Figures 7(a)-7(d) show that for all four events, the angle of anisotropy is detected. From this volume, the principal moveout directions are readily determined, the fast and slow axes are resolved, and the apparent anisotropy (the ratio or difference of V_{fast} and V_{slow}) is measurable.

We next demonstrate a potential application of the automatic moveout correction on a real data example. In this case, we apply the automatic elliptically anisotropic moveout correction as a residual moveout correction. A subset of the McElroy data set from West Texas was formed into a supergather seen in Figure 8. Figure **??** shows a common offset (3.6-4.0 km) time-versus-azimuth display of the same data, where azimuthal anisotropy is evident for several events. In Figure **??**, the magnitude of the

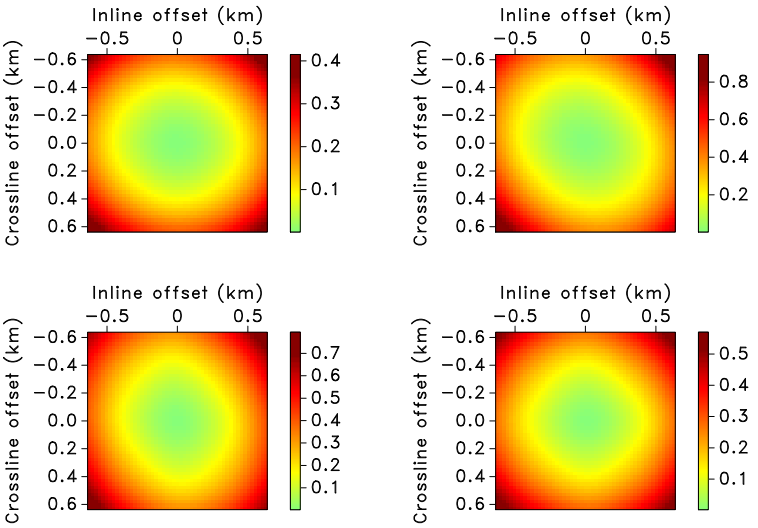


Figure 5: Traveltime squared shifts $(t_\Delta(t_0, x, y))$ for each event. (a.) Event A. (b.) Event B. (c.) Event C. (d.) Event D. nmo3/azimuthtest deltaTslice1,deltaTslice2,deltaTslice3,deltaTslice4

local slope is shown for the initial data. The areas with higher slope values highlight areas that were not ideally flattened by the prior isotropic NMO correction. The region of the highest slopes along the top of the figure is due to the proximity of the prior NMO mute at about 0.8 s. Figure **??** shows the results of applying two iterations of the proposed moveout correction. Further iterations will continue to flatten later events as distortions from overlying layers are removed. The events in the results are already noticeably flatter, and will therefore produce a cleaner stack.

DISCUSSION

Many advancements have been made in semi-automated traveltime picking schemes which have made the velocity analysis phase of a conventional seismic data processing flow much more efficient. However, a great deal of time is still required to manually check the quality of the assisted picking, and this remains as a time-consuming step in the conventional processing flow, especially in 3D. A similar procedure can be used for our velocity-independent approach. In production applications, the automatically measured slope fields from a subset of CMP gathers should be inspected manually. Slopes are very intuitive to understand and easy to compare to the input data. An overlay or side-by-side display of the two, combined with the NMO performance provide efficient and accurate quality control criteria.

Plane-wave destruction filters provide a truly automated approach to velocity analysis, as they can be used without any user-selected input parameters. Here, we have used the finite-difference plane-wave destructors, which, as described by

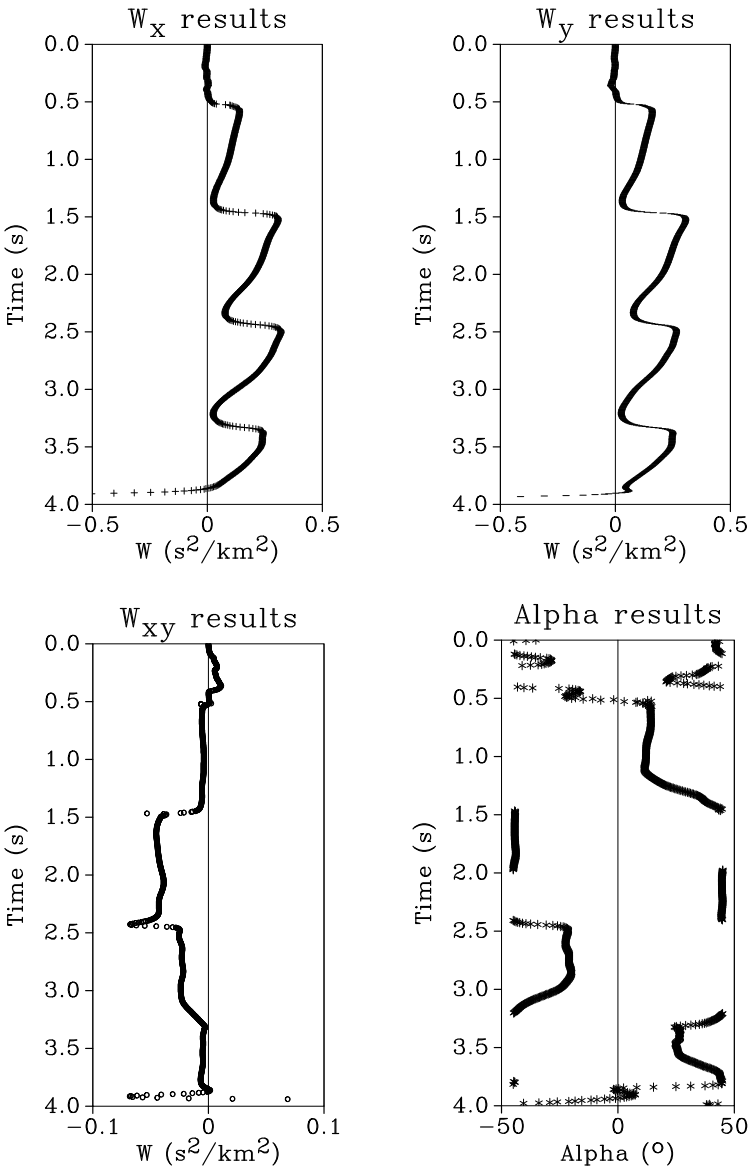


Figure 6: Elements of $\mathbf{W}(t_0)$ inverted from $t_\Delta(t_0, x, y)$ surfaces: (a.) W_x. (b.) W_y. (c.) W_{xy}. (d.) Azimuth angle $\alpha(t_0)$ computed from $\mathbf{W}(t_0)$. nmo3/azimuthtest Wxinv,Wyinv,Wxyinv,InvAlpha

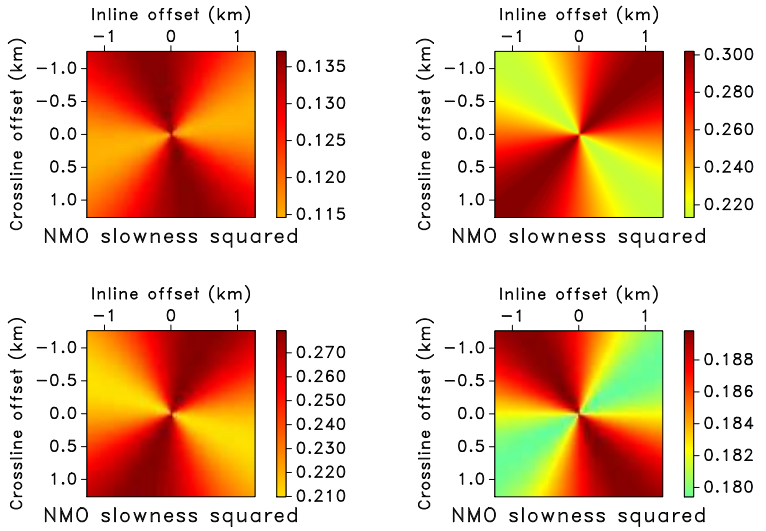


Figure 7: Time-slice views of NMO slowness-squared values computed for each event.(a.) Event A. (b.) Event B. (c.) Event C. (d.) Event D. nmo3/azimuthtest slowsqrdslice1,slowsqrdslice2,slowsqrdslice3,slowsqrdslice4

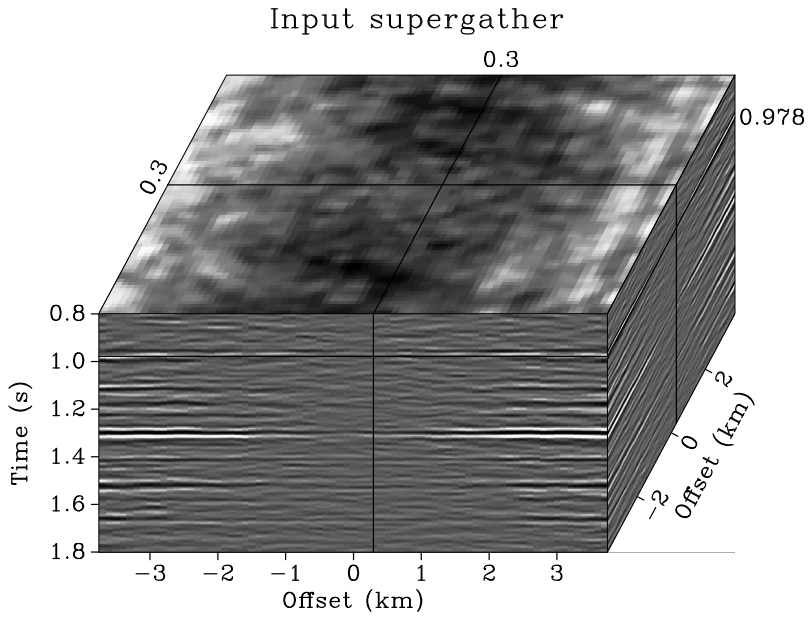


Figure 8: 3D view of a supergather from the McElroy data set, West Texas, US. Although the data has been isotropically NMO corrected, the time-slice view shows a subtle directional trend to the flatness of an event at 0.978 s. nmo3/mcelroy supercube

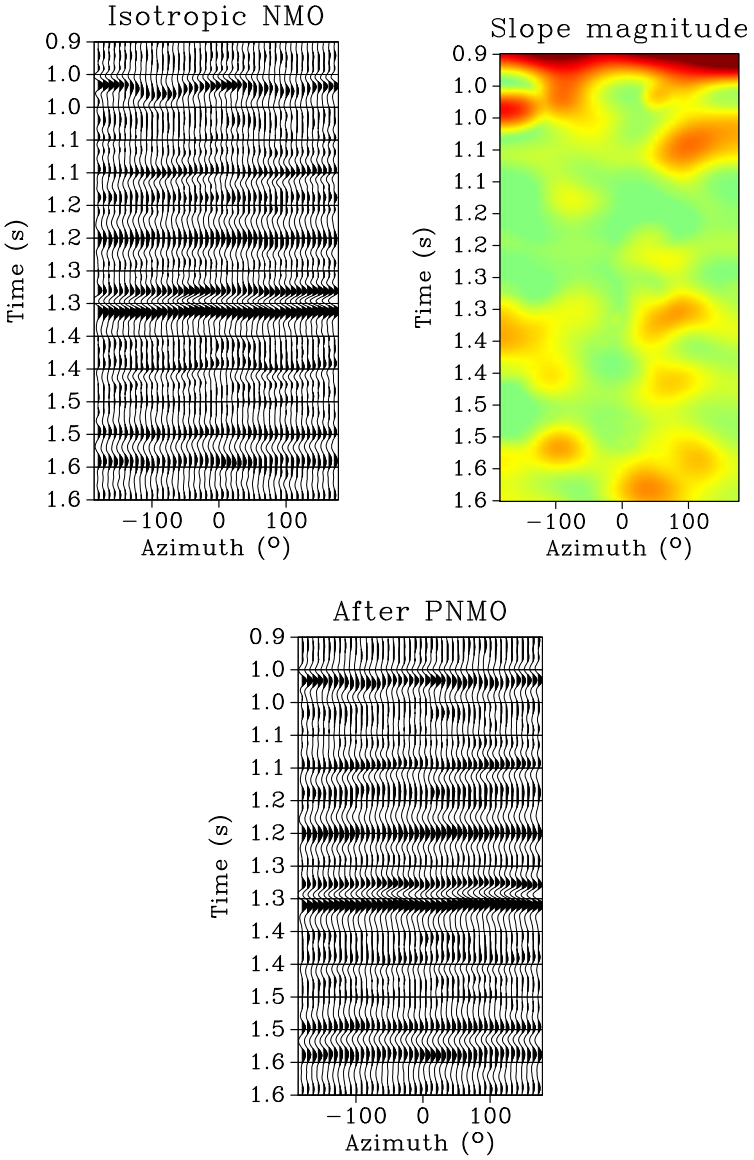


Figure 9: A time-versus-azimuth panel of traces for a range of offsets from 3.6-4.0 km from the McElroy data set. The central panel shows local slope magnitudes corresponding to the left input panel. The slope magnitude was computed as $\sqrt{p_x^2 + p_y^2}$. The same data after an automatic moveout correction has been applied as a residual moveout correction is shown on the right. nmo3/mcelroy super21-anisow,pmag3-anisow,pnmo-oa2-anisow

(2), can be given a user-supplied initial estimate of the slope field. Providing an initial slope estimate helps improve the efficiency of the slope-detection and can help estimate conflicting slopes (2). In all of the examples above, no initial slope field was provided. The output slope fields are computed using smoothing regularization, which helps make the moveout correction more robust, and provides a way for the user to interact with the slope detection performance. If the seismic data is particularly noisy, a more aggressive smoothing can help make a more consistent automatic NMO correction, while for clean data, less smoothing yields a better resolved localized slope field.

We would like to comment here on the performance of the method for realistic cases containing a stack of layers, each with a different orientation of azimuthal anisotropy. The azimuthally-dependent traveltime variations caused by wave propagation in the upper layers will be superimposed on the reflection events corresponding to underlying layers. While inverting for NMO parameters is shown to be straightforward through the velocity-independent approach, solving for interval parameters would require these effects to be unraveled through the use of layer-stripping (7) or a Dix-type inversion (6; 5). If the effects from overlying layers distort later traveltime surfaces enough such that they are no longer elliptically hyperbolic as suggested by equation 1, then the moveout correction will not be complete for the entire section. However, as seen in the second example in the previous section, the velocity-independent moveout method can be used as a residual correction, with no changes to the procedure. The later events with incomplete moveout correction can therefore be corrected with iterated applications of the method. Another complication arises in the residual moveout case though, if one wants to extract parameters such as the azimuth angle or moveout slownesses. The equations presented here for parameter extraction were derived for a single pass NMO correction; it remains for further work to extend the parameter estimation methods to cases where residual moveout correction is necessary.

CONCLUSIONS

Measuring local slopes is a thorough and azimuthally flexible way to characterize traveltime surface geometry, which, in the 3D case, provides useful information about azimuthal variations in moveout velocity. We have demonstrated an application for this feature in performing an elliptically anisotropic moveout correction in 3D. No velocities are picked in order to perform this moveout correction, and since we use plane-wave destruction filters to measure local slopes, the entire process is automated. Local moveout velocities can be calculated as a function of the local slopes, and the azimuthal angle of anisotropy can be estimated locally if one measures the first mixed-derivative of the traveltime surfaces at each point. By recording the traveltime shifts applied by our automated method, we formulate a highly overdetermined linear system to solve for moveout parameters as a function of time. This inversion scheme was shown to be very accurate on a synthetic data example, but remains to be tested on field data. The only practical limitation in the synthetic example comes from

steeply dipping events which introduce aliased slope measurements. Although this type of aliasing can be mitigated to some extent with additional processing, further testing will be required on typical field geometries with relatively coarse crossline spacing.

Even in multi-layer cases, where conflicting azimuthal anisotropies are present, the proposed moveout correction itself can be performed accurately and automatically without velocity or parameter estimation. In these cases, the effective moveout parameters can be estimated from our method. Extensions of this method following an iterative scheme analogous to a layer-stripping or Dix-type inversion strategy may provide a powerful option to automatically recover interval parameters as well.

ACKNOWLEDGMENTS

We would like to thank Thomas Browaeys, Vladimir Bashkardin, Xiaolei Song, Paul Murray, and Jim Simmons for helpful discussions and suggestions. We would also like to thank Chevron for providing the McElroy data set.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Crampin, S., 1984, Effective anisotropic elastic constants for wave propagation through cracked solids: Geophysical Journal of the Royal Astronomical Society, **76**, 135–145.

[2] Fomel, S., 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.

[3] ———, 2007, Velocity-independent time-domain seismic imaging using local event slopes: Geophysics, **72**, S139–S147.

[4] Grechka, V., and I. Tsvankin, 1998, 3-D description of normal moveout in anisotropic inhomogeneous media: Geophysics, **63**, 1079–1092.

[5] ———, 2002, NMO-velocity surfaces and Dix-type formulas in anisotropic heterogeneous media: Geophysics, **67**, 939–951.

[6] Grechka, V., I. Tsvankin, and J. K. Cohen, 1999, Generalized Dix equation and analytic treatment of normal-moveout velocity for anisotropic media: Geophysical Prospecting, **47**, 117–148.

[7] Hake, H., 1986, Slant stacking and its significance for anisotropy: Geophysical Prospecting, **34**, 595–608.

[8] Levin, F. K., 1985, Apparent velocity from dipping interface reflections: Geophysics, **50**, 2026–2032.

[9] Lynn, W., 2007, Uncertainty implications in azimuthal velocity analysis: 77th Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstracts, 84–88.

[10] Ottolini, R., 1983, Velocity independent seismic imaging, *in* SEP-37: Stanford Exploration Project, 59–68.

[11] Sicking, C., and S. Nelan, 2008, Azimuthal attribute methodologies: comparison of NMO super gathers, sectored isotropic migration, iterative azimuthal migration: 78th Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstracts, 983–987.

[12] Sicking, C., S. Nelan, and W. McLain, 2007, 3D azimuthal imaging: 77th Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstracts, 2364–2367.

[13] Siliqi, R., D. L. Meur, F. Gamar, L. Smith, J. Toure, and P. Herrmann, 2003, High-density moveout parameter fields V and Eta, Part 1: Simultaneous automatic picking: 73rd Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstracts, 2088–2091.

[14] Treadgold, G., C. Sicking, V. Sublette, and G. Hoover, 2008, Azimuthal processing for fracture prediction and image improvement: 78th Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstracts, 988–992.

[15] Weisstein, E. W., 2009, "Quadratic Curve" from MathWorld–A Wolfram Web Resource. (http://mathworld.wolfram.com/QuadraticCurve.html, accessed 14 April 2009).

[16] Williams, M., and E. Jenner, 2002, Interpreting seismic data in the presence of azimuthal anisotropy; or azimuthal anisotropy in the presence of the seismic interpretation: The Leading Edge, **21**, 771–774.

[17] Wolf, K., D. A. Rosales, A. Guitton, and J. Claerbout, 2004, Robust moveout without velocity picking: 74th Annual International Meeting, Society of Exploration Geophysicists, Expanded Abstracts, 2423–2426.

Structural uncertainty of time-migrated seismic images

*Sergey Fomel[1] and Evgeny Landa[2]**

ABSTRACT

Structural information in seismic images is uncertain. The main cause of this uncertainty is uncertainty in velocity estimation. We adopt the technique of velocity continuation for estimating velocity uncertainties and corresponding structural uncertainties in time-migrated images. Data experiments indicate that structural uncertainties can be significant even when both structure and velocity variations are mild.

INTRODUCTION

The usual outcome of seismic data processing is an image of the subsurface (26). In the conventional data analysis workflow, the image is passed to the seismic interpreter, who makes geological interpretation, often by extracting structural information, such as positions of horizons and faults in the image. Hidden in this process is the fact that structural information is fundamentally uncertain, mainly because of uncertainties in estimating seismic velocity parameters, which are required for imaging. Apart from the trivial case of perfectly flat seismic reflectors, which are positioned correctly in time even when incorrect stacking or migration velocities are used, seismic images can be and usually are structurally distorted because of inevitable errors in velocity estimation (17).

Understanding and quantifying uncertainty in geophysical information can be crucially important for resource exploration (5). The issue of structural uncertainty in seismic images was analyzed previously by (24) and (21). (25) studied the impact of velocity uncertainties on migrated images and AVO attributes. (2; 3) studied the influence of velocity and anisotropy uncertainties on structural uncertainties.

In this paper, we propose a constructive procedure for estimating the degree of structural uncertainty in seismic images obtained by prestack time migration. The basis for our approach is the method of velocity continuation (12; 18; 14; 15; 4), which constructs seismic images by an explicit continuation in migration velocity. Velocity continuation generalizes the earlier ideas of residual and cascaded migrations (22; 23; 20). In addition to generating accurate time-migration images, it provides a direct access to measuring the structural dependence (sensitivity) of these images on

*e-mail: sergey.fomel@beg.utexas.edu

migration velocities. We define structural uncertainty as a product of velocity picking uncertainty and structural sensitivity.

We use a simple data example to illustrate our approach and to show that structural uncertainty can be significant even when both structure and velocity variations are mild. Although the proposed approach is directly applicable only to prestack time migration, it can be extended in principle to prestack depth migration using velocity-ray approaches for extending the velocity continuation concept (1; 19; 11).

VELOCITY CONTINUATION AND STRUCTURAL SENSITIVITY

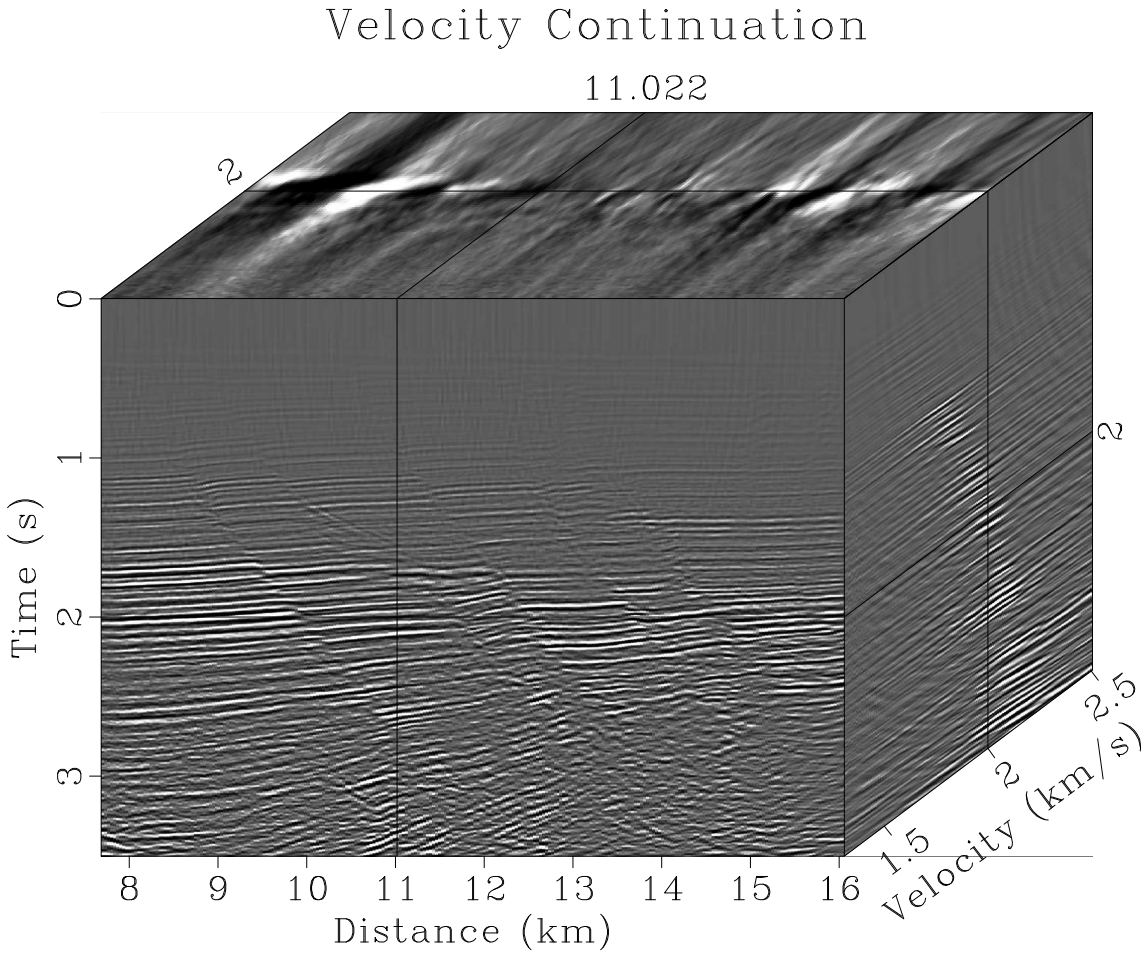


Figure 1: Velocity continuation cube for prestack time migration of the Gulf of Mexico dataset. uncert/beivc vlf

Velocity continuation is defined as the process of image transformation with changes in migration velocity (12; 15). Its output is equivalent to the output of

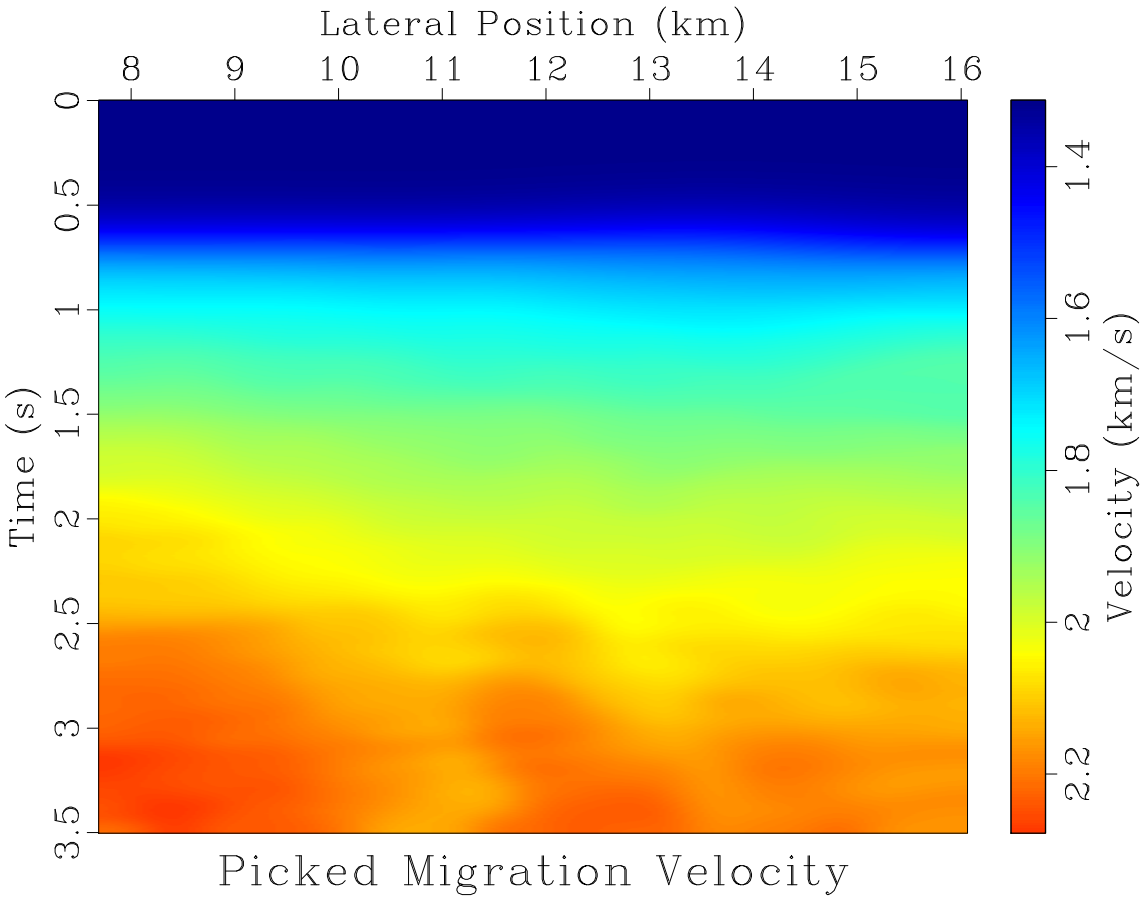


Figure 2: Migration velocity picked from velocity continuation. uncert/beivc npk

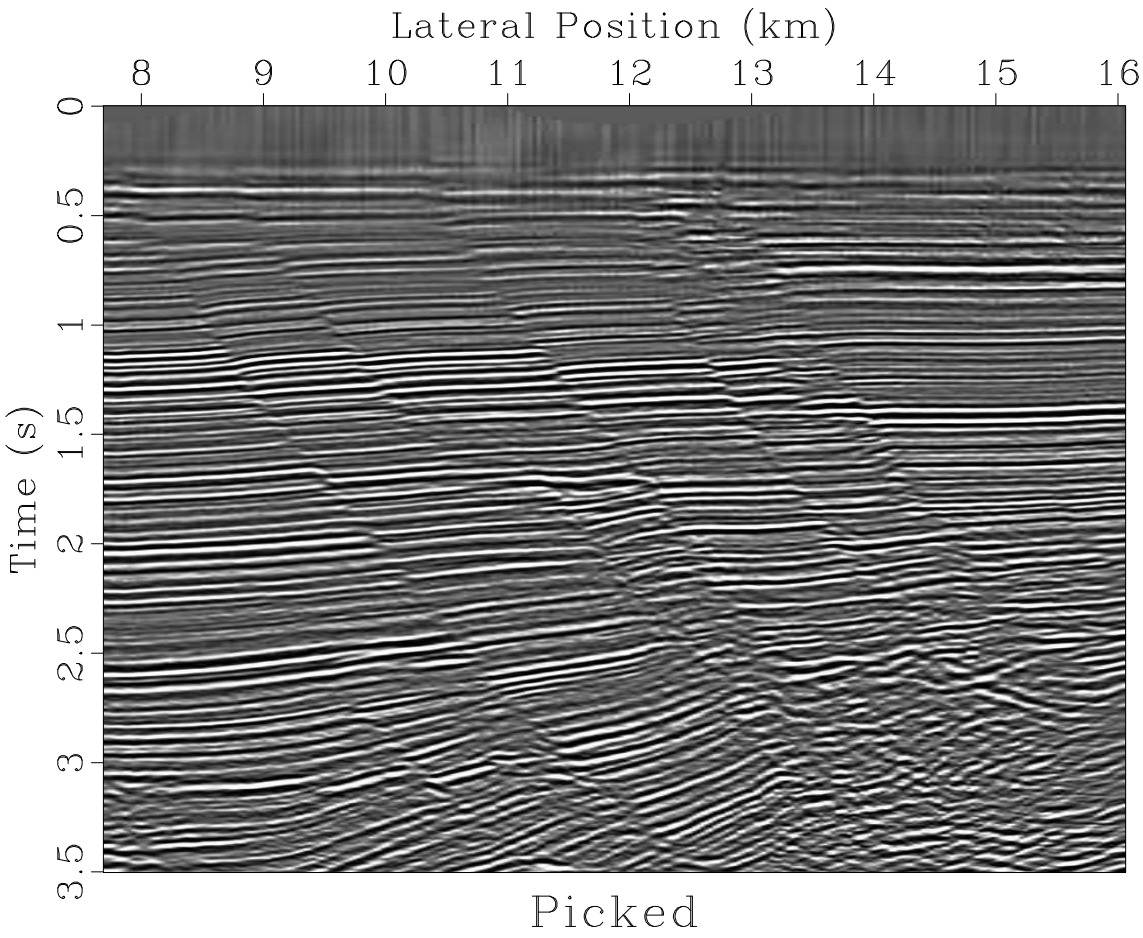


Figure 3: Seismic prestack time-migration image generated by velocity continuation.
uncert/beivc bei-agc

repeated migrations with different migration velocities (27) but produced more efficiently by using propagation of images in velocity (18). If we denote the output of velocity continuation as $C(t, x, v)$, where t and x are time-migration coordinates and v is migration velocity, the time-migrated image is simply

$$I(t, x) = C(t, x, v_M(t, x)) \ , \tag{1}$$

where $v_M(t, x)$ is the picked migration velocity. Figure 1 shows the velocity continuation cube $C(t, x, v)$ generated from a benchmark 2-D dataset from the Gulf of Mexico (10). Migration velocity $v_M(t, x)$ picked from the semblance analysis is shown in Figure 2. The velocity variations reflect a dominantly vertical gradient typical for the Gulf of Mexico and only mild lateral variations, which justifies the use of prestack time migration. The corresponding migration image $I(t, x)$ is shown in Figure 3 and exhibits mild, nearly-horizontal reflectors and sedimentary structures.

The structural sensitivity of an image can be described through derivatives $\partial t/\partial v$ and $\partial x/\partial v$, which correspond to slopes of events in the $C(t, x, v)$ volume evaluated at $v = v_M(t, x)$. These slopes are easy to measure experimentally from the $C(t, x, v_M)$ volume, using, for example, the plane-wave destruction algorithm (13; 7; 8). Figure 4 shows one common-image gather $G(t, v) = C(t, x_0, v)$ for $x_0 = 10\,\mathrm{km}$ and the time slice $S(x, v) = C(t_0, x, v)$ for $t_0 = 2\,\mathrm{s}$. Measuring the slope of events $\partial t/\partial v$ in this gather and evaluating it at the picked migration velocity produces the slope

$$p_t(t, x) = \left.\frac{\partial t}{\partial v}\right|_{v=v_M(t,x)} \ . \tag{2}$$

We measure the slope $p_x(t, x)$ analogously by evaluating local slopes in time slices of constant t:

$$p_x(t, x) = \left.\frac{\partial x}{\partial v}\right|_{v=v_M(t,x)} \ . \tag{3}$$

Figure 5 shows the estimated p_t and p_x, which comprise the structural sensitivity of our image.

Theoretically, structural sensitivity can be inferred from the zero-offset velocity ray equations (9; 15)

$$\frac{dt}{dv} = v_M \, t \, t_x^2 = \frac{t}{v_M} \, \tan^2 \theta \ , \tag{4}$$

$$\frac{dx}{dv} = -2 \, v_M \, t \, t_x = -2 \, t \frac{t}{v_M} \, \tan^2 \theta \ , \tag{5}$$

where t_x corresponds to the slope of the reflector, and θ is the reflector dip angle. According to equations 4-5, the reflector dip is the dominant factor in structural sensitivity.

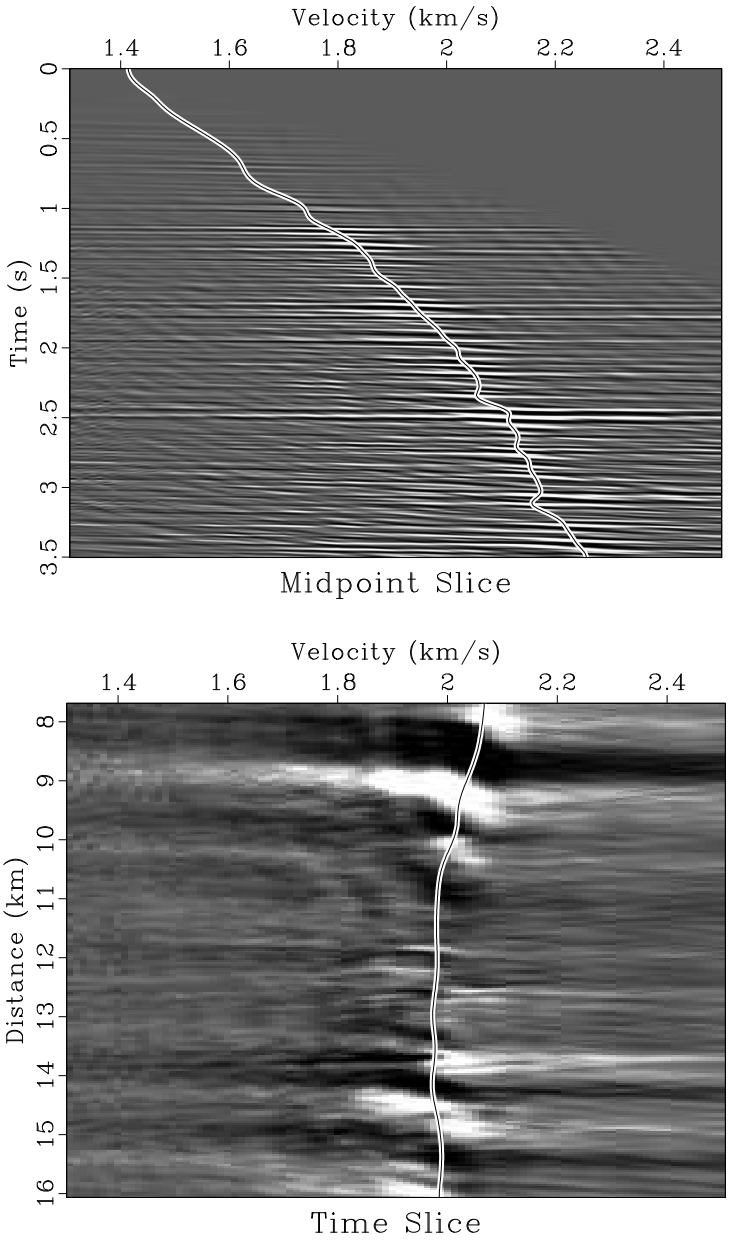


Figure 4: Common-image gather (a) and time slice (b) from velocity continuation with overlaid time-migration velocity. uncert/beivc slice,tslice

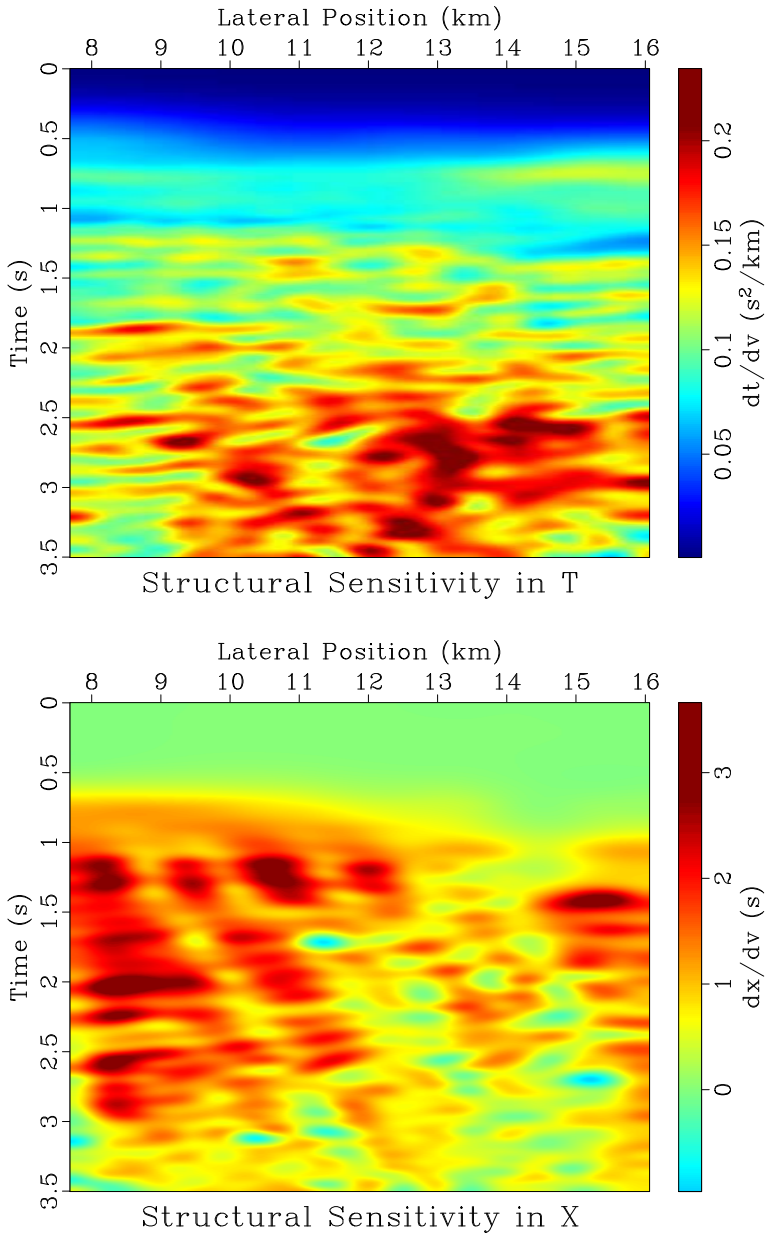


Figure 5: Estimated structural sensitivity in time (a) and lateral position (b) with respect to velocity. uncert/beivc bei-dtdv,bei-dxdv

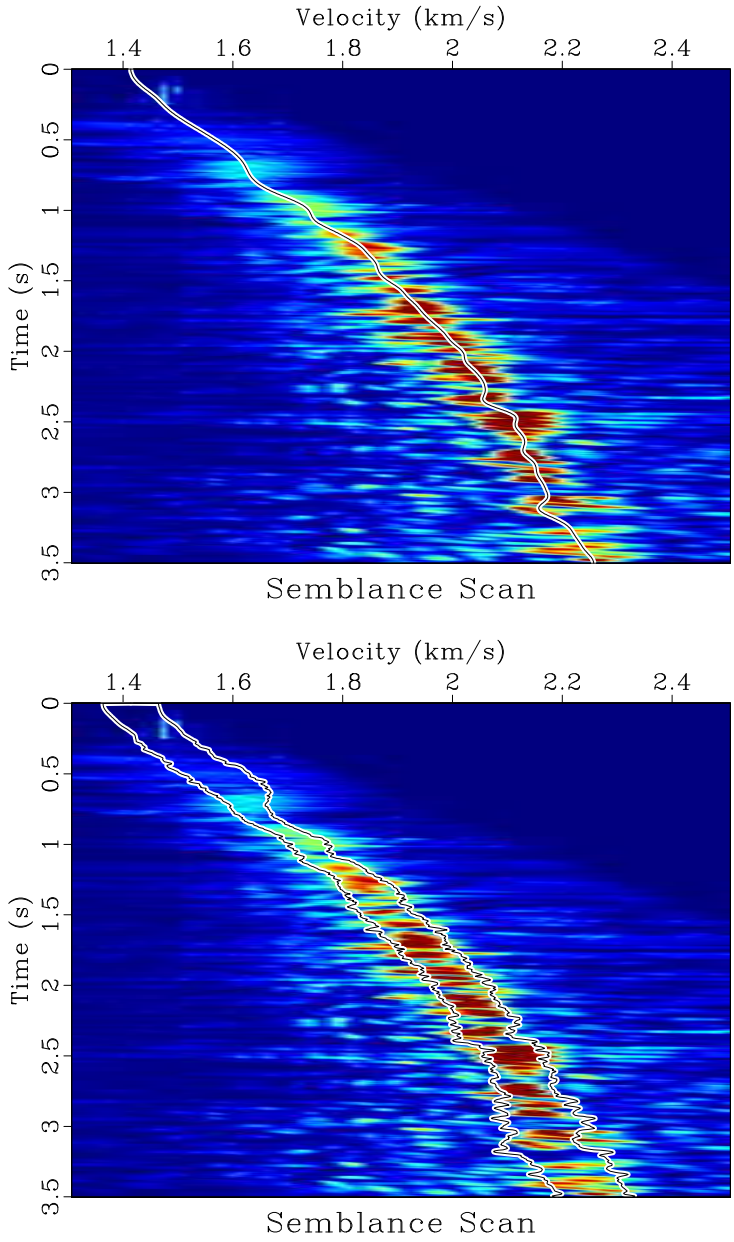


Figure 6: Velocity scan at 10 km image gather. The curve in (a) corresponds to the automatically picked velocity trend. The curves in (b) identify an approximate range of velocity uncertainty around the picked trend. uncert/beivc scan,scan2

UNCERTAINTY IN VELOCITY PICKING

Figure 6(a) shows a semblance scan produced in the process of velocity continuation. A common procedure in migration velocity analysis is picking a velocity trend from the semblance, either manually or automatically. In this example, we use automatic picking with the algorithm described by (16).

While picking may select the most probable velocity function, its probability is less than 100%. If we view normalized semblance as a probability distribution and determine a confidence interval corresponding roughly to one standard deviation, it provides an approximate range of uncertainty in velocity determination. This range is shown in Figure 6(b) and computed according to

$$\delta v(t,x) = \sqrt{\frac{\int_{v_{min}}^{v_{max}} [v - v_M(t,x)]^2 \; S(t,x,v)\, dv}{\int_{v_{min}}^{v_{max}} S(t,x,v)\, dv}} \; , \tag{6}$$

where $S(t,x,v)$ is the semblance volume that corresponds to $C(t,x,v)$, and $[v_{min}, v_{max}]$ is the full range of velociies. The interpretation of semblance picks as probability distributions is heuristic but helps in quantifying uncertainties in velocity picking.

STRUCTURE UNCERTAINTY

Putting structural sensitivity and velocity uncertainty together, we can define *structural uncertainty* simply as their product:

$$\delta t = \frac{\partial t}{\partial v}\, \delta v \; , \tag{7}$$

$$\delta x = \frac{\partial x}{\partial v}\, \delta v \; . \tag{8}$$

The uncertainty $\{\delta t, \delta x\}$ is the main output of our study. It is shown as small line segments in Figure 7 and as uncertainty in horizons in Figure 8. The estimated uncertainty varies inside the image space and generally increases with depth. It is surprisingly large, given the mild variations in structure and velocity. We believe that, when making quantitative estimates related to structural interpretation, it is important to take this kind of uncertainty into account.

When converting seismic images from time to depth, it is also important to realize that the time-to-depth conversion itself is a mathematically ill-posed problem (6) and has its own significant uncertainties.

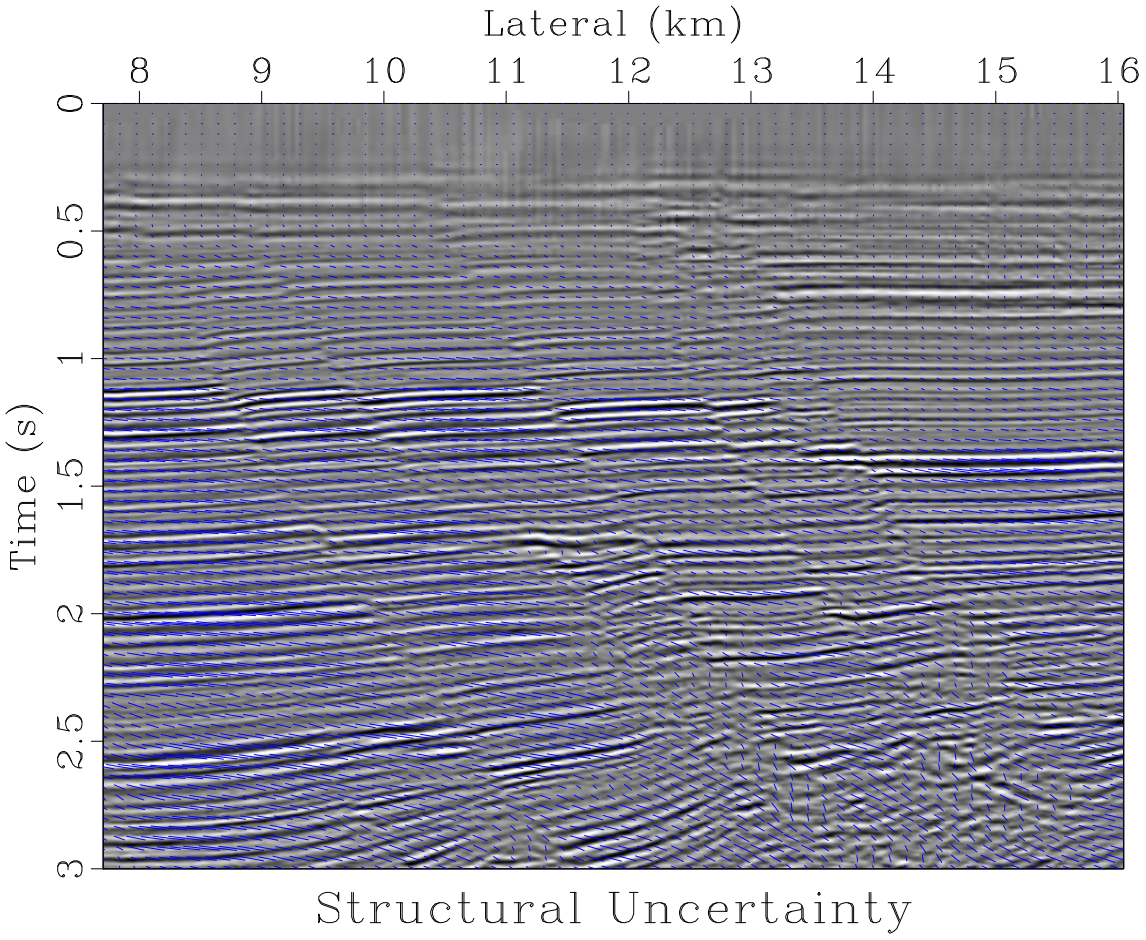


Figure 7: Estimated structural uncertainty in the seismic image from Figure 3, displayed as displacements. uncert/beivc arr

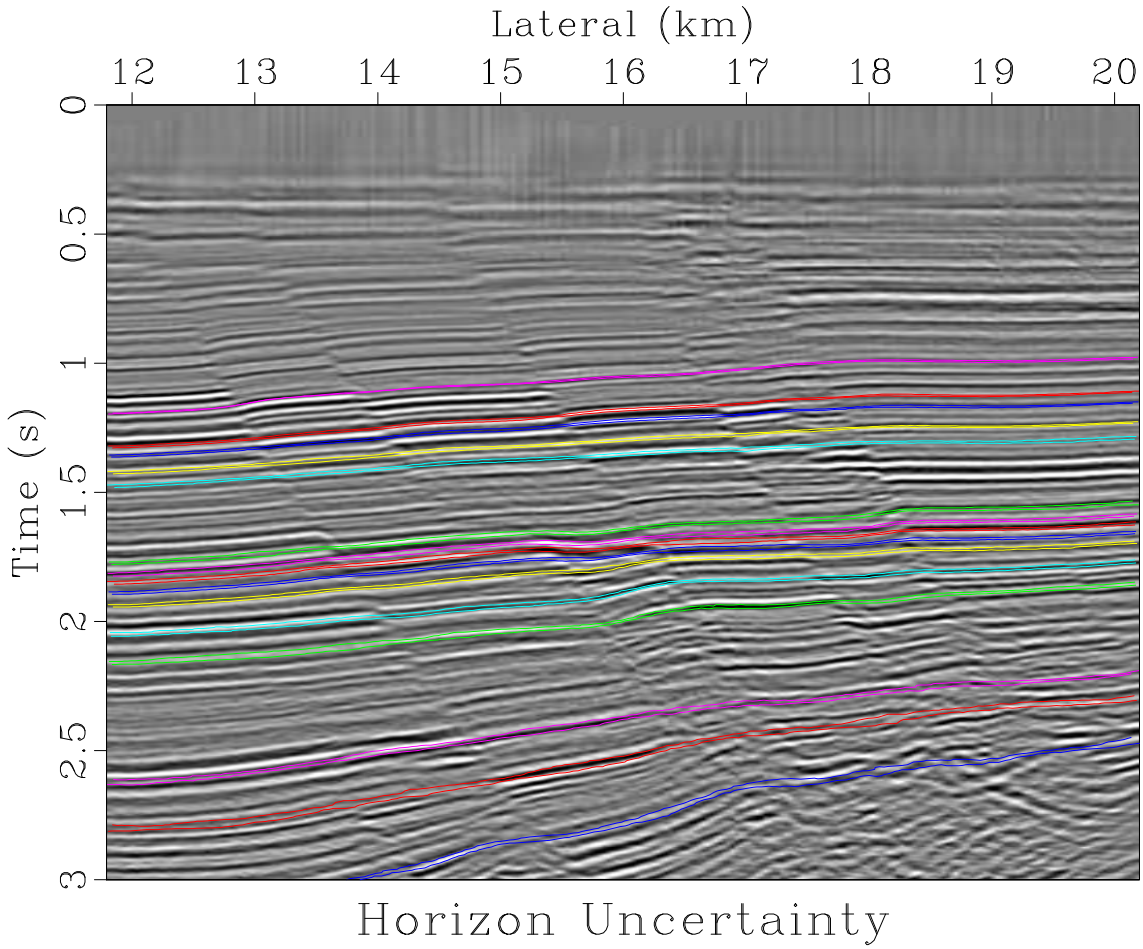


Figure 8: Estimated structural uncertainty in the seismic image from Figure 3, displayed as horizon uncertainties. uncert/beivc hors

CONCLUSIONS

We have estimated structural uncertainty in seismic time-domain images simultaneously with performing prestack time migration. To accomplish this task, we projected the uncertainty in migration velocity picking into the structural uncertainty by measuring the structural sensitivity of seismic images to velocity. The latter measure is provided by velocity continuation, which serves both as an imaging tool and as a tool for sensitivity analysis. Field data examples show that structural uncertainties can be significant even in the case of mild structures and slow velocity variations. Taking these uncertainties into account should improve the practice of seismic structural interpretation by making it more compliant with risk-management assessment in reservoir characterization.

REFERENCES

[1] Adler, F., 2002, Kirchhoff image propagation: Geophysics, **67**, 126–134.
[2] Bube, K. P., J. A. Kane, T. Nemeth, D. Medwedeff, and O. Mikhailov, 2004a, The influence of stacking velocity uncertainties on structural uncertainties: 74th Ann. Internat. Mtg., Soc. of Expl. Geophys., 2188–2191.
[3] Bube, K. P., T. Nemeth, O. Mikhailov, D. Medwedeff, and J. A. Kane, 2004b, The influence of uncertainties in anisotropy on structural uncertainties: 74th Ann. Internat. Mtg., Soc. of Expl. Geophys., 2192–2195.
[4] Burnett, W., and S. Fomel, 2011, Azimuthally anisotropic 3D velocity continuation: International Journal of Geophysics, **2011**, Article ID 484653.
[5] Caers, J., 2011, Modeling uncertainty in the Earth sciences: Wiley.
[6] Cameron, M. K., S. Fomel, and J. A. Sethian, 2007, Seismic velocity estimation from time migration velocities: Inverse Problems, **23**, 1329–1369.
[7] Chen, Z., S. Fomel, and W. Lu, 2013a, Accelerated plane-wave destruction: Geophysics, **78**, V1–V9.
[8] ——, 2013b, Omnidirectional plane-wave destruction: Geophysics, **78**, V171–V179.
[9] Chun, J. H., and C. A. Jacewitz, 1981, Fundamentals of frequency-domain migration: Geophysics, **46**, 717–733.
[10] Claerbout, J. F., 2005, Basic Earth imaging: Stanford Exploration Project, `http://sepwww.stanford.edu/sep/prof/`.
[11] Duchkov, A. A., and M. V. De Hoop, 2009, Velocity continuation in the downward continuation approach to seismic imaging: Geophysical Journal International, **176**, 909–924.
[12] Fomel, S., 1994, Method of velocity continuation in the problem of seismic time migration: Russian Geology and Geophysics, **35**, 100–111.
[13] ——, 2002, Applications of plane-wave destruction filters: Geophysics, **67**, 1946–1960.
[14] ——, 2003a, Time-migration velocity analysis by velocity continuation: Geophysics, **68**, 1662–1672.

[15] ———, 2003b, Velocity continuation and the anatomy of residual prestack time migration: Geophysics, **68**, 1650–1661.
[16] ———, 2009, Velocity analysis using AB semblance: Geophysical Prospecting, **57**, 311–321.
[17] Glogovsky, V., E. Landa, S. Langman, and T. J. Moser, 2009, Validating the velocity mode: the Hamburg Score: First Break, **27**, 77–85.
[18] Hubral, P., M. Tygel, and J. Schleicher, 1996, Seismic image waves: Geophysical Journal International, **125**, 431–442.
[19] Iversen, E., 2006, Velocity rays for heterogeneous anisotropic media: Theory and implementation: Geophysics, **71**, T117–T127.
[20] Larner, K., and C. Beasley, 1987, Cascaded migrations - Improving the accuracy of finite-difference migration: Geophysics, **52**, 618–643. (Errata in GEO-52-8-1165).
[21] Pon, S., and L. R. Lines, 2005, Sensitivity analysis of seismic depth migrations: Geophysics, **70**, S39–S42.
[22] Rocca, F., and L. Salvador, 1982, Residual migration: 52nd Ann. Internat. Mtg, Soc. of Expl. Geophys., Session:S1.4.
[23] Rothman, D. H., S. A. Levin, and F. Rocca, 1985, Residual migration - Applications and limitations: Geophysics, **50**, 110–126.
[24] Thore, P., A. Shtuka, M. A. Lecour, T. Ait-Ettajer, and R. Cognot, 2002, Structural uncertainties: Determination, management, and applications: Geophysics, **67**, 840–852.
[25] Tura, A., and C. Hanitzsch, 2001, Estimating and interpreting velocity uncertainty in migrated images and AvO attributes: Geophysics, **66**, 1208–1216.
[26] Yilmaz, O., 2001, Seismic data analysis: Soc. of Expl. Geophys.
[27] Yilmaz, O., I. Tanir, and C. Gregory, 2001, A unified 3-D seismic workflow: Geophysics, **66**, 1699–1713.

Time-lapse image registration using the local similarity attribute

*Sergey Fomel and Long Jin**

ABSTRACT

We present a method for registration of time-lapse seismic images based on the local similarity attribute. We define registration as an automatic point-by-point alignment of time-lapse images. Stretching and squeezing a monitor image and computing its local similarity to the base image allows us to detect an optimal registration even in the presence of significant velocity changes in the overburden. A by-product of this process is an estimate of the ratio of the interval seismic velocities in the reservoir interval. We illustrate the proposed method and demonstrate its effectiveness using both synthetic experiments and real data from the Duri time-lapse experiment in Indonesia.

INTRODUCTION

Time-lapse seismic monitoring is an important technology for enhancing hydrocarbon recovery (10). At the heart of the method is comparison between repeated seismic images with an attempt to identify changes indicative of fluid movements in the reservoir.

In general, time-lapse image differences contain two distinct effects: shifts of image positions in time caused by changes in seismic velocities and amplitude differences caused by changes in seismic reflectivity. The data processing challenge is to isolate changes in the reservoir itself from changes in the surrounding areas. Cross-equalization is a popular technique for this task (14; 15). A number of different cross-equalization techniques have been successfully applied in recent years to estimate and remove time shifts between time-lapse images (2; 1). An analogous task exists in medical imaging, where it is known as the *image registration* problem (12).

In this paper, we propose to use the local similarity attribute (3) for automatic quantitative estimation and extraction of variable time shifts between time-lapse seismic images. A similar technique has been applied previously to multicomponent image registration (5). As a direct quantitative measure of image similarity, local attributes are perfectly suited for measuring nonstationary time-lapse correlations. The extracted time shifts also provide a direct estimate of the seismic velocity changes in

*e-mail: sergey.fomel@beg.utexas.edu

the reservoir. We demonstrate an application of the proposed method with synthetic and real data examples.

THEORY

The correlation coefficient between two data sequences a_t and b_t is defined as

$$c = \frac{\sum_t a_t\, b_t}{\sqrt{\sum_t a_t^2 \sum_t b_t^2}} \tag{1}$$

and ranges between 1 (perfect correlation) and -1 (perfect correlation of signals with different polarity). The definition of the local similarity attribute (3) starts with the observation that the squared correlation coefficient can be represented as the product of two quantities $c^2 = p\, q$, where

$$p = \frac{\sum_t a_t\, b_t}{\sum_t b_t^2}$$

is the solution of the least-squares minimization problem

$$\min_p \sum_t \left(a_t - p\, b_t\right)^2 \ , \tag{2}$$

and

$$q = \frac{\sum_t a_t\, b_t}{\sum_t a_t^2}$$

is the solution of the least-squares minimization

$$\min_q \sum_t \left(b_t - q\, a_t\right)^2 \ . \tag{3}$$

Analogously, the local similarity γ_t is a variable signal defined as the product of two variable signals p_t and q_t that are the solutions of the regularized least-squares problems

$$\min_{p_t} \left(\sum\nolimits_t \left(a_t - p_t\, b_t\right)^2 + R\left[p_t\right]\right) \ , \tag{4}$$

$$\min_{q_t} \left(\sum\nolimits_t \left(b_t - q_t\, a_t\right)^2 + R\left[q_t\right]\right) \ , \tag{5}$$

where R is a regularization operator designed to enforce a desired behavior such as smoothness. Shaping regularization (4) provides a particularly convenient method of

enforcing smoothness in iterative optimization schemes. If shaping regularization is applied iteratively with Gaussian smoothing as a shaping operator, its first iteration is equivalent to the fast local cross-correlation method of (6). Further iterations introduce relative amplitude normalization and compensate for amplitude effects on the local image similarity. Choosing the amount of regularization (smoothness of the shaping operator) affects the results. In practice, we start with strong smoothing and decrease it when the results stop changing and before they become unstable.

The application of local similarity to the time-lapse image registration problem consists of squeezing and stretching the monitor image with respect to the base image while computing the local similarity attribute. Next, we pick the strongest similarity trend from the attribute panel and apply the corresponding shift to the image.

In addition to its use for image registration, the estimated local time shift is a useful attribute by itself. Time shift analysis has been widely applied to infer reservoir compaction (7; 16; 8; 13). Since the time shift has a cumulative effect, it is helpful to compute the derivative of time shift, which can relate the time shift change to the corresponding reservoir layer. (13) define the derivative of time shift as *time strain* and find it to be an intuitive attribute for studying reservoir compaction.

What is the exact physical meaning of the warping function $w(t)$ that matches the monitor image $I_1(t)$ with the base image $I_0(t)$ by applying the transformation $I_1[w(t)]$? One can define the base traveltime as an integral in depth, as follows:

$$t = 2 \int_0^{H_0} \frac{dz}{v_0(z)} \ , \tag{6}$$

where $v_0(z)$ is the base velocity, and H_0 is the base depth. A similar event in the monitor image appears at time

$$w(t) = 2 \int_0^{H_1} \frac{dz}{v_1(z)} = \int_0^{t+\Delta t} \frac{\hat{v}_0(\tau)}{\hat{v}_1(\tau)} \, d\tau \ , \tag{7}$$

where H_1 is the monitor depth, $\hat{v}_0(t)$ and $\hat{v}_1(t)$ are seismic velocities as functions of time rather than depth, and Δt is the part of the time shift caused by the reflector movement:

$$\Delta t = 2 \int_{H_0}^{H_1} \frac{dz}{v_1(z)} \tag{8}$$

In a situation where the change of Δt with t can be neglected, a simple differentiation of the function $w(t)$ detected by the local similarity analysis provides an estimate of the local ratio of the velocities:

$$\frac{dw}{dt} \approx \frac{\hat{v}_0(t)}{\hat{v}_1(t)} \ . \tag{9}$$

If the registration is correct, the estimated velocity ratio outside of the reservoir should be close to one. One can connect the local velocity ratio to other physical attributes that are related to changes in saturation, pore pressure, or compaction.

We demonstrate the proposed procedure in the next section using several examples.

EXAMPLES

1-D synthetic data

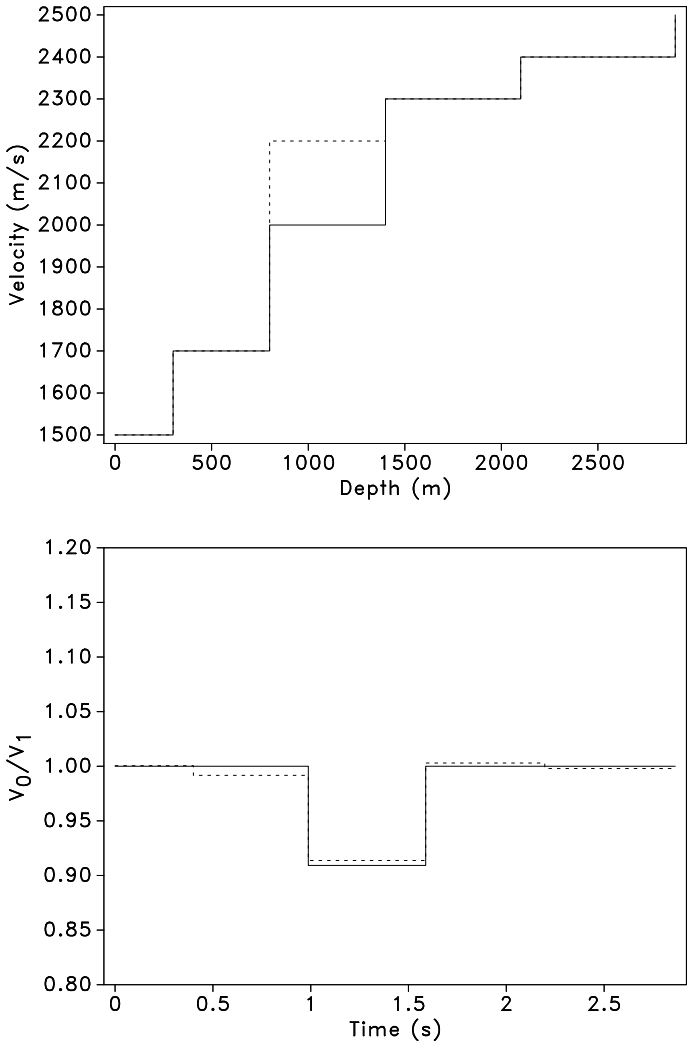


Figure 1: (a) 1-D synthetic velocity model before (solid line) and after (dashed line) reservoir production. (b) True (solid line) and estimated (dashed line) interval velocity ratio. timelapse/timelapse modl,rat

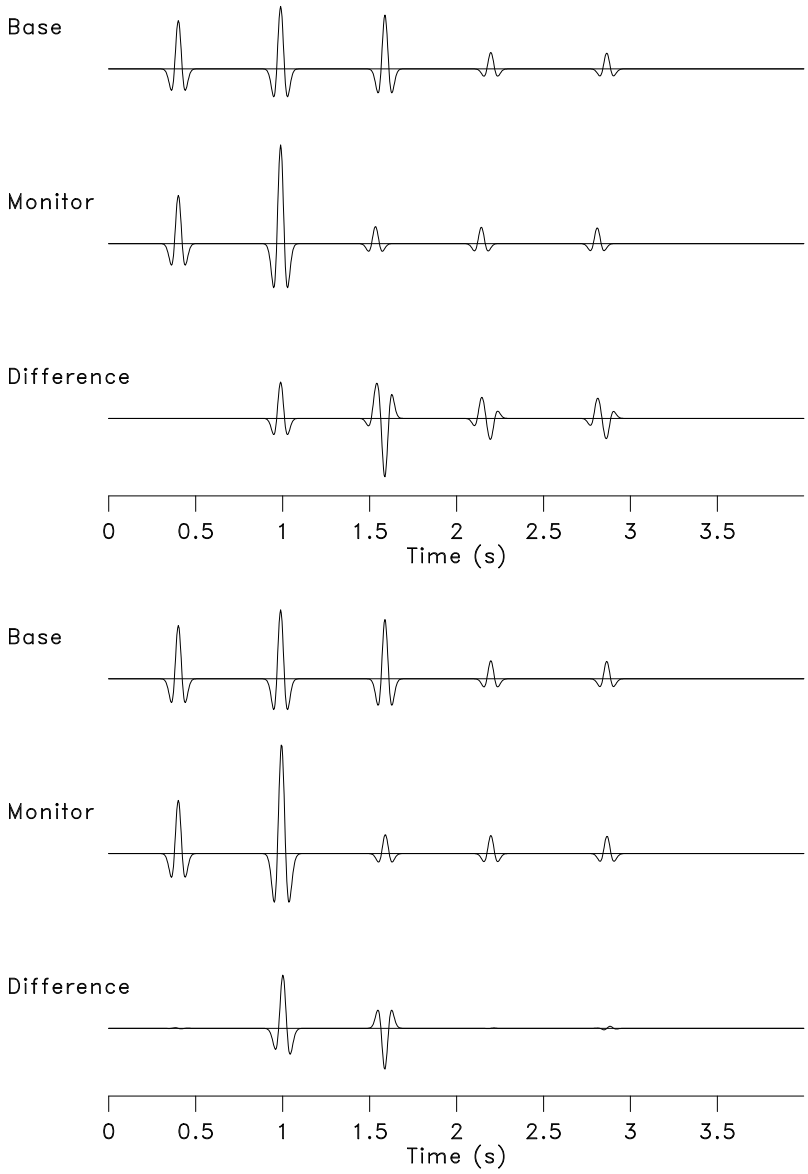


Figure 2: 1-D synthetic seismic images and the time-lapse difference initially (a) and after image registration (b). timelapse/timelapse data,warp

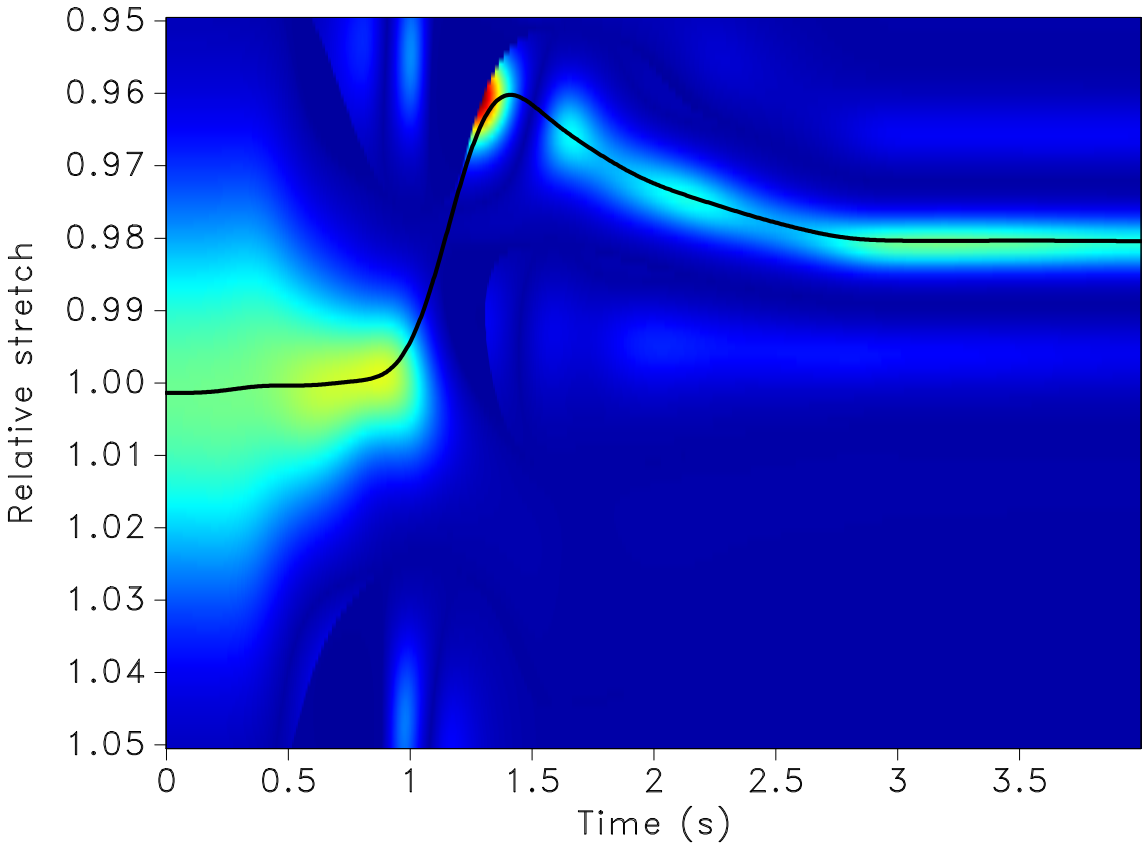


Figure 3: (a) Local similarity scan for detecting the warping function in the 1-D synthetic model. Red colors indicate large similarity. The black curve shows an automatically detected trend. timelapse/timelapse scan100

Figure 5 shows a simplistic five-layer velocity model, where we introduce a velocity increase in one of the layers to simulate a time-lapse effect. After generating synthetic image traces, we can observe, in Figure 9(a), that the time-lapse difference contains changes not only at the reservoir itself but also at interfaces below the reservoir. Additionally, the image amplitude and the wavelet shape at the reservoir bottom are incorrect. These artifact differences are caused by time shifts resulting from the velocity change. After detecting the warping function $w(t)$ from the local similarity scan, shown in Figure 3, and applying it to the time-lapse image, the difference correctly identifies changes in reflectivity only at the top and the bottom of the producing reservoir [Figure 2(b)]. To implement the local similarity scan, we use the *relative stretch* measure $s(t) = w(t)/t$. When the two images are perfectly aligned, $s(t) = 1$. Deviations of $s(t)$ from one indicate possible misalignment. Finally, we apply equation 9 to estimate interval velocity changes in the reservoir and observe a reasonably good match with the exact synthetic model [Figure 1(b)].

2-D synthetic data

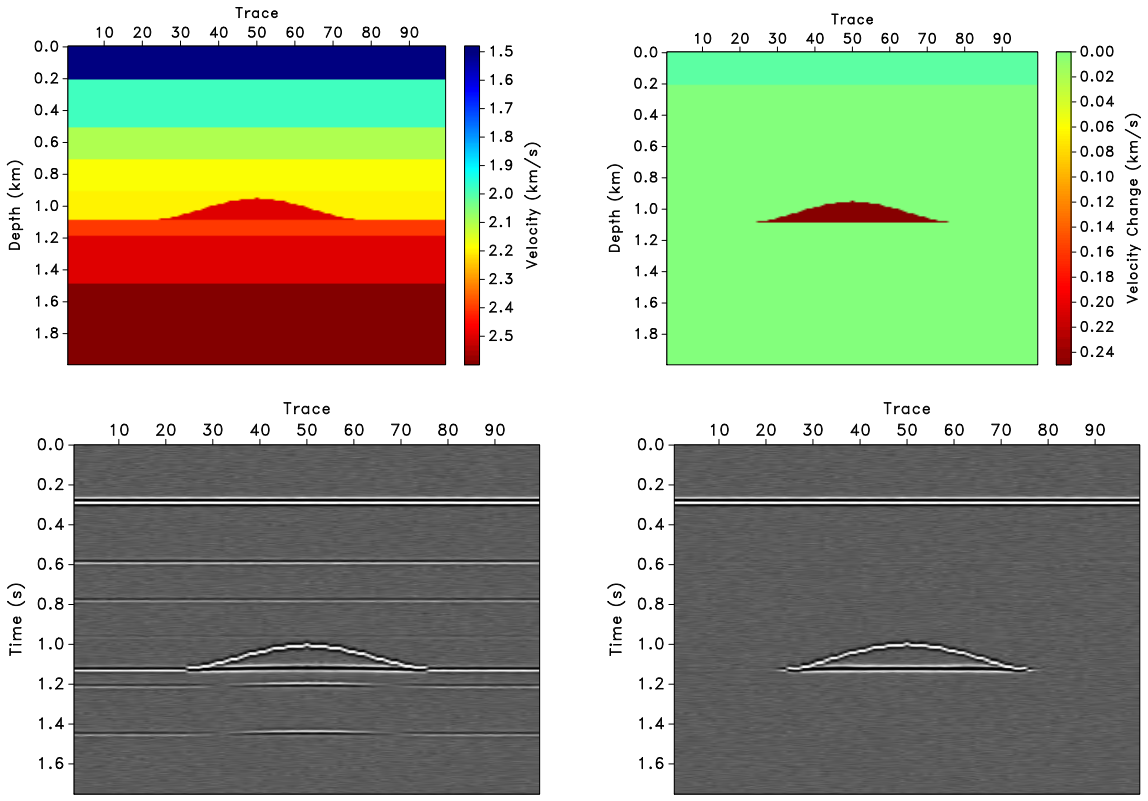


Figure 4: (a) Synthetic model. (b) Time-lapse change containing differences in both the reservoir interval and the shallow overburden. (c) Initial time-lapse difference image. (d) Time-lapse difference image after registration. timelapse/long1 time2-over,mdif-time2-over,dif-time2-over,dif2-time2-over

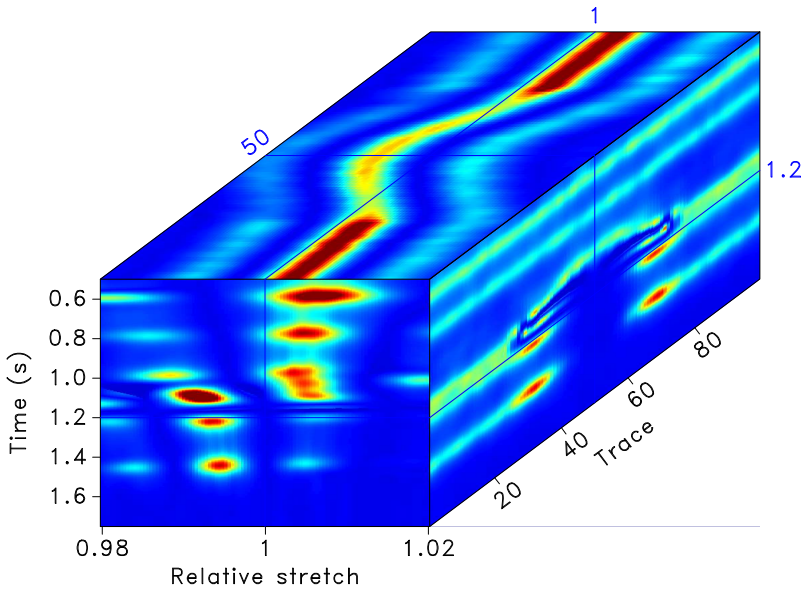


Figure 5: Local similarity scan for the 2-D synthetic model. timelapse/long1 scan-time2-over

Figure 4(a) shows a more complicated 2-D synthetic example. In this experiment, we assume that the changes occur both in the reservoir and in the shallow subsurface [Figure 4(b)]. The synthetic data were generated by convolution modeling. After computing local similarity between the two synthetic time-lapse images [Figure 5], we apply the extracted stretch factor to register the images. Figures 4(c) and 4(d) compare time-lapse difference images before and after registration. Similarity-based registration effectively removes artifact differences both above and below the synthetic reservoir. As mentioned before, the local similarity cube is an important attribute by itself and includes information on uncertainty bounds for the local stretch factor, which reflects the uncertainty of the reservoir parameter estimation.

3-D field data

Finally, Figure 6 shows an application of the proposed method to time-lapse images from steam flood monitoring in the Duri field, reproduced from (9; 11). Before registration, real differences in the monitor surveys after 2 months and 19 months are obscured by coherent artifacts, which are caused by velocity changes both in the shallow overburden and in the reservoir interval [Figure 6(a)]. Similarly to the results of the synthetic experiments, local-similarity registration succeeds in removing artifact differences both above and below the reservoir level [Figure 6(b)]. After separating the time-shift effect from amplitude changes, one can image the steam front propagation more accurately using time-lapse seismic data. We expect our method to work even better on higher-quality marine data.

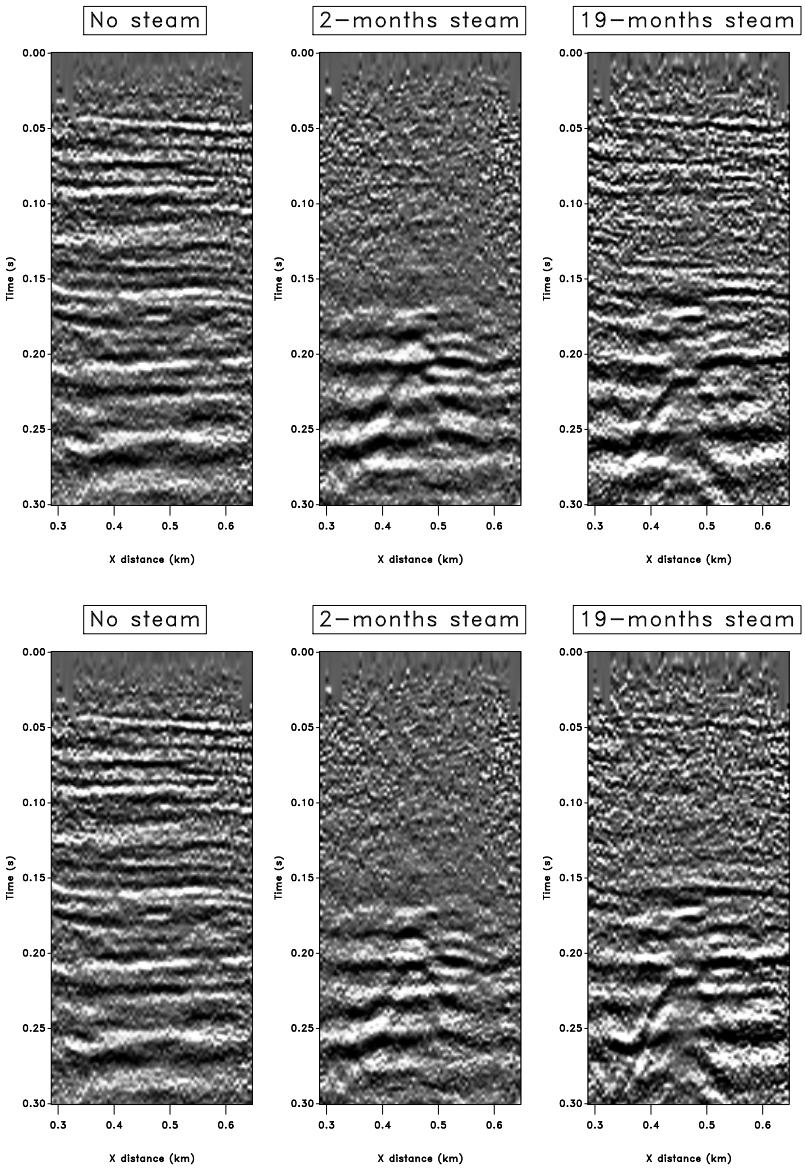


Figure 6: Application to the Duri field data. Base image and time-lapse differences: (a) before registration, reproduced from (9), (b) after registration. timelapse/duri difc,difc2

CONCLUSIONS

We propose a method of time-lapse image registration based on an application of the local similarity attribute. The local attribute provides a smooth continuous measure of similarity between two images. Perturbing the monitor image by stretching and squeezing it in time while picking its best match to the base image enables an effective registration algorithm. The by-product of this process is an estimate of the time-lapse seismic velocity ratios in the reservoir interval.

Using synthetic and real data examples, we have demonstrated the ability of our method to achieve an accurate time-domain image registration and to remove artifact time-lapse differences caused by velocity changes. Unlike some of the alternative cross-equalization methods, the proposed method is not influenced by amplitude differences and can account for velocity changes in the shallow overburden.

ACKNOWLEDGMENTS

We thank David Lumley for the permission to use results from his Ph.D. thesis. We also thank Aaron Janssen, Brackin Smith, and Ali Tura for inspiring discussions and two anonymous reviewers for helpful suggestions.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Aarre, V., 2006, Estimating 4D velocity changes and contact movement on the Norne field: 76th Ann. Internat. Mtg, Soc. of Expl. Geophys., 3115–3119.

[2] Bertrand, A., S. McQuaid, R. Bobolecki, S. Leiknes, and H. Ro, 2005, A high resolution workflow for 4D-friendly analysis: application to gas-oil contact monitoring at Troll West: 75th Ann. Internat. Mtg, Soc. of Expl. Geophys., 2422–2426.

[3] Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.

[4] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[5] Fomel, S., M. Backus, K. Fouad, B. Hardage, and G. Winters, 2005, A multistep approach to multicomponent seismic image registration with application to a West Texas carbonate reservoir study: 75th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1018–1021.

[6] Hale, D., 2006, Fast local cross-correlations of images: 76th Ann. Internat. Mtg, Soc. of Expl. Geophys., 3160–3163.

[7] Hatchell, P., and S. Bourne, 2005, Rocks under strain: Strain-induced time-lapse time shifts are observed for depleting reservoirs: The Leading Edge, **24**, 1222–1225.

[8] Janssen, A. L., B. A. Smith, and G. W. Byerley, 2006, Measuring velocity sensi-

tivity to production-induced strain at the Ekofisk Field using time-lapse time-shifts and compaction logs: 76th Ann. Internat. Mtg, Soc. of Expl. Geophys., 3200–3203.
[9] Lumley, D., 1995a, Seismic time-lapse monitoring of subsurface fluid flow: PhD thesis, Stanford University.
[10] ———, 2001, Time-lapse seismic reservoir monitoring: Geophysics, **66**, 50–53.
[11] Lumley, D. E., 1995b, 4-D seismic monitoring of an active steamflood: 65th Ann. Internat. Mtg, Soc. of Expl. Geophys., 203–206.
[12] Modersitzki, J., 2004, Numerical methods for image registration: Oxford University Press.
[13] Rickett, J., L. Duranti, T. Hudson, and B. Regel, 2007, 4D time strain and the seismic signature of geomechanical compaction at Genesis: The Leading Edge, **26**, 644–647.
[14] Rickett, J., and D. E. Lumley, 2001, Cross-equalization data processing for time-lapse seismic reservoir monitoring: A case study from the Gulf of Mexico: Geophysics, **66**, 1015–1025.
[15] Stucchi, E., A. Mazzotti, and S. Ciuffi, 2005, Seismic preprocessing and amplitude cross-calibration for a time-lapse amplitude study on seismic data from the Oseberg reservoir: Geophysical Prospecting, **53**, 265–282.
[16] Tura, A., T. Barker, P. Cattermole, C. Collins, J. Davis, P. Hatchell, K. Koster, P. Schutjens, and P. Wills, 2005, Monitoring primary depletion reservoirs using amplitudes and time shifts from high-repeat seismic surveys: The Leading Edge, **24**, 1214–1221.

Fractal heterogeneities in sonic logs and low-frequency scattering attenuation

*Thomas J. Browaeys and Sergey Fomel**

ABSTRACT

Cycles in sedimentary strata exist at different scales and can be described by fractal statistics. We use von Kármán's autocorrelation function to model heterogeneities in sonic logs from a clastic reservoir and propose a nonlinear parameter estimation. Our method is validated using synthetic signals, and when applied to real sonic logs, it extracts both the fractal properties of high spatial frequencies and one dominant cycle between 2.5 and 7 m. Results demonstrate non-Gaussian and antipersistent statistics of sedimentary layers. We derive an analytical formula for the scattering attenuation of scalar waves by 3D isotropic fractal heterogeneities using the mean field theory. Penetration of waves exhibits a high-frequency cutoff sensitive to heterogeneity size. Therefore shear waves can be more attenuated than compressional waves because of their shorter wavelength.

INTRODUCTION

Propagation of waves in heterogeneous media involves attenuation and dispersion by scattering. Theoreticians are still challenged by the phenomenon of wave propagation in random media. The mean field theory (7; 20; 50) is commonly used and provides both dispersion and attenuation, depending on scattering cross-sections of the heterogeneities (55; 60; 19), which are described by their statistical spatial autocorrelation. Higher order correlations have been more recently incorporated in a frequency-dependent effective medium theory (8). One major advance, pointed out by (57; 60), is the restriction of the validity of the mean field formalism to low frequency. The theory in fact includes destructive interferences, caused by averaging different realizations of the random medium, and overestimates attenuation at high frequencies. Alternative solutions have been proposed to remove this artificial decoherence of the phase. Two important examples are radiative transfer theory (59; 15) and the Rytov approximation (37), which is more adequate than the Born approximation when phase fluctuations are important.

In the area of seismic imaging, the layered structure of sediments led (33) to introduce the fundamental concept of stratigraphic filtering. The empirical formula

*e-mail: sergey.fomel@beg.utexas.edu

postulated by O'Doherty and Anstey was demonstrated in 1D using mean field formalism (4; 35) and, alternately, using wave localization theory (42; 41; 40), with a recent extension to a larger frequency band for acoustic waves in 3D (32). The wave localization method utilizes phase, and logarithm of the amplitude, which have the property of self-averaging over some distance called the localization length, in a stationary random medium. These quantities are exactly the ones used in the Rytov method and avoid the phenomenon of artificial phase decoherence at high frequencies. Multiple scattering of seismic waves remains a complex and active research area.

The fractal property of subsurface heterogeneities was initially discussed by (16) for hydrocarbon reservoirs and has been confirmed in both vertical and horizontal wells (45). The study of seismic scattering by 2D numerical wave propagation (10) demonstrates the necessity of self-similar heterogeneities for modeling both observations of coda waves and of traveltime anomalies. A $1/f$ spectrum of heterogeneities was modeled using the von Kármán spatial autocorrelation function (53) in order to obtain a constant quality factor at high frequencies. For scatterers larger than the wavelength, multipathing was observed, whereas 3D effects were revealed to be important for scattering loss at low frequency. (11) tried to explain seismic-wave attenuation in VSP surveys by 3D scattering from fractal heterogeneities. One of the first attempts to relate the statistics of well log data to seismic scattering used wave localization theory (56). A more detailed study of acoustic-wave localization effects in 1D fractal media (51) shows that a constant quality factor is possible only for the $1/f$ fractal spectrum and that localization can not occcur if the medium contains periodic layers involving resonance and violating the ergodicity assumption. Convergence of the localization effect in realistic 3D seismic surveys seems questionable. Presence of strong cycles in well log data is causing difficulties when the fractal exponent is being estimated (9) and is commonly attributed to Milankovitch cycles (1).

Further investigation of the relationship between cycles, fractal properties, and correlation lengths is necessary and low-frequency scattering theories in 3D fractal media can be appropriate for conventional seismic surveys. In this paper, we propose a nonlinear estimation method for fractal statistics of sonic-log heterogeneities using von Kármán's model. We attempt to identify different scales of the sedimentation process as proposed by (33) and (1). The inversion captures small-scale heterogeneities while larger local cycles exist. We use the mean field theory to calculate analytical solution of low-frequency attenuation by scattering from 3D fractal heterogeneities and predict a shift of the dominant frequency with depth in seismic surveys.

The paper is organized into two parts. In the first part we present a description of cycles in sediments in connection with fractal statistics. The von Kármán spatial autocorrelation function is introduced, and we briefly review some features of fractal statistics. We present our estimation method, validate it on synthetic signals, apply it to our sonic-log data, and show that one can detect the high-frequency part of the superposition of different geological scales. The second part explains the derivation of scattering attenuation for low-frequency acoustic waves by 3D isotropic fractal heterogeneities using the mean field theory. Results comply with the Rayleigh regime

and the Backus effective medium for very low frequency. We present analytical predictions of scattering attenuation and show the existence of a cutoff frequency for the penetration of waves. We use the frequency dependence of the penetration depth to calculate the shift of the dominant frequency of a Ricker wavelet. We conclude by suggesting further improvements.

STATISTICAL MODEL OF HETEROGENEITIES

Let us consider the spatial fluctuations of seismic velocities to be small and to constitute a second-order stochastic process. We describe the fluctuations by using different realizations of the random function $f(\mathbf{x})$ with the expectation value $\langle f \rangle = 0$ and with the spatial covariance depending on the relative distance $\mathbf{r}$ defined by

$$\langle f(\mathbf{x}) f(\mathbf{x}+\mathbf{r}) \rangle = \sigma^2 N(\mathbf{r}), \tag{1}$$

where σ is the standard deviation and $N(\mathbf{r})$ is the spatial autocorrelation function with $N(\mathbf{0}) = 1$. The energy spectrum $E^{(s)}(\mathbf{k})$ of the fluctuations in s dimensions ($s = 1, 2, 3$) is related to the autocorrelation by the Wiener-Khintchine theorem (6):

$$E^{(s)}(\mathbf{k}) = |F(\mathbf{k})|^2 = \sigma^2 \int N(\mathbf{r}) e^{-i\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}, \tag{2}$$

$$F(\mathbf{k}) = \int f(\mathbf{x}) e^{-i\mathbf{k}\cdot\mathbf{x}} d\mathbf{x}, \tag{3}$$

where $\mathbf{k}$ is the spatial wave vector and $F(\mathbf{k})$ is the Fourier transform of $f(\mathbf{x})$. The energy spectrum in equation 2 can be simplified, for an isotropic correlation function, to

$$E^{(1)}(k) = 2\,\sigma^2 \int_0^\infty N(r) \cos(kr) dr, \tag{4}$$

$$E^{(3)}(k) = \frac{4\pi}{k}\,\sigma^2 \int_0^\infty r N(r) \sin(kr) dr, \tag{5}$$

where $k = |\mathbf{k}|$. The von Kármán autocorrelation function $N_{H,b}(\mathbf{r})$ describes a self-affine medium relevant for geological structures (13; 17; 9; 38; 21; 12). This function was initially derived by (53) while studying the velocity field in a turbulent fluid and has been used to describe heterogeneous media (47; 10). The Fourier transform of $N_{H,b}(\mathbf{r})$ was given by (26). The statistical autocorrelation $N_{H,b}(\mathbf{r})$ and the energy spectrum $E^{(s)}_{H,b}(\mathbf{k})$ in the Fourier domain are

$$N_{H,b}(\mathbf{r}) = \frac{2^{(1-H)}}{\Gamma(H)}\,(r/b)^H K_H(r/b), \tag{6}$$

$$E^{(s)}_{H,b}(\mathbf{k}) = \sigma^2\, C^{(s)}_H \frac{(2b)^s}{(1+b^2k^2)^{H+\frac{s}{2}}}, \tag{7}$$

$$\text{with} \qquad C^{(s)}_H = \left| \frac{\Gamma(H+\frac{s}{2})}{\Gamma(H)} \right| \pi^{\frac{s}{2}}, \tag{8}$$

where $r = |\mathbf{r}|$, K_H is the modified Bessel function of the second kind with order H, and Γ is the Gamma function. Parameters describing the heterogeneities are characteristic distance b, below which the distribution is fractal, and exponent H, characterizing the roughness of the medium. We use the energy spectrum in equation 7 with $s = 1$ to analyze sonic logs and with $s = 3$ to predict 3D scattering attenuation.

Fractal statistics

Among different concepts introduced by the theory of fractals (28), self-affine property accounts for invariance of roughness of a curve observed at different scales. Self-affine fractals can be characterized by the power-law dependence of their energy spectrum $E(f)$ on frequency f:

$$E(f) \quad \propto \quad f^{-\beta}. \tag{9}$$

The exponent β, in the energy spectrum $E^{(s)}_{H,b}(\mathbf{k})$ from equation 7, is

$$\beta \quad = \quad 2H + s. \tag{10}$$

For $\beta = 0$, energy spectrum is constant and describes the familiar white noise. Causal integration of Gaussian white noise produces the classical Brownian motion, or random walk, characteristic of diffusion processes, and results in an energy spectrum with $\beta = 2$. The autocorrelation function of Brownian motion signals is a decreasing exponential and the autocorrelation in equation 6 properly reduces to $\exp[-r/b]$ for $H = 0.5$. Another interesting form of spectrum is for $\beta = 1$. The associated signal is called Flicker noise (39; 9) and can be interpreted as the superposition of different relaxation processes. For geological layers, such form of spectrum was interpreted as the expression of quasi-cyclicity and blocky layering (44). Generalization, including Gaussian white noise and Brownian motion, leads to two types of fractal signals (44; 49; 25), namely

- fractional Gaussian noise (fGn) defined as filtered Gaussian white noise with $-1 \leq \beta \leq 1$,
- fractional Brownian motion (fBm) built by causal integration of fGn above and resulting in $1 \leq \beta \leq 3$.

The fGn is stationary and Gaussian, whereas the fBm is neither stationary nor Gaussian. Exponent β_{fBm} of the fBm is related to exponent β_{fGn} of the fGn, used for integration, by $\beta_{fBm} = \beta_{fGn} + 2$.

The significance of parameter H in equation 10 is delicate and connected to the Hurst exponent Hu, which measures the correlation of time series (18) by

$$Hu \quad = \quad \frac{\log\left(R/\sqrt{S}\right)}{\log(T)}, \tag{11}$$

where R and S are respectively range of variations and variance calculated for the length T of the signal. The meaning of the value of Hu is

- antipersistence for $0 \leq Hu \leq 0.5$,
- random process for $Hu = 0.5$, and
- persistence for $0.5 \leq Hu \leq 1$.

Estimation of Hu using formula 11 is relevant only for fGn signals (49). For example, Gaussian white noise produces $Hu = 0.5$. Parameter H defined in equation 10 is associated with the Hurst exponent by

- $H = Hu - 1$ for fGn with $-1 \leq \beta \leq 1$;
- H equals the Hurst exponent of incremented fGn (25) for fBm with $1 \leq \beta \leq 3$.

The different self-affine 1D fractal models are presented in Table 1 according to the nature and persistency of the signal. A previous analysis of the logarithm of acoustic impedance from well data by (54) shows $1/2 \leq \beta \leq 3/2$, promoting the so-called $1/f$ geology. An improved solution, reproducing the complete well log sequence in sedimentary rocks, uses a similar random process based on fractional Lévy motion (34), but fBm can be adequate at small scales, inside different facies (27).

Fractal exponent	Von Kármán exponent H	*Description*	Geology
$\beta = 0$	-0.5	Gaussian white noise	Random process
$0 < \beta < 1$	$-0.5 < H < 0$	Persistent fGn	
$\beta = 1$	0	Flicker noise	Blocky layers
$1 < \beta < 2$	$0 < H < 0.5$	Antipersistent fBm	Quasi-cyclic deposition
$\beta = 2$	0.5	Brownian walk	Random deposition
$2 < \beta < 3$	$0.5 < H < 1$	Persistent fBm	Transitional deposition

Table 1: Classification of 1D fractal statistics according to the exponent β and geological interpretation.

Synthetic realizations

Our method of synthesizing correlated random media is summarized in Table 2 for fractional Gaussian noise (fGn) and fractional Brownian motion (fBm). The Gaussian nature of the initial white noise is contained in the phase of its Fourier transform $G(\mathbf{k})$, whereas the amplitude is constant with frequency because the noise is white. The causal integration, to produce the fBm, is performed by a phase rotation in the

Steps	*Domain*	*Operation for fGn*	*Operation for fBm*
1	Space	Generate Gaussian white noise $g(\mathbf{x})$ with zero mean and unit variance	
2	Fourier	Generate energy spectrum $E^{(s)}_{H,b}(\mathbf{k})$	
3	Fourier	$F(\mathbf{k}) = \sqrt{E^{(s)}_{H,b}(\mathbf{k})}\ G(\mathbf{k})$	$F(\mathbf{k}) = -i\ \mathrm{sign}(\mathbf{k})\ \sqrt{E^{(s)}_{H,b}(\mathbf{k})}\ G(\mathbf{k})$
4	Space	Obtain correlated fGn $f(\mathbf{x})$	Obtain correlated fBm $f(\mathbf{x})$

Table 2: Synthesis of correlated heterogeneous media by generating fractional Gaussian noise (fGn) or fractional Brownian motion (fBm) with the spectrum $E^{(s)}_{H,b}(\mathbf{k})$. The spatial Fourier transforms in s dimensions of the distributions $f(\mathbf{x})$ and $g(\mathbf{x})$ are respectively $F(\mathbf{k})$ and $G(\mathbf{k})$ with $|G(\mathbf{k})| = 1$, and $\mathrm{sign}(\mathbf{k}) = \mathbf{k}/|\mathbf{k}|$.

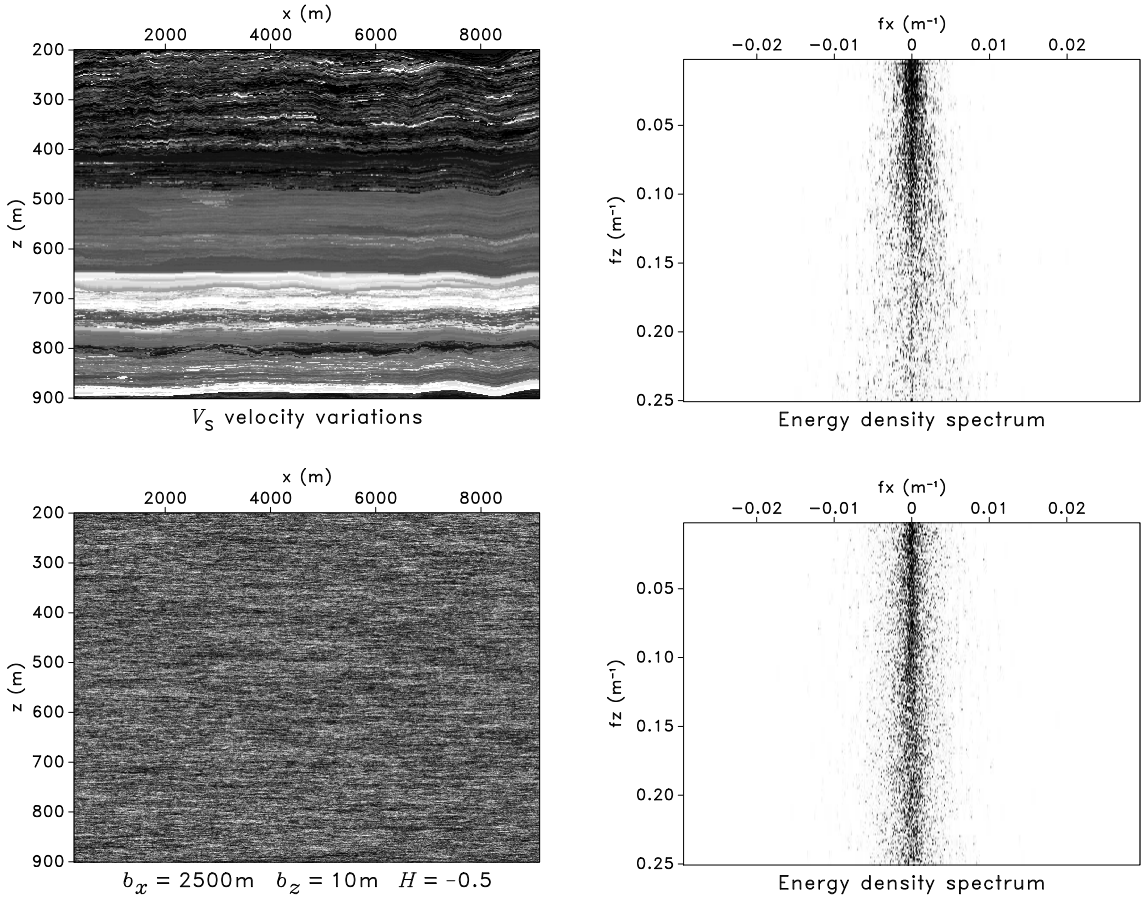


Figure 1: Variations of V_S in a high-resolution reservoir model based on seismic and well data from a field in Canada (a) and its spectral energy density (b). Synthetic realization of 2D fGn using the von Kármán spectral amplitude with the exponent $\beta = 1$ and elliptical anisotropy (c, d). atten/enerd signal,f2dfile,cgaussb,fcgaussbp

Fourier domain, strictly equivalent to the Hilbert transform. The fractal property, below spatial scale b, is imposed by the amplitude $\sqrt{E^{(s)}_{H,b}}$ of the von Kármán model.

For dimension $s = 2$, the exponent in equation 7 is $\beta = 2H + 2$. Therefore, the energy spectrum of the 2D fGn with $H = -0.5$ in Figure 1 is $E(f) \propto 1/f$. (21) in fact proposed using $-0.5 \leq H \leq 0$ to synthesize geologically realistic 2D models with the von Kármán function. Because sediments are made up of layers, we consider the autocorrelation function to be vertical transverse isotropic. We use two different correlation lengths b_x and b_z for horizontal and vertical directions, and define the Riemannian relative distance (13):

$$r/b = \sqrt{(r_x/b_x)^2 + (r_z/b_z)^2}. \tag{12}$$

Figure 1 shows, for comparison, the signal and associated energy spectrum of the synthetic fGn and of a 2D section from a high-resolution model of a clastic reservoir in Canada. The spectrum of the synthetic heterogeneous medium is similar to the one from the reservoir model, but the synthetic fGn, although exhibiting some comparable roughness in the space domain, does not contain coherent, large geological structures, i.e. folded beds.

Nonlinear parameter estimation on sonic well logs

We propose to use the synthesis of a random medium detailed in Table 2 for s=1 as a basis for the procedure to estimate heterogeneity parameters from sonic logs. We achieve optimization by using a weighted least-squares method in the spectral domain on the logarithm of the amplitude, with the model derived from equation 7 :

$$\ln|F(k)| = \ln|F(0)| - p \ln\left[1 + (kb)^2\right], \tag{13}$$

$$p = \frac{H}{2} + \frac{1}{4} \quad \text{and} \tag{14}$$

$$|F(0)| = \sigma \sqrt{2b\, C^{(1)}_H}\,. \tag{15}$$

We estimate the three parameters, b, H, and σ, using a separable least-squares method (14) for $\ln|F(0)|$ and the slope p, and a Gauss Newton optimization algorithm on the nonlinear parameter b^2. Parameter $\ln|F(0)|$ is included in the optimization algorithm because it is difficult to estimate directly from the zero-frequency component in the data. Standard deviation σ, extracted from relation 15, is confirmed by direct evaluation on the spatial signal. When applying the method, we first substract the signal expectation and use it as a scaling factor. We have tested the efficiency of the algorithm on synthetic fGn and fBm generated by the procedure in Table 2 with a discrete length of 4056 points. Three synthetic fractal signals and their parameter estimations are shown in Figure 2. Results of the validation tests are presented in Table 3.

Well log data come from a sandy channel reservoir with a clastic overburden, and the facies evolves from silty sandstone to mudstone, which is characteristic of alluvial

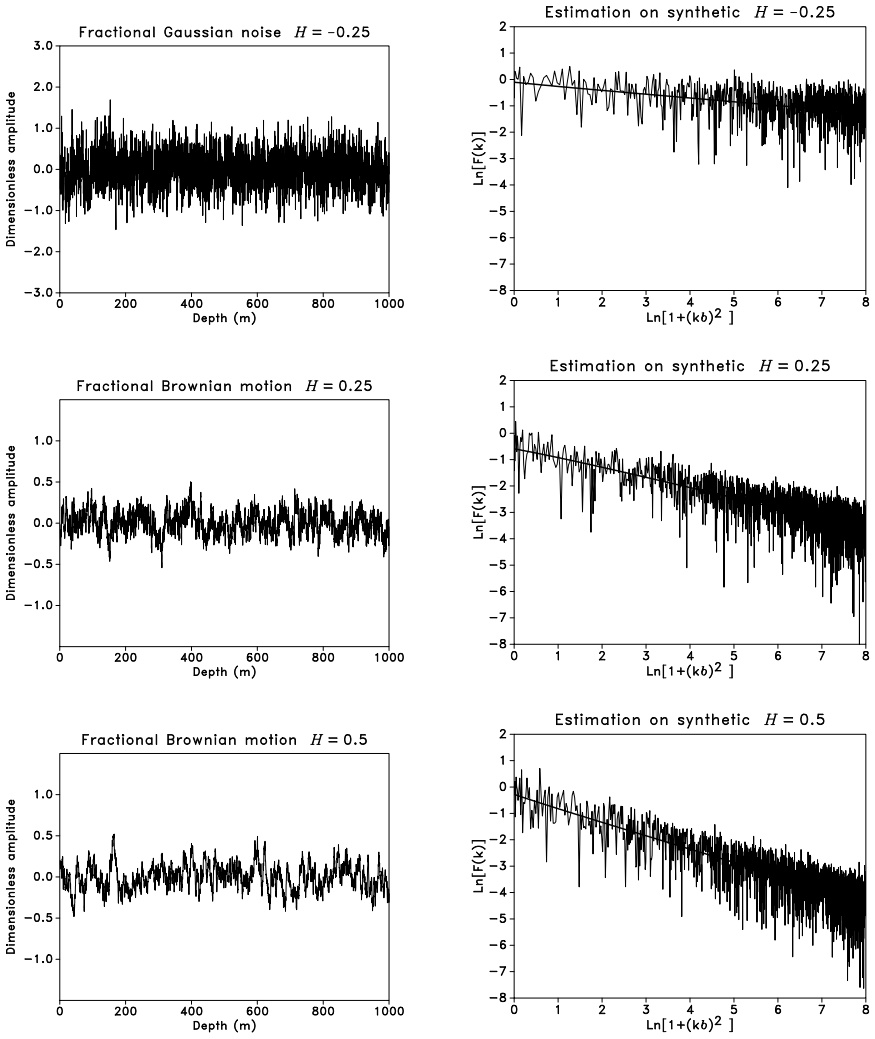


Figure 2: Synthetic signals generated as fGn with $H = -0.25$, $b = 10\,\mathrm{m}$, $\sigma = 20$ % in (a); fBm with $H = 0.25$, $b = 5\,\mathrm{m}$, $\sigma = 30$ % in (c); and fBm with $H = 0.5$, $b = 5\,\mathrm{m}$, $\sigma = 30$ % in (e). Parameter estimations on the logarithm of the spectral amplitude are shown on the right (b, d, f).

atten/karman1 cgaussM025,lligaussfM025,cgauss025,lligaussf025,cgauss05,lligaussf05

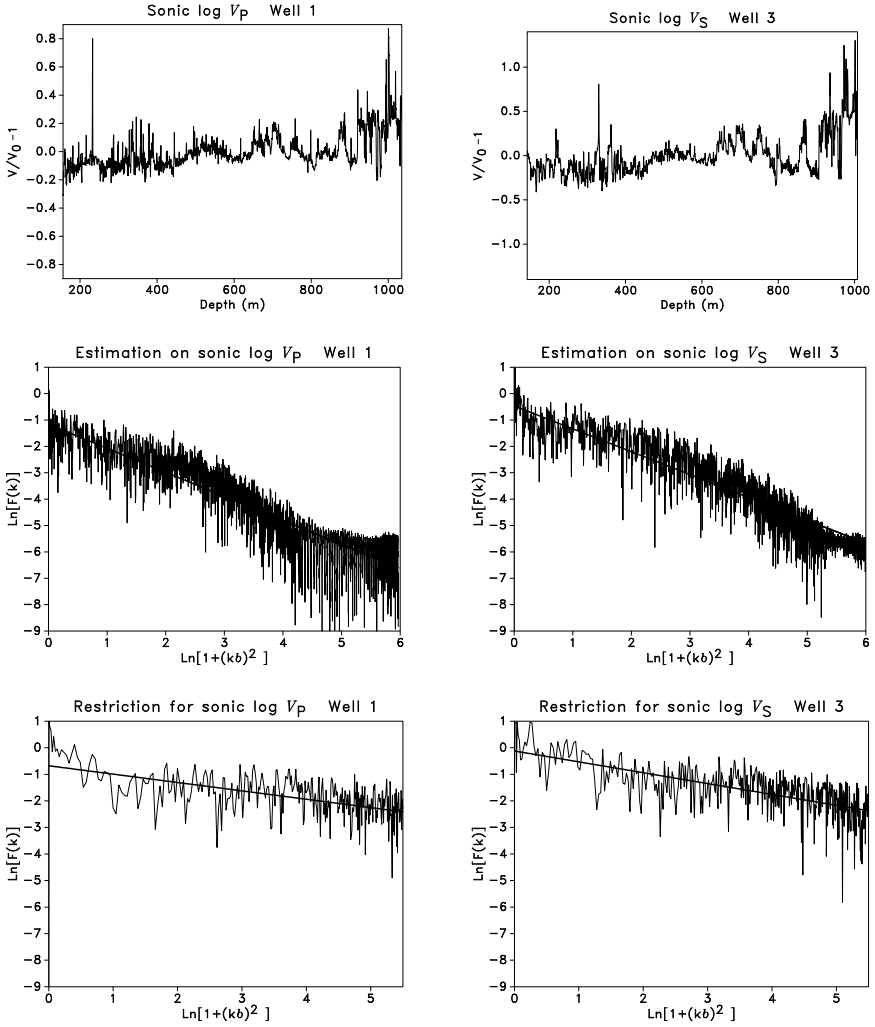


Figure 3: Sonic log V_P (a) from well N°1 and V_S (b) from well N°3, scaled by their respective average value V_0. Parameter estimation on the logarithm of the spectral amplitude (c, d) shows the existence of different slopes for low, medium, and high frequencies. These tool artefacts are removed by restricting the estimation method to low frequency (e, f).

atten/karmand signalA1,signalC2,llisignalfA1,llisignalfC2,rllisignalfA1L,rllisignalfC2L

Parameters	H	b (m)	σ (%)
Generated fGn	-0.25	10.0	20
Recovered	-0.21	11.0	18
Generated fGn	-0.25	5.0	20
Recovered	-0.21	5.9	17
Generated fBm	0.25	10.0	30
Recovered	0.26	10.2	23
Generated fBm	0.50	5.0	40
Recovered	0.51	5.3	32
Generated fBm	0.75	3.0	40
Recovered	0.79	2.7	30

Table 3: Comparison of the stochastic medium parameters used to generate synthetics fGn and fBm and their recovery by the nonlinear estimation method.

deposition. Velocities V_P and V_S were both measured with a spatial sampling of 0.125 m. Figure 3 shows the parameter estimation for two sonic logs. Comparison with the method applied to the synthetics in Figure 2 uncovers the existence of different slopes for different frequencies in Figures 3(c) and 3(d). We can reasonably delimit three domains, denoted (A) for low frequencies, (B) for medium frequencies, and (C) for very high frequencies. These domains can be identified by parameters r_S and r_I, representing specific values of relative distance r, namely

$$\text{(A) for } \; r \geq r_S, \quad \text{(B) for } \; r_S \geq r \geq r_I, \quad \text{(C) for } \; r_I \geq r,$$

where $1\,\text{m} < r_S < 2\,\text{m}$ and $r_I < 1\,\text{m}$. The sharp break (B) in the medium frequencies followed by a white noise (C) at high frequencies is characteristic of the tool artefact. Data acquisition involves a convolution with a box-car window (43; 9). Application of the estimation method is thus restricted to relative distances $r > r_S$, and results are shown in Figures 3(e) and 3(f).

Results are summarized in Table 4 for the four different sonic logs V_P and V_S. The ratio $\langle V_P \rangle / \langle V_S \rangle$ is almost constant for the four well logs and roughly equal to two. The updated estimation in the spatial wavelength domain (A) produces reasonable results in Table 4. Standard deviation σ varies from 20 to 45 % and is larger for V_S logs than for V_P logs. Correlation length b is about 5 m for both V_P and V_S, except for the well N°4, which is 2.5 m. Exponent H for V_P varies from 0.1 to 0.4 and for V_S from 0.2 to 0.6. In Figure 4, comparison of the frequency content of one real sonic log with one realization of a synthetic fBm, generated using similar parameters, shows that the sonic-log data contain higher peaks for very large wavelengths. We detected in the different sonic well logs the recurrence of some particular spatial cycles at 2.5 m, 5 m, 10 m, and 20 m.

Well	Log	b (m)	H	σ (%)	$\langle V \rangle$ (m/s)
N°1	V_P full	0.79	1.32	17	2791
	$r > 1.5$ m	6.70	0.13	22	
	V_S full	1.05	1.16	32	1218
	$r > 2.0$ m	5.92	0.21	35	
N°2	V_P full	1.90	0.92	27	2842
	$r > 1.6$ m	5.34	0.38	29	
	V_S full	2.84	0.83	45	1240
	$r > 1.5$ m	3.08	0.62	44	
N°3	V_P full	1.34	1.16	20	2787
	$r > 1.9$ m	7.22	0.18	21	
	V_S full	1.25	1.23	32	1216
	$r > 1.8$ m	5.01	0.32	36	
N°4	V_P full	0.64	1.98	18	2745
	$r > 1.4$ m	2.58	0.39	38	
	V_S full	0.57	2.26	32	1247
	$r > 1.3$ m	2.46	0.56	33	

Table 4: Parameters estimation from four wells in a clastic overburden, for the full sonic logs, and for limited spatial frequency bandwidths using the indicated restriction on the relative distance r to remove the tool artefacts. Relevant physical values are underlined.

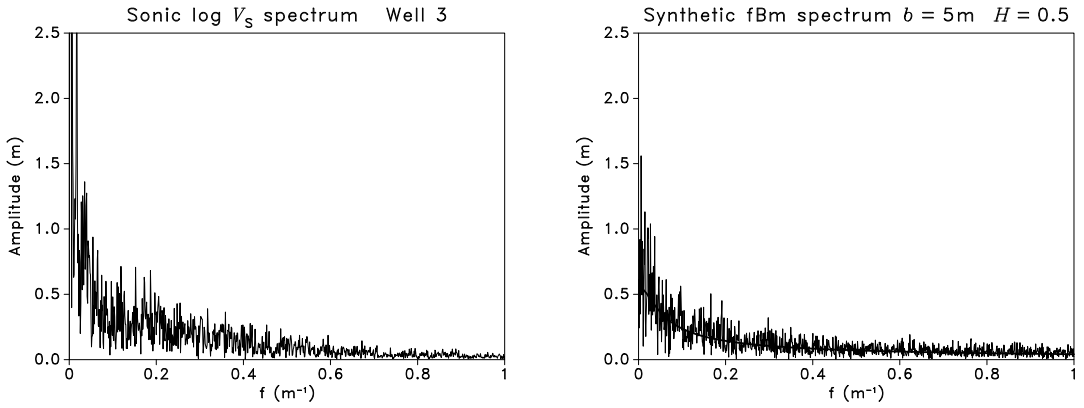


Figure 4: Fourier spectrum of the scaled V_S sonic log from well N° 3 (a). The shape of the low-frequency content is different from that of the Fourier spectrum of the fractional Brownian motion (b) synthesized with the von Kármán model using $H = 0.25$, $b = 5\,\mathrm{m}$, and $\sigma = 30$ %. atten/spectra rfsignalC2,fitfiltb025

Fractal heterogeneities and cycles in sediments

(33) and (1) described variations in well logs by the superposition of different types of deposition, leading to "layers inside layers". Their classification includes

1. A large number of small thickness layers (≤ 1 m) for weakly transitional depositions with small reflection coefficients;
2. Cyclic layers of thicknesses from 1 to 10 m with sharp interfaces, corresponding to fine layering depositions inside a facies for short-period sea cycles; and
3. Horizons imaged by seismic reflection, i.e. different facies for a small number of thicker blocky layers associated with low-order cycles.

They suggested that transmission losses could be compensated by multiple reflections, depending on seismic wavelength. This classification is in agreement with the fact that high exponents, H, appear for shorter scales, b, in Table 4. The estimation performed on the sonic logs indicates fractal properties for distances shorter than $b \simeq 5\,\mathrm{m}$. Acccording to (1), well log signals are the superposition of several processes with different scales. The von Kármán model captures part of it. Parameters extracted by our analysis describe heterogeneities corresponding to type 2 of the O'Doherty-Anstey classification, which is a fractal behavior inside major geological units, at least from 10 down to 1 m, with a correlation length of 5 m. Previous estimations of the correlation length on well logs were produced by direct calculation of the spatial autocorrelation (56; 43). (56) suggested the possibility of superposition of two correlation lengths at 5 and 20 m. The wavelet detection analysis of gamma-ray and resistivity well logs for a sandstone confirmed the strong evidence of local cyclicity in the stratigraphic sequences (36). We think that direct estimation of correlation distance b using the autocorrelation function, or our estimation method, captures the shortest dominant cycle in the sedimentary layers. This would explain why the fractal behavior seems to hold for larger scales in Figures 3(e) and 3(f).

Parameter H, estimated from well logs, is $0 < H \leq 0.5$ and consistent with an antipersistent fractional Brownian motion characteristic of cyclicity (see Table 1). The Hurst exponent commonly exhibits some antipersistence in sediments with values from 0.2 to 0.5 for sandstones (9; 27). High values of 0.5 and 0.6 could be interpreted, in a clastic context, to be caused by a transitional deposition involving persistency, as in natural floods. Natural flood records exhibit a Hurst exponent, $0.5 \leq Hu \leq 1.0$, associated with so-called black noise (18; 29).

Seismic-scale heterogeneities

Inside a main facies, we can not recover the short wavelengths of type 1 in the O'Doherty-Anstey classification because of well-logging tool limitation. Type 3 of the classification includes Milankovitch cycles of about 10 to 20 m and third-order

sea-level cycles from 15 to 300 m (1). Seismic reflectors are conventionally identified as being chronostratigraphic horizons separating different geological units. They correspond to the wavelength of conventional seismic surveys, up to 100 Hz, and should induce some resonant scattering with "friendly" multiples. Multipathing is observed for this ratio of seismic wavelength to the size of heterogeneities, as shown by numerical experiments (10). This domain should be treated using wave-localization theory and the Rytov method. These could nevertheless fail to explain the data because of quasi-periodicity of the medium at this scale, violating the ergodicity assumption required by wave-localization theory. Statistical methods using the autocorrelation function seem to be adapted to describe quasi-periodic media when the ratio of b over the seismic wavelength is small. When the wavelength is of the same size, local quasi-cyclicity of the sedimentary sequence should not be ignored (31; 46).

The non-Gaussian nature and non-stationarity of sedimentary layers call for more sophisticated methods to be used, especially in order to capture larger scale pseudo-cyclic heterogeneities, as, for example, a multifractal analysis (30) or a local cyclicity detection by wavelet analysis (36). The wavelet transform was indeed introduced so that seismic signals in locally cyclic sedimentary layers could be analyzed (31).

SCATTERING ATTENUATION IN 3D

Different scattering regimes exist when waves propagate in heterogeneous media, according to the ratio of the wavelength, λ, to the size, b, of heterogeneities. The formalism including the different scattering regimes, when heterogeneities are modeled by spherical inclusions, is the Mie scattering theory. Recent experimental results (23) on sonic-wave reflectivity in a granular medium, made up of beads of size b in a water tank, illustrate this classification :

1. for low frequencies, when $\lambda > \pi b$, backward scattering is dominant, the Born approximation can be used, and the regime is Rayleigh scattering;
2. for wavelengths similar in size to heterogeneity, when $\pi b > \lambda > \pi b/2$, lateral scattering is important, multiples should not be neglected, and the regime is called resonant scattering; and
3. for high frequencies, when $\pi b/2 > \lambda$, waves are scattered mainly forward, and localization theory and the Rytov formalism are appropriate.

Experiments have demonstrated that wave reflectivity strongly decreases when the wavelength of the incident wave is twice the diameter of the beads, for which lateral scattering starts to be dominant. Mean wavefield formalism is valid only for a low frequency, when the wavelength is larger than the size of heterogeneities because of the assumptions in the derivation (20). The Born approximation can describe the Rayleigh regime and the approach of resonant phase scattering (10; 38). For wavelengths shorter than the size of heterogeneities, artificial decoherence by phase

randomization occurs (57; 38). We intend to describe the 3D attenuation in a stochastic fractal medium, when $kb \leq 1$, which is relevant for seismic survey frequencies. The limit of validity corresponds to wavelengths approaching the size of heterogeneities. We assume heterogeneities to be isotropic. The schemes in Figure 5 compare a realistic geological structure with two different end-member models. Scattering calculations in 1D underestimate the scattering loss by small-scale heterogeneities.

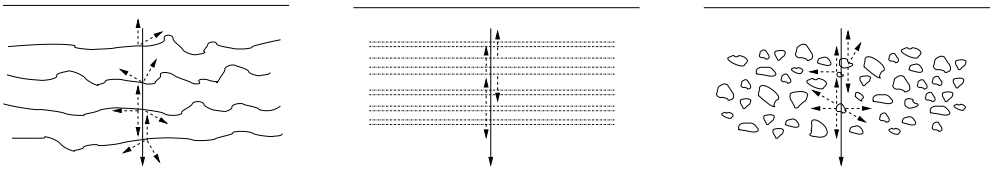


Figure 5: Schematic comparison of single scattering effects, during a vertical wave propagation in sediments, between a realistic geological structure (a) and two end-member models: horizontal layers with propagation including 1D scattering (b) and isotropic heterogeneities with 3D scattering (c). atten/XFig schemescatter3dgg,schemescatterd,schemescatter3d

Low-frequency waves in 3D isotropic heterogeneous media

A scalar wave $u(\mathbf{x},\omega)$ in a weakly inhomogeneous medium (7; 47; 20) satisfies the Helmholtz wave equation

$$\Delta u(\mathbf{x},\omega) + k_0^2 \left[1 + f(\mathbf{x})\right]^2 u(\mathbf{x},\omega) \quad = \quad 0\,, \tag{16}$$

where $f(\mathbf{x})$ is a small perturbation of the medium from homogeneity, and $k_0 = \omega/c_0$. Phase velocity c_0 is the background velocity. Assuming a second-order stationary statistical distribution for fluctuations $f(\mathbf{x})$ and a zero expectation value $\langle f \rangle = 0$, spatial covariance of the velocity variations is defined by relation 1. Expectation $\langle u(\mathbf{x},\omega) \rangle$ of random plane-wave realizations is calculated (20) using a perturbation theory to the second order in $f(\mathbf{x})$ by

$$\left[\Delta + k_0^2(1+\sigma^2)\right] \langle u(\mathbf{x},\omega)\rangle - 4\,k_0^4\,\sigma^2 \int N(\mathbf{x}'-\mathbf{x})\, G(\mathbf{x},\mathbf{x}',\omega)\, \langle u(\mathbf{x}',\omega)\rangle \; d\mathbf{x}' = 0\,, \tag{17}$$

where $G(\mathbf{x},\mathbf{x}',\omega)$ is Green's function of the operator $[\Delta + k_0^2]$, and integration is performed over the 3D space. The dispersion relation for a plane wave propagating in the heterogeneous medium follows as

$$\frac{k^2}{k_0^2} \quad = \quad 1 + \sigma^2 \left[1 - 4\,k_0^2 \int N(\mathbf{r})\, G(\mathbf{r},\omega)\, e^{i\,\mathbf{k}\cdot\mathbf{r}} d\mathbf{r}\right], \tag{18}$$

$$G(\mathbf{r},\omega) \quad = \quad \frac{e^{ik_0 r}}{4\pi r}\,, \tag{19}$$

where $\mathbf{r} = \mathbf{x}' - \mathbf{x}$ is the relative position, with an absolute value $r = |\mathbf{r}|$, and $G(\mathbf{r}, \omega)$ is the 3D isotropic free-space Green's function with the outward radiation condition (5). The path of waves should be sufficiently long to significantly sample medium heterogeneities statistically (11). The Born approximation is present because of Green's function. Heterogeneities with the isotropic correlation function $N(r)$ produce an isotropic wave vector $\mathbf{k}$. Combining Green's function in equation 19 with the isotropic integral in equation 5 reduces the squared dispersion relation of equation 18 to

$$\frac{k^2}{k_0^2} = 1 + \sigma^2 \left[1 - \frac{4\,k_0^2}{k} \int_0^\infty N(r)\, e^{ik_0 r} \sin(kr)\, dr \right]. \tag{20}$$

Second-order expansion $k/k_0 = 1 + O(\sigma^2)$ in the solution constrains validity to the domain $k_0 b \ll 1/\sigma$, where b is the characteristic length scale of the heterogeneities. The second-order approximation for the 3D dispersion relation is finally

$$\frac{k}{k_0} = 1 + \frac{\sigma^2}{2} + i\,\sigma^2\, k_0\, [S(0) - S(2k_0)]\,. \tag{21}$$

Quantity $S(k)$, introduced above, is related to the real and even function $E^{(1)}(k)$, defined by the isotropic integral of equation 4 :

$$S(k) = \int_0^{+\infty} N(r) e^{ikr} dr\,, \tag{22}$$

$$E^{(1)}(k)\ \sigma^{-2} = \int_{-\infty}^{+\infty} N(r) e^{ikr} dr = 2\ \mathrm{Re}[S(k)]\,, \tag{23}$$

$$2\,i\,\mathrm{Im}[S(k)] = S(k) - S(-k). \tag{24}$$

Connection to the O'Doherty-Anstey formula is detailed in Appendix A.

Attenuation in 3D fractal media

The energy spectrum, $E^{(1)}_{H,b}(k)$, of von Kármán's autocorrelation function $N_{H,b}(r)$ in equation 7 is real and even :

$$E^{(1)}_{H,b}(k)\ \sigma^{-2} = \int_{-\infty}^{+\infty} N_{H,b}(r)\, e^{ikr} dr\,, \tag{25}$$

$$= C_H^{(1)}\, \frac{2b}{(1 + b^2k^2)^{H+\frac{1}{2}}}\,. \tag{26}$$

Values of $S(k)$ defined by equation 22 are

$$S(0) = C_H^{(1)}\, b\,, \tag{27}$$

$$Re[S(2k_0)] = \frac{C_H^{(1)}\, b}{(1 + 4\,b^2k_0^2)^{H+\frac{1}{2}}}\,. \tag{28}$$

Coefficient $C_H^{(1)}$, defined by equation 8, is an increasing function of exponent H and has to be calculated numerically, except for some specific values :

$$\begin{array}{lll} C_H^{(1)} \sim \pi/\Gamma(H) \to 0 & \text{for} \quad H \to 0\,, & \\ C_{-0.25}^{(1)} = 1.3110\ldots\,, & C_{0.25}^{(1)} = 0.5991\ldots\,, & C_{0.5}^{(1)} = 1\,, \\ & C_{0.75}^{(1)} = 1.3317\ldots\,, & C_1^{(1)} = \pi/2\,. \end{array}$$

The dispersion relation of equation 21 solves for an explicit solution of attenuation and dispersion :

$$Re[k/k_0] \;=\; 1 + \frac{\sigma^2}{2}\left(1 + 2k_0\; Im[S(2k_0)]\right)\,, \tag{29}$$

$$Im[k/k_0] \;=\; \sigma^2 k_0 b\; C_H^{(1)} \left[1 - \frac{1}{(1+4\,b^2k_0^2)^{H+\frac{1}{2}}}\right]. \tag{30}$$

When $H = 0.5$, the derivation produces simple expressions as detailed in Appendix B. The use of $S^*(k) = S(-k)$ with the Kramers-Krönig relation can be used to determine the real part of k. In the context of the second-order approximation, scattering attenuation in a von Kármán isotropic medium is

$$\frac{1}{Q} = \frac{2\,Im[k]}{Re[k]} = 2\;\sigma^2\,k_0 b\;C_H^{(1)} \left[1 - \frac{1}{(1+4\,b^2k_0^2)^{H+\frac{1}{2}}}\right]. \tag{31}$$

For $k_0 b \ll 1$, the scattering attenuation reduces to the Rayleigh diffusion regime :

$$\frac{1}{Q} \;\simeq\; 8\,\sigma^2\,C_H^{(1)} \left(H + \frac{1}{2}\right)(k_0 b)^3\,. \tag{32}$$

Penetration depth

Waves propagating in disordered media are exponentially attenuated by scattering (33; 56). We define penetration depth $d(f)$ to be the skin depth (52) for low-frequency waves propagating in the heterogeneous medium :

$$\frac{1}{d(f)} \;=\; \frac{k_0}{2Q} = Im[k]\,, \tag{33}$$

$$d(f) \;=\; \frac{b\;(1+4\,b^2k_0^2)^{H+\frac{1}{2}}}{\sigma^2\,(k_0 b)^2\,C_H^{(1)}\left[(1+4\,b^2k_0^2)^{H+\frac{1}{2}} - 1\right]}\,, \tag{34}$$

where f is frequency. Penetration depth $d(f)$ corresponds to a decrease of wave amplitude by $1/e$. For two-way traveltime, recorded amplitude is 14 % of the initial one. Figure 6 shows the frequency dependence of the penetration depth for different values of parameters b and H and the scattering attenuation for acoustic P- and S-waves. Seismic background velocities are $V_P = 2700$ m/s and $V_S = 1230$ m/s.

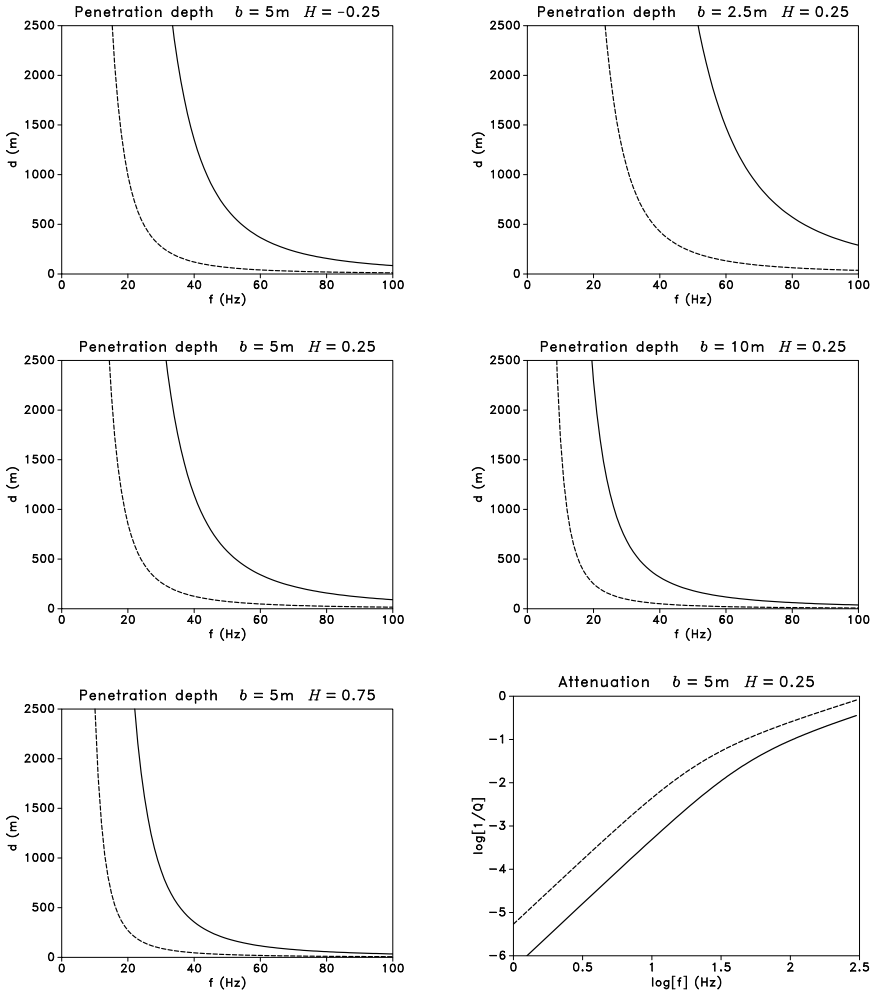


Figure 6: Penetration depth for P (solid) and S (dashed) scalar waves in heterogeneous media described by the von Kármán model with $\sigma = 30$ %. For $b = 5\,\mathrm{m}$, a higher exponent H decreases the penetration (a, c, e). The value of b strongly influences the penetration of waves (b, d). The slope break in the attenuation curve (f) determines the frequency below which our scattering model is valid.

atten/qcurve depthlfb05M025,depthlfb25025,depthlfb05025,depthlfb10025,depthlfb05075,llqfb05025

Scattering attenuation $1/Q$ is proportional to $1/\lambda^3$ at low frequencies and corresponds to the Rayleigh diffusion regime. Attenuation increases at higher frequencies, where $1/Q$ is proportional to $1/\lambda$ and wavelength is comparable to the size of the heterogeneities. Nevertheless, the validity of our scattering theory is constrained to the low-frequency bandwidth until the attenuation curves in Figure 6(f) reach the change of slope at 45 Hz for S-waves and 70 Hz for P-waves. For a conventional seismic survey and for correlation length $b = 5\,\mathrm{m}$, previously estimated, wavelengths of P- and S-waves and ratio λ/b are indicated in Table 5.

Wave	Frequency (Hz)	λ (m)	λ/b
P	10	270	54
	90	30	6
S	10	123	24
	50	25	5

Table 5: Ratio of the wavelength λ over the size of heterogeneities $b = 5\,\mathrm{m}$ for realistic seismic frequencies, when $V_P = 2700$ m/s, $V_S = 1230$ m/s.

Scattering is more important for seismic wavelengths with a dimension similar to that of heterogeneities : high frequencies and S-waves, because their wavelengths are shorter than P-waves, are more attenuated. Penetration depth is close to infinity at very low frequencies but decreases drastically in a narrow frequency band, depending on parameters H and b (see Figure 6). This steep descent shifts to higher frequencies when the fractal exponent decreases, corresponding to a stronger cyclicity of the layers. A shorter correlation length of heterogeneities highly improves penetration of high frequencies for both types of wave (see Figures 6(b) and 6(d)). For large-size heterogeneities, i.e. $b > 20$ m, the scattering theory we use is not valid because the seismic frequencies statisfy $k_0 b \geq 1$. Scattering regimes and the suggested description of heterogeneities are summarized in Figure 7. Our results therefore ignore the effects of large cycles in sediments. We refer the reader to (46) for more information on the effects of cycles on wave progagation.

Dominant frequency versus depth

If seismic pulse is defined as a Ricker wavelet, a relation can be derived for modification of the frequency content of P and S acoustic waves by scattering attenuation. Dominant frequency $f_{dom}(z)$ with depth z and initial spectrum $S_I(f)$ of the source are defined by

$$S_I(f) = \left(\frac{f}{f_0}\right)^2 e^{-f^2/f_0^2}, \tag{35}$$

$$f_{dom}(z) = \frac{d}{df}\left(S_I(f)\, e^{-z/d(f)}\right), \tag{36}$$

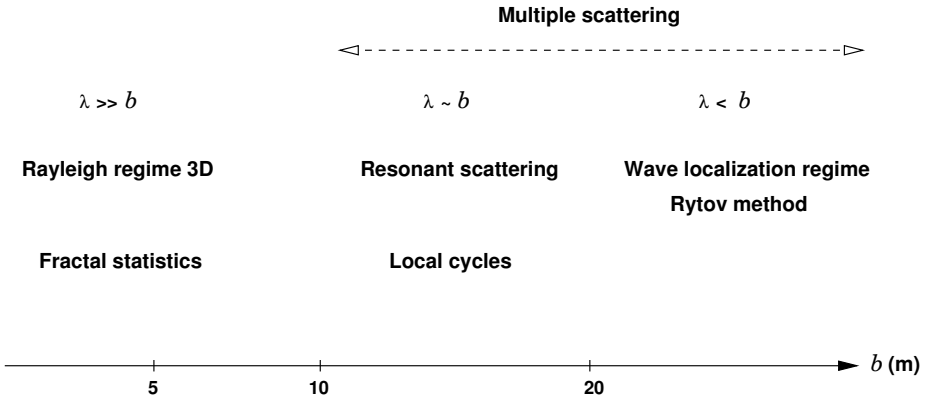


Figure 7: Schematic representation, including the suggested description of heterogeneities, of the different scattering regimes depending on the ratio between the seismic wavelength λ and the size of heterogeneities b. The scale is for indication only and depends on the frequency band of the survey. atten/XFig schemeshetb

with initial condition $f_{dom}(0) = f_0$, and where $d(f)$ is the penetration depth defined in equation 34. Dispersion involves different traveltimes at different frequencies but does not modify the frequency content or amplitude. For convenience, we estimate dominant frequency as frequency expectation :

$$f_{dom}(z) = \frac{1}{\langle S_I \rangle_z} \int f\, S_I(f)\, e^{-z/d(f)}\, df\,, \tag{37}$$

$$\langle S_I \rangle_z = \int S_I(f)\, e^{-z/d(f)}\, df\,. \tag{38}$$

Figure 8 shows the evolution of the dominant frequency with depth in fractal media, with $V_P = 2700$ m/s, $V_S = 1230$ m/s, and standard deviation $\sigma = 30\,\%$. The value of correlation length b again has a very high impact, whereas the fractal exponent moderately influences results. For a multicomponent seismic survey in a clastic reservoir, evolution of the peak frequency should show a more important decrease with depth for PS data than for PP data.

DISCUSSION

The attenuation caused by anticorrelated 3D small-scale heterogeneities can be explained by a low-frequency scattering theory. The length scales that we estimated from sonic logs justify this approach for conventional seismic frequencies. The intensity of scattering attenuation and the value of frequency cutoff strongly depend on the size of the heterogeneities, and S-waves are more attenuated than P-waves at the same frequency. Using low-frequency P-waves provides a better depth of penetration. More reflectors can be detected and imaged, but, of course, with less resolution. This phenomenon was observed in sub-basalt imaging (61). Our analysis of sonic

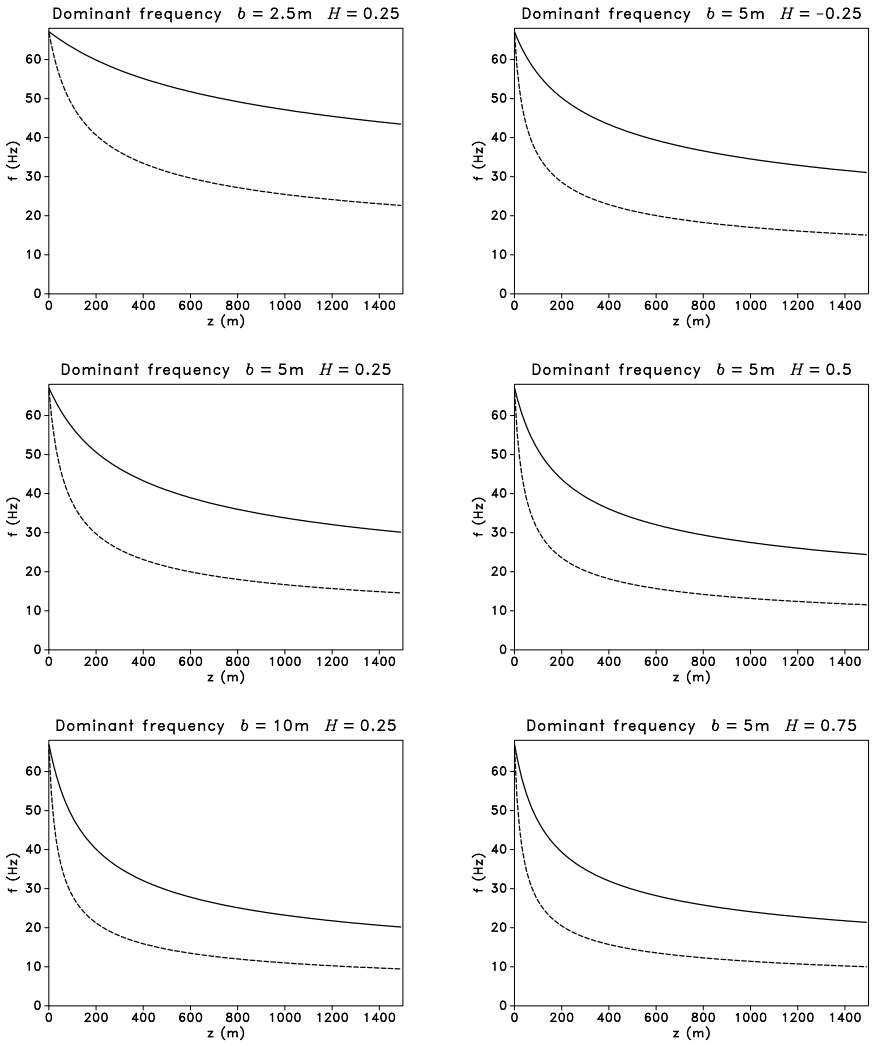


Figure 8: Evolution of the dominant frequency with depth for P (solid line) and S (dashed line) scalar waves modeled by a Ricker wavelet ($f_0 = 60\,\mathrm{Hz}$) in heterogeneous media with $\sigma = 30$ %. For a constant exponent $H = 0.25$, the dominant frequency shifts to lower frequencies faster for larger values of b (a, c, e). The exponent H weakly influences the evolution of the dominant frequency (b, d, f).

atten/pdepth fdomfb25025,fdomfb05M025,fdomfb05025,fdomfb0505,fdomfb10025,fdomfb05075

logs confirms the relevance of a fractal description for the high-frequency content of quasi-periodic geological layers. Because sediments are highly stratified, their layered structure has previously motivated use of 1D models for seismic scattering attenuation, but a realistic estimate needs to be conducted in 3D. More generally, description of geological heterogeneities and use of scattering theories should be as depicted in Figure 7.

Several limitations should be pointed out in our study. Well logs constitute 1D samples of the geological medium in the near vertical direction. Use of 3D isotropy is the simplest assumption consistent with our limited knowledge. We consequently ignore the anisotropic effect of layering, which undoubtly affects lateral scattering when $\lambda \sim b$. Also note that the relation between fractal exponent β and parameter H depends on spatial dimension s. We have therefore proposed different values of parameter H for depth of penetration and evolution of dominant frequency in Figures 6 and 8 and have attempted to extract some general trends.

Our analysis is limited by the fact that we used variations of the seismic velocity, but not of density, in order to be consistent with the mean field theory. This theory retrieves the Backus limit and the Rayleigh diffusion regime. Fortunately, densities commonly exhibit fewer variations. Meanwhile, analysis of the logarithm of impedance Z relates directly to the reflection coefficients (44). Backscattering is known to be related to impedance fluctuations (4; 58). Modification of the Helmholtz equation (16) in 1D to incorporate $\ln Z$ was succesfully achieved by (4), and they provided a proof of O'Doherty-Anstey formula. Impedance $Z(z)$, depending on depth z, and reflection coefficient series $R(z)$ are connected by

$$\lim_{dz\to 0} R(z+dz/2) = \lim_{dz\to 0} \frac{Z(z+dz)-Z(z)}{Z(z+dz)+Z(z)}, \tag{39}$$

$$R(z) = \frac{1}{2}\, d\ln Z(z). \tag{40}$$

Reflection series can reasonably be considered to be Gaussian and stationary only inside blocky layers, using the segmentation method (48). Incrementation of the fractional Gaussian noise, corresponding to reflection series $R(z)$, produces the nonstationary and non-Gaussian fBm describing quantity $\ln Z(z)$ (44). A white spectrum of reflectivity coefficients occurs for $\beta = 2$ and generates a Brownian walk describing $\ln Z(z)$ and involving an exponential autocorrelation. We advocate the use of a nonwhite reflectivity hypothesis, as previously recommended by several authors (48; 22; 2).

CONCLUSIONS

The high-frequency quasi-cyclic variations of seismic velocities can be described as an antipersistent fractional Brownian motion as demonstrated by our sonic-log data from a clastic reservoir. The correlation length, estimated for von Kármán's model, is about 5 m, but the sonic logs contain larger local cycles at 10 and 20 m : our method

extracts one dominant cycle of deposition. Conventional seismic surveys contain frequencies as high as 100 Hz, with typical peak frequencies at 25 Hz. Our statistical description of geological heterogeneities below 10 m can thus be used consistently in our low-frequency scattering theory to estimate the scattering loss caused by small-scale heterogeneities.

Shear waves have shorter wavelengths than compressional waves and can be more attenuated because they are more sensitive to heterogeneities. We showed the existence of a high-frequency cutoff for the depth of penetration of waves, whose position in frequency depends on the maximum size of fractal heterogeneities. The dominant frequency of a wavelet decreases faster for higher fractal exponents and for larger characteristic sizes of heterogeneities. This loss of high-frequency content influences resolution in seismic imaging. Our study recommends using low-frequency P-waves for deep targets under a strongly heterogeneous overburden.

Agreement of our results with the Backus limit and the Rayleigh diffusion regime is due to the use of velocity fluctuations. Nevertheless, proper connection with the O'Doherty-Anstey formula requires use of the logarithm of impedance. Moreover, more complex, multiple scattering occurs when sizes of heterogeneities are similar to that of the seismic wavelength. Large-scale local cycles, present in the sonic-log data, call for incorporation of resonant scattering effects into high-frequency scattering theories.

ACKNOWLEDGMENTS

We thank Apache Canada Ltd. for providing the data used in this study and Mark Tomasso from the Bureau of Economic Geology for the velocity model. We acknowledge Apache Corporation, GX Technology Corporation, and the Jackson School of Geosciences for financial support. This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A

O'DOHERTY-ANSTEY FORMULA AND THE MEAN FIELD THEORY

O'Doherty and Anstey (33; 35) proposed that local transmission coefficient $T(\omega)$ for traveltime Δt in sedimentary layers should be

$$T(\omega) = \exp\left[-\frac{\omega\,\Delta t}{2\,Q(\omega)}\right] = e^{-R(\omega)\,\Delta t}, \tag{A-1}$$

where $R(\omega)$ is the spectrum of reflection coefficients, which is related to the spectrum of impedance fluctuations (4). Because we have used velocity fluctuations, attenuation

in the dispersion relation of equation 21 depends on the velocity spectrum :

$$\frac{1}{Q(\omega)} = 2\,\frac{Im[k]}{Re[k]}, \tag{A-2}$$

$$= k_0\left[E^{(1)}(0) - E^{(1)}(2k_0)\right], \tag{A-3}$$

$$T(\omega) = \exp\left\{-\frac{k_0\,\omega}{2}\,\left[E^{(1)}(0) - E^{(1)}(2k_0)\right]\Delta t\right\}. \tag{A-4}$$

These formulae have been previously derived and analyzed (24; 58; 38). The term $E^{(1)}(2k_0)$ can be interpreted as backward scattering with exchange wavenumber $2k_0$, whereas the term $E^{(1)}(0)$ is forward scattering with exchange wavenumber 0. Both energy terms reduce to 1D expressions because of isotropic integration, whereas one symmetry axis is imposed by propagation direction of the wave. Further interpretation can be found using a dynamic effective model for multiple scattering (55) : the scattered waves interfere with the main wavefield, and their relative phase continuously changes in all directions, except for significant interferences in forward and backward directions. Previous 1D derivations (38) have incorporated a traveltime correction, corresponding to neglecting forward scattering in order to reproduce the O'Doherty-Anstey formula. This approach extends validity of the analytical expressions to higher frequencies, but no simple traveltime phase correction exists for the mean field theory in 3D. Constant $E^{(1)}(0)$ ensures recovery of the Backus effective medium and the Rayleigh diffusion regime for very low frequencies.

APPENDIX B

EXPONENTIALLY CORRELATED HETEROGENEITIES

Results have been derived several times (20; 38) using exponential correlation function $N(r) = \exp[-r/b]$ and are added here as a specific case with simple analytical expressions within the general framework of von Kármán's media. Values of the integral in equation 22 are

$$S(0) = b\,, \tag{B-1}$$

$$S(2k_0) = \frac{b}{1 - 2\,i\,k_0 b}\,. \tag{B-2}$$

Using these expressions in the dispersion relation of equation 21, the new dispersion relation, phase velocity $c(\omega)$, and attenuation are

$$\frac{k}{k_0} = \frac{c_0}{c(\omega)} + \frac{i}{2\,Q(\omega)} = 1 + \frac{\sigma^2}{2}\left[1 + \frac{(2\,k_0 b)^2}{1 - 2\,i\,k_0 b}\right] + O(\sigma^4)\,, \tag{B-3}$$

$$\frac{1}{c(\omega)} = \frac{1}{c_0}\left[1 + \sigma^2\,\frac{1/2 + (2\,k_0 b)^2}{1 + (2\,k_0 b)^2}\right], \tag{B-4}$$

$$\frac{1}{Q(\omega)} = \sigma^2\,\frac{8\,(k_0 b)^3}{1 + (2\,k_0 b)^2}\,. \tag{B-5}$$

In the limit of very long wavelengths, i.e. $k_0 \to 0$, attenuation and velocity reduce respectively to the Rayleigh diffusion regime and the effective medium theory of (3) :

$$\frac{1}{c(\omega)} \quad \to \quad \frac{1}{c_0}\left(1+\frac{\sigma^2}{2}\right) , \tag{B-6}$$

$$\frac{1}{Q(\omega)} \quad \sim \quad 8\,\sigma^2\,(k_0 b)^3 \to 0\,. \tag{B-7}$$

REFERENCES

[1] Anstey, N. A., and R. F. O'Doherty, 2002a, Cycles, layers, and reflections: part 1: The Leading Edge, **21**, 44–51.

[2] ———, 2002b, Cycles, layers, and reflections: part 2: The Leading Edge, **21**, 152–158.

[3] Backus, G. E., 1962, Long-wave elastic anisotropy produced by horizontal layering: Journal of Geophysical Research, **67**, 4427–4440.

[4] Banik, N. C., I. Lerche, and R. T. Shuey, 1985, Stratigraphic filtering, part I: derivation of the O'Doherty-Anstey formula: Geophysics, **50**, 2768–2774.

[5] Bleistein, N., J. K. Cohen, and J. W. Stockwell, 2001, Mathematics of multidimensional seismic imaging, migration, and inversion: Springer. Interdisciplinary applied mathematics.

[6] Born, M., and E. Wolf, 1964, Principles of optics: electromagnetic, theory of propagation, interference and diffraction of light: Pergamon Press.

[7] Chernov, L. A., 1960, Wave propagation in a random medium: McGraw-Hill.

[8] Chesnokov, E. M., Y. A. Kukharenko, and P. Y. Kukharenko, 1998, Frequency dependence of physical parameters of microinhomogeneous media, space statistics: Revue de l'Institut Français du Pétrole, **53**, 729–734.

[9] Dolan, S. S., C. J. Bean, and B. Riollet, 1998, The broad-band fractal nature of heterogeneity in the upper crust from petrophysical logs: Geophysical Journal International, **132**, 489–507.

[10] Frankel, A., and R. W. Clayton, 1986, Finite difference simulations of seismic scattering: implications for the propagation of short period seismic waves in the crust and models of crustal heterogeneity: Journal of Geophysical Research, **91**, 6465–6489.

[11] Gist, G. A., 1994, Seismic attenuation from 3-D heterogeneities: a possible resolution of the VSP attenuation paradox: 64th Annual International Meeting, SEG, Expanded Abstracts, Society of Exploration Geophysicists, 1042–1045.

[12] Goff, J. A., and K. Holliger, 2003, Heterogeneity in the crust and upper mantle: nature, scaling, and seismic properties: Kluwer Academic/Plenum Publishers.

[13] Goff, J. A., and T. H. Jordan, 1988, Stochastic modeling of seafloor morphology: inversion of sea beam data for second-order statistics: Journal of Geophysical Research, **93**, 13,589–13,608.

[14] Golub, G. H., and V. Pereira, 1973, The differentiation of pseudo-inverse and nonlinear least squares problems whose variables separate: SIAM Journal of Numerical Analysis, **10**, 413–432.

[15] Haney, M. M., K. van Wijk, and R. Snieder, 2005, Radiative transfer in layered media and its connection to the O'Doherty-Anstey formula: Geophysics, **70**, no. 1, T1–T11.
[16] Hewett, T. A., 1986, Fractal distributions of reservoir heterogeneity and their influence on fluid transport: Presented at the 61st Technical Conference, Society of Petroleum Engineering, Society of Petroleum Engineering, SPE 15386.
[17] Holliger, K., and A. R. Levander, 1992, A stochastic view of lower crustal fabric based on evidence from the Ivrea zone: Geophysical Research Letters, **19**, 1153–1156.
[18] Hurst, H. E., 1951, Long-term storage capacity of reservoirs: Transaction of American Society of Civil Engineering, **116**, 770–808.
[19] Kanaun, S. K., and V. Levin, 2008, Self-consistent methods for composites, vol.2: wave propagation in heterogeneous materials: Springer-Verlag.
[20] Karal, F. C., and J. B. Keller, 1964, Elastic, electromagnetic, and other waves in a random medium: Journal of Mathematical Physics, **5**, 537–547.
[21] Klimeš, L., 2002, Correlation functions of random media: Pure and Applied Geophysics, **159**, 1811 –1831.
[22] Lancaster, S., and D. Whitcombe, 2000, Fast-track 'coloured' inversion: 60th Annual International Meeting, SEG, Expanded Abstracts, Society of Exploration Geophysicists, 1572–1575.
[23] Le Gonidec, Y., and D. Gibert, 2007, Multiscale analysis of waves reflected by granular media: acoustic experiments on glass beads and effective medium theories: Journal of Geophysical Research, **112**.
[24] Lerche, I., 1986, Stratigraphic filter theory: combined effects of parallel bedding and random inhomogeneities: Geophysics, **51**, 1531–1537.
[25] Li, C. F., 2003, Rescaled-range and power spectrum analyses on well-logging data: Geophysical Journal International, **153**, 201–212.
[26] Lord, R. D., 1954, The use of the Hankel transform in statistics I. General theory and examples: Biometrika, **41**, 44–55.
[27] Lu, S., F. J. Molz, G. E. Fogg, and J. W. Castle, 2002, Combining stochastic facies and fractal models for representing natural heterogeneity: Hydrogeology Journal, **10**, 475–482.
[28] Mandelbrot, B. B., 1983, The fractal geometry of nature: W. H. Freeman and Co.
[29] Mandelbrot, B. B., and J. R. Wallis, 1969, Some long-run properties of geophysical records: Water Resources Research, **5**, 321–340.
[30] Marsan, D., and C. J. Bean, 2003, Multifractal modeling and analyses of crustal heterogeneity, *in* Heterogeneity in the crust and upper mantle: nature, scaling, and seismic properties: Kluwer Academic/Plenum Publishers, 207–236.
[31] Morlet, J., G. Arens, E. Fourgeau, and D. Giard, 1982, Wave propagation and sampling theory - part I: complex signal and scattering in multilayered media: Geophysics, **47**, 203–221.
[32] Müller, T. M., and S. A. Shapiro, 2001, Most probable seismic pulses in single realizations of two- and three-dimensional random media: Geophysical Journal International, **144**, 83–95.

[33] O'Doherty, R. F., and N. A. Anstey, 1971, Reflections on amplitudes: Geophysical Prospecting, **19**, 430–458.
[34] Painter, S., and L. Paterson, 1994, Fractional Lévy motion as a model for spatial variability in sedimentary rocks: Geophysical Research Letters, **21**, 2857–2860.
[35] Resnick, J. R., 1990, Stratigraphic filtering: Pure and Applied Geophysics, **132**, 49–65.
[36] Rivera, N. A., S. Ray, J. L. Jensen, A. K. Chan, and W. B. Ayers, 2004, Detection of cyclic patterns using the wavelets: an example study in the Ormskirk sandstone, Irish sea: Mathematical Geology, **36**, 529–543.
[37] Rytov, S. M., Y. A. Kravtsov, and V. I. Tatarski, 1989, Wave propagation through random media: Springer-Verlag. Principles of statistical radio physics.
[38] Sato, H., and M. Fehler, 1998, Seismic wave propagation and scattering in the heterogeneous Earth: Springer-Verlag.
[39] Schottky, W., 1926, Small-shot effect and Flicker effect: Physical Reviews, **28**, 74–104.
[40] Shapiro, S. A., and P. Hubral, 1999, Elastic waves in random media: Springer-Verlag.
[41] Shapiro, S. A., and H. Zien, 1993, The O'Doherty-Anstey formula and localization of seismic waves: Geophysics, **58**, 736–740.
[42] Sheng, P., B. White, Z. Q. Zhang, and G. Papanicolaou, 1986, Multiple scattering noise in one dimension: universality through localization length scaling: Physical Review Letters, **57**, 1000–1003.
[43] Shiomi, K., H. Sato, and M. Ohtake, 1997, Broad-band power-law spectra of well log data in Japan: Geophysical Journal International, **130**, 57–64.
[44] Shtatland, E. S., 1991, Fractal stochastic models for acoustic impedance: an explanation of scaling or 1/f geology and stochastic inversion: 51st Annual International Meeting, SEG, Expanded Abstracts, Society of Exploration Geophysicists, 1598–1601.
[45] Stefani, J., and S. D. Gopa, 2001, On the power-law behavior of subsurface heterogeneity: 61st Annual International Meeting, SEG, Expanded Abstracts, Society of Exploration Geophysicists, 2033–2036.
[46] Stovas, A., and B. Ursin, 2007, Equivalent time-average and effective medium for periodic layers: Geophysical Prospecting, **55**, 871–882.
[47] Tatarski, V. I., 1961, Wave propagation in a turbulent medium: McGraw-Hill.
[48] Todoeschuck, J. P., O. G. Jensen, and S. Labonte, 1990, Gaussian scaling noise model of seismic reflection sequences: evidence from well logs: Geophysics, **55**, 480–484.
[49] Turcotte, D. L., 1997, Fractals and chaos in geology and geophysics: Cambridge University Press.
[50] Uscinski, N. J., 1977, The elements of wave propagation in random media: McGraw-Hill.
[51] van der Baan, M., 2001, Acoustic wave propagation in one-dimensional random media: the wave localization approach: Geophysical Journal International, **145**, 631–646.
[52] van der Baan, M., J. Wookey, and D. Smit, 2007, Stratigraphic filtering and

source penetration depth: Geophysical Prospecting, **55**, 679–684.
[53] von Kármán, T., 1948, Progress in the statistical theory of turbulence: Journal of Marine Research, **7**, 252–264.
[54] Walden, A. T., and J. W. J. Hosken, 1985, An investigation of the spectral properties of primary reflection coefficients: Geophysical Prospecting, **33**, 400–435.
[55] Waterman, P. C., and R. Truell, 1961, Multiple scattering of waves: Journal of Mathematical Physics, **2**, 512–537.
[56] White, B., P. Sheng, and B. Nair, 1990, Localization and backscattering spectrum of seismic waves in stratified lithology: Geophysics, **55**, 1158–1165.
[57] Wu, R. S., 1982, Attenuation of short period seismic waves due to scattering: Geophysical Research Letters, **9**, 9–12.
[58] ———, 1988, Discussion on:"Stratigraphic filter theory: combined effects of parallel bedding and random inhomogeneities," I. Lerche, author: Geophysics, **53**, 995–997.
[59] ———, 1993, Separation of scattering and absorption in 1-D random media from spatio-temporal distribution of seismic energy: 53rd Annual International Meeting, SEG, Expanded Abstracts, Society of Exploration Geophysicists, 1014–1017.
[60] Wu, R. S., and K. Aki, 1985, Elastic wave scattering by a random medium and the small scale inhomogeneities in the lithosphere: Journal of Geophysical Research, **90**, 10261–10273.
[61] Ziolkowski, A., P. Hanssen, R. Gatliff, H. Jakubowicz, A. Dobson, G. Hampson, X. Y. Li, and E. Liu, 2003, Use of low frequencies for sub-basalt imaging: Geophysical Prospecting, **51**, 169–182.

Adaptive multiple subtraction using regularized nonstationary regression

*Sergey Fomel**

ABSTRACT

Stationary regression is the backbone of different seismic data processing algorithms including match filtering, which is commonly applied for adaptive multiple subtraction. However, the assumption of stationarity is not always adequate for describing seismic signals. I present a general method of nonstationary regression and show its application to nonstationary match filtering. The key idea is the use of shaping regularization for constraining the variability of nonstationary regression coefficients.

As shown by simple computational experiments, shaping regularization has clear advantages over conventional Tikhonov's regularization, incuding a more intuitive selection of parameters and a faster iterative convergence.

Using benchmark synthetic data examples, I demonstrate successful applications of this method to the problem of adaptive subtraction of multiple reflections.

INTRODUCTION

Many natural phenomena, including geological events and geophysical data, are fundamentally nonstationary. They may exhibit stationarity on the short scale but eventually change their behavior in space and time. Nonstationary adaptive filtering is a well-developed field in signal processing (13). In seismic signal processing, nonstationary filters were analyzed by (15) and applied to many important problems, including multiple suppression (17), data interpolation (3; 4), migration deconvolution (10; 21).

In this paper, I present a general approach to designing nonstationary operators, including the case of nonstationary matching and prediction-error filters. The key idea is the application of shaping regularization (8) for constraining the continuity and smoothness of the filter coefficients. Regularization makes the estimation problem well-posed and leads to fast numerical algorithms. Advantages of shaping regularization in comparison with the classic Tikhonov regularization include an easier control of the regularization parameters and a faster iterative convergence resulting from the better conditioning of the inverted matrix.

Adaptive subtraction is a method for matching and removing coherent noise, such as multiple reflections, after their prediction by a data-driven technique (24). Adap-

*e-mail: sergey.fomel@beg.utexas.edu

tive subtraction involves a matching filter to compensate for the amplitude, phase, and frequency distortions in the predicted noise model. Different techniques for matching filtering and adaptive subtraction have been developed and discussed by a number of authors (24; 16; 18; 22; 25; 12; 14; 1; 5). The regularized non-stationary regression technique, proposed in this paper, allows the matching filter to become smoothly nonstationary without the need to break the input data into local windows.

The paper is organized as follows. I start with an overview of stationary and nonstationary regression theory and introduce a method of regularized nonstationary regression. Next, I demonstrate this method using toy examples of line fitting and nonstationary deconvolution. Finally, I apply it to the adaptive multiple suppression problem and test its performance using a number of benchmark synthetic data examples.

STATIONARY AND NONSTATIONARY REGRESSION

Consider a "master" signal $m(\mathbf{x})$, where $\mathbf{x}$ represents the coordinates of a multidimensional space, and a collection of "slave" signals $s_k(\mathbf{x})$, $k = 1, 2, \ldots, N$. The goal of stationary regression is to estimate coefficients a_k, $k = 1, 2, \ldots, N$ such that the prediction error

$$e(\mathbf{x}) = m(\mathbf{x}) - \sum_{k=1}^{N} a_k \, s_k(\mathbf{x}) \tag{1}$$

is minimized in the least-squares sense. Particular examples include:

Line fitting

Let $\mathbf{x}$ be one-dimensional (denoted by x), $N = 2$, $s_1(x) = 1$, and $s_2(x) = x$. The problem of minimizing $e(x)$ amounts to fitting a straight line $a_1 + a_2\, x$ to the master signal.

Match filtering

If the slave signals are translates of the same signal $s(\mathbf{x})$, the regression problem corresponds to the problem of match filtering between $s(\mathbf{x})$ and $m(\mathbf{x})$. In the 1-D case, one can take, for example,

$$s_k(x) = s(x - k + N/2) \ ,$$

which turns the sum in equation 1 into a convolution with an unknown matching filter.

Prediction-error filtering

If the slave signals are causal translates of the master signal, the regression problem corresponds to the problem of autoregressive prediction-error filtering. In the 1-D case, one can take, for example,

$$s_k(x) = m(x-k) \ .$$

Nonstationary regression

Non-stationary regression uses a definition similar to equation 1 but allows the coefficients a_k to change with $\mathbf{x}$. The error turns into

$$e(\mathbf{x}) = m(\mathbf{x}) - \sum_{k=1}^{N} b_k(\mathbf{x})\, s_k(\mathbf{x}) \ , \tag{2}$$

and the problem of its minimization becomes ill-posed, because one can get more unknown variables than constraints. The remedy is to include additional constraints that limit the allowed variability of the estimated coefficients.

The classic regularization method is Tikhonov's regularization (20; 6), which amounts to minimization of the following functional:

$$F[\mathbf{b}] = \|e(\mathbf{x})\|^2 + \epsilon^2 \sum_{k=1}^{N} \|\mathbf{D}\,[b_k(\mathbf{x})]\,\|^2 \ , \tag{3}$$

where $\mathbf{D}$ is the regularization operator (such as the gradient or Laplacian filter) and ϵ is a scalar regularization parameter. If $\mathbf{D}$ is a linear operator, least-squares estimation reduces to linear inversion

$$\mathbf{b} = \mathbf{A}^{-1}\,\mathbf{d} \ , \tag{4}$$

where $\mathbf{b} = \left[\ b_1(\mathbf{x}) \quad b_2(\mathbf{x}) \quad \cdots \quad b_N(\mathbf{x}) \ \right]^T$,
$\mathbf{d} = \left[\ s_1(\mathbf{x})\, m(\mathbf{x}) \quad s_2(\mathbf{x})\, m(\mathbf{x}) \quad \cdots \quad s_N(\mathbf{x})\, m(\mathbf{x}) \ \right]^T$, and the elements of matrix $\mathbf{A}$ are

$$A_{ij}(\mathbf{x}) = s_i(\mathbf{x})\, s_j(\mathbf{x}) + \epsilon^2\, \delta_{ij}\, \mathbf{D}^T\, \mathbf{D} \ . \tag{5}$$

Shaping regularization (8) formulates the problem differently. Instead of specifying a penalty (roughening) operator $\mathbf{D}$, one specifies a shaping (smoothing) operator $\mathbf{S}$. The regularized inversion takes the form

$$\mathbf{b} = \widehat{\mathbf{A}}^{-1}\,\widehat{\mathbf{d}} \ , \tag{6}$$

where

$$\widehat{\mathbf{d}} = \left[\ \mathbf{S}\,[s_1(\mathbf{x})\, m(\mathbf{x})] \quad \mathbf{S}\,[s_2(\mathbf{x})\, m(\mathbf{x})] \quad \cdots \quad \mathbf{S}\,[s_N(\mathbf{x})\, m(\mathbf{x})] \ \right]^T \ ,$$

the elements of matrix $\widehat{\mathbf{A}}$ are

$$\widehat{A}_{ij}(\mathbf{x}) = \lambda^2 \, \delta_{ij} + \mathbf{S} \left[s_i(\mathbf{x}) \, s_j(\mathbf{x}) - \lambda^2 \, \delta_{ij} \right] \ , \tag{7}$$

and λ is a scaling coefficient.The main advantage of this approach is the relative ease of controlling the selection of λ and $\mathbf{S}$ in comparison with ϵ and $\mathbf{D}$. In all examples of this paper, I define $\mathbf{S}$ as Gaussian smoothing with an adjustable radius and choose λ to be the median value of $s_i(\mathbf{x})$. As demonstrated in the next section, matrix $\widehat{\mathbf{A}}$ is typically much better conditioned than matrix $\mathbf{A}$, which leads to fast inversion with iterative algorithms.

In the case of $N = 1$ (regularized division of two signals), a similar construction was applied before to define local seismic attributes (7).

TOY EXAMPLES

In this section, I illustrate the general method of regularized nonstationary regression using simple examples of nonstationary line fitting and nonstationary deconvolution.

Nonstationary line fitting

Figure 1a shows a classic example of linear regression applied as a line fitting problem. When the same technique is applied to data with a non-stationary behavior (Figure 1b), stationary regression fails to produce an accurate fit and creates regions of consistent overprediction and underprediction.

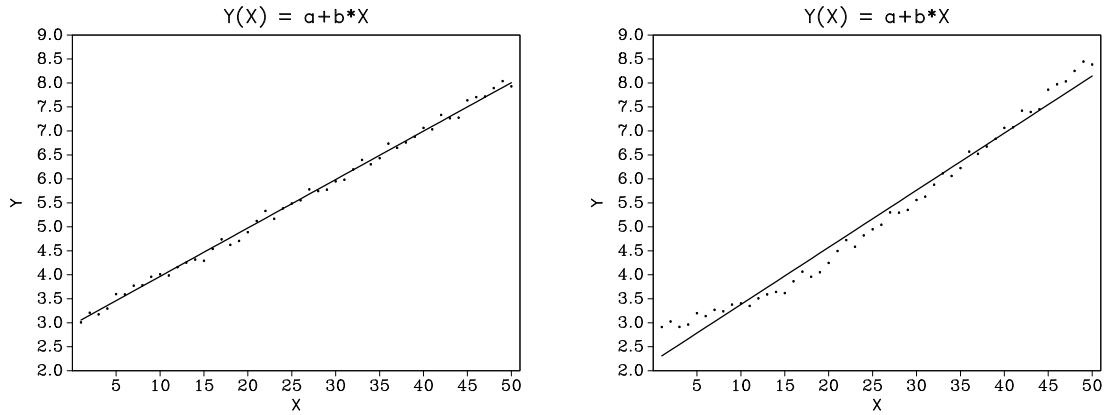


Figure 1: Line fitting with stationary regression works well for stationary data (a) but poorly for non-stationary data (b). lpf/regr pred,pred2

One remedy is to extend the model by including nonlinear terms (Figure 2a), another is to break the data into local windows (Figure 2b). Both solutions work

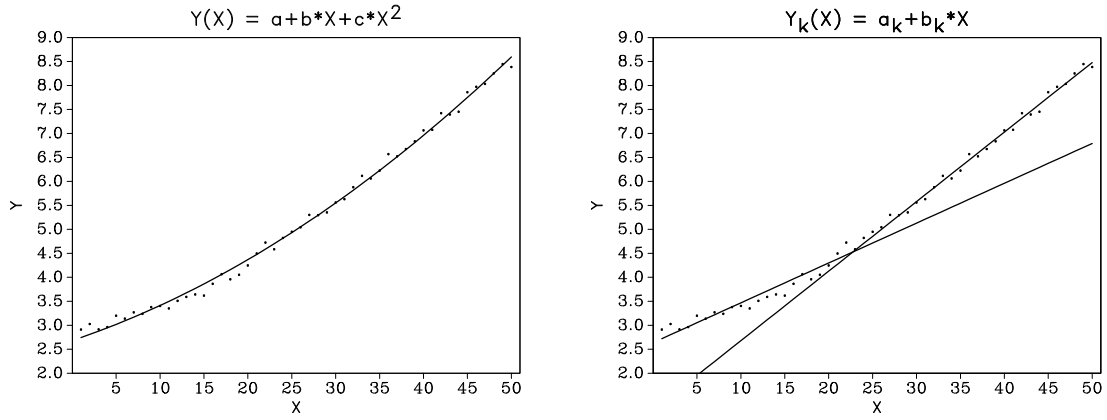


Figure 2: Nonstationary line fitting using nonlinear terms (a) and local windows (b). lpf/regr pred3,pred4

to a certain extent but are not completely satisfactory, because they decreases the estimation stability and introduce additional non-intuitive parameters.

The regularized nonstationary solution, defined in the previous section, is shown in Figure 3. When using shaping regularization with smoothing as the shaping operator, the only additional parameter is the radius of the smoothing operator.

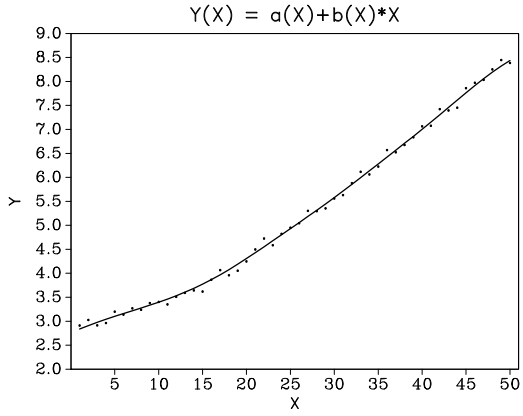


Figure 3: Nonstationary line fitting by regularized nonstationary regression. lpf/regr pred5

This toy example makes it easy to compare shaping regularization with the more traditional Tikhonov's regularization. Figures 4 and 5 show inverted matrix $\mathbf{A}$ from equation 5 and the distribution of its eigenvalues for two different values of Tikhonov's regularization parameter ϵ, which correspond to mild and strong smoothing constraints. The operator $\mathbf{D}$ in this case is the first-order difference. Correspondingly, Figures 6 and 7 show matrix $\widehat{\mathbf{A}}$ from equation 7 and the distribution of its eigenvalues for mild and moderate smoothing implemented with shaping. The operator $\mathbf{S}$ is Gaussian smoothing controlled by the smoothing radius.

When a matrix operator is inverted by an iterative method such as conjugate gradients, two characteristics control the number of iterations and therefore the cost of inversion (9; 23):

1. the condition number κ (the ratio between the largest and the smallest eigenvalue)
2. the clustering of eigenvalues.

Large condition numbers and poor clustering lead to slow convergence. Figures 4–7 demonstrate that both the condition number and the clustering of eigenvalues are significantly better in the case of shaping regularization than in the case of Tikhonov's regularization. In toy problems, this difference in behavior is not critical, because one can easily invert both matrices exactly. However, this difference becomes important in large-scale applications, where inversion is iterative and saving the number of iterations is crucial for performance.

As the smoothing radius increases, matrix $\widehat{\mathbf{A}}$ approaches the identity matrix, and the result of non-stationary regression regularized by shaping approaches the result of stationary regression. This intuitively pleasing behavior is difficult to emulate with Tikhonov's regularization.

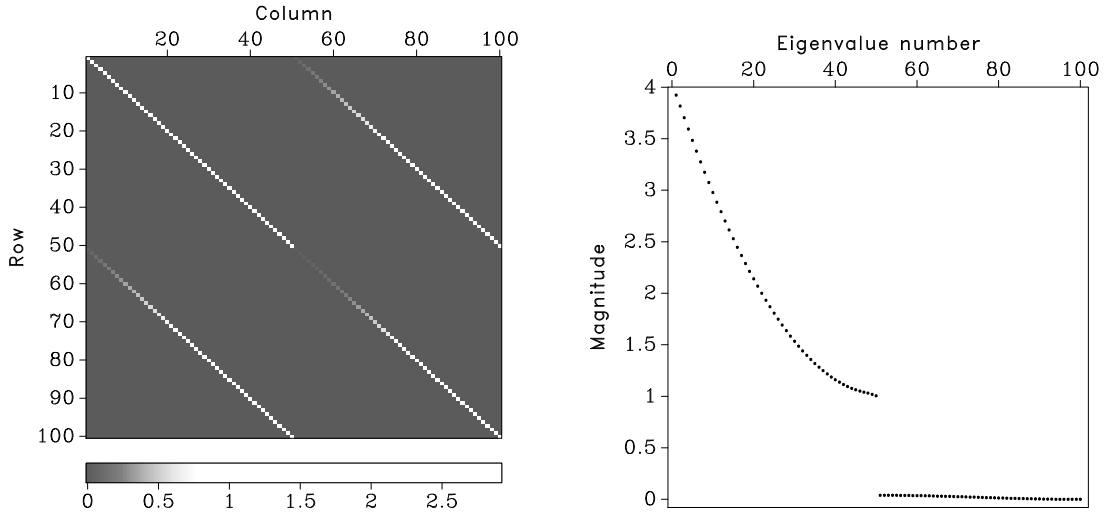


Figure 4: Matrix inverted in Tikhonov's regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The regularization parameter $\epsilon = 0.1$ corresponds to mild smoothing. The condition number is $\kappa \approx 888888$. lpf/regr tmat0,teig0

Nonstationary deconvolution

Figure 8 shows an application of regularized nonstationary regression to a benchmark deconvolution test from (2). The input signal is a synthetic trace that contains events

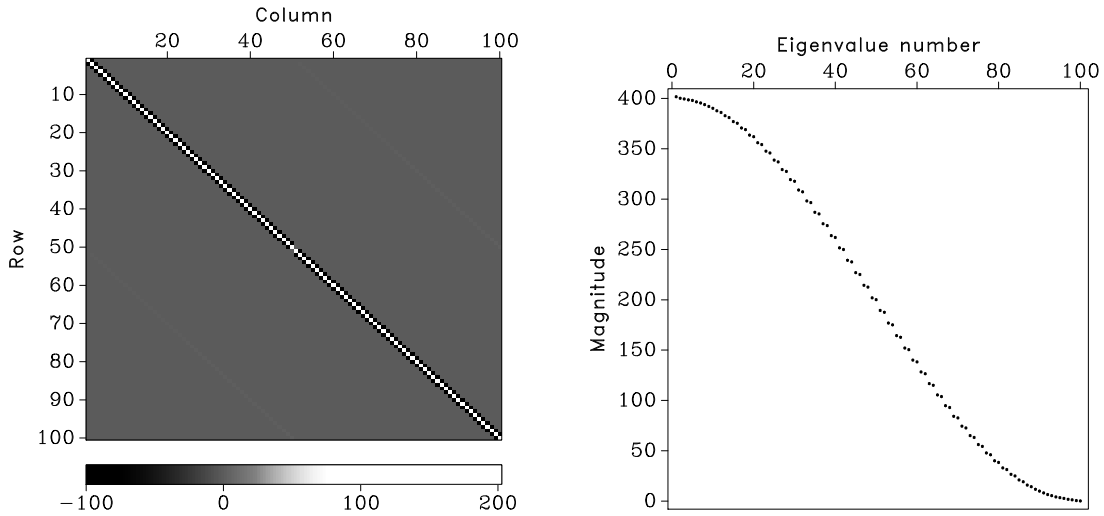


Figure 5: Matrix inverted in Tikhonov's regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The regularization parameter $\epsilon = 10$ corresponds to strong smoothing. The condition number is $\kappa \approx 14073$. Eigenvalues are poorly clustered. lpf/regr tmat2,teig2

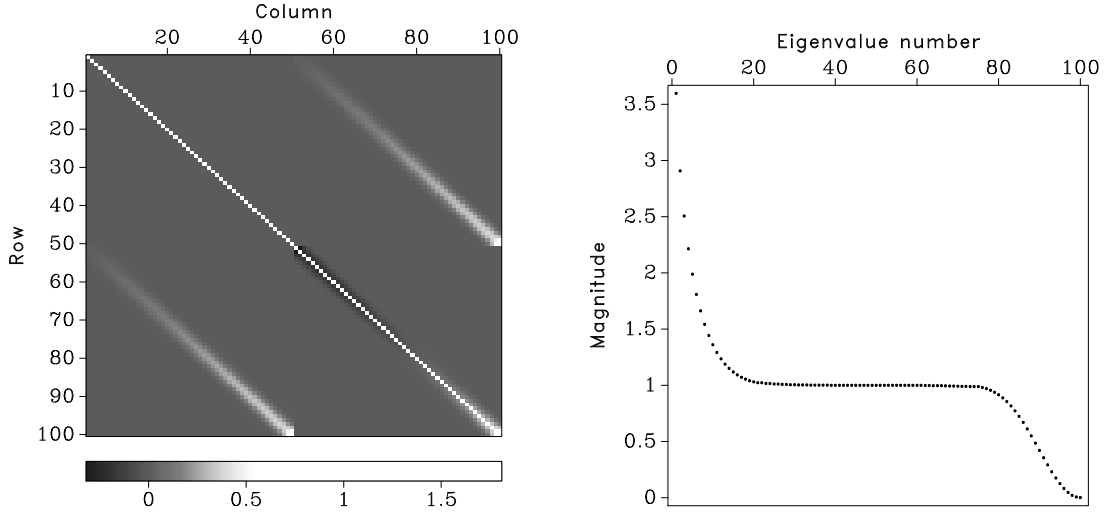


Figure 6: Matrix inverted in shaping regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The smoothing radius is 3 samples (mild smoothing). The condition number is $\kappa \approx 6055$. lpf/regr smat0,seig0

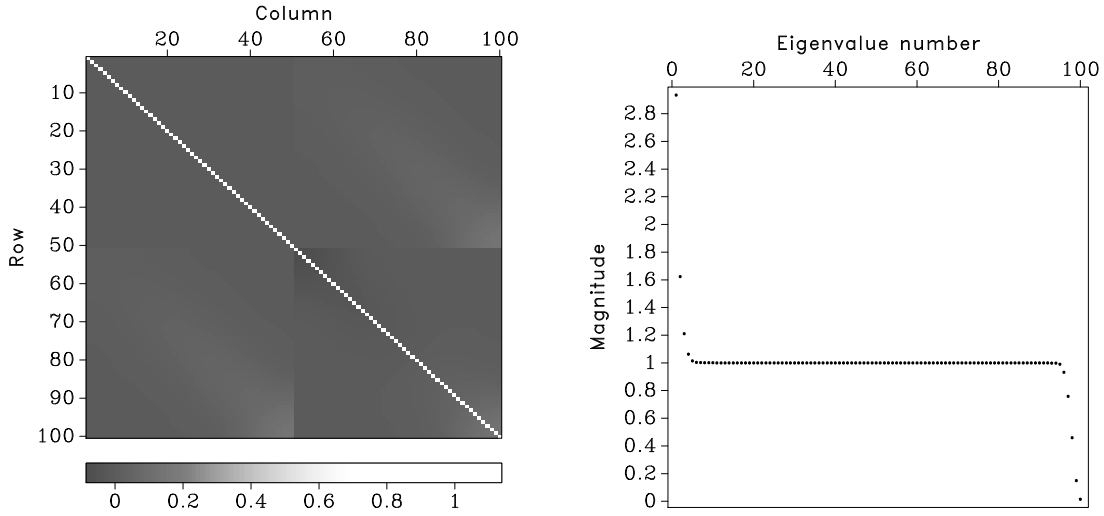


Figure 7: Matrix inverted in shaping regularization applied to nonstationary line fitting (a) and the distribution of its eigenvalues (b). The smoothing radius is 15 samples (moderate smoothing). The condition number is $\kappa \approx 206$. Eigenvalues are well clustered. lpf/regr smat1,seig1

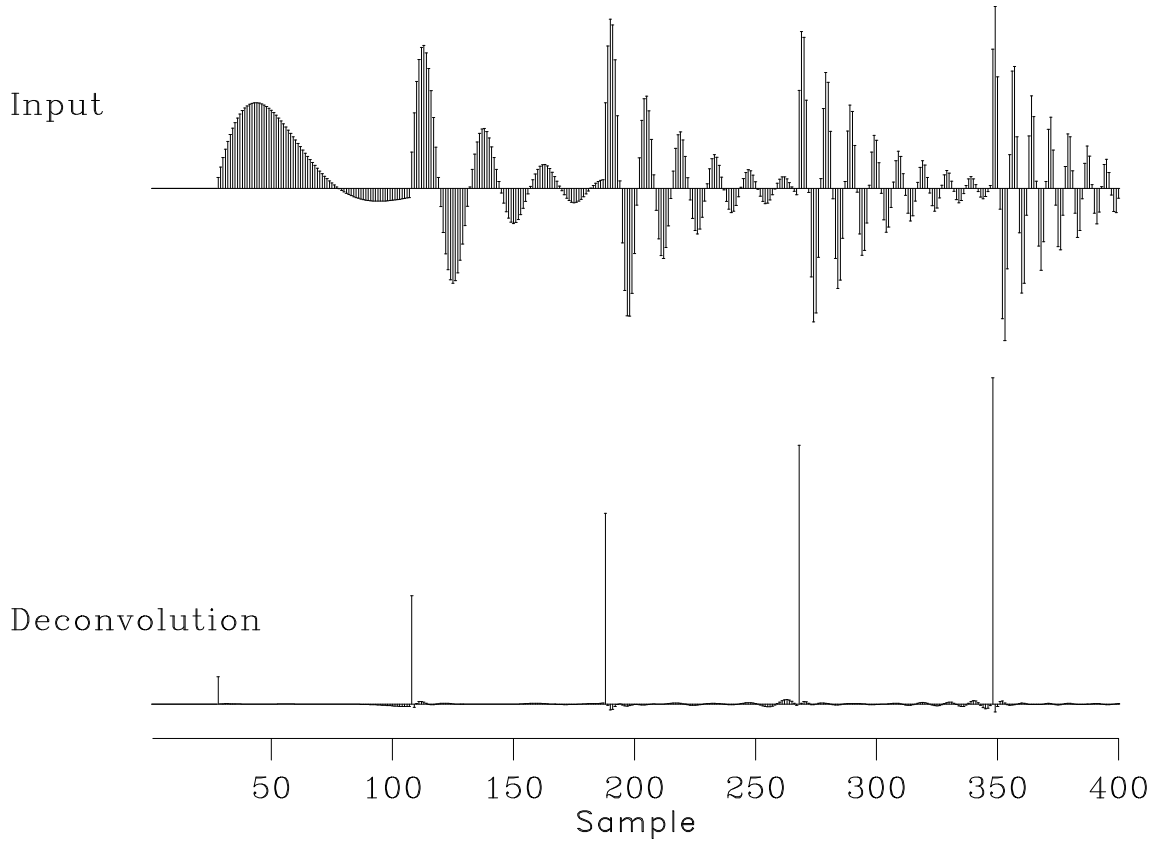


Figure 8: Benchmark test of nonstationary deconvolution from (2). Top: input signal, bottom: deconvolved signal using nonstationary regularized regression. lpf/lpf lpf

with variable frequencies. A prediction-error filter is estimated by setting $N = 2$, $s_1(x) = m(x-1)$ and $s_2(x) = m(x-2)$. I use triangle smoothing with a 5-sample radius as the shaping operator $\mathbf{S}$. The deconvolved signal (bottom plot in Figure 8) shows the nonstationary reverberations correctly deconvolved.

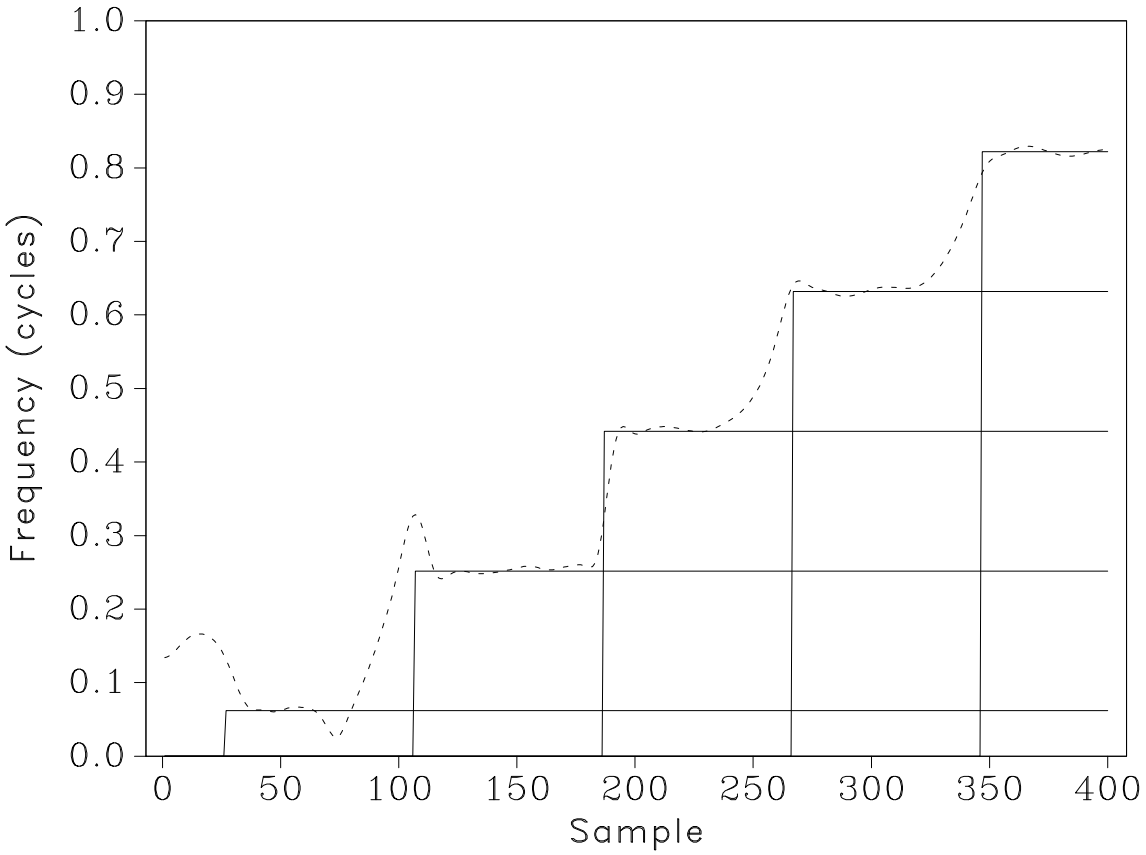


Figure 9: Frequency of different components in the input synthetic signal from Figure 8 (solid line) and the local frequency estimate from non-stationary deconvolution (dashed line). lpf/lpf freq

A three-point prediction-error filter $\{1, a_1, a_2\}$ can predict an attenuating sinusoidal signal

$$m(x) = \rho^x \, cos(\omega \, x) \; , \tag{8}$$

where ω is frequency, and ρ is the exponent factor, provided that the filter coefficients are defined as

$$a_1 = -2 \, \rho \, \cos(\omega) \; ; \tag{9}$$

$$a_2 = \rho^2 \; . \tag{10}$$

This fact follows from the simple identity

$$\begin{aligned} m(x) + a_1 \, m(x-1) + a_2 \, m(x-2) &= \\ \rho^x \, [\cos(\omega \, x) - 2 \, \cos(\omega) \, \cos(\omega \, x - \omega) + \cos(\omega \, x - 2 \, \omega)] &= 0 \; . \end{aligned} \tag{11}$$

According to equations 9–10, one can get an estimate of the local frequency $\omega(x)$ from the non-stationary coefficients $b_1(x) = -a_1(x)$ and $b_2(x) = -a_2(x)$ as follows:

$$\omega(x) = \arccos\left(\frac{b_1(x)}{2\sqrt{-b_2(x)}}\right) . \tag{12}$$

Figure 9 shows frequency of the different components in the input non-stationary signal from Figure 8 and the local frequency estimate using equation 12. A reasonably accurate match between the true nonstationary frequencies and their estimate from nonstationary regression can be observed. The local frequency attribute (7) provides a different approach to the same problem.

ADAPTIVE MULTIPLE SUBTRACTION

In this section, I apply regularized nonstationary regression to the problem of adaptive subtraction and evaluate its performance using benchmark synthetic tests.

One possible variation of non-stationary regression applied to adaptive subtraction is an introduction of the signal prediction filter (18). The signal prediction applies to the residual error $e(\mathbf{x})$ in equation 2, thus modifying the objective function (11).

Abma test

(1) present a comparison of different adaptive subtraction algorithms used for multiple suppression. I use Abma's benchmark examples to illustrate an application of nonstationary regression to the adaptive subtraction problem. Figure 10 shows the first test: the input data (Figure 10a) contain a horizontal "signal" event and dipping "noise" events. We are provided with a model of the noise (Figure 10b). However, the model events are slightly shifted and have a different amplitude behavior. This test simulates a typical situation in adaptive surface-related multiple elimination, where there are phase and amplitude distortions in the multiple model, caused by an imperfect prediction. To handle the variability in noise amplitudes, I design a non-stationary matching filter with 13 coefficients and 3-sample smoothing radius and apply it to match the noise model to the data. Subtracting matched noise (Figure 10d) produces the desired and nearly perfect signal estimate (Figure 10c). The variability of filter coefficients is illustrated in Figure 11, which displays the zero-lag coefficient and the mean coefficient of the non-stationary matching filter.

Another benchmark from (1) is shown in Figure 12. This time, the noise part of the data is a curved event which has a predictive model (Figure 12b) with some amplitude and phase differences. Non-stationary regularized regression correctly predicts the noise signal (Figure 12d) using match filtering and produces an accurate signal estimate (Figure 12c). The variability of non-stationary filter coefficients is shown in Figure 13.

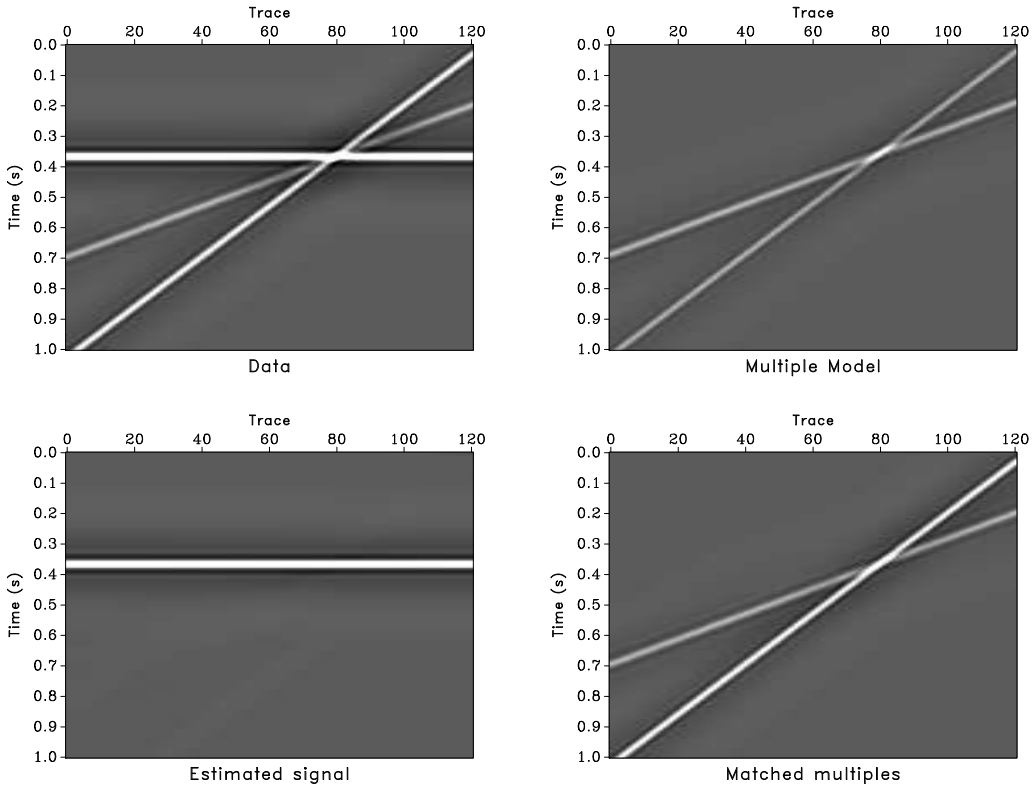


Figure 10: Benchmark test on adaptive subtraction from (1). a – Input synthetic data. b – Model of the noise containing amplitude and phase differences with respect to the corresponding part of the data. c – Extracted signal. d – Estimated noise. lpf/ray planes

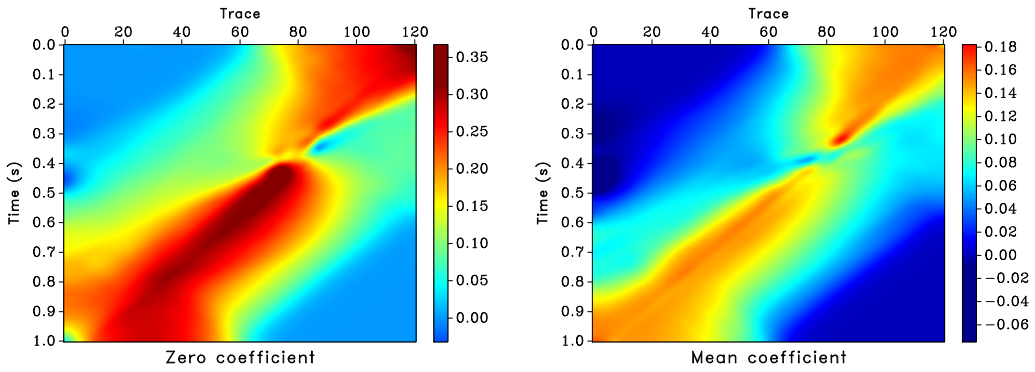


Figure 11: Variability of non-stationary match filter coefficients for the example shown in Figure 10. a – Zero-lag coefficient. b – Mean coefficient. lpf/ray filt

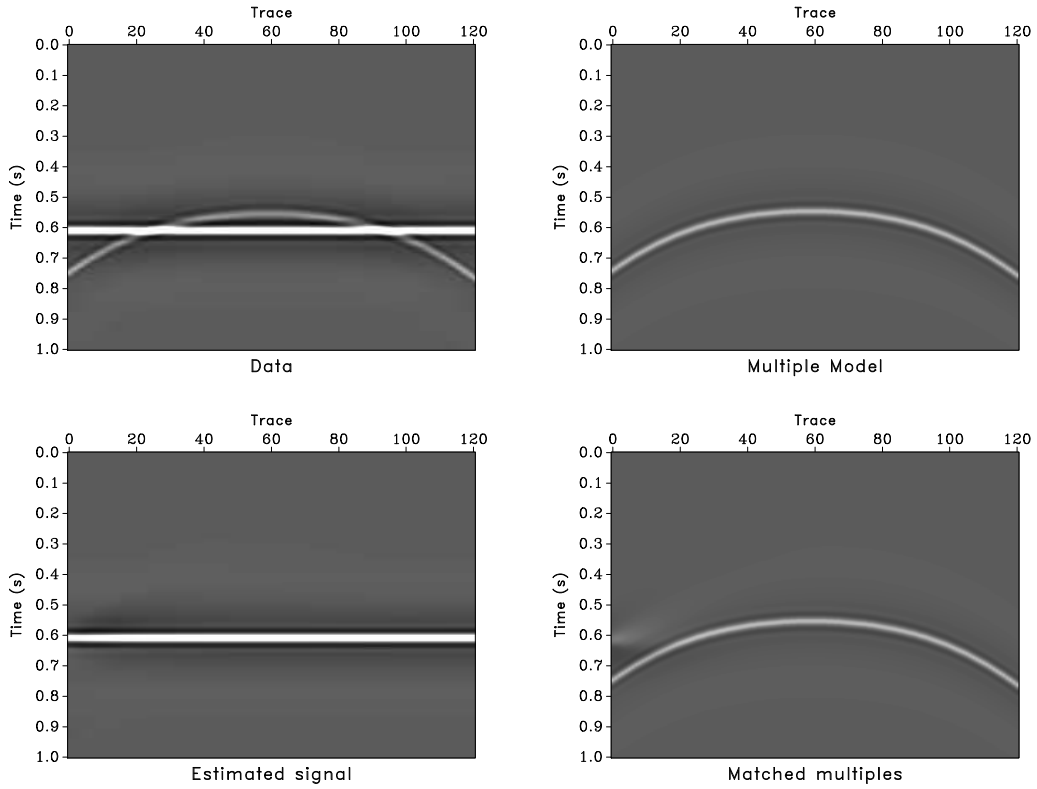


Figure 12: Benchmark test on adaptive subtraction from (1). a – Input synthetic data. b – Model of the noise containing amplitude and phase differences with respect to the corresponding part of the data. c – Extracted signal. d – Estimated noise. lpf/ray curves

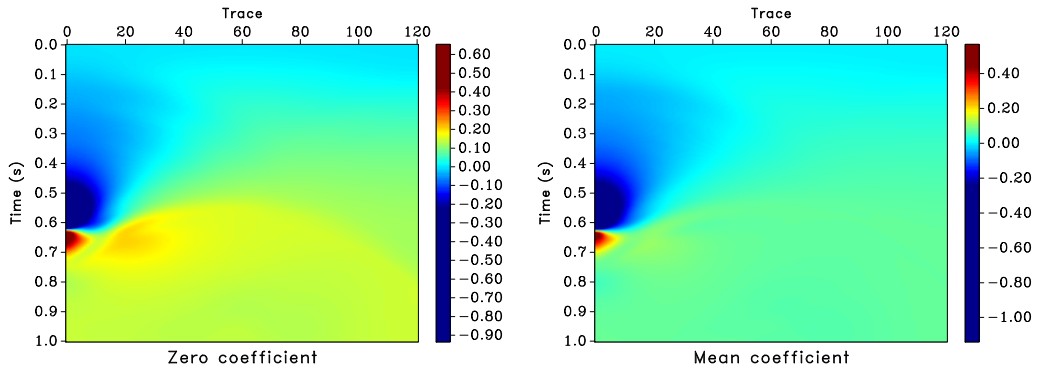


Figure 13: Variability of non-stationary match filter coefficients for the example shown in Figure 12. a – Zero-lag coefficient. b – Mean coefficient. lpf/ray dfilt

Spitz test

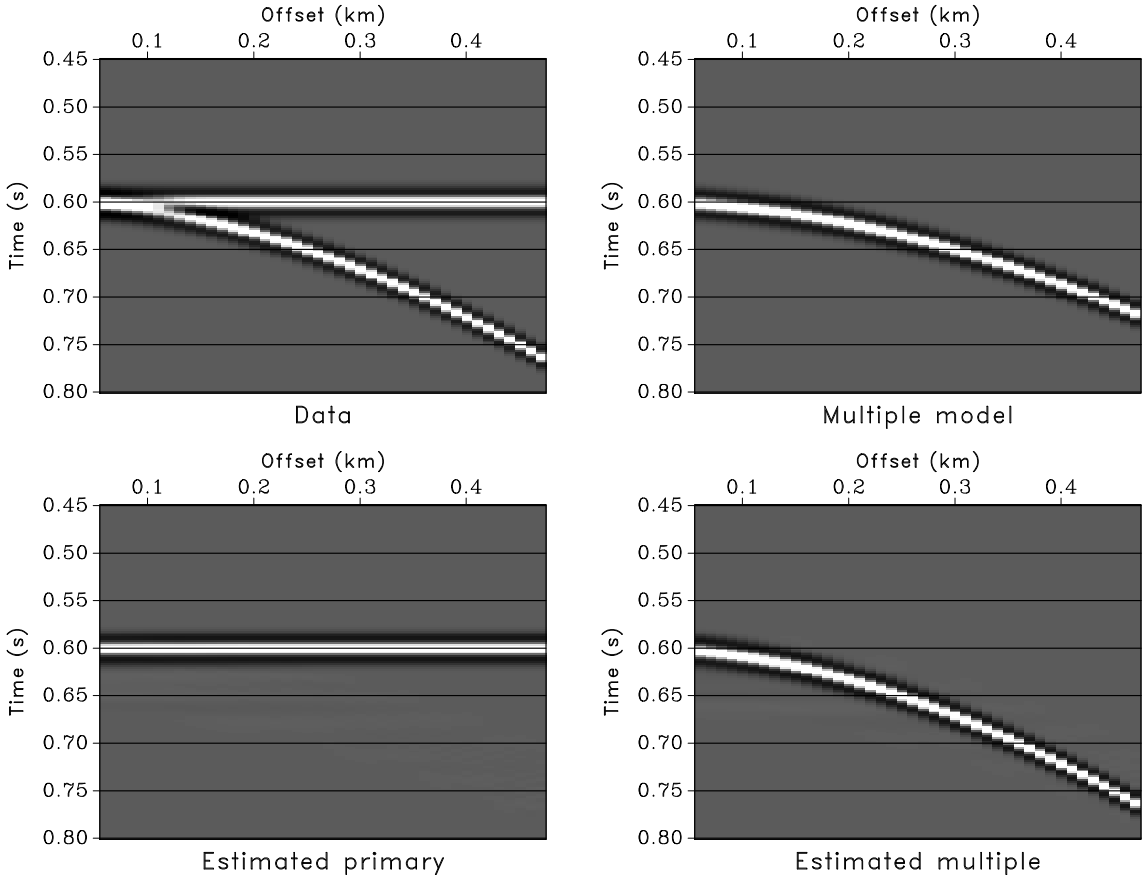


Figure 14: Benchmark test on multiple attenuation from (19). a – Input synthetic data containing a signal (primary reflection) and noise (multiple reflection) events. b – Model of the noise containing phase differences with respect to the corresponding part of the data. c – Extracted signal. d – Estimated noise. lpf/simon simon

The next benchmark example (Figure 14) reproduces a test created by (19), which simulates a seismic gather after migration or normal moveout correction. The signal (primary reflection) event is horizontal while the noise (multiple reflection) event exhibits curvature and overlaps with the signal at small offsets (Figure 14a). Multiple prediction (Figure 14b) contains a curved event but with incorrect curvature. As in the previous examples, non-stationary regularized regression correctly predicts the noise signal (Figure 14d) using match filtering and produces an accurate signal estimate (Figure 14c).

Pluto test

Finally, Figure 15 shows an application of the nonstationary matching technique to the Pluto synthetic dataset, a well-known benchmark for adaptive multiple sub-

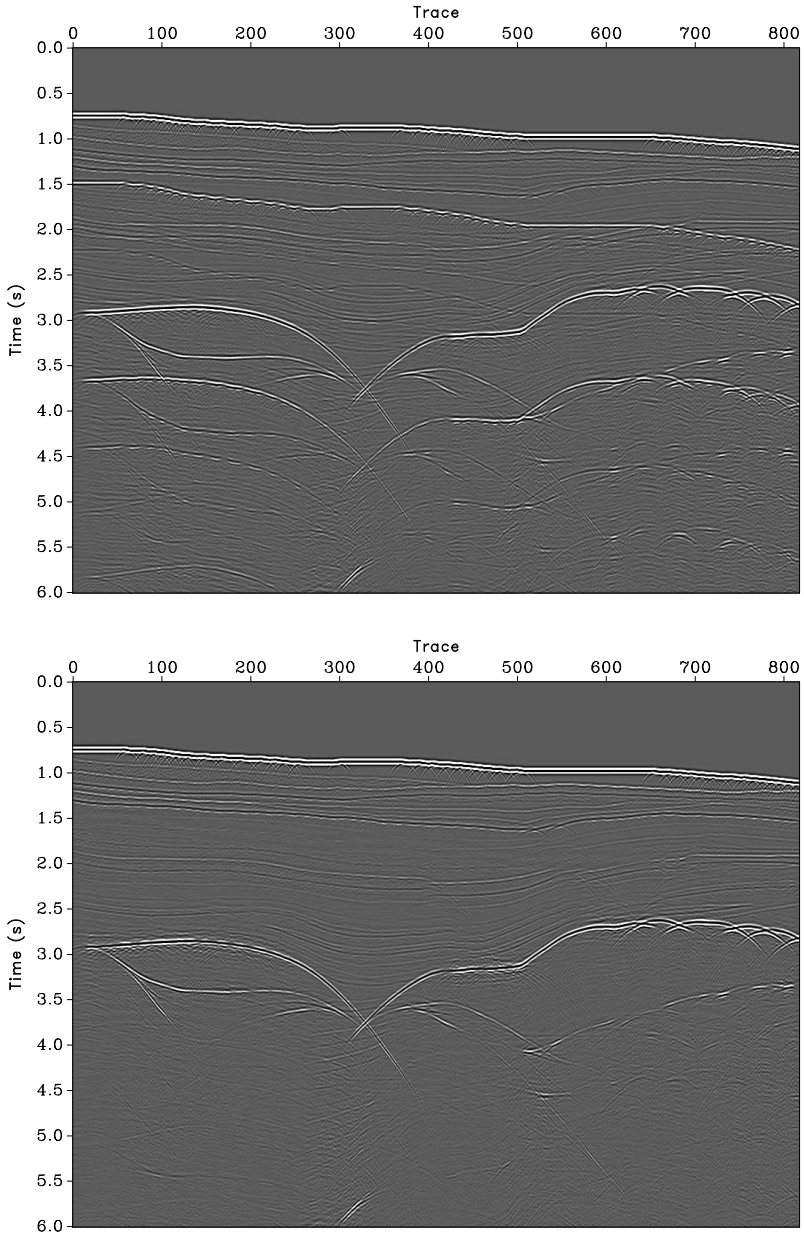


Figure 15: Adaptive multiple subtraction in the Pluto synthetic dataset. (a) Input data. (b) Extracted Signal. Surface-related multiples are successfully subtracted. lpf/plut ref,sig

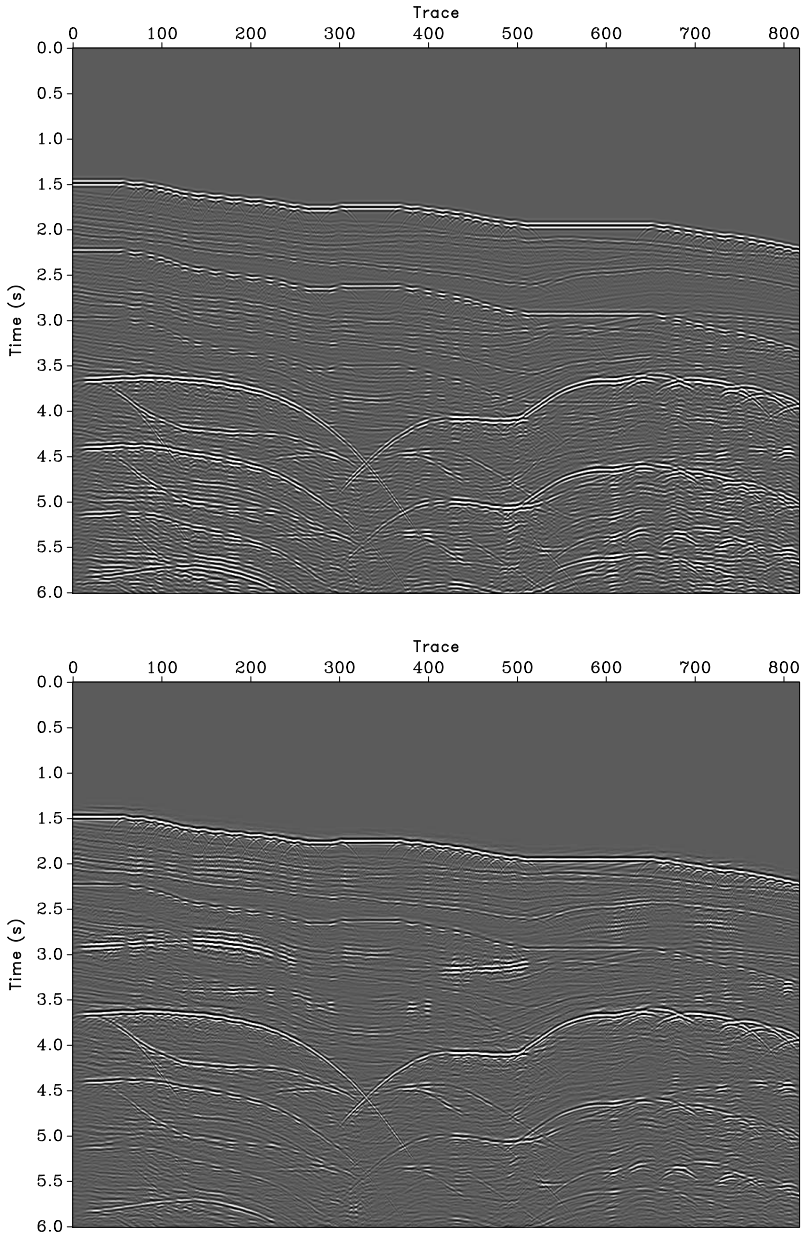


Figure 16: Multiple model from surface-related prediction (a) and estimated multiples (b) for the Pluto synthetic dataset. lpf/plut mod,pre

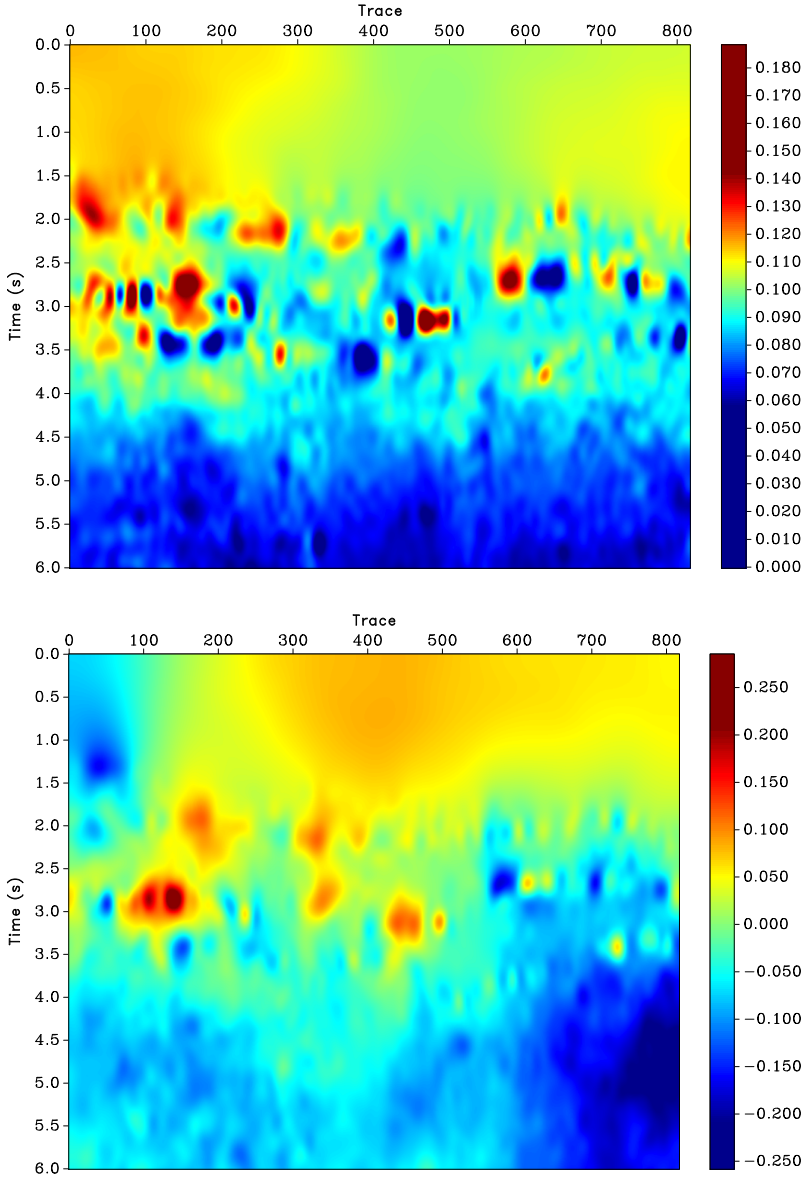


Figure 17: Variability of non-stationary match filter coefficients for the Pluto test. (a) Zero-lag coefficient. (b) Mean coefficient. lpf/plut zero,csum

traction. Matching and subtracting an imperfect model of the multiples created by the surface-related multiple elimination approach of (24) leaves a clean estimate of primary reflections. Figure 16 shows a comparison between the multiple model obtained by surface-related prediction and the multiple model generated by nonstationary matching. The matching filter non-stationarity is depicted in Figure 17, which shows the variability of filter coefficients with time and space.

CONCLUSIONS

I have presented a general method for regularized nonstationary regression. The key idea is the application of shaping regularization for constraining the variability of nonstationary coefficients. Shaping regularization has clear advantages over conventional Tikhonov's regularization: a more intuitive selection of regularization parameters and a faster iterative convergence because of better conditioning and eigenvalue clustering of the inverted matrix.

I have shown examples of applying regularized regression to benchmark tests of adaptive multiple subtraction. When the signal characteristics change with time or space, regularized nonstationary regression provides an effective description of the signal. This approach does not require breaking the input data into local windows, although it is analogous conceptually to sliding (maximum overlap) spatial-temporal windows.

ACKNOWLEDGMENTS

I thank StatoilHydro for a partial financial support of this work. I would like also to thank Dave Hale, Long Jin, Dmitry Lokshtanov, and Gary Sitton for inspiring discussions and Ray Abma for providing benchmark datasets for adaptive subtraction. The Pluto synthetic was generated by the SMAART consortium.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

REFERENCES

[1] Abma, R., N. Kabir, K. H. Matson, S. Michell, S. Shaw, and B. McLain, 2005, Comparisons of adaptive subtraction methods for multiple attenuation: The Leading Edge, **24**, 277–280.

[2] Claerbout, J., 2008, Image estimation by example: Environmental soundings image enhancement: Stanford Exploration Project.

[3] Crawley, S., J. Claerbout, and R. Clapp, 1999, Interpolation with smoothly nonstationary prediction-error filters: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1154–1157.

[4] Curry, W., 2003, Interpolation of irregularly sampled data with nonstationary, multiscale prediction-error filters: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 1913–1916.
[5] Denisov, M. S., M. A. Polubojarinov, and D. B. Finikov, 2006, Robust methods of adaptive seismic multiple suppression: EAGE St. Petersburg, Expanded Abstracts, European Association of Geoscientists & Engineers, B033.
[6] Engl, H., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.
[7] Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.
[8] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.
[9] Golub, G. H., and C. F. Van Loan, 1996, Matrix computations: The John Hopkins University Press.
[10] Guitton, A., 2004, Amplitude and kinematic corrections of migrated images for nonunitary imaging operators: Geophysics, **69**, 1017–1024.
[11] ———, 2005, Multiple attenuation in complex geology with a pattern-based approach: Geophysics, **70**, V97–V107.
[12] Guitton, A., and D. Verschuur, 2004, Adaptive subtraction of multiples using the L1-norm: Geophys. Prosp., **52**, 27–38.
[13] Haykin, S., 2001, Adaptive filter theory: Prentice Hall.
[14] Lu, W., and F. Mao, 2005, Adaptive multiple subtraction using independent component analysis: The Leading Edge, **24**, 282–284.
[15] Margrave, G. F., 1998, Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering: Geophysics, **63**, 244–259.
[16] Monk, D. J., 1993, Wave-equation multiple suppression using constrained gross-equalization: Geophys. Prosp., **41**, 725–736.
[17] Rickett, J., A. Guitton, and D. Gratwick, 2001, Adaptive Multiple Subtraction with Non-Stationary Helical Shaping Filters: 63rd Mtg., Eur. Assn. Geosci. Eng., Session: P167.
[18] Spitz, S., 1999, Pattern recognition, spatial predictability, and subtraction of multiple events: The Leading Edge, **18**, 55–58.
[19] ———, 2007, Perturbed prediction-driven multiple elimination: A PEF scheme: 77th Ann. Internat. Mtg, Soc. of Expl. Geophys., 2649–2653.
[20] Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Sovet Math. Dokl., 1035–1038.
[21] Valenciano, A. A., B. Biondi, and A. Guitton, 2006, Target-oriented wave-equation inversion: Geophysics, **71**, A35–A38.
[22] van Borselen, R., G. Fookes, and J. Brittan, 2003, Target-oriented adaptive subtraction in data-driven multiple removal: The Leading Edge, **22**, 340–343.
[23] van der Vorst, H. A., 2003, Iterative Krylov methods for large linear systems: Cambridge University Press.
[24] Verschuur, D. J., A. J. Berkhout, and C. P. A. Wapenaar, 1992, Adaptive surface-related multiple elimination: Geophysics, **57**, 1166–1177.
[25] Wang, Y., 2003, Multiple subtraction using an expanded multichannel matching filter: Geophysics, **68**, 346–354.

First-break traveltime tomography with the double-square-root eikonal equation

*Siwei Li[1], Alexander Vladimirsky[2] and Sergey Fomel[1]**

ABSTRACT

First-break traveltime tomography is based on the eikonal equation. Since the eikonal equation is solved at fixed shot positions and only receiver positions can move along the ray-path, the adjoint-state tomography relies on inversion to resolve possible contradicting information between independent shots. The double-square-root eikonal equation allows not only the receivers but also the shots to change position, and thus describes the prestack survey as a whole. Consequently, its linearized tomographic operator naturally handles all shots together, in contrast with the shot-wise approach in the traditional eikonal-based framework. The double-square-root eikonal equation is singular for the horizontal waves, which require special handling. Although it is possible to recover all branches of the solution through post-processing, our current forward modeling and tomography focus on the diving wave branch only. We consider two upwind discretizations of the double-square-root eikonal equation and show that the explicit scheme is only conditionally convergent and relies on non-physical stability conditions. We then prove that an implicit upwind discretization is unconditionally convergent and monotonically causal. The latter property makes it possible to introduce a modified fast marching method thus obtaining first-break traveltimes both efficiently and accurately. To compare the new double-square-root eikonal-based tomography and traditional eikonal-based tomography, we perform linearizations and apply the same adjoint-state formulation and upwind finite-differences implementation to both approaches. Synthetic model examples justify that the proposed approach converges faster and is more robust than the traditional one.

INTRODUCTION

The first-break traveltime tomography (46; 26; 22; 36; 25) has been an established tool for estimating near-surface macro-feature seismic velocities. Starting from a prior model, tomographic inversion gradually modifies the velocities such that the misfits between predicted and observed first-breaks decrease. Since the problem is nonlinear, several linearization iterations may be required until convergence. Moreover, inversion must be carried out with careful choice of regularization in order to avoid local

*e-mail: sergey.fomel@beg.utexas.edu

minima (35; 34; 14). The estimated model has a direct influence on subsequent applications, for example static corrections (24; 9; 3) where it provides a medium-to-long wavelength near-surface model, and waveform tomography (33; 5; 38) where it serves as a low-frequency prior.

The traditional first-break traveltime tomography is based on the eikonal equation that arises from high-frequency approximation of the wave equation (8). During forward modeling, the first-breaks computed through the eikonal equation are naturally shot-indexed because only receiver coordinates move while the source is fixed. At tomography stage, one may formulate the minimization of cost function as a sequence of explicitly linearized problems or directly as a nonlinear optimization problem. The first choice (42; 45; 10; 28) requires computation of Fréchet derivatives, which is usually carried out by combining an eikonal solver with posterior ray tracing. Then an algorithm such as LSQR (27) is applied to solve the linearized tomographic system iteratively. While this approach accounts for information from both source and receiver dimensions, it faces computational limitations when the Fréchet derivative matrix becomes difficult to handle because of a large number of model parameters. The nonlinear optimization approach, on the other hand, can be combined with the adjoint-state method (29) and avoids an explicit computation of Fréchet derivatives (36). The cost of computing gradient is equivalent to twice the solution of the forward modeling problem, regardless of the size of input data. However, one major drawback of this approach, as we will show later, is that the resulting gradient disregards information available along the shot dimension.

The drawback of eikonal-based adjoint-state tomographies is that they always face conflicting information that propagates across different shots. Such conflicts must be resolved during inversion, or else an erroneous model update may appear. In practice, the inversion may be less robust and may take more iterations to converge, compared to the situation where we replace the eikonal equation with another governing equation that allows both source and receiver positions to change along ray-paths. The double-square-root (DSR) eikonal equation is a promising candidate in this regard, because it describes the prestack data as a whole by linking the evolution of traveltimes to both sub-surface source and receiver positions. In this paper, we investigate the feasibility of using the DSR eikonal equation for first-break traveltime tomography with the adjoint-state method.

DSR eikonal was analyzed previously by (2), (13) and (1). Ray-tracing methods applied to DSR are capable of providing multi-arrivals by extrapolating isochron rays (17) or using perturbation theory, but their extra costs in computing non-first-breaks are not necessary for first-break tomography purpose. We first prove that an implicit discretization of the DSR eikonal equation is causal and thus can be solved by a Dijkstra-like non-iterative method (12). The DSR singularity and two DSR branches that are non-causal need special treatment. Our current implementation employs a modified fast-marching (32) DSR eikonal solver. We first test its accuracy by DSR forward modeling. Next, we linearize the DSR eikonal equation and use the resulting operators in adjoint-state tomography. For comparison, we apply an

analogous linearization and adjoint-state formulation to the traditional tomography based on shot-indexed eikonal equation. Then we demonstrate the differences between the proposed and traditional approaches and justify advantages of the new method using several synthetic model examples. We conclude by discussing possible further improvements and extensions of our method.

THEORY

DSR eikonal equation

The DSR eikonal equation can be derived by considering a ray-path and its segments between two depth levels. Figure 1 illustrates a diving ray (46) in 2-D with velocity $v = v(z, x)$. We denote $T(z, r, s)$ as the total traveltime of the ray-path beneath depth z, where r and s are *sub-surface* receiver and source lateral positions, respectively.

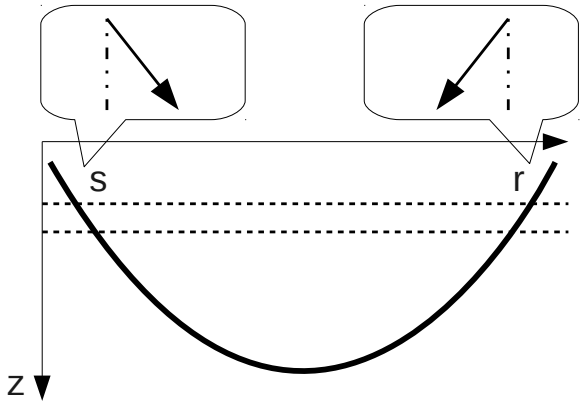


Figure 1: A diving ray and zoom-in of the ray segments between two depth levels. dsrtomo/. raypath

At both source and receiver sides the traveltime satisfies the eikonal equation, therefore

$$\frac{\partial T}{\partial z} = -\sqrt{\frac{1}{v^2(z,s)} - \left(\frac{\partial T}{\partial s}\right)^2} - \sqrt{\frac{1}{v^2(z,r)} - \left(\frac{\partial T}{\partial r}\right)^2}\,. \tag{1}$$

The negative signs before the two square-roots in equation 1 correspond to a decrease of traveltime with increasing depth, or geometrically a *downward* pointing of slowness vectors on both s and r sides. Since the slowness vectors could also be pointing *upward* and the directions may be different at r and s, the DSR eikonal equation (2) should

account for all the possibilities (Figure 2):

$$\frac{\partial T}{\partial z} = \pm\sqrt{\frac{1}{v^2(z,s)} - \left(\frac{\partial T}{\partial s}\right)^2} \pm \sqrt{\frac{1}{v^2(z,r)} - \left(\frac{\partial T}{\partial r}\right)^2} \ . \tag{2}$$

The boundary condition for DSR eikonal equation is that traveltimes at the subsurface zero-offset plane, i.e. $r = s$, are zero: $T(z, r = s) = 0$.

Equation 2 has a *singularity* when $\partial T/\partial z = 0$, in which case the slowness vectors at s and r sides are both horizontal and equation 2 reduces to

$$\left(\frac{\partial T}{\partial s}\right)^2 = \frac{1}{v^2(z,s)} \ ; \ \ \left(\frac{\partial T}{\partial r}\right)^2 = \frac{1}{v^2(z,r)} \ . \tag{3}$$

The two independent equations in 3 are not in conflict according to the source-receiver reciprocity, because they are the same with an exchange of s and r.

Note that equations 2 and 3 describe T in full prestack domain (z, r, s) by allowing not only receivers but also sources to change positions. In contrast, the eikonal equation

$$\left(\frac{\partial T}{\partial z}\right)^2 + \left(\frac{\partial T}{\partial x}\right)^2 = \frac{1}{v^2(z,x)} \tag{4}$$

with boundary condition $T(z = 0, x = s) = 0$ can be used only for one fixed source position at a time and thus traveltimes of different shots are independent of each other. In equation 4 s is *surface* source lateral position. In the 3-D case, the scalars s, r and x in equations 2, 3 and 4 become 2-D vectors $\mathbf{s}$, $\mathbf{r}$ and $\mathbf{x}$ that contain in-line and cross-line positions. The prestack traveltime is then in a 5-D space. Our current work is restricted to 2-D and we consider the 3-D extension in the Discussion section.

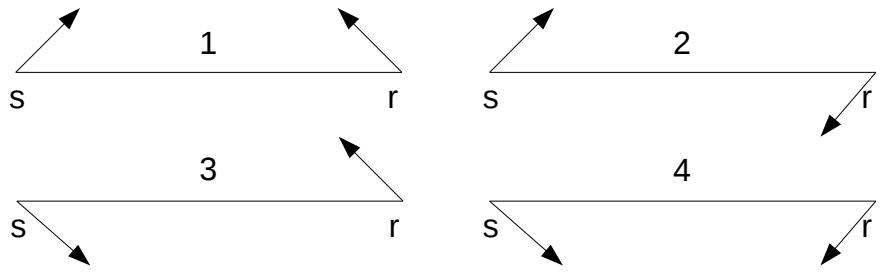


Figure 2: All four branches of DSR eikonal equation from different combination of upward or downward pointing of slowness vectors. Whether the slowness vector is pointing leftward or rightward does not matter because the partial derivatives with respect to s and r in equation 2 are squared. Figure 1 and equation 1 belong to the last situation. dsrtomo/. root

Similarly to the eikonal equation, the DSR eikonal equation is a nonlinear first-order partial differential equation. Its solutions include in general not only first-breaks but all arrivals, and can be computed by solving separate eikonal equations for each sub-surface source-receiver pair followed by extracting the traveltime and putting the value into prestack volume. However, such an implementation is impractical due to the large amount of computations. Meanwhile, for first-break tomography purposes, we are only interested in the first-arrival solutions but require an efficient and accurate algorithm. In this regard, a finite-difference DSR eikonal solver analogous to the fast-marching (32) or fast-sweeping (43) eikonal solvers is preferable.

In upwind discretizations of the DSR eikonal equation on the grid in (z, r, s) domain, one has to make a decision about the z-slice, in which the finite differences are taken to approximate $\partial T/\partial s$ and $\partial T/\partial r$. In Figure 1, it appears natural to approximate these partial derivatives in the z-slice below $T(z, r, s)$. We refer to the corresponding scheme as *explicit*, since it allows to directly compute the grid value $T(z, r, s)$ based on the already known T values from the next-lower z. An alternative *implicit* scheme is obtained by approximating $\partial T/\partial s$ and $\partial T/\partial r$ in the same z-slice as $T(z, r, s)$, which results in a coupled system of nonlinear discretized equations. In Appendix A, we prove the following:

1. The explicit scheme is very efficient to use on a fixed grid, but only conditionally convergent. This property is also confirmed numerically in Synthetic Model Examples section.

2. The implicit scheme is *monotone causal*, meaning $T(z, r, s)$ depends on the smaller neighboring grid values only. This enables us to apply a Dijkstra-like method (12) to solve the discretized system efficiently. Importantly, the DSR singularity requires a special ordering in the selection of upwind neighbors, which switches between equations 1 and 3 when necessary. We provide a modified fast-marching (32) DSR eikonal solver along with such an ordering strategy in Numerical Implementation section.

3. The causality analysis in Appendix A applies only to the first and last *causal branches* out of all four shown in Figure 2. Additional post-processings, albeit expensive, can be used to recover the rest two non-causal branches as they may be decomposed into summations of the causal ones.

In practice, we find that, for moderate velocity variations, the first-breaks correspond only to causal branches. An example in Synthetic Model Examples section serves to illustrate this observation. Therefore, for efficiency, we turn off the non-causal branch post-processings in forward modeling and base the tomography solely on equations 1 and 3.

DSR tomography

The first-break traveltime tomography with DSR eikonal equation (DSR tomography) can be established by following a procedure analogous to the traditional one with the shot-indexed eikonal equation (standard tomography). To further reveal their differences, in this section we will derive both approaches.

For convenience, we use slowness-squared $w \equiv 1/v^2$ instead of velocity v in equations 1, 3 and 4. Based on analysis in Appendix A, the velocity model $w(z,x)$ and prestack cube $T(z,r,s)$ are Eulerian discretized and arranged as column vectors $\mathbf{w}$ of size $nz \times nx$ and $\mathbf{t}$ of size $nz \times nx \times nx$. We denote the observed first-breaks by $\mathbf{t}^{obs}$, and use $\mathbf{t}^{std}$ and $\mathbf{t}^{dsr}$ whenever necessary to discriminate between $\mathbf{t}$ computed from shot-indexed eikonal equation and DSR eikonal equation.

The tomographic inversion seeks to minimize the l_2 (least-squares) norm of the data residuals. We define an objective function as follows:

$$E(\mathbf{w}) = \frac{1}{2}(\mathbf{t} - \mathbf{t}^{obs})^T(\mathbf{t} - \mathbf{t}^{obs}) \ , \tag{5}$$

where the superscript T represents transpose. A Newton method of inversion can be established by considering an expansion of the misfit function 5 in a Taylor series and retaining terms up to the quadratic order (4):

$$E(\mathbf{w} + \delta\mathbf{w}) = E(\mathbf{w}) + \delta\mathbf{w}^T\nabla_w E(\mathbf{w}) + \frac{1}{2}\delta\mathbf{w}^T\mathbf{H}(\mathbf{w})\delta\mathbf{w} + O(|\delta\mathbf{w}|^3) \ . \tag{6}$$

Here $\nabla_w E$ and $\mathbf{H}$ are gradient vector and Hessian matrix, respectively. We may evaluate the gradient by taking partial derivatives of equation 5 with respect to $\mathbf{w}$, yielding

$$\nabla_w E \equiv \frac{\partial E}{\partial \mathbf{w}} = \mathbf{J}^T(\mathbf{t} - \mathbf{t}^{obs}) \ , \tag{7}$$

where $\mathbf{J}$ is the Frechét derivative matrix and can be found by further differentiating $\mathbf{t}$ with respect to $\mathbf{w}$.

We start by deriving the Frechét derivative matrix of standard tomography. Denoting

$$D_m \equiv \frac{\partial}{\partial m} \ ; \ \ m = z, x, r, s \tag{8}$$

as the partial derivative operator in the m's direction, equation 4 can be re-written as

$$D_z t_k^{std} \cdot D_z t_k^{std} + D_x t_k^{std} \cdot D_x t_k^{std} = w \ ; \ \ k = 1, 2, 3, ..., nx \ . \tag{9}$$

Here we assume that there are in total nx shots and use t_k^{std} for first-breaks of the kth shot. Applying $\partial/\partial w$ to both sides of equation 9, we find

$$J_k^{std} \equiv \frac{\partial t_k^{std}}{\partial w} = \frac{1}{2}(D_z t_k^{std} \cdot D_z + D_x t_k^{std} \cdot D_x)^{-1} \ ; \ \ k = 1, 2, 3, ..., nx \ . \tag{10}$$

Kinematically, each J_k^{std} contains characteristics of the kth shot. Because shots are independent of each other, the full Frechét derivative is a concatenation of individual J_k^{std}, as follows:

$$J^{std} = \begin{bmatrix} J_1^{std} & J_2^{std} & \cdots & J_{nx}^{std} \end{bmatrix}^T \ . \tag{11}$$

Inserting equation 11 into equation A-1, we obtain

$$\nabla_w E = \sum_{k=1}^{nx} \left(\mathbf{J}_k^{std}\right)^T \left(\mathbf{t}_k^{std} - \mathbf{t}_k^{obs}\right) , \tag{12}$$

where, similar to t_k^{std}, t_k^{obs} stands for the observed first-breaks of the kth shot.

Figure 3 illustrates equation 12 schematically, i.e. the gradient produced by standard tomography. The first step on the left depicts the transpose of the kth Frechét derivative acting on the corresponding kth data residual. It implies a back-projection that takes place in the $z-r$ plane of a fixed s position. The second step on the right is simply the summation operation in equation 12.

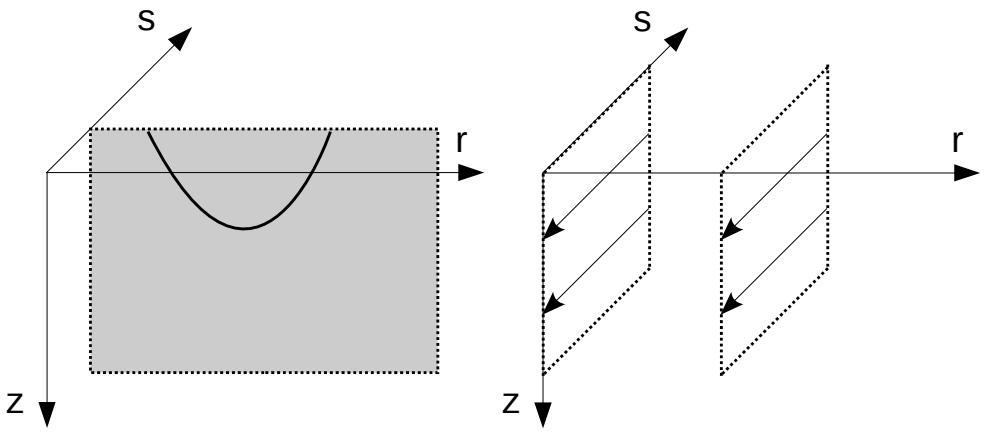


Figure 3: The gradient produced by standard tomography. The solid curve indicates a shot-indexed characteristic. dsrtomo/. cartonstd

To derive the Frechét derivative matrix associated with DSR tomography, we first re-write equation 1 with definition 8

$$D_z t^{dsr} = -\sqrt{w_s - D_s t^{dsr} \cdot D_s t^{dsr}} - \sqrt{w_r - D_r t^{dsr} \cdot D_r t^{dsr}} \ , \tag{13}$$

where w_s and w_r are w at sub-surface source and receiver locations, respectively. Note that in equation 13 w appears twice. Thus a differentiation of t^{dsr} with respect to w must be carried out through the chain-rule:

$$J^{dsr} \equiv \frac{\partial t^{dsr}}{\partial w} = \left.\frac{\partial t^{dsr}}{\partial w_s}\right|_{w_r} \frac{\partial w_s}{\partial w} + \left.\frac{\partial t^{dsr}}{\partial w_r}\right|_{w_s} \frac{\partial w_r}{\partial w} \ . \tag{14}$$

We recall that w and t^{dsr} are of different lengths. Meanwhile in equation 13, both w_s and w_r have the size of t^{dsr}. Clearly in equation 14 $\partial w_s/\partial w$ and $\partial w_r/\partial w$ must achieve dimensionality enlargement. In fact, according to Figure 1, w_s and w_r can be obtained by spraying w such that $w_s(z,r,s) = w(z,s)$ and $w_r(z,r,s) = w(z,r)$. Therefore, $\partial w_s/\partial w$ and $\partial w_r/\partial w$ are essentially spraying operators and their adjoints perform stackings along s and r dimensions, respectively.

In Appendix B, we prove that J^{dsr} has the following form:

$$J^{dsr} = B^{-1}(C_s + C_r) \; . \tag{15}$$

Combining equations A-1 and 15 results in

$$\nabla_w E = \left(\mathbf{C}_s^T + \mathbf{C}_r^T\right) \mathbf{B}^{-T}(\mathbf{t}^{dsr} - \mathbf{t}^{obs}) \; . \tag{16}$$

Note that unlike equation 12, equation 16 can not be expressed as an explicit summation over shots.

Figure 4 shows the gradient of DSR tomography. Similarly to the standard tomography, the gradient produced by equation 16 is a result of two steps. The first step on the left is a back-projection of prestack data residuals according to the adjoint of operator B^{-1}. Because B contains DSR characteristics that travel in prestack domain, this back-projection takes place in (z,r,s) and is different from that in standard tomography, although the data residuals are the same for both cases. The second step on the right follows the adjoint of operators C_s and C_r and reduces the dimensionality from (z,r,s) to (z,x). However, compared to standard tomography this step involves summations in not only s but also r.

NUMERICAL IMPLEMENTATION

Following the analysis in Appendix A, we consider an implicit Eulerian discretization. For forward modeling, we solve the DSR eikonal equation by a version of the fast-marching method (FMM) (32). First, a plane-wave with $T = 0$ at subsurface zero-offset $r = s$ is initialized. Next, in the update stage the traveltime at a grid point is computed from its upwind neighbors. A priority queue keeps track of the first-break wave-front, and the computation is non-recursive.

To properly handle the DSR singularity, we design an ordering of the combination of upwind neighbors during the update stage. Assuming that T^i is the upwind neighbor of T in the i's direction for $i = z, r, s$, we summarize the ordering as follows:

1. First try a three-sided update:

   Solve equation A-9, return T if $T \geq \max(T^z, T^r, T^s)$;

2. Next try a two-sided update: solve equations A-10, A-12 and A-13 and keep the results as T_{rs}, T_{zr} and T_{zs}, respectively.

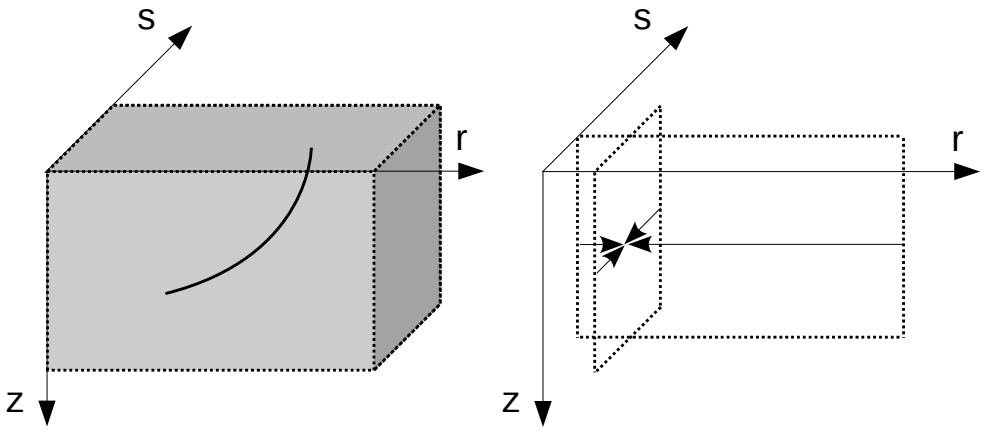


Figure 4: The gradient produced by DSR tomography. The solid curve indicates a DSR characteristic, which has one end in plane $z = 0$ and the other in plane $s = r$. Compare with Figure 3. dsrtomo/. cartondsr

If $T_{zr} \geq \max(T^z, T^r)$ and $T_{zs} \geq \max(T^z, T^s)$, return $\min(T_{zr}, T_{zs}, T_{rs})$;

If $T_{zr} < \max(T^z, T^r)$ and $T_{zs} \geq \max(T^z, T^s)$, return $\min(T_{zs}, T_{rs})$;

If $T_{zr} \geq \max(T^z, T^r)$ and $T_{zs} < \max(T^z, T^s)$, return $\min(T_{zr}, T_{rs})$;

3. Finally try a one-sided update:

   Solve equation A-14, return $\min(T, T_{rs})$.

An optional search routine A-17 may be added after the update to recover all branches of the DSR eikonal equation. The overall cost can be reduced roughly by half by acknowledging the source-receiver reciprocity and thus computing only the positive (or negative) subsurface offset region.

For an implementation of linearized tomographic operators 12 and 16, we choose upwind approximations (15; 23; 21) for the difference operators in equation 8. In Appendix C, we show that the upwind finite-differences result in triangularization of matrices 11 and 15. Therefore, the costs of applying $\mathbf{J}^{std}$ and $\mathbf{J}^{dsr}$ and their transposes are inexpensive. Moreover, although our implementation belongs to the family of adjoint-state tomographies, we do not need to compute the adjoint-state variable as an intermediate product for the gradient.

Additionally, the Gauss-Newton approach approximates the Hessian in equation 6 by $\mathbf{H} \approx \mathbf{J}^T\mathbf{J}$. An update $\delta\mathbf{w}$ at current $\mathbf{w}$ is found by taking derivative of equation

6 with respect to $\delta\mathbf{w}$, which results in the following normal equation:

$$\delta\mathbf{w} = \left[\mathbf{J}^T\mathbf{J}\right]^{-1}\mathbf{J}^T(\mathbf{t}^{obs} - \mathbf{t}) \; . \tag{17}$$

To add model constraints, we combine equation 17 with Tikhonov regularization (37) with the gradient operator and use the method of conjugate gradients (16) to solve for the model update $\delta\mathbf{w}$.

SYNTHETIC MODEL EXAMPLES

The numerical examples in this section serve several different purposes. The first example will test the accuracy of modified FMM DSR eikonal equation solver (DSR FMM) and show the drawbacks of the alternative explicit discretization. The second example will demonstrate effect of considering non-causal branches of DSR eikonal equation in forward modeling. The third example will compare the sensitivity kernels of DSR tomography and standard tomography in a simple model. The last example will present a tomographic inversion and demonstrate advantages of DSR method over the traditional method.

Figure 5 shows a 2-D velocity model with a constant-velocity-gradient background plus a Gaussian anomaly in the middle. We use Δ to denote the grid spacing in z and δ in x. The traveltimes on the surface $z = 0$ km of a shot at $(0, 0)$ km are computed by DSR FMM at a gradually refined Δ or δ while fixing the other one. For reference, we also calculate first-breaks by a second-order FMM (31; 30) for the same shot at a very fine grid spacing of $\Delta = \delta = 1$ m. In Figure 6, a grid refinement in both Δ and δ helps reducing errors of the implicit discretization, although improvements in the Δ refinement case are less significant because the majority of the ray-paths are non-horizontal. The results are consistent with the analysis in Appendix A, which shows that the implicit discretization is unconditionally convergent. On the other hand, as shown in Figure 7, the explicit discretization is only conditionally convergent when $\Delta/\delta \to 0$ under grid refinement in order to resolve the flatter parts of the ray-paths. This explains why its accuracy deteriorates when refining δ and fixing Δ. A more detailed error analysis remains open for future research.

Next, we use a smoothed Marmousi model (Figure 8) and run two DSR FMMs, one with the search process for non-causal DSR branches turned-on and the other turned-off. In Figure 9, again we compute reference values by a second-order FMM. The three groups of curves are traveltimes of shots at $(0, 0)$ km, $(0.75, 0)$ km and $(1.5, 0)$ km, respectively. The maximum absolute differences between the two DSR FMMs, for all three shots, are approximately 5 ms at the largest offset. This shows that, if the near-surface model is moderately complex, then the first-breaks are of causal types described by equations 1 and 3, and we therefore can use their linearizations 15 for tomography.

According to equations 11 and 15, the sensitivity kernels (a row of Frechét derivative matrix) of standard tomography and DSR tomography are different. Figure 10

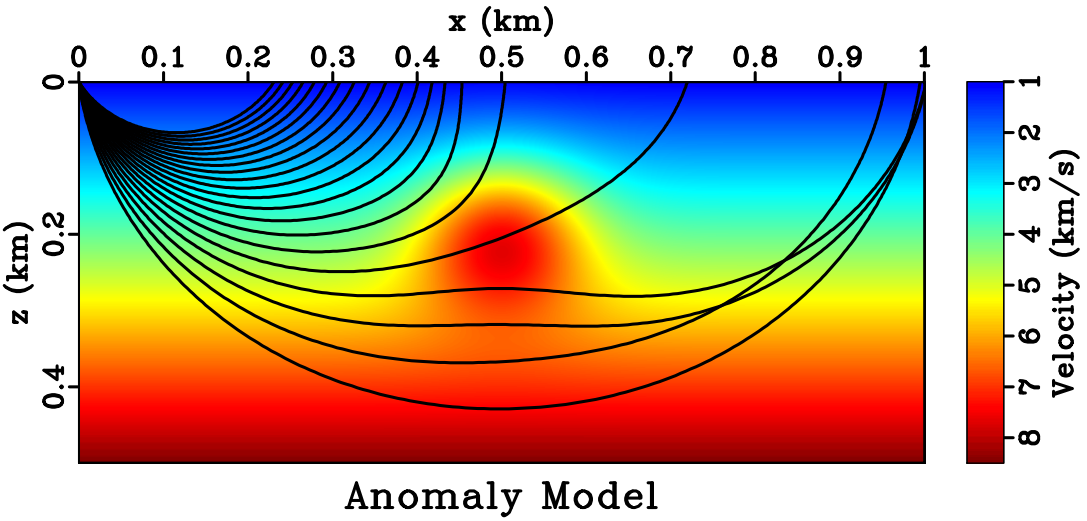


Figure 5: The synthetic model used for DSR FMM accuracy test. The overlaid curves are rays traced from a shot at $(0, 0)$ km. dsrtomo/accuracy modl

compares sensitivity kernels for the same source-receiver pair in a constant velocity-gradient model. We use a fine model sampling of $\Delta = \delta = 2.5$ m. The standard tomography kernel appears to be asymmetric. Its amplitude has a bias towards the source side, while the width is broader on the receiver side. These phenomena are related to our implementation, as described in Appendix C. Note in the top plot of Figure 10, the curvature of first-break wave-front changes during propagation. Upwind finite-differences take the curvature variation into consideration and, as a result, back-project data-misfit with different weights along the ray-path. Meanwhile, the DSR tomography kernel is symmetric in both amplitude and width, even though it uses the same discretization and upwind approximation as in standard tomography. The source-receiver reciprocity may suggest averaging the standard tomography kernel with its own mirroring around $x = 1$ km, however the result will still be different from the DSR tomography kernel as the latter takes into consideration all sources at the same time.

Finally, Figure 9(a) illustrates a prestack first-break traveltime modeling of the Marmousi model by DSR FMM. We use a constant-velocity-gradient model as the prior for inversion. There are 287 shots evenly distributed on the surface, each shot has a maximum absolute receiver offset of 6 km. Figure 12 shows a zoom-in of the exact model that is within the tomographic aperture. The DSR tomography and standard tomography are performed with the same parameters: 10 conjugate gradient iterations per linearization update and 4 linearization updates in total. Figure 13 shows the convergence histories. While both inversions converge, the relative l_2 data misfits of DSR tomography decreases faster than that of standard tomography. Figure 14 compares the recovered models. Although both results resemble the exact model in Figure 12 at the large scale, the standard tomography model exhibits sev-

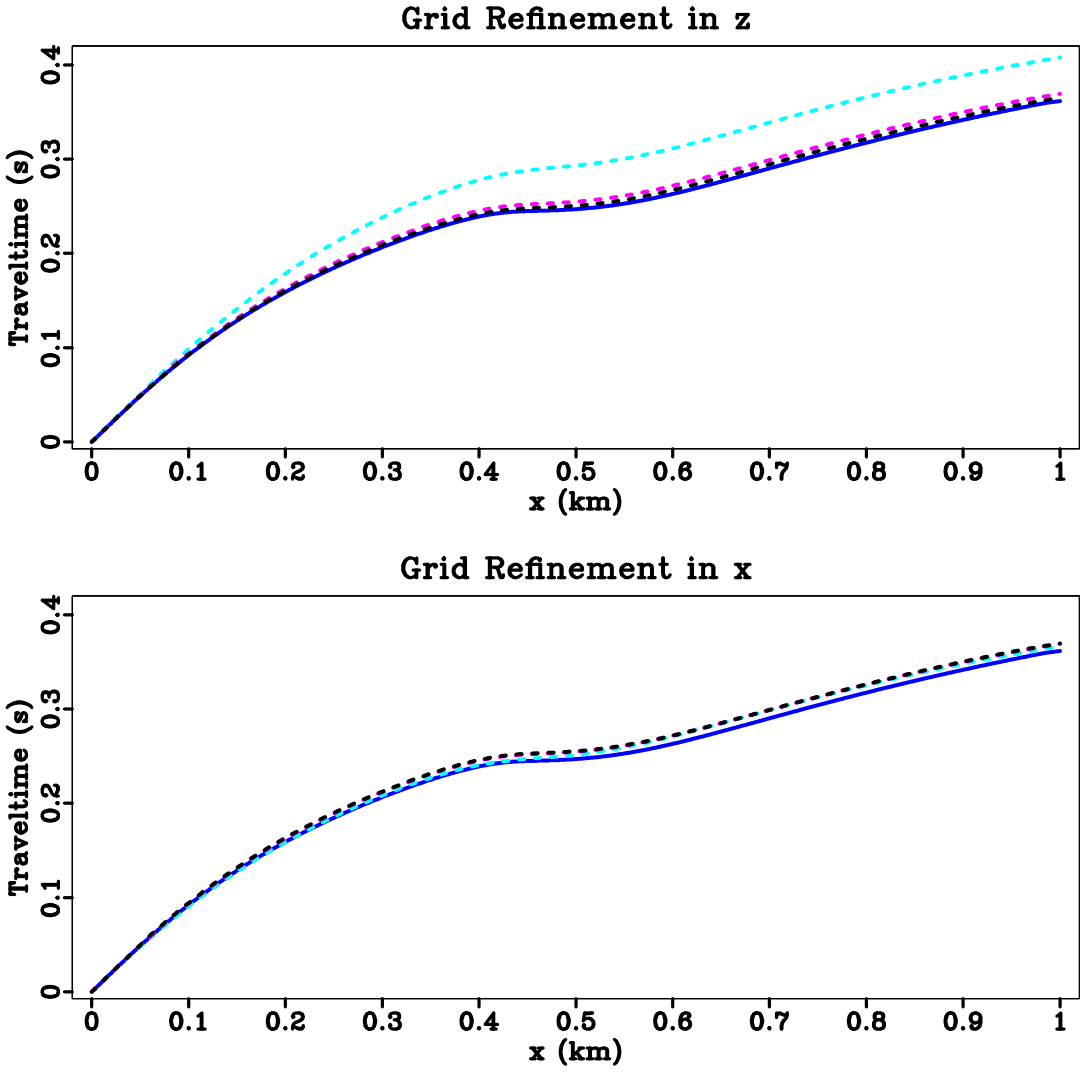


Figure 6: Grid refinement experiment (implicit discretization). In both figures, the solid blue curve is the reference values and the dashed curves are computed by DSR FMM. Top: fixed $\delta = 10$ m and $\Delta = 50$ m (cyan), 10 m (magenta), 5 m (black). Bottom: fixed $\Delta = 10$ m and $\delta = 50$ m (cyan), 10 m (magenta), 5 m (black). dsrtomo/accuracy imp

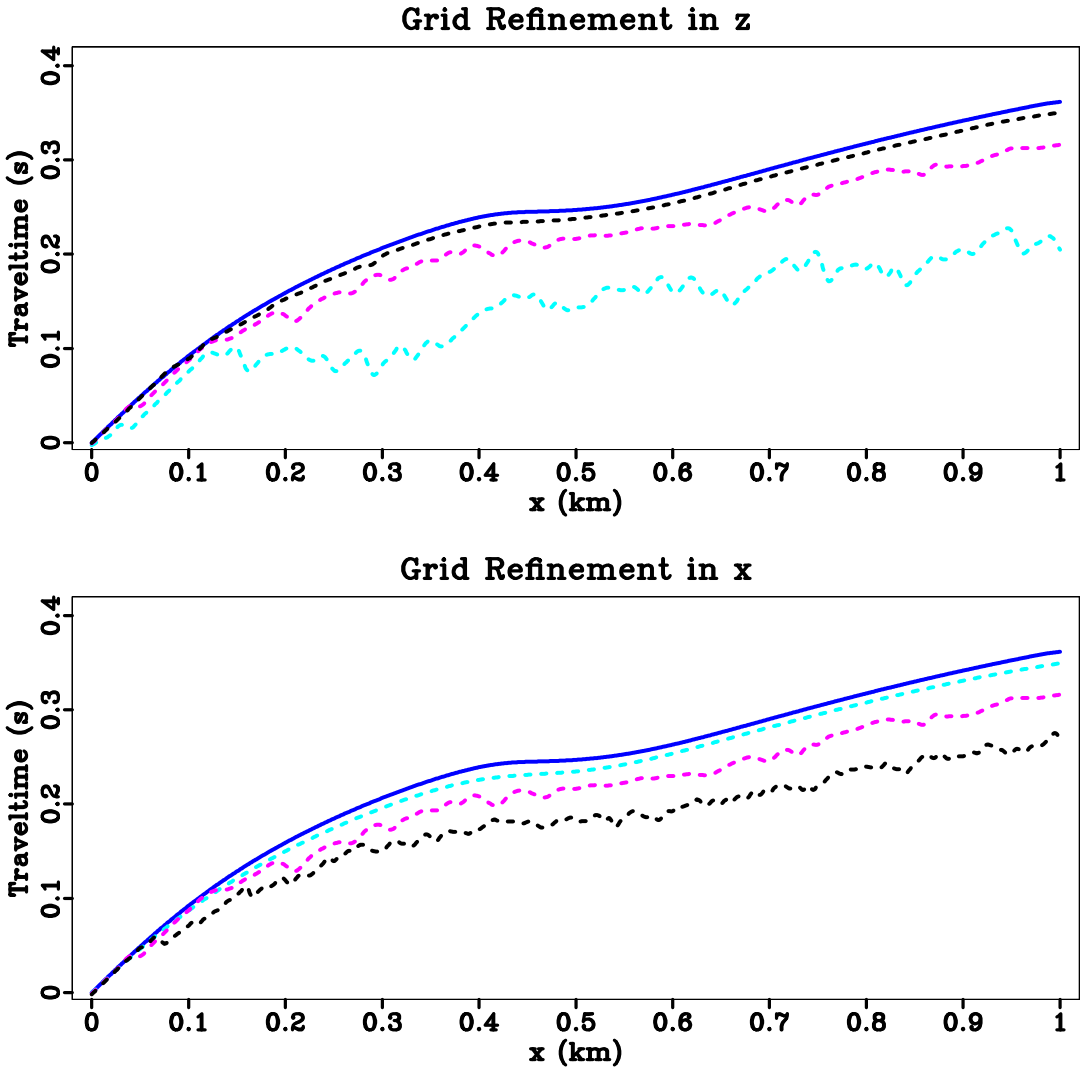


Figure 7: Grid refinement experiment (explicit discretization). The experiment setups are the same as in Figure 6. dsrtomo/accuracy exp

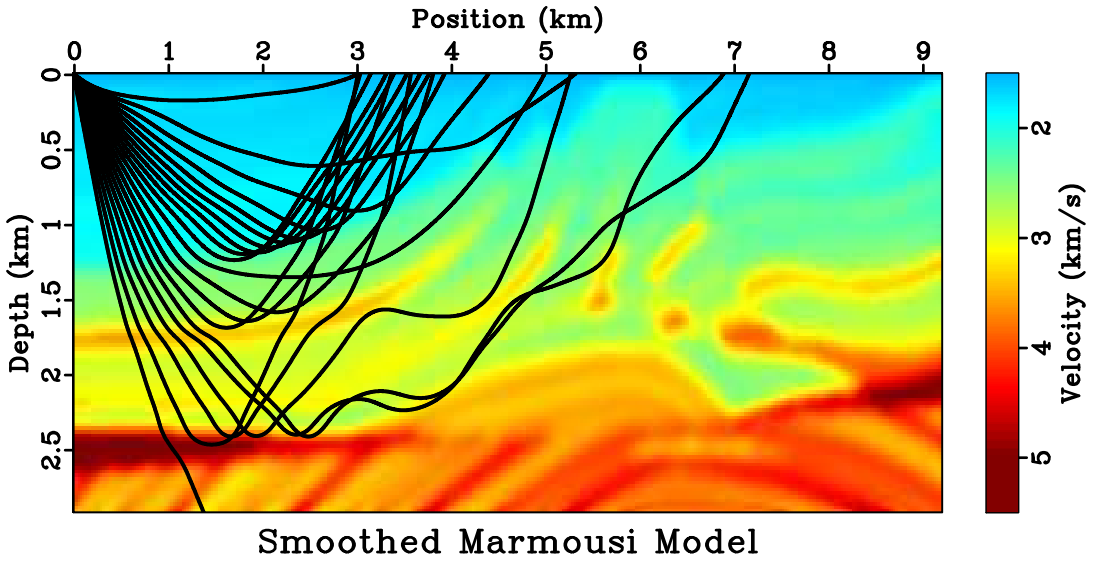


Figure 8: A smoothed Marmousi model overlaid with rays traced from a shot at $(0, 0)$ km. Because of velocity variations, multi-pathing is common in this model, especially at large offsets. dsrtomo/causal marmsmooth

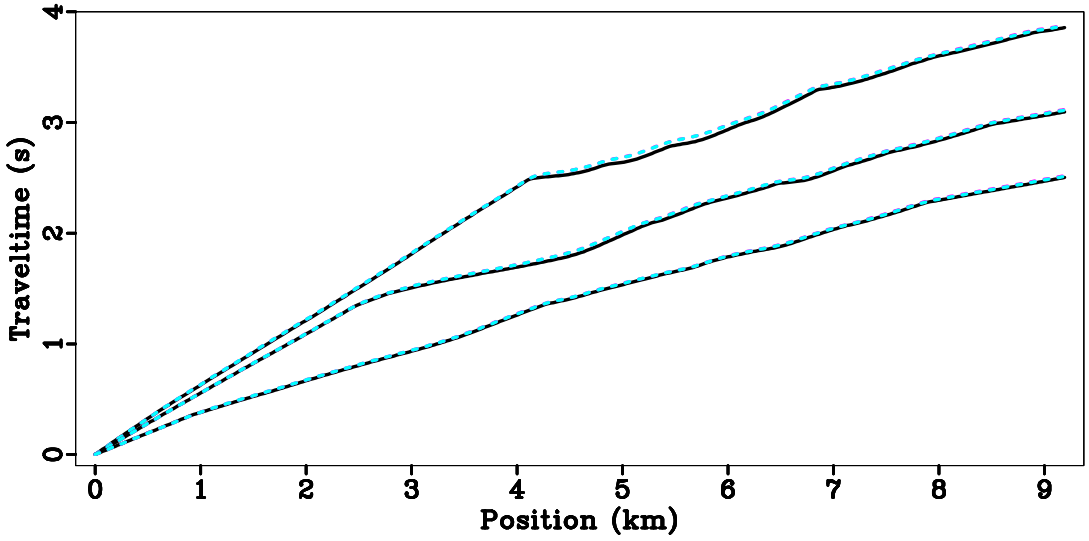


Figure 9: DSR FMM with non-causal branches. The solid black lines are reference values. There are two groups of dashed lines, both from DSR FMM but one with the optional search process turned-on and the other without. The differences between them are negligible and hardly visible. dsrtomo/causal causal

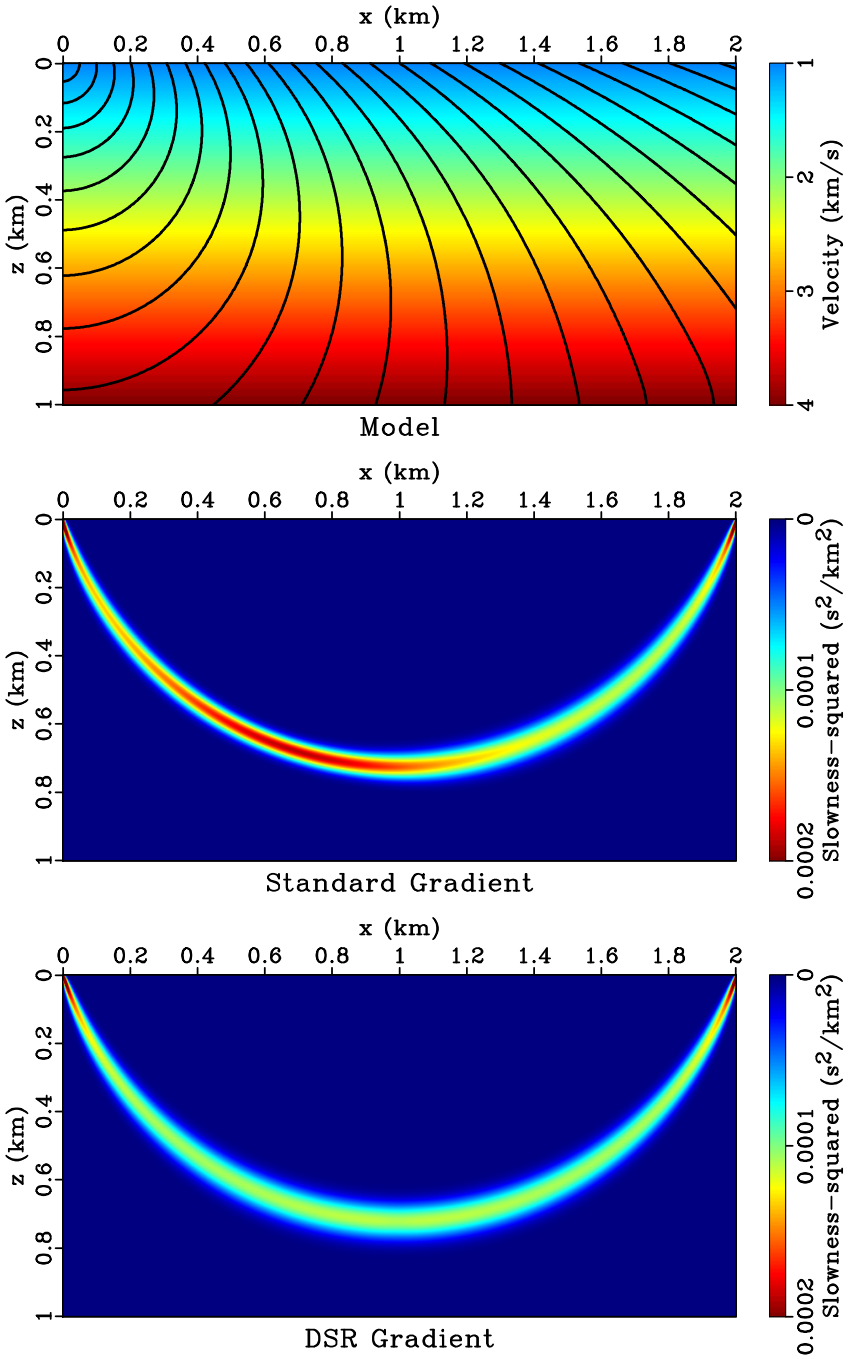


Figure 10: (Top) model overlaid with traveltime contours of a source at $(0, 0)$ km and sensitivity kernels of (middle) the standard tomography and (bottom) the DSR tomography. dsrtomo/hessian grad

eral undesired structures. For example, a near-horizontal structure with a velocity of around 2.75 km/s at location $(0.85, 4.8)$ km is false. It indicates the presence of a local minimum that has trapped the standard tomography. In practice, it is helpful to tune the inversion parameters so that the standard tomography takes more iterations with a gradually reducing regularization. The inversion parameters are usually empirical and hard to control. Our analysis in preceeding sections suggests that part of the role of regularization is to deal with conflicting information between shots. In contrast, we find DSR tomography less dependent on regularization and hence more robust.

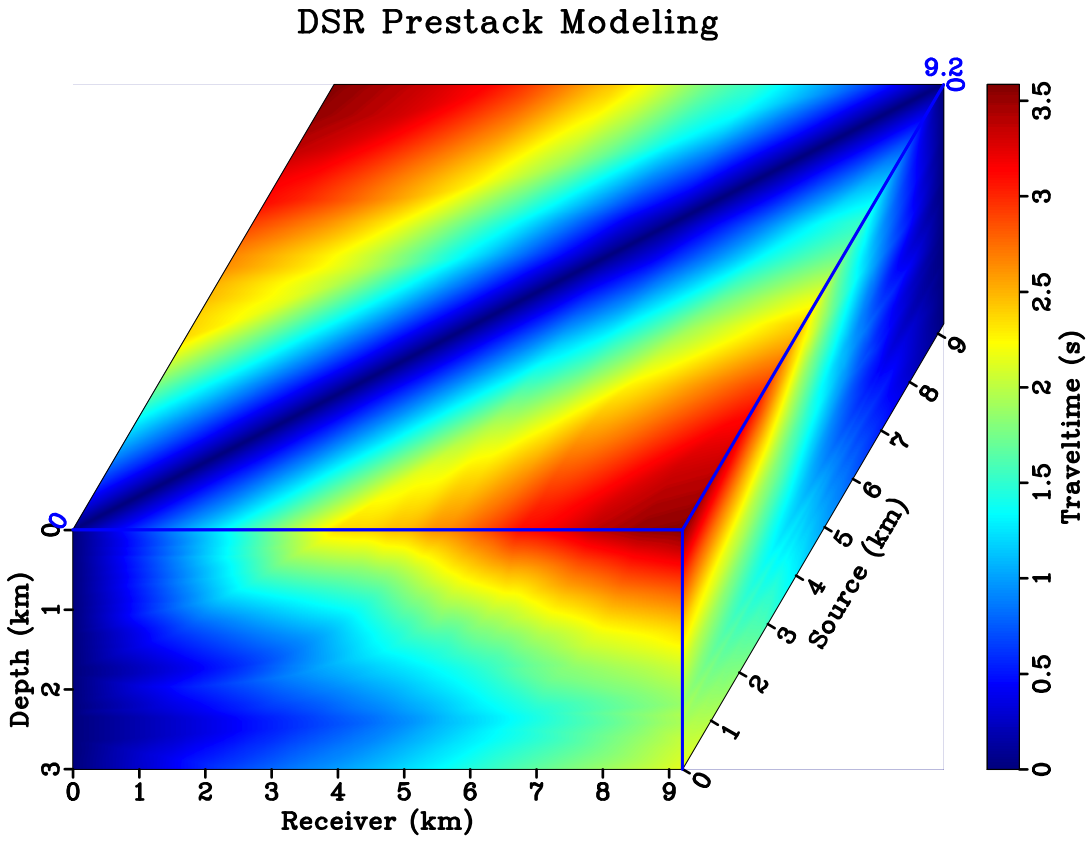


Figure 11: DSR first-break traveltimes in the Marmousi model. The original model is decimated by 2 in both vertical and lateral directions, such that $nz = 376$, $nx = 1151$ and $\Delta = \delta = 8$ m. dsrtomo/marm data

The advantage of DSR tomography becomes more significant in the presence of noise in the input data. We generate random noise of normal distribution with zero mean and a range between ±600 ms, then threshold the result with a minimum absolute value of 250 ms. This is to mimic the spiky errors in first-breaks estimated from an automatic picker. After adding noise to the data, we run inversions with the same parameters as in Figures 13 and 14. Figures 15 and 16 show the convergence history and inverted models. Again, the standard tomography seems to provide a model with higher resolution, but a close examination reveals that many small scale

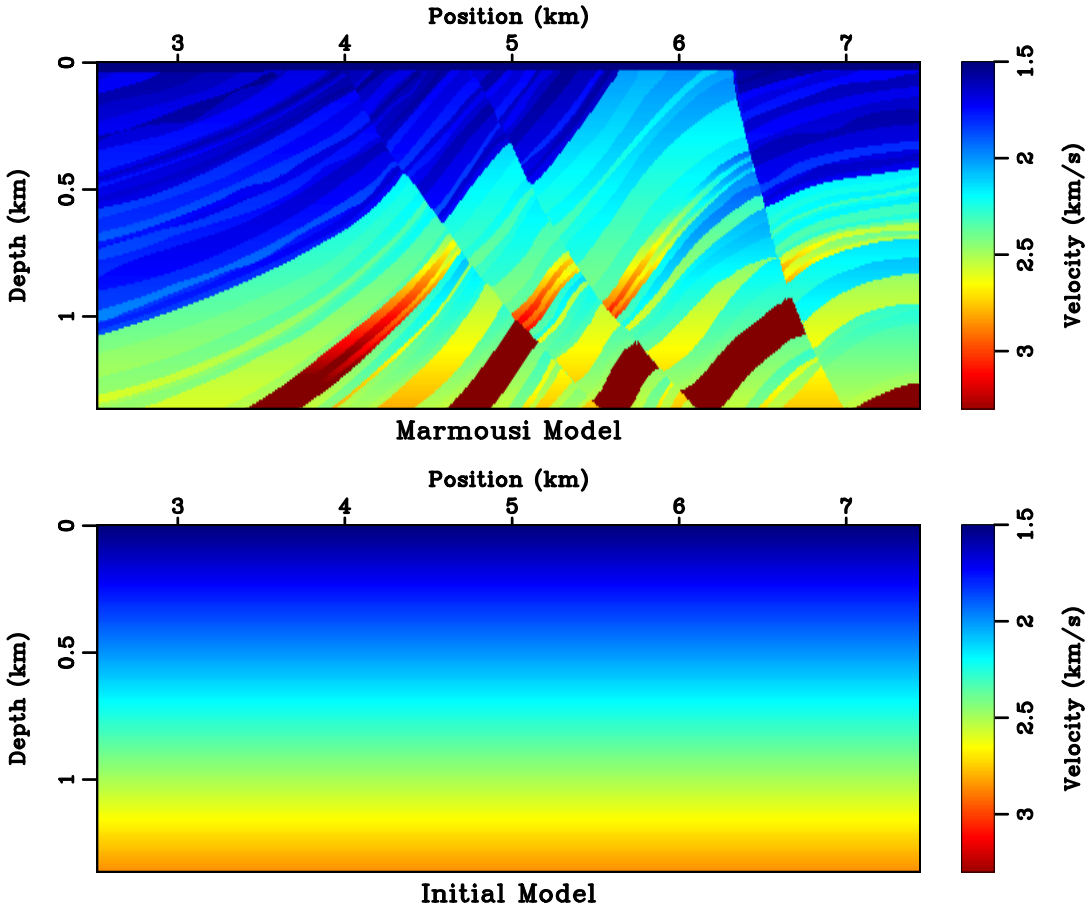


Figure 12: (Top) a zoom-in of Marmousi model and (bottom) the initial model for tomography. dsrtomo/marm marm

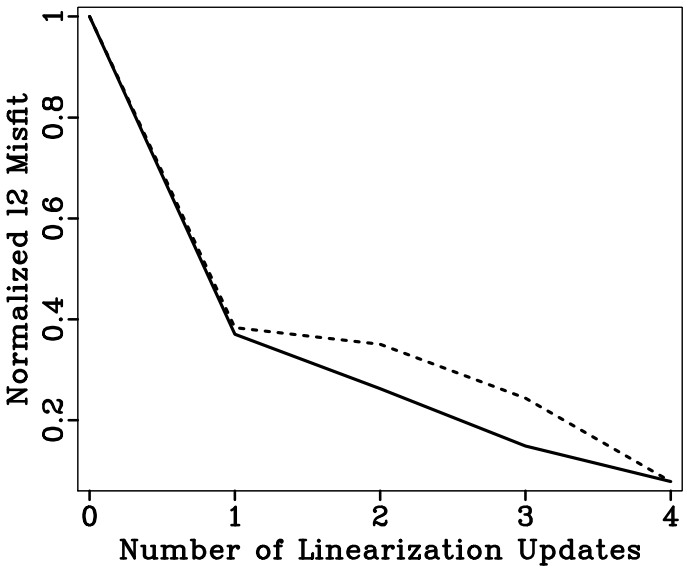


Figure 13: Convergency history of DSR tomography (solid) and standard tomography (dashed). There is no noticeable improvement on misfit after the fourth update. dsrtomo/marm conv

details are in fact non-physical. On the other hand, DSR tomography suffers much less from the added noise. Adopting a $l1$ norm in objective function 5 can improve the inversion, especially for standard tomography. However, it also raises the difficulty in selecting appropriate inversion parameters.

DISCUSSION

There are three main issues in the DSR tomography. The first issue comes from a large dimensionality of the prestack space, which results in a considerable computational domain size after discretization. The memory consumption becomes an immediate problem for 3-D models, where the prestack traveltime belongs to a 5-D space and may require distributed storage.

The second issue is related to the computational cost. The FMM DSR we have introduced in this paper has a computational complexity of $O(N \log N)$, where N is the total number of grid points after discretization, $N = nz \times nx^2$. The $\log N$ factor arises in the priority queue used in FMM for keeping track of expanding wave-fronts. Some existing works could accelerate FMM to an $O(N)$ complexity and may be applicable to the DSR eikonal equation (19; 41). A number of other fast methods developed for the eikonal equation might be similarly applicable to the DSR eikonal equation. These include fast sweeping (43), hybrid two-scale marching-sweeping methods and various label-correcting algorithms (see (6) and references therein).

The last issue is possible parallelization of the proposed method. Our current

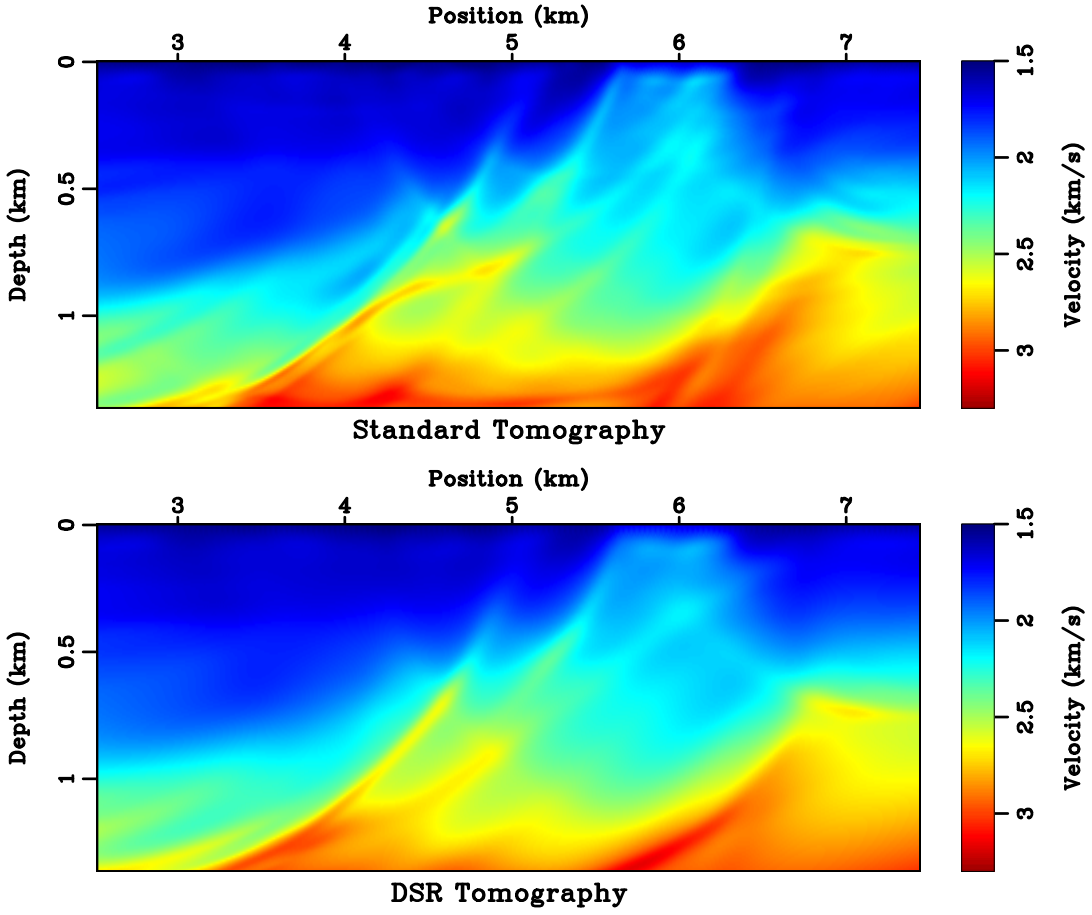


Figure 14: Inverted model of (top) standard tomography and (bottom) DSR tomography. Compare with Figure 12. dsrtomo/marm tomo

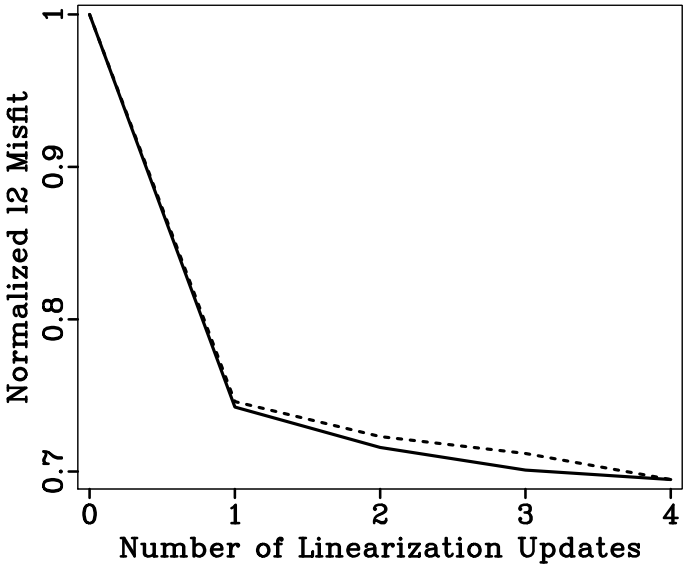


Figure 15: Inversion with noisy data. Convergency history of DSR tomography (solid) and standard tomography (dashed). No significant decrese in misfit appears after the fourth update. dsrtomo/marm nconv

implementation of the FMM DSR tomography algorithm is sequential, while the traditional tomography could be parallelized among different shots. However, we notice that the DSR eikonal equation has a plane-wave source, therefore a distributed wave-front propagating at the beginning followed by a subdomain merging is possible. A number of parllelizable algorithms for the eikonal equation have been developed (44; 18; 40; 7; 11). Extending these methods to the DSR eikonal equation would be the first step in parallelizing DSR tomography.

CONCLUSIONS

We propose to use the DSR eikonal equation for the first-break traveltime tomography. The proposed method relies on an efficient DSR solver, which is realized by a version of the fast-marching method based on an implicitly causal discretization. Since the DSR eikonal equation allows changing of source position along the ray-path, its linearization results in a tomographic inversion that naturally handles possible conflicting information between different shots. Our numerical tests show that, compared to the traditional tomography with a shot-indexed eikonal equation, the DSR tomography converges faster and is more robust. Its benefits may be particularly significant in the presence of noise in the data.

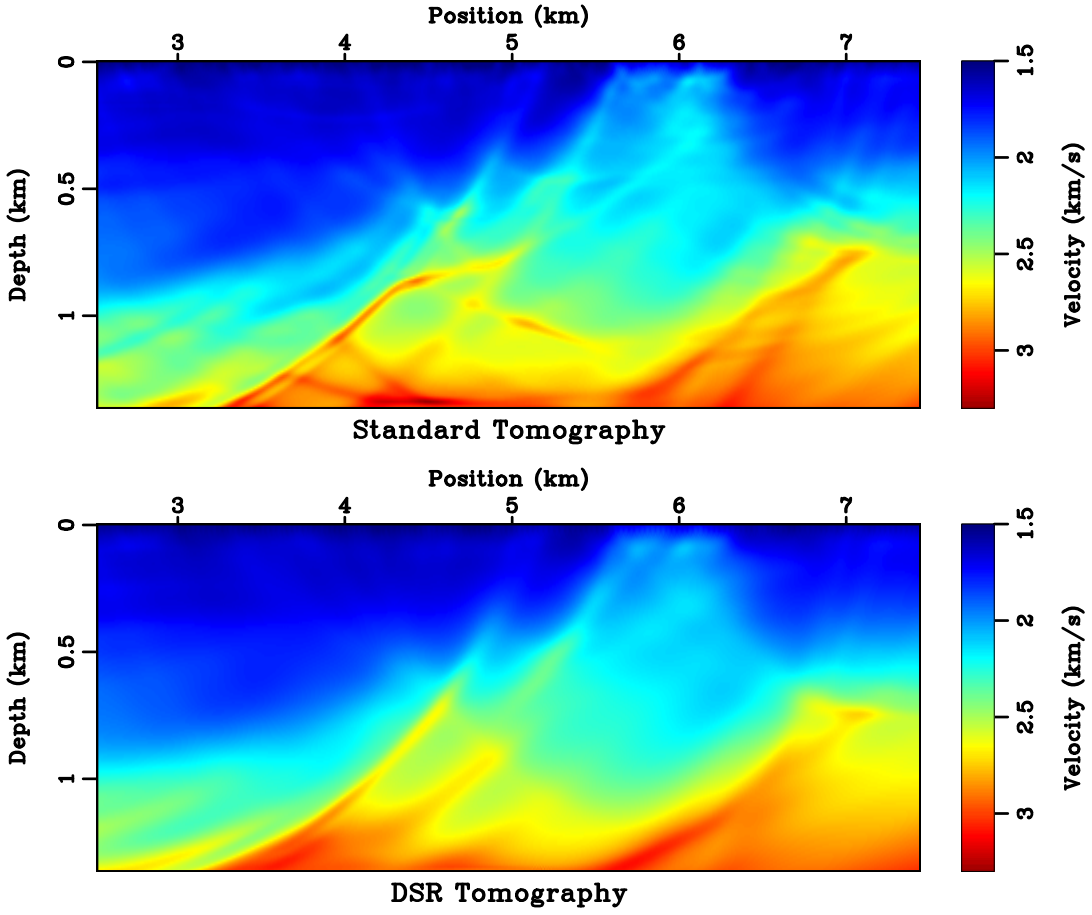


Figure 16: Inversion with noisy data. Inverted model of (top) standard tomography and (bottom) DSR tomography. Compare with Figure 14. dsrtomo/marm ntomo

ACKNOWLEDGMENTS

This research was supported in part by Saudi Aramco. The second author's work was also partially supported by the NSF grant DMS-1016150. We thank Anton Duchkov, Mauricio Sacchi, Jörg Schleicher, Aldo Vesnaver, and four anonymous reviewers for valuable comments and suggestions. We thank Tariq Alkhalifah and Tim Keho for useful discussions. This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A: CAUSAL DISCRETIZATION OF DSR EIKONAL EQUATION

To simplify the analysis, we consider first the DSR branch as shown in Figure 1 and described by equation 1. We assume a rectangular 2-D velocity model $v(z, x)$ and thus a cubic 3-D prestack volume $T(z, r, s)$ with r and s axes having the same dimension as x. After an Eulerian discretization of both v and T, we denote the grid spacing in z as Δ, and in x, r and s as δ.

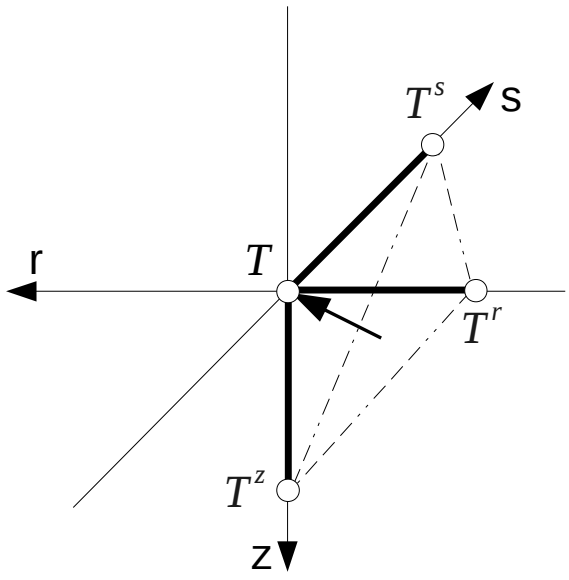


Figure A-1: An implicit discretization scheme. The arrow indicates a DSR characteristic. Its root is located in the simplex $T^z T^r T^s$. dsrtomo/. update1

In Figure A-1, we study the traveltime at grid point $\mathbf{y} = (z, r, s)$ and its relationship with neighboring grid points $\mathbf{y}^z = (z + \Delta, r, s)$, $\mathbf{y}^r = (z, r - \delta, s)$ and

$\mathbf{y}^s = (z, r, s+\delta)$ with a semi-Lagrangian scheme. According to the geometry in Figure 1, in the (z, r, s) space the DSR characteristic (13) is straddled by $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$.

In order to compute T, we could continue along this characteristic up until its intersection with the simplex $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$. Suppose the intersection point is $\tilde{\mathbf{y}} = (\tilde{z}, \tilde{r}, \tilde{s})$ and α_i's are its barycentric coordinates, i.e.

$$\alpha_z, \alpha_r, \alpha_s \in [0,1]; \;\; \alpha_z + \alpha_r + \alpha_s = 1; \;\; \tilde{\mathbf{y}} = \alpha_z\mathbf{y}^z + \alpha_r\mathbf{y}^r + \alpha_s\mathbf{y}^s \; . \tag{A-1}$$

This leads to the following discretization:

$$T \equiv T(\mathbf{y}) = \min_{\tilde{\mathbf{y}} \in \mathbf{y}^z\mathbf{y}^r\mathbf{y}^s} \left\{ T(\tilde{\mathbf{y}}) + \frac{\sqrt{(z-\tilde{z})^2 + (r-\tilde{r})^2}}{v(z,r)} + \frac{\sqrt{(z-\tilde{z})^2 + (s-\tilde{s})^2}}{v(z,s)} \right\} . \tag{A-2}$$

Here we further assume that $v(z,r)$ and $v(z,s)$ are locally constant and that ray-segments between z and $z+\Delta$ are well approximated by straight lines. This also means that a linear interpolation in T within the simplex $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$ is exact, i.e. $T(\tilde{\mathbf{y}}) = \alpha_z T^z + \alpha_r T^r + \alpha_s T^s$, where $T^i = T(\mathbf{y}^i)$ for $i = z, r, s$. The minimization over all possible intersection points in equation A-2 guarantees a first-arrival traveltime.

Defining the ratio in grid spacing as $\mu \equiv \Delta/\delta$ and denoting $v_r = v(z,r)$ and $v_s = v(z,s)$, equation A-2 can be re-written with the barycentric coordinates in A-1 as

$$T = \min_{(\alpha_z, \alpha_r, \alpha_s)} \left\{ (\alpha_z T^z + \alpha_r T^r + \alpha_s T^s) + \frac{\delta}{v_r}\sqrt{\alpha_r^2 + \mu^2\alpha_z^2} + \frac{\delta}{v_s}\sqrt{\alpha_s^2 + \mu^2\alpha_z^2} \right\} . \tag{A-3}$$

Figure A-1 is based on a particular direction of the diving wave: rightward from the source and leftward from the receiver, as in Figure 1. This yields the above positions of $\mathbf{y}^s$ and $\mathbf{y}^r$, and the formula A-3 becomes an update from the $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$ quadrant. Since in general the direction of a diving wave is not a priori known, we compute one such update from each of the lower quadrants and take the smallest amongst them as a value of T.

To explore the causal properties of equation A-3, we first assume that the minimum is attained at some $\tilde{\mathbf{y}}^* = \xi_z\mathbf{y}^z + \xi_r\mathbf{y}^r + \xi_s\mathbf{y}^s$ such that $\xi_i > 0$ for $i = z, r, s$. From the Kuhn-Tucker optimality conditions (20), there exists a Lagrange multiplier λ such that

$$\lambda = T^z + \delta\left(\frac{\mu^2\xi_z}{v_r\sqrt{\xi_r^2 + \mu^2\xi_z^2}} + \frac{\mu^2\xi_z}{v_s\sqrt{\xi_s^2 + \mu^2\xi_z^2}} \right) \; ; \tag{A-4}$$

$$\lambda = T^i + \delta\left(\frac{\xi_i}{v_i\sqrt{\xi_i^2 + \mu^2\xi_z^2}} \right) \; ; \;\; i = r, s \; . \tag{A-5}$$

Taking a linear combination of A-4 and A-5 to match the right-hand side of A-3, we find that $\lambda = T$ and thus

$$T - T^z = \delta\left(\frac{\mu^2\xi_z}{v_r\sqrt{\xi_r^2 + \mu^2\xi_z^2}} + \frac{\mu^2\xi_z}{v_s\sqrt{\xi_s^2 + \mu^2\xi_z^2}} \right) > 0 \; ; \tag{A-6}$$

$$T - T^i = \delta \left(\frac{\xi_i}{v_i \sqrt{\xi_i^2 + \mu^2 \xi_z^2}} \right) > 0 \; ; \;\; i = r, s \; . \tag{A-7}$$

This means that if T defined by equation A-3 depends on T^i then $T > T^i$ for $i = z, r, s$, or

$$T > \max(T^z, T^r, T^s) \tag{A-8}$$

and a Dijkstra-like method (12) is applicable to solve the discretized system.

A direct substitution from equations A-6 and A-7 results in

$$\frac{T - T^z}{\Delta} = \sqrt{\frac{1}{v_r^2} - \left(\frac{T - T^r}{\delta} \right)^2} + \sqrt{\frac{1}{v_s^2} - \left(\frac{T - T^s}{\delta} \right)^2} \; . \tag{A-9}$$

If T^z, T^r and T^s are known, then T can be recovered by solving the 4th degree polynomial equation A-9 and choosing the smallest real root that satisfies condition A-8. This gives a three-sided update at T. The discretization is *implicitly causal* and provides unconditional consistency and convergence.

If there is no real root or none of the real roots satisfy A-8, the minimizer $\tilde{\mathbf{y}}$ can not lie in the interior of simplex $\mathbf{y}^z \mathbf{y}^r \mathbf{y}^s$ and at least one of the ξ_is must be zero. If $\xi_z = 0$, it is easy to show that one of the other barycentric coordinates is also zero and equation A-3 simplifies to

$$T = \min \left(T^r + \frac{\delta}{v_r}, T^s + \frac{\delta}{v_s} \right) \; , \tag{A-10}$$

which is a causal discretization of the DSR singularity in equation 3. On the other hand, if $\xi_z \neq 0$ and $\xi_r \neq 0$ but $\xi_s = 0$, i.e. the slowness vector at s is vertical, then

$$T = (\xi_z T^z + \xi_r T^r) + \frac{\delta}{v_r} \sqrt{\xi_r^2 + \mu^2 \xi_z^2} + \frac{\Delta}{v_s} \xi_z \; . \tag{A-11}$$

A similar Kuhn-Tucker-type argument shows that equation A-11 is also causal: if $\xi_z, \xi_r > 0$, then $T > \max(T^z, T^r)$. In this case, T can be computed by solving

$$\frac{T - T^z}{\Delta} = \sqrt{\frac{1}{v_r^2} - \left(\frac{T - T^r}{\delta} \right)^2} + \frac{1}{v_s} \; . \tag{A-12}$$

Equation A-12 is equivalent to setting $\partial T / \partial s = 0$ in equation 1. Analogously, when $\xi_z \neq 0$ and $\xi_s \neq 0$ but $\xi_r = 0$, we have $\partial T / \partial r = 0$ and

$$\frac{T - T^z}{\Delta} = \frac{1}{v_r} + \sqrt{\frac{1}{v_s^2} - \left(\frac{T - T^s}{\delta} \right)^2} \; , \tag{A-13}$$

with the causal solution satisfying $T > \max(T^z, T^s)$. Equations A-12 and A-13 provide a two-sided update at T. Finally, if $\xi_z \neq 0$ but $\xi_s = 0$ and $\xi_r = 0$, i.e. $\partial T / \partial s = 0$ and $\partial T / \partial r = 0$, we obtain the simplest one-sided update:

$$\frac{T - T^z}{\Delta} = \frac{1}{v_r} + \frac{1}{v_s} \; . \tag{A-14}$$

We note that the one-sided update A-14 could be considered a special case of two-sided updates: if $T = T^r$ (or T^s), then A-14 becomes equivalent to A-12 (or, respectively, A-13). Similarly, the two-sided updates can be viewed as special versions of the three-sided one: e.g., if $T = T^r = \max(T^z, T^r, T^s)$, then A-9 becomes equivalent to A-13. This means that the causal criteria for formulas A-9, A-12 and A-13 can be relaxed (the inequalities do not have to be strict). This relaxation is used to streamline the update strategy in the Numerical Implementation section.

In Figure A-1 and the corresponding semi-Lagrangian discretization A-2, the ray-path is linearly approximated up to its intersection with the simplex $\mathbf{y}^z\mathbf{y}^r\mathbf{y}^s$ at a priori unknown depth $z+\xi_z\Delta$. An alternative *explicit* semi-Lagrangian discretization can be obtained in the spirit of Figure 1 by tracing the ray up to the pre-specified depth $z + \Delta$. In Figure A-2, we consider the DSR characteristic being straddled by $\mathbf{y}^z_*\mathbf{y}^r_*\mathbf{y}^s_*$, where $\mathbf{y}^z_* = (z + \Delta, r, s)$, $\mathbf{y}^r_* = (z + \Delta, r - \delta, s)$ and $\mathbf{y}^s_* = (z + \Delta, r, s + \delta)$. Denoting $\tilde{\mathbf{y}}_* = (z + \Delta, \tilde{r}_*, \tilde{s}_*)$ for the intersection point between DSR characteristic and simplex $\mathbf{y}^z_*\mathbf{y}^r_*\mathbf{y}^s_*$, we obtain the following discretization:

$$T = \min_{\tilde{\mathbf{y}}_* \in \mathbf{y}^z_*\mathbf{y}^r_*\mathbf{y}^s_*} \left\{ T(\tilde{\mathbf{y}}_*) + \frac{\sqrt{\Delta^2 + (r - \tilde{r}_*)^2}}{v(z, r)} + \frac{\sqrt{\Delta^2 + (s - \tilde{s}_*)^2}}{v(z, s)} \right\} . \tag{A-15}$$

One could perform the same analysis of A-3 through A-9 to equation A-15. For the sake of brevity, we omit the derivation and show the resulting explicit discretization scheme:

$$\frac{T - T^z_*}{\Delta} = \sqrt{\frac{1}{v_r^2} - \left(\frac{T^z_* - T^r_*}{\delta}\right)^2} + \sqrt{\frac{1}{v_s^2} - \left(\frac{T^z_* - T^s_*}{\delta}\right)^2} , \tag{A-16}$$

where $T^i_* = T(\mathbf{y}^i_*)$ for $i = z, r, s$. More generally, to account for various possible directions of the diving wave, we can set $T^r_* = \min\left(T(z + \Delta, r - \delta, s), T(z + \Delta, r + \delta, s)\right)$ and $T^s_* = \min\left(T(z + \Delta, r, s - \delta), T(z + \Delta, r, s + \delta)\right)$.

Compared with equation A-9, equation A-16 does not require solving a polynomial equation. Moreover, T depends only on values in lower z-slices, which means that the system of equations can be solved in a single sweep in the $-z$ direction. Unfortunately, despite this efficiency on a fixed grid, the explicit discretization has a major disadvantage stemming from the requirement that the characteristic should be straddled by $y^z_* y^r_* y^s_*$. This imposes an upper bound on μ based on the slope of the diving wave. Moreover, since every diving ray is horizontal at its lowest point, the convergence is possible only if $\mu \to 0$ under mesh refinement. In practice, this means that the results are meaningful only if Δ is significantly smaller than δ. We note that restrictive stability conditions also arise for time-dependent Hamilton-Jacobi equations of optimal control, where sufficiently strong inhomogenieties can make nonlinear/implicit schemes preferable to the usual linear/explicit approach (39).

The above analysis also applies to the first branch of the DSR eikonal equation in Figure 2. However, in the discretized (z, r, s) domain, the slowness vectors at s and r are always aligned in the z direction, either upward or downward. For this reason, there is no DSR characteristic that accounts for the second and third scenarios. We

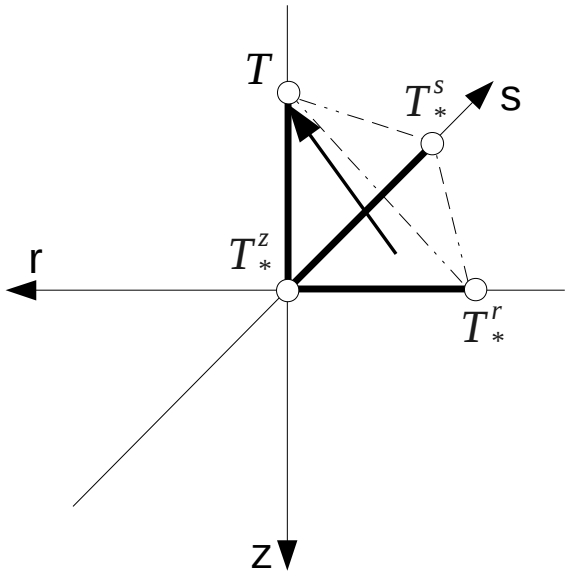


Figure A-2: An explicit discretization scheme. Compare with Figure A-1. The arrow again depicts a DSR characteristic with its root confined in the simplex $T^z_* T^r_* T^s_*$. dsrtomo/. update2

will refer to the first and last branches in Figure 2 as *causal branches* of DSR eikonal equation, and the left-over two as *non-causal branches*.

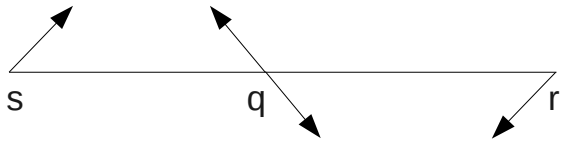


Figure A-3: When slowness vectors at s and r are pointing in the opposite directions, the ray-path must intersect with line $s - r$ at certain point q. dsrtomo/. search

Note that when the slowness vectors at s and r are pointing in opposite directions, there must be at least one intersection of the ray-path with the z depth level in-between. As shown in Figure A-3, ray segments between these intersections fall into the category of causal branches. Thus a search process for the intersections is sufficient in recovering the non-causal branches during forward modeling. Moreover, because we are interested in first-breaks only, the minimum traveltime requirement allows us to search for only one intersection, such as q denoted in Figure A-3:

$$T(z,r,s) = \min_{q\in(s,r)} \{T(z,q,s) + T(z,r,q)\} \ . \tag{A-17}$$

Other possible intersections in intervals (s,q) and (q,r) have already been recovered when computing $T(z,q,s)$ and $T(z,r,q)$, as long as we enable the intersection searching from the beginning of forward modeling. The traveltime of non-causal branches from equation A-17 is then compared with that from causal branches, and the smaller one should be kept.

Unfortunately, this search routine induces considerable computational cost. Moreover, we note that, under a dominant diving waves assumption, the first DSR branch, despite being causal, becomes useless if A-17 is turned-off.

APPENDIX B: FRECHÉT DERIVATIVE OF DSR TOMOGRAPHY

To derive the Frechét derivative, we start from equations 13 and 14. Applying $\partial/\partial w_s$ to both sides of equation 13 results in

$$\begin{aligned} D_z \frac{\partial t^{dsr}}{\partial w_s} &= -\frac{1}{2\sqrt{w_s - D_s t^{dsr}\cdot D_s t^{dsr}}} \\ &+ \left(\frac{D_s t^{dsr}\cdot D_s}{\sqrt{w_s - D_s t^{dsr}\cdot D_s t^{dsr}}} + \frac{D_r t^{dsr}\cdot D_r}{\sqrt{w_r - D_r t^{dsr}\cdot D_r t^{dsr}}} \right) \frac{\partial t^{dsr}}{\partial w_s} \ . \end{aligned} \tag{B-1}$$

Analogously

$$D_z \frac{\partial t^{dsr}}{\partial w_r} = -\frac{1}{2\sqrt{w_r - D_r t^{dsr} \cdot D_r t^{dsr}}} + \left(\frac{D_s t^{dsr} \cdot D_s}{\sqrt{w_s - D_s t^{dsr} \cdot D_s t^{dsr}}} + \frac{D_r t^{dsr} \cdot D_r}{\sqrt{w_r - D_r t^{dsr} \cdot D_r t^{dsr}}} \right) \frac{\partial t^{dsr}}{\partial w_r} . \quad \text{(B-2)}$$

Inserting equations B-1 and B-2 into 14 and regrouping the terms, we prove equation 15

$$J^{dsr} = B^{-1}(C_s + C_r) , \quad \text{(B-3)}$$

where

$$B = D_z - \left(\frac{D_s t^{dsr} \cdot D_s}{\sqrt{w_s - D_s t^{dsr} \cdot D_s t^{dsr}}} \right) - \left(\frac{D_r t^{dsr} \cdot D_r}{\sqrt{w_r - D_r t^{dsr} \cdot D_r t^{dsr}}} \right) , \quad \text{(B-4)}$$

and

$$C_s = -\frac{1}{2\sqrt{w_s - D_s t^{dsr} \cdot D_s t^{dsr}}} \frac{\partial w_s}{\partial w} ; \quad \text{(B-5)}$$

$$C_r = -\frac{1}{2\sqrt{w_r - D_r t^{dsr} \cdot D_r t^{dsr}}} \frac{\partial w_r}{\partial w} . \quad \text{(B-6)}$$

At the singularity of DSR eikonal equation, the operators B, C_s and C_r take simpler forms and can be derived directly from equation 3.

APPENDIX C: ADJOINT-STATE TOMOGRAPHY WITH UPWIND FINITE-DIFFERENCES

Following Appendix A, we let $T_i^{j,k}$ in the DSR case be the traveltime at vertex (z_i, r_j, s_k) and approximate D_z in equation 8 by a one-sided finite-difference

$$D_z^{\pm} T_i^{j,k} = \pm \frac{T_{i\pm1}^{j,k} - T_i^{j,k}}{\Delta} , \quad \text{(C-1)}$$

where the $\pm$ sign corresponds to the two neighbors of $T_i^{j,k}$ in z direction. An upwind scheme (15) picks the sign by

$$D_z T_i^{j,k} = \max \left(D_z^- T_i^{j,k}, -D_z^+ T_i^{j,k}, 0 \right) . \quad \text{(C-2)}$$

The above strategy can be applied to D_r and D_s straightforwardly. For the shot-indexed eikonal equation 9, we approximate D_x with the same upwind method while T in this case is indexed for z and x.

For a Cartesian ordering of the discretized T, i.e. vector $\mathbf{t}$, the discretized operators $D_m T \cdot D_m$ with $m = z, x, r, s$ are matrices. Thanks to upwind finite-differences,

these matrices are sparse and contain only two non-zero entries per row. For instance, suppose $T_i^{j,k}$ has its upwind neighbor in z at $T_{i-1}^{j,k}$, then

$$D_z T \cdot D_z = \begin{bmatrix} \ddots & & & & \\ & \ddots & & & \\ & & \ddots & & \\ & -\kappa_z & & \kappa_z & \\ & & & & \ddots \end{bmatrix} , \tag{C-3}$$

where

$$\kappa_z \equiv \frac{D_z T_i^{j,k}}{\Delta} = \frac{T_i^{j,k} - T_{i-1}^{j,k}}{\Delta^2} . \tag{C-4}$$

Definitions of κ_r, κ_s and κ_x follow their upwind approximations, respectively. In equation C-3, $\pm\kappa_z$ are located in the same row as that of $T_i^{j,k}$ in $\mathbf{t}$. While κ_z sits on the diagonal, $-\kappa_z$ has a column index equals to the row of $T_{i-1}^{j,k}$ in $\mathbf{t}$. At $T = 0$, there is no upwind neighbor and the corresponding row contains all zeros.

We can sort entries of $\mathbf{t}$ by their values in an increasing order, which equivalently performs column-wise permutations to $D_m T \cdot D_m$. The results are lower triangular matrices. In fact, during FMM forward modeling, such an *upwind ordering* is maintained and updated by the priority queue and thus can be conveniently imported for usage here.

Note that the summation and subtraction of two (or more) $D_m T \cdot D_m$ matrices are still lower triangular. These matrices are also invertible, except for a singularity at $T = 0$ where we may set the entries to be zero. Naturally, the inverted matrices are also lower triangular. One example is the linearized eikonal equation that gives rise to equation 10. Following notation C-4 and assuming the upwind neighbors of T_i^j are T_{i-1}^j and T_i^{j-1}, the linearized equation 9 reads

$$2\kappa_z(\delta T_i^j - \delta T_{i-1}^j) + 2\kappa_x(\delta T_i^j - \delta T_i^{j-1}) = \delta w_i^j . \tag{C-5}$$

After regrouping the terms, we get

$$\delta T_i^j = \frac{2\kappa_z \delta T_{i-1}^j + 2\kappa_x \delta T_i^{j-1} + \delta w_i^j}{2\left(\kappa_z + \kappa_x\right)} . \tag{C-6}$$

Equation C-6 means the inverse of the operator $D_z T \cdot D_z + D_x T \cdot D_x$ does not need to be computed by an explicit matrix inversion. Instead, we can perform its application to a vector by a single sweep based on causal upwind ordering. The same conclusion can be drawn for operator B-4.

Lastly, the adjoint-state calculations implied by equations 12 and 16 multiply the transpose of these inverse matrices with the data residual. The matrix transposition leads to upper triangular matrices. Accordingly, we solve the linear system with anti-causal *downwind ordering* that follows a decrease in values of $\mathbf{t}$.

REFERENCES

[1] Alkhalifah, T., 2011, Prestack traveltime approximations: 81st Annnual International Meeting, SEG, Expanded Abstracts, 3017–3021.

[2] Belonosova, A. V., and A. S. Alekseev, 1974, About one formulation of the inverse kinematic problem of seismics for a two-dimensional continuously heterogeneous medium: Some methods and algorithms for interpretation of geophysical data (in Russian), Nauka, 137–154.

[3] Bergman, B., A. Tryggvason, and C. Juhlin, 2004, High-resolution seismic traveltime tomography incorporating static corrections applied to a till-covered bedrock environment: Geophysics, **69**, 1082–1090.

[4] Bertsekas, D., 1982, Enlarging the region of convergence of Newton's method for constrained optimization: Journal of Optimization Theory and Applications, **36**, 221–251.

[5] Brenders, A. J., and R. G. Pratt, 2007, Efficient waveform tomography for lithospheric imaging: implications for realistic, two-dimensional acquisition geometries and low-frequency data: Geophysical Journal International, **168**, 152–170.

[6] Chacon, A., and A. Vladimirsky, 2012a, Fast two-scale methods for Eikonal equations: SIAM Journal on Scientific Computing, **33**, no. 3, A547–A578.

[7] ———, 2012b, A parallel two-scale methods for Eikonal equations: submitted to SIAM Journal on Scientific Computing.

[8] Chapman, C., 2002, Fundamentals of seismic wave propagation: Cambridge University Press.

[9] Cox, M., 1999, Static corrections for seismic reflection surveys: Society of Exploration Geophysics.

[10] Dessa, J. X., S. Operto, A. Nakanishi, G. Pascal, K. Uhira, and Y. Kaneda, 2004, Deep seismic imaging of the eastern Nankai Trough, Japan, from multifold ocean bottom seismometer data by combined traveltime tomography and prestack depth migration: Journal of Geophysical Research, **109**, B02111.

[11] Detrixhe, M., C. Min, and F. Gibou, 2013, A parallel fast sweeping method for the eikonal equation: Journal of Computational Physics, **237**, 46–55.

[12] Dijkstra, E. W., 1959, A note on two problems in connexion with graphs: Numerische Mathematik, **1**, 269–271.

[13] Duchkov, A., and M. V. de Hoop, 2010, Extended isochron ray in prestack depth (map) migration: Geophysics, **75**, no. 4, S139–S150.

[14] Engl, H. W., M. Hanke, and A. Neubauer, 1996, Regularization of inverse problems: Kluwer Academic Publishers.

[15] Franklin, J. B., and J. M. Harris, 2001, A high-order fast marching scheme for the linearized eikonal equation: Journal of Computational Acoustics, **9**, 1095–1109.

[16] Hestenes, M. R., and E. Stiefel, 1952, Method of conjugate gradients for solving linear systems: Journal of Research of the National Bureau of Standards, **49**, 409–436.

[17] Iversen, E., 2004, The isochron ray in seismic modeling and imaging: Geophysics, **69**, 1053–1070.

[18] Jeong, W., and R. T. Whitaker, 2008, A fast iterative method for eikonal equa-

tions: SIAM Journal on Scientific Computing, **30**, 2512–2534.
[19] Kim, S., 2001, An $O(N)$ level set method for eikonal equations: SIAM Journal on Scientific Computing, **22**, 2178–2193.
[20] Kuhn, H. W., and A. W. Tucker, 1951, Nonlinear programming: Proceedings of 2nd Berkeley Symposium, 481–492.
[21] Lelièvre, P. G., C. G. Farquharson, and C. A. Hurich, 2011, Inversion of first-arrival seismic traveltimes without rays, implemented on unstructured grids: Geophysical Journal International, **185**, 749–763.
[22] Leung, S., and J. Qian, 2006, An adjoint state method for three-dimensional transmission traveltime tomography using first arrivals: Communications in Mathematical Sciences, **4**, 249–266.
[23] Li, S., S. Fomel, and A. Vladimirsky, 2011, Improving wave-equation fidelity of Gaussian beams by solving the complex eikonal equation: 71st Annnual International Meeting, SEG, Expanded Abstracts, 3829–3834.
[24] Marsden, D., 1993, Static corrections—a review: The Leading Edge, **12**, 43–49.
[25] Noble, M., P. Thierry, C. Taillandier, and H. Calandra, 2010, High-performance 3D first-arrival traveltime tomography: The Leading Edge, **29**, 86–93.
[26] Osypov, K., 2000, Robust refraction tomography: 70th Annnual International Meeting, SEG, Expanded Abstracts, 2032–2035.
[27] Paige, C. C., and M. A. Saunders, 1982, LSQR: an algorithm for sparse linear equations and sparse least squares: ACM Transactions on Mathematical Software, **8**, 43–71.
[28] Pei, D., 2009, Three-dimensional traveltime tomography via LSQR with regularization: 79th Annnual International Meeting, SEG, Expanded Abstracts, 4004–4008.
[29] Plessix, R. E., 2006, A review of the adjoint-state method for computing the gradient of a functional with geophysical applications: Geophysical Journal International, **167**, 495–503.
[30] Popovici, A. M., and J. Sethian, 2002, 3-D imaging using higher order fast marching traveltimes: Geophysics, **67**, 604–609.
[31] Rickett, J., and S. Fomel, 1999, A second-order fast marching eikonal solver: Stanford Exploration Project Report, **100**, 287–292.
[32] Sethian, J. A., 1999, Level set methods and fast marching methods: Evolving interfaces in computational geometry, fluid mechanics, computer vision and material sciences: Cambridge University Press.
[33] Sheng, J., A. Leeds, M. Buddensiek, and G. T. Schuster, 2006, Early arrival waveform tomography on near-surface refraction data: Geophysics, **71**, no. 4, U47–U57.
[34] Simmons, J. L., and N. Bernitsas, 1994, Nonlinear inversion of first-arrival times: 64th Annnual International Meeting, SEG, Expanded Abstracts, 992–995.
[35] Stefani, J. P., 1993, Possibilities and limitations of turning ray tomography—a synthetics study: 63rd Annnual International Meeting, SEG, Expanded Abstracts, 610–612.
[36] Taillandier, C., M. Noble, H. Chauris, and H. Calandra, 2009, First-arrival traveltime tomography based on the adjoint-state method: Geophysics, **74**, no. 6,

WCB57–WCB66.

[37] Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: Soviet Mathematics Doklady, 1035–1038.

[38] Virieux, J., and S. Operto, 2009, An overview of full-waveform inversion in exploration geophysics: Geophysics, **74**, no. 6, WCC1–WCC26.

[39] Vladimirsky, A., and C. Zheng, 2013, A fast implicit method for time-dependent Hamilton-Jacobi PDEs: submitted to SIAM Journal on Scientific Computing.

[40] Weber, O., Y. Devir, A. Bronstein, M. Bronstein, and R. Kimmel, 2008, Parallel algorithms for the approximation of distance maps on parametric surfaces: ACM Transactions on Graphics, **27**, no. 4, 104.

[41] Yatziv, L., A. Bartesaghi, and G. Sapiro, 2006, A fast $O(N)$ implementation of the fast marching algorithm: Journal of Computational Physics, **212**, 393–399.

[42] Zelt, C. A., and P. J. Barton, 1998, 3D seismic refraction tomography: A comparison of two methods applied to data from the Faeroe Basin: Journal of Geophysical Research, **103**, 7187–7210.

[43] Zhao, H., 2005, A fast sweeping method for eikonal equations: Mathematics of Computation, **74**, 603–627.

[44] ———, 2007, Parallel implementations of the fast sweeping method: Journal of Computational Mathematics, **25**, 421–429.

[45] Zhu, T., S. Cheadle, A. Petrella, and S. Gray, 2000, First-arrival tomography: method and application: 70th Annnual International Meeting, SEG, Expanded Abstracts, 2028–2031.

[46] Zhu, X., D. P. Sixta, and B. G. Angstman, 1992, Tomostatics: turning-ray tomography + static corrections: The Leading Edge, **11**, 15–23.

Generalized nonhyperbolic moveout approximation

*Sergey Fomel[1] and Alexey Stovas[2]**

ABSTRACT

Reflection moveout approximations are commonly used for velocity analysis, stacking, and time migration. We introduce a novel functional form for approximating the moveout of reflection traveltimes at large offsets. While the classic hyperbolic approximation uses only two parameters (the zero-offset time and the moveout velocity), our form involves five parameters, which can be determined, in a known medium, from zero-offset computations and from tracing one non-zero-offset ray. We call it a generalized approximation because it reduces to some known three-parameter forms (the shifted hyperbola of Malovichko, de Baziliere, and Castle; the Padé approximation of Alkhalifah and Tsvankin; and others) with a particular choice of coefficients. By testing the accuracy of the proposed approximation with analytical and numerical examples, we show that it can bring several-orders-of-magnitude improvement in accuracy at large offsets compared to known analytical approximations, which makes it as good as exact for many practical purposes.

INTRODUCTION

Reflection moveout approximations are commonly used for velocity analysis, stacking, and time migration (31). The reflection traveltime as a function of the source-receiver offset has a well-known hyperbolic form in the case of plane reflectors in homogeneous isotropic (or elliptically anisotropic) overburden. A hyperbolic behavior of the PP moveout is always valid around the zero offset thanks to the source-receiver reciprocity and the first-order Taylor series expansion. However, any deviations from this simple model may cause nonhyperbolic behavior at large offsets (18).

Considerable research effort has been devoted to developing nonhyperbolic moveout approximations in both isotropic and anisotropic media. The work on isotropic approximations goes back to (8), (27), (23), (12), (9), and others. (20) provides a comprehensive review of many different approximations developed for non-hyperbolic moveout in anisotropic (VTI – vertically transversally isotropic) media. A particularly simple "velocity acceleration" model for nonhyperbolic moveout is suggested by (28; 29). (10) approximates nonhyperbolic moveout by expanding it into a sum of basis functions. (13) and (14) build an accurate moveout approximation by using rational interpolation between several rays.

*__e-mail:__ sergey.fomel@beg.utexas.edu

In this paper, we propose a general functional form for nonhyperbolic approximations that can be applied to different kinds of seismic media. The proposed form includes five coefficients as opposed to two coefficients in the classic hyperbolic approximation. In certain cases, the number of coefficients can be reduced. In the case of a homogeneous VTI medium and the "acoustic approximation" of (3), our approximation becomes identical to the one proposed previously by (17). In the general case, determining the optimal coefficients requires tracing of only one non-zero-offset ray.

Using analytical ray-tracing solutions and numerical experiments, we compare the accuracy of our approximation with the accuracy of other known approximations and discover an improvement in accuracy of several orders of magnitude. Potential applications of the new approximation include velocity analysis and time-domain imaging.

NONHYPERBOLIC MOVEOUT APPROXIMATION

Let $t(x)$ represent the reflection traveltime as a function of the source-receiver offset x. We propose the following general form of the moveout approximation:

$$t^2(x) \approx (1-\xi)\,(t_0^2 + a\,x^2) + \xi\,\sqrt{t_0^4 + 2\,b\,t_0^2\,x^2 + c\,x^4}\;. \tag{1}$$

The five parameters a, b, c, ξ, and t_0 describe the moveout behavior. By simple algebraic manipulations, one can also rewrite equation 1 as

$$t^2(x) \approx t_0^2 + \frac{x^2}{v^2} + \frac{A\,x^4}{v^4\left(t_0^2 + B\,\dfrac{x^2}{v^2} + \sqrt{t_0^4 + 2\,B\,t_0^2\,\dfrac{x^2}{v^2} + C\,\dfrac{x^4}{v^4}}\right)}\;, \tag{2}$$

where the new set of parameters A, B, C, v, and t_0 is related to the previous set by the equalities

$$a = \frac{A\,B + B^2 - C}{v^2\,(A + B^2 - C)}\;; \tag{3}$$

$$b = \frac{B}{v^2}\;; \tag{4}$$

$$c = \frac{C}{v^4}\;; \tag{5}$$

$$\xi = \frac{A}{C - B^2}\;. \tag{6}$$

The inverse transform is given by

$$v^2 = \frac{1}{a\,(1-\xi)+b\,\xi}\;; \tag{7}$$

$$A = \frac{\xi\;(c-b^2)}{\left[a\,(1-\xi)+b\,\xi\right]^2}\;; \tag{8}$$

$$B = \frac{b}{a\,(1-\xi)+b\,\xi}\;; \tag{9}$$

$$C = \frac{c}{\left[a\,(1-\xi)+b\,\xi\right]^2}\;. \tag{10}$$

The existence of the nonhyperbolic part in the traveltime approximation 1 and 2 is controlled by parameter A. When A is zero (which implies that $\xi = 0$ or $c = b^2$), approximation 1 is hyperbolic. When both B and C are very large, approximation 2 also reduces to the hyperbolic form.

Connection with other approximations

Equations 1-2 reduce to some well-known approximations with special choices of parameters.

- If $A = 0$, the proposed approximation reduces to the classic hyperbolic form

$$t^2(x) \approx t_0^2 + \frac{x^2}{v^2}\;, \tag{11}$$

  which is a two-parameter approximation.

- The choice of parameters $A = (1-s)/2$; $B = s/2$; $C = 0$ reduces the proposed approximation to the shifted hyperbola (23; 12; 9), which is the following three-parameter approximation:

$$t(x) \approx t_0\,\left(1-\frac{1}{s}\right) + \frac{1}{s}\,\sqrt{t_0^2 + s\,\frac{x^2}{v^2}}\;. \tag{12}$$

- The choice of parameters $A = -4\,\eta$; $B = 1+2\,\eta$; $C = (1+2\,\eta)^2$ reduces approximation 2 to the form proposed by (4) for VTI media, which is the following three-parameter approximation:

$$t^2(x) \approx t_0^2 + \frac{x^2}{v^2} - \frac{2\,\eta\,x^4}{v^4\,\left[t_0^2 + (1+2\,\eta)\,\dfrac{x^2}{v^2}\right]}\;. \tag{13}$$

- The choice of parameters $A = -2\,\gamma\,t_0^2\,v^2$; $B = -A/2$; $C = A^2/4$ reduces approximation 2 to the following three-parameter approximation suggested by (5) and reminiscent of the "velocity acceleration" equation proposed by (28; 29):

$$t^2(x) \approx t_0^2 + \frac{x^2}{v^2\,(1+\gamma\,x^2)}\;. \tag{14}$$

- The choice of parameters $A=(1-s)/2$; $B=1$; $C=2-s$ reduces the proposed approximation to the following three-parameter approximation suggested by (6):

$$t(x) \approx \frac{1}{2}\sqrt{t_0^2+\left(1-\sqrt{s-1}\right)\,\frac{x^2}{v^2}}+\frac{1}{2}\sqrt{t_0^2+\left(1+\sqrt{s-1}\right)\,\frac{x^2}{v^2}}\;. \tag{15}$$

- The choice of parameters $B=0$; $C=2\,A$ reduces the proposed approximation to the following three-parameter approximation also suggested by (6):

$$t^2(x) \approx \frac{t_0^2}{2}+\frac{x^2}{v^2}+\frac{1}{2}\sqrt{t_0^4+\frac{2\,A\,x^4}{v^4}}\;. \tag{16}$$

- The choice of parameters $A=2\,\tan^2\theta$, $B=1-\tan^2\theta$, $C=1/\cos^4\theta$ reduces the proposed approximation to the double-square-root expression

$$\begin{aligned} t(x) &\approx \frac{1}{2}\sqrt{t_0^2+\frac{x\,(x+t_0\,v\,\sin 2\theta)}{v^2\,\cos^2\theta}}+\frac{1}{2}\sqrt{t_0^2+\frac{x\,(x-t_0\,v\,\sin 2\theta)}{v^2\,\cos^2\theta}} \\ &= \frac{\sqrt{z^2+(y+x/2)^2}}{V}+\frac{\sqrt{z^2+(y-x/2)^2}}{V}\;, \end{aligned} \tag{17}$$

where $V=v\,\cos\theta$, $z=(t_0\,V/2)\,\cos\theta$, and $y=(t_0\,V/2)\,\sin\theta$. Equation 17 describes moveout precisely for the case of a diffraction point in a constant velocity medium.

Thus, the proposed approximation encompasses some other known forms but introduces more degrees of freedom for optimal fitting.

General method for parameter selection

Zero-offset ray

The Taylor expansion of approximation 2 around the zero offset

$$t^2(x)=t_0^2+\frac{x^2}{v^2}+\frac{A}{2}\,\frac{x^4}{v^4\,t_0^2}+O(x^6) \tag{18}$$

provides a convenient method for evaluating coefficients t_0, v, and A by matching expansion 18 to the corresponding expansion of the exact traveltime. This is the method used previously for deriving approximations 2 and 12.

In the special case of an isotropic $V(z)$ medium, the coefficients are readily available and reduce to statistical averages of the velocity distribution (8)

$$t_0 = 2\,m_{-1}\;, \tag{19}$$

$$v^2 = \frac{m_1}{m_{-1}}\;, \tag{20}$$

$$A = \frac{1}{2}\left(1-\frac{m_3\,m_{-1}}{m_1^2}\right)\;, \tag{21}$$

where

$$m_k = \int_0^z V^k(\zeta)\, d\zeta$$

Equations 19-21 are easily extensible to the vertical transverse isotropy (VTI) case (22; 7; 2; 30).

Nonzero-offset ray

To determine uniquely the remaining coefficients B and C, we propose to use just one additional ray reflected at a nonzero offset. Suppose that a reflection ray with the ray parameter P arrives at the offset X and traveltime T. Substituting approximation 2 into equations $t(X) = T$ and $dt/dX = P$ and solving for B and C produces the explicit analytical solution

$$B = \frac{t_0^2\,(X - P\,T\,v^2)}{X\,(t_0^2 - T^2 + P\,T\,X)} - \frac{A\,X^2}{X^2 + v^2\,(t_0^2 - T^2)}\,, \tag{22}$$

$$C = \frac{t_0^4\,(X - P\,T\,v^2)^2}{X^2\,(t_0^2 - T^2 + P\,T\,X)^2} + \frac{2\,A\,v^2\,t_0^2}{X^2 + v^2\,(t_0^2 - T^2)}\,. \tag{23}$$

Horizontal ray

If the reference ray happens to be horizontal, both X and T are infinite, and equations 22-23 are not directly applicable. However, one can use the same principle and match two terms for the behavior of the traveltime at infinitely large offsets. If the traveltime behaves as

$$t^2(x) \approx T_\infty^2 + P_\infty^2\, x^2 \tag{24}$$

for x approaching infinity, then, matching the corresponding behavior of approximation 2, we find that

$$B = \frac{t_0^2\,(1 - v^2\,P_\infty^2)}{t_0^2 - T_\infty^2} - \frac{A}{1 - v^2\,P_\infty^2}\,, \tag{25}$$

$$C = \frac{t_0^4\,(1 - v^2\,P_\infty^2)^2}{(t_0^2 - T_\infty^2)^2}\,. \tag{26}$$

ACCURACY TESTS

To illustrate the applicability of the proposed approximation, we try several analytical and numerical models. Using these models, we test the proposed approximation against the hyperbolic approximation 2, the shifted hyperbola approximation 12, and the Alkhalifah-Tsvankin approximation 13.

Analytical examples

Linear velocity and linear sloth

We start with two analytical isotropic three-parameter models: linear velocity model (described in Appendix A) and linear sloth model (described in Appendix B). In both models, it is possible to compute the exact moveout analytically and thus to compare directly the accuracy of different approximations with the exact moveout. We show this comparison in Figures 1 and 2, where the relative absolute approximation error is plotted for different approximations against a large range of the offset-to-depth ratio and the maximum-to-minimum velocity ratio. As evident from the figures, three-parameter approximations (shifted-hyperbola and Alkhalifah-Tsvankin) improve the accuracy of the two-parameter hyperbolic approximation. However, the proposed five-parameter generalized approximation brings a more significant improvement and reduces the error by several orders of magnitude.

Curved reflector in a constant-velocity medium

Our next analytical example is a curved reflector under a constant-velocity overburden. The reflector curvature is one of the possible causes of non-hyperbolic moveout (18). The Taylor expansion around zero offset for the case of a curved reflector has the form of equation 18 with the following set of parameters (15)

$$t_0 = \frac{2\,L}{V}\,, \tag{27}$$

$$v = \frac{V}{\cos\beta}\,, \tag{28}$$

$$A = 2\,\tan^2\beta\,G\,, \tag{29}$$

where L is the length of the normal (zero-offset) ray, V is the true velocity, β is the reflector dip angle at the normal reflection point, $G = K\,L/(1+K\,L)$, and K is the reflector curvature at the normal reflection point.

The two additional parameters B and C depend on the particular shape of the reflector. In the case of a hyperbolic reflector, analyzed in Appendix C, equation 2 happens to be exact. In this case,

$$T_\infty^2 = t_0^2\,\frac{G}{G+\tan^2\beta}\,, \tag{30}$$

$$P_\infty = \frac{1}{V} = \frac{1}{v\,\cos\beta}\,, \tag{31}$$

which, after substitution in equations 25-26, produce

$$B = G - \tan^2\beta\,, \tag{32}$$

$$C = \left(G+\tan^2\beta\right)^2\,. \tag{33}$$

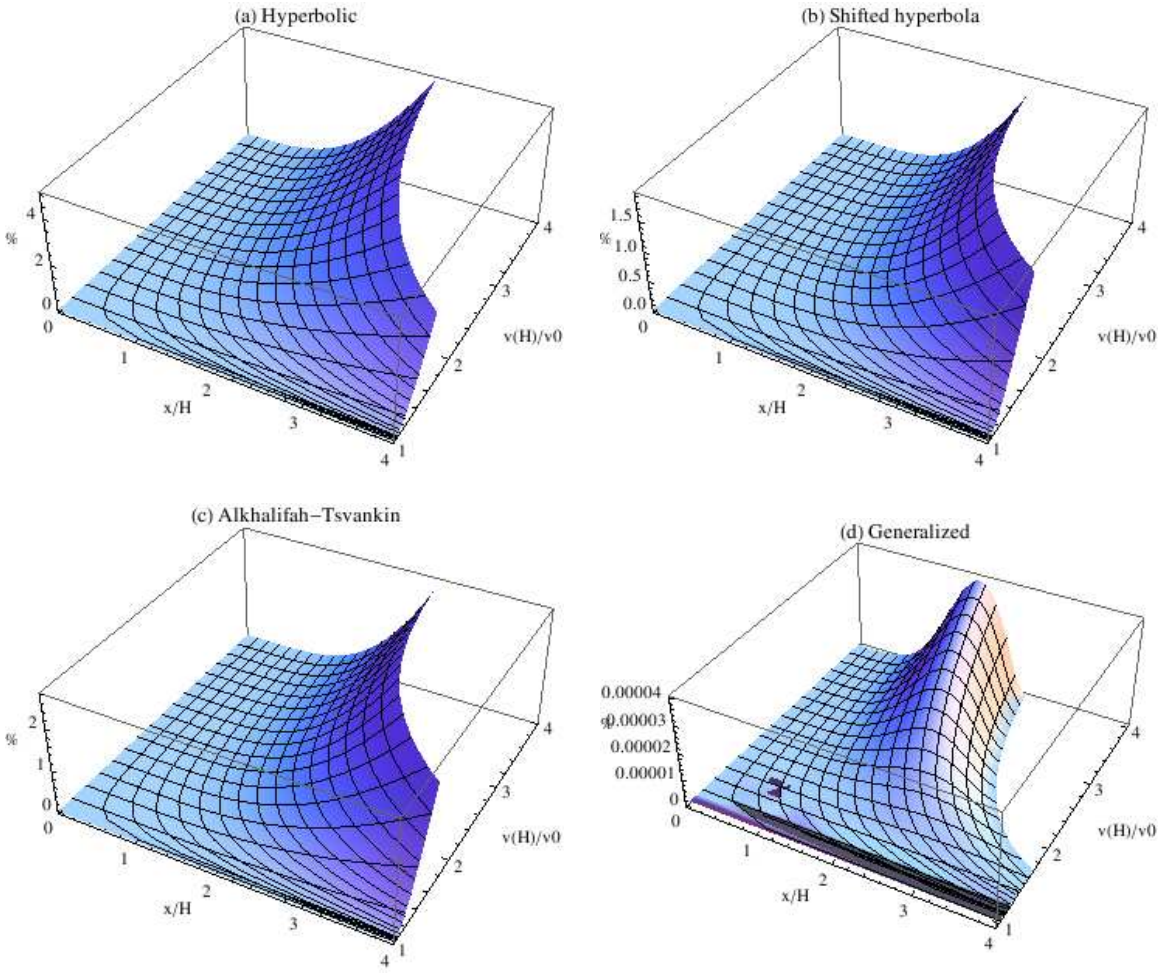


Figure 1: Relative absolute error of different traveltime approximations as a function of velocity contrast and offset/depth ratio for the case of a linear velocity model. (a) Hyperbolic approximation, (b) Shifted hyperbola approximation, (c) Alkhalifah-Tsvankin approximation, (d) Generalized nonhyperbolic approximation. The proposed generalized approximation reduces the maximum approximation error by several orders of magnitude. hyper/Math linvel

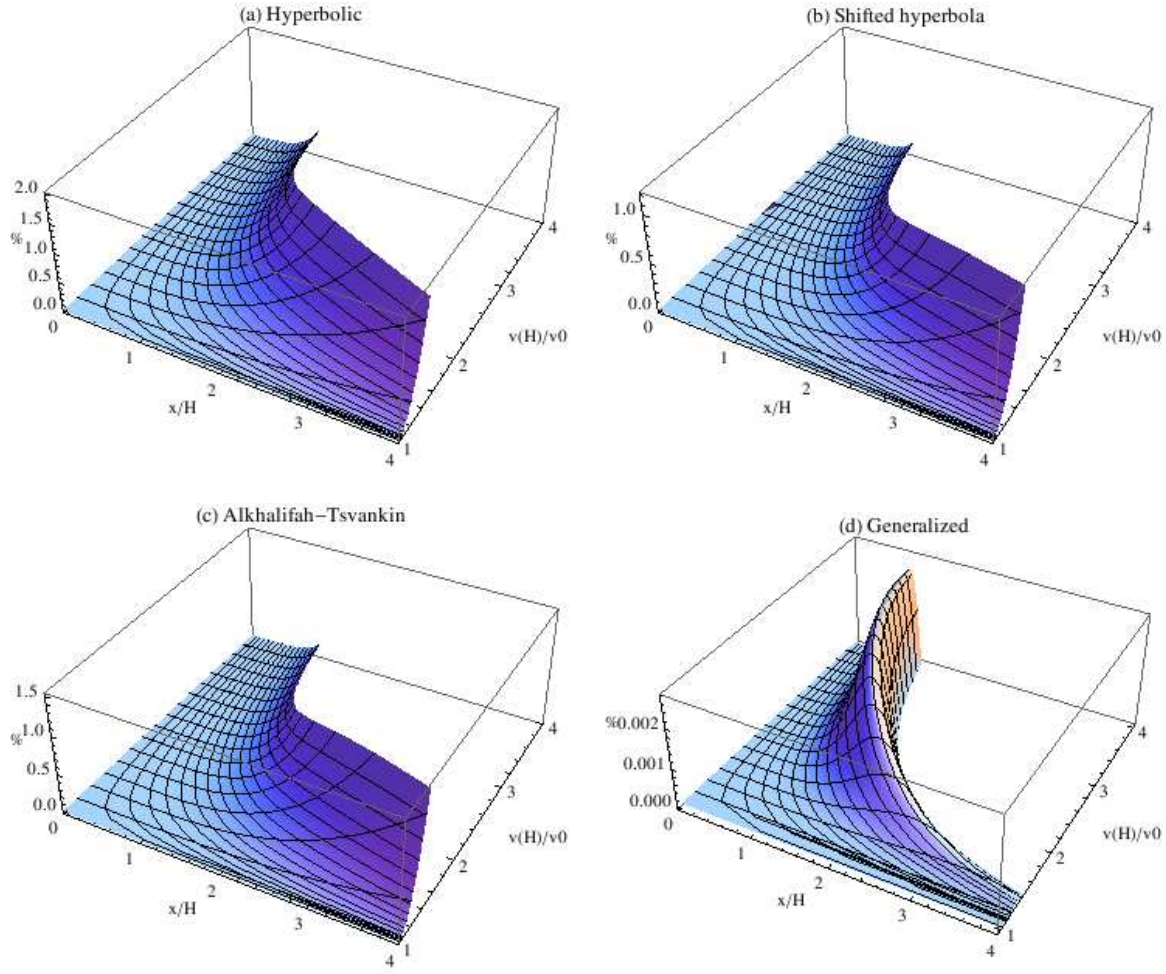


Figure 2: Relative absolute error of different traveltime approximations as a function of velocity contrast and offset/depth ratio for the case of a linear sloth model. (a) Hyperbolic approximation, (b) Shifted hyperbola approximation, (c) Alkhalifah-Tsvankin approximation, (d) Generalized nonhyperbolic approximation. The proposed generalized approximation reduces the maximum approximation error by several orders of magnitude. hyper/Math linsloth

In the special case of a plane (zero curvature) reflector, $G = 0$, and the generalized approximation reduces to a hyperbolic form. In the special case of a diffraction point (infinite curvature), $G = 1$, and the generalized approximation reduces to the double-square-root equation 17. In both of those cases, as well as in the case of a hyperbolic reflector, the generalized approximation is simply exact.

Figure 3 shows a comparison between different approximations for the case of a circular reflector, analyzed in Appendix D. As in the other examples, the proposed five-parameter generalized approximation brings an improvement in accuracy in several orders of magnitude in comparison with the three-parameter approximations.

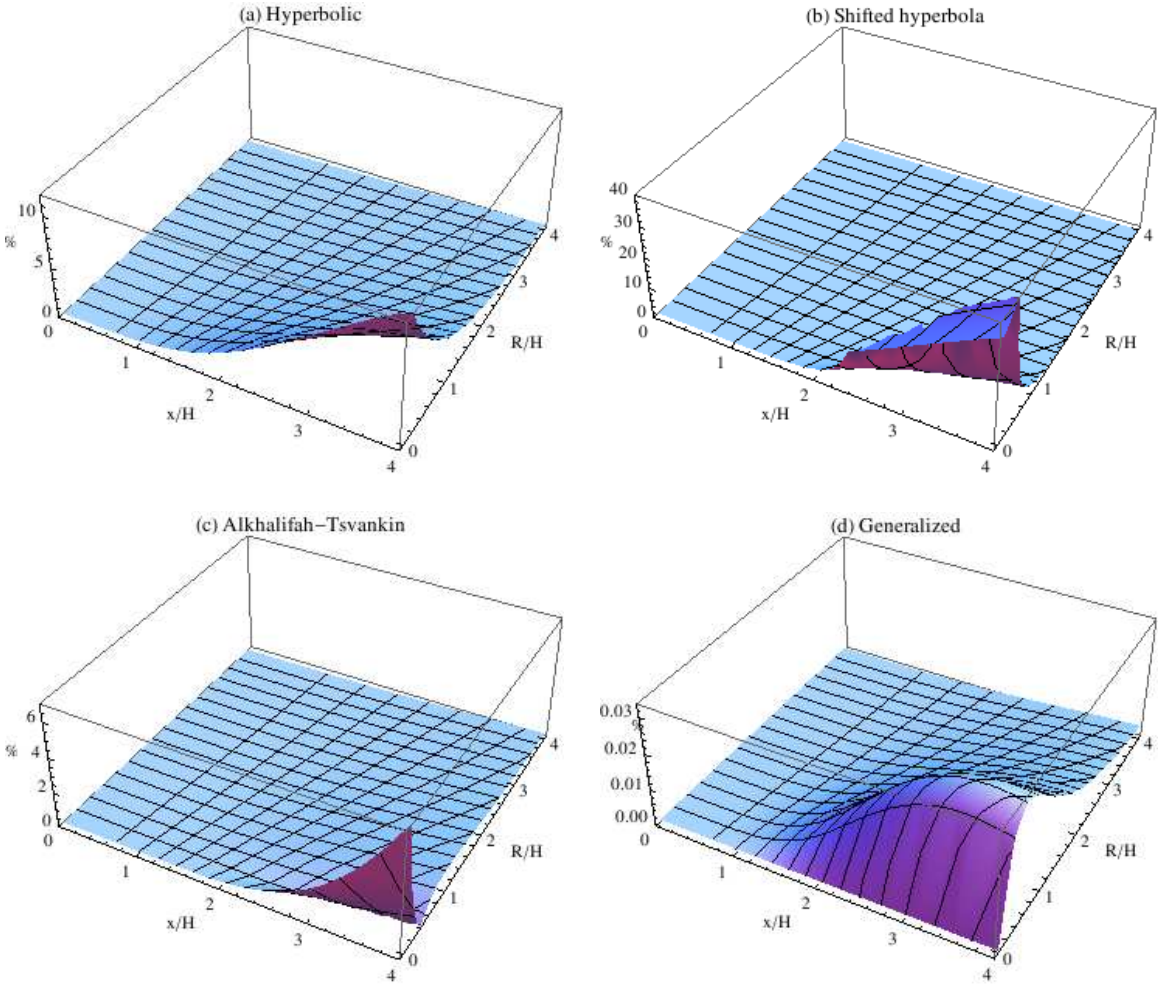


Figure 3: Relative absolute error of different traveltime approximations for the case of a circular reflector as a function of the radius/depth ratio and the offset/depth ratio. The midpoint location with respect to the center of the circle is equal to the depth of the reflector. (a) Hyperbolic approximation, (b) Shifted hyperbola approximation, (c) Alkhalifah-Tsvankin approximation, (d) Generalized nonhyperbolic approximation. The proposed generalized approximation reduces the maximum approximation error by several orders of magnitude. hyper/Math circle

Homogeneous VTI layer

Our next analytical example is a horizontal reflector in a homogeneous VTI (vertically transverse isotropic) medium. As derived in Appendix E, the approximation coefficients, under the assumption of the acoustic approximation of (3), take the form

$$A = -4\,\eta\,, \tag{34}$$

$$B = \frac{1+8\,\eta+8\,\eta^2}{1+2\,\eta}\,, \tag{35}$$

$$C = \frac{1}{(1+2\,\eta)^2}\,. \tag{36}$$

Equation 2 with coefficients given by equations 34-36 is precisely equivalent to the traveltime approximation suggested previously by (17). (17) shows comparisons with alternative non-hyperbolic approximations, which demonstrate superior accuracy of equation 2 in case of strongly anisotropic media.

Numerical example

For a numerical test, we create a one-dimensional velocity model by extracting a depth column out of the anisotropic Marmousi model, created by (1) and shown in Figure 4. We evaluate exact reflection traveltimes by ray tracing (Figure 5). Next, we compare the exact time for different reflection rays with values predicted by different traveltime approximations. As shown in Figure 6, only the proposed generalized approximation is able to predict the true traveltime accurately over the full range of offsets. To define approximation parameters, we used equations 19-21 and 22-23 and the ray with the largest offset as the reference ray.

DISCUSSION

Approximation is more of an art than a science. We don't have a justification for suggesting equations 1 or 2 other than pointing out that they reduce to known forms with particular choices of parameters.

The choice of a proper functional form is important for the approximation accuracy. Suppose that we try to replace the five-parameter approximation in equation 2 with the four-parameter equation

$$t^2(x) \approx t_0^2 + \frac{x^2}{v^2} + \frac{A\,x^4}{v^4\left(2\,t_0^2 + D\,\dfrac{x^2}{v^2}\right)}\,, \tag{37}$$

where $D = B + \sqrt{C}$. Equation 37 has the same behavior as equation 2 at small offsets and the same asymptote as x approaches infinity. However, its accuracy is not nearly as spectacular (Figure 7).

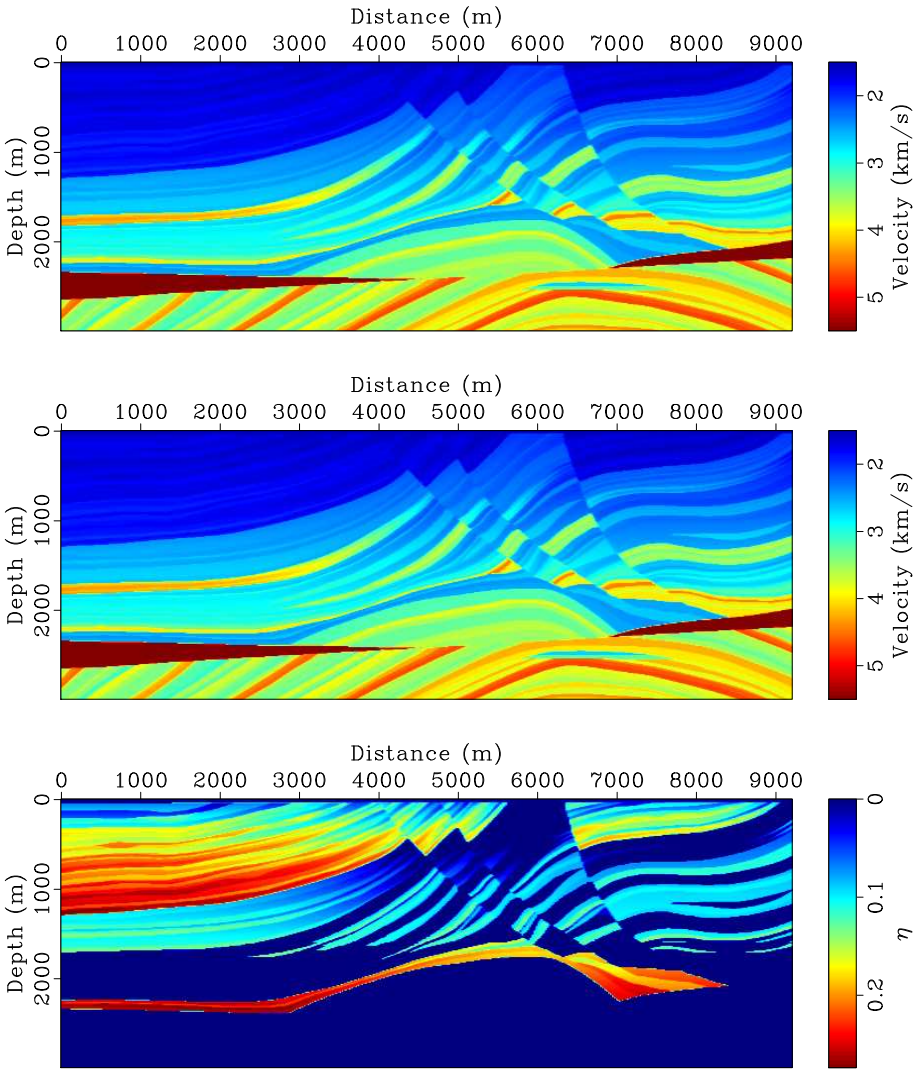


Figure 4: Anisotropic Marmousi model. (a) Vertical velocity. (b) NMO velocity. (c) η parameter. hyper/amarm vz,v,eta

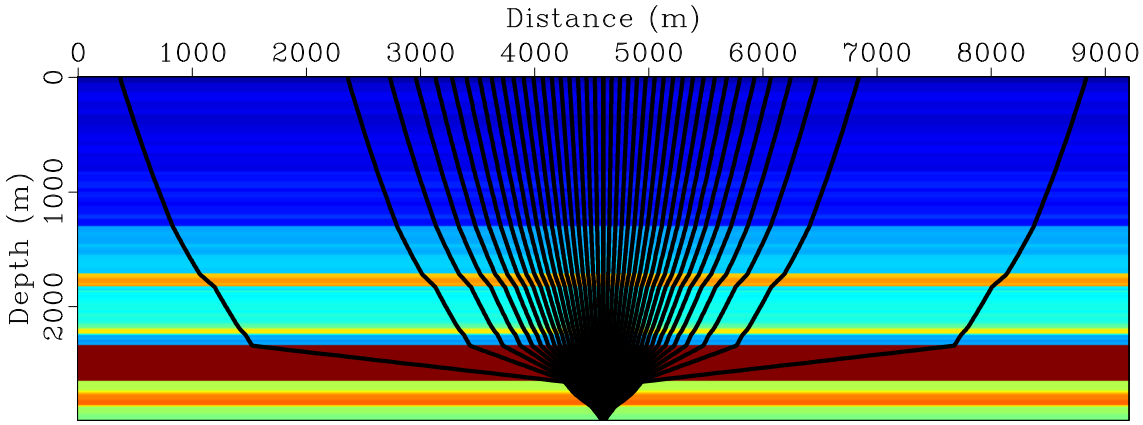


Figure 5: One-dimensional model extracted from the left column of the anisotropic Marmousi model and corresponding reflection rays. hyper/amarm vz1

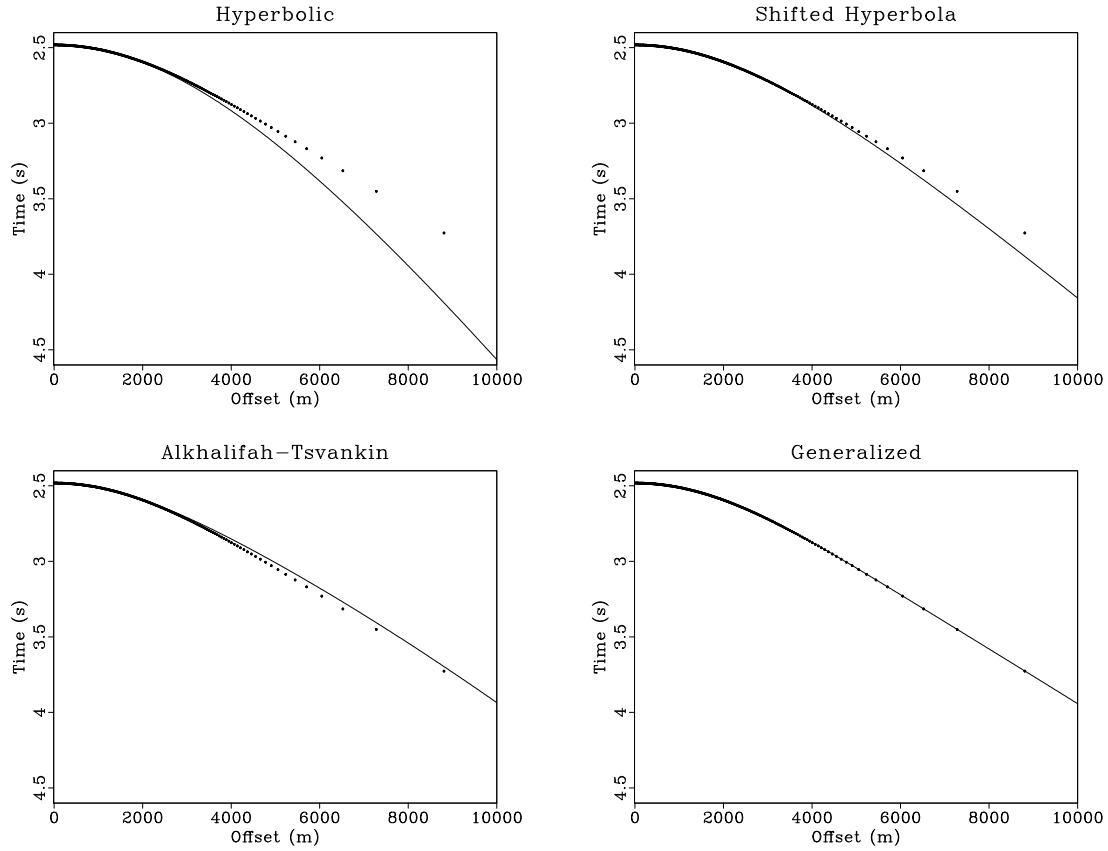


Figure 6: Exact moveout from ray tracing in the one-dimensional anisotropic Marmousi model (dots) and different approximations (solid lines). (a) Hyperbolic approximation, (b) Shifted hyperbola approximation, (c) Alkhalifah-Tsvankin approximation, (d) Generalized nonhyperbolic approximation. hyper/amarm hynmo,shnmo,atnmo,fsnmo

A proper selection of the reference ray for equations 22 and 23 is also important for approximation accuracy. If this ray is taken not at the largest possible offset, the accuracy will deteriorate. As an extreme example, suppose that we try to define B and C by fitting subsequent terms of the Taylor expansion 18 near the zero offset rather than the behavior of the approximation at large offsets. Figure 8 shows the result for the case of a linear sloth model: the approximation is more accurate than alternatives (shown in Figure 2) but not nearly as accurate as the generalized approximation fitted at the critical offset.

Possible extensions of this work may include nonhyperbolic approximations for diffraction traveltimes (for use in prestack time migration) and reflection surfaces (for use in common-reflection-surface methods) as well as approximations for anisotropic phase and group velocities in ray tracing and wave extrapolation.

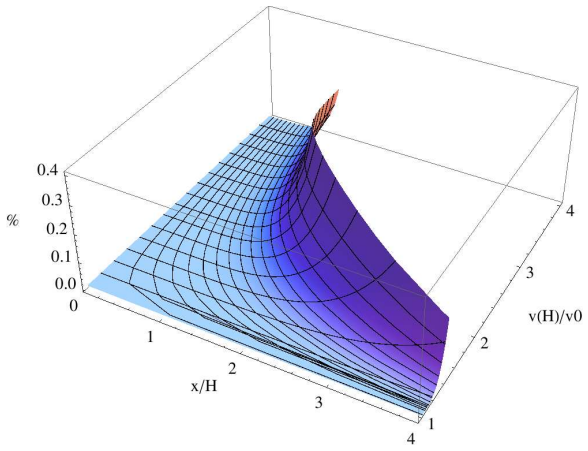


Figure 7: Relative absolute error of Padé approximation in equation 37 as a function of velocity contrast and offset/depth ratio for the case of a linear sloth model. Compare with Figure 2. hyper/Math linsloth2

CONCLUSIONS

We propose a five-parameter nonhyperbolic moveout approximation that generalizes the classic two-parameter hyperbolic approximation as well as some known three-parameter approximations. We propose a method for selecting the approximation parameters, which involves only two rays: the normal-incident ray and one additional ray, preferably at a large offset. The special case of the additional ray being horizontal can be handled as well.

A comparison with the classic hyperbolic approximation, the shifted hyperbola approximation and the Alkhalifah-Tsvankin approximation for analytical and numerical isotropic and transversely isotropic models shows that the proposed generalized nonhyperbolic approximation can bring an improvement of several orders of magnitude in approximation accuracy. Based on these experiments, we claim that, for

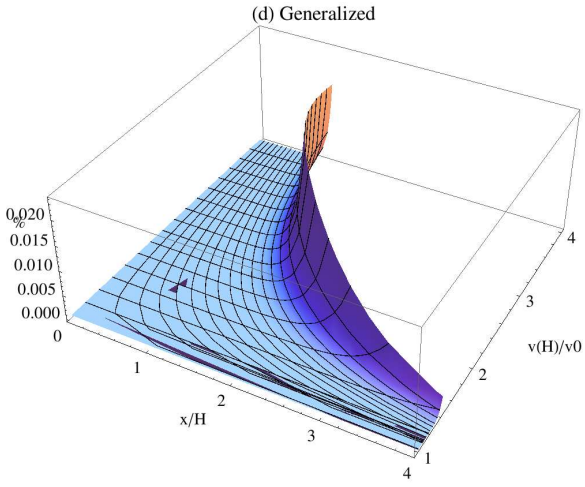


Figure 8: Relative absolute error of the generalized approximation fitted to the zero offset as opposed to the critical offset. Compare with Figure 2. hyper/Math linsloth1

many practical purposes, the proposed approximation can be used in place of the exact moveout.

ACKNOWLEDGMENTS

We would like to thank Tariq Alkhalifah, Emil Blias, Houb Douma, and Bin Wang for many helpful suggestions that improved the paper.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APPENDIX A: LINEAR VELOCITY MODEL

The linear velocity model is defined by

$$V(z) = V_0\,(1 + g\,z) \tag{A-1}$$

where g is the velocity gradient and V_0 is velocity at zero depth.

The reflection traveltime can be expressed in an analytical form as a function of offset (24)

$$t(x) = \frac{2\,H}{V_0\,(r-1)}\,\mathrm{arccosh}\left[1 + \frac{(r-1)^2}{2\,r}\,\left(1 + \frac{x^2}{4\,H^2}\right)\right], \tag{A-2}$$

where H is the depth of the reflector, and $r = V(H)/V_0$ is the ratio of velocity at the

bottom and the top of the model. The traveltime parameters are given by

$$t_0 = \frac{2\,H}{V_0}\,\frac{\ln r}{r-1}\,, \tag{A-3}$$

$$v^2 = V_0^2\,\frac{r^2-1}{2\,\ln r}\,, \tag{A-4}$$

$$A = \frac{1}{2}\left(1-\frac{r^2+1}{r^2-1}\,\ln r\right)\,. \tag{A-5}$$

This model has maximum (critical) offset and traveltime that are defined by

$$X = 2\,H\,\sqrt{\frac{r+1}{r-1}}\,, \tag{A-6}$$

$$T = \frac{2\,H}{V_0}\,\frac{\operatorname{arccosh} r}{r-1}\,. \tag{A-7}$$

Substituting equations A-6 and A-7 into equations 22-23 and also using the expressions for traveltime parameters A-3, A-4, and A-5 results in complicated but analytical expressions for additional parameters B and C.

APPENDIX B: LINEAR SLOTH MODEL

The linear sloth model is defined by

$$\frac{1}{V^2(z)} = \frac{1}{V_0^2}\,(1+G\,z)\,. \tag{B-1}$$

where G is the sloth gradient and V_0 is velocity at zero depth.

The equation for traveltime can be computed analytically, as follows (11):

$$t^2(x) = t_0^2 + \frac{x^2}{v^2} - \frac{x^4\,(r^2-1)^2\,(2\,Q+1)}{144\,Q^3\,H^2\,v^2\,(Q+1)^2}\,, \tag{B-2}$$

where H is the depth of the reflector, $r = V(H)/V_0$ is the ratio of velocity at the bottom and the top of the model and

$$Q = \sqrt{1-\frac{x^2\,(r^2-1)^2}{16\,r^2\,H^2}}\,.$$

The main traveltime parameters are given by

$$t_0 = \frac{4\,H}{3\,V_0}\,\frac{1+r+r^2}{r\,(r+1)}\,, \tag{B-3}$$

$$v^2 = V_0^2\,\frac{3\,r^2}{1+r+r^2}\,, \tag{B-4}$$

$$A = -\frac{(r-1)^2}{6\,r}\,. \tag{B-5}$$

The maximum offset and traveltime are defined by

$$X = \frac{4\,H}{\sqrt{r^2-1}}\,, \tag{B-6}$$

$$T = \frac{4\,H}{3\,V_0}\,\frac{r^2+2}{r\,\sqrt{r^2-1}}\,. \tag{B-7}$$

Substituting equations B-6 and B-7 into equations 22-23 and using the expressions for traveltime parameters B-3, B-4, and B-5 results in the following analytical expressions for additional parameters B and C:

$$B = -\frac{(r-1)^2\,(1+r+r^2)}{2\,r\,(r+2)\,(2\,r+1)}\,, \tag{B-8}$$

$$C = -\frac{(r-1)^4\,(1+r+r^2)^2}{3\,r\,(r+2)\,(2\,r+1)^2}\,. \tag{B-9}$$

APPENDIX C: REFLECTION FROM A HYPERBOLIC REFLECTOR IN A HOMOGENEOUS VELOCITY MODEL

In this appendix, we derive an analytical expression for reflection traveltime from a hyperbolic reflector in a homogeneous velocity model (Figure C-1). Similar derivations apply to an elliptic reflector and were used previously in the theory of dip moveout, offset continuation, and non-hyperbolic common-reflection surface (25; 16; 19).

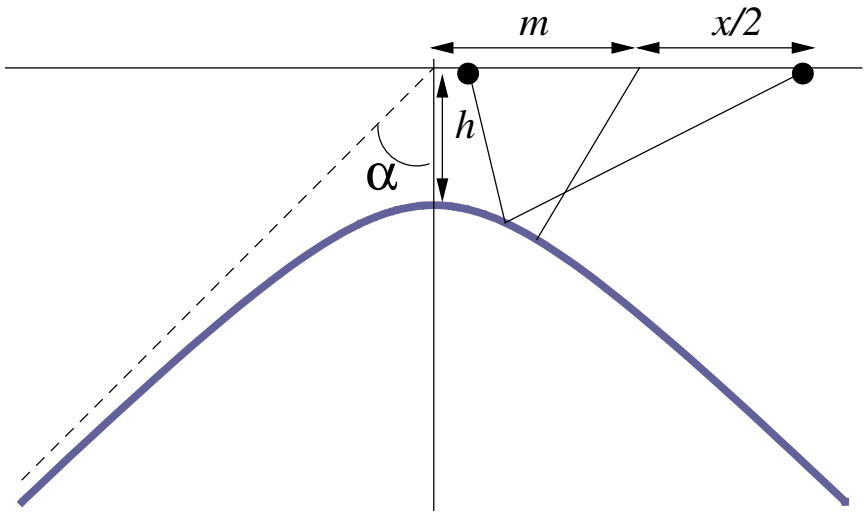


Figure C-1: Reflection from a hyperbolic reflector in a homogeneous velocity model (a scheme). hyper/XFig hyper

Consider the source point x_s and the receiver point x_r at the surface $z=0$ above a 2-D constant-velocity medium and a hyperbolic reflector defined by the equation

$$z(x) = \sqrt{h^2 + x^2\,\tan^2\alpha}\,. \tag{C-1}$$

The reflection traveltime as a function of the reflection point location y is

$$t = \frac{\sqrt{(x_s - y)^2 + z^2(y)} + \sqrt{(x_r - y)^2 + z^2(y)}}{V} \; . \tag{C-2}$$

According to Fermat's principle, the traveltime should be stationary with respect to the reflection point y:

$$0 = \frac{\partial t}{\partial y} = \frac{y - x_s + y \, \tan^2 \alpha}{V \sqrt{(x_s - y)^2 + h^2 + y^2 \, \tan^2 \alpha}} + \frac{y - x_r + y \, \tan^2 \alpha}{V \sqrt{(x_r - y)^2 + h^2 + y^2 \, \tan^2 \alpha}} \; . \tag{C-3}$$

Putting two terms in equation A-3 on different sides of the equation, squaring them, and reducing their difference to a common denominator, we arrive at the equation

$$\begin{aligned} 0 &= \left[\frac{y}{\cos^2 \alpha} - x_s\right]^2 \left[(x_r - y)^2 + h^2 + y^2 \, \tan^2 \alpha\right] \\ &- \left[\frac{y}{\cos^2 \alpha} - x_r\right]^2 \left[(x_s - y)^2 + h^2 + y^2 \, \tan^2 \alpha\right] \end{aligned} \tag{C-4}$$

which simplifies to the following quadratic equation with respect to y:

$$y^2 \, (x_s + x_r) \, \tan^2 \alpha - 2 \, y \, \left(x_s \, x_r \, \sin^2 \alpha - h^2\right) - h^2 \, (x_s + x_r) \, \cos^2 \alpha = 0 \; . \tag{C-5}$$

The discriminant is

$$D = \left(x_s \, x_r \, \sin^2 \alpha - h^2\right)^2 + h^2 \, (x_s + x_r)^2 \, \sin^2 \alpha = (h^2 + x_s^2 \, \sin^2 \alpha) \, (h^2 + x_r^2 \, \sin^2 \alpha) \; . \tag{C-6}$$

Only one of the two branches of the solution

$$\begin{aligned} y &= \frac{x_s \, x_r \, \sin^2 \alpha - h^2 + \sqrt{(h^2 + x_s^2 \, \sin^2 \alpha) \, (h^2 + x_r^2 \, \sin^2 \alpha)}}{(x_s + x_r) \, \tan^2 \alpha} \\ &= \frac{h^2 \, (x_s + x_r) \, \cos^2 \alpha}{h^2 - x_s \, x_r \, \sin^2 \alpha + \sqrt{(h^2 + x_s^2 \, \sin^2 \alpha) \, (h^2 + x_r^2 \, \sin^2 \alpha)}} \end{aligned} \tag{C-7}$$

has physical meaning. Substituting equation into equation A-2, we obtain, after a number of algebraic simplifications,

$$t = \frac{\sqrt{2h^2 + x_s^2 + x_r^2 - 2 \, x_s \, x_r \, \cos^2 \alpha + 2 \, \sqrt{(h^2 + x_s^2 \, \sin^2 \alpha) \, (h^2 + x_r^2 \, \sin^2 \alpha)}}}{V} \; . \tag{C-8}$$

Making the variable change in equation A-5 from x_s and x_r to the midpoint and offset coordinates m and x according to $x_s = m - x/2$, $x_r = m + x/2$, we notice that this equation is exactly equivalent to equation 1 with the following definition of

parameters:

$$t_0 = \frac{2\sqrt{h^2+m^2\,\sin^2\alpha}}{V}\,, \tag{C-9}$$

$$a = \frac{2-\sin^2\alpha}{V^2}\,, \tag{C-10}$$

$$b = \frac{\sin^2\alpha}{V^2}\,\frac{h^2-m^2\,\sin^2\alpha}{h^2+m^2\,\sin^2\alpha}\,, \tag{C-11}$$

$$c = \frac{\sin^4\alpha}{V^4}\,, \tag{C-12}$$

$$\xi = \frac{1}{2}\,. \tag{C-13}$$

APPENDIX D: REFLECTION FROM A CIRCULAR REFLECTOR IN A HOMOGENEOUS VELOCITY MODEL

In the case of a circular (cylindrical or spherical) reflector in a homogeneous velocity model, there is no closed-form analytical solution. However, the moveout can be described analytically by parametric relationships (21).

Consider the geometry of the reflection shown in Figure B-1. According to the trigonometry of the reflection triangles, the source and receiver positions can be expressed as

$$x_s = R\,\sin\alpha + (H+R-R\,\cos\alpha)\,\tan(\alpha-\theta)\,, \tag{D-1}$$

$$x_r = R\,\sin\alpha + (H+R-R\,\cos\alpha)\,\tan(\alpha+\theta)\,, \tag{D-2}$$

where R is the reflector radius, H is the minimum reflector depth, α is the reflector dip angle at the reflection point, and θ is the reflection angle. Correspondingly, the midpoint and offset coordinates can be expressed as

$$m = \frac{x_s+x_r}{2} = R\,\sin\alpha + (H+R-R\,\cos\alpha)\,\frac{\cos\alpha\,\sin\alpha}{\cos^2\theta-\sin^2\alpha}\,, \tag{D-3}$$

$$x = x_r - x_s = 2\,(H+R-R\,\cos\alpha)\,\frac{\cos\theta\,\sin\theta}{\cos^2\theta-\sin^2\alpha}\,, \tag{D-4}$$

and the reflection traveltime can be expressed as

$$\begin{aligned} t &= \frac{H+R-R\,\cos\alpha}{V}\left[\frac{1}{\cos(\alpha-\theta)}+\frac{1}{\cos(\alpha+\theta)}\right] \\ &= 2\,\frac{H+R-R\,\cos\alpha}{V}\,\frac{\cos\alpha\,\cos\theta}{\cos^2\theta-\sin^2\alpha}\,, \end{aligned} \tag{D-5}$$

where V is the medium velocity. Expressing the reflection angle θ from equation A-7 and substituting it into equations D-4 and B-4, we obtain a pair of parametric equa-

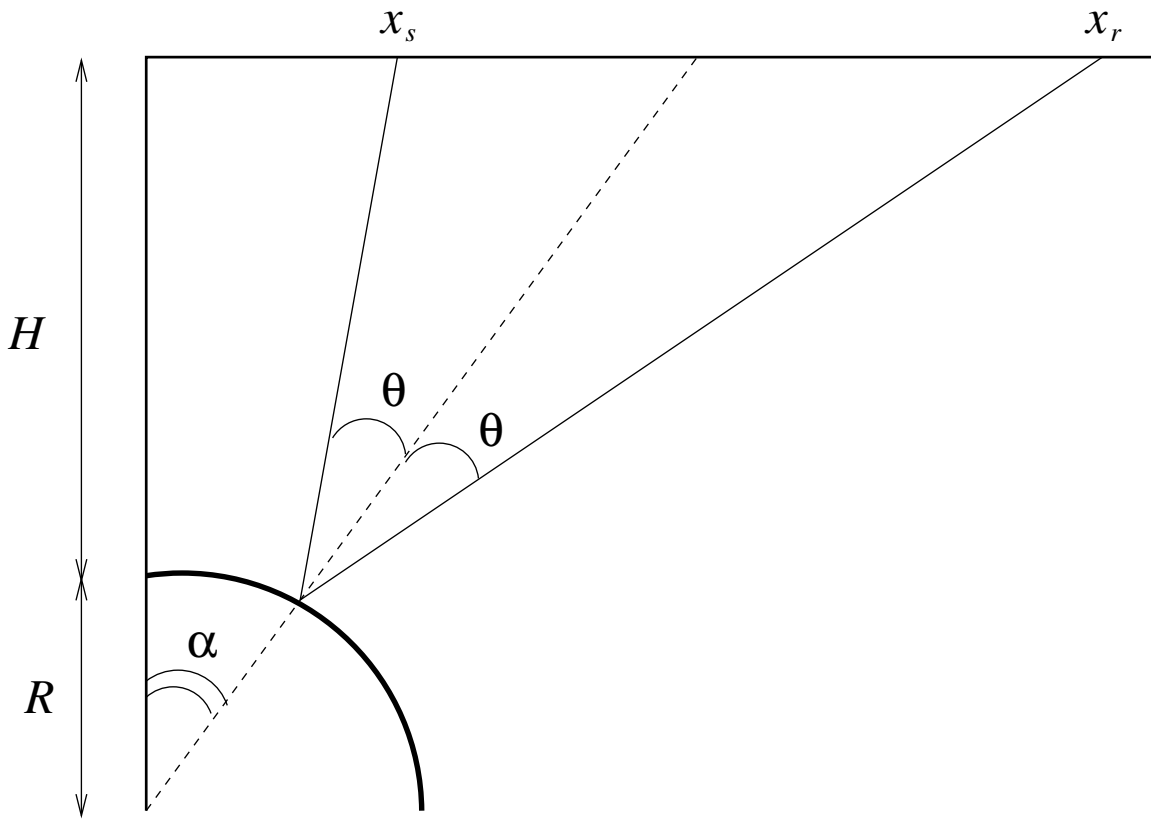


Figure D-1: Reflection from a circular reflector in a homogeneous velocity model (a scheme). hyper/XFig crefl

tions

$$x^2(\alpha) = 4\,\frac{[m\,\cos\alpha-(H+R)\,\sin\alpha]\,[m\,\sin\alpha+(H+R)\,\cos\alpha-R]}{\cos\alpha\,\sin\alpha}\,, \quad \text{(D-6)}$$

$$t^2(\alpha) = \frac{4}{V^2}\,\frac{(m-R\,\sin\alpha)\,[m\,\sin\alpha+(H+R)\,\cos\alpha-R]}{\sin\alpha}\,, \quad \text{(D-7)}$$

which define the exact reflection moveout for the case of a circular reflector in a homogeneous medium.

The connection with parameters of equations 27-29 is given by

$$L = \sqrt{m^2+(H+R)^2}-R\,, \quad \text{(D-8)}$$

$$\cos\beta = \frac{H+R}{\sqrt{m^2+(H+R)^2}}\,, \quad \text{(D-9)}$$

$$G = \frac{L}{L+R} = 1-\frac{R}{\sqrt{m^2+(H+R)^2}}\,. \quad \text{(D-10)}$$

The behavior of the moveout at infinitely large offsets is controlled by $P_\infty = 1/V$ and

$$T_\infty = \frac{2\,H}{V} = t_0\,\frac{G+\cos\beta-1}{G\,\cos\beta}\,. \quad \text{(D-11)}$$

After substitution in equations 25-26, we obtain somewhat complicated but analytical expressions for parameters B and C.

APPENDIX E: HOMOGENEOUS VTI MODEL

According to the acoustic approximation of (3), one can use the following parametric equations to define the traveltime-offset relationship in a homogeneous VTI model:

$$x(p) = \frac{2\,H}{v_z}\,\frac{p\,v^2}{(1-2\,\eta\,p^2\,v^2)^2\,\sqrt{1-\frac{p^2\,v^2}{1-2\,\eta\,p^2\,v^2}}}\,, \quad \text{(E-1)}$$

$$t(p) = \frac{2\,H}{v_z}\,\frac{(1-2\,\eta\,p^2\,v^2)^2+2\,\eta\,p^4\,v^4}{(1-2\,\eta\,p^2\,v^2)^2\,\sqrt{1-\frac{p^2\,v^2}{1-2\,\eta\,p^2\,v^2}}}\,, \quad \text{(E-2)}$$

where p is the ray parameter, H is the depth of the reflector, v_z is the vertical velocity, v is the NMO velocity, and η is the dimensionless parameter introduced by (4).

At small offsets, the homogeneous VTI traveltime behaves as

$$t^2(x) \approx t_0^2+\frac{x^2}{v^2}-\frac{2\,\eta\,x^4}{t_0^2\,v^4}\,, \quad \text{(E-3)}$$

which allows us to define $A=-4\,\eta$ according to equation 18.

At large offsets, the homogeneous VTI traveltime behaves as

$$t^2(x) \approx t_0^2\,(1+2\,\eta) + \frac{x^2}{v^2\,(1+2\,\eta)}\;. \tag{E-4}$$

Comparing with equation 24, we note that $T_\infty = t_0\,\sqrt{1+2\,\eta}$ and $P_\infty = 1/(v\,\sqrt{1+2\,\eta})$. Substituting into equations 25-26, we derive the coefficients B and C to be

$$B = \frac{1+8\,\eta+8\,\eta^2}{1+2\,\eta}\;, \tag{E-5}$$

$$C = \frac{1}{(1+2\,\eta)^2}\;. \tag{E-6}$$

REFERENCES

[1] Alkhalifah, T., 1997a, An anisotropic Marmousi model, *in* SEP-95: Stanford Exploration Project, 265–282.

[2] ———, 1997b, Seismic data processing in vertically inhomogeneous TI media: Geophysics, **62**, 662–675.

[3] ———, 1998, Acoustic approximations for processing in transversely isotropic media: Geophysics, **63**, 623–631.

[4] Alkhalifah, T., and I. Tsvankin, 1995, Velocity analysis for transversely isotropic media: Geophysics, **60**, 1550–1566.

[5] Blias, E., 2007, Long-spreadlength approximations to NMO function for a multilayered subsurface: Recorder, **3**, 36–42.

[6] ———, 2009, Long-offset NMO approximations for a layered VTI model: Model study: Presented at the 79th Ann. Internat. Mtg, Soc. of Expl. Geophys.

[7] Blias, E. A., 1983, Reflected wave's traveltime curve in flat-bedded medium with transverse layers and their interpretation: Soviet Geology and Geophysics, 91–95.

[8] Bolshykh, S. F., 1956, About an approximate representation of the reflected wave traveltime curve in the case of a multi-layered medium: Applied Geophysics (in Russian), **15**, 3–15.

[9] Castle, R. J., 1994, Theory of normal moveout: Geophysics, **59**, 983–999.

[10] Causse, E., 2004, Approximations of reflection travel times with high accuracy at all offsets: Journal of Geophysics and Engineering, **1**, 28–45.

[11] Červený, V., 2001, Seismic ray theory: Cambridge University Press.

[12] de Bazelaire, E., 1988, Normal moveout revisited - Inhomogeneous media and curved interfaces: Geophysics, **53**, 143–157.

[13] Douma, H., and A. Calvert, 2006, Nonhyperbolic moveout analysis in VTI media using rational interpolation: Geophysics, **71**, D59–D71.

[14] Douma, H., and M. van der Baan, 2008, Rational interpolation of qP-traveltimes for semblance-based anisotropy estimation in layered VTI media: Geophysics, **73**, D53–D62.

[15] Fomel, S., 1994, Recurrent formulas for derivatives of a CMP traveltime curve: Russian Geology and Geophysics, **35**, 118–126.

[16] ———, 2003, Theory of differential offset continuation: Geophysics, **68**, 718–732.
[17] ———, 2004, On anelliptic approximations for qP velocities in VTI media: Geophys. Prosp., **52**, 247–259.
[18] Fomel, S., and V. Grechka, 2001, Nonhyperbolic reflection moveout of P waves. An overview and comparison of reasons: Technical Report CWP-372, Colorado School of Mines.
[19] Fomel, S., and R. Kazinnik, 2009, Non-hyperbolic common reflection surface: Presented at the 79th Ann. Internat. Mtg, Soc. of Expl. Geophys.
[20] Fowler, P. J., 2003, Practical VTI approximations: a systematic anatomy: Journal of Applied Geophysics, **54**, 347–367.
[21] Glaeser, G., 1999, Reflections on spheres and cylinders of revolution: Journal for Geometry and Graphics, **3**, 121–139.
[22] Lyakhovitsky, F. M., and M. V. Nevskiy, 1971, Traveltime curves of reflected waves for transversely-isotropic medium: Proceedings of the USSR Academy of Sciences, **196**, 327–330.
[23] Malovichko, A. A., 1978, A new representation of the traveltime curve of reflected waves in horizontally layered media: Applied Geophysics (in Russian), **91**, 47–53, English translation in (26).
[24] Slotnick, M. M., 1959, Lessons in Seismic Computing: Soc. of Expl. Geophys. (Edited by R. A. Geyer).
[25] Stovas, A. M., and S. B. Fomel, 1996, Kinematically equivalent integral DMO operators: Russian Geology and Geophysics, **37**, 102–113.
[26] Sword, C. H., 1987, A Soviet look at datum shift, *in* SEP-51: Stanford Exploration Project, 313–316.
[27] Taner, M. T., and F. Koehler, 1969, Velocity spectra - Digital computer derivation and applications of velocity functions: Geophysics, **34**, 859–881. (Errata in GEO-36-4-0787).
[28] Taner, M. T., S. Treitel, and M. Al-Chalabi, 2005, A new travel time estimation method for horizontal strata: 75th Ann. Internat. Mtg, Soc. of Expl. Geophys., 2273–2276.
[29] Taner, M. T., S. Treitel, M. Al-Chalabi, and S. Fomel, 2007, An offset dependent NMO velocity model: EAGE 69th Conference and Exhibition, EAGE, P036.
[30] Ursin, B., and A. Stovas, 2006, Traveltime approximations for a layered transversely isotropic medium: Geophysics, **71**, D23–D33.
[31] Yilmaz, O., 2000, Seismic data analysis: Soc. of Expl. Geophys.

Lowrank seismic wave extrapolation on a staggered grid

*Gang Fang[1], Sergey Fomel[2], Qizhen Du[3], and Jingwei Hu[4]**

ABSTRACT

We propose a new spectral method and a new finite-difference method for seismic wave extrapolation in time. Using staggered temporal and spatial grids, we derive a wave extrapolation operator using a lowrank decomposition for a first-order system of wave equations and design the corresponding finite-difference scheme. The proposed methods extend previously proposed lowrank and lowrank finite-difference wave extrapolation methods from the cases of constant density to those of variable density. Dispersion analysis demonstrates that the proposed methods have high accuracy for a wide wavenumber range and significantly reduce the numerical dispersion. The method of manufactured solutions coupled with mesh refinement is used to verify each method and to compare numerical errors. 2-D synthetic examples demonstrate that the proposed method is highly accurate and stable. The proposed methods can be used for seismic modeling or reverse time migration.

INTRODUCTION

Wave extrapolation in time is an essential part of seismic imaging, modeling, and full waveform inversion. Finite-difference methods (8; 37) and spectral methods (33; 24) are the two most popular and straightforward ways of implementing wave extrapolation in time. The finite-difference (FD) methods are highly efficient and easy to implement. However, the traditional FD methods are only conditionally stable and suffer from numerical dispersion (10). Thanks to advances in supercomputer technology, spectral methods have become feasible for large-scale problems. Compared with FD methods, spectral methods have superior accuracy and are able to suppress dispersion artifacts (9).

Several spectral methods have been developed for seismic wave extrapolation in variable-velocity media (14; 6). (39) and (38) proposed a one-step extrapolation algorithm, which is derived from an optimized separable approximation (OSA). This algorithm formulates the two-way wave equation as a first-order partial differential equation in time without suffering from numerical instability or dispersion problems. (30) proposed a two-step extrapolation method that is based on a high-order differential operator, which allows for large time steps in extrapolation. However, the

*e-mail: sergey.fomel@beg.utexas.edu

decomposition algorithm in OSA can be expensive, particularly in anisotropic media. (13) presented an approach to approximating the mixed-domain operator using a lowrank decomposition, which reduces computational cost by optimally selecting reference velocities and weights. (27) developed a related method, Fourier finite-differences (FFD), by cascading a Fourier transform operator and a finite-difference operator to form a chain operator.

In practice, first-order wave equations are often involved in handling wave extrapolation in media with both velocity and density variations. (20) provided a derivation of the k-space method for solving the ultrasonic wave equation. (31) extended this method to solving coupled first-order differential equations for wave propagation, efficiently accounting for velocity and density variations. In the k-space method, dispersion errors from a second-order time integration operator are compensated by a modified spectral operator in the wavenumber domain. This correction is exact for a medium with constant velocity in particular. (29) modified the k-space method with a mixed-domain operator and applied FFD to hand these operators. This method has highly accurate for variable velocity and density.

In the FD methods, the FD coefficients are conventionally determined through a Taylor-series expansion around the zero wavenumber (5; 15). Traditional FD methods are therefore particularly accurate for long-wavelength components. Several approaches have been proposed to improve the performance of FD method in practice. Implicit FD operators (18; 3) can be used to achieve high numerical accuracy. Another way to control numerical errors is to use optimized FD operators (32; 4; 17). (28) derived optimized coefficients of the FD operator from a lowrank approximation (13) of the space-wavenumber extrapolation matrix.To improve the accuracy and stability, the FD methods have been developed on a staggered grid (19; 34; 35; 16). (21) investigated the stability and grid dispersion in the 3D fourth-order staggered grid FD scheme. In the past years, the viscous wave modeling using staggered grid FD methods (25; 2; 23)have also been studied and reported.

In this paper, we use modified staggered grid k-space method (31; 29) to handle the derivative operator in mixed-domain for variable velocity and density. We introduce lowrank decomposition (13) to approximate the modified k-space extrapolation operator and reduce the computational cost. Inspired by lowrank finite-differences (28), we derive optimized finite-difference coefficients for coupled first-order wave-propagation equations using staggered spatial and temporal grids (34; 35; 16). We apply dispersion analysis and use the method of manufactured solutions to evaluate the accuracy of the proposed methods. Numerical tests demonstrate that the proposed SGL(staggered grid lowrank) and SGLFD(staggered grid lowrank finite-differences) methods are highly accurate and applicable for variable velocity and density modeling and reverse time migration (RTM) in complicated models. Our implementation of the new methods and the numerical examples are based on *Madagascar* software (11) and can be reproduced using the latest version of *Madagascar*.

THEORY

Second-order and first-order mixed-domain methods

We consider first-order acoustic wave equations for a medium of variable velocity and density. For a lossless 2-D medium,

$$\begin{aligned} \rho(\mathbf{x})\frac{\partial \mathbf{u}(\mathbf{x},t)}{\partial t} &= -\nabla p(\mathbf{x},t), \\ \frac{1}{\rho(\mathbf{x})v^2(\mathbf{x})}\frac{\partial p(\mathbf{x},t)}{\partial t} &= -\nabla\cdot\mathbf{u}(\mathbf{x},t), \end{aligned} \tag{1}$$

where $\mathbf{u}(\mathbf{x},t)$ is acoustic particle velocity with components $u_x(\mathbf{x},t)$ and $u_z(\mathbf{x},t)$; $p(\mathbf{x},t)$ is the acoustic pressure; $\rho(\mathbf{x})$ is density of the medium; $v(\mathbf{x})$ is seismic wave velocity of the medium; and $\mathbf{x}=(x,z)$ denotes the space location in vector coordinate.

The second-order wave equation corresponding to equation 1 is

$$\nabla\cdot\frac{1}{\rho(\mathbf{x})}\nabla p(\mathbf{x},t) - \frac{1}{\rho(\mathbf{x})v^2(\mathbf{x})}\frac{\partial^2 p(\mathbf{x},t)}{\partial t^2} = 0. \tag{2}$$

In the case of homogeneous velocity and density, equation 2 can be written in the spatial-frequency domain as

$$\frac{\partial^2 \hat{p}(\mathbf{k},t)}{\partial t^2} = -v_0^2\mathbf{k}^2\hat{p}(\mathbf{k},t), \tag{3}$$

where $\hat{p}(\mathbf{k},t)$ is the 2-D spatial Fourier transform of $p(\mathbf{x},t)$. Equation 3 has an analytical solution,

$$\hat{p}(\mathbf{k},t+\Delta t) = e^{\pm iv_0|\mathbf{k}|\Delta t}\hat{p}(\mathbf{k},t). \tag{4}$$

Applying the second-order time-marching scheme leads to the k-space scheme (31),

$$\begin{aligned} &\frac{\hat{p}(\mathbf{k},t+\Delta t) - 2\hat{p}(\mathbf{k},t) + \hat{p}(\mathbf{k},t-\Delta t)}{\Delta t^2} \\ &= -(v_0\,|\mathbf{k}|)^2\mathrm{sinc}^2(v_0\,|\mathbf{k}|\,\Delta t/2)\hat{p}(\mathbf{k},t), \end{aligned} \tag{5}$$

where $\mathrm{sinc}(x) = \sin(x)/x$. In general, velocity and density vary in space. When both the gradient of velocity and the time step are small, replacing v_0 with $v(\mathbf{x})$ in equation 5 provides a new approximation. Applying inverse Fourier transform to this approximation leads to the scheme

$$\begin{aligned} &\frac{p(\mathbf{x},t+\Delta t) - 2p(\mathbf{x},t) + p(\mathbf{x},t-\Delta t)}{\Delta t^2} \\ &= -v^2(\mathbf{x})\mathbf{F}^{-1}\left[|\mathbf{k}|^2\,\mathrm{sinc}^2(v(\mathbf{x})\,|\mathbf{k}|\,\Delta t/2)\mathbf{F}\left[p(\mathbf{x},t)\right]\right], \end{aligned} \tag{6}$$

where $\mathbf{F}$ denotes a spatial Fourier transform. The operator on the right-hand side of equation 6 depends on both $\mathbf{x}$ and $\mathbf{k}$. Following (31) we call it second-order kx-space operator,

$$\left[\nabla^{v(\mathbf{x})\Delta t}\right]^2 p(\mathbf{x},t) \equiv -\mathbf{F}^{-1}\left[|\mathbf{k}|^2\,\mathrm{sinc}^2(v(\mathbf{x})\,|\mathbf{k}|\,\Delta t/2)\mathbf{F}\left[p(\mathbf{x},t)\right]\right], \tag{7}$$

where the operator $\left[\nabla^{v(\mathbf{x})\Delta t}\right]^2$ is analogous to the standard gradient operator, but it is a function of parameter $v(\mathbf{x})\Delta t$. Similar to the definition of the standard gradient operator, we can define

$$\left[\nabla^{v(\mathbf{x})\Delta t}\right]^2 p(\mathbf{x},t) \equiv \left(\frac{\partial}{\partial^+ x}\frac{\partial}{\partial^- x} + \frac{\partial}{\partial^+ z}\frac{\partial}{\partial^- z}\right) p(\mathbf{x},t). \tag{8}$$

(31) suggested a factorization, which can factor their second-order k-space operator into parts associated with each spatial direction. This factorization can also be applied to kx-space operator by replacing constant velocity with variable velocity. The factored operators are called first-order kx-space operators:

$$\begin{aligned}
\frac{\partial p(\mathbf{x},t)}{\partial^+ x} &\equiv \mathbf{F}^{-1}\left[ik_x e^{ik_x\Delta x/2}\mathrm{sinc}(v(\mathbf{x})\,|\mathbf{k}|\,\Delta t/2)\mathbf{F}[p(\mathbf{x},t)]\right],\\
\frac{\partial p(\mathbf{x},t)}{\partial^- x} &\equiv \mathbf{F}^{-1}\left[ik_x e^{-ik_x\Delta x/2}\mathrm{sinc}(v(\mathbf{x})\,|\mathbf{k}|\,\Delta t/2)\mathbf{F}[p(\mathbf{x},t)]\right],\\
\frac{\partial p(\mathbf{x},t)}{\partial^+ z} &\equiv \mathbf{F}^{-1}\left[ik_z e^{ik_z\Delta z/2}\mathrm{sinc}(v(\mathbf{x})\,|\mathbf{k}|\,\Delta t/2)\mathbf{F}[p(\mathbf{x},t)]\right],\\
\frac{\partial p(\mathbf{x},t)}{\partial^- z} &\equiv \mathbf{F}^{-1}\left[ik_z e^{-ik_z\Delta z/2}\mathrm{sinc}(v(\mathbf{x})\,|\mathbf{k}|\,\Delta t/2)\mathbf{F}[p(\mathbf{x},t)]\right].
\end{aligned} \tag{9}$$

The spatial frequency components k_x and k_z are defined so that $k^2 = k_x^2 + k_z^2$. Application of the exponential coefficient in equation 9 requires the corresponding wavefield to be evaluated on grid points staggered by distance of $\Delta x/2$ along the positive or negative x direction and $\Delta z/2$ along the positive or negative z direction. The spatial staggering in equation 9 is implicitly incorporated into the spatial derivative by the shift property of the Fourier transform. Using operators in equation 9 within equation 1 enables a new construction of spectral method. The first-order coupled equations for acoustic-wave extrapolation in variable velocity and density media with staggered spatial and temporal grids are therefore:

$$\begin{aligned}
\frac{u_x(\mathbf{x}_1,t^+) - u_x(\mathbf{x}_1,t^-)}{\Delta t} &= -\frac{1}{\rho(\mathbf{x}_1)}\frac{\partial p(\mathbf{x},t)}{\partial^+ x},\\
\frac{u_z(\mathbf{x}_2,t^+) - u_z(\mathbf{x}_2,t^-)}{\Delta t} &= -\frac{1}{\rho(\mathbf{x}_2)}\frac{\partial p(\mathbf{x},t)}{\partial^+ z},\\
\frac{p(\mathbf{x},t+\Delta t) - p(\mathbf{x},t)}{\Delta t} &= -\rho(\mathbf{x})v^2(\mathbf{x})\left(\frac{\partial u_x(\mathbf{x}_1,t^+)}{\partial^- x} + \frac{\partial u_z(\mathbf{x}_2,t^+)}{\partial^- z}\right),
\end{aligned} \tag{10}$$

where $\mathbf{x}_1 \equiv (x+\Delta x/2, z)$, $\mathbf{x}_2 \equiv (x, z+\Delta z/2)$, $t^+ \equiv t+\Delta t/2$, $t^- \equiv t-\Delta t/2$.

The partial derivative operators in equation 10 are defined by equation 9. Note that the ordering of $\partial/\partial^+ x$ and $\partial/\partial^- x$ is arbitrary depending on the configuration of staggered grid. However, these operators should be used in pairs, such that the spatial shifting cancel out over any temporal interval length Δt. The spatial and temporal staggered grids used in equation 10 are analogous to staggered scheme employed in previous finite-difference methods(19; 34; 35), which are know to increase accuracy and stability by halving spatial and time interval without increasing the number of computational points. Equation 9 can be solved with a localized Fourier transform

(36). However, this kind of solution has a high computational cost. (29) proposed to apply FFD method to calculate the first-order kx-space operators in equation 9, which can handle the variable velocity and density accurately and efficiently. Another possible ways to speed up the computation is to represent the extrapolation operator with a lowrank matrix.

Lowrank approximation for first-order extrapolation operators

In this section, we apply lowrank decomposition to approximate the extrapolation operator in equation 9. As indicated by (12; 13), the mixed-domain matrix in equation 9, taking $\frac{\partial p(\mathbf{x},t)}{\partial^+ x}$ as an example,

$$W_x(\mathbf{x},\mathbf{k}) = k_x \text{sinc}(v(\mathbf{x})\,|\mathbf{k}|\,\Delta t/2), \tag{11}$$

can be efficiently decomposed into a separated representation as follows:

$$W_x(\mathbf{x},\mathbf{k}) \approx \sum_{m=1}^{M}\sum_{n=1}^{N} W_x(\mathbf{x},\mathbf{k}_m) a_{mn} W_x(\mathbf{x}_n,\mathbf{k}), \tag{12}$$

where $W_x(\mathbf{x},\mathbf{k}_m)$ is a submatrix of $W_x(\mathbf{x},\mathbf{k})$ which consists of selected columns associated with $\mathbf{k}_m$; $W_x(\mathbf{x}_n,\mathbf{k})$ is another submatrix that contains selected rows associated with $\mathbf{x}_n$; and a_{nm} stands for middle matrix coefficients. The numerical construction of the separated representation in equation 12 follows the method of (7).

Using representation 12, we can calculate $\frac{\partial p(\mathbf{x},t)}{\partial^+ x}$ using a small number of fast Fourier transforms (FFTs), because

$$\frac{\partial p(\mathbf{x},t)}{\partial^+ x} \approx \sum_{m=1}^{M}\sum_{n=1}^{N} W_x(\mathbf{x},\mathbf{k}_m) a_{mn} \mathbf{F}^{-1}\left[ie^{ik_x\Delta x/2} W_x(\mathbf{x}_n,\mathbf{k})\mathbf{F}\left[p(\mathbf{x},t)\right]\right]. \tag{13}$$

The same lowrank decomposition approach can be applied to the remaining three partial derivative operators in equation 9. Evaluation of equation 13 only needs N inverse FFTs, whose computational cost is $O(NN_x log N_x)$. However a straightforward application of equation 9 needs computational cost of $O(N_x^2)$, where N_x is the total size of the space grid. N is related to the rank of the decomposed mixed-domain matrix 12, which is usually significantly smaller than N_x. Note that the number of FFTs N also depends on the given error level of lowrank decomposition with a predetermined Δt. Thus a complex model or increasing the time interval size Δt may increase the rank of the approximation matrix and correspondingly N. In the numerical examples in this paper, the values of rank are usually between 2 and 4. Lowrank decomposition saves cost in calculating equations 9 and 10. We propose to apply it for seismic wave extrapolation in variable velocity and density media on a staggered grid. We call this method *staggered grid lowrank*(SGL) method.

Lowrank FD for first-order extrapolation operators

Approximation 12 can also be used to design accurate finite-difference schemes. Here we extend the lowrank finite-difference method (28) to first-order kx-space operators. Note that $W_x(\mathbf{x}_n, \mathbf{k})$ in equation 13 is a matrix related only to wavenumber $\mathbf{k}$. It can be further decomposed as follows:

$$W_x(\mathbf{x}_n, \mathbf{k}) \approx \sum_{l=1}^{L} C(\mathbf{x}_n, \xi_l) B(\xi_l, \mathbf{k}), \tag{14}$$

where B is an $L \times N_x$ matrix. Specifically, we can define $B(\xi_l, \mathbf{k})$ to take the form of $\sin(\sum_{j=1}^{3} \xi_l^j k_j \Delta_j)$, in which ξ_l^j is the j-th component of a 3-D vector, $\xi_l = (\xi_l^1, \xi_l^2, \xi_l^3)$, k_j is the j-th component of wavenumber $\mathbf{k}$, Δ_j is the space grid size in the j-th direction, $j = 1, 2, 3$ corresponds to x, y, z direction in space. C is the matrix product of W_x and the pseudo-inverse of B. If we define

$$G(\mathbf{x}, l) = \sum_{m=1}^{M} \sum_{n=1}^{N} W_x(\mathbf{x}, \mathbf{k}_m) a_{mn} C(\mathbf{x}_n, \xi_l), \tag{15}$$

then equation 13 can be described as

$$\begin{aligned}
\frac{\partial p(\mathbf{x}, t)}{\partial^+ x} &\approx \sum_{l=1}^{L} G(\mathbf{x}, l) \mathbf{F}^{-1} \left[i e^{ik_x \Delta x/2} B(\mathbf{x}_n, \xi_l) \mathbf{F} \left[p(\mathbf{x}, t) \right] \right] \\
&\approx \sum_{l=1}^{L} G(\mathbf{x}, l) \mathbf{F}^{-1} \left[i e^{ik_x \Delta x/2} \sin(\sum_{j=1}^{3} \xi_l^j k_j \Delta_j) \mathbf{F} \left[p(\mathbf{x}, t) \right] \right] \\
&\approx \frac{1}{2} \sum_{l=1}^{L} G(\mathbf{x}, l) \mathbf{F}^{-1} [e^{ik_x \Delta x/2} (e^{i \sum_{j=1}^{3} \xi_l^j k_j \Delta_j} \\
&\qquad - e^{-i \sum_{j=1}^{3} \xi_l^j k_j \Delta_j}) \mathbf{F} \left[p(\mathbf{x}, t) \right]].
\end{aligned} \tag{16}$$

For a staggered grid, in which pressure is defined on main grid points and partial velocity on half-grid points, we can choose ξ_l as $\xi_l^1 = (2l^1 - 1)/2$, $\xi_l^2 = l^2$, $\xi_l^3 = l^3$ to calculate partial derivative in x-direction, where l^1, l^2, $l^3 = 1, 2, \cdots, L$, and L is the length of the stencil. According to the shift property of FFTs, we can finally obtain the following expression in the space domain:

$$\frac{\partial p(\mathbf{x}, t)}{\partial^+ x} \approx \frac{1}{2} \sum_{l=1}^{L} G(\mathbf{x}, l) [P(\mathbf{x}_R, t) - P(\mathbf{x}_L, t)], \tag{17}$$

where $\mathbf{x}_R = (x_1 + l^1 \Delta_1, x_2 + l^2 \Delta_2, x_3 + l^3 \Delta_3)$ and $\mathbf{x}_L = (x_1 - (l^1 - 1)\Delta_1, x_2 - l^2 \Delta_2, x_3 - l^3 \Delta_3)$.

Equation 17 corresponds to the procedure of finite-difference scheme for calculating the kx-space operator. The vector $\xi_l = (\xi_l^1, \xi_l^2, \xi_l^3)$ provides stencil information, and $G(\mathbf{x}, l)$ stores the corresponding coefficients. A similar derivation can be applied

to the remaining partial derivative operators in equation 9. We call this *staggered grid lowrank finite-differences* (SGLFD) method.

While the SGL method (equation 13) is proposed by applying lowrank approximation to the $kx-$space method on a staggered grid (equation 9), SGLFD (equation 17) is a further approximation of SGL. Theoretically, the SGLFD method using a longer stencil reaches higher accuracy. It is hard to derive stability condition for SGL. However, applying Von Neumann stability analysis, we can easily obtain a sufficient condition of stability for SGLFD as

$$\left| \Delta t v_{max} \sum_{l=1}^{L} G(x,l) sin(\sum_{j=1}^{3} \xi_l^j k_j \Delta_j) \right| \leq 1, \tag{18}$$

where v_{max} is the maximum value of velocity. Once we obtain the finite difference coefficient $G(\mathbf{x}, l)$ for certain velocity $v(\mathbf{x})$ and the predefined parameters Δ_j and Δt, we can use condition 18 to estimate the stability of the SGLFD scheme in equation 17.

Next, we use the plane wave theory to evaluate numerical dispersion for SGLFD method. Inserting the plane wave solution,

$$\begin{aligned} p(\mathbf{x}, t) &= p_0 e^{i\mathbf{k}\cdot\mathbf{x}-\omega t} \\ \mathbf{u}(\mathbf{x}, t) &= \mathbf{u}_0 e^{i\mathbf{k}\cdot\mathbf{x}-\omega t} \end{aligned} \tag{19}$$

into equation 17, and also adopting the dispersion relation $\omega = |\mathbf{k}|v$, the relative error of phase velocity is defined as

$$\varepsilon = \frac{v_{LFD}}{v} - 1 = \frac{1}{\omega \Delta t} \arcsin \left(\sum_{l=1}^{L} G(\mathbf{x}, l) \sum_{j=1}^{3} \sin(\xi_l^j k_j \Delta_j) \right) - 1. \tag{20}$$

The relative error ε describes the numerical dispersion of SGLFD method. If ε equals 0, there is no dispersion. If ε is far from 0, a large dispersion will occur. Here we define the order of SGLFD as that of conventional FD, which has the same stencil length (L). Next we compare the conventional SGFD method with the SGLFD method by the dispersion curves for different orders, time intervals and velocities.

Figure 1 shows the variation of ε with frequency for different order. This figure demonstrates that dispersion decrease with the increase of the order for both SGFD and SGLFD method. Note that for SGFD method increase of order decreases the magnitude of the dispersion error without increasing the area where ε nearly equals 0. Compared with the SGFD method, the SGLFD method is high accurate in a wider range of wavenumber. Figure 2 shows the variation of the ε with frequency for difference time interval. From this figure, we can see that the dispersion becomes stronger when the SGFD method uses larger time interval. Moreover, if a large time interval is used, like $\Delta t = 2.5ms$ in this example, the SGFD method will be unstable. However, for the SGLFD method, its dispersion mainly depends on the frequency. Compared with the SGFD method, the SGLFD method keeps high accuracy for different time intervals (up to 70% of the Nyquist frequency). Figure 3 illustrates the

effect of velocity on dispersion. Note that for the SGFD method, its dispersion curves change greatly with the variation of velocity. Compared with the SGFD method, the SGLFD method is more stable and accurate in a wider range of frequency(up to 70% of the Nyquist frequency). In the previous examples, we used least squares to fit to nearly the 67% of the Nyquist frequency.

COMPARISON OF ACCURACY BETWEEN THE CONVENTIONAL SGFD METHOD AND THE LOWRANK METHODS

The study of the accuracy is important especially for the heterogeneous media. However, for lowrank methods, it is hard to derive the theoretical accuracy order as what the Taylor-series expansion based SGFD method usually did. In this section, we compare the accuracy of the conventional SGFD method and the new lowrank methods numerically. We focus on the variable velocity case in the following analysis.

A simple illustration of the accuracy

We first use a simple 1-D example shown in Figure 4 to demonstrate the accuracy of the SGL method and SGLFD method when they are used to calculate the partial derivatives in equation 9. The velocity increases linearly from 1000 to 2275 m/s. The rank is 2 for lowrank decomposition, assuming 1 ms time step. The exact kx-space operator $\partial/\partial^+ x$ in equation 9 is shown in Figure 4(a). Figures 4(b), 4(c) and 4(d) display errors of approximation operators of SGL, SGLFD and conventional staggered grid finite-difference (SGFD), respectively. Figure 5 shows the middle column of the error matrix. The errors of SGL and SGLFD are significantly smaller than that of SGFD.

Error analysis using the Method of Manufactured Solutions

Next we use the method of manufactured solutions (MMS) (26) to analyze numerical error of the proposed methods. MMS provides an approach to designing exact reference solutions for wave equations in heterogeneous media. We couple the use of manufactured solutions with mesh refinement to plot the error curve for each method. To begin our investigation, we consider the system of 1D first order wave equations,

$$\begin{aligned} \frac{\partial u(x,t)}{\partial t} &= -\frac{1}{\rho(x)}\frac{\partial p(x,t)}{\partial x} + s_u(x,t), \\ \frac{\partial p(x,t)}{\partial t} &= -\rho(x)v^2(x)\frac{\partial u(x,t)}{\partial x} + s_p(x,t), \end{aligned} \tag{21}$$

which are subjected to appropriate boundary and initial conditions, where s_u and s_p represent inject source terms for the particle velocity u and pressure p, respectively.

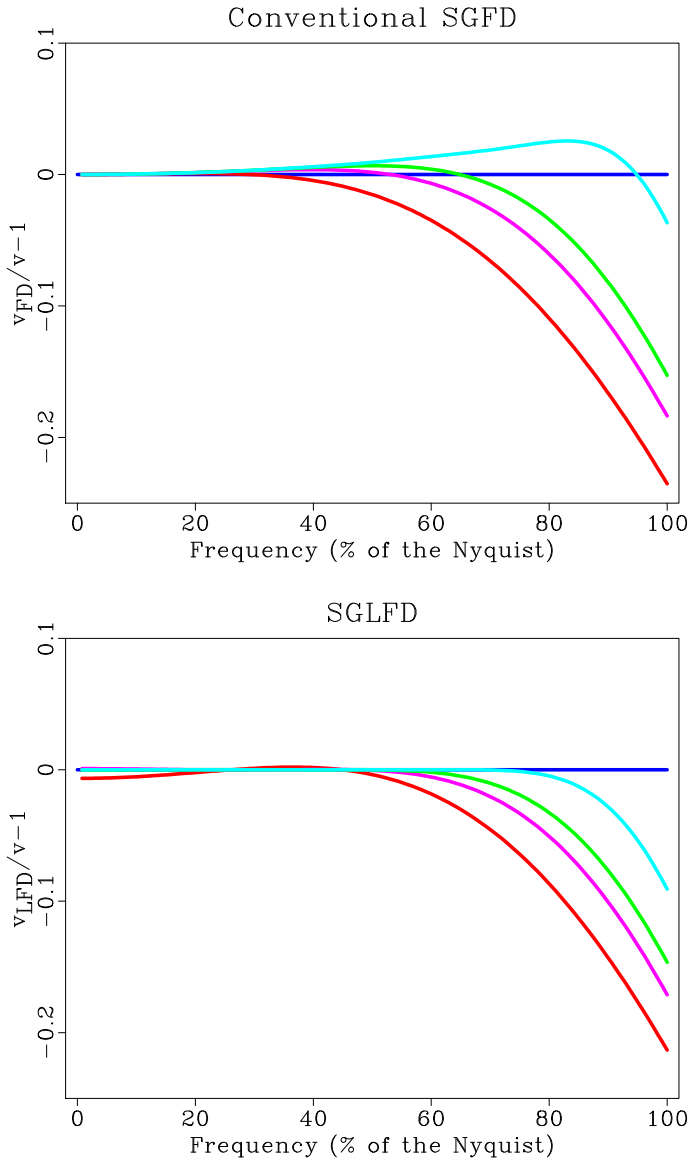


Figure 1: Plot of 1-D dispersion curves of (a) the conventional SGFD method and (b) the SGLFD method for different orders, $4th-$order(red, $2L = 4$), $6th-$order(pink, $2L = 6$), $8th-$order(green, $2L = 8$), $16th-$order(blue, $2L = 16$), time interval $\Delta t = 1ms$, space interval $\Delta_x = 10m$, velocity $v = 3000m/s$. sglowrank/dispersion Mfd,Mlr

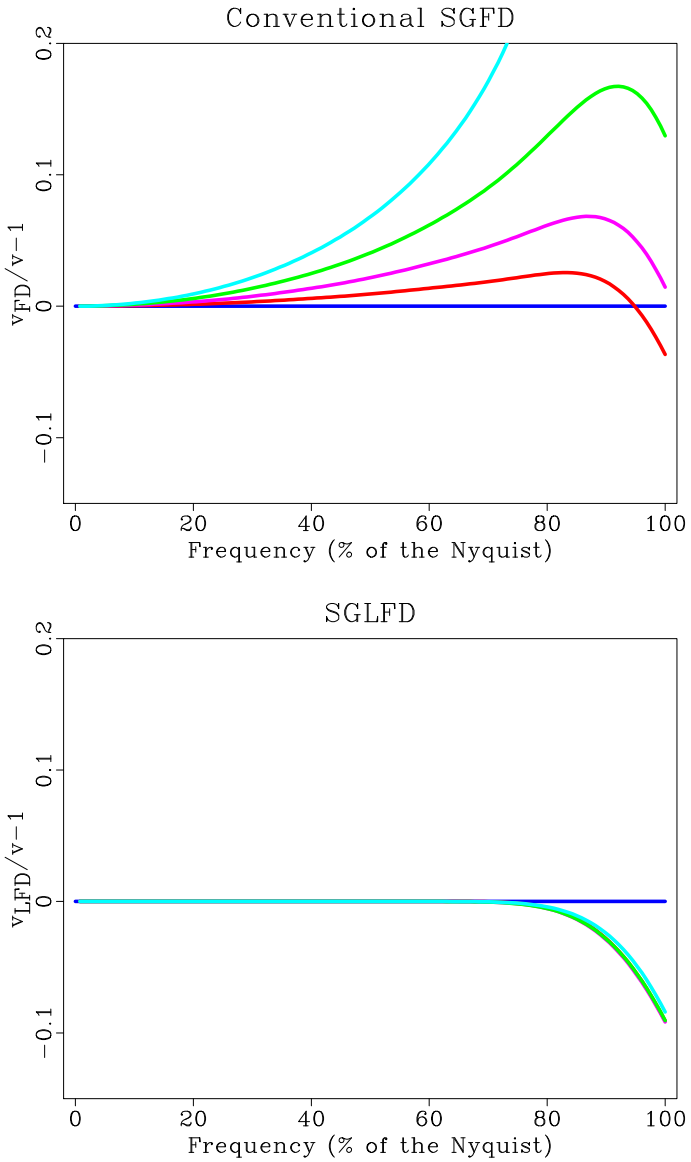


Figure 2: Plot of 1-D dispersion curves of (a) the conventional SGFD method and (b) the SGLFD method for different time interval, $\Delta t = 1ms$(red), $\Delta t = 1.5ms$(pink), $\Delta t = 2ms$(green), $\Delta t = 2.5ms$(blue), $\Delta x = 10m$, $v = 3000m/s$, $2L = 16$. sglowrank/dispersion Mfdt,Mlrt

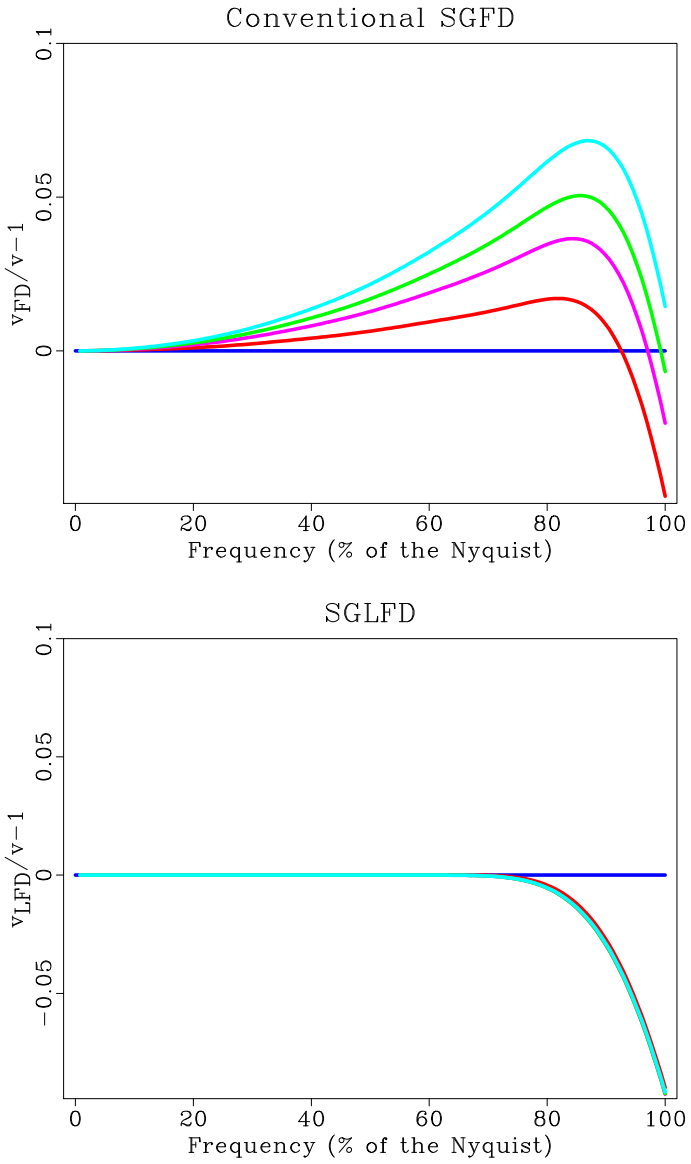


Figure 3: Plot of 1-D dispersion curves of (a) the conventional SGFD method and (b) the SGLFD method for different velocity, $v = 2500m/s$(red), $v = 3500m/s$(pink), $v = 4000m/s$(green), $\Delta t = 4500m/s$(blue), $\Delta t = 1ms$, $\Delta x = 10m$, $2L = 16$. sglowrank/dispersion Mfdv,Mlrv

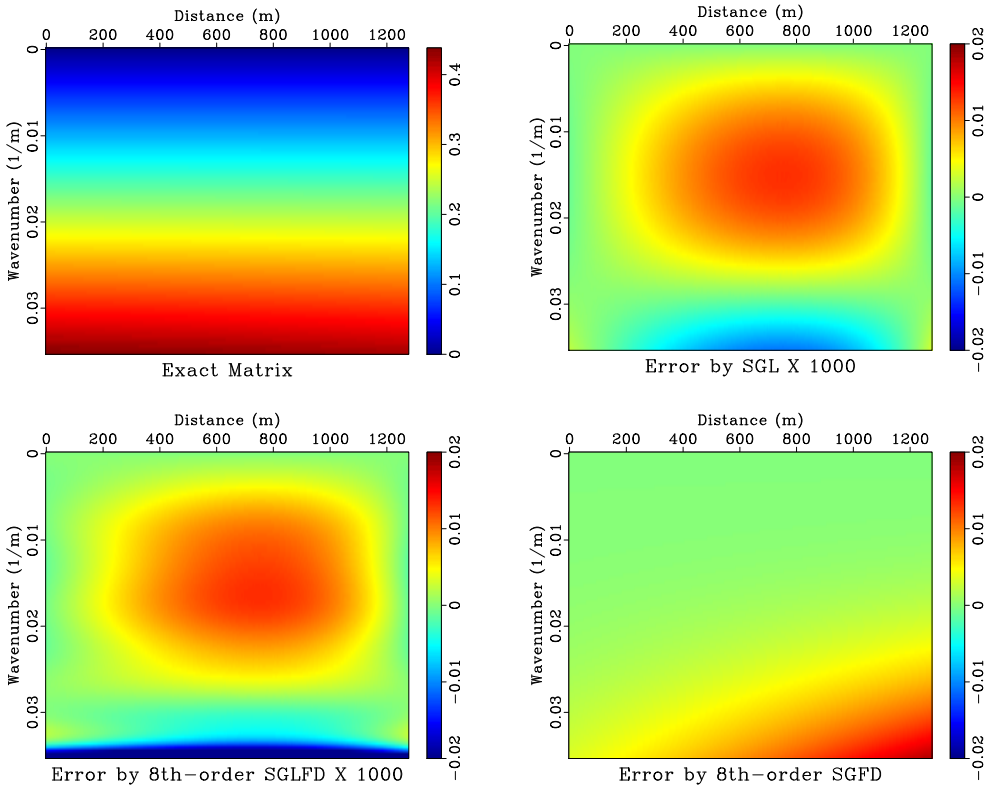


Figure 4: (a) kx-space operator $\partial/\partial^+ x$ for 1-D linearly increasing velocity model. (b) Error of SGL operator. (c) Error of 8th order SGLFD operator. (d) Error of 8th order SGFD operator. sglowrank/oned Mexact,Mlrerr,Mapperr,Mfd10err

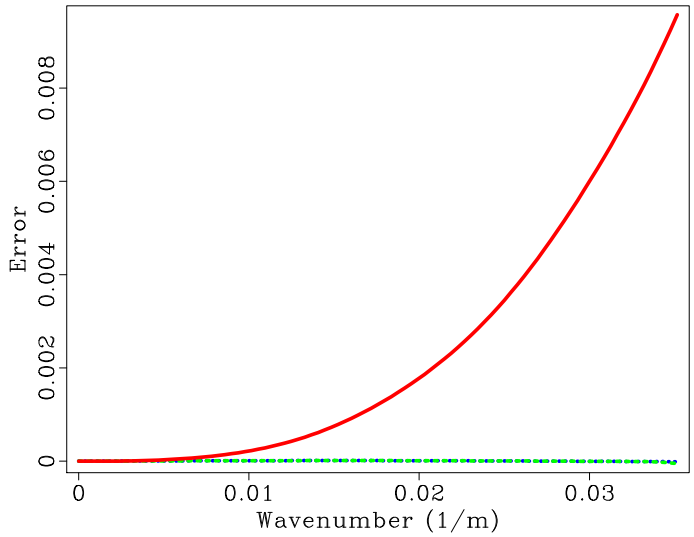


Figure 5: Middle column of the error matrix. Blue dashed line: SGL operator. Green dotted line: the 8th order SGLFD operator. Red solid line: the 8th order SGFD operator. sglowrank/oned slicel

For this study, we are interested in the governing equation. We choose a Gaussian pulse as the solution of the pressure and particle velocity in variable velocity media (22),

$$u(x,t) = p(x,t) = e^{-\lambda^2(x-x_0-v(x)t)^2}, \tag{22}$$

where λ is the wavelength of the Gaussian pulse. x_0 is the source location and $v(x)$ is the variable velocity. It is easy to derive the corresponding source terms from equations 21 and 22,

$$\begin{aligned} s_u(x,t) &= 2\lambda^2[x-x_0-v(x)t][v(x)+\frac{v'(x)t-1}{\rho(x)}]p(x,t), \\ s_p(x,t) &= 2\lambda^2[x-x_0-v(x)t][v(x)+\rho(x)v^2(x)(v'(x)t-1)]p(x,t). \end{aligned} \tag{23}$$

For numerical solution, accuracy is often affected by various factors. In this study, we examine the numerical error of the proposed methods when using different temporal and spatial discretization and wavelets of different dominant frequencies. The velocity we used here increases with x, defined as $v(x) = 2.1 + 0.1x^2(km/s)$. Its gradient is $v'(x) = 0.2x$ correspondingly. We use constant density in this experiment. Figure 6(a) shows the RMS errors of the wavefields for different time interval. For all these calculations, we keep space interval $\Delta x = 25m$ as constant. The blue dashed line and green dotted line indicate the errors of SGL method and SGLFD method respectively, while red solid line plots the errors of the traditional SGFD method. Compared with the traditional SGFD method, the proposed lowrank methods correct the distortion caused by increasing Δt. They exhibit a high accuracy in time. Figure 6(b) shows the recorded trace at the x of $4km$ for $\Delta t = 4ms$.The black solid line corresponds to the exact solution generated by MMS. The colors of remaining lines has the same meaning as in Figure 6(a). Numerical dispersion is more visible when Δt is increased; this effect is much less significant in our methods.

Figure 7(a) shows the RMS errors of the wavefields for different space intervals. We use a small time interval $\Delta t = 0.4ms$ to keep time error small. Figure 7(b) shows the recorded data for $\Delta x = 40m$. The first-order kx-space operators (equation 9) have a high accuracy in space which make the error of SGL method increase slowly when Δx increase. The coefficients of the SGLFD are obtained by applying least-squares fitting to the SGL method. The error from the least-squares fitting make SGLFD method less accurate than SGL method. On the other hand, the coefficients of lowrank FD are optimized and auto-adapted to variations in velocity, which makes it significantly more accurate than SGFD method.

In applications of seismic wave extrapolation, temporal and spatial intervals are often increased at a certain ratio for saving computational cost. Figure 8 show the error curves for a temporal and spatial refinement study where both Δx and Δt are increased simultaneously. We define the Courant-Friedrichs-Lewy (CFL) number as $\alpha = v_{max}\Delta t/\Delta x$ to specify the stability condition, where v_{max} is the maximum velocity. Figure 8(a) shows the RMS errors for $\alpha = 0.2$. Figure 8(b) shows the RMS error of $\alpha = 0.4$, while figure 8(c) for $\alpha = 0.8$. The color of the lines has

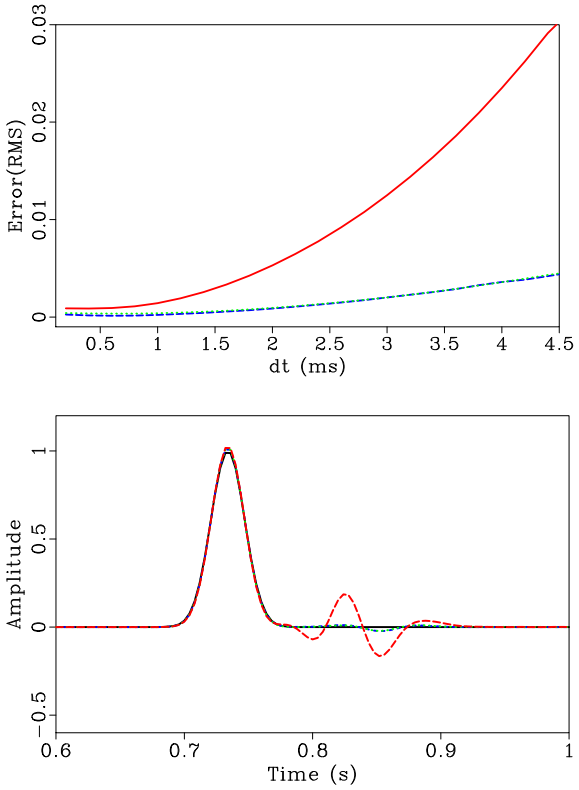


Figure 6: (a) RMS errors as function of time interval. (b) Recorded data for $\Delta t = 4ms$. Blue dashed line: SGL method. Green dotted line: SGLFD method. Red solid line: SGFD method. Black solid line: MMS. sglowrank/mms1d-dt-vv et,trec19

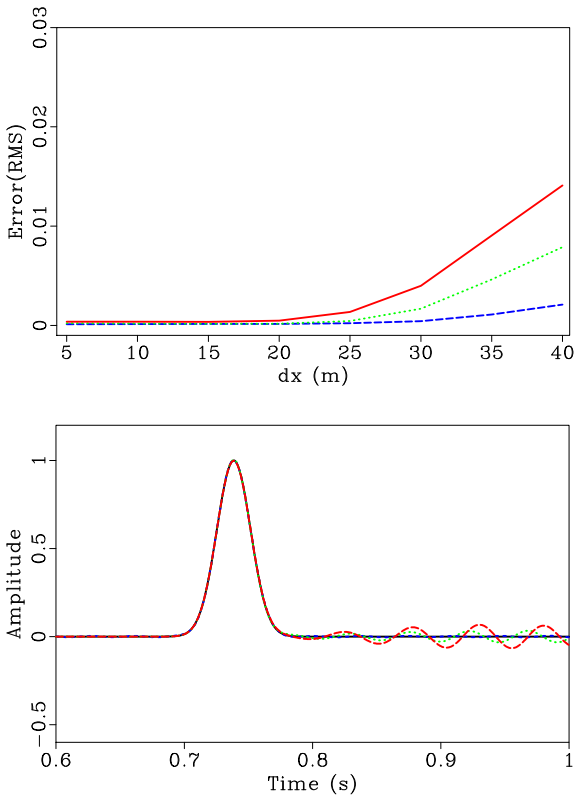


Figure 7: (a) RMS errors as function of space interval. (b) Recorded data for $\Delta x = 40m$. Blue dashed line: SGL method. Green dotted line: SGLFD method. Red solid line: SGFD method. Black solid line: MMS. sglowrank/mms1d-dx-vv em,mrec7

the same meaning as indicated before. We can see that the errors increase with the CFL number α. This increase is especially significant for SGFD method. Both SGL and SGLFD methods keep high accuracy for a larger scale of temporal and spatial intervals than SGFD.

Seismic exploration techniques, such as prestack depth migration or FWI, need the modeling engine which remains highly accurate for a wide band seismic wavelet. The numerical dispersion can be serious for a high frequency source. Figure 9(a) compares the accuracy of different methods for different source frequencies. The meanings of different line colors are the same as before. Figure 9(b) shows the recorded data with source frequency of $40Hz$. It can be seen that the SGFD method has a visible dispersion, while SGL and SGLFD methods remain almost dispersion free.

The accuracy estimation for planar interface

For seismic modeling, it is important to estimate the accuracy of the proposed methods for heterogeneous media, especially for the amplitude variation with offset(AVO) or amplitude variation with angle(AVA) effects of reflected or transmitted wavefields along the interface. The theoretical analysis of this feature can be complicated. Here, we provide a simple numerical test to illustrate the accuracy of the proposed methods. We design a planar interface model which is defined on a grid system of 601×501 with a space interval of $10m$ in both horizontal and vertical direction, as shown in figure 10. The planar interface is aligned with the vertical $251th$ grid. The velocities of upper and lower layer are designed as $4000m/s$ and $2000m/s$ to avoid critical reflection. We use constant density $\rho = 1700kg/m^3$. We synthesize a shot record to examine the accuracy of our methods. The source is located at the position of $3000m$ in horizontal direction and $1500m$ in vertical direction. Thus the maximum incident angle is $71.6°$. We place two receiver lines above and below the interface and measure the amplitudes of incident, reflected and transmitted wavefields. The reflection coefficient is given by the ratio of amplitudes of the reflected wavefield and the incident wavefield. The transmission coefficient is given by the ratio of amplitudes of transmitted wavefield and reflected wavefield. Figure 11 compares the reflection and transmission coefficients calculated by the SGL and SGLFD methods with the theoretical values calculated by solving Zoeppritz equations. From this figure, we see that the reflection and transmission coefficients calculated by the SGL and SGLFD method match well with the theoretical values. Thus, both the SGL and SGLFD methods appear sufficiently accurate to provide correct dynamic information of wavefields.

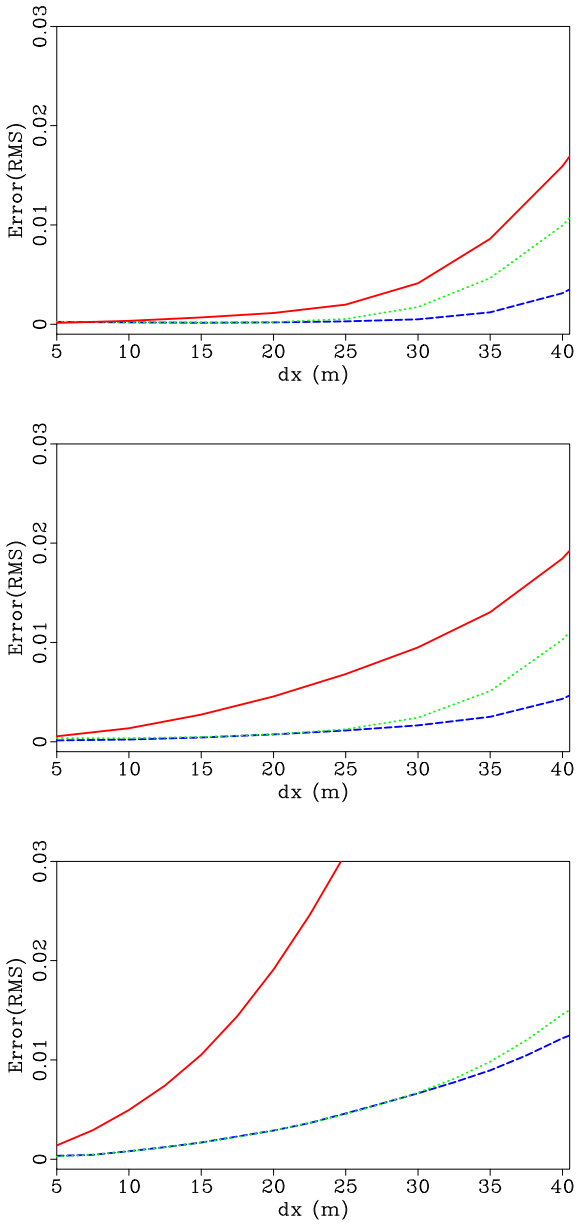


Figure 8: RMS errors for CFL number (a) $\alpha = 0.2$, (b) $\alpha = 0.4$, (c) $\alpha = 0.8$. Blue dashed line: SGL method. Green dotted line: SGLFD method. Red solid line: SGFD method. sglowrank/mms1d-all-vv ea2,ea4,ea8

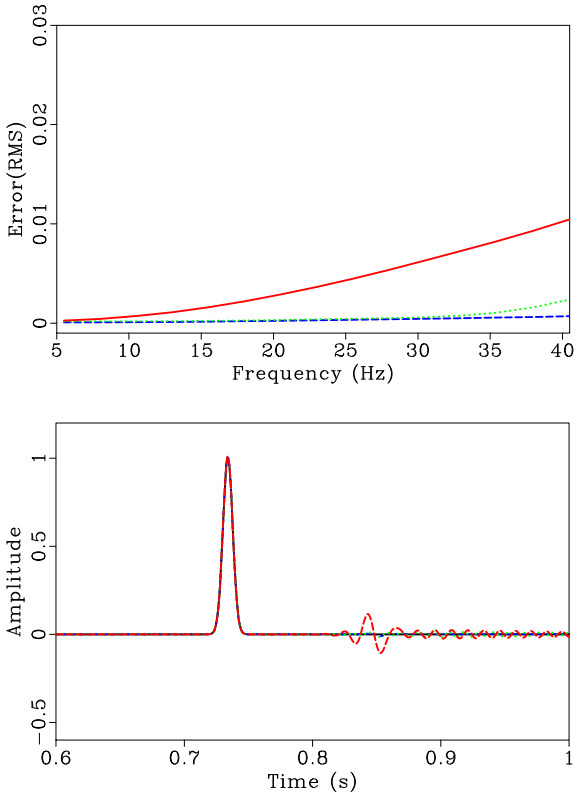


Figure 9: (a) RMS errors as function of dominant frequency of source wavelet. (b) Recorded data for frequency of $40Hz$. Blue dashed line: SGL method. Green dotted line: SGLFD method. Red solid line: SGFD method. Black solid line: MMS. sglowrank/mms1d-fre fre,frec15

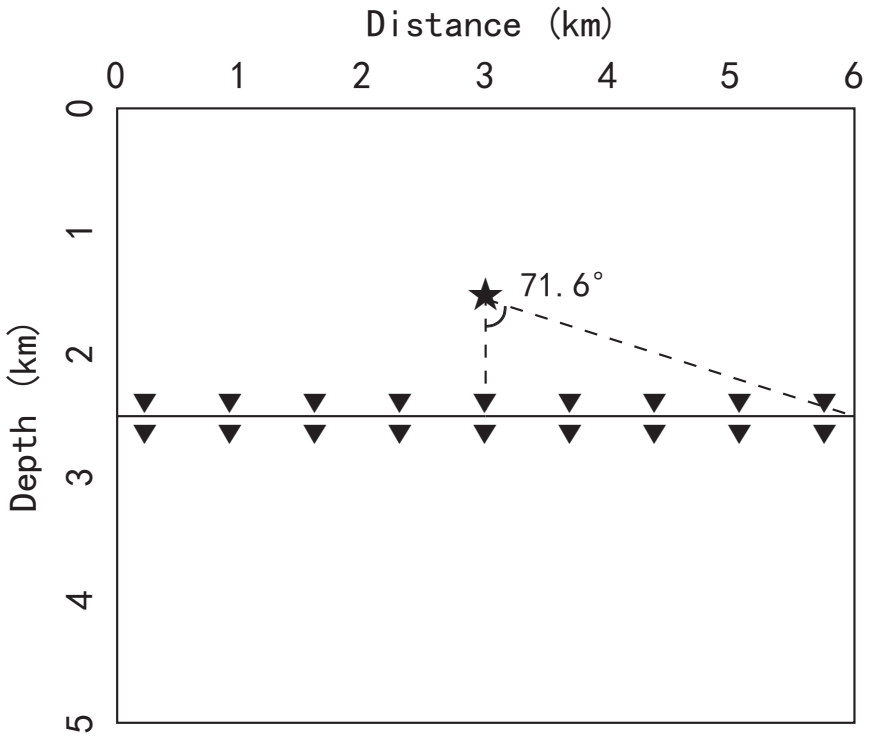


Figure 10: The geometry of the planar interface model. The star denotes the source location and the triangles denote the receiver locations. The values of incident angle along the planar interface are between 0 and 71.6 degree. sglowrank/. geo

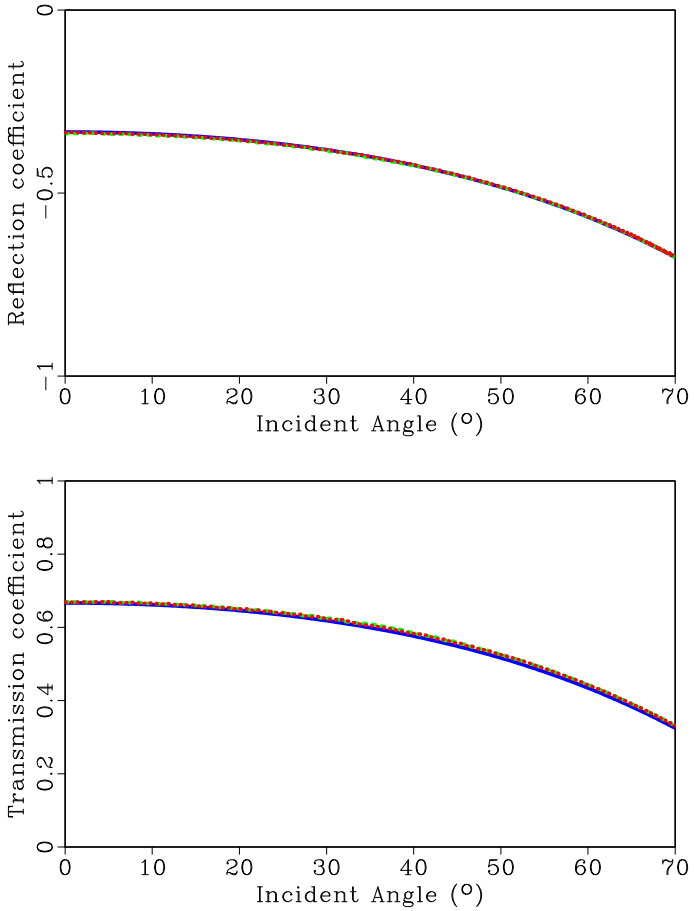


Figure 11: Comparison of reflection (a) and transmission (b) coefficients calculated by the SGL method (red dashed line) and the SGLFD method (green dashed line) with the theoretical values (blue solid line). sglowrank/avo rpp,tpp

NUMERICAL EXAMPLES

2D examples of a two-layer model

To test the performance of the proposed methods with a rough velocity model, we use a two-layer model with high velocity and density contrasts. The model is defined on a 501×501 grid, with $\Delta x = \Delta z = 10m$ and $\Delta t = 1.5ms$. Velocities of the upper and lower layers are $1.3km/s$ and $3.2km/s$. The densities of upper and lower layers are $1.7g/cm^3$ and $2.7g/cm^3$ respectively. A point source of a Ricker-wavelet with dominant frequency of 20 Hz is located in the center of the model at a depth of $0.2km$.The maximum frequency (f_{max}) is around $60Hz$. Following (28), we still use the CFL number α to specify the stability and define dispersion factor as $\beta = v_{min}/(f_{max}\Delta x)$ to indicate the sample points per wavelength, where v_{min} is the minimum velocity of the model. For modeling with above parameters, $\alpha = 0.32$ and $\beta = 2.2$.

Figure 12 shows a wavefield snapshot generated by lowrank FD with a time interval equal to $2ms$. At this time interval, the SGFD method becomes unstable.

Example of BP model

Finally we test the proposed methods in a complex velocity model. Figure 13 shows a part of the BP-2004 model, which is a complicated model containing a salt body and sharp velocity and density contrasts on the flanks of the salt body (1). We use a Ricker-wavelet at a point source with a dominant frequency of $20Hz$ ($f_{max} \approx 60Hz$). Both of the horizontal grid size Δx and vertical grid size Δz are $12.5m$, and the time step is $1.5ms$, thus, $\alpha \approx 0.57$ and $\beta \approx 1.98$. Figure 14 shows the wavefield snapshot in the BP model generated by SGL method and 6th order SGLFD with a rank of 3. This experiment confirms that the staggered grid lowrank methods are able to handle sharp velocity and density variations.

CONCLUSIONS

FD methods and spectral methods are the two most popular wavefield extrapolation approaches for seismic modeling and seismic wavefield imaging. To handle variable density and velocity in seismic modeling or RTM, we have proposed SGL method by applying lowrank decomposition to first-order kx-space propagation operators on a staggered grid. The cost of the new method amounts to using a small number of FFTs, which corresponds to the approximation rank. On the basis of the SGL method, we have also designed the SGLFD method, which extends the lowrank finite-differences from a case of constant density to one of variable density. This approach promises higher accuracy and better stability than those of the traditional, explicit staggered grid finite-difference methods. We tested the proposed methods using the MMS solutions and concluded that they retained high accuracy for large temporal and

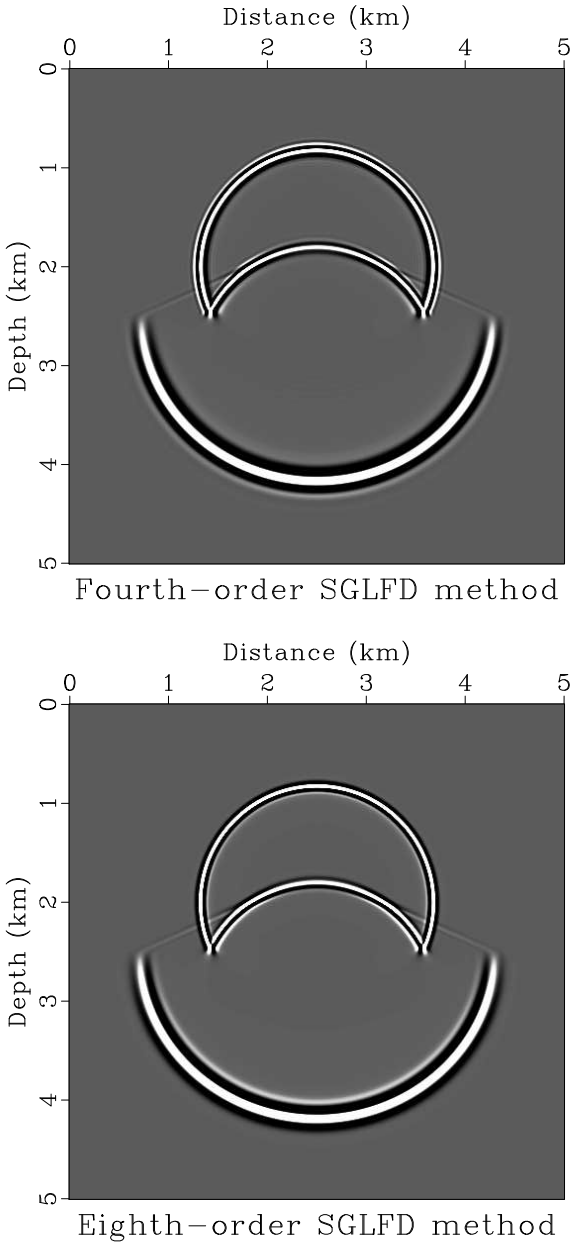


Figure 12: Wavefield snapshot in a two-layer model with variable density and velocity using (a) 4th order SGLFD method and (b) 8th order SGLFD method. sglowrank/tlayer3 lfd4snap2,lfd8snap2

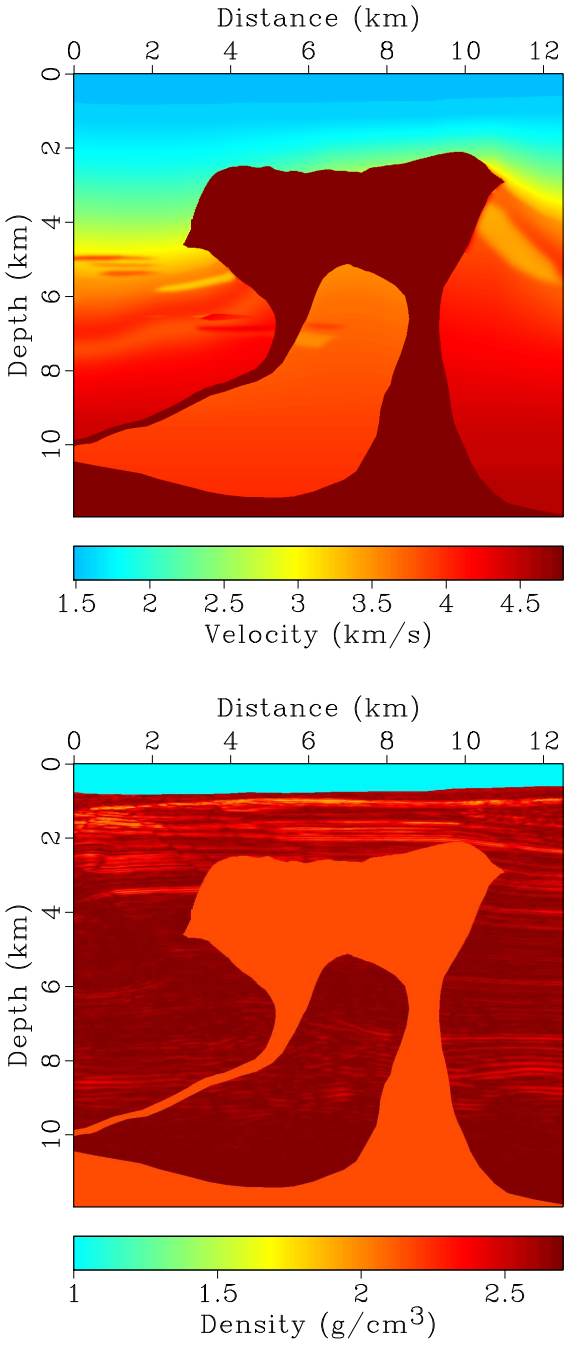


Figure 13: BP model: (a) velocity, (b) density. sglowrank/bp3 vel,den

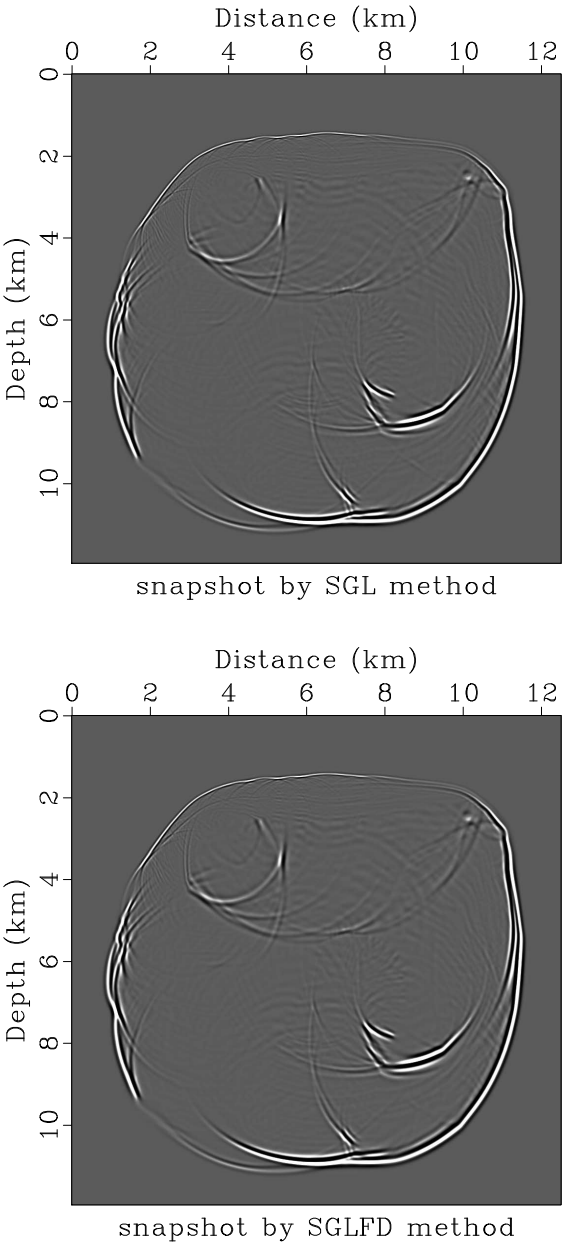


Figure 14: Wavefield snapshot modeled through for BP Model using (a) SGL method and (b) SGLFD method. sglowrank/bp3 lrsnap,lfdsnap

spatial intervals or high frequency sources. Results for synthetic models illustrate that our proposed methods are highly accurate for heterogeneous media and can handle sharp velocity and density variations. Although the proposed methods are focused on the acoustic case, they can also be extended in principle to elastic, anisotropic or attenuating media. The methods can be used for seismic modeling or RTM.

ACKNOWLEDGMENTS

We appreciate the support of the China State Scholarship Fund, the Natural Science Foundation of China (41074087, 41174100), the Fundamental Research Funds for the Central Universities(11CX06002A). We thank KAUST and the Texas Consortium for Computational Seismology (TCCS) for partial financial support. We also thank all developers of the Madagascar open-source software package (`http://ahay.org`). We thank TACC (Texas Advanced Computing Center) for providing supercomputing resources used in this study.

REFERENCES

[1] Billette, F. J., and S. Brandsberg-Dahl, 2004, The 2004 BP velocity benchmark: 67th Annual EAGE Meeting, EAGE, Expanded Abstracts, Soc. Expl. Geophys., B305.

[2] Bohlen, T., 2002, Parallel 3-d viscoelastic finite difference seismic modelling: Computers & Geosciences, **28**, 887–899.

[3] Chu, C., and P. L. Stoffa, 2011, Application of normalized pseudo-Laplacian to elastic wave modeling on staggered grids: Geophysics, **76**, T113–T121.

[4] Chu, C., P. L. Stoffa, and R. Seif, 2009, 3d elastic wave modeling using modified high-order time stepping schemes with improved stability condition: 79th Ann. Internat. Mtg., Soc. Expl. Geophys., 2662–2666.

[5] Dablain, M. A., 1986, The application of high-order differencing to the scalar wave equation: Geophysics, **51**, 54–66.

[6] Du, X., P. Fowler, and R. Fletcher, 2014, Recursive integral time-extrapolation methods for waves: A comparative review: Geophysics, **79**, T9–T26.

[7] Engquist, B., and L. Ying, 2009, A fast directional algorithm for high frequency acoustic scattering in two dimensions: Commun. Math. Sci., **7**, 327–345.

[8] Etgen, J., 1986, High order finite-difference reverse time migration with the two way nonreflecting wave equation: SEP-48, 133–146.

[9] Etgen, J., and S. Brandsberg-Dahl, 2009, The pseudo-analytical method: application of pseudo-Laplacians to acoustic and acoustic anisotropic wave propagation: 79nd Annual International Meeting, SEG, Expanded Abstracts, 2552–2556.

[10] Finkelstein, B., and R. Kastner, 2007, Finite difference time domain dispersion reduction schemes: Journal of Computational Physics, **221**, 422–438.

[11] Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013a, Madagascar:

open-source software project for multidimensional data analysis and reproducible computational experiments: Journal of Open Research Software, **1**, e8.

[12] Fomel, S., L. Ying, and X. Song, 2010, Seismic wave extrapolation using a lowrank symbol approximation: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3092–3096.

[13] ———, 2013b, Seismic wave extrapolation using lowrank symbol approximation: Geophysical Prospecting, **61**, 526–536.

[14] Fowler, P., X. Du, and R. Fletcher, 2010, Recursive integral time extrapolation methods for scalar waves: 80th Ann. Internat. Mtg., Soc. Expl. Geophys., 3210–3215.

[15] Kindelan, M., A. Kamel, and P. Sguazzero, 1990, On the construction and efficiency of staggered numerical differentiators for the wave equation: Geophysics, **55**, 107–110.

[16] Levander, A. R., 1988, Fourth-order finite-difference p-sv seismograms: Geophysics, **53**, 1425–1436.

[17] Liu, Y., 2013, Globally optimal finite-difference schemes based on least squares: Geophysics, **78**, T113–T132.

[18] Liu, Y., and M. K. Sen, 2009, An implicit staggered-grid finite-difference method for seismic modelling: Geophysical Journal International, **179**, 459–474.

[19] Madariaga, R., 1976, Dynamics of an expanding circular fault: Bulletin of the Seismological Society of America, **66**, 639–666.

[20] Mast, T. D., L. P. Souriau, D. Liu, M. Tabei, A. I. Nachman, and R. C. Waag, 2001, A k-space method for large-scale models of wave propagation in tissue: IEEE Trans. Ultrason. Ferroelect. Freq. Contr., **48**, 341–354.

[21] Moczo, P., J. Kristek, and V. Vavrycuk, 2002, 3d heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli and densities: Bulletin of the Seismological Society of America, **92**, 3042–3066.

[22] Ober, C. C., S. S. Collis, B. B. Waanders, and C. Marcinkovich, 2009, Method of manufactured solutions for the acoustic wave equation: 79th Ann. Internat. Mtg., Soc. Expl. Geophys., 3615–3619.

[23] Operto, S., J. Virieux, P. Amestoy, J. Y. Amestoy, J. Y. Giraud, and H. B. H. Ali, 2007, 3d finite-difference frequency-domain modeling of visco-acoustic wave propagation using a massively parallel direct solver: A feasibility study: Geophysics, **72**, SM195–SM211.

[24] Reshef, M., D. Kosloff, M. Edwards, and C. Hsiung, 1988, Three-dimensional acoustic modeling by the Fourier method: Geophysics, **53**, 1175–1183.

[25] Robertsson, J. O. A. ., J. O. Blanch, and W. W. Symes, 1994, Viscoelastic finite-difference modeling: Geophysics, **59**, 1444–1456.

[26] Salari, K., and P. Knupp, 2000, Code verification by the method of manufactured solutions: Sandia National Labs., Albuquerque, NM (US); Sandia National Labs., Livermore, CA (US), No. SAND2000–1444.

[27] Song, X., and S. Fomel, 2011, Fourier finite-difference wave propagation: Geophysics, **76**, T123–T129.

[28] Song, X., S. Fomel, and L. Ying, 2013, Lowrank finite-differences and lowrank

fourier finite-differences for seismic wave extrapolation: Geophysical Journal International, **193**, 960–969.
[29] Song, X., K. Nihei, and J. Stefani, 2012, Seismic modeling in acoustic variable-density media by fourier finite differences: 82th Ann. Internat. Mtg., Soc. Expl. Geophys., 1–6.
[30] Soubaras, R., and Y. Zhang, 2008, Two-step explicit marching method for reverse time migration: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2272–2276.
[31] Tabei, M., T. D. Mast, and R. C. Waag, 2002, A k-space method for coupled first-order acoustic propagation equations: J. Acoust. Soc. Am., **111**, 53–63.
[32] Takeuchi, N., and R. J. Geller, 2000, Optimally accurate second order time-domain finite difference scheme for computing synthetic seismograms in 2-D and 3-D media: Phys. Earth Planet. Int., **119**, 99–131.
[33] Tal-Ezer, H., D. Kosloff, and Z. Koren, 1987, An accurate scheme for seismic forward modeling: Geophysical Prosecting, **35**, 479–490.
[34] Virieux, J., 1984, Sh-wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **49**, 1933–1942.
[35] ———, 1986, P-sv wave propagation in heterogeneous media: Velocity-stress finite-difference method: Geophysics, **51**, 889–901.
[36] Wards, B., G. Margrave, and M. Lamoureux, 2008, Phase-shift time-stepping for reverse-time migration: 78th Ann. Internat. Mtg., Soc. Expl. Geophys., 2262–2266.
[37] Wu, W., L. R. Lines, and H. Lu, 1996, Analysis of high-order, finite-difference schemes in 3-d reverse-time migration: Geophysics, 845–856.
[38] Zhang, Y., and G. Zhang, 2009, one-step extrapolation method for reverse time migration: Geophysics, **74**, A29–A33.
[39] Zhang, Y., G. Zhang, D. Yingst, and J. Sun, 2007, Explicit marching method for reverse time migration: 76th Ann. Internat. Mtg., Soc. Expl. Geophys., 2300–2304.

Deblending using normal moveout and median filtering in common-midpoint gathers

*Yangkang Chen[1], Jiang Yuan[2], Zhaoyu Jin[3], Keling Chen[4] and Lele Zhang[5]**

ABSTRACT

The benefits of simultaneous source acquisition are compromised by the challenges of dealing with intense blending noise. In this paper, we propose a processing workflow for blended data. The incoherent property of blending noise in the common-midpoint (CMP) gathers is utilized for applying median filtering along the spatial direction after normal moveout (NMO) correction. The key step in the proposed workflow is that we need to obtain a precise velocity estimation which is required by the subsequent NMO correction. Because of the intense blending noise, the velocity scan can not be obtained in one step. We can recursively polish both deblended result and velocity estimation by deblending using the updated velocity estimation and velocity scanning using the updated deblended result. We use synthetic and field data examples to demonstrate the performance of the proposed approach. The migrated image of deblended data is cleaner than that of blended data, and is similar to that of unblended data.

INTRODUCTION

The technique of simultaneous-source (sometimes called *multisource*) acquisition means firing more than one shot at nearly the same time regardless of their interference. In conventional acquisition, however, either the temporal shooting intervals or the spatial sampling intervals are large enough so that the interference between successive shots can be ignored. The multisource technique can reduce the acquisition period and at the same time can improve data quality because of the decreased spatial sampling interval (5). The benefits from simultaneous-source acquisition are compromised by the challenges in removing strong blending interference. Because of its economic benefits and technical challenges, this technique has attracted the attention of researchers in both industry and academia (16; 15).

In blended acquisition, more than one source is shot simultaneously, regardless of their interactive interference. The term *source* denotes a shot array, which can contain all the shots in a conventional acquisition system. When more than one source is involved in acquisition, either a denser or a wider shot coverage can be obtained

*__e-mail:__ sergey.fomel@beg.utexas.edu

for a given constant acquisition period. Figure 1(a) depicts two simultaneous sources shooting from the same position towards the same direction. In this case, a two-times denser coverage can be obtained. Figure 1(b) depicts two simultaneous sources far from each other shooting along the same direction, in which case, a two-times wider shot coverage can be obtained. The number of simultaneous sources can be even larger, yielding a much denser and wider shot coverage. The shooting sequence of the shots and the direction of each source can also be variable. It's natural, that the observed data will contain strong interference, and the more simultaneous sources involved, the more severe the interference will be. For the blended acquisition geometry as shown in Figure 1(b) we acquire blended data with strong interference, as shown in Figure 2(b). Figure 2(a) shows the acquired data using conventional acquisition, supposing that the two sources in Figure 1(b) are fired with large-enough time interval.

There are two main ways to deal with the challenges posed by simultaneous-source acquisition. The first is to use a first-separate and second-process strategy (7), which is also known as "deblending" (13). The other is to use direct imaging and waveform inversion by applying some constraints to attenuate the artifacts caused by interference (21; 9). Although the direct imaging approach has achieved some encouraging results, the preferable way so far is still to focus on the separation of blended data into individual sources as if acquired conventionally.

Different filtering and inversion methods have been used previously to deblend seismic data. Filtering methods utilize the property that the coherency of the simultaneous-source data is not the same in different domains, thus we can get the unblended data by filtering out the randomly distributed blending noise in a particular domain, in which one source record is coherent and the other is not (14; 17; 15). One choice is to transform seismic data from the common-shot domain to common-receiver, common-offset or common-midpoint domain. Inversion methods treat the separation problem as an estimation problem that aims at estimating the desired unblended data. Because of the ill-posed property of such estimation problems, a regularization term is usually required (12). (19), (2) and (18) use a sparsity constraint in the Radon domain to regularize the inversion. A sparsity constraint is also used by (1) to minimize the energy of incoherent events present in the blended data. (3) compared two separation techniques for dithered slip-sweep (DSS) data using the sparse inversion method (18) and f-x predictive filtering (6; 8), and found the advantage of inversion methods over random noise attenuation techniques. (20) proposed to distribute all energy in the simultaneous shot records by reconstructing the individual shot records at their respective locations. (17) introduced an iterative estimation and subtraction scheme that combines the properties of filtering and inversion methods and exploits the fact that characteristics of blending noise differs in different domains. In order to deal with the aliasing problem, (4) proposed the alternating projection method (APM), which chooses corrective projections to exploit data characteristics and claims to be less sensitive to aliasing than other approaches.

Median filtering is notable for its ability to remove spiky noise and is also suitable

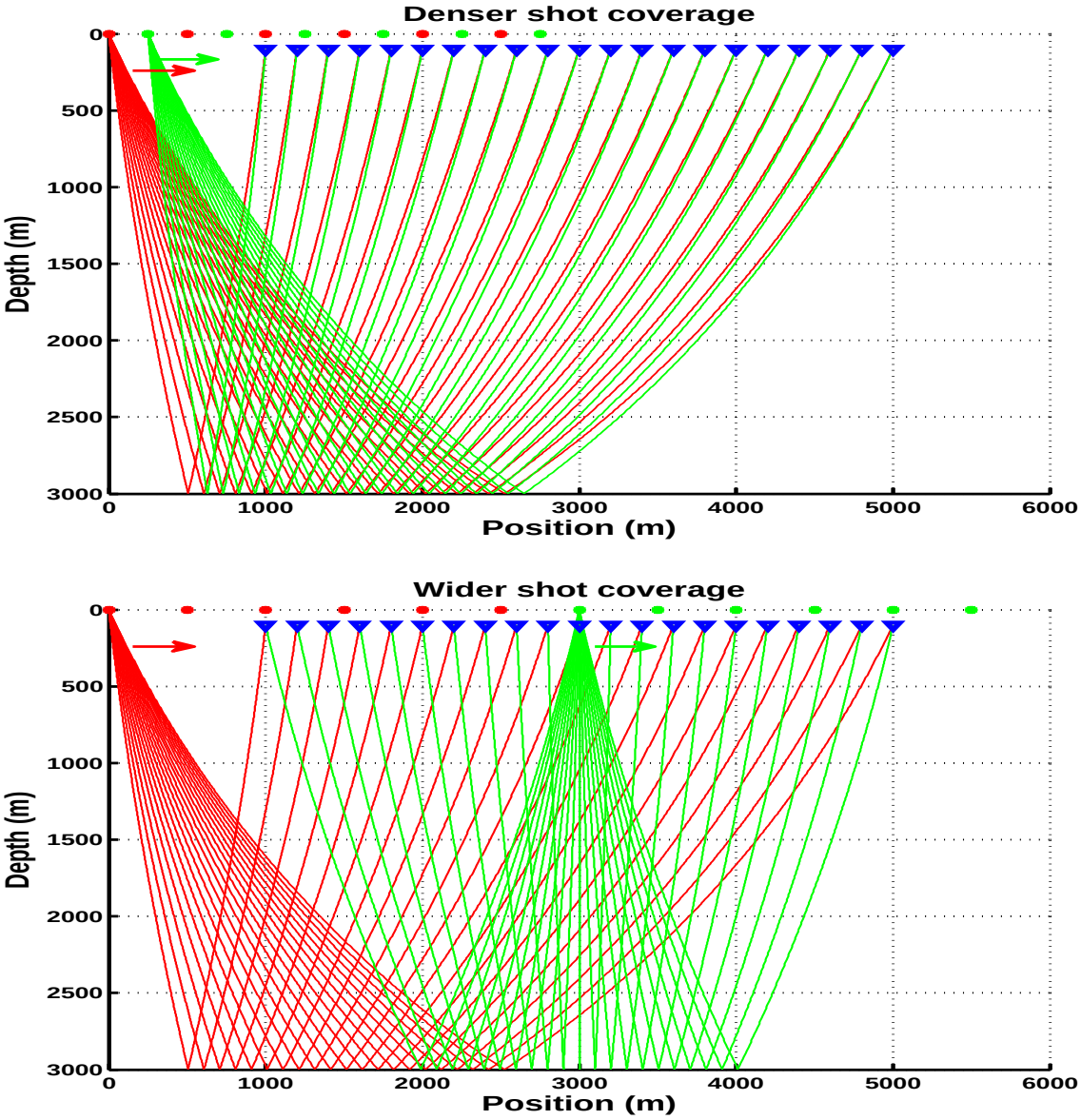


Figure 1: Demonstration for denser shot coverage (a) and wider shot coverage (b). Red points denote shot positions for source 1. Green points denote shot positions for source 2. Blue points denote receiver positions. Red and green strings denote the shooting rays. Arrows denote the shooting directions. nmodeblend/. demo1,demo2

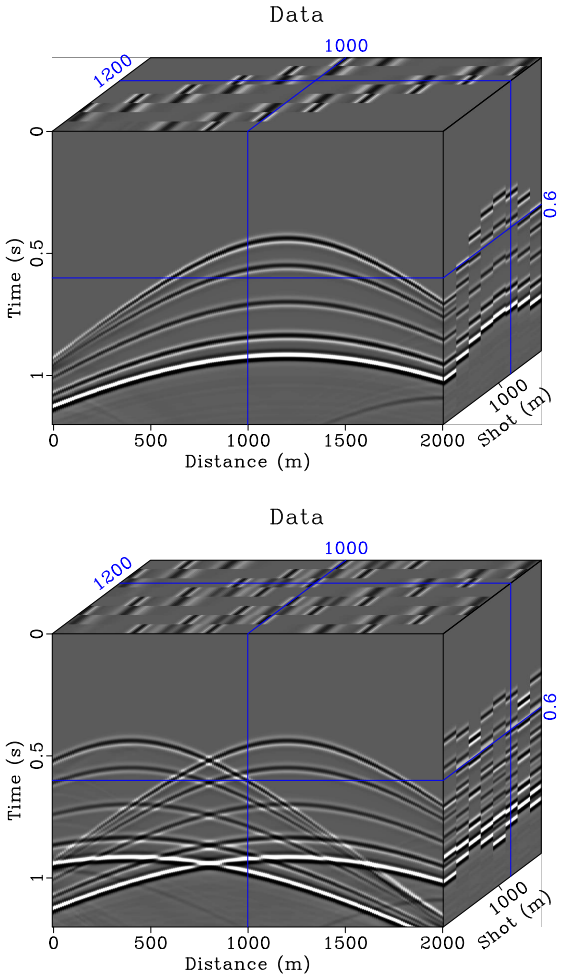


Figure 2: (a) Unblended data. (b) Blended data. nmodeblend/. datademo,blenddemo

to remove blending noise. However, median filtering can be applied only to seismic profiles containing horizontal events, otherwise it will harm much of the useful energy. In this paper, we propose to implement median filtering to attenuate blending noise after normal moveout in common-midpoint (CMP) gathers. The deblending can be inserted into a conventional processing workflow . The benefits of the proposed deblending approach are its easy implementation and efficient improvement for the migrated image without use of iterative deblending, which is extremely time-consuming. Synthetic and field data examples demonstrate the effectiveness of the deblending method and improvement for the final migrated image.

METHOD

Independent marine-streamer simultaneous shooting (IMSSS)

As shown in Figure 3, our blended survey is based on independent marine-streamer simultaneous shooting (IMSSS). Figure 3 shows four independent marine-streamer sources. Note that the number of simultaneous sources need not be limited to four. They shoot independently as in the conventional way and yield their own data. For a 3D seismic survey, the IMSSS can help cover a 2D surface area efficiently. With four sources, the efficiency is increased by four times. The increased efficiency, however, is compromised by interference among the four sources.

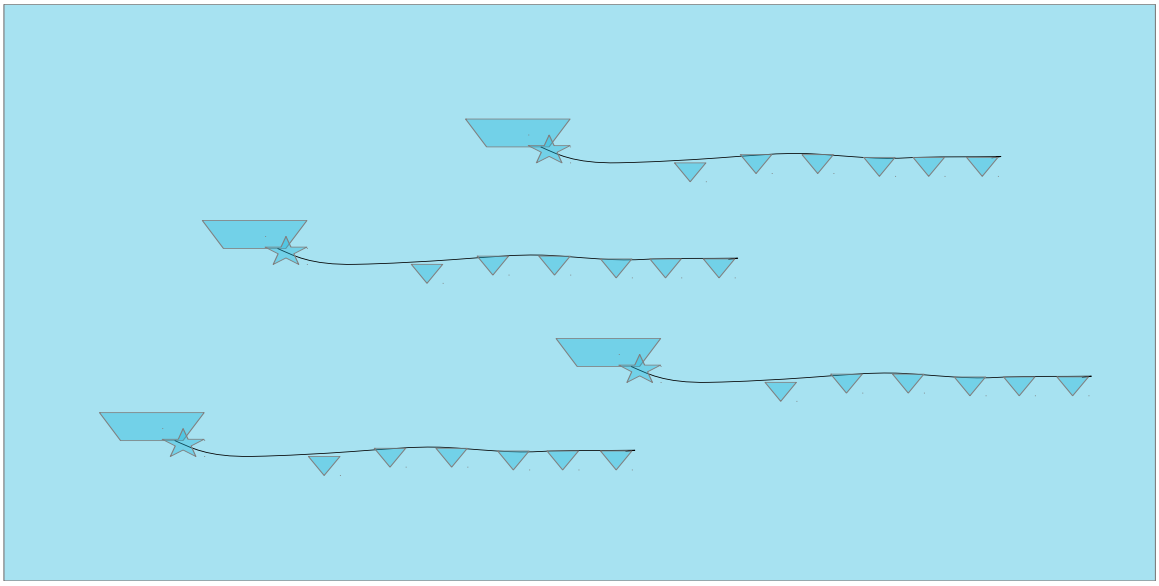


Figure 3: Demonstration of independent marine-streamer simultaneous shooting (IMSSS). nmodeblend/. streamer

Transformation between different domains

The coherency of simultaneous-source data is not the same in different domains. The interference appears coherent in common-shot gathers (CSG); however, the interference turns out to be incoherent when observed in common-midpoint gathers (CMG), common-offset gathers (COG) and common-receiver gathers (CRG).The transformation from shot-receiver domain to midpoint-half-offset domain can be realized by the following equations:

$$\mathbf{m} = \frac{\mathbf{s} + \mathbf{r}}{2}, \tag{1}$$

$$\mathbf{h} = \frac{\mathbf{r} - \mathbf{s}}{2}, \tag{2}$$

where $\mathbf{m}$, $\mathbf{s}$ and $\mathbf{r}$ denote the coordinates of midpoint, shot, and receiver, respectively. $\mathbf{h}$ denotes the half-offset (with sign).

The transformation from shot-offset domain (the acquired data from marine streamer) to midpoint-half-offset domain can be realized by the following equations:

$$\mathbf{m} = \mathbf{s} + \frac{\mathbf{h}'}{2}, \tag{3}$$

$$\mathbf{h} = \frac{\mathbf{h}'}{2}, \tag{4}$$

where $\mathbf{h}'$ denotes the full offset.

Figure 4(a) shows synthetic shot-offset domain unblended data simulated from marine-streamer acquisition. Figure 4(b) shows the corresponding shot-offset-domain blended data using the IMSSS blended acquisition (with two sources). We can observe that the interference from the other source appears to be coherent in CSG. After transformation from the shot-offset domain to the midpoint-half-offset domain, the blending noise becomes random and spike-like, as shown in Figure 4(c). Then deblending problem thus turns into a common denoising problem. Thus, in this section of examples, we focus on removing spiky noise in CMG. In the following sections, CSG also refers to the shot-offset domain and CMG also refers to the midpoint-half-offset domain.

Median filtering

Conventional median filtering is based on a scalar-value sorting process. When a set of scalars is sorted into an ascending or descending sequence, the middle value is chosen as standard for this sequence. In signal-processing or geophysical data analysis fields, this filter is commonly used to remove spiky noise. The more general mathematical

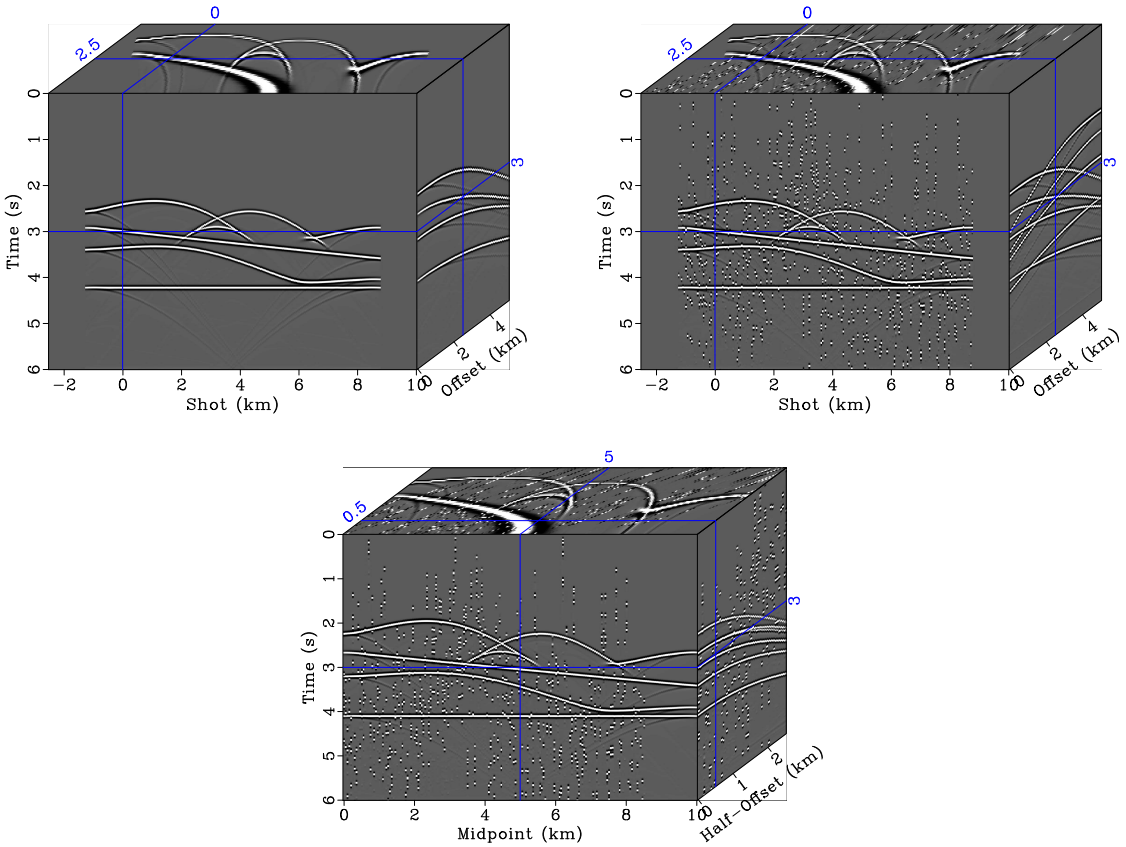


Figure 4: (a) Unblended data in the shot-offset domain. (b) Blended data in the shot-offset domain. (c) Blended data in the midpoint-half-offset domain. nmodeblend/class data-csg,datas-csg,data-blend

formulation of median filtering is given as:

$$\hat{u}_{i,j} = \arg \min_{u_m \in U_{i,j}} \sum_{l=1}^{L} \|u_m - u_l\|_p, \tag{5}$$

where $\hat{u}_{i,j}$ is the output value for location $x_{i,j}$, $U_{i,j} = \{u_1, u_2, \cdots, u_L\}$, i, j are the position indices in a 2-D profile, and l and m are both indices in the filtering window. L is the length of the filtering window, and p denotes L_p norm. Commonly $p = 1$ corresponds to standard median filtering.

Unlike 2D signalprocessing , where the signal is multi-dimensionally coherent, geophysical data are only spatially coherent. Because of the temporally sparse property, the useful signal turns into a spike-like form. This spatial coherent characteristic requires that conventional median filtering should be taken along the spatial direction. Here, spatial direction means the horizontal direction for a 2D seismic profile. In addition, the local slope of an event should be small in order to preserve more useful energy. Figure 5 shows the effectiveness of median filtering in attenuating blending noise and preserving useful events. Figure 5(b) corresponds to median filtering along the time direction with a filter length of 11. With this choice, most of the useful energy has been removed. Figure 5(c) corresponds to median filtering along the spatial direction with a filter length of 7. Figure 5(d) corresponds to median filtering along the spatial direction with a filter length of 11. When the filter length is set at 7, most of horizontal energy is preserved and some dipping energy is lost. When the filter length is set at 11, most of the dipping energy is lost. Thus, we conclude that using median filtering in a profile with dipping events is dangerous; if the filter length is not appropriately chosen, the energy loss is heavy.

For dipping events, multidimensional median filtering can be used as a substitute (15). Multidimensional median filtering, however, requires a precise estimation of the local slope of the seismic events, which may be difficult in field-data processing in the presence of intense random or spiky noise. Besides, using multidimensional median filtering needs much more computational cost and memory. Here, we propose using median filtering after NMO in the CMG, where the events are flatten and the effectiveness of median filtering is maximized.

Median filtering after NMO

By utilizing the effectiveness of median filtering in removing spike-like noise and inserting median filtering into a common seismic processing workflow, we propose the following new processing workflow:

1. Transform the blended data from CSG to CMG.
2. Apply velocity scan and pick the NMO velocity.
3. Apply normal moveout.

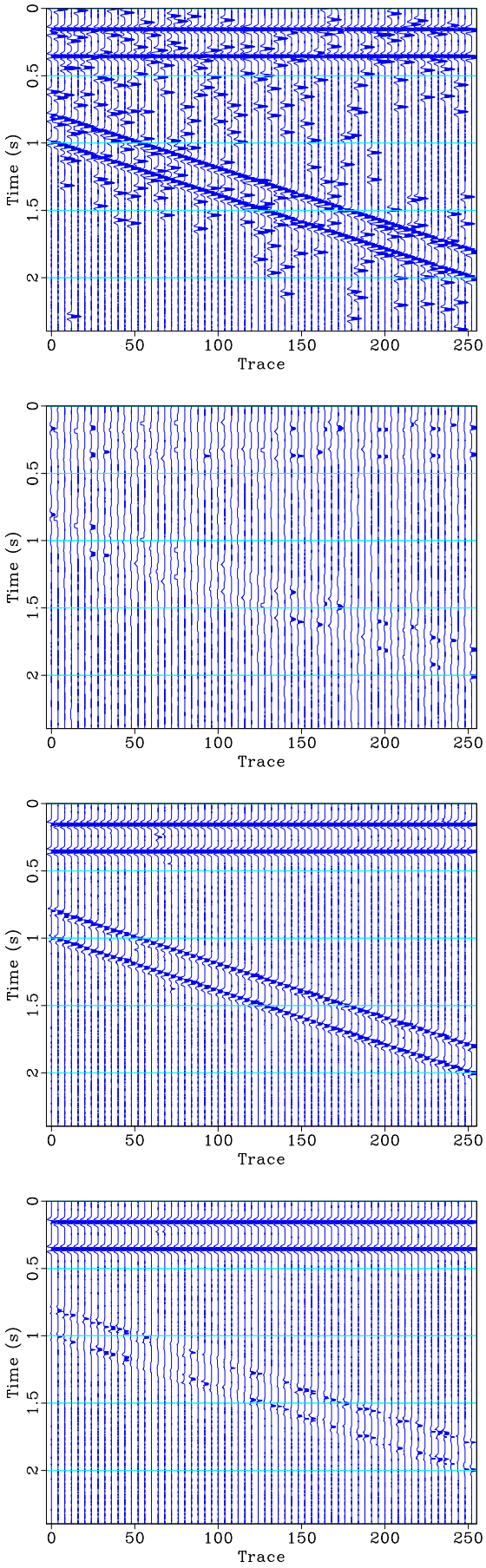


Figure 5: (a) Blended data. (b) Median filtering along the time direc-

4. Apply median filtering along the offset direction in CMG.
5. Apply inverse normal moveout.
6. Prestack migration in CMG or transform back to CSG for other processing tasks.

The key point in the above workflow is flattening the seismic events in order to apply median filtering, which results from obtaining a convincing velocity scan. Because of the intense blending noise, the velocity scan could not be obtained in one step. To solve the problem, we may need to implement steps 2-4 recursively in order to get a better velocity scan. Two velocity scanning iterations are usually adequate to get an acceptable velocity. The proposed processing flow can recursively polish both deblended result and velocity estimation. On one hand, the better velocity estimation can help to make the NMO-corrected events flatter, which improve the performance of median filtering to remove blending noise and to preserve useful energy. On the other hand, the better deblended result can also help to improve the velocity estimation.

EXAMPLES

The first example is a single synthetic CMP gather. The clean unblended data, blended data, and velocity scan of the blended data are shown in Figures 6(a), 1, and 6(c), respectively. Figure 7 shows the blended data after NMO correction and deblended data after NMO correction and median filtering. The median filtering is effective in that most of the interferences have been removed. After inverse NMO on Figure 7(b), the deblended data in CMP gather is shown in Figure 8(a). The blending noise section is shown in Figure 8(b). From the deblending error section as shown in Figure 8(c), we conclude that the proposed method can achieve an excellent result, because the deblending error is small.

We now provide two examples to demonstrate the performance of the proposed workflow described previously. In the next two examples, we use two sources to simulate the blended data. The second example is based on a simple synthetic dataset, which contains four reflectors. The velocity in the model is linearly increasing along the depth axis. We use Kirchhoff modeling to simulate the CSG and blend the data according to IMSSS. After common velocity-semblance scanning, we can pick the NMO velocity and apply NMO. Applying median filtering with a 9-point filter length along the offset direction removes the blending noise. By inverse NMO, we get the deblended dataset in the common-midpoint domain. Figure 9 shows the comparison of CMGs. Figure 10 shows the comparison of velocity scan, in which we see the velocity scan for the blended CMG is smeared along with the noise. After scanning the deblended data coming from the first rough velocity scan and first NMO correction, we however obtain a convincing velocity map. Using the updated NMO with new velocities, we get flatter events that are more suitable for median filtering. Figure 11 shows the comparison of migrated images. In this example, we use prestack

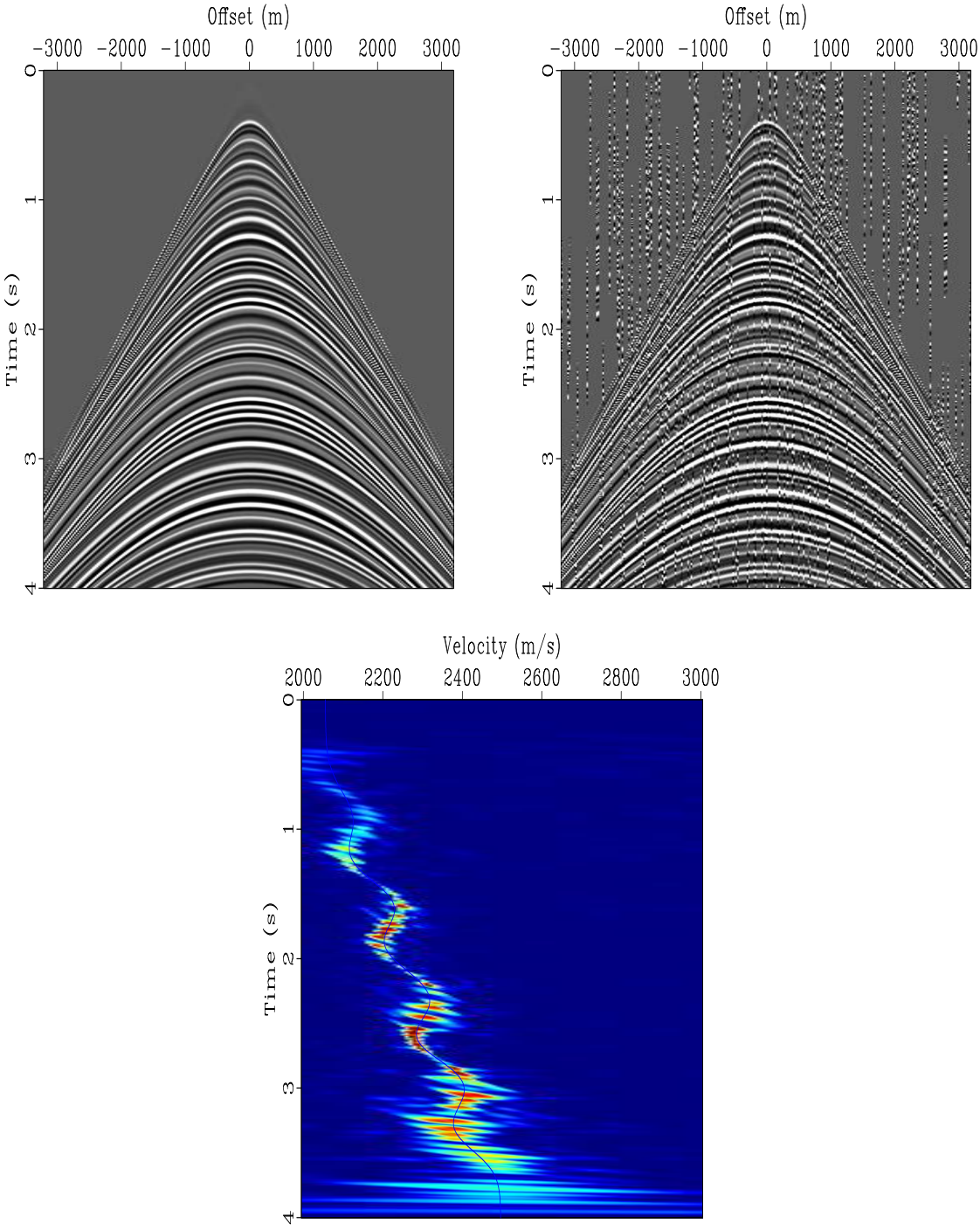


Figure 6: (a) Clean CMP gather. (b) Blended CMP gather. (c) Velocity scan for blended CMP gather. nmodeblend/simple cmpa,cmp,semblancescn

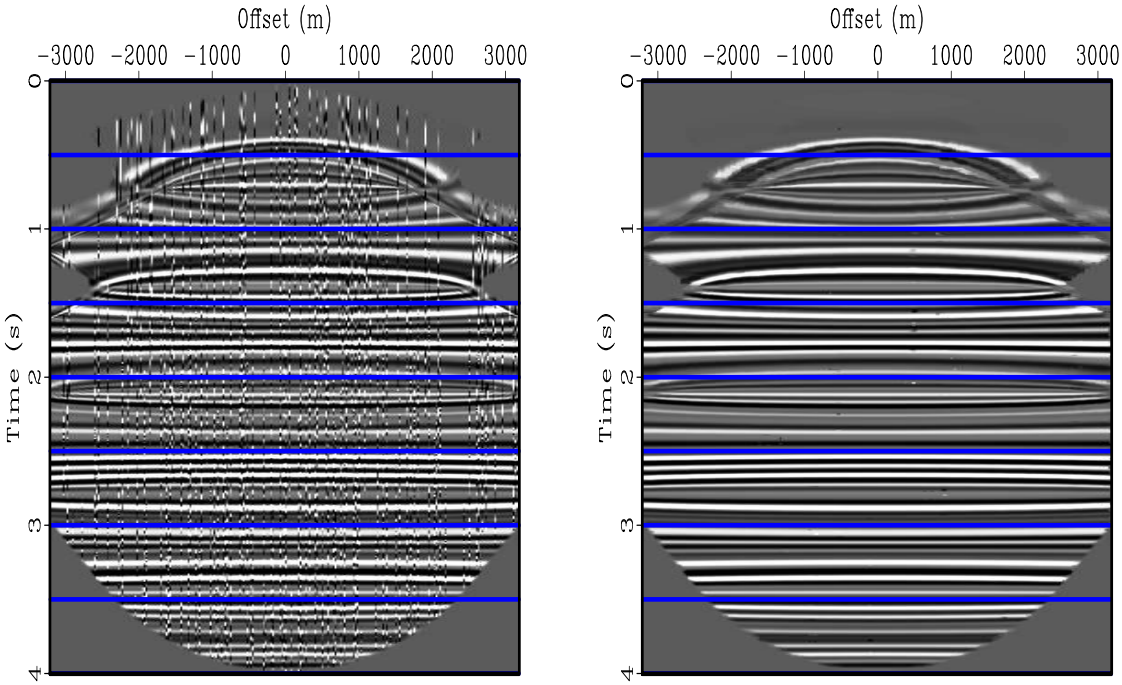


Figure 7: (a) CMP gather after NMO. (b) CMP gather after NMO and median filtering. nmodeblend/simple semblancenmo,semblancenmomf

kirchhoff time migration (PSKTM) (11) as the migration operator. We then use Dix inversion (10) to convert the time image to depth image. The migrated image after deblending is much cleaner than that of the blended data.

The third example is a marine field dataset from the Gulf of Mexico. Figure 12 shows the comparison of CMGs. Figure 13 shows the comparison of velocity scans. Figure 14 shows the comparison of the migrated image. In this example, we use the same migration approach to obtain the seismic images. We have the similar observation to that for the second example: the migrated image for the deblended data is cleaner, especially for the shallow part, indicated by the arrows and frame boxes. For a better view, we zoom the parts indicated by frame boxes and show them in Figure 15. It's obvious to see the improvement for the final migrated image after using the proposed deblending approach.

CONCLUSIONS

We have proposed a basic workflow for dealing with blended data. The processing domain is in CMG, where blending noise appears to be incoherent and seismic reflections appear as hyperbolic events. By applying median filtering along the spatial direction after NMO in CMG, we can easily remove the blending noise without harming useful signal. By applying inverse NMO, we can get the deblended data. In order

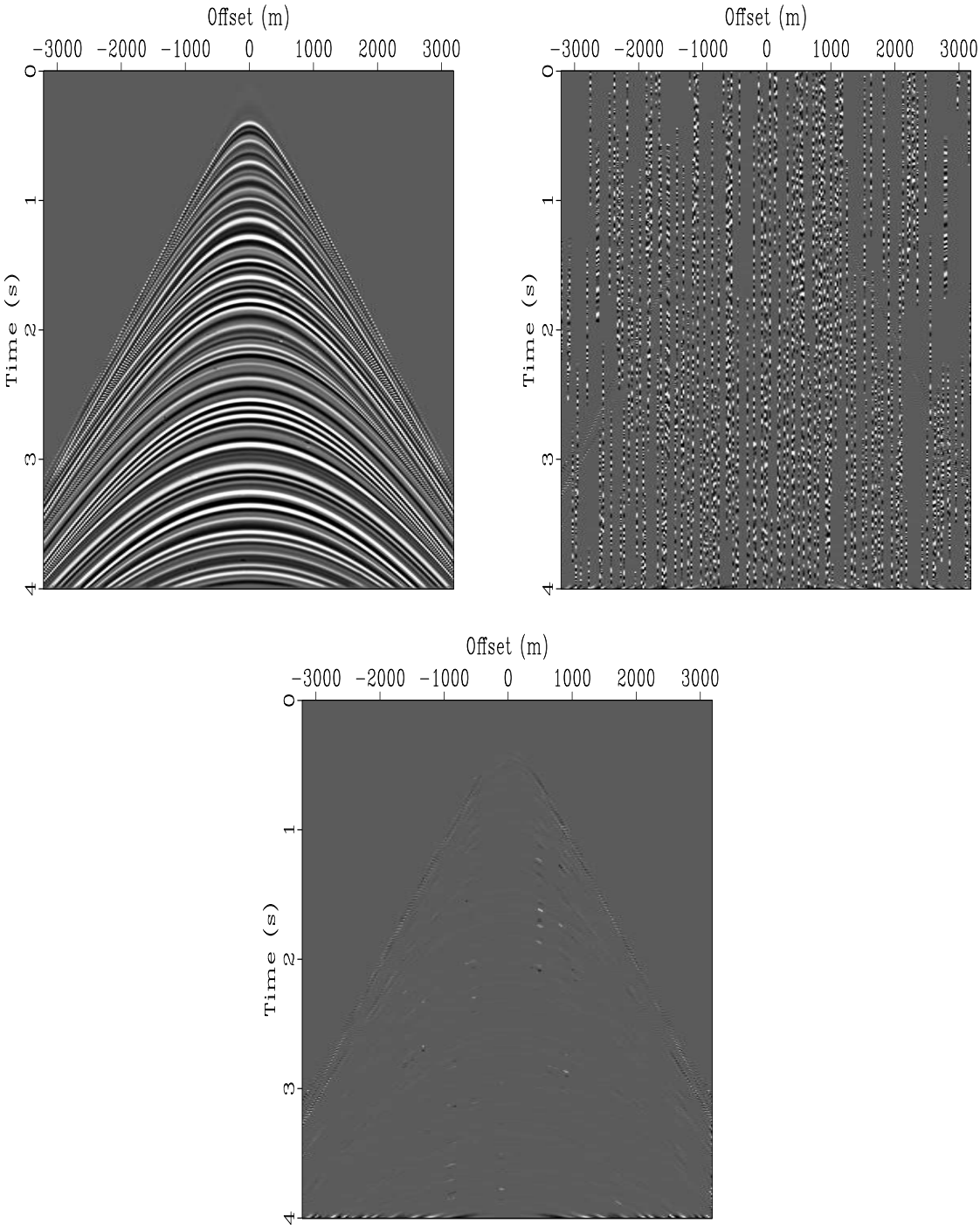


Figure 8: (a) CMP gathering after deblending. (b) Blending noise section. (c) Deblending error section. nmodeblend/simple semblanceinmo,semblancediffmf,semblanceerrormf

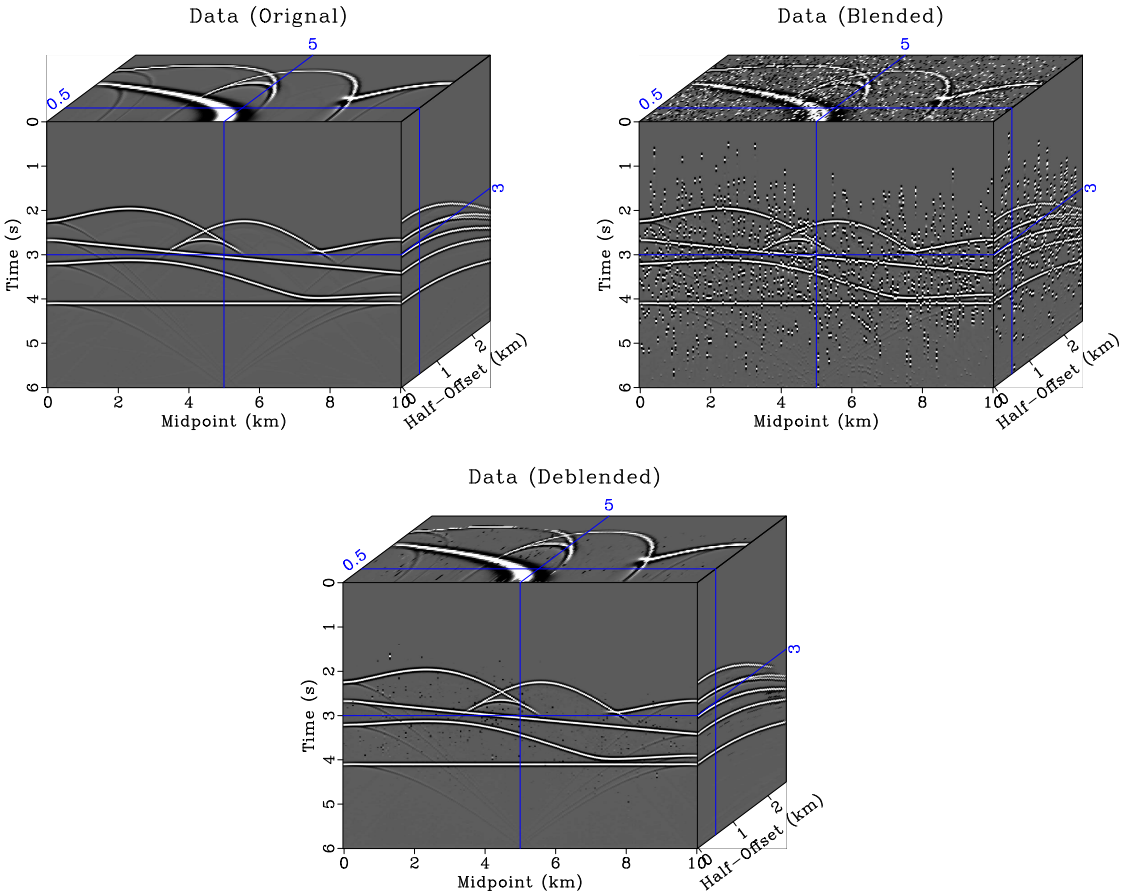


Figure 9: Comparison of CMP gathers. (a) Original unblended CMP gathers. (b) Blended CMP gathers. (c) Deblended CMP gathers. nmodeblend/synth data,data-b,data-db

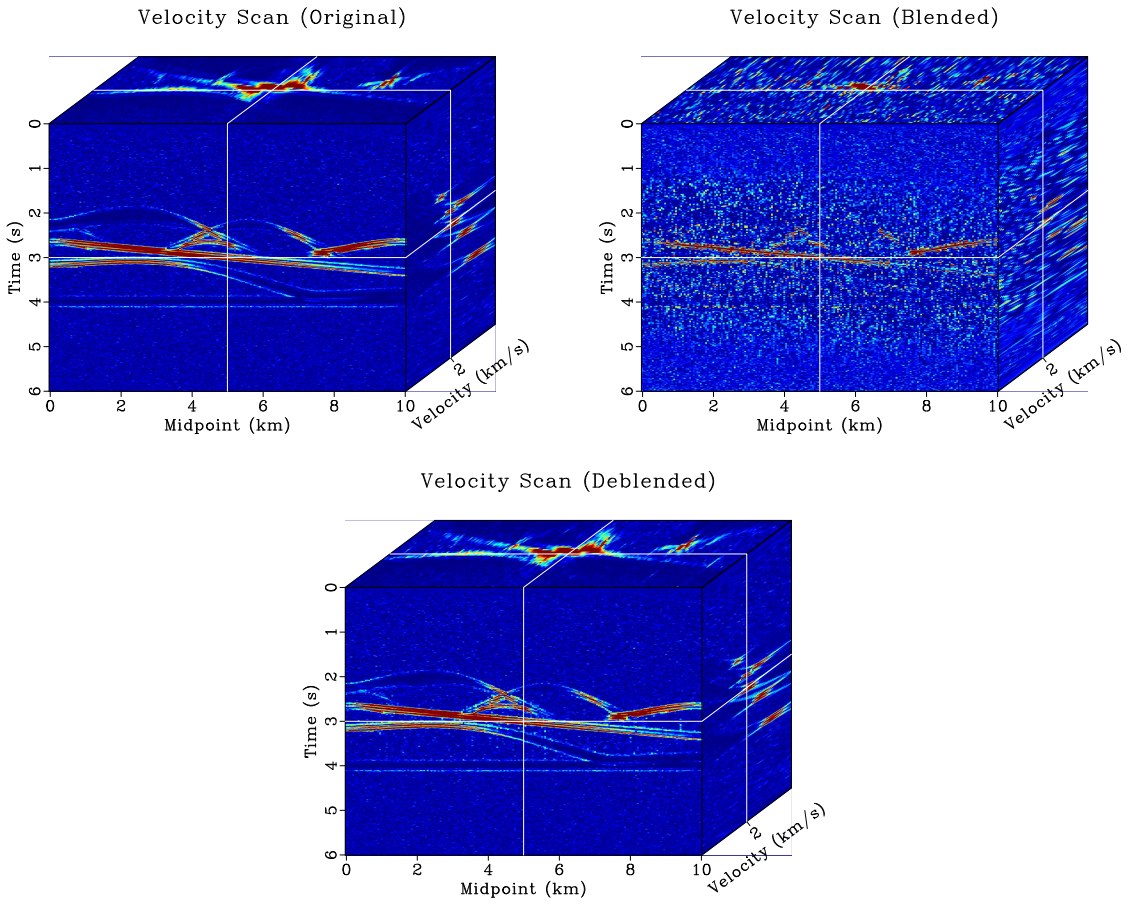


Figure 10: Comparison of velocity scans. (a) Velocity scan for original unblended data. (b) Velocity scan for blended data. (c) Velocity scan for deblended data. nmodeblend/synth svscan,svscan-b,svscan-db

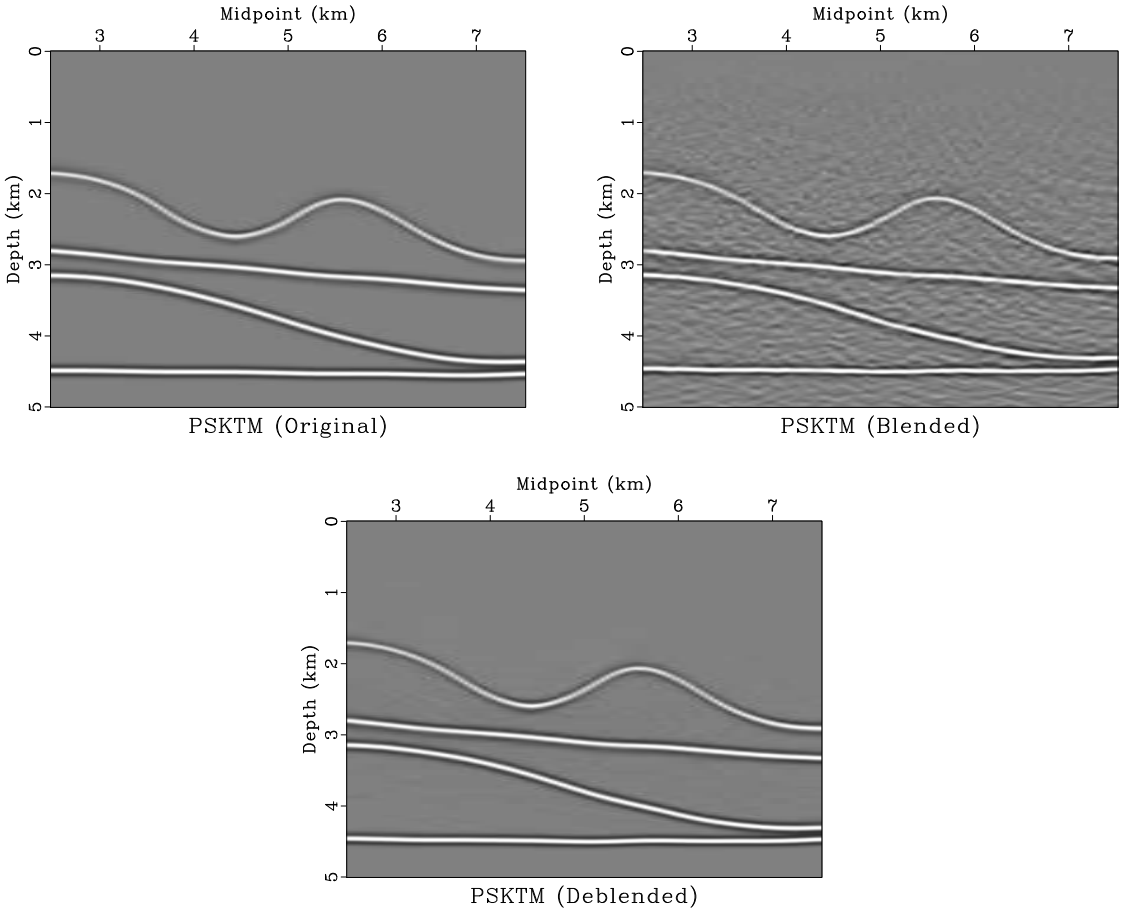


Figure 11: Comparison of migrated images using prestack Kirchhoff time migration (PSKTM). (a) Migrated image for unblended data. (b) Migrated image for blended data. (c) Migrated image for deblended data. nmodeblend/synth spstm,spstm-b,spstm-db

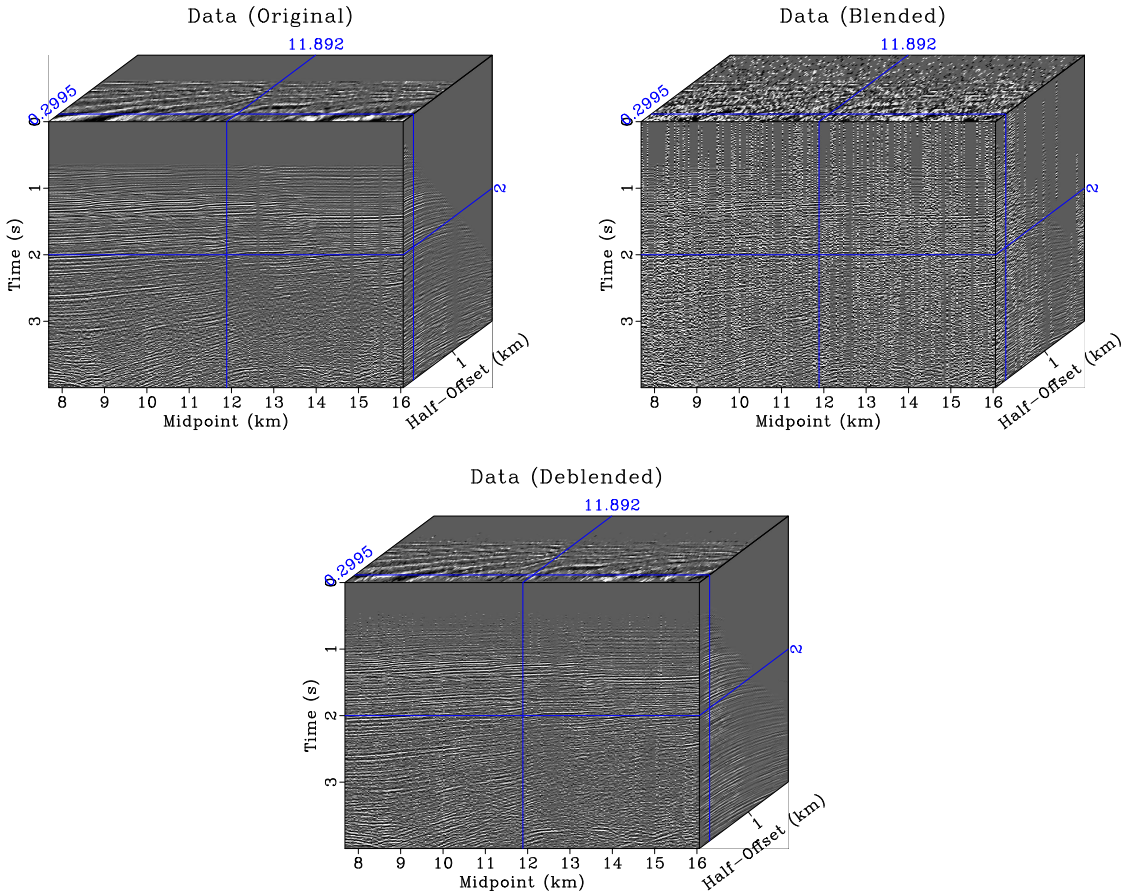


Figure 12: Comparison of CMP gathers. (a) Original unblended CMP gathers. (b) Blended CMP gathers. (c) Deblended CMP gathers. nmodeblend/bei gulf,gulf-b,cmps-db

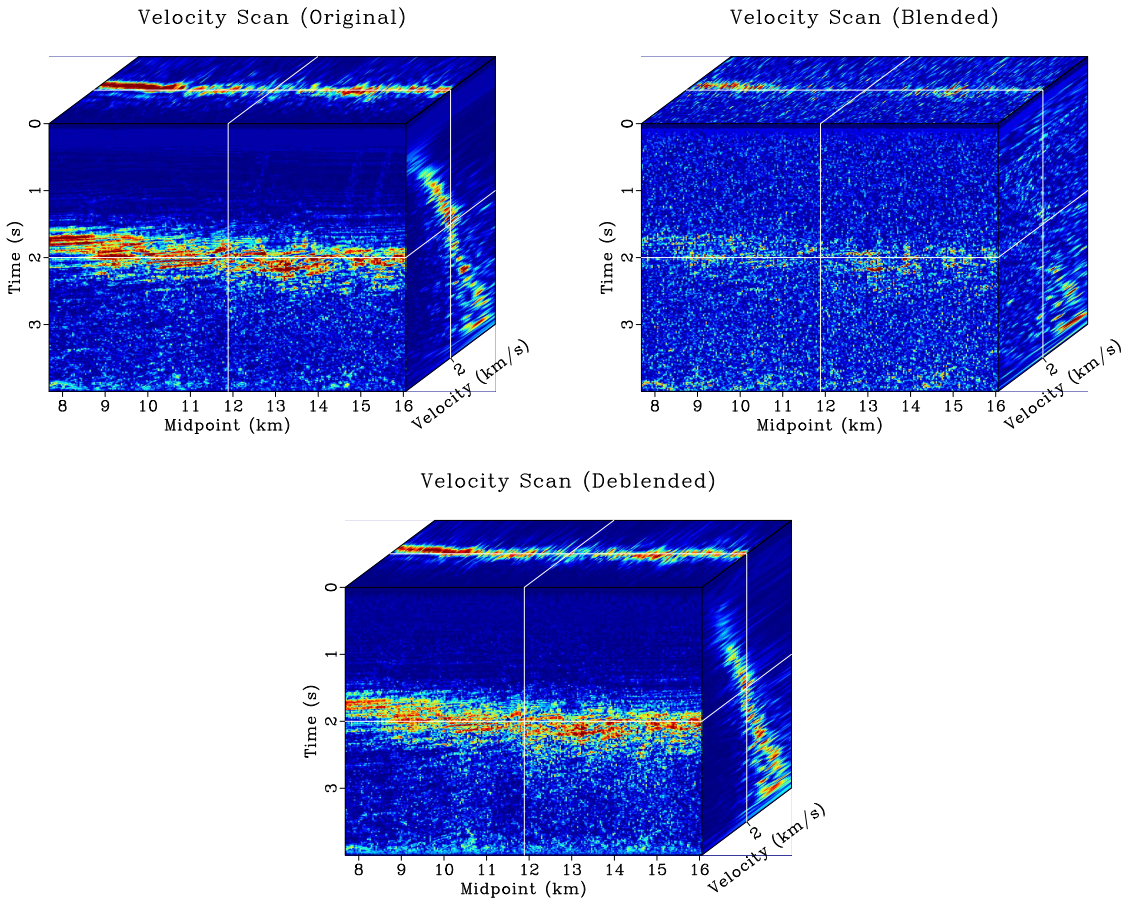


Figure 13: Comparison of velocity scans. (a) Velocity scan for original unblended data. (b) Velocity scan for blended data. (c) Velocity scan for deblended data. nmodeblend/bei vscan-gulf,vscan-b,vscan-db

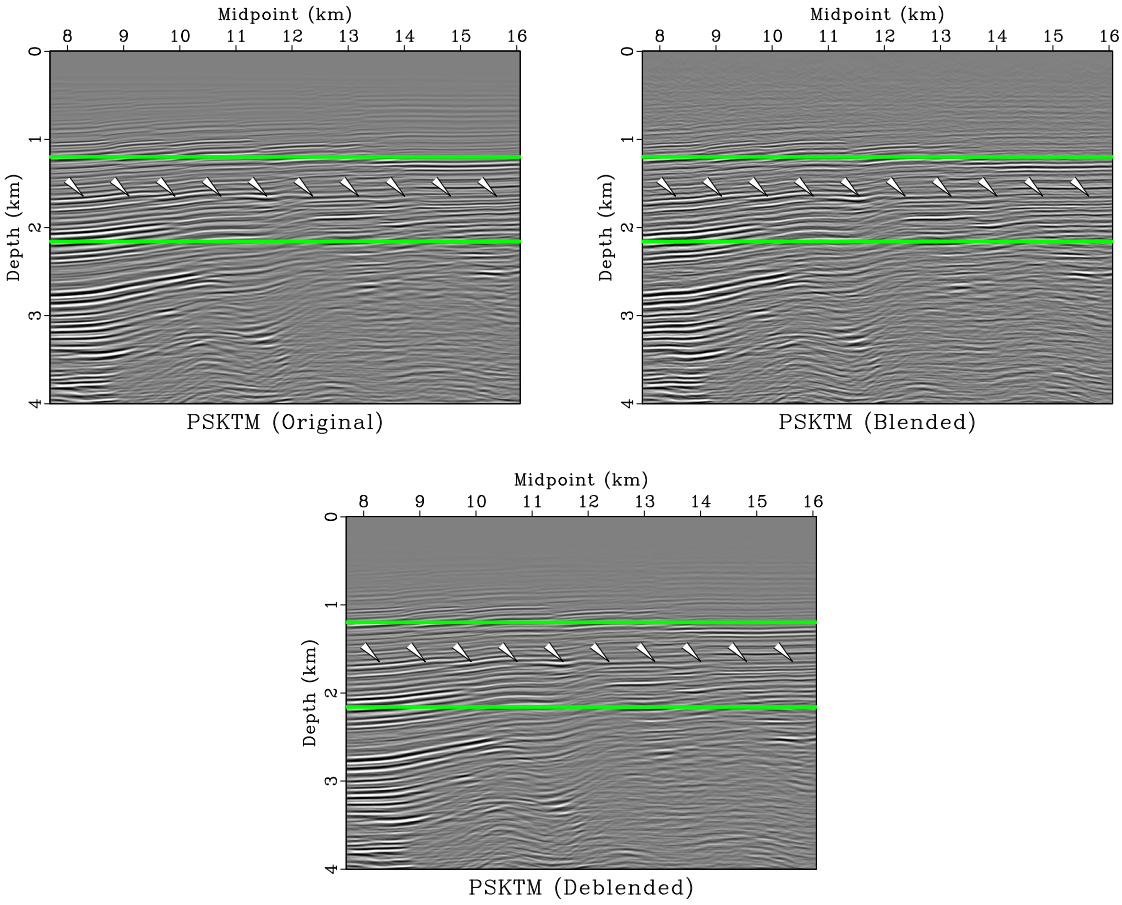


Figure 14: Comparison of migrated images using prestack Kirchhoff time migration (PSKTM). (a) Migrated image for unblended data. (b) Migrated image for blended data. (c) Migrated image for deblended data. nmodeblend/bei pstm,pstm-b,pstm-db

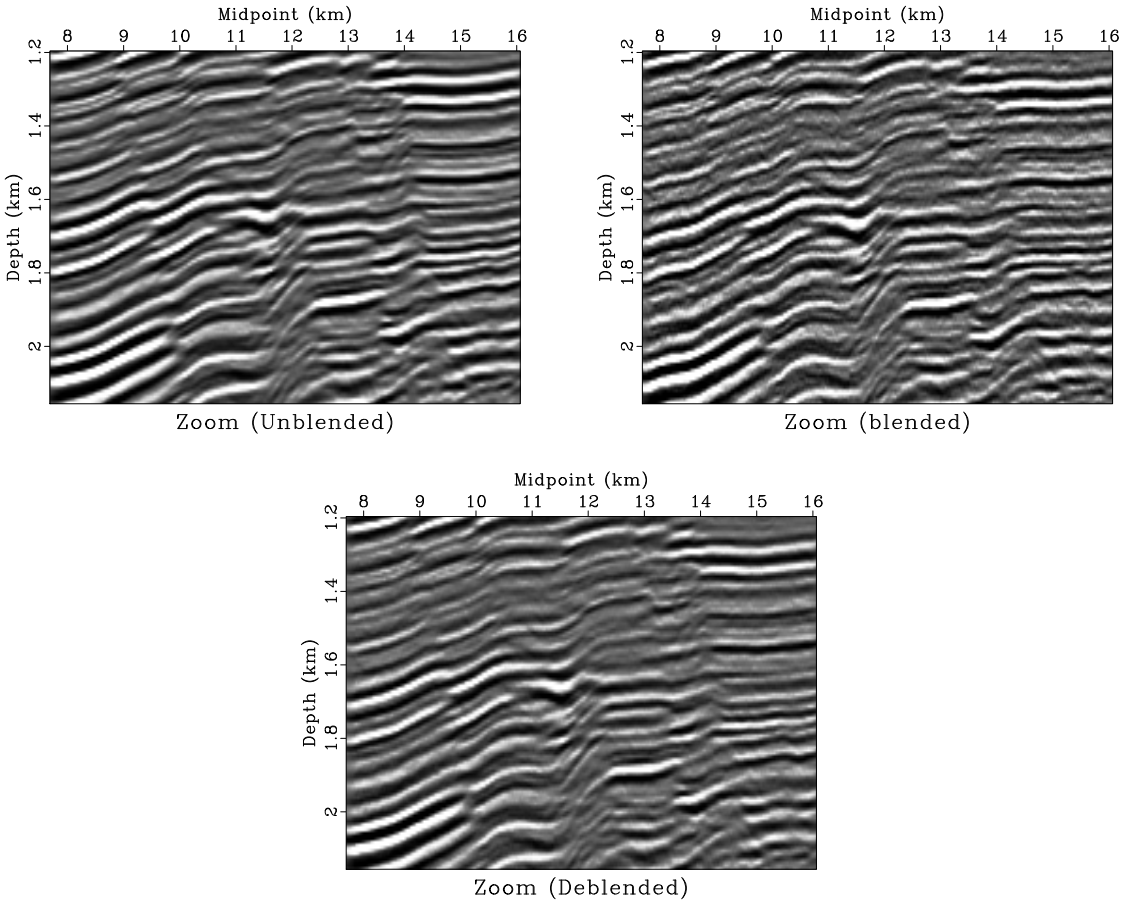


Figure 15: Comparison of zoomed migrated images using prestack Kirchhoff time migration (PSKTM). (a) Zoomed migrated image for unblended data. (b) Zoomed migrated image for blended data. (c) Zoomed migrated image for deblended data. nmodeblend/bei pstmzoom,pstm-bzoom,pstm-dbzoom

to obtain convincing velocity estimation, we use a recursive strategy. We polish both deblended result and velocity estimation by deblending using the updated velocity estimation and velocity scanning using the updated deblended result. It is readily implemented because we use a conventional version of median filtering and don't use any other sophisticated technique to aid in the deblending process. The migrated image of the deblended data shows cleaner structure than that of blended data, which confirms the effectiveness of the proposed deblending approach.

ACKNOWLEDGMENTS

Yangkang Chen would like to thank FairfieldNodal for the opportunity for a summer internship. We also thank Sergey Fomel, Josef Paffenholz and Araz Mahdad for inspiring discussions and helpful suggestions. We thank developers of Madagascar open-source platform for providing accessible codes.

REFERENCES

[1] Abma, R. L., T. Manning, M. Tanis, J. Yu, and M. Foster, 2010, High quality separation of simultaneous sources by sparse inversion: 72nd Annual International Conference and Exhibition, EAGE, Extended Abstracts.

[2] Akerberg, P., G. Hampson, J. Rickett, H. Martin, and J. Cole, 2008, Simultaneous source separation by sparse radon transform: 78th Annual International Meeting, SEG, Expanded Abstracts, 2801–2805.

[3] Bagaini, C., M. Daly, and I. Moore, 2012, The acquisition and processing of dithered slip-sweep vibroseis data: Geophysical Prospecting, **60**, 618–639.

[4] Beasley, C. J., B. Dragoset, and A. Salama, 2012, A 3d simultaneous source field test processed using alternating projections: a new active separation method: Geophysical Prospecting, **60**, 591–601.

[5] Berkhout, A. J., 2008, Changing the mindset in seismic data acquisition: The Leading Edge, **27**, 924–938.

[6] Canales, L., 1984, Random noise reduction: 54th Annual International Meeting, SEG, Expanded Abstracts, 525–527.

[7] Chen, Y., S. Fomel, and J. Hu, 2013, Iterative deblending of simultaneous-source seismic data using shaping regularization: 83rd Annual International Meeting, SEG, Expanded Abstracts, 119–125.

[8] Chen, Y., and J. Ma, 2014, Random noise attenuation by *f-x* empirical mode decomposition predictive filtering: Geophysics, **79**, V81–V91.

[9] Dai, W., P. Fowler, and G. T. Schuster, 2012, Multi-source least-squares reverse time migration: Geophysical Prospecting, **60**, 681–695.

[10] Dix, C. H., 1955, Seismic velocities from surface measurements: Geophysics, **20**, 68–86.

[11] Docherty, P., 1991, A brief comparison of some kirchhoff integral formulas for migration and inversion: Geophysics, **56**, 1164–1169.

[12] Doulgeris, P., and K. Bube, 2012, Analysis of a coherency-constrained inversion for the separateion of blended data: discovering the leakage subspace: 74th Annual International Conference and Exhibition, EAGE, Extended Abstracts.
[13] Doulgeris, P., K. Bube, G. Hampson, and G. Blacquiere, 2012, Covergence analysis of a coherency-constrained inversion for the separation of blended data: Geophysical Prospecting, **60**, 769–781.
[14] Hampson, G., J. Stefani, and F. Herkenhoff, 2008, Acquisition using simultaneous sources: The Leading Edge, **27**, 918–923.
[15] Huo, S., Y. Luo, and P. G. Kelamis, 2012, Simultaneous sources separation via multidirectional vector-median filtering: Geophysics, **77**, V123–V131.
[16] Mahdad, A., P. Doulgeris, and G. Blacquiere, 2011, Separation of blended data by iterative estimation and subtraction of blending interference noise: Geophysics, **76**, Q9–Q17.
[17] ———, 2012, Iterative method for the separation of blended seismic data: discussion on the algorithmic aspects: Geophysical Prospecting, **60**, 782–801.
[18] Moore, I., 2010, Simultaneous source - processing and applications: 72th Annual International Conference and Exhibition, EAGE, Extended Abstracts.
[19] Moore, I., B. Dragoset, T. Ommundsen, D. Wilson, C. Ward, and D. Eke, 2008, Simultaneous source separation using dithered sources: 78th Annual International Meeting, SEG, Expanded Abstracts, 2806–2809.
[20] van Borselen, R., R. Baardman, T. Martin, B. Goswami, and E. Fromyr, 2012, An inversion approach to separating sources in marine simultaneous shooting acquision-application to a gulf of mexico data set: Geophysical Prospecting, **60**, 640–647.
[21] Verschuur, D. J., and A. J. Berkhout, 2011, Seismic migration of blended shot records with surface-related multiple scattering: Geophysics, **76**, A7A13.

Azimuthally anisotropic 3D velocity continuation

*William Burnett and Sergey Fomel**

ABSTRACT

We extend time-domain velocity continuation to the zero-offset 3D azimuthally anisotropic case. Velocity continuation describes how a seismic image changes given a change in migration velocity. This description turns out to be of a wave propagation process, in which images change along a velocity axis. In the anisotropic case, the velocity model is multi-parameter. Therefore, anisotropic image propagation is multi-dimensional. We use a three-parameter slowness model, which is related to azimuthal variations in velocity, as well as their principal directions. This information is useful for fracture and reservoir characterization from seismic data. We provide synthetic diffraction imaging examples to illustrate the concept and potential applications of azimuthal velocity continuation and to analyze the impulse response of the 3D velocity continuation operator.

INTRODUCTION

Velocity continuation (12; 14) provides a framework for describing how a seismic image changes given a change in the migration velocity model. Similar in concept to residual migration (28) and cascaded migrations (24), velocity continuation is a continuous formulation of the same process. Velocity continuation has found applications in migration velocity analysis (13; 30) and diffraction imaging (27; 16).

(12) and (21) point out that velocity continuation is a wave propagation process. Instead of wavefronts propagating as a function of time, images propagate as a function of migration velocity. Recent work has extended the concept to heterogeneous and anisotropic velocity models (4; 1; 22; 29; 31; 10). To account for anisotropy, the seismic velocity model must become multi-parameter. Consequentially, velocity continuation generalizes to a process of implementing image transformations caused by changes in multiple parameters rather than the single isotropic velocity alone.

Accounting for azimuthal variations in seismic velocity results in better event focusing and improved imaging in such media (32). Azimuthal variation in velocity has been shown to be an indicator of preferentially aligned vertical fractures (8), lateral heterogeneity (25), regional stress (33), or a combination of these factors. However, velocity analysis is commonly first performed on pre-stack common midpoint (CMP)

*e-mail: sergey.fomel@beg.utexas.edu

gathers, where the geologic cause of any observed azimuthal velocity variation is ambiguous. Without the help of additional diagnostic gathers such as hybrid or cross-spread gathers (11), or an interpretive comparison between picked root-mean-square (RMS) and interval velocities (23), the cause of azimuthal variations in velocity can be identified only after migration.

Azimuthal seismic imaging commonly requires iterations between velocity analysis and imaging. Residual azimuthal variations in traveltime after migration can be measured by using migration binning schemes which preserve both offset and azimuth information (5; 37). After the first pass of (isotropic) migration, azimuthal variations in velocity are detected from residual moveout, which then provides the velocity model for anisotropic migration. Iterative processing flows that use these strategies are popular not only because they are fairly efficient and intuitive, but also because they can be implemented with minimal modification to existing software. However, iterative imaging flows cannot guarantee convergence to the correct or optimal velocity model (9). Velocity continuation has the underlying strategy of performing velocity analysis and imaging simultaneously, and can thus be used to directly find an optimal velocity model without iteration. (33) have demonstrated the success of a similar strategy of using imaging as a velocity analysis tool for 3D multiazimuth reflection seismic data. Azimuthal velocity continuation can provide a theoretical framework for this approach. With these benefits as motivation, we extend time-domain velocity continuation to 3D, accounting for the case of azimuthally variable migration velocity.

THEORY

The theory of velocity continuation formulates the connection between the seismic velocity model and the seismic image as a wavefield evolution process. In doing so, the process can be implemented in the same variety of ways as seismic migration. Seismic migration in its many forms is commonly derived starting at the wave equation, which is approximated by its time and amplitude components by the eikonal and transport equations, and if necessary, a system of ray tracing equations. Velocity continuation is derived in the opposite order (14). Starting with a geometrical description of the image, a corresponding kinematic equation for traveltime is derived to describe how the image moves according to changes in imaging parameters. Subsequently, the kinematic equation is used to derive a corresponding wave equation, which describes the dynamic behavior of the image as an evolution through imaging parameter coordinates. This section outlines the key steps of this derivation, starting with a traveltime equation that permits azimuthal variations in velocity.

(18) truncate a two-dimensional Taylor series expansion for a generally inhomogeneous anisotropic media to derive the "NMO ellipse" moveout equation. Geometrically, the NMO ellipse model still assumes that events have hyperbolic moveout with offset, but it allows the velocity to change with azimuth. We start here by using the same truncated 2D Taylor series expansion to describe an azimuth-dependent traveltime equation for the summation surface of zero-offset time migration,

$$T^2(\mathbf{x},\mathbf{y},\mathbf{W}) = 4\left(\tau^2 + (\mathbf{x}-\mathbf{y})^{\mathrm{T}}\,\mathbf{W}\,(\mathbf{x}-\mathbf{y})\right), \tag{1}$$

where τ is the one-way vertical traveltime after migration, $\mathbf{x}$ is the (x_1, x_2) surface position of the zero-offset receiver in survey coordinates, $\mathbf{y}$ is the surface position of the point source image, and superscript $^{\mathrm{T}}$ denotes transpose. The three independent elements of the symmetric slowness matrix,

$$\mathbf{W} = \begin{pmatrix} W_{11} & W_{12} \\ W_{12} & W_{22} \end{pmatrix}, \tag{2}$$

have units of slowness-squared, and the eigenvalues and eigenvectors of $\mathbf{W}$ determine the symmetry axes of the effective anisotropic medium (18). In most common geologic situations, the eigenvalues of $\mathbf{W}$ are positive (36), and equation 1 describes an elliptical-hyperbolic traveltime surface in 3D—hyperbolic in cross-section view and elliptical in map-view. The fast and slow moveout velocities are aligned with the major and minor axes of this ellipse. W_{11} and W_{22} are the squared moveout slownesses along their respective survey coordinates, x_1 and x_2. The third parameter, W_{12}, arises from observing the ellipse in the x_1-x_2 survey coordinates, which are generally rotated relative to its major and minor axes.

The three-parameter moveout model of equation 1 is analytically convenient and practical, but the parameters themselves are not intuitive to interpret in terms of more common geophysical or geological parameters. However, some simple geometric observations can help convert the three elements of $\mathbf{W}$ into more intuitive measurements. If the ellipse happens to be aligned with the survey coordinates, $W_{12} = 0$. Finding the rotation angle which properly diagonalizes $\mathbf{W}$ therefore allows one to predict the orientation of the symmetry axes. This amounts to an eigenvalue problem, where the fast and slow velocities can be found as the eigenvalues and eigenvectors of $\mathbf{W}$. The eigenvalues, W_{fast} and W_{slow}, of the slowness matrix can be found following (18),

$$W_{slow,fast} = \frac{1}{2}\left[W_{11} + W_{22} \pm \sqrt{(W_{11}-W_{22})^2 + 4W_{12}^2}\right]. \tag{3}$$

Since the eigenvalues have units of slowness squared, the smaller eigenvalue is $W_{fast} = 1/v_{fast}^2$. One can solve for the angle β between the acquisition coordinates and the symmetry axes by using the formula found by (18),

$$\beta = \tan^{-1}\left[\frac{W_{22} - W_{11} + \sqrt{(W_{22}-W_{11})^2 + 4W_{12}^2}}{2W_{12}}\right]. \tag{4}$$

The eigenvalues can then be used together with β to solve for the zero-offset migration slowness S as a function of source-receiver azimuth θ:

$$S^2(\theta) = W_{slow}\cos^2(\theta-\beta) + W_{fast}\sin^2(\theta-\beta). \tag{5}$$

Equations 3-5 allow one to convert the mathematically convenient parameters of $\mathbf{W}$ to more intuitive parameters, such as the fastest and slowest propagation velocities (V_{fast},V_{slow}), the azimuth of the slowest velocity (β), and the percent anisotropy ($\sigma = 100 \times (1 - V_{slow}/V_{fast})$). Alternatively, $\mathbf{W}$ can be converted into other common geophysical parametrizations. For example, (18) show that once the effective parameters $\mathbf{W}$ have been converted to slowness as a function of azimuth by equation 5, they can be expressed in terms of horizontal transverse isotropy parameters as,

$$S^2(\theta) = \frac{1}{V_{P0}^2}\frac{1+2\delta^{(v)}\sin^2(\theta)}{1+2\delta^{(v)}}, \tag{6}$$

where $\delta^{(v)}$ is the Thomsen-style parameter (34), introduced by (35), and V_{P0} is the vertical P-wave velocity.

Conventionally, one assumes that equation 1 characterizes a particular event defined in image coordinates $(\mathbf{x},\tau)$, but one can also describe how that event would transform given a change in the image parameters $\mathbf{W}$. Regardless of the velocity model, the traveltime T must remain unchanged between different images. From this observation, we arrive at the following set of conditions:

$$\nabla_{\mathbf{x}} T^2 = \begin{pmatrix} \frac{\partial T^2}{\partial x_1} \\ \frac{\partial T^2}{\partial x_2} \end{pmatrix} = 8\tau\nabla_x\tau + 8\mathbf{W}(\mathbf{x}-\mathbf{y}) = 0, \tag{7}$$

and,

$$\nabla_{\mathbf{W}} T^2 = \begin{pmatrix} \frac{\partial T^2}{\partial W_{11}} & \frac{\partial T^2}{\partial W_{12}} \\ \frac{\partial T^2}{\partial W_{12}} & \frac{\partial T^2}{\partial W_{22}} \end{pmatrix} = 8\tau\nabla_{\mathbf{W}}\tau + 4(\mathbf{x}-\mathbf{y})(\mathbf{x}-\mathbf{y})^{\mathrm{T}} = 0. \tag{8}$$

Combining and reducing these conditions yields a system of equations that are defined only in the image parameter coordinates,

$$2\frac{\partial\tau}{\partial W_{11}} + \frac{\tau\left(W_{22}\frac{\partial\tau}{\partial x_1} - W_{12}\frac{\partial\tau}{\partial x_2}\right)^2}{\left(W_{12}^2 - W_{11}W_{22}\right)^2} = 0, \tag{9}$$

$$2\frac{\partial\tau}{\partial W_{22}} + \frac{\tau\left(W_{12}\frac{\partial\tau}{\partial x_1} - W_{11}\frac{\partial\tau}{\partial x_2}\right)^2}{\left(W_{12}^2 - W_{11}W_{22}\right)^2} = 0, \tag{10}$$

and,

$$2\frac{\partial\tau}{\partial W_{12}} - \frac{2\tau\left(W_{12}\frac{\partial\tau}{\partial x_1} - W_{11}\frac{\partial\tau}{\partial x_2}\right)\left(W_{22}\frac{\partial\tau}{\partial x_1} - W_{12}\frac{\partial\tau}{\partial x_2}\right)}{\left(W_{12}^2 - W_{11}W_{22}\right)^2} = 0. \tag{11}$$

The system of kinematic equations describing azimuthally anisotropic velocity continuation is then found by combining equations 9-11. In a vector notation, this becomes

$$\nabla_{\mathbf{W}}\tau + \frac{\tau}{2}\mathbf{W}^{-1}\nabla_{\mathbf{x}}\tau\left(\nabla_{\mathbf{x}}\tau\right)^{\mathrm{T}}\mathbf{W}^{-1} = 0, \tag{12}$$

where $\nabla_{\mathbf{x}}$ and $\nabla_{\mathbf{W}}$ are in the form given by equations 7 and 8.

The method of characteristics (7) provides a link between a kinematic equation (such as 12) and its corresponding wave-type equation. (14) demonstrates specifically how the method can be used to derive a velocity continuation wave equation from its kinematic counterpart. By first setting the characteristic surface condition,

$$\psi = t - \tau(\mathbf{x}, \mathbf{W}) = 0, \tag{13}$$

and replacing τ with ψ and t, we obtain an alternative form of equation 12,

$$\psi_t\nabla_{\mathbf{W}}\psi + \frac{t}{2}\mathbf{W}^{-1}\nabla_{\mathbf{x}}\psi\left(\nabla_{\mathbf{x}}\psi\right)^{\mathrm{T}}\mathbf{W}^{-1} = 0. \tag{14}$$

Equation 13 guarantees that the wavefronts of time-domain image wavefield P exist only where the arrival time τ is equal to the recorded time t at a given location. Now take both ξ_i and ξ_j to represent each of t, W_{11}, W_{12}, W_{22}, x_1, and x_2. According to the method of characteristics, if Λ_{ij} is the coefficient in front of $\frac{\partial\psi}{\partial\xi_i}\frac{\partial\psi}{\partial\xi_j}$ from kinematic equation 14, then the corresponding wave equation will have the same coefficients Λ_{ij} in front of each $\frac{\partial^2 P}{\partial\xi_i\partial\xi_j}$ derivative. The time-derivative ψ_t is equal to 1 given equation 13, and is included in the first term of equation 14 to facilitate the use of the method of characteristics. Then, by introducing $\mathbf{P}_{xx}$ as the spatial Hessian matrix of the wavefield,

$$\mathbf{P}_{xx} = \begin{pmatrix} \frac{\partial^2 P}{\partial x_1^2} & \frac{\partial^2 P}{\partial x_1\partial x_2} \\ \frac{\partial^2 P}{\partial x_2\partial x_1} & \frac{\partial^2 P}{\partial x_2^2} \end{pmatrix}, \tag{15}$$

we arrive at the azimuthally anisotropic post-stack velocity continuation wave equation,

$$\nabla_{\mathbf{W}}P_t = -\frac{t}{2}\mathbf{W}^{-1}\mathbf{P}_{xx}\mathbf{W}^{-1}. \tag{16}$$

In the isotropic case, $\mathbf{W}$ is diagonal and $W_{11} = W_{22}$. Equation 16 then reduces to the isotropic velocity continuation equation first derived by (6).

METHOD

Since velocity continuation is described by a wave equation, it can be implemented in analogous ways to seismic migration. Here, we demonstrate a spectral implementation of equation 16. By first stretching the time coordinate of an input image from t to $\tilde{t} = t^2/2$, and then taking its 3D Fourier transform, equation 16 becomes the reduced partial differential equation,

$$i\Omega\nabla_{\mathbf{W}}\hat{P} = \frac{1}{2}\mathbf{W}^{-1}\mathbf{k}\mathbf{k}^{\mathrm{T}}\mathbf{W}^{-1}\hat{P}, \tag{17}$$

where Ω is the Fourier dual of $\tilde{t}$ and $\mathbf{k}$ is the wavenumber vector (Fourier dual of $\mathbf{x}$). Equation 17 has the analytical solution,

$$\hat{P}(\Omega, k_1, k_2, \mathbf{W}) = \hat{P}(\Omega, k_1, k_2, \mathbf{W}_0)e^{\frac{-i}{2\Omega}\mathbf{k}^{\mathrm{T}}(\mathbf{W}^{-1}-\mathbf{W}_0^{-1})\mathbf{k}}, \tag{18}$$

which shows that continuation of an image from an arbitrary $\mathbf{W}_0$ to $\mathbf{W}$ can be achieved by multiplication with a shifting exponential in the Fourier domain. One can also directly migrate an unmigrated image by using the 2×2 matrix $\mathbf{W}_0^{-1} = \mathbf{0}$ for the initial velocity. In practice, the coordinate stretch from t to $\tilde{t}$ should be carefully applied as data will be compressed along the time-axis for early samples.

With a range of slowness matrices $\mathbf{W}$, equation 18 can be used to quickly generate the corresponding range of anisotropically migrated images. When the correct velocity model is used, diffractions collapse to points, which we recognize as the image coming into focus. Although constant velocity models are used for each image, this type of spectral implementation can still be useful in the heterogeneous case, as different parts of the image will come into focus locally as the appropriate velocity is used (20; 17). Once the range of images is generated, we search for the best-focused image at each output location. We use the image attribute of kurtosis, defined as,

$$\phi(\mathbf{W}) = \frac{\int\int P^4(\mathbf{x}, t, \mathbf{W})\, d\mathbf{x}\, dt}{\left[\int\int P^2(\mathbf{x}, t, \mathbf{W})\, d\mathbf{x}\, dt\right]^2}, \tag{19}$$

to quantify how well a location is focused in a particular image (38; 16). Including integration limits specifies a window size for locally measuring kurtosis in the image. In application, the integration limits control either the size of a "sliding window", or when viewing kurtosis as a local attribute (15), they can be used to control the smoothness enforced by shaping regularization. In either case, the integration limits control a trade-off between the robustness of the focusing measurement and the resolution. From experience, typical limits for field data correspond to window sizes on the order of 10^1 samples in each dimension. It should be noted that the traveltime approximation of equation 1 loses accuracy in the presence of strong lateral heterogeneity, but is commonly used to estimate smooth effective parameter models. Following the maximum values through the resulting six-dimensional kurtosis

hypercube, $\phi(t, \mathbf{x}, \mathbf{W})$, and then slicing corresponding pieces from the output images volume, $P(t, \mathbf{x}, \mathbf{W})$, reveals an effective medium based heterogeneous velocity model and a well-focused image. This spectral implementation and slicing procedure is similar to searching through a set of constant-velocity $f - k$ migrations, and can be completed without any prior knowledge of the velocity model (17; 26).

EXAMPLES

Two simple synthetic examples are provided below to illustrate 3D velocity continuation over a range of velocity models. In the first example, we apply velocity continuation to a point diffractor. In the second example, we apply the method to a synthetic post-stack image of a set of faults. The second example illustrates fracture characterization through diffraction imaging as a potential application for 3D azimuthal velocity continuation. The data in both examples are modeled using equation 1, which geometrically approximates a diffraction surface as an elliptical-hyperbolic surface. Field data and more accurately modeled data will generally also exhibit nonhyperbolic moveout, for which equation 1 does not account. The physical validity and limitations of equation 1 are thoroughly discussed by (18), but we focus here on how well diffractions can be collapsed, and how well the velocity parameters can be measured, if the data are ideally described by equation 1.

Figure 1a shows a single diffraction event, modeled using equation 1. The fastest direction of propagation is at β=105° counter-clockwise from the x_1 axis, with V_{fast}=3.50 km/s. The data in Figure 1a were modeled with σ=7% anisotropy, which may be quite high for most field cases, but it was chosen to allow the azimuthal variations in diffraction moveout to be visibly pronounced. As described above, we first stretch the time axis from t to $\tilde{t}$ and take the 3D Fourier transform of the data. Then we apply the phase-shift prescribed by equation 18 for a range of $\mathbf{W}$. We found it more intuitive to specify the parameter ranges in terms of V_{x_1}, β, and σ, and then convert them at each step into the three parameters of $\mathbf{W}$ for use in equation 18. The inverse of the in-line velocity squared $1/V_{x_1}^2$ is equivalent to W_{11}, which, along with a given fast azimuth β and percent anisotropy σ, can be used to calculate W_{12} and W_{22} using equations 3-5. Last, we apply the 3D inverse Fourier transform and unstretch from $\tilde{t}$ to t to obtain the 6D image volume. Examples from the image volume using incorrect parameters are shown in Figures 1b-1c. The correct parameters are used in Figure 1d, where the image is well-focused.

Since only a single diffraction is present in this example, we can measure kurtosis over a window spanning the entirety of each 3D image, reducing the kurtosis volume from 6D to 3D. Figure 2 is a 2D slice of the kurtosis volume at the correct $W_{11} = 1/V_{x_1}^2$ value of $0.0935 = 1/(3.27 \text{ km/s})^2$. The peak of the kurtosis map is near the correct values of σ=7 and β=105°. Once the peak of the kurtosis map is identified, one could refine the increments around the peak to yield more accurate estimates. The physical limitations of resolving azimuthal velocity parameters are discussed by (2).

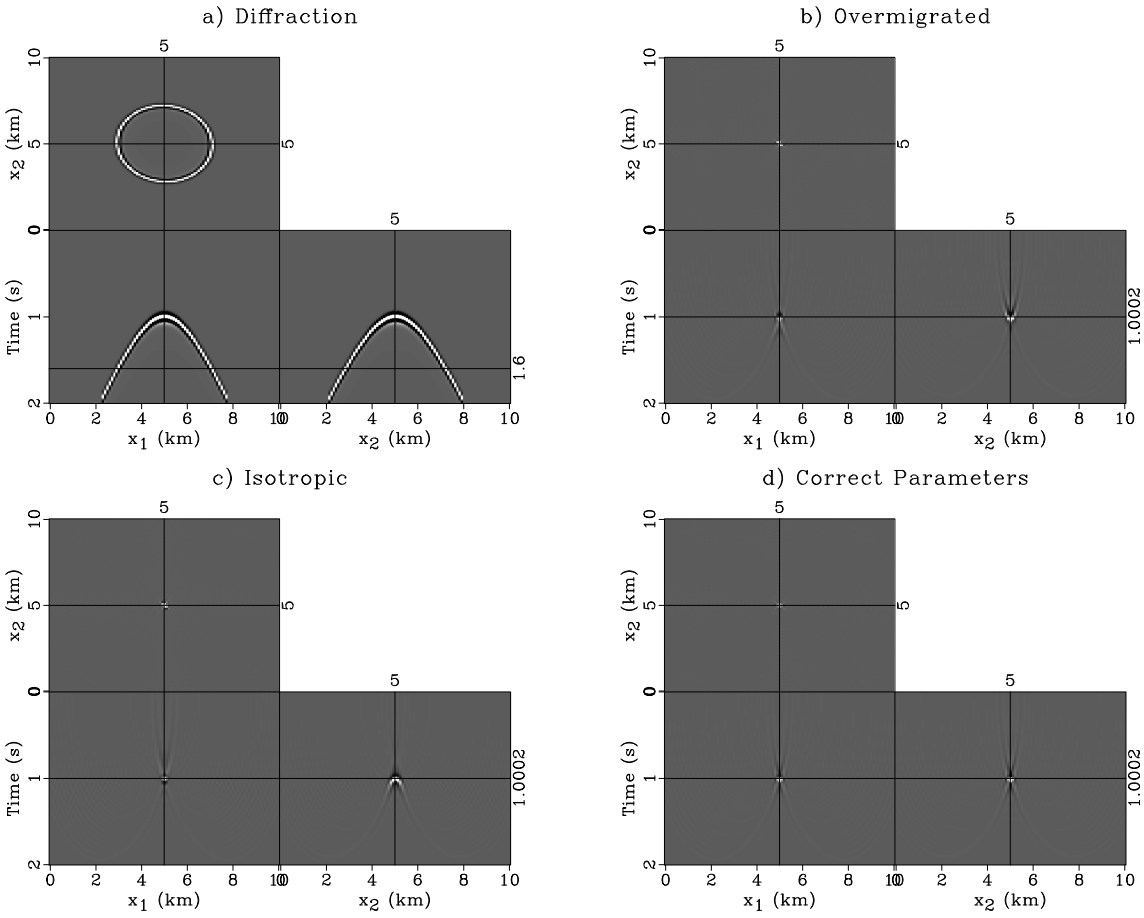


Figure 1: (a) A single azimuthally anisotropic diffraction. (b) The diffraction migrated by velocity continuation using correct parameters except σ=10, resulting in overmigration along x_2. (c) Migration using the correct W_{11}, but assuming isotropy. The result is now undermigrated along x_2. (d) Migration using correct parameters. The image is well focused in both directions. vc3/threedim images

In practice, a conventional in-line 2D velocity analysis directly yields W_{11} from $1/V_{x_1}^2$, so Figure 2 could illustrate a realistic scenario for using 3D velocity continuation to improve upon a previous isotropic velocity model. In such a case, one would use previous V_{x_1} picks to hold W_{11} constant, and then effectively test a variety of W_{12} and W_{22} values. Since W_{11} and W_{22} are measured with respect to the survey coordinates, either (or both) can be measured independently via a single-azimuth semblance scan, along x_1 or x_2, respectively. The best isotropic velocity based on a fully multiazimuth semblance scan will generally not represent either W_{11} or W_{22}, but it can help limit the range of test parameters. Note that our method does not require prior knowledge of the velocity model, but without prior knowledge, the kurtosis measure remains a 6D volume. Although more difficult to visualize, the 6D kurtosis volume is computationally just as easily scanned for optimal imaging parameters as the 2D map in Figure 2.

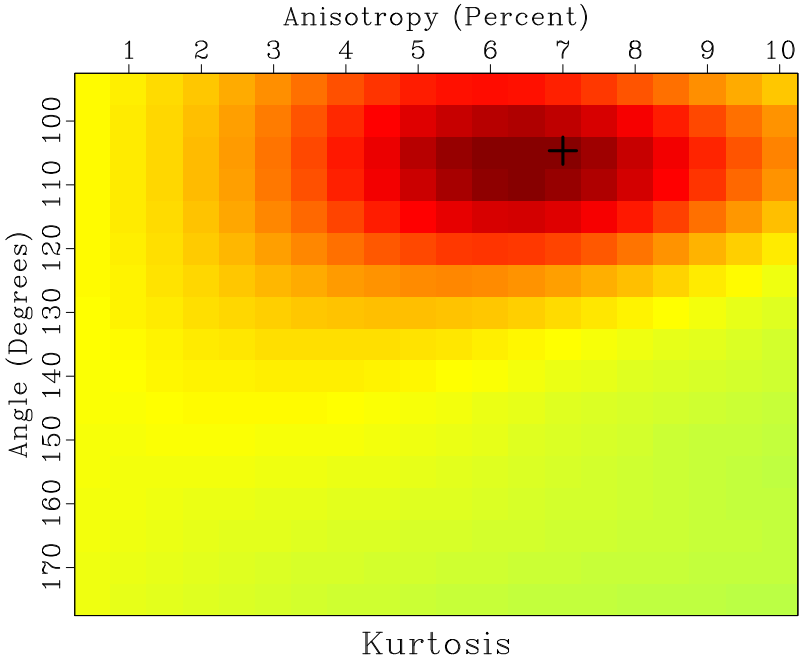


Figure 2: Kurtosis values for the velocity continuation of the diffraction in Figure 1a. The map covers a range of anisotropy and angle values with an increment in β of 5° and an increment in σ of 0.5%. The correct values at 105° and 7% anisotropy (indicated by crosshairs) coincide with the peak of the kurtosis map. vc3/threedim focus

In the next example, we illustrate the concept of applying 3D anisotropic velocity continuation to diffraction imaging and fracture characterization. Figure 3a shows a 3D synthetic post-stack diffraction data set, equivalent to the ideal separation of diffractions from specular reflections in post-stack data following (16). A fault map from (19) (shown in Figure 3a) was digitized and used to create a 3D fracture model. Each fault location was used to generate a point diffraction in a homogeneous anisotropic background via equation 1. A timeslice of the modeled diffraction data is shown in Figure 3b. The faults in the model typically have a strike of 112°, and

in cases where faults and nearby fractures (which more likely influence the seismic velocity) are similarly aligned, the fast direction of seismic wave propagation tends to align with their strike. By assuming a typical tight sandstone velocity of V_{fast}=4.0 km/s with 3% anisotropy, we choose the modeling $\mathbf{W}$ to be comprised of W_{11}=0.0659, W_{22}=0.0631, and W_{12}=0.0014 (all in s^2/km^2). This results in a fast velocity direction along the strike of the faults. In Figure 3d, we see that 3D velocity continuation using the correct parameters (again found by maximum kurtosis) allows the faults to be clearly imaged. If an intermediate isotropic velocity model is used, as in Figure 3c, the diffractions are still imaged, but they are not as well-focused compared to the anisotropically migrated diffractions in Figure 3d. Conventionally, diffraction arrivals such as those in Figure 3a may be viewed as noise, but by separating them and treating them as signal, we can see here that imaging of steep and detailed features while simultaneously extracting anisotropy information may be possible.

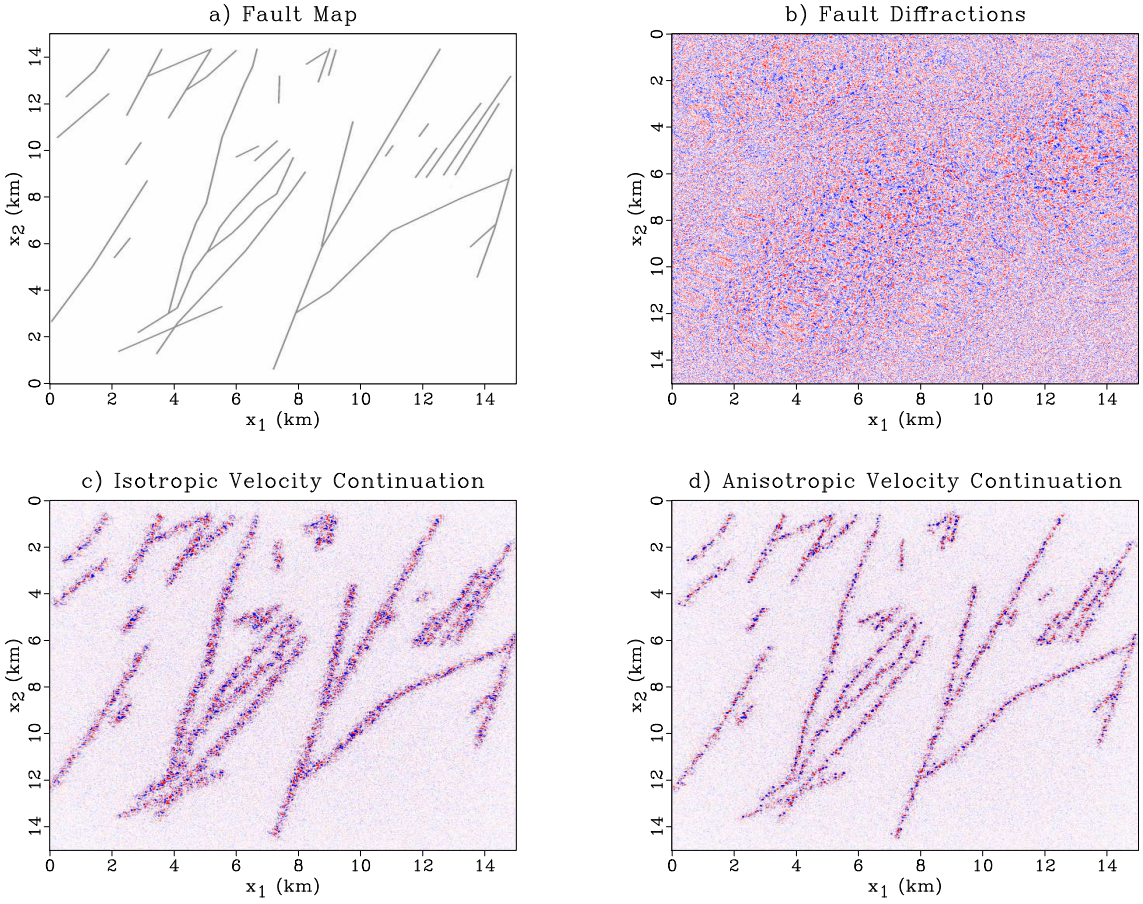


Figure 3: (a) Fault map from Northwest Scotland (19) used to model diffraction data. (b) Synthetic post-stack diffraction data modeled using equation 1 and a 3D model based on the fault map in (a). (c) Diffractions from (b) migrated using an isotropic velocity model. (d) Diffractions from (b) migrated by anisotropic 3D velocity continuation. vc3/fracs images-mig-all

DISCUSSION AND CONCLUSIONS

By extending time-domain velocity continuation to the azimuthally anisotropic 3D case, we have combined the concepts of azimuthal imaging and diffraction imaging. We assume a three-parameter migration slowness model that allows velocity to vary elliptically with azimuth. We have provided simple examples to illustrate the potential application of our method to fracture characterization through diffraction imaging. By treating diffractions as signal, our method performs azimuthal analysis on post-stack data, without the requirement for common-offset-vector or offset-vector-tile binning schemes. This is possible because, unlike reflections, diffractions can preserve azimuthal information post-stack. Post-stack data volumes have obvious advantages over pre-stack or vector-binned data for analysis, including smaller memory size, and improved signal-to-noise ratio.

Allowing azimuthal variation in the migration velocity will result in improved imaging, which is clearly a benefit of 3D velocity continuation. However, the potential for fracture characterization may be even more useful. Our method has many of the same ideas as the azimuthal imaging and fracture characterization flow proposed by (33) for reflection data. Under the velocity continuation framework, we can extend the azimuthal imaging idea to 3D diffraction imaging. Since diffraction-generating fractures and faults are often nearly vertical and preferentially aligned, they can be associated with azimuthal anisotropy. (16) demonstrate that it is possible to separate diffractions from specular reflections, and then image their associated discontinuities through the use of velocity continuation. Their method operates on post-stack data, as they show that diffractions are highly sensitive to migration velocity, even in the zero-offset case. (3) have successfully demonstrated zero-offset diffraction image focusing as a fracture detection attribute on azimuth-sectored 3D field data. Our proposed method uses multi-azimuth image focusing primarily as a velocity analysis criterion, but kurtosis could also be used as an image attribute. In cases where subsurface fractures cause azimuthal anisotropy, kurtosis as an attribute may be indicative of fracture intensity (3). By applying velocity continuation to 3D diffraction imaging, one may be able to estimate both the orientation and the intensity of fractures from the resulting anisotropic velocity model and maximum kurtosis volumes, respectively. This information can be useful in reservoir development, as it can provide insight to subsurface fluid flow behavior.

Although the spectral implementation of our method allows a range of possible images to be computed efficiently, it demands large amounts of memory to store a suite of images as well as the kurtosis volume. Modern computational hardware makes our approach feasible as-is, especially for target-oriented imaging and analysis strategies. Future studies may lead to better optimization-based approaches to finding local kurtosis maxima, in which case, our method could be practical for dense parameter estimation throughout full 3D volumes.

The underlying strategy of velocity continuation is to simultaneously estimate the velocity model as the data are imaged. This is beneficial in the case of azimuthal

anisotropy discussed here, as the ambiguity between structural heterogeneity and anisotropy is handled without the need for iteration. Other multi-parameter seismic imaging problems, such as converted-wave imaging, which are also conventionally handled by iterative flows, could also benefit from pre-stack versions of the 3D velocity continuation strategy.

ACKNOWLEDGMENTS

We would like to thank Vladimir Bashkardin, Hesam Kazemeini, Siwei Li, Yang Liu, and Xiaolei Song for helpful suggestions and Peter Eichhubl and Pete Hargrove for providing the example fracture map. We would also like to thank the three anonymous reviewers for comments helpful in improving this paper. Funding for this project is provided by RPSEA through the "Ultra-Deepwater and Unconventional Natural Gas and Other Petroleum Resource" program authorized by the U.S. Energy Policy Act of 2005. RPSEA (www.rpsea.org) is a nonprofit corporation whose mission is to provide a stewardship role in ensuring the focused research, development and deployment of safe and environmentally responsible technology that can effectively deliver hydrocarbons from domestic resources to the citizens of the United States. RPSEA, operating as a consortium of premier U.S. energy research universities, industry, and independent research organizations, manages the program under a contract with the U.S. Department of Energy's National Energy Technology Laboratory.

REFERENCES

[1] Adler, F., 2002, Kirchhoff image propagation: Geophysics, **67**, 126–134.

[2] Al-Dajani, A., and T. Alkhalifah, 2000, Reflection moveout inversion for horizontal transverse isotropy: Accuracy, limitation, and acquisition: Geophysics, **65**, 222–231.

[3] Al-Dajani, A., and S. Fomel, 2010, Fractures detection using multi-azimuth diffractions focusing measure: is it feasible?: Expanded Abstracts 80^{th} SEG International Convention, 287–291.

[4] Alkhalifah, T., and S. Fomel, 1997, Residual migration in VTI media using anisotropy continuation, *in* SEP-94: Stanford Exploration Project, 327–337.

[5] Cary, P. W., 1999, Common-offset-vector gathers: an alternative to cross-spreads for wide-azimuth 3-D surveys: Expanded Abstracts 69^{th} SEG International Convention, 1496–1499.

[6] Claerbout, J. F., 1986, Velocity extrapolation by cascaded 15 degree migration, *in* SEP-48: Stanford Exploration Project, 79–84.

[7] Courant, R., and D. Hilbert, 1989, Methods of mathematical physics: John Wiley & Sons.

[8] Crampin, S., 1984, Effective anisotropic elastic constants for wave propagation through cracked solids: Geophysical Journal of the Royal Astronomical Society, **76**, 135–145.

[9] Deregowski, S. M., 1990, Common-offset migrations and velocity analysis: First Break, **08**, 224–234.
[10] Duchkov, A., and M. de Hoop, 2009, Velocity continuation in the downward continuation approach to seismic imaging: Geophysical Journal International, **176**, 909–924.
[11] Dunkin, J. W., and F. K. Levin, 1971, Isochrons for a three-dimensional seismic system: Geophysics, **36**, 1099–1137. (Discussion in GEO-37-03-0545-0545 with reply by authors).
[12] Fomel, S., 1994, Method of velocity continuation in the problem of temporal seismic migration: Russian Geology and Geophysics, **35**, 100–111.
[13] ——, 2003a, Time-migration velocity analysis by velocity continuation: Geophysics, **68**, 1662–1672.
[14] ——, 2003b, Velocity continuation and the anatomy of residual prestack time migration: Geophysics, **68**, 1650–1661.
[15] ——, 2007, Local seismic attributes: Geophysics, **72**, A29–A33.
[16] Fomel, S., E. Landa, and T. Taner, 2007, Poststack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, U89–U94.
[17] Fowler, P., 1984, Velocity independent imaging of seismic reflectors: 54th Ann. Internat. Mtg, Soc. of Expl. Geophys., Session:S1.8.
[18] Grechka, V., and I. Tsvankin, 1998, 3-D description of normal moveout in anisotropic inhomogeneous media: Geophysics, **63**, 1079–1092.
[19] Hargrove, P., 2010, Fault-related fracture systems in the Cambrian Eriboll Formation, Northwest Scotland: A field and petrographic study of a tight gas sandstone analog, m.Sc. Geo.Sci. thesis: The University of Texas at Austin.
[20] Harlan, W. S., J. F. Claerbout, and F. Rocca, 1984, Signal/noise separation and velocity estimation: Geophysics, **49**, 1869–1880.
[21] Hubral, P., M. Tygel, and J. Schleicher, 1996, Seismic image waves: Geophysical Journal International, **125**, 431–442.
[22] Iversen, E., 2006, Velocity rays for heterogeneous anisotropic media: Theory and implementation: Geophysics, **71**, T117–T127.
[23] Jenner, E., 2008, Recognizing apparent P-wave azimuthal velocity anisotropy caused by lateral velocity variations: The 13^{th} Annual International Workshop on Seismic Anisotropy, 71–72.
[24] Larner, K., and C. Beasley, 1987, Cascaded migrations - Improving the accuracy of finite-difference migration: Geophysics, **52**, 618–643. (Errata in GEO-52-8-1165).
[25] Levin, F. K., 1985, Apparent velocity from dipping interface reflections: Geophysics, **50**, 2026–2032.
[26] Mikulich, W., and D. Hale, 1992, Steep-dip v(z) imaging from an ensemble of Stolt-like migrations: Geophysics, **57**, 51–59.
[27] Novais, A., J. Costa, and J. Schleicher, 2006, Velocity determination by image-wave remigration: Expanded Abstracts 68^{th} EAGE International Conference, P157:1–4.
[28] Rothman, D. H., S. A. Levin, and F. Rocca, 1985, Residual migration - Applications and limitations: Geophysics, **50**, 110–126.
[29] Schleicher, J., and R. Alexio, 2007, Time and depth remigration in elliptically

anisotropic media using image-wave propagation: Geophysics, **72**, S1–S9.
[30] Schleicher, J., J. Costa, and A. Novais, 2008a, Time-migration velocity analysis by image-wave propagation of common-image gathers: Geophysics, **73**, VE161–VE171.
[31] Schleicher, J., A. Novais, and J. Costa, 2008b, Vertical image waves in elliptically inhomogeneous media: Studia Geophysica et Geodaetica, **52**, 101–122.
[32] Sicking, C., and S. Nelan, 2008, Azimuthal attribute methodologies: comparison of NMO super gathers, sectored isotropic migration, iterative azimuthal migration: Expanded Abstracts 78^{th} SEG International Convention, 983–987.
[33] Sicking, C., S. Nelan, and W. McLain, 2007, 3D azimuthal imaging: Expanded Abstracts 77^{th} SEG International Convention, 2364–2367.
[34] Thomsen, L., 1986, Weak elastic anisotropy: Geophysics, **51**, 1954–1966. (Discussion in GEO-53-04-0558-0560 with reply by author).
[35] Tsvankin, I., 1997, Reflection moveout and parameter estimation for horizontal transverse isotropy: Geophysics, **62**, 614–629.
[36] ———, 2005, Seismic signatures and analysis of reflection data in anisotropic media: Elsevier.
[37] Vermeer, G. J. O., 1999, Fundamentals of 3-D seismic survey design: PhD thesis, Delft University of Technology.
[38] Wiggins, R., 1978, Minimum entropy deconvolution: Geoexploration, **16**, 21–25.

Time-frequency analysis of seismic data using local attributes

*Guochang Liu[1], Sergey Fomel[2], Xiaohong Chen[1]**

ABSTRACT

Time-frequency analysis is an important technology in seismic data processing and interpretation. To localize frequency content in time, we have developed a novel method for computing a time-frequency map for nonstationary signals using an iterative inversion framework. We calculated time-varying Fourier coefficients by solving a least-squares problem that uses regularized nonstationary regression. We defined the time-frequency map as the norm of time-varying coefficients. Time-varying average frequency of the seismic data can also be estimated from the time-frequency map calculated by our method. We tested the method on benchmark synthetic signals and compared it with the well-known Stransform. Two field data examples showed applications of the proposed method for delineation of sand channels and for detection of low-frequency anomalies.

INTRODUCTION

Time-frequency decomposition maps a 1D signal into a 2D signal of time and frequency, and describes how the spectral content of the signal changes with time. Time-frequency analysis has been used extensively in seismic data processing and interpretation, including attenuation measurement (30), direct hydrocarbon detection (3), and stratigraphic mapping (28). The widely used short-time Fourier transform (STFT) method produces a time-frequency spectrum by taking the Fourier transform of data windows (7), which leads to a tradeoff between temporal and spectral resolution.

Over the past two decades, wavelet-based methods have been applied to time-frequency analysis of seismic data. (4) compare wavelet-based with Fourier-based methods for performing time-frequency analysis on seismic data and showed that the wavelet-based method improves spectral resolution. The continuous-wavelet transform (CWT) provides a time-scale map, known as a scalogram (31), rather than a time-frequency spectrum. Because a scale represents a frequency band, (15) choose their scale to be inversely proportional to the center frequency of the wavelet, which allowed them to transform their scalogram into a time-frequency map. (34) provide a novel method of obtaining such a time-frequency map by taking a Fourier transform of the inverse CWT.

*e-mail: sergey.fomel@beg.utexas.edu

Wigner-Ville distribution (WVD) (40) represents time-frequency components by using the reverse of the signal as an analysis window function. WVD varies resolution in the time-frequency plane by providing good temporal resolution at high frequencies and good frequency resolution at a low-frequency (6). Applications of WVD are hindered by cross-term interference, which can be suppressed by using appropriate kernel functions. A smoothed pseudo-WVD with independent time and frequency functions as kernels achieves a relatively good resolution in both time and frequency (18). (41) employed smoothed pseudo-WVD with a Gaussian kernel function to reduce cross-term interference.

The S-transform is an invertible time-frequency analysis technique that combines elements of wavelet transforms and short-time Fourier transforms. The S-transform with an arbitrary and varying shape was applied to seismic data analysis by (29). Matching pursuit (26), which was also applied in seismic signal analysis (4; 3; 23; 22; 39), decomposes a seismic trace into a series of wavelets that belong to a comprehensive dictionary of functions. These wavelets are selected so as to best match signal structures. The spectrum of the signal is then the time-shifted sum of each of the wavelets.

Local attributes (9) can adaptively measure timevarying seismic signal characteristics in the neighborhood of each data point. Local attributes have been successfully applied to seismic image registration (13), phase detection (38; 14), and stacking (21; 20). A natural extension of local attributes, regularized nonstationary regression, decomposes input data into a number of nonstationary components (11; 24; 19).

In this paper, we propose a new method of time-frequency analysis in which we use time-varying Fourier coefficients to define a time-frequency map. As in regularized nonstationary regression, shaping regularization (10) constrains continuity and smoothness of the coefficients. Given the close connection between Fourier transforms and the least-squares norm, the least-squares approach to time-frequency analysis is not new, having been used previously, for example, by (42). What is novel about our approach is the use of regularization for explicitly controlling the time resolution of time-frequency representations.

The paper is organized as follows.We first describe the proposed method for time-frequency analysis. Then we show how to compute the time-varying average frequency from the time-frequency map. We use benchmark synthetic examples to test the performance of the proposed method. Finally, we apply the proposed method to channel detection and low-frequency anomaly detection in field seismic data.

TIME-FREQUENCY ANALYSIS USING LOCAL ATTRIBUTES

The Fourier transform has found various applications in signal analysis. The classic Fourier transform indicates the presence of different frequencies within the analysis window, but does not show where in that window the particular frequency components

reside. Localized frequency information can be obtained by computing the Fourier transform with a temporally shifted window. Such an approach to time-frequency analysis is known as the STFT (1). The window function is commonly parameterized by window size, overlap, and taper. Once the window function has been chosen for the STFT, temporal and spectral resolutions are fixed for the entire time-frequency map.

The S-transform (36) is similar to the STFT, but with a Gaussian-shaped window whose width scales inversely with frequency. The expression of the S-transform is

$$S(\tau, f) = \int_{-\infty}^{\infty} S(t) \left\{ \frac{|f|}{\sqrt{2\pi}} \exp\left[\frac{-f^2(\tau - t)^2}{2} \right] \exp(-2\pi i f t) \right\} dt, \tag{1}$$

where $s(t)$ is a signal and τ is a parameter which controls the position of the Gaussian window. The S-transform is conceptually a hybrid of the STFT and wavelet analysis, containing elements of both but having its own properties. Like STFT, the S-transform uses a window to localize the complex Fourier sinusoid, but, unlike the STFT and analogously to the wavelet transform, the width of the window scales with frequency.

Consider a signal $f(x)$ on $[0, L]$. The Fourier series, assuming a periodic extension of the boundary conditions, can be expressed as

$$f(t) \approx a_0 + \sum_{k=1}^{\infty} \left[a_k \cos\left(\frac{2k\pi t}{L} \right) + b_k \sin\left(\frac{2k\pi t}{L} \right) \right], \tag{2}$$

where a_k and b_K are the series coefficients. The relationship between k and frequency f is $k = Lf$. In the case of the discrete Fourier transform, frequency is finite. Letting $\mathbf{\Psi}_k(t)$ represent the Fourier basis

$$\mathbf{\Psi}_k(t) = \begin{bmatrix} \mathbf{\Psi}_{1k}(t) \\ \mathbf{\Psi}_{2k}(t) \end{bmatrix} = \begin{bmatrix} \cos\left(\frac{2k\pi t}{L}\right) \\ \sin\left(\frac{2k\pi t}{L}\right) \end{bmatrix}, \tag{3}$$

and $\mathbf{C}_k$ represent the series coefficients

$$\mathbf{C}_k = \begin{bmatrix} a_k & b_k \end{bmatrix}, \tag{4}$$

where $b_0 = 0$, equation 1 can be written as

$$f(t) = \sum_{k=0}^{\infty} \mathbf{C}_k \mathbf{\Psi}_k(t). \tag{5}$$

If frequency is finite, the range of k becomes $[0, N]$, $N = k_{max} = Lf_{max}$. In linear notation, $\mathbf{C}_k$ can be obtained by solving the least-squares problem

$$\min_{\mathbf{C}_k} \left|\left| f(t) - \sum_{k=0}^{N} \mathbf{C}_k \mathbf{\Psi}_k(t) \right|\right|_2^2, \tag{6}$$

where ||||denotes the squared $L-2$ norm of a function. By allowing coefficients $\mathbf{C}_k$ to change with time t, we can define the timevarying coefficients $\mathbf{C}_k(t)$ via the following least-squares problem

$$\min_{\mathbf{C}_k} \sum_{k=0}^{N} ||f(t) - \mathbf{C}_k \mathbf{\Psi}_k(t)||_2^2 . \tag{7}$$

The Fourier coefficient $\mathbf{C}_k(t)$ in equation 7 is a function of time t and frequency $f = kL$ and $\mathbf{C}_k(t) = [a_k \quad b_k]$. The numerical support of frequency f can be between zero and the Nyquist frequency (7), and the interval of frequency can be $\Delta f = 1L$. In practical applications, the range of frequencies can also be assigned by the user. In matrix notation, equation 7 can be written as

$$[f(t) \quad f(t)... \quad f(t)]^T \approx [D\{\mathbf{\Psi}_1(t) \quad ... \quad \mathbf{\Psi}_N(t)\}] [\mathbf{C}_1(t) \quad ... \quad \mathbf{C}_N(t)]^T , \tag{8}$$

where $D\{...\}$ denotes a diagonal matrix which is composed from the elements of $\mathbf{Psi}_k(t)$.

The problem of minimization in equation 7 is mathematically ill-posed because it is severely underconstrained: There are more unknown variables than constraints. To solve this ill-posed problem, we force the coefficients $\mathbf{C}_k(t)$ to have a desired behavior, such as smoothness. With the addition of a regularization term, equation 7 becomes

$$\min_{\mathbf{C}_k(t)} \sum_{k=0}^{N} ||f(t) - \mathbf{C}_k(t) \mathbf{\Psi}_k(t)||_2^2 + R[\mathbf{C}_k(t)] , \tag{9}$$

where R denotes the regularization operator. If we use classic Tikhonov regularization (37), equation 9 can be written as

$$\min_{\mathbf{C}_k(t)} \sum_{k=0}^{N} ||f(t) - \mathbf{C}_k(t) \mathbf{\Psi}_k(t)||_2^2 + \varepsilon^2 \sum_{k=0}^{N} ||\mathbf{D}[\mathbf{C}_k(t)]||_2^2 , \tag{10}$$

where $\mathbf{D}$ is the Tikhonov regularization operator (roughening operator) and ε is a scalar regularization parameter.

Shaping regularization (10) provides a particularly convenient method of enforcing smoothness in iterative optimization schemes. In the appendix, we review the method of shaping regularization in detail. (11) used shaping regularization to constrain the problem of nonstationary regression. In this paper, we use shaping regularization, analogous to the problem of nonstationary regression, to constrain estimated coefficients. We choose our shaping operator to be Gaussian smoothing with an adjustable radius. In shaping regularization, the radius of the Gaussian smoothing operator controls the smoothness of coefficients $\mathbf{C}_k(t)$.

Once we obtain time-varying Fourier coefficients $\mathbf{C}_k(t) = [a_k(t) \quad b_k(t)]$, the time-frequency map is defined as

$$F(t, f = k/L) = \sqrt{a_k^2(t) + b_k^2(t)}. \tag{11}$$

It is also possible to do an invertible time-frequency transform, as shown by Liu and (24).

Consider a simple signal (Figure 1(a)), which includes three monophonic components at 10, 20, and 30 Hz and two broad-band spikes at 2 and 2.3 s. Time-frequency maps generated by the S-transform and our method are shown, respectively, in Figure 1(b) and 1(c). Frequency components appear to be represented well from low to high frequencies in the proposed method. In comparison with the S-transform, the proposed method provides superior resolution for monophonic waves. At the lower frequencies, the S-transform is as good as the proposed method on the spectral resolution for monophonic waves. Note that the S-transform represents spikes better at high frequencies. However, because the width of analysis window is large at low frequencies, the S-transform provides poor time resolution at low frequencies for spikes. When we use the shaping operator as a regularization term, there is an edge effect from smoothing, which appears near 0 s and 4 s (Figure 1(c)). Figure 1(d) displays the time-frequency map using a different parameter (30-point smoothing radius) to demonstrate adjustable time-frequency representation of the proposed method.

To show the effect of varying frequencies, we provide a composite chirp signal (Figure 2(a)), which includes two parabolic frequencies, each having a constant amplitude. Figure 2(b) and 2(c) shows the results of the S-transform and the proposed method with 10-point smoothing radius, respectively. The two frequency components are detected with higher time and frequency resolution by our method. The S-transform has high spectral resolution near low frequencies, but loses resolution at high frequencies. Resolution of the time-frequency map deteriorates in both methods when the curvature of the time-frequency curve becomes large (as indicated by the arrow).

In the proposed method, the smoothing radius used in shaping regularization is a parameter which controls the smoothness of the model. It is different from the method of the STFT and the S-transform, in which division into local windows is applied to the data to localize frequency content in time. In contrast to the wavelet-based methods (4; 34; 39), our method, as a straightforward extension of the classic Fourier analysis, is different because of the choice of basis functions. The wavelet-based methods need to decide the wavelets, which can be regarded as patterns of seismic data. The closer the selected wavelets are to the true pattern of the input data, the better the result of the time-frequency analysis. In this paper, we use more general Fourier basis functions to compute the time-frequency map.

ESTIMATION OF TIME-VARYING AVERAGE FREQUENCY

Seismic instantaneous frequency is the derivative of the instantaneous phase

$$f_i(t) = \frac{1}{2\pi}\frac{d\phi(t)}{dt}, \tag{12}$$

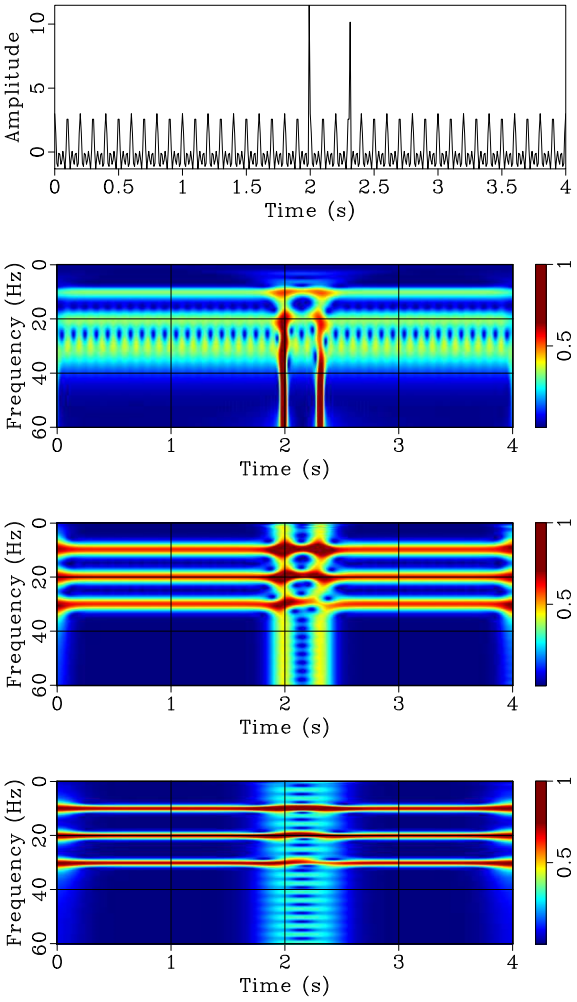


Figure 1: (a) Synthetic signal with three constant frequency components and two spikes. (b) Time-frequency map of the S-transform. (c) Time-frequency map of the proposed method with smoothing radius of 15 points. (d) Time-frequency map of the proposed method with smoothing radius of 30 points. timefreq/syn s-0,st-0,proj-0,proj-0-30

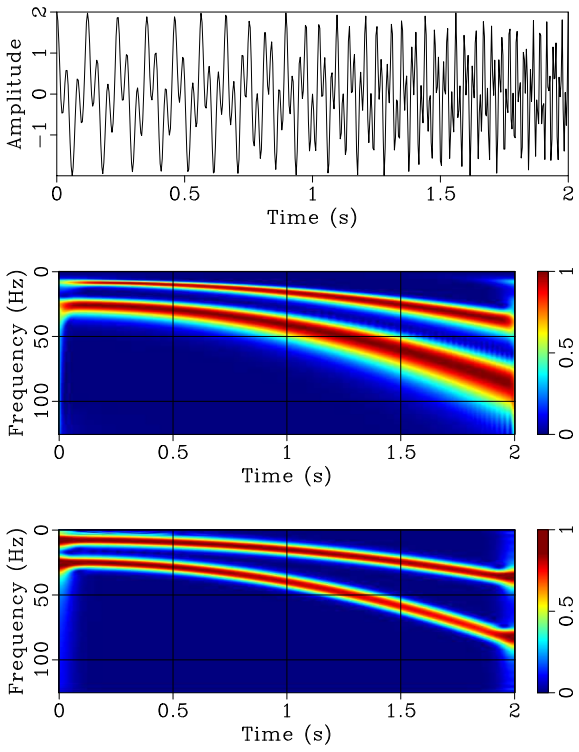


Figure 2: (a) Synthetic chirp signal with two parabolic frequency changes. (b) Time-frequency map of the S-transform. (c) Time-frequency map of the proposed method. timefreq/syn s-1,st-1,proj-1

where $\phi(t)$ is the instantaneous phase. Instantaneous frequency can be estimated directly using a discrete form of equation 12. This estimate is highly susceptible to noise. (9) modified the definition of instantaneous frequency to that of a local frequency by recognizing it as a form of regularized inversion, and by using regularization to constrain continuity and smoothness of the output.

Average frequency can be estimated from the time-frequency map (5; 6; 15; 35; 33). Average frequency at a given time is

$$f_a(t) = \frac{\int f F^2(f,t) df}{\int F^2(f,t) df}, \tag{13}$$

where $F(f,t)$ is the time-frequency map. Average frequency measured by equation 13 is the first moment along the frequency axis of a time-frequency power spectrum. (32) and (2) analyzed the relationship between instantaneous frequency and the time-frequency map in detail. Extraction of the attributes from the time-frequency map of the seismic trace leads to considerable improvement of the signal-to-noise ratio of the attributes (35). We therefore propose applying equation 13 to our time-frequency map to compute the time-varying average frequency.

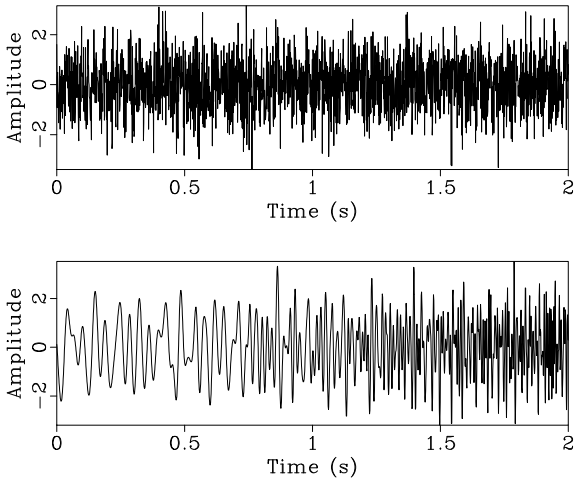


Figure 3: (a) Random reflectivity series. (b) Synthetic seismic trace. timefreq/syn ref-6,s-6

We used a synthetic nonstationary seismic trace to illustrate our approach to estimating time-varying average frequency. Figure 3(b) shows a synthetic seismic trace generated by nonstationary convolution (27) of a random reflectivity series (Figure 3(a)) using a Ricker wavelet, the dominant frequency of which is a function of time, $f_d = 25t^2 + 15$. Figure 9 shows the scaled spectrum of the Ricker wavelet. Both the dominant frequency (white line in Figure 9 and the bandwidth increase with time. We computed the average frequency (black line in Figure 9 using equation 13 from the scaled spectrum of Ricker wavelets. We note that average frequency is larger than the dominant frequency at high frequencies for Ricker wavelets.

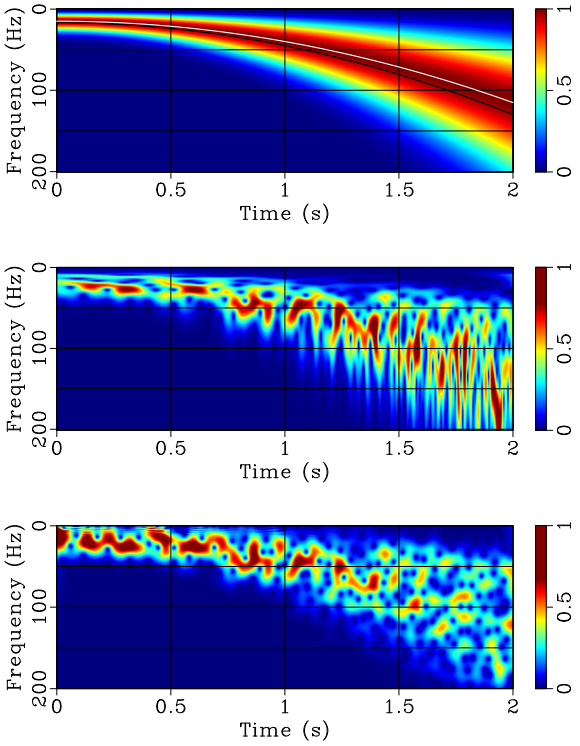


Figure 4: (a) Theoretical time-frequency map, which is scaled by a maximum in the frequency axis. White and black lines indicate dominant frequency and average frequency of Ricker wavelet, respectively. (b) Time-frequency map of the S-transform. (c) Time-frequency map of the proposed method. timefreq/syn spec,st-6,proj-6

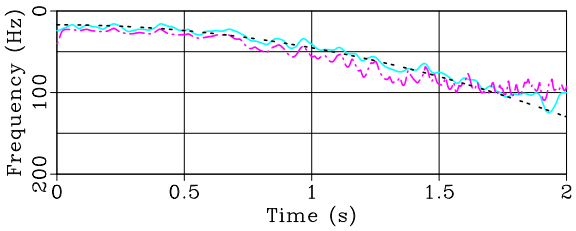


Figure 5: Time-varying average-frequency estimation (blue solid curve: estimated by our method; pink dashed curve: estimated by S-transform; black dot-dashed curve: theoretical curve, which is denoted by black line in Figure 9). timefreq/syn lcf-6

We generated the time-frequency map of the synthetic nonstationary seismic trace using the S-transform (Figure 4(b)) and the proposed method with a 10-point smoothing radius (Figure 4(c)). We observe that the time-frequency map by the proposed method has a bandwidth more similar to that of the time-frequency map of the Ricker wavelet (Figure 9), especially at the high frequencies. We estimated time-varying average frequency curves from the time-frequency map by the proposed method (blue solid curve in Figure 5) and S-transform (pink dashed curve in Figure 5), respectively. Compared with the theoretical curve (black dashed curve in Figure 5), which was computed using equation 13 on the scaled spectrum of Ricker wavelets (Figure 9), the time-varying average frequency estimated by the proposed method is closer to the theoretical one.

EXAMPLES

We demonstrate the effectiveness of the proposed time-frequency analysis on a benchmark synthetic data set and two field data sets. We first use a synthetic seismic trace to illustrate how the proposed method obtains a time-frequency map, and then we test our method on two field data sets with applications to detecting channels and low-frequency anomalies.

Synthetic data

A synthetic seismic trace (Figure 6(a)) was obtained by adding Ricker wavelets with different frequencies and time shifts. The first event is an isolated wavelet with a frequency of 30 Hz, and the second event consists of a 15-Hz wavelet and a 60-Hz wavelet overlapping in time. The last event is a superposition of two 50-Hz wavelets with different arrival times. Figure 6(b) and 6(c) show the time-frequency maps by the S-transform and our method with 15-point smoothing radius, respectively. From the time-frequency map of the first event at 0.2 s, we found that time duration is long at low-frequency in the S-transform. The proposed method produces a more temporally limited ellipsoid spectrum for the first event. Note that the proposed method has higher spatial resolution at 0.6 s and temporal resolution at 1 s. This simple test shows that our method can be effective in representing

3D Gulf of Mexico data for channel detection

Detecting channel structures is a common application of spectral decomposition (28). Different frequency slices show different stratigraphic features. Figure 7 shows field data from the Gulf of Mexico (25). To generate stratal slices (43), we flattened this data set using predictive painting (12). The flattened data set is shown in Figure 8. The flattening procedure can remove structural distortions and allows the interpreter

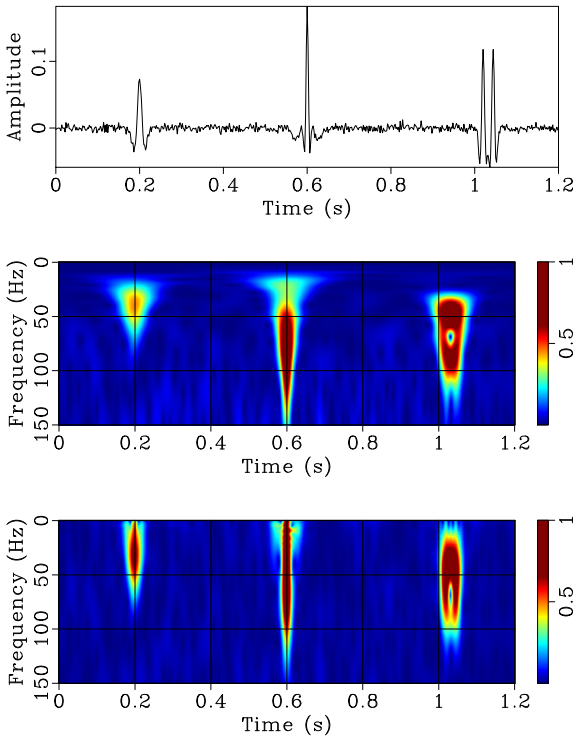


Figure 6: (a) Synthetic seismic trace. (b) Time-frequency map of the S-transform. (c) Time-frequency map of the proposed method. timefreq/syn s-2,st-2,proj-2

to see geologic features as they were emplaced (25). Predictive painting does a careful job of restoring true geological frequencies, as is evident from vertical sections. Figure 9 shows the average spectrum of the data set before flattening (dashed blue curve) and after flattening (solid red curve). We calculated the spectral decomposition of flattened data using the proposed method with a five-point smoothing radius. Figure 10 shows horizon slices from six different volumes at the same level in reference time. The horizontal blue lines on Figure 8b and 8c identify where the time slice is. Comparing seismic amplitudes (Figure 10a) with spectral decomposition at different frequencies (Figures 10b-f), several channels in 30 Hz slice are easily visible. The high-frequency spectral-decomposition maps, such as the 30-, 40-, and 50-Hz slices, can highlight detailed geologic features.

2D data example for low-frequency anomaly detection

Low-frequency anomalies are often attributed to abnormally high attenuation in gas-filled reservoirs and can be used as a hydrocarbon indicator (3). A number of studies have investigated possible mechanisms of low-frequency anomalies associated with hydrocarbon reservoirs, but no adequate explanation is accepted absolutely (8; 16).

Figure 11 shows a poststack field data set with a bandwidth of about 10150 Hz

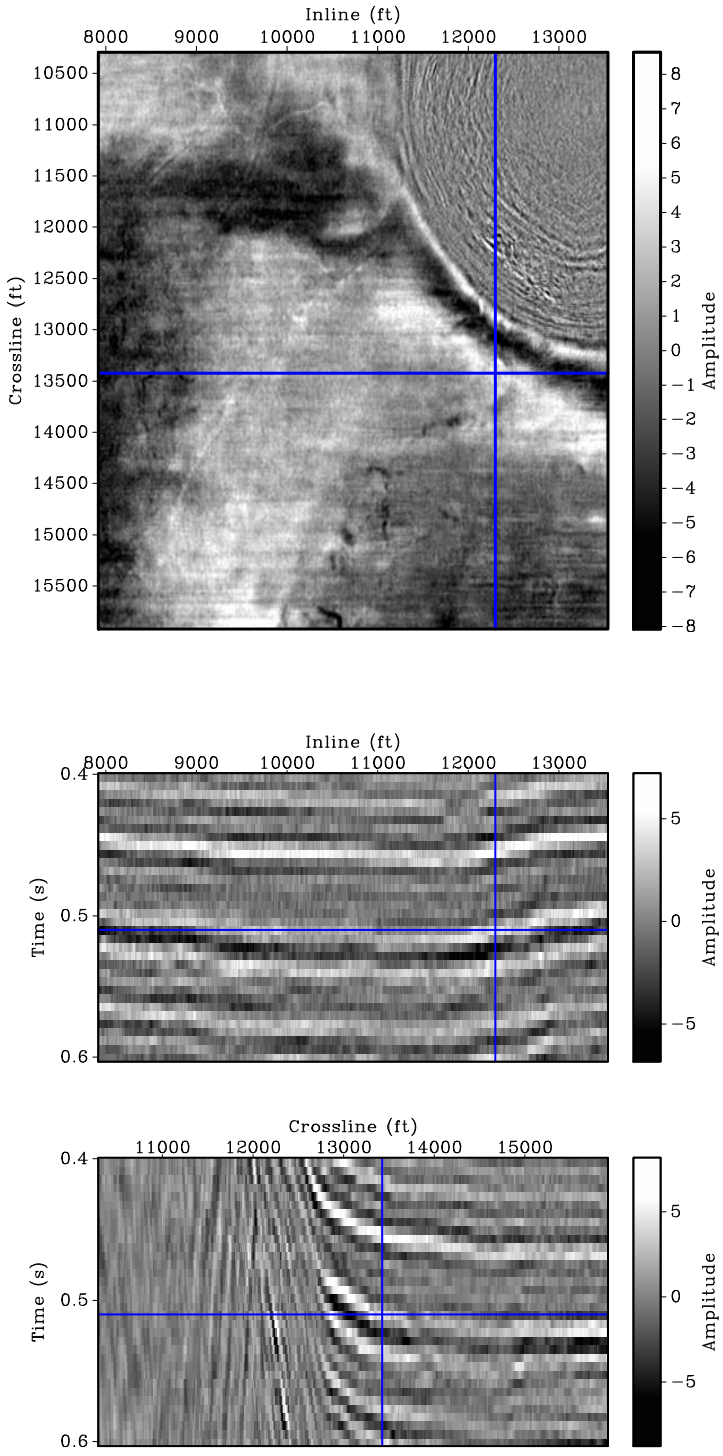


Figure 7: Seismic image from Gulf of Mexico. (a) Time slice. (b) Inline section. (c) Crossline section. Blue lines in (a) identify the location of inline and crossline sections. The horizontal blue lines in (b) and (c) identify where the time slice is. timefreq/chev chev

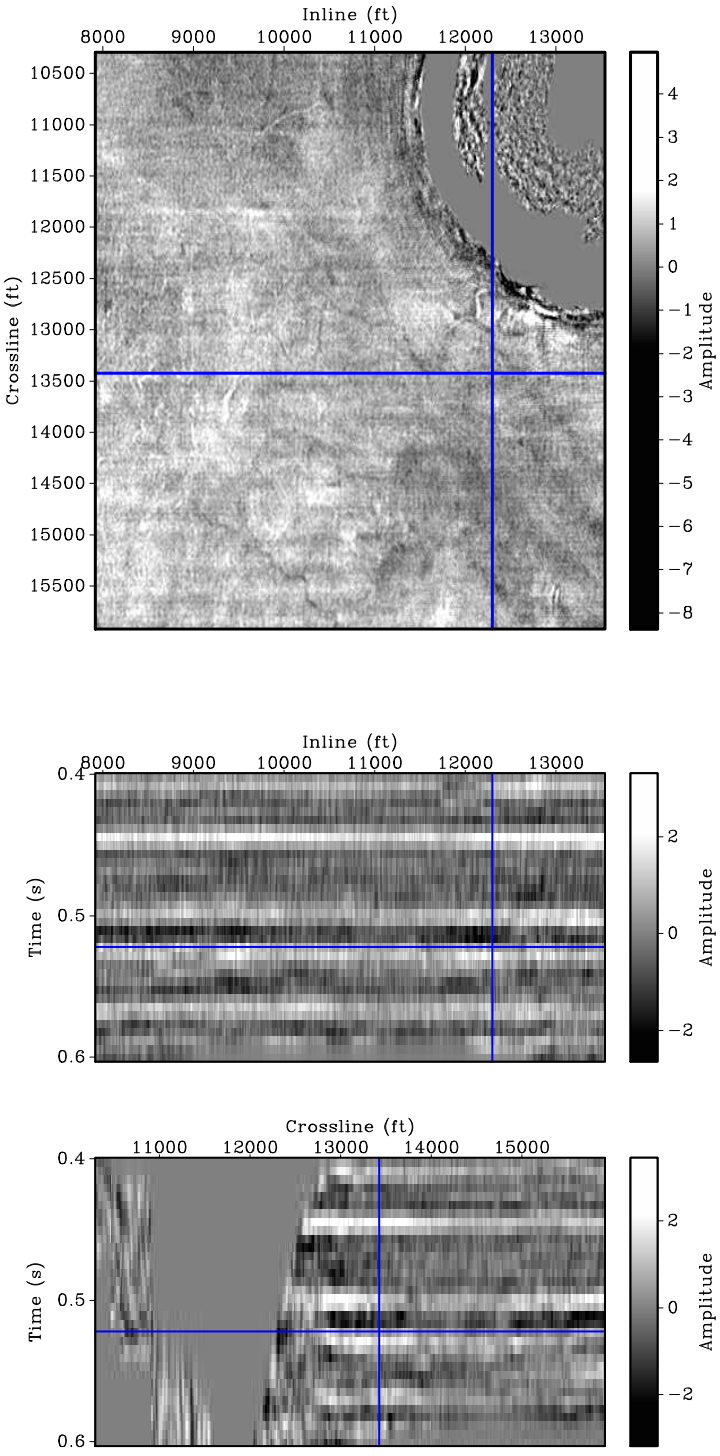


Figure 8: Seismic image from Figure 7 after flattening by predictive painting. (a) Time slice. (b) Inline section. (c) Crossline section. The blue lines on (a) identify the location of inline and crossline sections. The horizontal blue lines on (b) and (c) identify where the time slice is. timefreq/chev flats

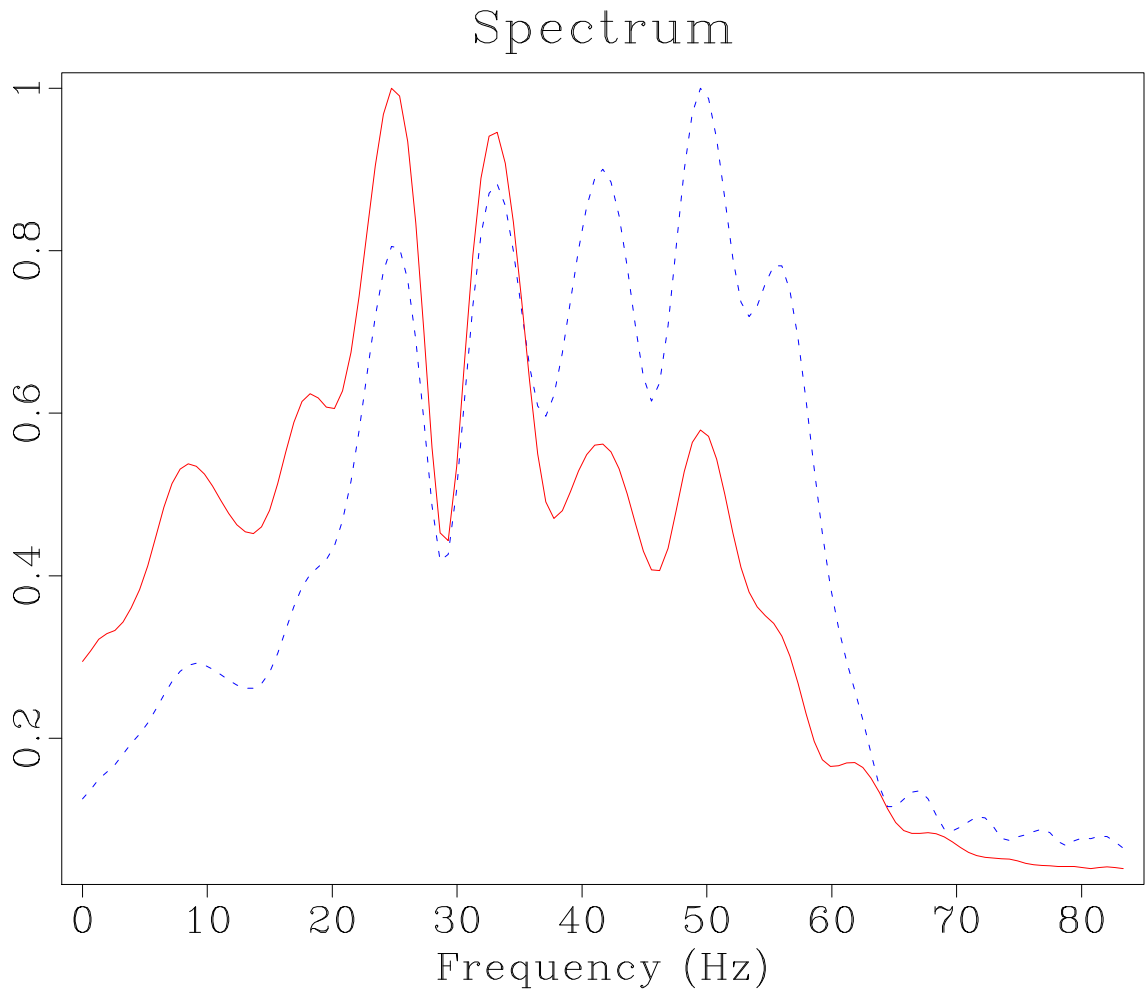


Figure 9: Average data spectrum before flattening (dashed blue curve) and after flattening (solid red curve). timefreq/chev spec

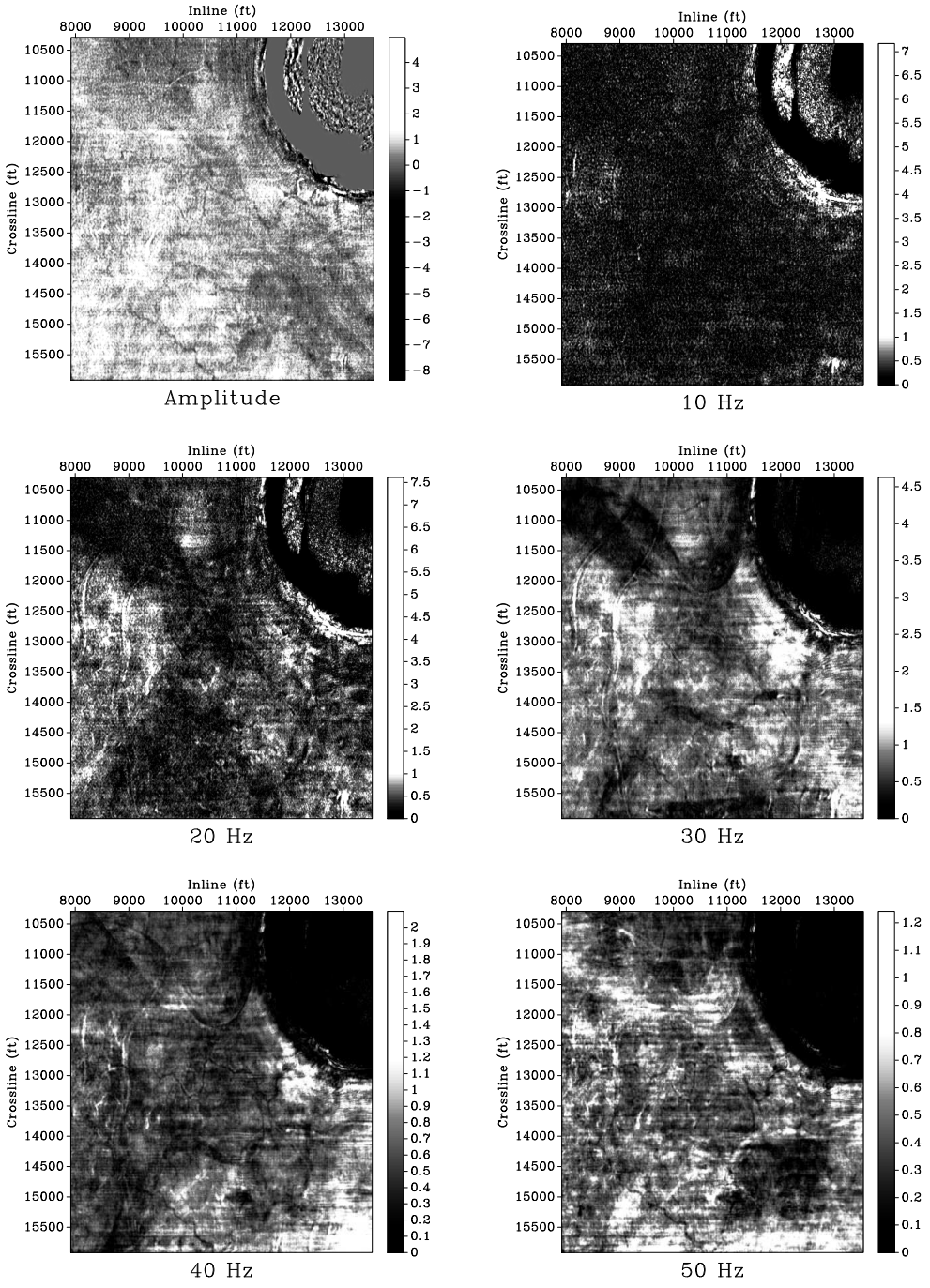


Figure 10: Comparison of horizon slices, from four different volumes at the same level: (a) conventional amplitude; and spectraldecomposition at (b) 10 Hz, (c) 20 Hz, (d) 30 Hz, (e) 40 Hz, and (f) 50 Hz. The slice of 30 Hz displays most clearly visible channel features. timefreq/chev slice,slice10,slice20,slice30,slice40,slice50

(Figure 12).We used our proposed method with a ten-point smoothing radius to generate a time-frequency spectral map. Figure 13 shows the time-frequency map of one trace from the field data set. Note that our method has a higher temporal resolution than that of the S-transform, especially at 0.6-0.8 s.We can observe a general decay of frequency with time caused by seismic attenuation from the time-frequency map of one trace (Figure 13). The low-frequency anomaly is shown at 1.2-1.3 s (arrows). Figures 14 and 15 show single-frequency sections at 15, 31, and 70 Hz from two methods, the S-transform and the proposed method. We further found that the proposed method provides higher temporal and spatial resolution, especially in low-frequency sections. Figures 14(a) and 15(a) both show high-amplitude, low-frequency anomalies (ellipse). However, these anomalies gradually disappear in the high-frequency section (Figures 14(b) and 14(c) and 15(b) and 14(c)). Note that the spatial resolution of low-frequency anomalies in the proposed method (Figure 15(a)) is higher than that in the S-transform (Figure 14(a)). We also computed the time-varying average frequency section for this example using equation 13 (Figure 16). We find that the average frequency is about 60 Hz in shallow layers, but 25 Hz in deep layers. The average frequency of low-frequency anomalies is very low, just about 15 Hz. This example shows that comparison of single low-frequency sections from the proposed method is able to detect low-frequency anomalies that might be caused by hydrocarbons, as demonstrated in other studies (3; 17; 44). What we demonstrate by this example is that the proposed method can be used to detect low-frequency anomalies in seismic sections. Well control is generally needed to interpret this section more accurately.

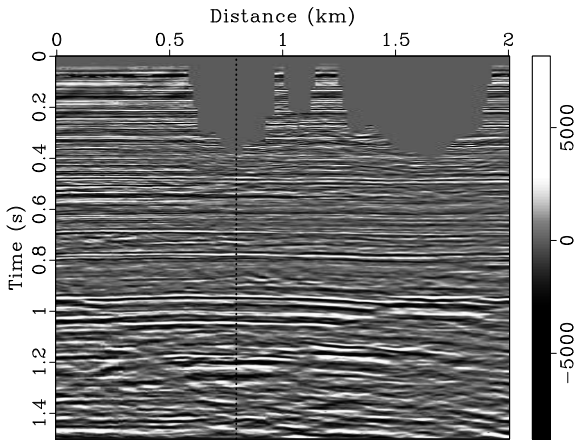


Figure 11: A field marine seismic data set. timefreq/lowf old-1

CONCLUSION

We have presented a novel numerical method for computing time-frequency representation using least-squares inversion with shaping regularization. This time-frequency-analysis technology can be applied to nonstationary signal analysis. The method is

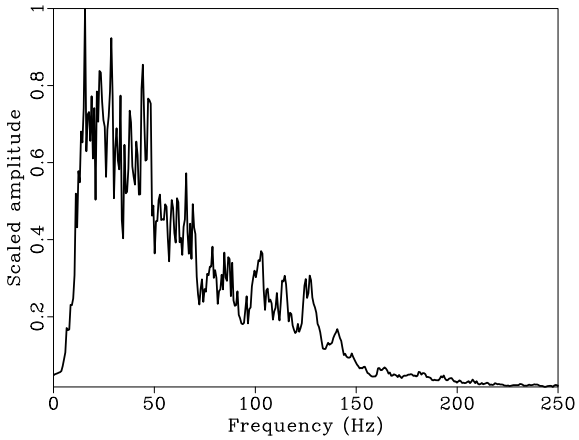


Figure 12: The averaged Fourier spectra of the field data. timefreq/lowf spectra

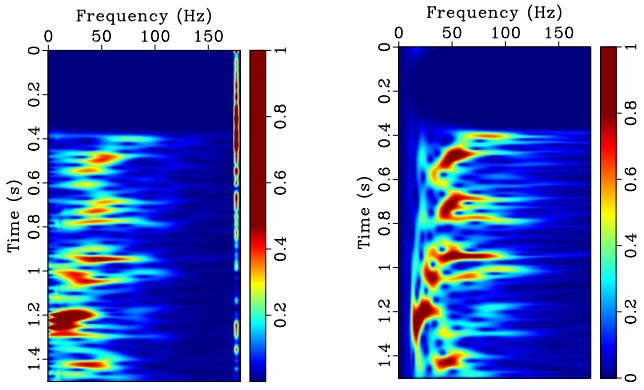


Figure 13: Time-frequency map of one trace (denoted by black line in Figure 12). (a) S-transform; (b) the proposed method. timefreq/lowf tf-l,tf-s

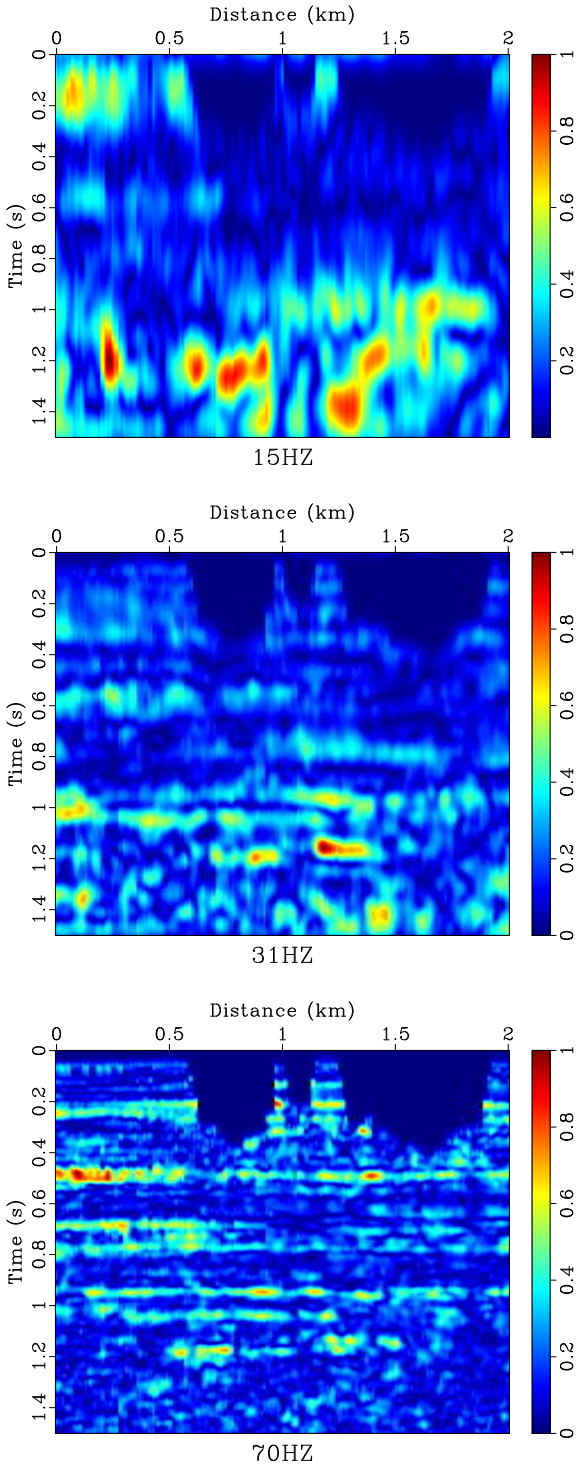


Figure 14: Comparison of common-frequency slices from S-transform; (a) 10 Hz, (b) 31 Hz, and (c) 70 Hz. The lowfrequency abnormality is indicated by an ellipse. timefreq/lowf s1,s2,s3

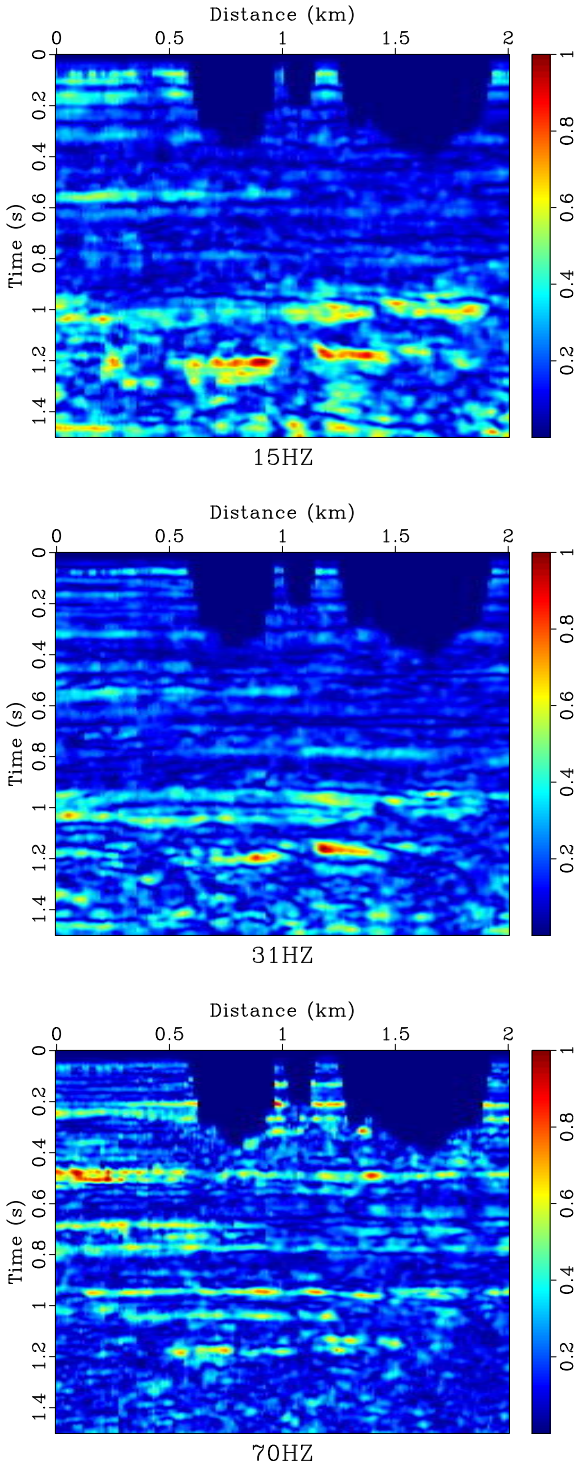


Figure 15: Comparison of common-frequency slices from the proposed method; (a) 10 Hz, (b) 31 Hz, and (c) 70 Hz. The lowfrequency abnormity is indicated by an ellipse. timefreq/lowf l1,l2,l3

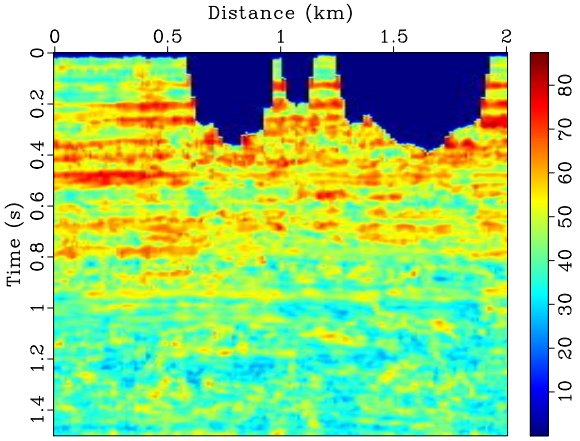


Figure 16: Estimated time-varying average frequency of the field data from time-frequency map. timefreq/lowf lcf-l

a straightforward extension of the classic Fourier analysis. The parameter used in shaping regularization, the radius of the Gaussian smoothing operator, controls the smoothness of time-varying Fourier coefficients. The smooth time-varying average frequency attribute can also be estimated from the proposed time-frequency analysis technology. We have demonstrated applications of the proposed time-frequency analysis for detecting channels and lowfrequency anomalies in seismic images. Finally, we realize that many different algorithms are capable of computing time-frequency representations. We have provided some comparisons of our method with one of them (the S-transform) but cannot make the comparison exhaustive and do not claim that our method will necessarily perform better in all practical situations. Both approaches to time-frequency analysis have advantages and disadvantages. What we see as the main advantages of our method are its conceptual simplicity, computational efficiency, and explicit controls on time and frequency resolution.

ACKNOWLEDGMENTS

We would like to thank Thomas J. Browaeys, Yang Liu, and Mirko van der Baan for inspiring discussions, and the China Scholarship Council (grant 2007U44003) for partial support of the first authors research. This work is financially supported by the National Basic Research Program of China (973 program, grant 2007CB209606) and by the National High Technology Research and Development Program of China (863 program, grant 2006AA09A102-09).We also thank associate editor Kurt J. Marfurt and three anonymous reviewers for their constructive comments, which improved the quality of the paper.

SHAPING REGULARIZATION FOR INVERSE PROBLEMS

In this appendix, we review the theory of shaping regularization in inversion problems. (10) introduces shaping regularization, a general method of imposing regularization constraints. A shaping operator provides an explicit mapping of the model to the space of acceptable models.

Consider a system of linear equation given as $\mathbf{Ax} = \mathbf{b}$, where $\mathbf{A}$ is a forward-modeling operator, $\mathbf{x}$ is the model, and $\mathbf{b}$ is the data. By equation 8, we can find that the proposed time-frequency decomposition can be written as the form $\mathbf{Ax} = \mathbf{b}$. The standard regularized least-squares approach to solving this equation seeks to minimize $||\mathbf{Ax} = \mathbf{b}||_2^2 + \varepsilon^2 ||\mathbf{Dx}||_2^2$, where $\mathbf{D}$ is the Tikhonov regularization operator (37) and ε is scaling.

The formal solution, denoted by $\hat{\mathbf{x}}$, is given by

$$\hat{x} = \left(\mathbf{A}^T\mathbf{A} + \varepsilon^2\mathbf{D}^T\mathbf{D}\right)^{-1}\mathbf{A}^T\mathbf{b}, \tag{A-1}$$

where $\mathbf{A}^T$ denotes the adjoint operator. (10) defined a relation between a shaping operator S and a regularization operator $\mathbf{D}$ as

$$\mathbf{S} = \left(\mathbf{I} + \varepsilon^2\mathbf{D}^T\mathbf{D}\right)^{-1}. \tag{A-2}$$

Substituting equation A-2 into equation A-1 yields a formal solution of the estimation problem regularization by shaping

$$\hat{\mathbf{x}} = \left(\mathbf{A}^T\mathbf{A} + \mathbf{S}^{-1} - \mathbf{I}\right)^{-1}\mathbf{A}^T\mathbf{b} = \left[\mathbf{I} + \mathbf{S}\left(\mathbf{A}^T\mathbf{A} - \mathbf{I}\right)\right]^{-1}\mathbf{S}\mathbf{A}^T\mathbf{b}. \tag{A-3}$$

Introducing scaling of $\mathbf{A}$ by 1 in equation A-3 , we obtain

$$\hat{\mathbf{x}} = \left[\lambda^2\mathbf{I} + \mathbf{S}\left(\mathbf{A}^T\mathbf{A} - \lambda^2\mathbf{I}\right)\right]^{-1}\mathbf{S}\mathbf{A}^T\mathbf{b}. \tag{A-4}$$

The conjugate-gradient method can be used to compute the inversion in equation A-4 iteratively. As shown by (11), the iterative convergence for inversion in equation A-4 can be dramatically faster that the one in equation A-1.

The main advantage of shaping regularization is the relative ease of controlling the selection of λ and $S\mathbf{S}$ in comparison with ε and $\mathbf{D}$ cite[]Fomel2009. In this paper, we choose λ to be the median value of $\mathbf{\Psi_k(t)}$ and the shaping operator to be a Gaussian smoothing operator. The Gaussian smoothing operator is a convolution operator with the Gaussian function that is used to smooth images and remove detail and noise. In this sense it is similar to the mean filter, but it uses a different kernel to represent the shape of a Gaussian (bell-shaped) hump. The Gaussian function in 1D has the form,

$$G(x) = \frac{1}{\sqrt{2\pi\sigma}}\exp\left(-\frac{x^2}{2\sigma^2}\right). \tag{A-5}$$

One can implement Gaussian smoothing by Gaussian filtering in either the frequency domain or time domain. (10) shows that repeated application of triangle smoothing can also be used to implement Gaussian smoothing efficiently, in which case the only additional parameter is the radius of the triangle smoothing operator. In this paper, we use the repeated triangle smoothing operator to implement the Gaussian smoothing operator.

The computational cost of generating a time-frequency representation with our method is $O\left(N_t N_f N_{iter}/N_p\right)$, where N_t is the number of time samples, N_f is the number of frequencies, N_{iter} is the number of conjugate-gradient iterations, N_p is the number of processors used to process different frequencies in parallel, and N_{iter} decreases with the increase of smoothing and is typically around 10.

REFERENCES

[1] Allen, J. G., 1977, Short term spectral analysis, synthetic and modification by discrete Fourier transform: IEEE Transactions on Signal Processing, 135–238.

[2] Brian, C., R. C. Williamson, and B. Boashash, 1993, The relationship between instantaneous frequency and time-frequency representations: IEEE Transactions on Acoustics, Speech, and Signal Processing, **41**, 1458–1461.

[3] Castagna, J., S. Sun, and R. W. Siegfried, 2003, Instantaneous spectral analysis: Detection of low-frequency shadows associated with hydrocarbons: The Leading Edge, **22**, 120–127.

[4] Chakraborty, A., and D. Okaya, 1995, Frequency-time decomposition of seismic data using wavelet-based methods: Geophysics, **60**, 1906–1916.

[5] Claasen, T., and W. Mecklenbruker, 1980, The Wigner distribution – A tool for time-frequency signal analysis, 3 parts: Philips Journal of Research.

[6] Cohen, L., 1989, Time-frequency distributions – A review: Proceedings of the IEEE, **77**, 941–981.

[7] ———, 1995, Time-frequency analysis: Prentice Hall, Inc.

[8] Ebrom, D., 2004, The low frequency gas shadows in seismic sections: The Leading Edge, **23**, 772.

[9] Fomel, S., 2007a, Local seismic attributes: Geophysics, **72**, A29–A33.

[10] ———, 2007b, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.

[11] ———, 2009, Adaptive multiple subtraction using regularized nonstationary regression: Geophysics, **74**, V25–V33.

[12] ———, 2010, Predictive painting of 3D seismic volumes: Geophysics, **75**, A25–A30.

[13] Fomel, S., and L. Jin, 2009, Time-lapse image registration using the local similarity attribute: Geophysics, **74**, A7–A11.

[14] Fomel, S., and van der Baan, 2009, Local similarity with the envelope as a seismic phase detector: 80th Annual International Meeting, Soc. of Expl. Geophys., 1555–2559.

[15] Hlawatsch, F., and G. Boudreaux-Bartels, 1992, Linear and quadratic timefrequency signal representations: IEEE Signal Processing Magazine, **9**, 21–67.
[16] Kazemeini, S. H., C. Juhlin, K. Z. Jorgensen, and B. Norden, 2009, Application of the continuous wavelet transform on seismic data for mapping of channel deposits and gas detection at the CO2SINK site, Ketzin, Germany: Geophysical Prospecting, **57**, 111–123.
[17] Korneev, V. A., G. M. Goloshubin, T. M. Daley, and B. S. Dmitry, 2004, Seismic low-frequency effects in monitoring fluid-saturated reservoirs: Geophysics, **69**, 522–532.
[18] Li, Y., and X. Zheng, 2008, Spectral decomposition using Wigner-Ville distribution with applications to carbonate reservoir characterization: The Leading Edge, **27**, 1050–1057.
[19] Liu, G., X. Chen, J. Li, J. Du, and J. Song, 2011a, Seismic noise attenuation using nonstationary polynomial fitting: Applied Geophysics, **8**, 18–26.
[20] Liu, G., S. Fomel, and X. Chen, 2011b, Stacking angle-domain common-image gathers for normalization of illumination: Geophysical Prospecting, **59**, 244–255.
[21] Liu, G., S. Fomel, L. Jin, and X. Chen, 2009, Seismic data stacking using local correction: Geophysics, **74**, V43–V48.
[22] Liu, J., and K. J. Marfurt, 2005, Matching pursuit decomposition using Morlet wavelet: 75th Annual International Meeting, Soc. of Expl. Geophys., 786–789.
[23] Liu, J., Y. Wu, D. Han, and X. Li, 2004, Time-frequency decomposition based on Ricker wavelet: 74th Annual International Meeting, Soc. of Expl. Geophys., 1937–1940.
[24] Liu, Y., and S. Fomel, 2010, Seismic data analysis using local time-frequency transform: 80th Annual International Meeting, Soc. of Expl. Geophys., 3711–3716.
[25] Lomask, J., A. Guitton, S. Fomel, J. Claerbout, and A. Valenciano, 2006, Flattening without picking: Geophysics, **71**, 13–20.
[26] Mallat, S., and Z. Zhang, 1993, Matching pursuit with time-frequency dictionaries: IEEE Transactions on Signal Processing, **41**, 3397–3415.
[27] Margrave, G., 1998, Theory of nonstationary linear filtering in the Fourier domain with application to time-variant filtering: Geophysics, **63**, 244–259.
[28] Partyka, G. A., J. M. Gridley, and J. Lopez, 1998, Interpretational applications of spectral decomposition in reservoir characterization: The Leading Edge, **18**, 353–360.
[29] Pinnegar, C., and L. Mansinha, 2003, The S-transform with windows of arbitrary and varying shape: Geophysics, **68**, 381–385.
[30] Reine, C., M. van der Baan, and R. Clark, 2009, The robustness of seismic attenuation measurements using fixed- and variable-window time-frequency transforms: Geophysics, **74**, 123–135.
[31] Rioul, O., and M. Vetterli, 1991, Wavelets and signal processing: IEEE Signal Processing Magazine, **8**, 14–38.
[32] Saha, J., 1987, Relationship between Fourier and instantaneous frequency: 57th Annual International Meeting, Soc. of Expl. Geophys., 591.
[33] Sinha, S., P. Routh, and P. Anno, 2009, Instantaneous spectral attributes using scales in continuous-wavelet transform: Geophysics, **74**, WA137–WA142.

[34] Sinha, S., P. S. Routh, P. D. Anno, and J. P. Castagna, 2005, Spectral decomposition of seismic data with continuous-wavelet transform: Geophysics, **70**, P19–P25.
[35] Steeghs, P., and G. Drijkoningen, 2001, Seismic sequence analysis and attribute extraction using quadratic time-frequency representations: Geophysics, **66**, 1947–1959.
[36] Stockwell, R. G., L. Mansinha, and R. P. Lowe, 1996, Localization of the complex spectrum: IEEE Transactions on Signal Processing, **44**, 998–1001.
[37] Tikhonov, A. N., 1963, Solution of incorrectly formulated problems and the regularization method: IEEE Transactions on Signal Processing.
[38] van der Baan, M., and S. Fomel, 2009, Nonstationary phase estimation using regularized local kurtosis maximization: Geophysics, **74**, A75–A80.
[39] Wang, Y., 2007, Seismic time-frequency spectral decomposition by matching pursuit: Geophysics, **72**, V13–V20.
[40] Wigner, E., 1932, On the quantum correction for thermodynamic equilibrium: Physical Review, **40**, 749–759.
[41] Wu, X., and T. Liu, 2009, Spectral decomposition of seismic data with reassigned smoothed pseudo Wigner-Ville distribution: Journal of Applied Geophysics, **68**, 386–393.
[42] Youn, D. H., and J. G. Kim, 1985, Short-time Fourier transform using a bank of low-pass filters: IEEE Transactions on Acoustics, Speech, and Signal Processing, **33**, 182–185.
[43] Zeng, H., M. M. Backus, K. T. Barrow, and N. Tyler, 1998, Stratal slicing, Part I: Realistic 3D seismic model: Geophysics, **63**, 502–513.
[44] Zhenhua, H., X. Xiaojun, and B. Lien, 2008, Numerical simulation of seismic low-frequency shadows and its application: Applied Geophysics, **5**, 301–306.

Velocity analysis using AB semblance

*Sergey Fomel**

ABSTRACT

I derive and analyze an explicit formula for a generalized semblance attribute, which is suitable for velocity analysis of prestack seismic gathers with distinct amplitude trends. While the conventional semblance can be interpreted as squared correlation with a constant, the AB semblance is defined as a correlation with a trend. This measure is particularly attractive for analyzing class II AVO anomalies and converted waves. Analytical derivations and numerical experiments show that the resolution of the AB semblance is approximately twice lower than that of the conventional semblance. However, this does not prevent it from being an effective attribute. I use synthetic and field data examples to demonstrate the improvements in velocity analysis from AB semblance.

INTRODUCTION

Since its introduction by (13), the semblance measure has been an indispensable tool for velocity analysis of seismic records. Conventional velocity analysis of seismic gathers scans different values of effective moveout velocity, computes semblance of flattened gathers and generates velocity spectra for later velocity picking (14) .

While effective in most practical situations, semblance becomes troublesome in the case of strong variation of amplitudes along seismic events (10). A particular example is class II AVO anomalies (8) that cause seismic amplitudes to go through a polarity reversal. To address this problem, (7) and (10; 9) developed algorithms for correcting the semblance measurement for amplitude variations.

In this paper, I interpret the semblance attribute as a correlation with a constant and derive an explicit mathematical expression for the measure which corresponds to correlation with an amplitude trend. This measure is equivalent to AB *semblance* proposed by (10; 9). It reduces, in the case of constant amplitudes, to the conventional semblance. I analyze the statistics of the AB semblance attribute and quantify the loss of resolution associated with it. Numerical experiments with synthetic and field data demonstrate the effectiveness of the AB semblance as a robust velocity analysis attribute, which is applicable even in the presence of strong amplitude variations and polarity reversals. Moreover, the ratio of the AB and conventional semblances serves as a useful AVO indicator attribute.

*e-mail: sergey.fomel@beg.utexas.edu

THEORY

I start by interpreting the meaning of the conventional semblance attribute as a correlation with a constant. Next, I define AB semblance as a correlation with a trend and analyze its statistical properties.

Semblance as correlation

The correlation coefficient γ between two sequences of numbers $\mathbf{a} = a_1, a_2, \ldots, a_N$ and $\mathbf{b} = b_1, b_2, \ldots, b_N$ is defined as

$$\gamma(\mathbf{a},\mathbf{b}) = \frac{\mathbf{a}\cdot\mathbf{b}}{|\mathbf{a}|\,|\mathbf{b}|} = \frac{\sum_{i=1}^{N} a_i\, b_i}{\sqrt{\sum_{i=1}^{N} a_i^2}\sqrt{\sum_{i=1}^{N} b_i^2}} \tag{1}$$

The correlation coefficient is analogous to the cosine of the angle between two vectors $\mathbf{a}$ and $\mathbf{b}$. It takes values in the range from -1 to 1. Taking a correlation of a sequence $\mathbf{a}$ with a constant sequence $\mathbf{c} = C, C, \ldots, C$ produces a measure β, defined as

$$\beta(\mathbf{a}) = \gamma(\mathbf{a},\mathbf{c}) = \frac{\sum_{i=1}^{N} a_i\, C}{\sqrt{\sum_{i=1}^{N} a_i^2}\sqrt{\sum_{i=1}^{N} C^2}} = \frac{\sum_{i=1}^{N} a_i}{\sqrt{N\sum_{i=1}^{N} a_i^2}} \tag{2}$$

Squaring the correlation with a constant yields the measure equivalent to semblance

$$\beta^2(\mathbf{a}) = \frac{\left(\sum_{i=1}^{N} a_i\right)^2}{N\sum_{i=1}^{N} a_i^2}\ . \tag{3}$$

Semblance is maximized when the sequence $\mathbf{a}$ has a uniform distribution. When seismic amplitude is uniformly distributed along a moveout curve, the semblance of a horizontal slice through the gather will be maximized when the event is flattened. This fact is the basis of the conventional velocity analysis originally developed by (13). The approach fails, however, when the amplitude variation is distinctly non-uniform.

AB semblance: correlation with a trend

Suppose that the reference sequence has a trend $b_i = A + B\,\phi_i$, where ϕ_i is a known function. The trend can be, for example, an expression of the PP reflection coefficient

in Shuey's approximation (11), where A and B are the AVO intercept and gradient, $\phi_i = \sin^2\theta_i$, and θ_i corresponds to the reflection angle at trace i. In examples of this paper, I use offset instead of angle. Relating offset and reflection angle can be done either by using approximate equations of by ray tracing once the velocity model is established.

Estimating A and B from least-square fitting of the trend amounts to the minimization of

$$F(A,B) = \sum_{i=1}^{N} (a_i - A - B\,\phi_i)^2 \tag{4}$$

Differentiating equation (4) with respect to A and B, setting the derivatives to zero, and solving the system of two linear equations produces the well-known linear fit equations

$$A = \frac{\displaystyle\sum_{i=1}^{N}\phi_i \sum_{i=1}^{N} a_i\,\phi_i - \sum_{i=1}^{N}\phi_i^2 \sum_{i=1}^{N} a_i}{\displaystyle\left(\sum_{i=1}^{N}\phi_i\right)^2 - N\sum_{i=1}^{N}\phi_i^2}\,, \tag{5}$$

$$B = \frac{\displaystyle\sum_{i=1}^{N}\phi_i \sum_{i=1}^{N} a_i - N\sum_{i=1}^{N} a_i\,\phi_i}{\displaystyle\left(\sum_{i=1}^{N}\phi_i\right)^2 - N\sum_{i=1}^{N}\phi_i^2}\,. \tag{6}$$

Substituting the trend $b_i = A + B\,\phi_i$ with A and B defined from the least-squares equations (5) and (6) into the correlation coefficient equation (A-8) and squaring the result leads to equation

$$\alpha^2(\mathbf{a}) = \frac{2\displaystyle\sum_{i=1}^{N} a_i \sum_{i=1}^{N}\phi_i \sum_{i=1}^{N} a_i\,\phi_i - \left(\sum_{i=1}^{N} a_i\right)^2 \sum_{i=1}^{N}\phi_i^2 - N\left(\sum_{i=1}^{N} a_i\,\phi_i\right)^2}{\displaystyle\sum_{i=1}^{N} a_i^2 \left[\left(\sum_{i=1}^{N}\phi_i\right)^2 - N\sum_{i=1}^{N}\phi_i^2\right]}\,. \tag{7}$$

Equation (7) generalizes the semblance measure β defined in equation (3) to a new measure α. In the absence of a trend (when the numerator in equation (6) is zero), α is equivalent to β.

(10) defined semblance using a normalized least-squares objective

$$\alpha^2(\mathbf{a}) = 1 - \frac{F(A,B)}{\displaystyle\sum_{i=1}^{N} a_i^2}\,. \tag{8}$$

Substituting equations (5) and (6) into (8) is an alternative way of deriving equation (7). This is the *AB semblance* in terminology of (10; 9).

Sensitivity analysis of AB semblance

(10; 9) observed a decrease of resolution in AB semblance in comparison with the conventional semblance when applying it on synthetic examples. The explicit equation (7) allows us to access the resolution limits of both methods by measuring their effect on random noise.

As shown in Appendix A, the semblance of a sequence of uncorrelated normally-distributed zero-mean noise samples has the expectation value of $1/N$, where N is the number of samples. The corresponding AB semblance has the expectation $2/N$ or twice higher. Moreover, while the standard deviation of the conventional semblance decreases as $1/N$, the deviation of the AB semblance decreases as $2/N$. This analysis shows that the AB measurement is generally twice as sensitive to noise and has a twice lower resolution. This is the price one has to pay for the ability to handle amplitude trends.

EXAMPLES

In this section, I demonstrate the behavior of AB semblance with a synthetic and a field data example.

Synthetic example

Figure 1 shows two synthetic CMP gathers generated by applying inverse normal moveout with a variable moveout velocity and adding a modest amount of random noise. The first gather contains no amplitude variations, while the second gather contains a region of polarity reversal. Figure 2 shows velocity analysis panels using conventional semblance scans and the corresponding automatic velocity picks. The picking algorithm is explained in Appendix B. We can observe that the AVO anomaly causes a signal loss in the semblance measure, which in turn leads to inaccurate velocity picking. The AB semblance, on the other hand, is not affected much by the amplitude variations and allows for accurate velocity picking in both cases, as shown in Figure 3.

Figures 4 and 5 compare NMO-corrected gathers using velocities picked from the conventional and AB semblance scans respectively. While Figures 4a and 5a are virtually identical, the residual curvature evident in the anomalous region in Figure 4b disappears in Figure 5b, which clearly demonstrates the advantage of the AB approach. Figure 6 shows NMO-corrected gathers colored according to the AVO-

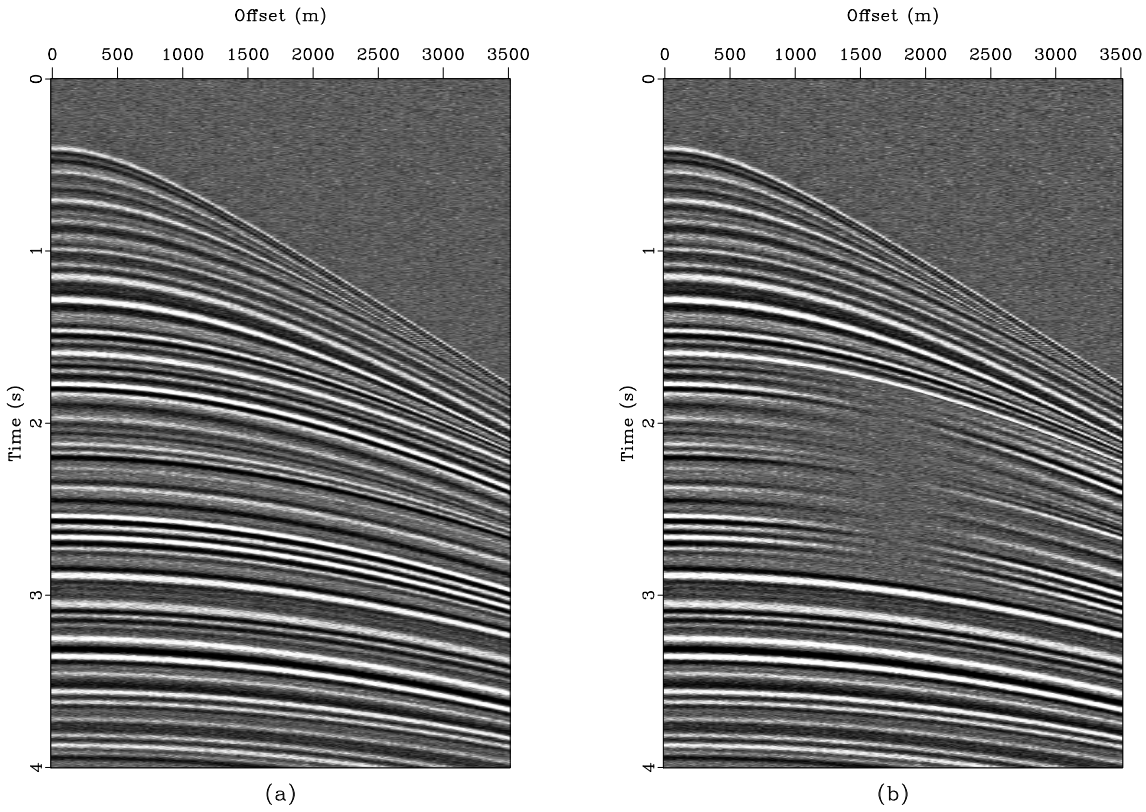


Figure 1: Synthetic CMP gathers. a: no AVO, b: AVO trend with polarity reversal in the middle section. avo/avo cmp

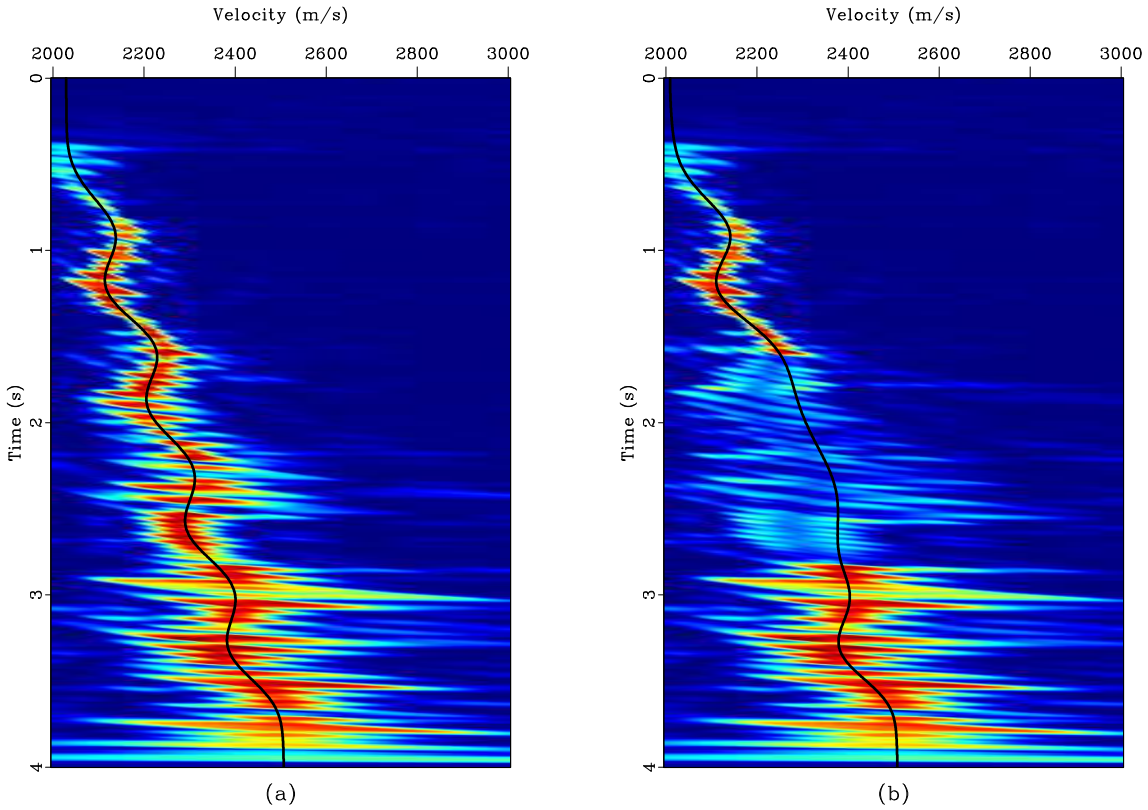


Figure 2: Conventional semblance scans for CMP gathers from Figure 1. Black curves indicate automatic velocity picks. The loss of signal in the right plot is caused by the AVO anomaly. a: no AVO, b: AVO trend with polarity reversal in the middle section. avo/avo scn

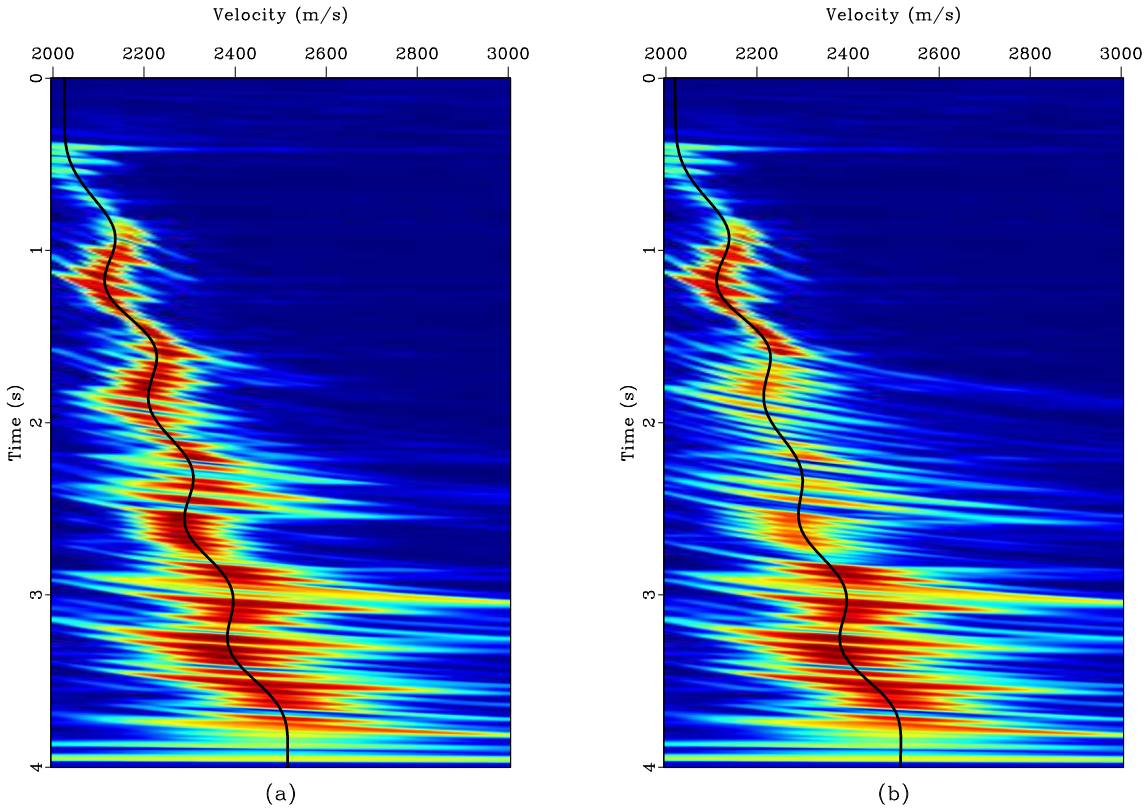


Figure 3: *AB* semblance scans for CMP gathers from Figure 1. Black curves indicate automatic velocity picks. Semblance remains strong despite the AVO anomaly. a: no AVO, b: AVO trend with polarity reversal in the middle section. Compare with Figure 2. avo/avo avoscn

indicator attribute, which I define as a ratio between the conventional and the AB semblances. The anomalous region is clearly visible in the attribute display.

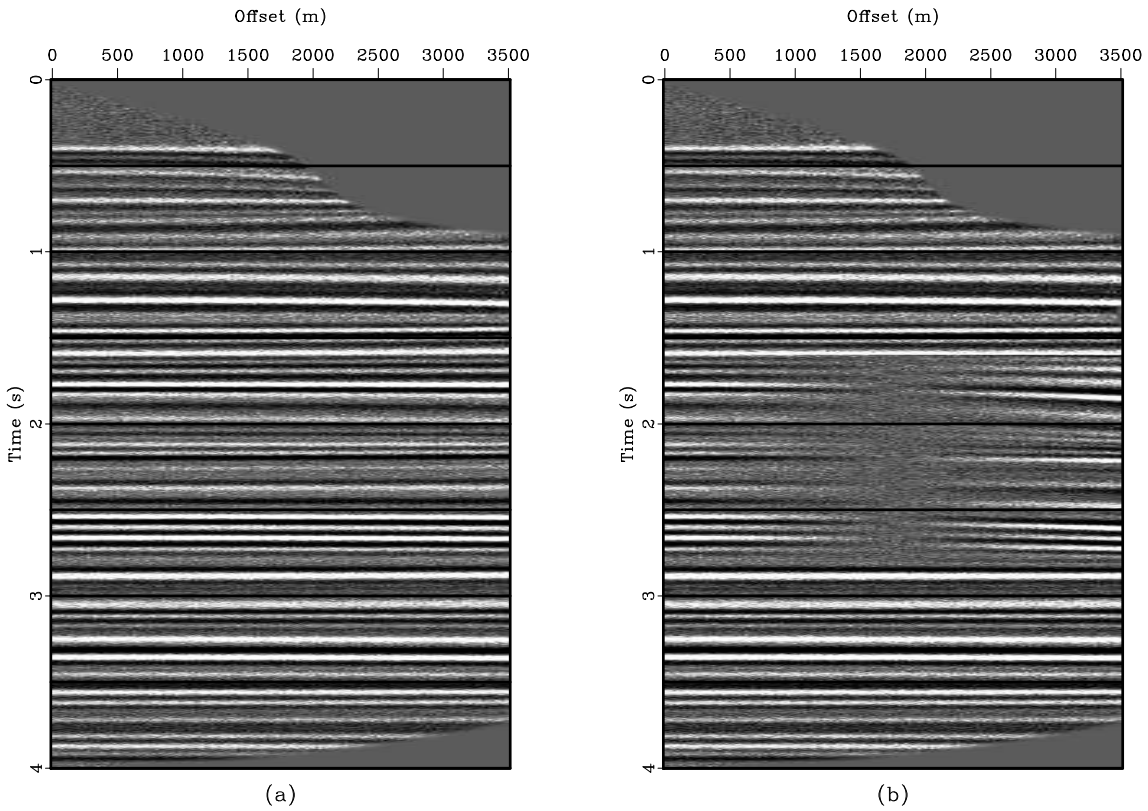


Figure 4: NMO-corrected CMP gathers from Figure 1 using velocity functions picked from conventional semblance in Figure 2. avo/avo nmo

Field data example

For a field data example, I select a gather already processed by a seismic contractor. The gather, shown in Figure 7, exhibits a clear polarity reversal around 3.8 s. The polarity reversal is the apparent cause of a visible residual moveout artifact. In order to correct the residual curvature, I apply semblance-based analysis. The comparison between the conventional and the AB semblance is shown in Figure 8. Similarly to the synthetic example, the AB semblance provides a better indicator of the residual velocity for the curved event with anomalous amplitude. Figure 9 shows NMO-corrected gather using a velocity trend picked automatically from the AB analysis. The curved reflection event is successfully flattened.

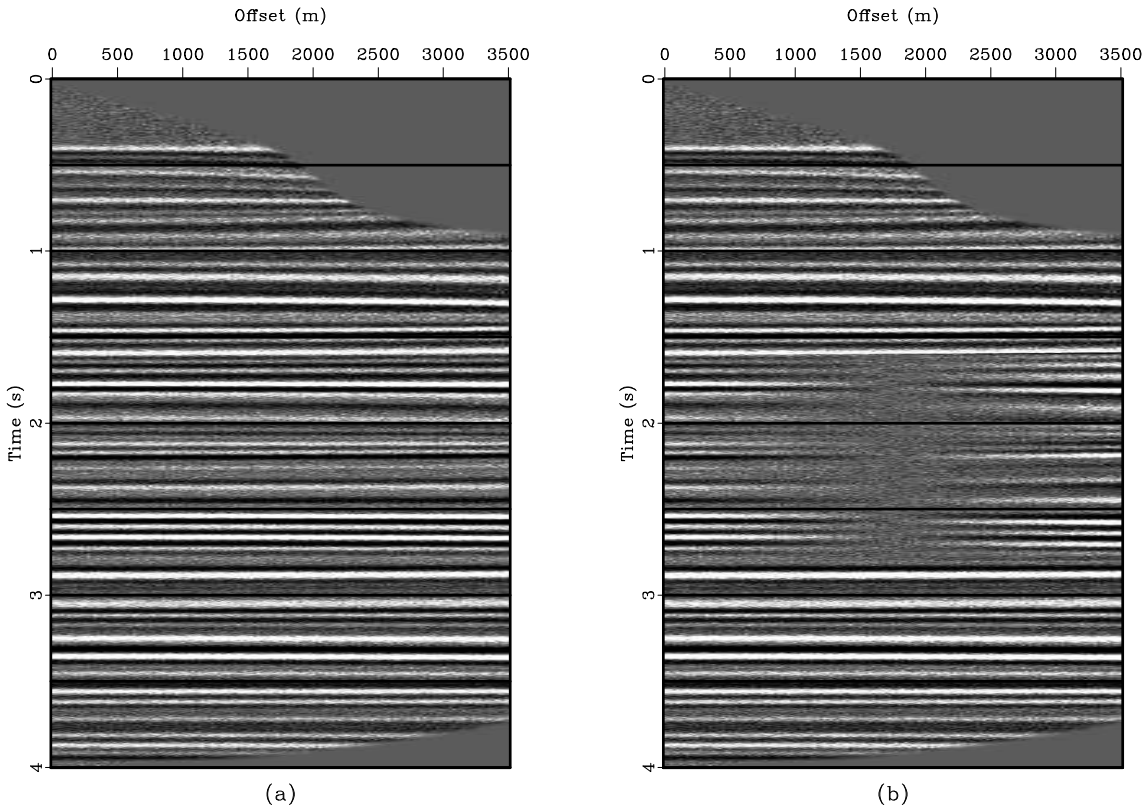


Figure 5: NMO-corrected CMP gathers from Figure 1 using velocity functions picked from *AB* semblance in Figure 3. avo/avo avonmo

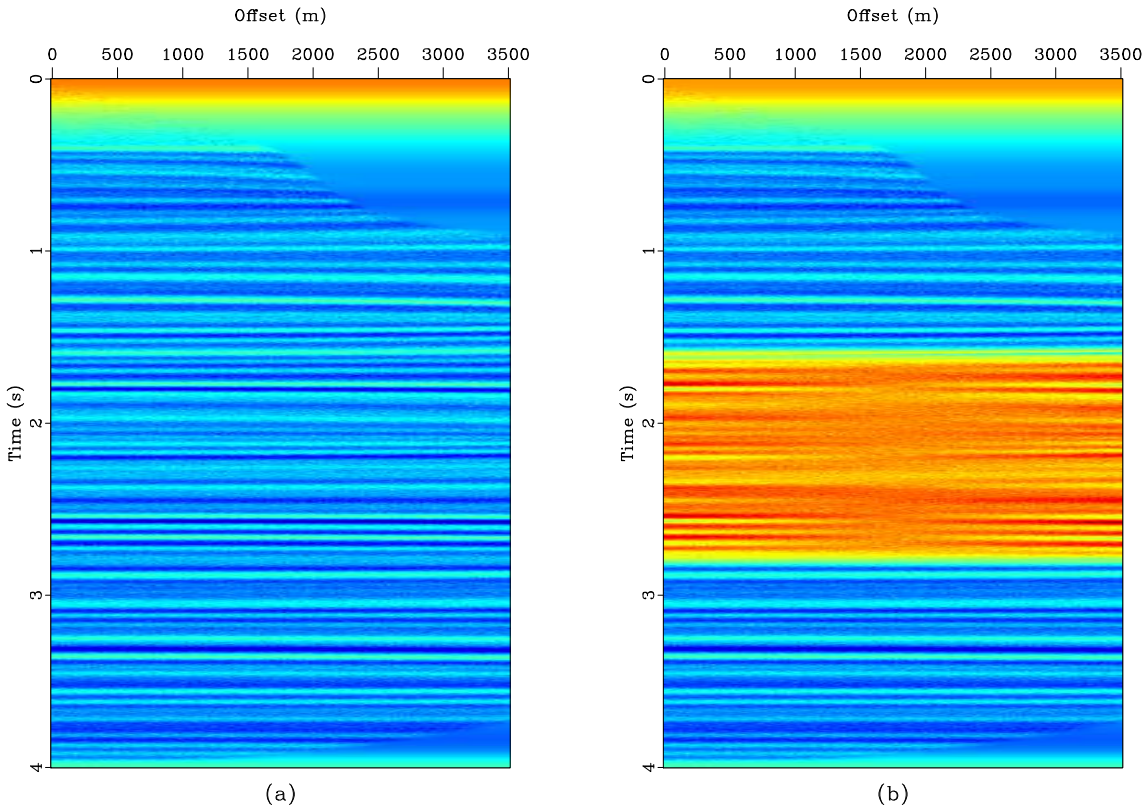


Figure 6: NMO-corrected CMP gathers from Figure 5 colored according to the AVO indication attribute. avo/avo attr

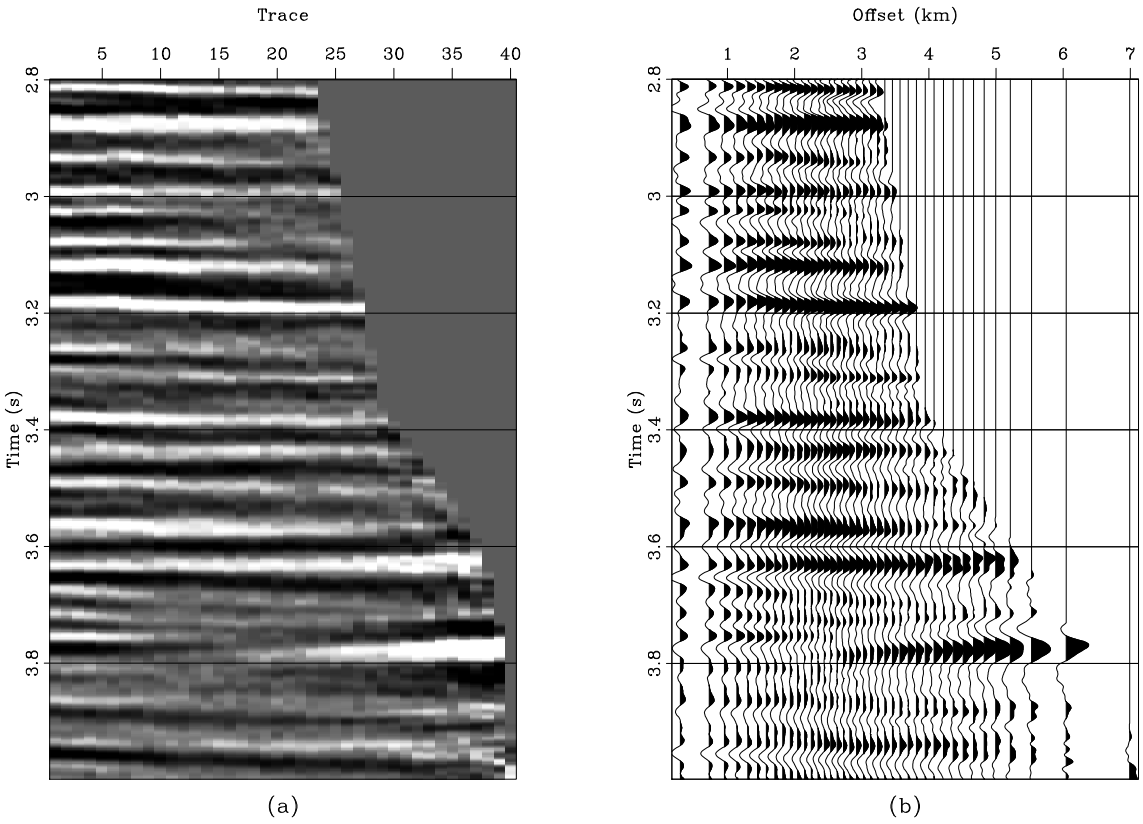


Figure 7: Input CMP gather after preprocessing. a: trace display, b: wiggle display. avo/avo2 gath4

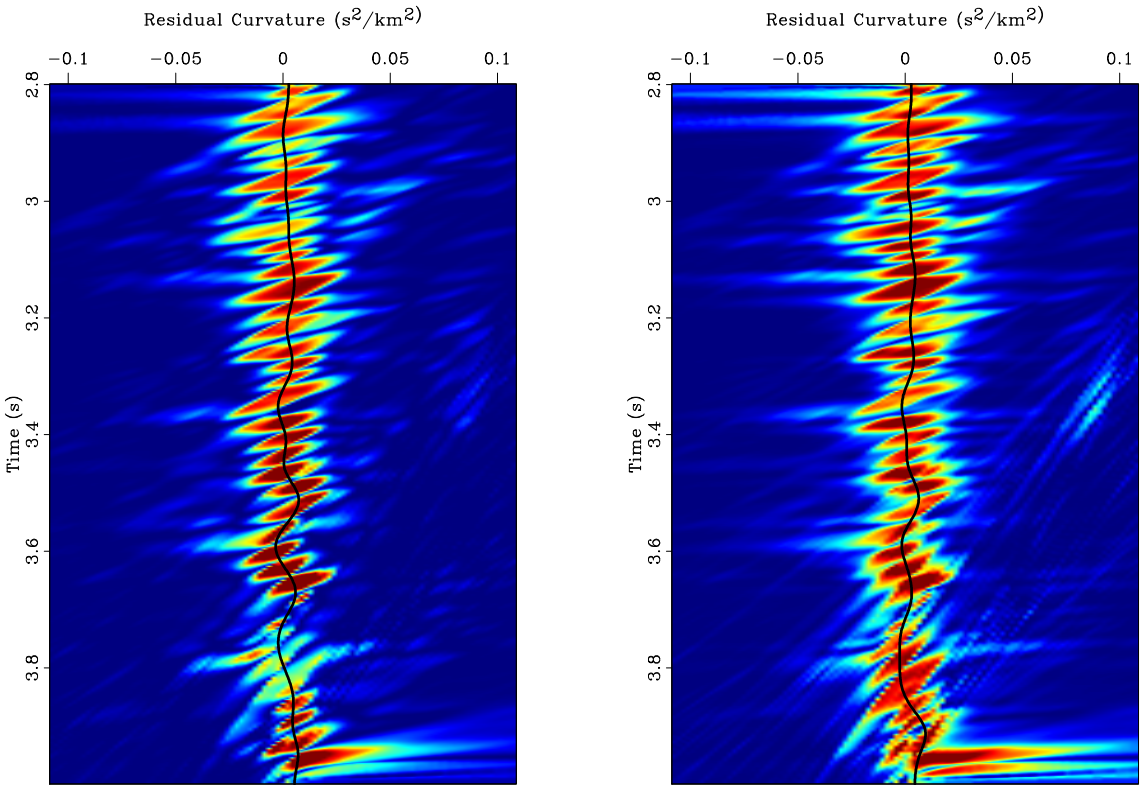


Figure 8: Residual moveout curvature scans using conventional semblance (a) and AB semblance (b). Black curves indicate automatically picked trends. avo/avo2 scan4

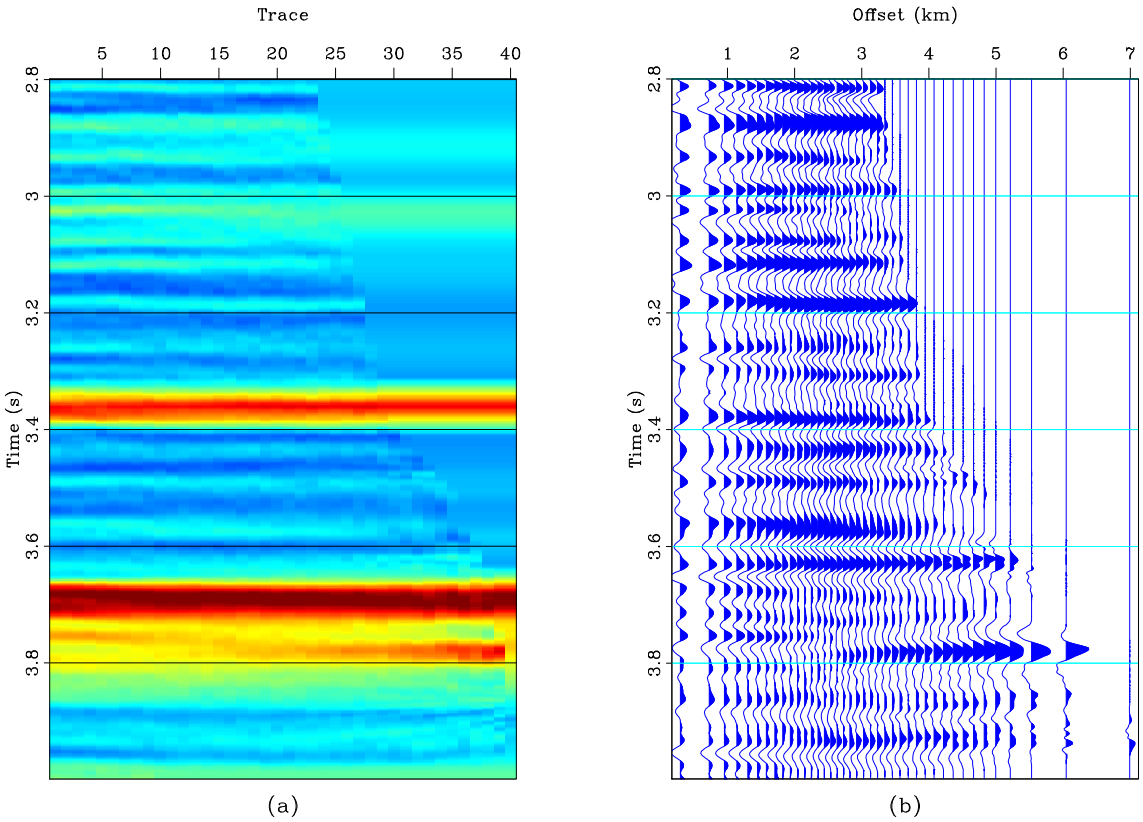


Figure 9: CMP gather from Figure 7 after residual normal moveout correction using trends picked from the *AB* semblance in Figure 8. a: trace display colored according to the AVO indicator attribute, b: wiggle display. avo/avo2 nmo4

CONCLUSIONS

If one interprets the conventional semblance as a correlation with a constant, the AB semblance is a correlation with an amplitude trend. I have derived an explicit expression for the AB semblance and analyzed it to quantify the observable loss of resolution. I have demonstrated the advantages of the AB semblance attribute using synthetic and field data examples. The ratio of the AB and conventional semblances serves as a useful AVO indicator attribute. The sensitivity of this attribute to non-linear variations in the amplitude trend can be a subject of further research. Further applications may also include velocity analysis of converted waves and seismic diffractions.

ACKNOWLEDGMENTS

I would like to thank Tury Taner and Richard Uden for inspiring discussions and for providing the field data example.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APENDIX A: STATISTICAL ANALYSIS OF SEMBLANCE MEASURES

In this appendix, I study the influence of noise on semblance measures. Let us assume that the signal $\mathbf{a} = a_1, a_2, \ldots, a_N$ is composed of random independent samples normally distributed with zero mean and σ^2 variance. In this case, the mathematical expectation for the semblance measure (3) is

$$E\left[\beta^2(\mathbf{a})\right] = \frac{E\left[\left(\sum_{i=1}^{N} a_i\right)^2\right]}{E\left[N\sum_{i=1}^{N} a_i^2\right]} = \frac{\sum_{i=1}^{N} E\left[a_i^2\right]}{N\sum_{i=1}^{N} E\left[a_i^2\right]} = \frac{1}{N}\,. \tag{A-1}$$

Correspondingly, the variance of the noise semblance is

$$\begin{aligned} V\left[\beta^2(\mathbf{a})\right] &= \frac{E\left[\left(\sum_{i=1}^{N} a_i\right)^4\right]}{E\left[\left(N\sum_{i=1}^{N} a_i^2\right)^2\right]} - \frac{1}{N^2} = \frac{\sum_{i=1}^{N} E\left[a_i^4\right] + 3\sum_{i=1}^{N}\sum_{j\neq i} E\left[a_i^2\,a_j^2\right]}{N^2\left(\sum_{i=1}^{N} E\left[a_i^4\right] + \sum_{i=1}^{N}\sum_{j\neq i} E\left[a_i^2\,a_j^2\right]\right)} - \frac{1}{N^2} \\ &= \frac{3\,N\,\sigma^4 + 3\left(N^2-N\right)\sigma^4}{N^2\left[3\,N\,\sigma^4 + \left(N^2-N\right)\sigma^4\right]} - \frac{1}{N^2} = \frac{2\left(N-1\right)}{N^2\left(N+2\right)}\,. \end{aligned} \tag{A-2}$$

Equations (A-1) and (A-2) show that both the mathematical expectation and the standard deviation (the square root of variance) of the random noise semblance decrease at the rate of $1/N$ with the increase in the number of traces. To derive these equations, I make an assumption that the terms in the numerator and denominator are statistically independent. Rather than proving this assumption mathematically, I test it by numerical experiments with multiple random number realizations. Figure A-1 compares the theoretical prediction with experimental measurements from 10,000 random realizations.

Applying similar analysis to the AB semblance (7), we deduce that

$$\begin{aligned} E\left[\alpha^2(\mathbf{a})\right] &= \frac{E\left[2\sum_{i=1}^{N} a_i \sum_{i=1}^{N} \phi_i \sum_{i=1}^{N} a_i\,\phi_i - \left(\sum_{i=1}^{N} a_i\right)^2 \sum_{i=1}^{N} \phi_i^2 - N\left(\sum_{i=1}^{N} a_i\,\phi_i\right)^2\right]}{E\left[\sum_{i=1}^{N} a_i^2\right]\left[\left(\sum_{i=1}^{N} \phi_i\right)^2 - N\sum_{i=1}^{N} \phi_i^2\right]} \\ &= \frac{2\,E\left[a_i^2\right]\left[\left(\sum_{i=1}^{N} \phi_i\right)^2 - N\sum_{i=1}^{N} \phi_i^2\right]}{N\;E\left[a_i^2\right]\left[\left(\sum_{i=1}^{N} \phi_i\right)^2 - N\sum_{i=1}^{N} \phi_i^2\right]} = \frac{2}{N}\,. \end{aligned} \tag{A-3}$$

and

$$\begin{aligned} V\left[\alpha^2(\mathbf{a})\right] &= \frac{E\left[\left\{2\sum_{i=1}^{N} a_i \sum_{i=1}^{N} \phi_i \sum_{i=1}^{N} a_i\,\phi_i - \left(\sum_{i=1}^{N} a_i\right)^2 \sum_{i=1}^{N} \phi_i^2 - N\left(\sum_{i=1}^{N} a_i\,\phi_i\right)^2\right\}^2\right]}{E\left[\left(\sum_{i=1}^{N} a_i^2\right)^2\right]\left[\left(\sum_{i=1}^{N} \phi_i\right)^2 - N\sum_{i=1}^{N} \phi_i^2\right]^2} - \frac{4}{N^2} \\ &= \frac{2\left[2\left(\sum_{i=1}^{N} \phi_i\right)^2 + N\sum_{i=1}^{N} \phi_i^2\right]^2 + 6\,N^2\left(\sum_{i=1}^{N} \phi_i^2\right)^2 - 24\,N\left(\sum_{i=1}^{N} \phi_i\right)^2 \sum_{i=1}^{N} \phi_i^2}{\left[3\,N + (N^2 - N)\right]\left[\left(\sum_{i=1}^{N} \phi_i\right)^2 - N\sum_{i=1}^{N} \phi_i^2\right]^2} - \frac{4}{N^2} \\ &= \frac{4\,(N-2)}{N^2\,(N+2)}\,. \end{aligned} \tag{A-4}$$

One can see that, in the case of the AB semblance, the mathematical expectation and the standard deviation of the random noise semblance decrease at the rate of $2/N$, twice higher than that for the conventional semblance. Figure A-2 compares the theoretical prediction with experimental measurements.

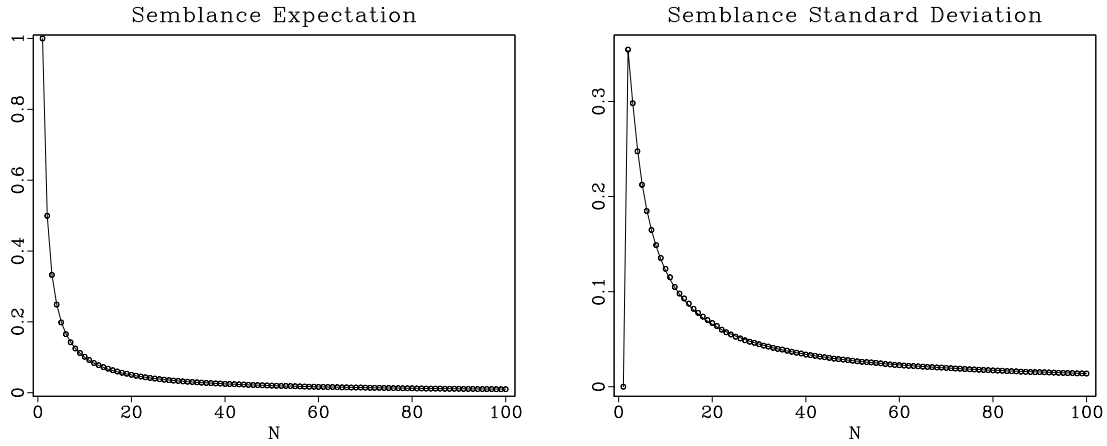


Figure A-1: Mathematical expectation (a) and standard deviation (b) of random-noise semblance as functions of the number of traces N. Solid lines are theoretical curves, circles are measurements from a numerical experiment. avo/stat mean,vari

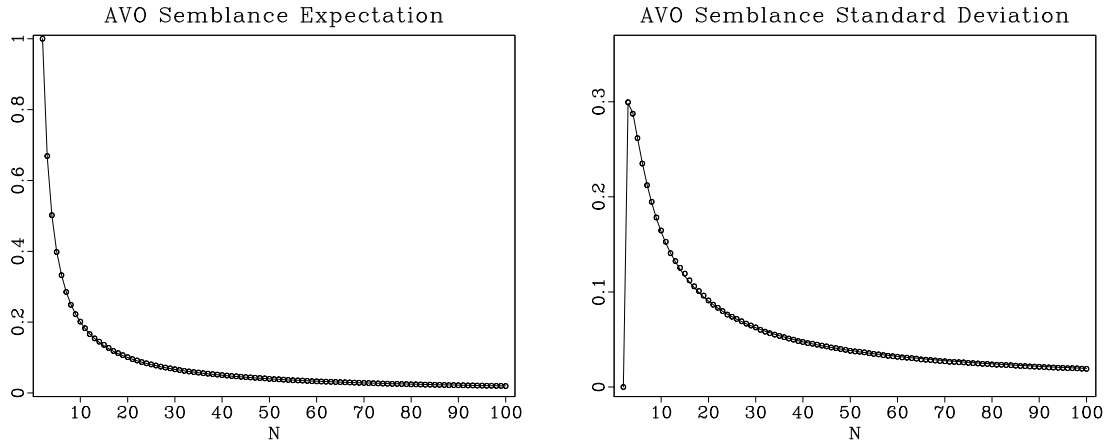


Figure A-2: Mathematical expectation (a) and standard deviation (b) of random-noise AB semblance as a function of the number of traces N. Solid lines are theoretical curves, circles are measurements from a numerical experiment. avo/stat amean,avari

APPENDIX B: AUTOMATIC VELOCITY PICKING FROM SEMBLANCE SCANS

The problem of automatic picking of velocities from semblance scans has been considered by many authors (1; 12; 2). The approach taken in this paper is inspired by the suggestion of (5) to look at velocity picking as a variational problem. According to (5), an optimally picked velocity trend $v(t)$ in the semblance field $\alpha(t, v)$ corresponds to the maximum of the variational integral

$$P_1[v(t)] = \int_{t_{min}}^{t_{max}} \alpha\left(t, v(t)\right)\, dt\ . \tag{B-1}$$

I take the variational formulation further by considering its analogy to the ray tracing problem. The first-arrival seismic ray is a trajectory corresponding to the minimum traveltime. The trajectory corresponding to an optimal velocity trend should minimize an analogous measure defined in the space of the velocity scan $\{t, v\}$. I use the variational measure

$$P_2[v(t)] = \int_{t_{min}}^{t_{max}} \exp[-\alpha\left(t, v(t)\right)]\, \sqrt{\lambda^2 + \left[v'(t)\right]^2}\, dt\ . \tag{B-2}$$

where λ is a scaling parameter. According to variational theory (6), an optimal trajectory can be determined by solving the eikonal equation

$$\left(\frac{\partial T}{\partial v}\right)^2 + \frac{1}{\lambda^2}\, \left(\frac{\partial T}{\partial t}\right)^2 = \exp[-2\,\alpha\left(t, v\right)] \tag{B-3}$$

with a finite-difference algorithm. The quantity in the right hand side of equation (B-3) plays the role of squared slowness. Small slowness corresponds to high semblance and attracts ray trajectories in a "wave guide". After obtaining a finite-difference solution, the picking trajectory can be extracted by tracking backward along the traveltime gradient direction. An analogous approach has been used in medical imaging in the method of virtual endoscopy (3). To remove random oscillations, I smooth the picked trajectory using the method of shaping regularization (4).

REFERENCES

[1] Adler, F., and S. Brandwood, 1999, Robust estimation of dense 3-D stacking velocities from automated picking: 69th Ann. Internat. Mtg, Soc. of Expl. Geophys., 1162–1165.

[2] Arnaud, J., D. Rappin, J.-P. Dunand, and V. Curinier, 2004, High density picking for accurate velocity and anisotropy determination: 74th Ann. Internat. Mtg., Soc. of Expl. Geophys., 1627–1629.

[3] Deschamps, T., and L. D. Cohen, 2001, Fast extraction of minimal paths in 3D images and applications to virtual endoscopy: Medical Image Analysis, **5**, 281–299.
[4] Fomel, S., 2007, Shaping regularization in geophysical-estimation problems: Geophysics, **72**, R29–R36.
[5] Harlan, W. S., 2001, Constrained automatic moveout picking from semblances: http://billharlan.com/pub/papers/autopick.pdf.
[6] Lanczos, C., 1966, The variational principles of mechanics: University of Toronto Press.
[7] Ratcliffe, A., and F. Adler, 2000, Accurate velocity analysis for Class Ii AvO events: 70th Ann. Internat. Mtg, Soc. of Expl. Geophys., 232–235.
[8] Rutherford, S. R., and R. H. Williams, 1989, Amplitude-versus-offset variations in gas sands: Geophysics, **54**, 680–688.
[9] Sarkar, D., R. T. Baumel, and K. L. Larner, 2002, Velocity analysis in the presence of amplitude variation: Geophysics, **67**, 1664–1672.
[10] Sarkar, D., J. P. Castagna, and W. Lamb, 2001, AVO and velocity analysis: Geophysics, **66**, 1284–1293.
[11] Shuey, R. T., 1985, A simplification of the Zoeppritz-equations: Geophysics, **50**, 609–614. (Errata in GEO-50-9-1522).
[12] Siliqi, R., D. L. Meur, F. Gamar, L. Smith, J. Toure, and P. Herrmann, 2003, High-density moveout parameter fields V and Eta, Part 1: Simultaneous automatic picking: 73rd Ann. Internat. Mtg., Soc. of Expl. Geophys., 2088–2091.
[13] Taner, M. T., and F. Koehler, 1969, Velocity spectra - Digital computer derivation and applications of velocity functions: Geophysics, **34**, 859–881. (Errata in GEO-36-4-0787).
[14] Yilmaz, O., 2000, Seismic data analysis: Soc. of Expl. Geophys.

Non-hyperbolic common reflection surface

*Sergey Fomel[1] and Roman Kazinnik[1][2]**

ABSTRACT

The method of common reflection surface (CRS) extend conventional stacking of seismic traces over offset to multidimensional stacking over offset-midpoint surfaces. We propose a new form of the stacking surface, derived from the analytical solution for reflection traveltime from a hyperbolic reflector. Both analytical comparisons and numerical tests show that the new approximation can be significantly more accurate than the conventional CRS approximation at large offsets or at large midpoint separations while using essentially the same parameters.

INTRODUCTION

Seismic data stacking is (together with deconvolution and migration) one of the fundamental operations in seismic data analysis (19). Conventional stacking operates on common-midpoint (CMP) gathers and stacks traces after a hyperbolic moveout. The method of multifocusing (MF), originally developed by (4; 5) and modified to the common-reflection-surface (CRS) method by (11), stacks data from multiple CMP locations. As a result, the signal-to-noise ratio is improved considerably. Both MF and CRS require estimation of multiple parameters in addition to the conventional stacking velocity. These parameters correspond to the slope and curvature of seismic events in the midpoint direction and have physical interpretation in terms of wavefront slopes and curvatures. Many successful applications of MF and CRS have been reported in the literature (12; 8; 14; 9; 6; 10).

The CRS method employs a multiparameter hyperbolic approximation of the reflection traveltime surface (17). The hyperbolic approximation can be justified from a truncated Taylor series expansion of the squared traveltime around a reference ray. As such, it is always accurate at small deviations from the central ray. However, it loses its accuracy at large offsets or large midpoint separations.

In this paper, we propose a new nonhyperbolic approximation. The form of this approximation follows from an analytical equation for reflection traveltime from a hyperbolic reflector. The idea of approximation reflection traveltimes by approximating reflector surfaces was first proposed by (15) and (13). However, these publications did not provide a closed-form representation of the stacking surface. By analyzing the accuracy of the proposed nonhyperbolic approximation on a number of examples, we

*e-mail: sergey.fomel@beg.utexas.edu

show that the proposed approximation can significantly extend the accuracy range of CRS.

HYPERBOLIC AND NONHYPERBOLIC CRS

If $P(t,m,h)$ represents the prestack seismic data as a function of time t, midpoint m and half-offset h, then conventional stacking can be described as

$$S(t_0,m_0) = \int P\left(\theta(h;t_0),m_0,h\right)\,dh\ , \tag{1}$$

where $S(t,m)$ is the stack section, and $\theta(h;t_0)$ is the moveout approximation, which may take a form of a hyperbola

$$\theta(h;t_0) = \sqrt{t_0^2 + \frac{4\,h^2}{v^2}} \tag{2}$$

with v as an effective velocity parameter or, alternatively, a more complicated non-hyperbolic functional form, which involves other parameters (3).

The MF or CRS stacking takes a different form,

$$\widehat{S}(t_0,m_0) = \iint P\left(\widehat{\theta}(m-m_0,h;t_0),m,h\right)\,dm\,dh\ , \tag{3}$$

where the integral over midpoint m is typically carried out only over a limited neighborhood of m_0. The multifocusing approximation of (4) takes the form

$$\widehat{\theta}_{MF}(d,h;t_0) = t_0 + T_{(+)}(d,h) + T_{(-)}(d,h)\ , \tag{4}$$

where, in the notation of (18),

$$T_{(\pm)} = \frac{\sqrt{1+2\,K_{(\pm)}\,(d\pm h)\,\sin\beta + K_{(\pm)}^2\,(d\pm h)^2} - 1}{V_0\,K_{(\pm)}}\ , \tag{5}$$

$$K_{(\pm)} = \frac{K_N \pm \sigma\,K_{NIP}}{1\pm\sigma}\ , \tag{6}$$

and

$$\sigma(d,h) = \frac{h}{d + K_{NIP}\,\sin\beta(d^2-h^2)}\ . \tag{7}$$

The four parameters $\{K_N, K_{NIP}, \beta, V_0\}$ have clear physical interpretations in terms of the wavefront and ray geometries (4). V_0 represents the velocity at the surface and is typically assumed known and constant around the central ray. One important property of the MF approximation is that, in a constant velocity medium with velocity V_0, it can accurately describe both reflections from a plane dipping interfaces and diffractions from point diffractors.

The CRS approximation (11) is

$$\widehat{\theta}_{CRS}(d,h;t_0) = \sqrt{F(d) + b_2\,h^2}\;, \tag{8}$$

where $F(d) = (t_0 + a_1\,d)^2 + a_2\,d^2$, and the three parameters $\{a_1, a_2, b_2\}$ are related to the multifocusing parameters as follows:

$$a_1 = \frac{2\,\sin\beta}{V_0}\;, \tag{9}$$

$$a_2 = \frac{2\,\cos^2\beta\,K_N\,t_0}{V_0}\;, \tag{10}$$

$$b_2 = \frac{2\,\cos^2\beta\,K_{NIP}\,t_0}{V_0}\;. \tag{11}$$

Equation (8) is equivalent to a truncated Taylor expansion of the squared traveltime in equation (4) around $d = 0$ and $h = 0$. In comparison with MF, CRS possesses a fundamental simplicity, which makes it easy to extend the method to 3-D. However, it looses the property of accurately describing diffractions in a constant-velocity medium.

We propose the following modification of approximation (8):

$$\widehat{\theta}(d,h;t_0) = \sqrt{\frac{F(d) + c\,h^2 + \sqrt{F(d-h)\,F(d+h)}}{2}}\;, \tag{12}$$

where $c = 2\,b_2 + a_1^2 - a_2$. Equation (12), which we call *non-hyperbolic common reflection surface*, is derived in Appendix A. A truncated Taylor expansion of the squared traveltime from equation (12) around $d = 0$ and $h = 0$ is equivalent to equation (8).

There are two important special cases:

1. If $a_2 = 0$ or $K_N = 0$, equation (12) becomes equivalent to equation (8), with $F(d) = (t_0 + a_1\,d)^2$. In a constant-velocity medium, this case corresponds to reflection from a planar reflector.

2. If $a_2 = b_2$ or $K_{NIP} = 0$, equation (12) becomes equivalent to

$$\widehat{\theta}(d,h;t_0) = \frac{\sqrt{F(d-h)} + \sqrt{F(d+h)}}{2}\;. \tag{13}$$

   In a constant-velocity medium, this case corresponds to a point diffractor.

3-D extension

In the case of 3-D multi-azimuth acquisition, both d and h become two-dimensional vectors. A natural way to extend approximation (8) is to replace it with

$$\widehat{\theta}_{CRS}(\mathbf{d},\mathbf{h};t_0) = \sqrt{F(\mathbf{d}) + \mathbf{h}^T\,\mathbf{B}_2\,\mathbf{h}}\;, \tag{14}$$

where $F(\mathbf{d}) = (t_0 + \mathbf{d}^T \mathbf{a}_1)^2 + \mathbf{d}^T \mathbf{A}_2 \mathbf{d}$, $\mathbf{a}_1$ is a two-dimensional vector, and $\mathbf{A}_2$ and $\mathbf{B}_2$ are two-by-two symmetric matrices (17). A similar approach works for extending approximation (12) to

$$\widehat{\theta}(\mathbf{d}, \mathbf{h}; t_0) = \sqrt{\frac{F(\mathbf{d}) + \mathbf{h}^T \mathbf{C} \mathbf{h} + \sqrt{F(\mathbf{d}-\mathbf{h})\, F(\mathbf{d}+\mathbf{h})}}{2}} \;, \tag{15}$$

where $\mathbf{C} = 2\,\mathbf{B}_2 + \mathbf{a}_1 \mathbf{a}_1^T - \mathbf{A}_2$. In the 3-D case, we have not found a simple connection between approximation (15) and the analytical reflection traveltime for a 3-D hyperbolic reflector.

ACCURACY COMPARISONS

Analytical Example

The first example we use to compare the accuracy of different approximations is that of a circular reflector under a homogeneous overburden. As shown in Appendix B, the exact traveltime in this case can be derived analytically in a parametric form. Obtaining a non-parametric closed-form expression in this case would require a solution of a high-order algebraic equation (13). Figure 1 compares the accuracy of CRS and nonhyperbolic CRS approximations for a range of offsets and midpoints. We display the relative absolute error as a function of offset to depth ratio and midpoint separation to depth ratio for a range of offsets and midpoints. The central midpoint is taken at the same horizontal distance from the center of the circle as the depth. The CRS approximation (8) develops an error both at large offsets and at large midpoint separations. The proposed non-hyperbolic CRS approximation (12) shows a significantly smaller error in the full range of offsets and midpoints. In our experiments, the multifocusing approximation (4) was even more accurate in this example. However, because of its different functional form, we focus our analysis on comparing CRS and non-hyperbolic CRS.

Numerical Example

In our next test, we generate a reflection traveltime surface by modeling reflection seismic data from a Gaussian-shape reflector, shown in Figure 2(a) by Kirchhoff modeling. The velocity changes linearly with depth. We extract the traveltime surface, shown in Figure 2(b), and fit it with different approximations by non-linear least-squares optimization.

Our method for fitting the multivariate time-correction function $\hat{\theta}(d, h; t_0, \mathbf{a})$ (either CRS or nonhyperbolic CRS) to a given experimental time data $t(d, h)$ by finding optimal parameters set $\mathbf{a}$ is a minimization approach, defined as follows:

$$\min_{\mathbf{a}} f(\mathbf{a}) \;, \tag{16}$$

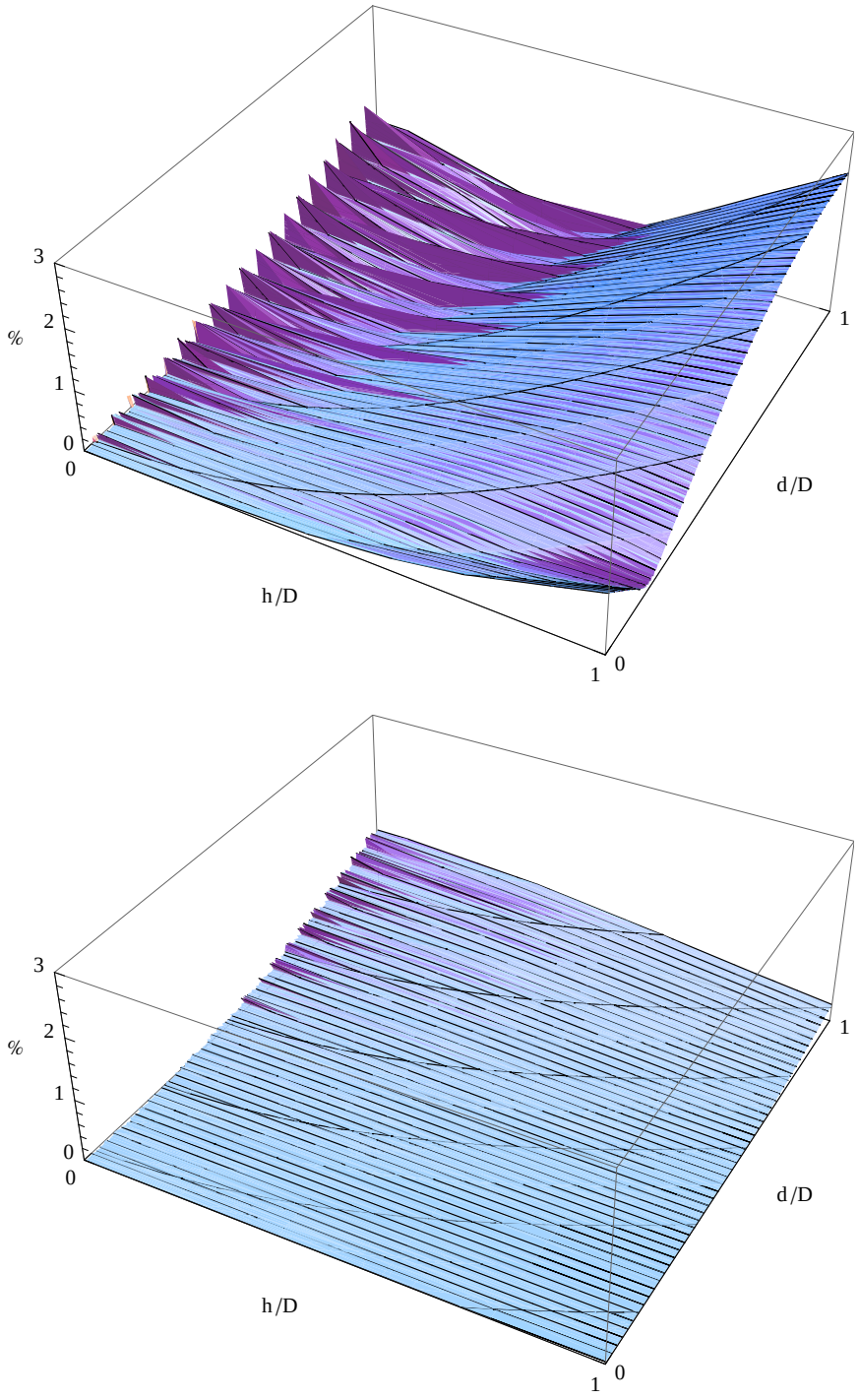


Figure 1: Relative error in prediction of reflection traveltime in the case of a circular reflector in a homogeneous medium using different approximations: (a) CRS, (b) non-hyperbolic CRS. The model parameters are $m = R = D$ (see Figure B-1.)
crs/. crs,ncrs

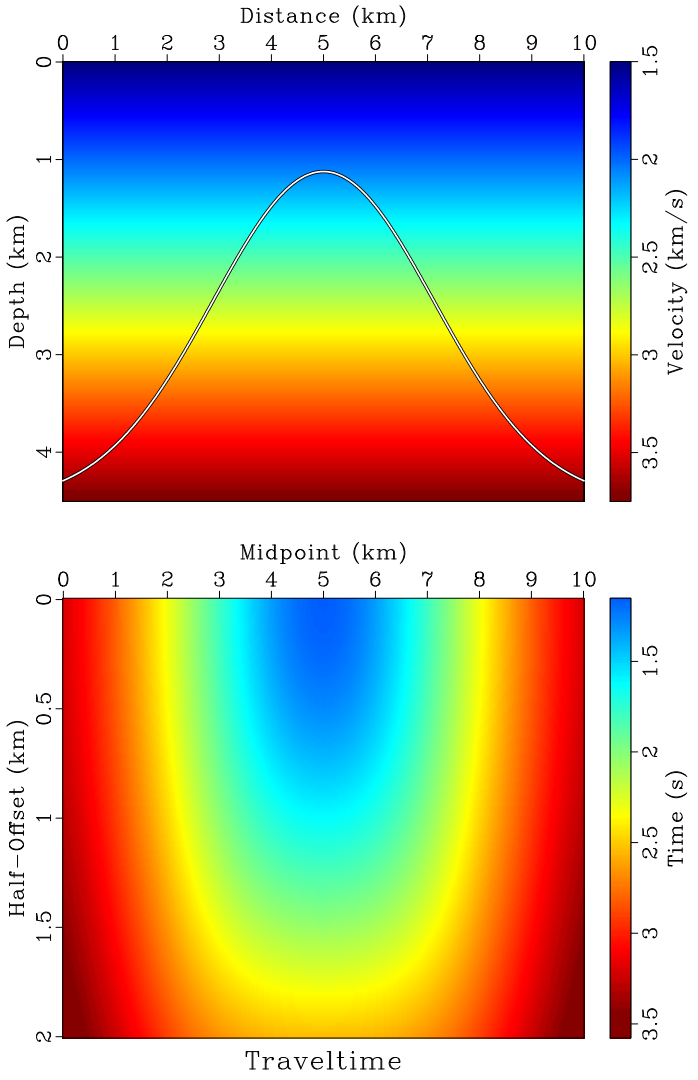


Figure 2: (a) Synthetic velocity model with a Gaussian-shape reflector. (b) Modeled reflection traveltime. crs/dome2 dome,pick

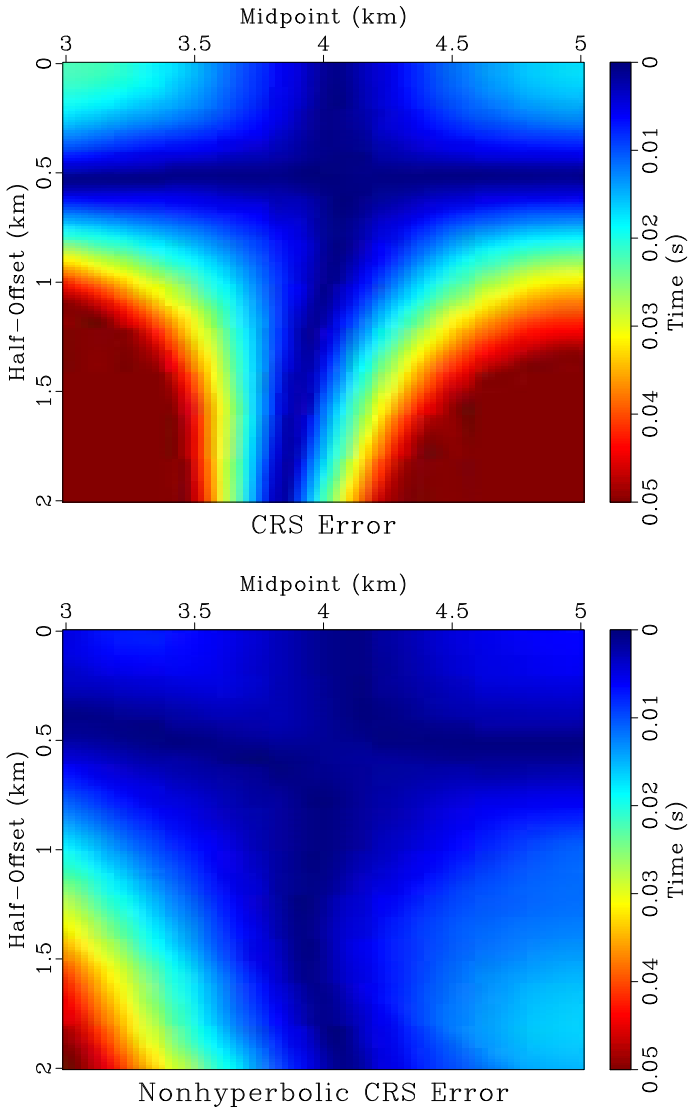


Figure 3: Absolute error of (a) CRS approximation, (b) nonhyperbolic CRS approximation. The reference midpoint is at 4 km. crs/dome2 crs-err4,ncrs-err4

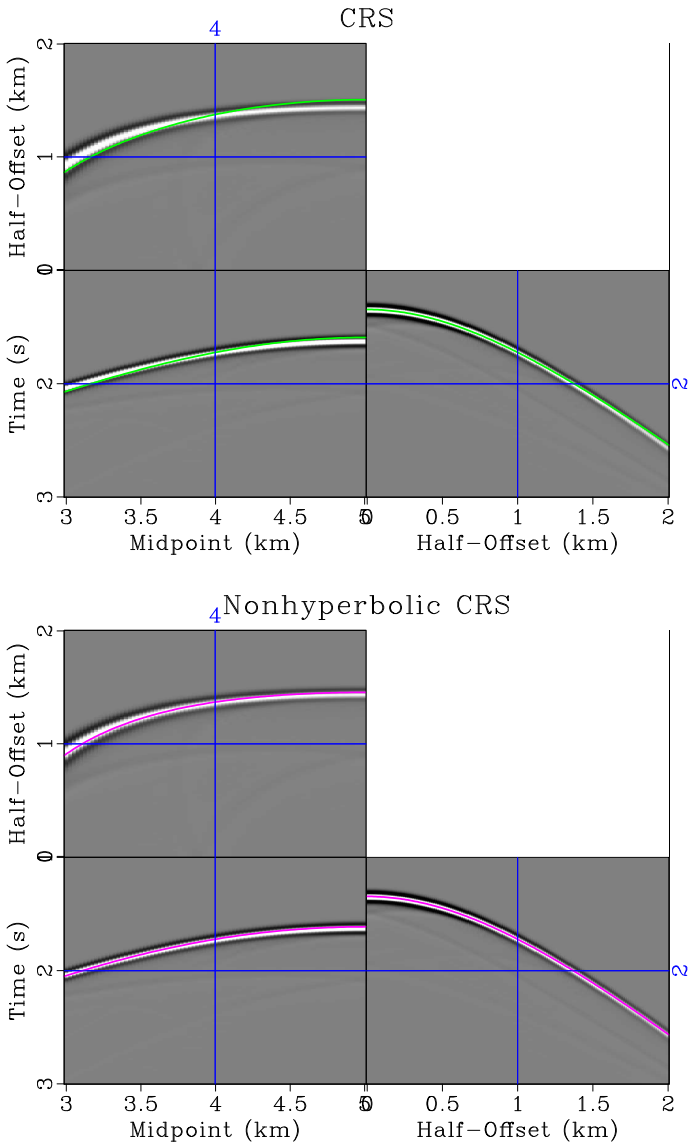


Figure 4: Modeled synthetic data with overlaid traveltime surfaces from least-squares fitting (colored curves). (a) CRS approximation. (b) Nonhyperbolic CRS approximation. The reference midpoint is at 4 km. crs/dome2 crs-data4,ncrs-data4

where f is the squared-error sum:

$$f(\mathbf{a}) = \frac{1}{2} \sum_d \sum_h |\hat{\theta}(d,h;t_0,\mathbf{a}) - t(d,h)|^2 \ . \tag{17}$$

The gradient of the objective function is defined by

$$\frac{\partial f}{\partial a_i} = \sum_d \sum_h \left[\hat{\theta}(d,h;t_0,\mathbf{a}) - t(d,h)\right] \frac{\partial \hat{\theta}}{\partial a_i} \ . \tag{18}$$

The partial derivatives of $\hat{\theta}$ are continuous functions. Therefore, classical minimization methods, such as the Gauss-Newton iteration, can be employed for finding a local minimum of $f(\mathbf{a})$ (1). In the case of CRS, the solution is unique because of the linear dependence of $\hat{\theta}^2$ on parameters.

The absolute error of CRS and nonhyperbolic CRS approximations, for a range of offsets and midpoints, is plotted in Figure 3. The nonhyperbolic CRS error is significantly smaller for a wide range of offsets and midpoints, which extends the applicability of the approximation. To effect of different approximations is shown additionally in Figure 4, which displays the modeled synthetic data with overlaid CRS and nonhyperbolic CRS approximations. The average relative error in different approximations, for the selected range of offsets and midpoints, is 1.24 % for the case of CRS and 0.41 % for the case of nonhyperbolic CRS. The average absolute error is 25 ms for the case of nonhyperbolic CRS and 8 ms for the case of nonhyperbolic CRS.

CONCLUSIONS

We have presented the non-hyperbolic common reflection surface, a new approximation for prestack reflection traveltimes. Non-hyperbolic CRS uses the same set of parameters as the hyperbolic CRS but in a different functional form, which can make the approximation significantly more accurate in a large range of offsets and midpoints. The proposed approximation is derived from the analytical expression of the reflection traveltime in the case of a hyperbolic reflector in a constant velocity medium.

Why use a hyperbolic reflector? A special property of this reflector is that it reduces to a plane reflector or a point diffractor with a special choice of parameters. ($z_0 = 0$ or $\alpha = \pi/2$ respectively). Thus, it encompasses two particularly important special cases.

Numerical experiments show that the new approximation can be significantly more accurate than the conventional hyperbolic CRS while using essentially the same set of parameters. The multifocusing approximation can be even more accurate but uses a different set of parameters, which makes it more difficult to extend it to 3-D.

ACKNOWLEDGMENTS

The first author is grateful to Emil Blias, Evgeny Landa, Tijmen Jan Moser, Alexey Stovas, and Martin Tygel for inspiring discussions.

This publication is authorized by the Director, Bureau of Economic Geology, The University of Texas at Austin.

APENDIX A: HYPERBOLIC REFLECTOR

In this appendix, we reproduce the derivation of an analytical expression for reflection traveltime from a hyperbolic reflector in a homogeneous velocity model (3). Similar derivations apply to an elliptic reflector and were used previously in the theory of offset continuation (16; 2).

Consider the source point s and the receiver point r at the surface $z = 0$ above a 2-D constant-velocity medium and a hyperbolic reflector defined by the equation

$$z(x) = \sqrt{z_0^2 + x^2 \, \tan^2 \alpha} \; . \tag{A-1}$$

The reflection traveltime as a function of the reflection point location y is

$$t = \frac{\sqrt{(s-y)^2 + z^2(y)} + \sqrt{(r-y)^2 + z^2(y)}}{V} \; . \tag{A-2}$$

According to Fermat's principle, the traveltime should be stationary with respect to the reflection point y:

$$\begin{aligned} 0 = \frac{\partial T}{\partial y} &= \frac{y - s + y \, \tan^2 \alpha}{V \sqrt{(s-y)^2 + z_0^2 + y^2 \, \tan^2 \alpha}} \\ &+ \frac{y - r + y \, \tan^2 \alpha}{V \sqrt{(r-y)^2 + z_0^2 + y^2 \, \tan^2 \alpha}} \; . \end{aligned} \tag{A-3}$$

Putting two terms in equation (A-3) on the different sides of the equation, squaring them, and reducing their difference to a common denominator, we arrive at the following quadratic equation with respect to y:

$$\begin{aligned} y^2 \, (s+r) \, \tan^2 \alpha &- 2 \, y \, \left(s \, r \, \sin^2 \alpha - z_0^2 \right) \\ &- z_0^2 \, (s+r) \, \cos^2 \alpha = 0 \; . \end{aligned} \tag{A-4}$$

Only one of the two branches of the solution

$$y = \frac{z_0^2 \, (s+r) \, \cos^2 \alpha}{z_0^2 - s \, r \, \sin^2 \alpha + \sqrt{\left(z_0^2 + s^2 \, \sin^2 \alpha\right) \left(z_0^2 + r^2 \, \sin^2 \alpha\right)}}$$

has physical meaning. Substituting this solution into equation (A-2), we obtain, after a number of algebraic simplifications,

$$\begin{aligned} t^2 &= \frac{2z_0^2 + s^2 + r^2 - 2\,s\,r\,\cos^2\alpha}{V^2} \\ &+ \frac{2\,\sqrt{(z_0^2 + s^2\,\sin^2\alpha)\,(z_0^2 + r^2\,\sin^2\alpha)}}{V^2}\,. \end{aligned} \tag{A-5}$$

Making the variable change in equation (A-5) from s and r to midpoint and half-offset coordinates m and h according to $s = m - h = m_0 + d - h$, $r = m + h = m_0 + d + h$, we transform this equation to form (12), where the following correspondence between parameters is applied:

$$z_0^2 = \frac{t_0^2\,a_2}{(a_1^2 + a_2)\,(a_1^2 + b_2)}\,, \tag{A-6}$$

$$m_0 = \frac{t_0\,a_1}{a_1^2 + a_2}\,, \tag{A-7}$$

$$\sin^2\alpha = \frac{a_1^2 + a_2}{a_1^2 + b_2}\,, \tag{A-8}$$

$$V^2 = \frac{4}{a_1^2 + b_2}\,. \tag{A-9}$$

The inverse relationships are given by

$$t_0 = \frac{2\,\sqrt{m_0^2\,\sin^2\alpha + z_0^2}}{V}\,, \tag{A-10}$$

$$a_1 = \frac{2\,m_0\,\sin^2\alpha}{V\,\sqrt{m_0^2\,\sin^2\alpha + z_0^2}}\,, \tag{A-11}$$

$$a_2 = \frac{4\,z_0^2\,\sin^2\alpha}{V^2\,\left(m_0^2\,\sin^2\alpha + z_0^2\right)}\,, \tag{A-12}$$

$$b_2 = \frac{4\,\left(m_0^2\,\sin^2\alpha\,\cos^2\alpha + z_0^2\right)}{V^2\,\left(m_0^2\,\sin^2\alpha + z_0^2\right)}\,. \tag{A-13}$$

The connection with the multifocusing parameters is summarized in Table 1 for the general case and three special cases (a plane dipping reflector, a flat reflector, and a point diffractor). The first two special cases turn the nonhyperbolic CRS equation into the hyperbolic form (8). The last case turns it into the double-square-root form (13).

APPENDIX B: CIRCULAR REFLECTOR

In the case of a circular (cylindrical or spherical) reflector in a homogeneous velocity model, the closed-form analytical solution is complicated, because it involves a solution of a high-order polynomial equation (13). However, the traveltime surface can be easily described analytically by parametric relationships (7).

	t_0	K_{NIP}	K_N	$\sin\beta$
Hyperbolic reflector	$\frac{2\sqrt{m_0^2\,\sin^2\alpha+z_0^2}}{V}$	$\frac{1}{\sqrt{m_0^2\,\sin^2\alpha+z_0^2}}$	$K_{NIP}\,\frac{z_0^2\,\sin^2\alpha}{m_0^2\,\sin^2\alpha\,\cos^2\alpha+z_0^2}$	$\frac{m_0\,\sin^2\alpha}{\sqrt{m_0^2\,\sin^2\alpha+z_0^2}}$
Plane dipping reflector $z_0=0$	$\frac{2\,m_0\,\sin\alpha}{V}$	$\frac{1}{m_0\,\sin\alpha}$	0	$\sin\alpha$
Flat reflector $\alpha=0$	$\frac{2\,z_0}{V}$	$\frac{1}{z_0}$	0	0
Point diffractor $\alpha=\pi/2$	$\frac{2\sqrt{m_0^2+z_0^2}}{V}$	$\frac{1}{\sqrt{m_0^2+z_0^2}}$	K_{NIP}	$\frac{m_0}{\sqrt{m_0^2+z_0^2}}$

Table 1: Multifocusing parameters for a hyperbolic reflector in a homogeneous medium ($V_0=V$). The general case and three special cases.

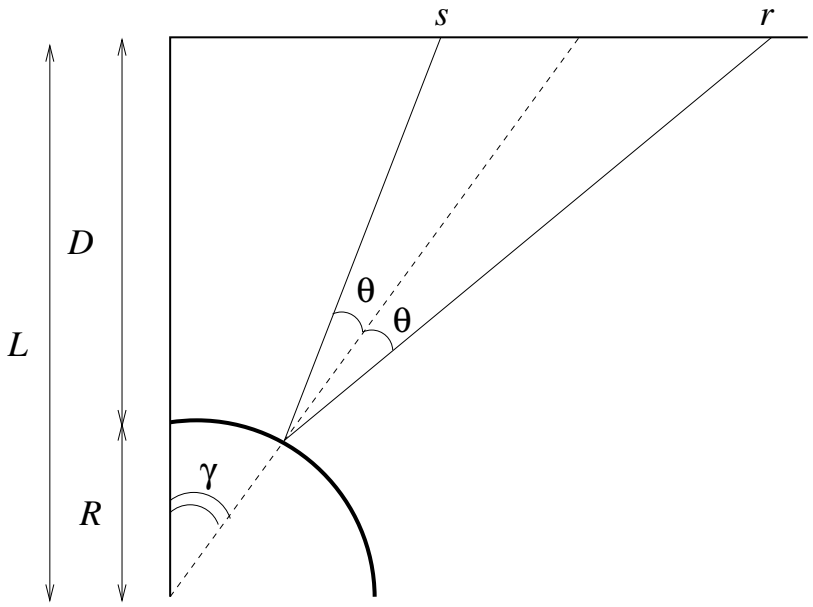


Figure B-1: Reflection from a circular reflector in a homogeneous velocity model (a scheme). crs/XFig crefl

Consider the reflection geometry shown in Figure B-1. According to the trigonometry of the reflection triangles, the source and receiver positions can be expressed as

$$s = R\sin\gamma + (L - R\cos\gamma)\tan(\gamma - \theta)\,, \tag{B-1}$$
$$r = R\sin\gamma + (L - R\cos\gamma)\tan(\gamma + \theta)\,, \tag{B-2}$$

where R is the reflector radius, D is the minimum reflector depth, $L = D + R$, γ is the reflector dip angle at the reflection point, and θ is the reflection angle. Correspondingly, the midpoint and half-offset coordinates are expressed as

$$\begin{aligned} m &= R\sin\gamma + (L - R\cos\gamma)\frac{\cos\gamma\,\sin\gamma}{\cos^2\theta - \sin^2\gamma}\,, \\ h &= (L - R\cos\gamma)\frac{\cos\theta\,\sin\theta}{\cos^2\theta - \sin^2\gamma}\,, \end{aligned} \tag{B-3}$$

and the reflection traveltime is expressed as

$$\begin{aligned} t &= \frac{L - R\cos\gamma}{V}\left[\frac{1}{\cos(\gamma - \theta)} + \frac{1}{\cos(\gamma + \theta)}\right] \\ &= 2\,\frac{L - R\cos\gamma}{V}\,\frac{\cos\gamma\,\cos\theta}{\cos^2\theta - \sin^2\gamma}\,, \end{aligned} \tag{B-4}$$

where V is the medium velocity. Equations (B-3-B-4) define the reflection traveltime surface $t(m,h)$ parametrically via the dependence $\{m(\gamma,\theta), h(\gamma,\theta), t(\gamma,\theta)\}$. The connection with the multifocusing parameters is given by

$$t_0 = \frac{2\left(\sqrt{m_0^2 + L^2} - R\right)}{V} \tag{B-5}$$
$$K_{NIP} = \frac{1}{\sqrt{m_0^2 + L^2} - R}\,, \tag{B-6}$$
$$K_N = \frac{1}{\sqrt{m_0^2 + L^2}}\,, \tag{B-7}$$
$$\sin\beta = \frac{m_0}{\sqrt{m_0^2 + L^2}}\,, \tag{B-8}$$

and $V_0 = V$.

REFERENCES

[1] Björck, A., 1996, Numerical methods for least squares problems: SIAM.

[2] Fomel, S., 2003, Theory of differential offset continuation: Geophysics, **68**, 718–732.

[3] Fomel, S., and A. Stovas, 2010, Generalized nonhyperbolic moveout approximation: Geophysics, **75**, U9–U18.

[4] Gelchinsky, B., A. Berkovitch, and S. Keydar, 1999a, Multifocusing homeomorphic imaging – Part 1. Basic concepts and formulas: Journal of Applied Geophysics, **42**, 229–242.

[5] ———, 1999b, Multifocusing homeomorphic imaging – Part 2. multifold data set and multifocusing: Journal of Applied Geophysics, **42**, 243–260.

[6] Gierse, G., J. Pruessmann, and R. Coman, 2006, CRS strategies for solving severe static and imaging issues in seismic data from Saudi Arabia: Geophysical Prospecting, **54**, 709–719.

[7] Glaeser, G., 1999, Reflections on spheres and cylinders of revolution: Journal for Geometry and Graphics, **3**, 121–139.

[8] Gurevich, B., S. Keydar, and E. Landa, 2002, Multifocusing imaging over an irregular topography: Geophysics, **67**, 639–643.

[9] Heilmann, Z., J. Mann, and I. Koglin, 2006, CRS-stack-based seismic imaging considering top-surface topography: Geophysical Prospecting, **54**, 681–695.

[10] Hoecht, G., P. Ricarte, S. Bergler, and E. Landa, 2009, Operator-oriented CRS interpolation: Geophysical Prospecting, **57**, 957–979.

[11] Jäger, R., J. Mann, G. Hocht, and P. Hubral, 2001, Common-reflection-surface stack: Image and attributes: Geophysics, **66**, 97–109.

[12] Landa, E., B. Gurevich, S. Keydar, and P. Trachtman, 1999, Application of multifocusing method for subsurface imaging: Journal of Applied Geophysics, **42**, 283–300.

[13] Landa, E., S. Keydar, and T. J. Moser, 2010, Multifocusing revisited – Inhomogeneous media and curved interfaces: Geophysical Prospecting, **58**, 925–938.

[14] Menyoli, E., D. Gajewski, and C. Hübscher, 2004, Imaging of complex basin structures with the common reflection surface (CRS) stack method: Geophysical Journal International, **157**, 1206–1216.

[15] Moser, T. J., and E. Landa, 2009, Multifocusing revisited – inhomogeneous media and curved interfaces: Presented at the 71st Mtg., Eur. Assn. Geosci. Eng.

[16] Stovas, A. M., and S. B. Fomel, 1996, Kinematically equivalent integral DMO operators: Russian Geology and Geophysics, **37**, 102–113.

[17] Tygel, M., and L. T. Santos, 2007, Quadratic normal moveouts of symmetric reflections in elastic media: A quick tutorial: Studia Geophysica et Geodaetica, **51**, 185–206.

[18] Tygel, M., L. T. Santos, and J. Schleicher, 1999, Multifocus moveout revisited: derivations and alternative expressions: Journal of Applied Geophysics, **42**, 319–331.

[19] Yilmaz, O., 2000, Seismic data analysis: Soc. of Expl. Geophys.

Carbonate reservoir characterization using seismic diffraction imaging

*Luke Decker, Xavier Janson, and Sergey Fomel**

ABSTRACT

Although extremely prolific worldwide, carbonate reservoirs are challenging to characterize using traditional seismic reflection imaging techniques. We use computational experiments with synthetic models to demonstrate the possibility seismic diffraction imaging has of overcoming common obstacles associated with seismic reflection imaging and aiding interpreters of carbonate systems. Diffraction imaging improves the horizontal resolution of individual voids in a karst reservoir model and identification of heterogeneous regions below the resolution of reflections in a reservoir scale model.

INTRODUCTION

Carbonate reservoirs contain a majority of remaining proven oil reserves, yet are much more difficult to evaluate than their siliciclastic counterparts (11; 27; 6; 30). Many aspects of carbonate rocks make their seismic signature complex and difficult to interpret both qualitatively and quantitatively. Because carbonate rocks generally have higher seismic velocities than siliciclastics, horizontal and vertical resolution is commonly low.

Carbonate sediments are also more prone to complex, rapid diagenetic alteration, where heat and pressure change rock chemistry, after deposition and continuing through the burial process (31). These diagenetic processes significantly affect the acoustic properties of carbonate rocks. Postdepositional alteration such as karst processes, where weathering creates steep sided valleys and cave networks in carbonate strata, or dolimitization, where calcium carbonate is replaced by calcium magnesium carbonate (dolomite), can further complicate already heterogeneous deposits. Carbonate heterogeneity exists at different scales. Carbonates often posses larger-scale voids, caves, and fracture networks, accompanied by small scale features such as microfractures, intergranular porosity, and chemical alteration (24).

The acoustic properties of carbonate rocks are not a simple function of mineralogy and porosity. Recent advances in rock mechanics have shown that carbonate rocks' acoustic properties also depend on their pore type, size, shape, and distribution (32; 6; 1; 33). Heterogeneities in carbonates scatter seismic energy, attenuating high

*e-mail: sergey.fomel@beg.utexas.edu

frequency signal and reducing resolution. High impedance contrasts typically exist between carbonate structures and surrounding rocks, leading to a strong reflective interface that can generate multiple reflections. Impedance contrasts for strata within the carbonate body tend to be relatively weak, and horizontal homogeneity means that reflections rarely have strong lateral continuity.

Rock physics models developed for siliciclastics often fail to effectively describe carbonate systems (29; 2). This makes relating velocities from core, sonic logs, and seismic very difficult, as these are sampled with different frequency waves. As a result, seismic images of carbonate deposits are usually not easily interpreted, especially at the reservoir scale (1-5 km). In addition, because of the intertwined control factors on the seismic response of carbonates, quantitative interpretation of the seismic signal is even more challenging. (16) and (14) used outcrop analogues and synthetic models to better understand the seismic response of carbonate deposits.

The difficulties associated with reflection imaging in carbonates encourage us to explore alternative imaging approaches, such as diffraction imaging. Seismic diffractions are a fundamentally different phenomenon than seismic reflections (20). They occur when seismc waves scatter from small-scale features. Diffractions may be caused by geologically significant features including voids, faults, fractures, karsts, pitchouts, salt flanks, and other small-scale heterogeneities (13; 19; 8; 25; 21). Rays associated with seismic diffractions take more diverse paths than those associated with reflection events, and thus can contain more information about the subsurface (26). These more diverse ray paths enable super-resolution with seismic diffraction imaging (19).

Seismic diffraction imaging can highlight features commonly observed in carbonates, such as karsts, voids, and small scale heterogeneities, with high resolution. These characteristics make seismic diffraction imaging well suited for use with carbonate imaging targets, where reflection resolution is typically limited. In this paper, we use two synthetic models to illustrate how seismic diffraction imaging can better constrain void geometry and detect heterogeneous zones that may not be immediately apparent in conventional reflection imaging.

SYNTHETIC MODELS

Ordovician Model

Our first synthetic model is based on the very deeply buried (5,500-6,500 m) Ordovician limestone strata in Northwest China's Tarim Basin, which features anomalous seismic amplitude bright spots. The amplitude bright spots correspond to high-gamma-ray, low-velocity zones in wireline logs and have been interpreted by (34) as paleokarst features. A geocellular model was built to study the seismic response of the paelokarst in detail (16; 35). The synthetic model uses the Ordovician unconformity surface (boundary between a basal Ordovician interval and an overlying Silurian

siliciclastic interval) that was mapped from subsurface seismic data (Figure 1). Collapsed paleocaves with cave sediments were modeled by randomly distributing low acoustic impedance (AI) circular geobodies that measured 300 × 300 m in the horizontal dimension and 18 m in vertical dimension. The AI (approximated by acoustic velocity using constant density) is distributed using a sequential Gaussian simulation with parameters derived from a sonic log in the cored well (16; 34; 35).

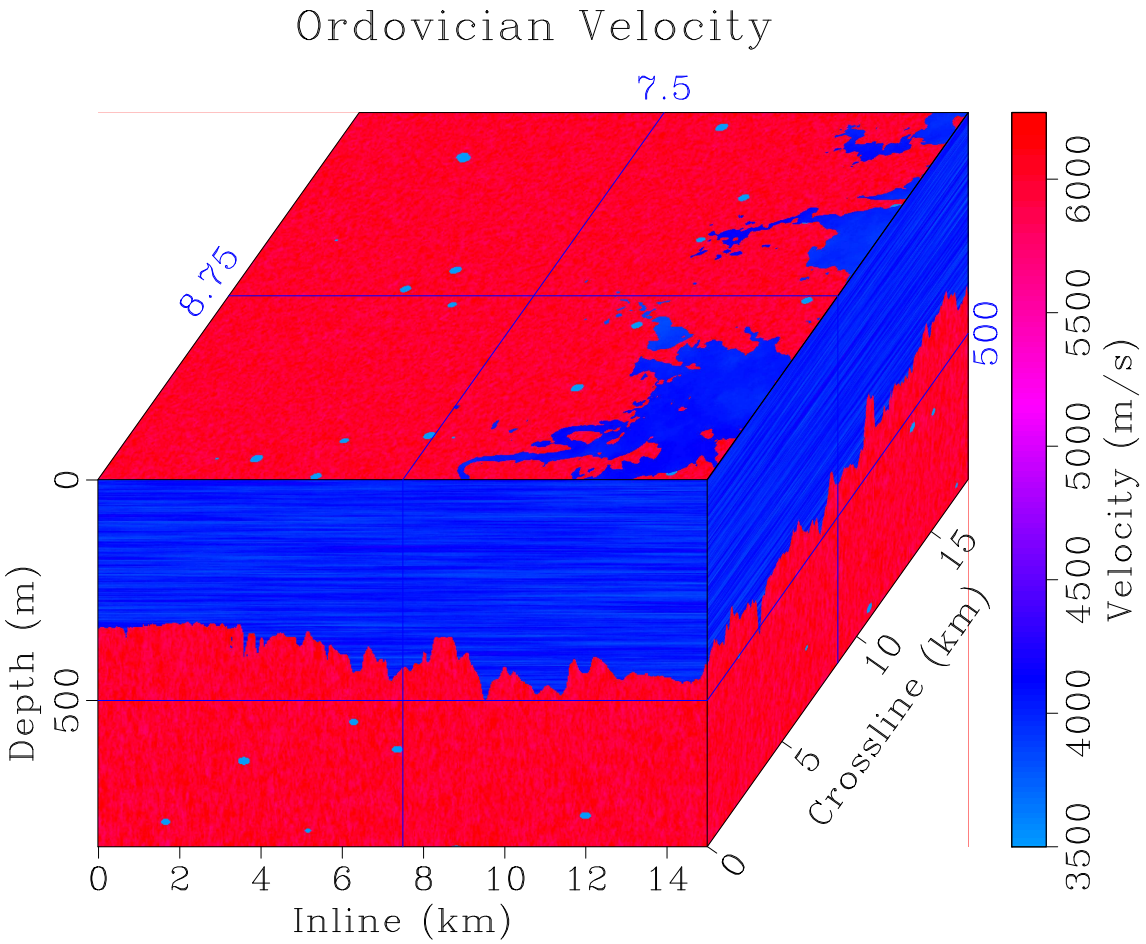


Figure 1: Ordovician velocity model. synth/ordexample20hz vp

Permo Triassic Khuff Model

Our second model examines rocks equivalent to the Permian-Triassic Khuff-A and -B reservoirs, which crop out near Buraydah in central Saudi Arabia. An outcrop-based geocellular model 600 m × 385 m × 30 m was built to investigate the effect of small-scale carbonate reservoir heterogeneities on subsurface flow models (15). In addition, the 3D geological geocellular model was converted into a acoustic impedence (AI) model using laboratory velocity (Figure 2(a)) and density (Figure 2(b)) measurement

from outcrop plugs. An average acoustic impedance value for each lithofacies present in the geological model was used to convert the lithofacies model into an impedance volume (Figure 2(c)) in order to maintain the realistic level and distribution of reservoir heterogeneities. Because of its limited size, the outcrop-based geocellular model was scaled up for seismic modeling by addition of a similar but simpler model of strata below it as well as acoustically constant buffer layers above and below to make the final model 110 m thick.

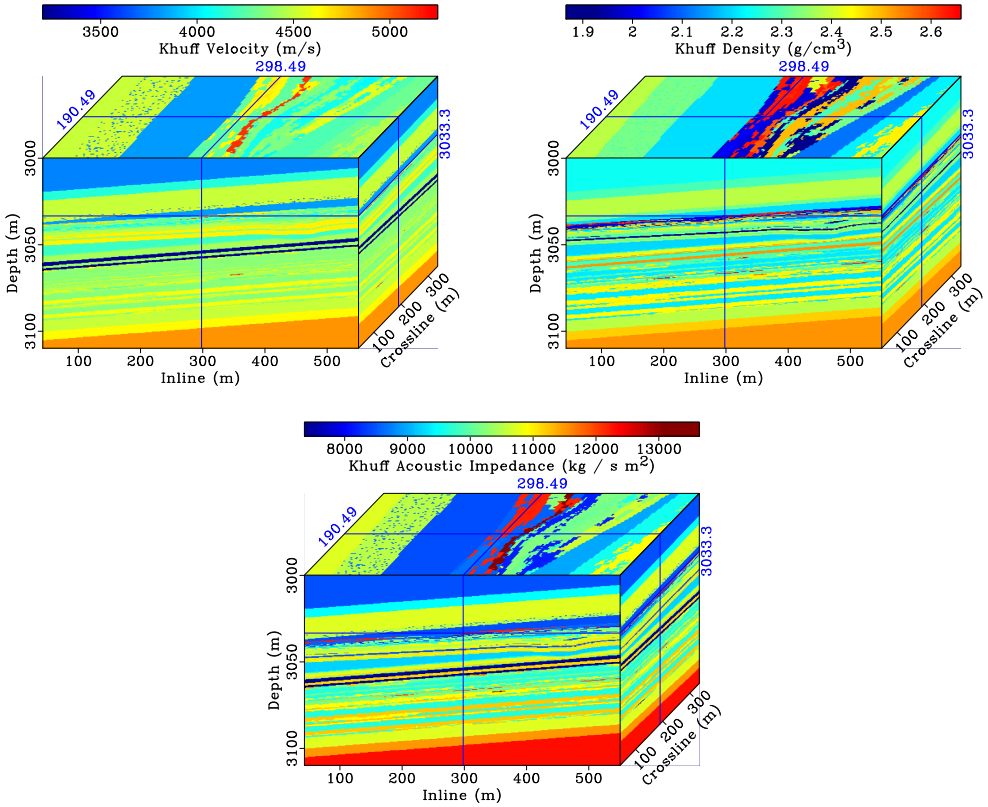


Figure 2: Khuff synthetic model: (a) velocity; (b) density; (c) acoustic impedance synth/small-khuff kvp,kden,kai

METHOD

Seismic diffraction events carry much less energy than reflection events, requiring that they be separated to be utilized. Several methods for seismic diffraction extraction exist (13; 23; 17; 19; 22; 21), including plane-wave destruction applied to common-offset data (8).

Plane-wave destruction (PWD) filters (3; 7), determine the dominant slope of seismic events as they attempt to map data to adjacent traces. Data not conforming to the local slope field is iteratively minimized. Because reflection events appear

planar in common-offset data while diffraction events appear hyperbolic, this residual will contain the set of diffractions along with random noise present in the data (13).

Zero offset data are modeled using methods described in the subsequent section. We use PWD to determine the dominant slope field of our modeled zero-offset data and remove the reflections that conform with local slope, providing us with zero-offset diffraction data. Zero-offset "conventional" data containing diffractions and reflections as well as zero-offset diffraction data are migrated, providing our conventional and diffraction images respectively. A workflow for the diffraction extraction and imaging process starting from common-offset data is displayed in Figure 3. Plane-wave destruction of common-offset data may face difficulties extracting diffractions in regions with complex geometry or velocity structure (5). The synthetic models we use in this paper have small enough lateral velocity variations for this to not be an issue. If the wavefield is sufficiently complicated to prevent common-offset data plane-wave destruction from functioning properly other methods of separating diffractions exist, including plane-wave destruction of angle-migrated partial images (4).

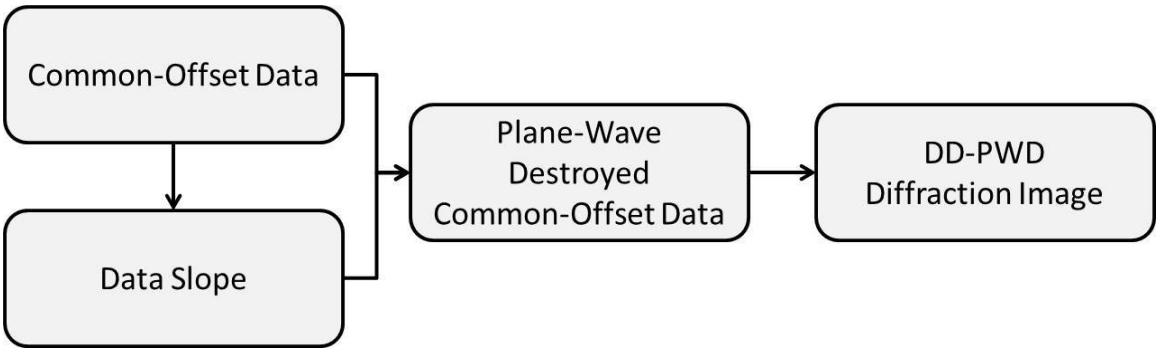


Figure 3: Our diffraction imaging workflow synth/. DD-PWD

Although we employ the same method of diffraction extraction on both the Orodovician and Khuff synthetic models, we adopt different methods of modeling and migration that are best suited for each model's scale and subsurface position. Reverse-time migration (36; 10) is used on the Ordovician model for greater accuracy while one-way wave-equation migration (12; 18) is utilized on the reservoir-scale Khuff model to allow for upward continuation of modeled data through an overburden to model the response of the interval at a geologically plausible depth.

RESULTS

Ordovician

We begin our experiment on the constant-density Ordovician synthetic velocity model, Figure 1, by calculating reflectivity. We transform this reflectivity to the time domain, convolve it with a 20 Hz ricker wavelet, and transform it back to the depth

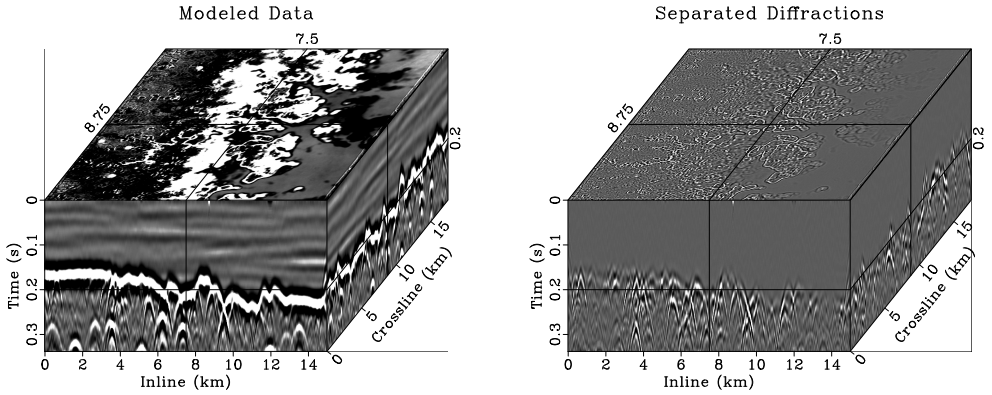


Figure 4: Modeled Ordovician zero-offset data: (a) conventional data; (b) separated diffraction data synth/ordexample20hz exp0,diffractions

domain to create an idealized seismic reflectivity image. Zero-offset data is generated by performing time domain low-rank forward modeling on the idealized image (10), providing us with conventional zero-offset data, Figure 4(a).

Data slopes are calculated using PWD, and events conforming with that slope are removed, providing the set of diffraction data shown in Figure 4(b). Conventional and diffraction data are migrated using low-rank RTM with a smoothed-slowness velocity field to provide a conventional image, shown in Figure 5(a), and a diffraction image, shown in Figure 5(b).

To highlight the improvement in horizontal feature resolution using diffraction imaging we take two depth slices from the images, which will be interpreted and discussed in the subsequent section. The depth slices represent the average of a 20 m interval centered around the target depth. Depth slices of the conventional and diffraction images for the first depth, 0.55 km, are visible in Figures 6(a) and 6(b). We zoom in on an interesting region of these slices featuring the karst interface and several superimposed voids to generate Figure 7.

Conventional and diffraction image slices for the second depth, 0.7 km, are shown in Figure 8. We zoom in on an area with several closely spaced voids, creating Figure 9.

Khuff

We use the Khuff synthetic model to illustrate how diffractions may be used to characterize features at the reservoir scale using higher frequency data.

The experiment begins with the Khuff velocity and density models, shown in Figures 2(a) and 2(b) respectively. We multiply density and velocity data to obtain acoustic impedance (Figure 2(c)). Reflectivity is calculated from this acoustic

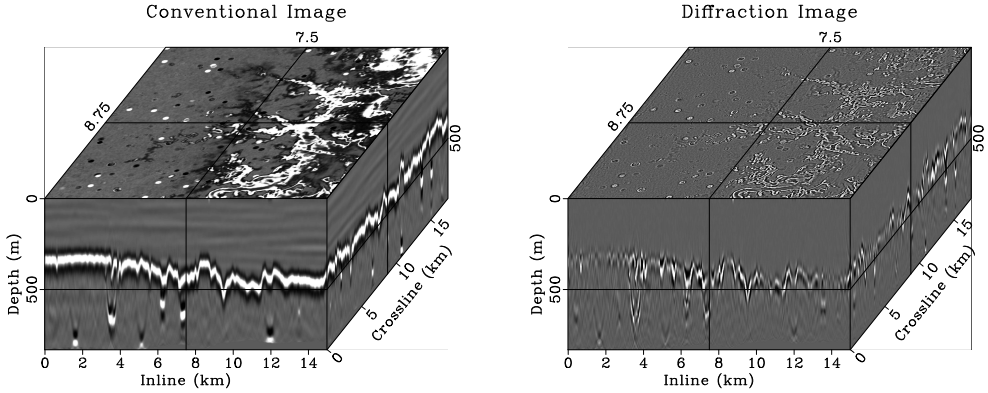


Figure 5: Low-rank RTM smoothed-slowness images for Ordovician model: (a) conventional image; (b) seismic diffraction image synth/ordexample20hz exp0-mig,diffractions-mig

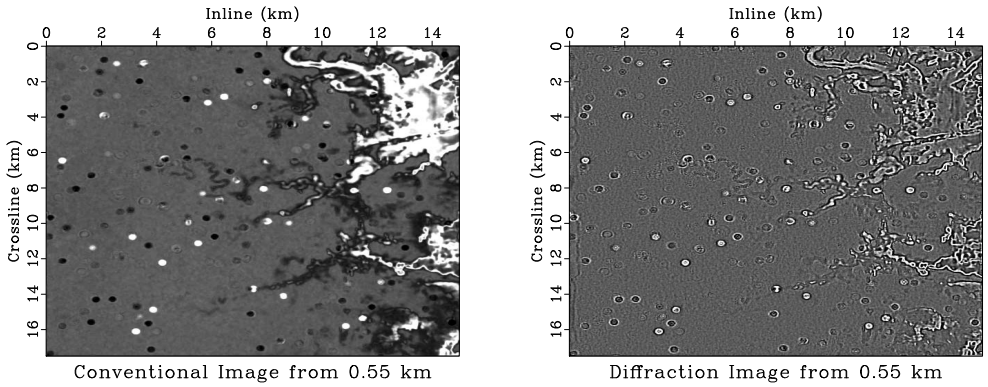


Figure 6: Ordovician depth slices from 0.55 km: (a) conventional image; (b) seismic diffraction image synth/ordexample20hz img2dstk-1,dif2dstk-1

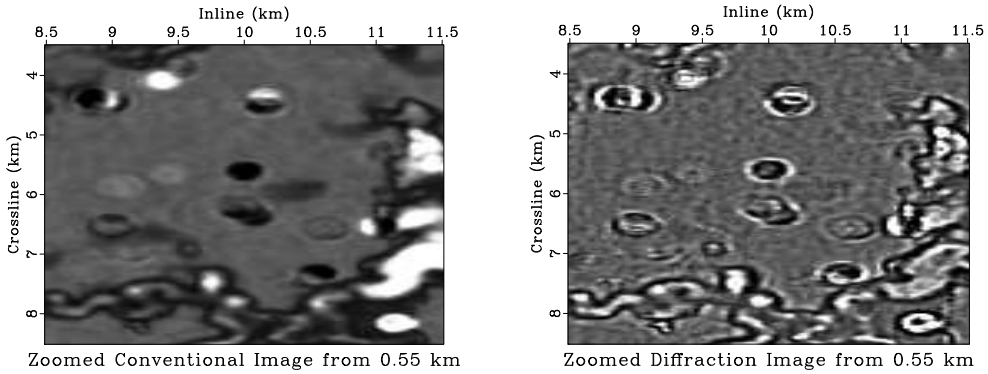


Figure 7: Zoomed Ordovician depth slices from 0.55 km: (a) conventional image; (b) seismic diffraction image synth/ordexample20hz img2dstka-1,dif2dstka-1

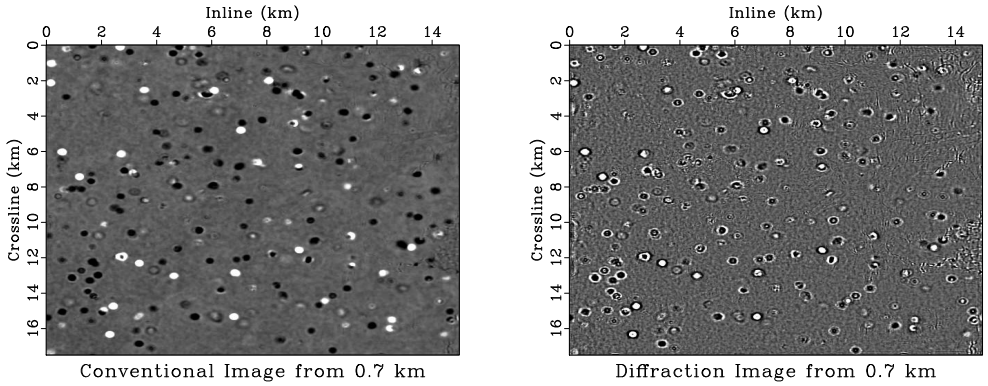


Figure 8: Ordovician depth slices from 0.7 km: (a) conventional image; (b) seismic diffraction image synth/ordexample20hz img2dstk-2,dif2dstk-2

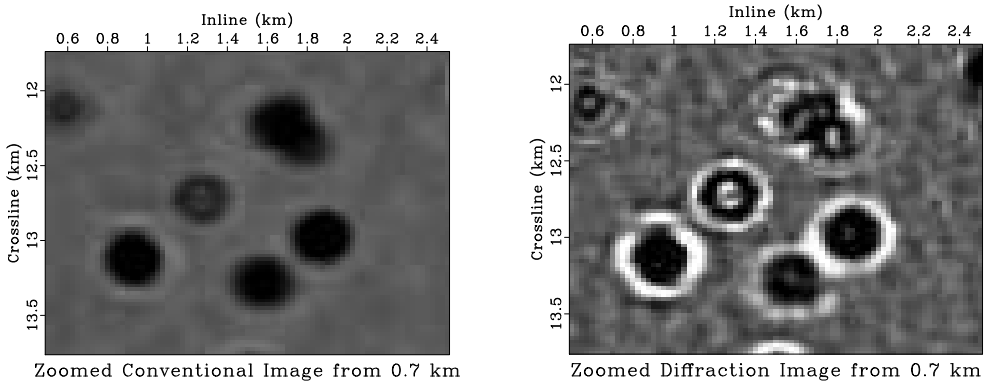


Figure 9: Zoomed Ordovician depth slices from 0.7 km: (a) conventional image; (b) seismic diffraction image synth/ordexample20hz img2dstkb-2,dif2dstkb-2

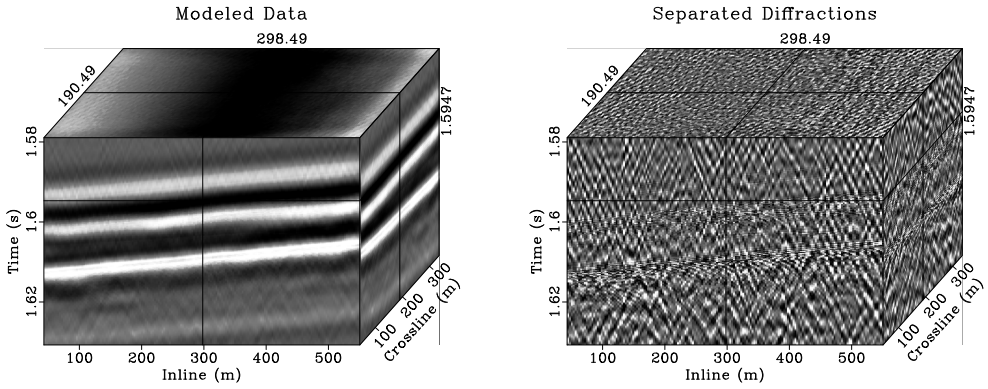


Figure 10: Zero-offset Khuff data: (a) conventional; (b) diffraction synth/small-khuff zo-ovr,k-diffractions

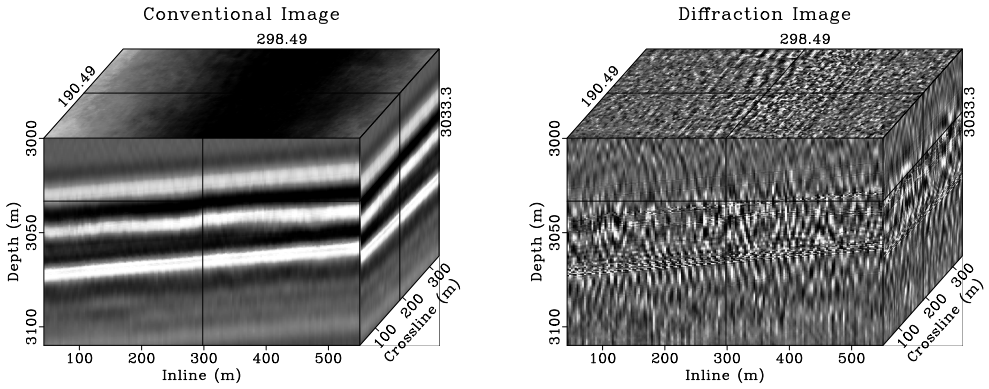


Figure 11: Migrated Khuff images: (a) conventional; (b) diffraction synth/small-khuff zo-ovr-mig,k-diffractions-mig

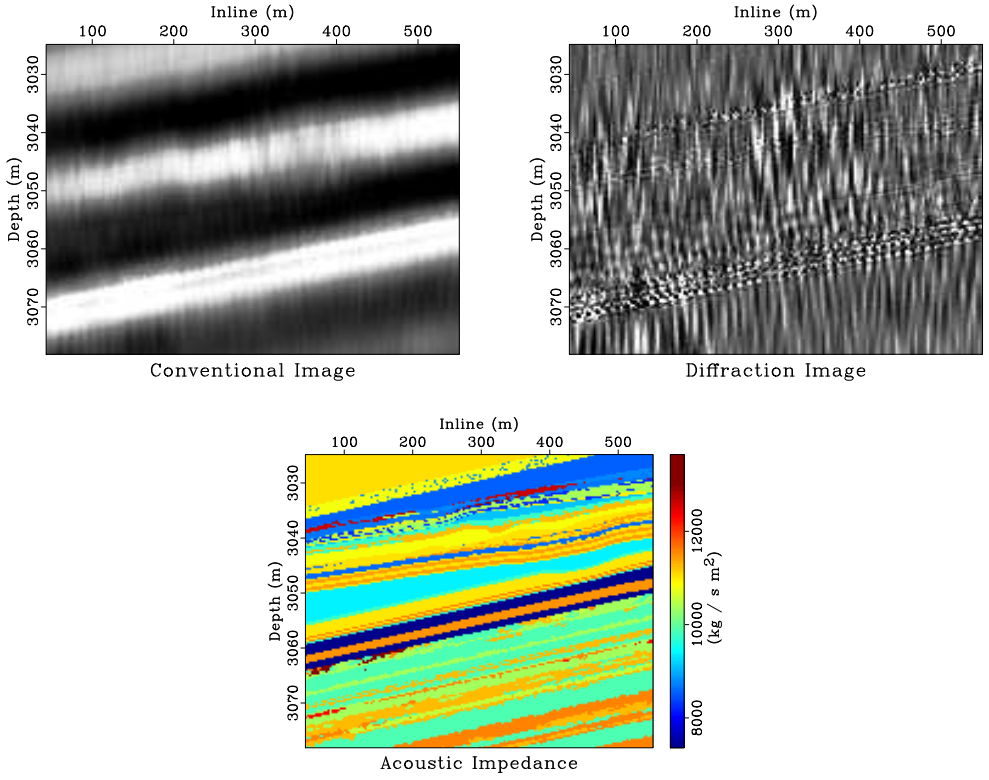


Figure 12: Khuff cross sections for 150 m crossline: (a) conventional image; (b) diffraction image; (c) acoustic impedance model synth/small-khuff zo-ovr-migI,k-diffractions-migI,aiI

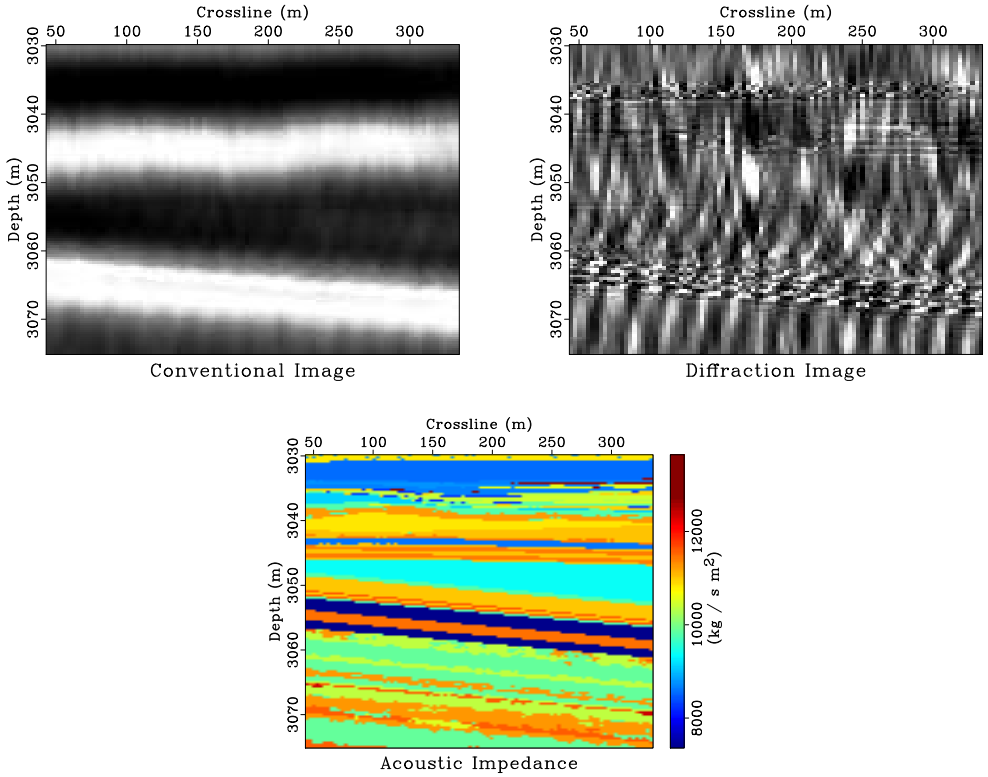


Figure 13: Khuff cross sections for 250 m inline: (a) conventional image; (b) diffraction image; (c) acoustic impedance model synth/small-khuff zo-ovr-migX,k-diffractions-migX,aiX

impedance, transformed to the time domain, convolved with a 100 Hz ricker wavelet, and transformed back to the depth domain to provide an idealized seismic reflection image. We model the zero-offset reservoir response using one-way wave equation modeling in the frequency domain (28), and then upward continue the reservoir response through a 3 km thick overburden to generate the zero-offset data, shown in Figure 10(a).

We separate diffractions using PWD. Data slopes are calculated and reflection events conforming to slope are removed, leaving zero-offset data with primarily diffractions (Figure 10(b)).

Conventional and diffraction zero-offset data are then downward continued through the smoothed-slowness overburden, and then depth migrated through the smoothed-slowness reservoir. This provides a conventional image (Figure 11(a)) and a diffraction image (Figure 11(b)). We zoom in on a horizontal cross section along crossline 150 m for the conventional and diffraction images as well as the acoustic impedance of the synthetic model, creating Figure 12. We also generate a cross section along inline 250 m (Figure 13).

INTERPRETATION

The following analysis shows that seismic diffraction imaging can better highlight small features in synthetic models than reflection imaging.

Ordovician

Seismic diffraction imaging improves the resolution of voids present in time slices relative to seismic reflection imaging. Examining the zoomed conventional and diffraction images from 0.55 km depth, Figures 7(a) and 7(b), we notice that diffraction imaging enables us to tell that what appears as single shapes in the reflection image are actually superpositions of multiple void responses. If we examine the features centered at inline 8.85 km, crossline 4.5 km; inline 10.15 km, crossline 4.25 km; and inline 10 km crossline 6.5 km in the conventional image, Figure 7(a), we see responses that may appear to be single voids. In the corresponding diffraction image, Figure 7(b), these shapes separate into joined rings, which define the edges of two overlaying voids.

The deeper slices from 0.7 km depth illustrate how diffraction imaging increases void edge resolution. If we compare the voids visible in the zoomed image (Figure 9), using the diffraction image we are better able to tell where void edges are located; they are marked by the reverse of seismic polarity. Additionally, diffraction imaging enables us to see that the feature centered at inline 1.65 km, crossline 12.25 km is actually a superposition of two nearby voids.

Therefore, using seismic diffraction imaging methods on the Ordovician model we are able to better distinguish between overlaying voids in depth slices, and better

spatially locate void edges.

Khuff

Seismic diffractions in the Khuff model highlight two strata with increased heterogeneity that are not immediately apparent in the conventional image.

If we examine the Khuff seismic diffraction image, Figure 11(b) we notice that amongst a chaotic diffraction background, there are two upward sloping linear features which intersect the left side of the image cube's inline axis near depths of 3040 m and 3065 m, and the right side of the image cube's inline axis near 3030 m and 3055 m. These layers correspond to the heterogeneous zones in the acoustic impedance model, Figure 2(c). These heterogeneous regions are lost in the reflection image, Figure 11(a), which features a series of parallel reflections.

Examining the diffraction image cross sections, Figures 12(b) and 13(b) provides a clearer view of the heterogeneous layers visible in the acoustic impedance cross sections, Figures 12(c) and 13(c), which remain less apparent in the corresponding conventional image cross sections, Figures 12(a) and 13(a). Although many of the features in the Khuff model are below diffraction resolution, the different scattering behavior and intensity is clearly helpful for detection of heterogeneous regions.

The heterogeneous strata are also apparent in the Khuff seismic diffraction data, Figure 10(b). These strata, located where at the left edge of the cube's inline axis near Time 1.6 s and Time 1.618 s slope upward to the right, and are rich in hyperbolic diffractions.

We can conclude that applying seismic diffraction imaging methods on the Khuff model enables us to more accurately determine regions of heterogeneity in a reservoir-scale model.

CONCLUSIONS

Using two synthetic models, we have investigated the potential of seismic diffraction imaging for aiding interpreters of carbonate systems. We use the first model to demonstrate how seismic diffraction imaging can better constrain the edges of voids and distinguish between the superposition of overlaying features. We use the second model to illustrate how diffraction imaging can detect reservoir-scale heterogeneous zones that might be indistinguishable in a conventional reflection image. The use of synthetic models is effective for comparing the differences between seismic diffraction and reflection imaging results, but additional case studies with field-data are required to verify the effectiveness of these promising methods on real carbonate systems.

ACKNOWLEDGMENTS

We thank Shell for partial financial support of this work. The computational experiments were completed using the Madagascar software package (9) and are reproducible in the Madagascar environment.

REFERENCES

[1] Adam, L., and M. Batzle, 2008, Elastic properties of carbonates from laboratory measurements at seismic and ultrasonic frequencies: The Leading Edge, **27**, no. 8, 1026–1032.

[2] Baechle, G., A. Colpaert, G. Eberli, and R. Weger, 2008, Effects of microporosity on sonic velocity in carbonate rocks: The Leading Edge, **27**, no. 8, 1012–1018.

[3] Claerbout, J. F., 1992, Earth sounding analysis: Processing versus inversion: Blackwell Scientific Publications, Inc.

[4] Decker, L., and A. Klokov, 2014, Diffraction extraction by plane-wave destruction of partial images: SEG Technical Program Expanded Abstracts, Accepted.

[5] Decker, L., A. Klokov, and S. Fomel, 2013, Comparison of seismic diffraction imaging techniques: Plane wave destruction versus apex destruction: SEG Technical Program Expanded Abstracts, 4054–4059.

[6] Eberli, G. P., G. T. Baechle, F. S. Anselmetti, and M. L. Incze, 2003, Factors contrillong elastic properties in carbonate sediments and rocks: The Leading Edge, **22**, 654–660.

[7] Fomel, S., 2002, Applications of plane wave destruction filters: Geophysics, **67**, no. 6, 1946–1960.

[8] Fomel, S., E. Landa, and M. T. Taner, 2007, Poststack velocity analysis by separation and imaging of seismic diffractions: Geophysics, **72**, no. 6, U89–U94.

[9] Fomel, S., P. Sava, I. Vlad, Y. Liu, and V. Bashkardin, 2013a, Madagascar: open-source software project for multidimensional data analysis and reproducible computational research: Journal of Open Research Software, **1**, e8.

[10] Fomel, S., L. Ying, and X. Song, 2013b, Seismic wave extrapolation using lowrank symbol approximation: Geophysical Prospecting, **61**, 526–536.

[11] Fontaine, J. M., R. Cussey, J. Lacaze, R. Lanaud, and L. Yapaudjian, 1987, Seismic interpretation of carbonate depositional environments: AAPG Bulletin, **71**, 281–297.

[12] Gazdag, J., and P. Sguazzero, 1984, Migration of seismic data by phase-shift plus interpolation: Geophysics, **49**, 124–131.

[13] Harlan, W., J. Claerbaut, and F. Rocca, 1984, Signal to noise separation and velocity estimation: Geophysics, **49**, no. 11, 1869–1880.

[14] Janson, X., and S. B. Fomel, 2011, 3-D forward seismic model of an outcrop-based geocellular model, *in* Outcrops revitalized: tools, techniques and applications: Martinsen, O., Pulham, A.J., Haughton, P.D.W., and Sullivan, M.D. eds, volume **10** *of* SEPM Concepts in Sedimentology and Paleontology, 87–106.

[15] Janson, X., F. J. Lucia, J. A. Bellian, A. A. AbuBshai, R. K. Al-Dukhayyail,

H. W. Mueller III, and D. Cantrell, 2013, Chapter 12: Outcrop-based 3d geological and reservoir model of the uppermost Khuff Formation in central Saudi Arabia, *in* Permo-Triassic Sequence of the Arabian Plate: Pöpelreiter, M., ed., EAGE Special Publication, 269–302.

[16] Janson, X., H. Zeng, B. Loucks, A. John, Y. G. Jackson, Q. Wang, and C. Wang, 2010, An Ultra-Deep Paleokarst System in the Ordovician North-Central Tarim Basin China: Outcrop Analog And Synthetic Seismic Models: 80th Annual International Meeting, SEG, Expanded Abstracts.

[17] Kanasewich, E. R., and S. M. Phadke, 1988, Imaging discontinuities on seismic sections: Geophysics, **53**, no. 3, 334–345.

[18] Kessinger, W., 1992, Extended split-step Fourier migration: 62nd Annual International Meeting, SEG, Expanded Abstracts, 917–920.

[19] Khaidukov, V., E. Landa, and T. Moser, 2004, Diffraction imaging by focusing-defocusing: an outlook on seismic super resolution: Geophysics, **56**, 1478–1490.

[20] Klem-Musatov, K., 1994, Theory of seismic diffractions: SEG.

[21] Klokov, A., and S. Fomel, 2012, Separation and imaging of seismic diffractions using migrated dip-angle gathers: Geophysics, **77**, S131–S143.

[22] Landa, E., S. Fomel, and M. Reshef, 2008, Separation, imaging, and velocity analysis of seismic diffractions using migrated dip-angle gathers: 78th Annual International Meeting, SEG, Expanded Abstracts, 2176–2180.

[23] Landa, E., V. Shtivelman, and B. Gelchinsky, 1987, A method for detection of diffracted waves on common-offset sections: Geophysical Prospecting, **34**, 359–373.

[24] Lucia, F. J., 1999, Carbonate Reservoir Caracterization: An Integrated Approach: Springer.

[25] Moser, T., and C. Howard, 2008, Diffraction imaging in depth: Geophysical Prospecting, **56**, 1365–2478.

[26] Neidell, N. S., 1997, Perceptions in seismic imaging, Part 2: Reflective and diffractive contributions to seismic imaging: The Leading Edge, **16**, no. 9, 1121–1123.

[27] Palaz, I., and K. J. Marfurt, 1997, Chapter 1: Carbonate Seismology: an overview, *in* Carbonate Seismology: I. Palaz and K. J. Marfurt, ed., volume **6** *of* SEG Geophysical Developments, 1–7.

[28] Sava, P., 2007, Stereographic imaging condition for wave-equation migration: Geophysics, **72**, no. 6, A87–A91.

[29] Sayers, C., 2008, The elastic properties of carbonates: The Leading Edge, **27**, no. 8, 1020–1024.

[30] Sayers, C., and R. Latimer, 2008, An introduction to this special section: Carbonates: The Leading Edge, **27**, no. 8, 1010–1011.

[31] Vanario, T., C. Scotellard, and G. Mavko, 2008, The effect of chemical and physical processes on the acoustic properties of carbonate rocks: The Leading Edge, **27**, no. 8, 1040–1048.

[32] Wang, Z., 1997, Seismic properties of carbonate rocks, *in* Carbonate Seismology: I. Palaz and K. J. Marfurt, ed., volume **6** *of* SEG Geophysical Developments, 29–52.

[33] Weger, R., G. P. Eberli, G. Baechle, J. L. Masaferro, and Y. F. Sun, 2009, Quantification of pore structure and its effect on sonic velocity and permeability

in carbonates: The Leading Edge, **93**, 1297–1317.

[34] Zeng, H., R. Loucks, X. Janson, G. Wang, Y. Xia, B. Yuan, and L. Xu, 2011a, Three-dimensional seismic geomorphology and analysis of the Ordovician paleokarst drainage system in the central Tabei Uplift, northern Tarim Basin, western China: AAPG Bulletin, **95**, no. 12, 2061–2083.

[35] Zeng, H., G. Wang, X. Janson, R. Loucks, Y. Xu, and B. Yuan, 2011b, Case History Characterizing seismic bright spots in deeply buried, Ordovician Paleokarst strata, Central Tabei uplift, Tarim Basin, Western China: Geophysics, **76**, no. 4, B127–B137.

[36] Zhang, Y., and J. Sun, 2009, Practical issues in reverse time migration: true amplitude gathers, noise removal and harmonic source encoding: First Break, **26**, 29–35.

Random noise attenuation by f-x empirical mode decomposition predictive filtering

*Yangkang Chen[1] and Jitao Ma[2]**

ABSTRACT

Random noise attenuation always plays an important role in seismic data processing. One of the most widely used methods for suppressing random noise is $f-x$ predictive filtering. When the subsurface structure becomes complex, this method suffers from higher prediction errors owing to the large number of different dip components that need to be predicted. In this paper, we propose a novel denoising method termed $f-x$ empirical mode decomposition predictive filtering (EMDPF). This new scheme solves the problem that makes $f-x$ empirical mode decomposition (EMD) ineffective with complex seismic data. Also, by making the prediction more precise, the new scheme removes the limitation of conventional $f-x$ predictive filtering when dealing with multi-dip seismic profiles. In this new method, we first apply EMD to each frequency slice in the $f-x$ domain and obtain several intrinsic mode functions (IMF). Then an autoregressive (AR) model is applied to the sum of the first few IMFs, which contain the high-dip-angle components, in order to predict the useful steeper events. Finally, the predicted events are added to the sum of the remaining IMFs. This process improves the prediction precision by utilizing an EMD based dip filter to reduce the dip components before $f-x$ predictive filtering. Both synthetic and real data sets demonstrate the performance of our proposed method in preserving more useful energy.

INTRODUCTION

Further development of exploration and production of reservoirs increases the demand for random noise suppression. Current random noise attenuation methods are realized in either the $t-x$ or a transformed domain (17). In the $t-x$ domain, denoising methods include stacking (19; 26; 15), polynomial fitting (28; 14), and median filtering (13). All these methods fully utilize the differences of both travel time and apparent velocity between signal and noise in the $t-x$ domain. In transformed domains, denoising methods include $f-x$ predictive filtering (4), the wavelet transform (27; 6), the curvelet transform (20), and the seislet transform (16; 5). These methods transform the seismic data from the $t-x$ domain to some other domain, where the

***e-mail:** sergey.fomel@beg.utexas.edu

signal and random noise can be separated. The noise is removed in the transformed domain prior to transformation back to the $t-x$ domain.

(4) first used $f-x$ predictive filtering to attenuate random noise. Since then, continuous efforts have been made to improve the predictive precision or to modify the conventional version to meet better the requirements set by various applications. (7) proposed $f-xy$ predictive filtering in order to improve the adaptation to both 2D and 3D post-stack seismic data processing. (22) suggested $f-xyz$ predictive filtering, which is mainly used for random noise attenuation in a pre-stack data set. (10) proposed a $f-x$ quasi-linear transform method adapted to the non-linear events of seismic data in complex regions. Unfortunately, when the subsurface is extremely complex, $f-x$ predictive filtering does not yield good results because of the large number of dip components that need to be predicted.

(9) proposed a new signal processing method, which uses empirical mode decomposition (EMD) to prepare stable input for the Hilbert Transform. The essence of EMD is to stabilize a non-stationary signal. That is, to decompose a signal into a series of intrinsic mode functions (IMF). Each IMF has a relatively local-constant frequency. The frequency of each IMF decreases according to the separation sequence of each IMF. EMD is a breakthrough in the analysis of linear and stable spectra. It adaptively separates non-linear and non-stationary signals, which are features of seismic data, into different frequency ranges. (2) applied $f-x$ EMD to attenuation of random and coherent noise, with good results. (3) suggested using $t-f-x$ EMD to denoise seismic data on the basis of a mixed time-frequency analysis. Nevertheless, for the purpose of random noise attenuation, the $f-x$ and $t-f-x$ domain EMD methods can only be applied on NMO corrected or post-stack seismic data. With profiles containing dipping events, these methods will suppress some of the useful energy.

In this paper, we propose a new approach , termed $f-x$ *empirical mode decomposition predictive filtering* (EMDPF), which combines both $f-x$ EMD and $f-x$ predictive filtering. This new noise attenuation methodology can adapt to more complex seismic profiles than $f-x$ EMD and preserve more useful energy than $f-x$ predictive filtering. $f-x$ EMDPF uses an EMD based dip filter to reduce the dip components for the subsequent predictive filtering in order to improve the predictive precision.

We start this paper by reviewing the conventional $f-x$ predictive filtering theory and point out its high-prediction-error problem when the number of dip components becomes large. Then we review basic EMD theory and its application, both in data processing and the exploration geophysical fields. Finally, we suggest a way to combine the properties of both $f-x$ predictive filtering and $f-x$ EMD in order to form the new denoising algorithm, $f-x$ EMDPF. Three synthetic data sets and one real data set demonstrate that $f-x$ EMDPF can preserve much more useful energy while removing slightly less random noise than $f-x$ EMD and $f-x$ predictive filtering.

F-X PREDICTIVE FILTERING

Let $s(t,h)(h=1,2,\cdots,H)$ be the signal of trace h and H be the number of traces. If the slope of a linear event with constant amplitude in a seismic section is ψ, then:

$$s(t,h+1)=s(t-h\psi\Delta x,1), \tag{1}$$

where Δx denotes the trace interval. Equation 1 can be transformed into the frequency domain in order to give:

$$S(f,h+1)=S(f,1)e^{-i2\pi fh\psi\Delta x}. \tag{2}$$

For a specific frequency f_0, from equation 2 we can obtain a linear recursion, which is given by:

$$S(f_0,h+1)=a(f_0,1)S(f_0,h), \tag{3}$$

where $a(f_0,1)=e^{-i2\pi f_0\psi\Delta x}$. This recursion is a first-order difference equation, also known as an auto-regressive (AR) model of order 1. Similarly, superposition of p linear events in the $t-x$ domain can be represented by an AR model of order p (24; 8) as the following equation:

$$S(f_0,h+1)=a(f_0,1)S(f_0,h)+a(f_0,2)S(f_0,h-1)+\cdots+a(f_0,p)S(f_0,h+1-p), \tag{4}$$

where $a(f_0,h)(h=1,2,\cdots,p)$ denotes the predictive error filter, with a length of p. The prediction error energy $E(f_0)$ is given by the following equation:

$$E(f_0)=\|a(f_0,h)*S(f_0,h)-S(f_0,h+1)\|_2^2, \tag{5}$$

where symbol $*$ denotes convolution, and $\|\cdot\|_2^2$ denotes the least-squares energy. By minimizing the prediction error energy $E(f_0)$, we can get the filtering operator $a(f_0,m)$. Applying this operator to the spatial trace yields the denoised results for the frequency slice f_0.

$f-x$ predictive filtering works perfectly on a single event. Figures 1(a)-1(c) show and compare the denoised results for a single flat synthetic event. The denoised result (Figure 1(b)) is quite good, with the random noise in Figure 1(a) largely removed and only a small amount of the useful component in the noise section, Figure 1(c). For a single dipping event, Figures 1(d)-1(f), the results are similar. However, when the number of different dips is increased, the seismic section becomes more complex and predictive filtering is not as effective. Figure 1(j) shows a synthetic section containing four events with differing dips. In the removed noise section, Figure 1(i), there remains a significant amount of residual useful energy.

The synthetic data shown in Figures 1(a), 1(d), and 1(j) were all generated by SeismicLab (21), with a signal-to-noise ratio (SNR) of 2.0 for all of them. Here we define the SNR as the ratio of maximum amplitude of useful energy and the maximum amplitude of Gaussian white noise. Note that the same parameters were used for the predictive filters in each case shown in Figure 1.

We now conclude that the effectiveness of $f-x$ predictive filtering deteriorates as the number of different dips increases, mainly because the total of leaked useful energy increases at the same time. In particular, when the number of dips is extremely large, as occurs with hyperbolic events, $f-x$ predictive filtering fails to achieve acceptable results. It is natural to infer that if we can first reduce the number of dips, or in other words pick the very steep events and total random noise out, then by applying the same $f-x$ predictive filtering, the predictive precision will improve. That is the subject of the section on $f-x$ empirical mode decomposition predictive filtering.

EMPIRICAL MODE DECOMPOSITION

1D EMD

The aim of empirical mode decomposition (EMD) is to empirically decompose a non-stationary signal into a finite set of subsignals, which are termed intrinsic mode functions (IMF) and are considered to be stable. The IMFs satisfy two conditions: (1) in the whole data set, the number of extrema and the number of zero crossings must either equal or differ at most by one; and (2) at any point, the mean value of the envelope defined by the local maxima and the envelope defined by the local minima is zero (9).

Provided that $s(t)$, $c_n(t)$, $r(t)$, and N denote the original non-stationary signal, the separated IMFs, the residual, and the number of IMFs, respectively, the mathematical principle of EMD can be expressed as:

$$s(t) = \sum_{n=1}^{N} c_n(t) + r(t). \tag{6}$$

For a non-stationary signal $s(t)$, using equation 6, we get a finite set of subsignals $c_n(t)$,$(n = 1, 2, \cdots, N)$.

A special property of EMD is that the IMFs represent different oscillations embedded in the data, where the oscillating frequency for each subsignal $c_n(t)$ decreases as the sequence number of the IMF becomes larger (we call it a frequency decreasing property in the following context). This property results from the sifting algorithm used to implement the decomposition. Appendix A gives a detailed instruction about the sifting process, which can be summarized as a process in which low-frequency components are gradually removed to generate a more local-constant-frequency mode, which is followed by the generation of the next mode.

Figure 2 gives a demonstration for a synthetic signal. The original synthetic signal is generated through $d(t) = \sin(0.2\pi t) + \sin(0.4\pi t) + \sin(0.8\pi t)$; in other words, it is constructed from three individual frequency components corresponding to 0.1 Hz, 0.2 Hz and 0.4 Hz, respectively. From Figure 2, we can see that, except for small edge imprecision and negligible residual, EMD successfully decompose this signal into three components with a frequency ratio of approximately 4:2:1.

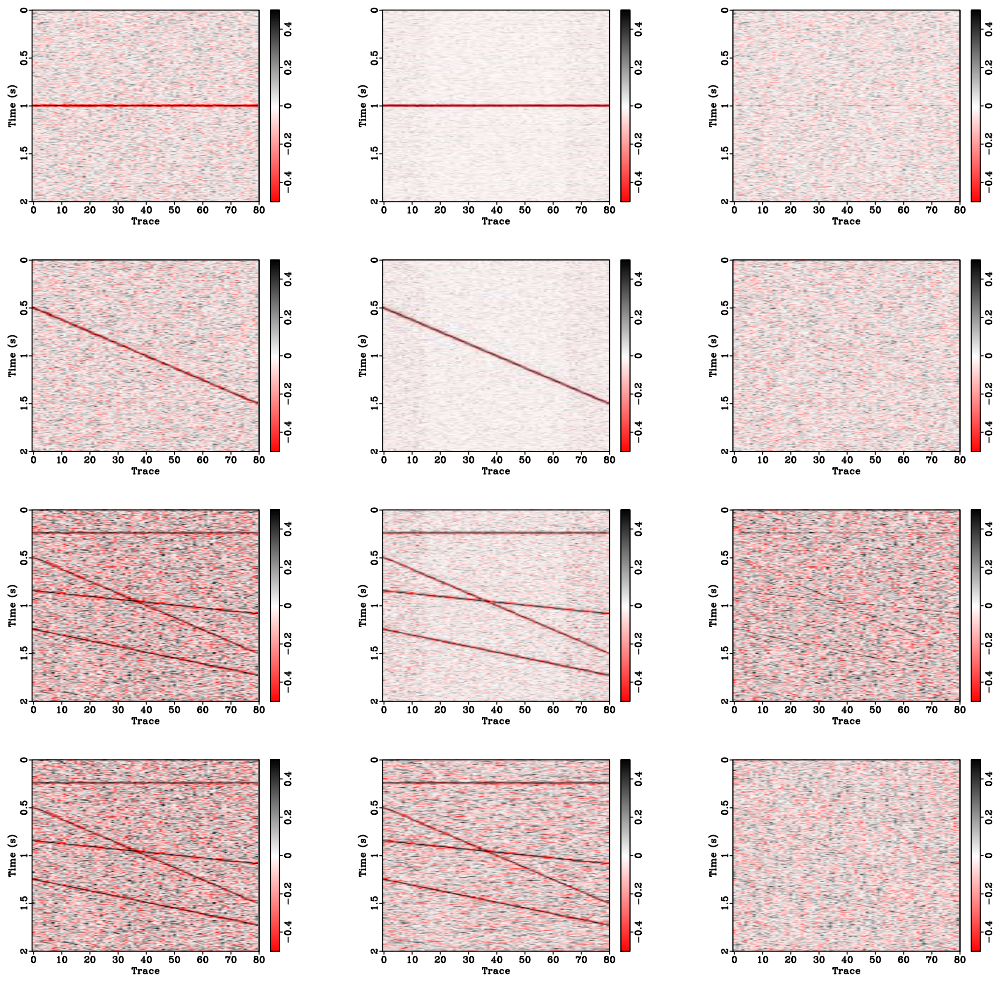


Figure 1: Demonstration of $f-x$ predictive filtering (a-i) and $f-x$ EMDPF (j-l) on synthetic section with different number of dip components. (a) Single flat event. (b) Denoised single flat event. (c) Removed noise section corresponding to (a) and (b). (d) Single dipping event. (e) Denoised single dipping event. (f) Removed noise section corresponding to (d) and (e). (g) Complex events section. (h) Denoised complex events section. (i) Removed noise section corresponding to (g) and (h). (j) Same as (g). (k) Denoised result by $f-x$ EMDPF. (l) Removed noise section corresponding to (j) and (k).

emdpf/fxdecon syn01-flat,syn01-flat-fxdecon,syn01-flat-fxdecon-noise,syn01-dip,syn01-dip-fxdecon,syn

Because of the frequency decreasing property, EMD has been used outside geophysics for noise attenuation (18; 12). Since random noise represents mainly the high-frequency components, by removing the IMFs with the highest frequency, we can attenuate this type of noise. However, in exploration geophysics, applying EMD to time traces is not effective because of the mode mixing problem. (11) defined mode mixing as any IMF consisting of frequencies of dramatically disparate scales. When mode mixing exists, the first one or two IMFs contain a lot of useful reflection energy. Extensions to EMD, such as ensemble empirical mode decomposition (EEMD) (25) and complete ensemble empirical mode decomposition (CEEMD) (23) have been proposed to solve the mode-mixing problem in signal processing and have been used in geophysics to analyze time-frequency properties, but have not been used for t-x domain seismic noise attenuation.

f-x EMD

Instead of $t-x$ domain EMD, a $f-x$ domain EMD method to attenuate random noise in seismic data has been proposed by (2). They apply EMD on each frequency slice in the $f-x$ domain, and suppress the higher wavenumber components, which mainly represent random noise. However, a problem occurs when applying $f-x$ EMD, because the high-wavenumber dipping events will also be removed. This problem occurs because, for many data sets, the random noise and any steeply dipping coherent energy make a significantly larger contribution to the high-wavenumber energy in the f-x domain than any desired signal (2).

(2) cleverly utilize this by-product of $f-x$ EMD to attenuate coherent noise such as ground roll.

The detailed algorithmic steps of $f-x$ EMD are given by (2) as:

1. Select a time window and transform the data to the $f-x$ domain.
2. For every frequency,
   (a) separate real and imaginary parts in the spatial sequence,
   (b) compute IMF1, for the real signal and subtract it to obtain the filtered real signal,
   (c) repeat for the imaginary part,
   (d) combine to create the filtered complex signal.
3. Transform data back to the $t-x$ domain.
4. Repeat for the next time window.

$f-x$ EMD can be used as an adaptive $f-k$ filter. The cutoff wavenumber is adaptively defined and does not need any apriori knowledge about the seismic data

in order to define the filter parameters. This adaptability makes $f-x$ EMD very convenient to utilize in real applications. The frequency-slice-dependent adaptability also makes $f-x$ EMD more precise than $f-x$ predictive filtering, because all the filter parameters in $f-x$ predictive filtering for each frequency slice are the same. Another advantage of $f-x$ EMD over $f-x$ predictive filtering is that the trace spacing does not need to be perfectly regular because no convolutional operator is used, a characteristic similar to local median and SVD filtering (1; 2).

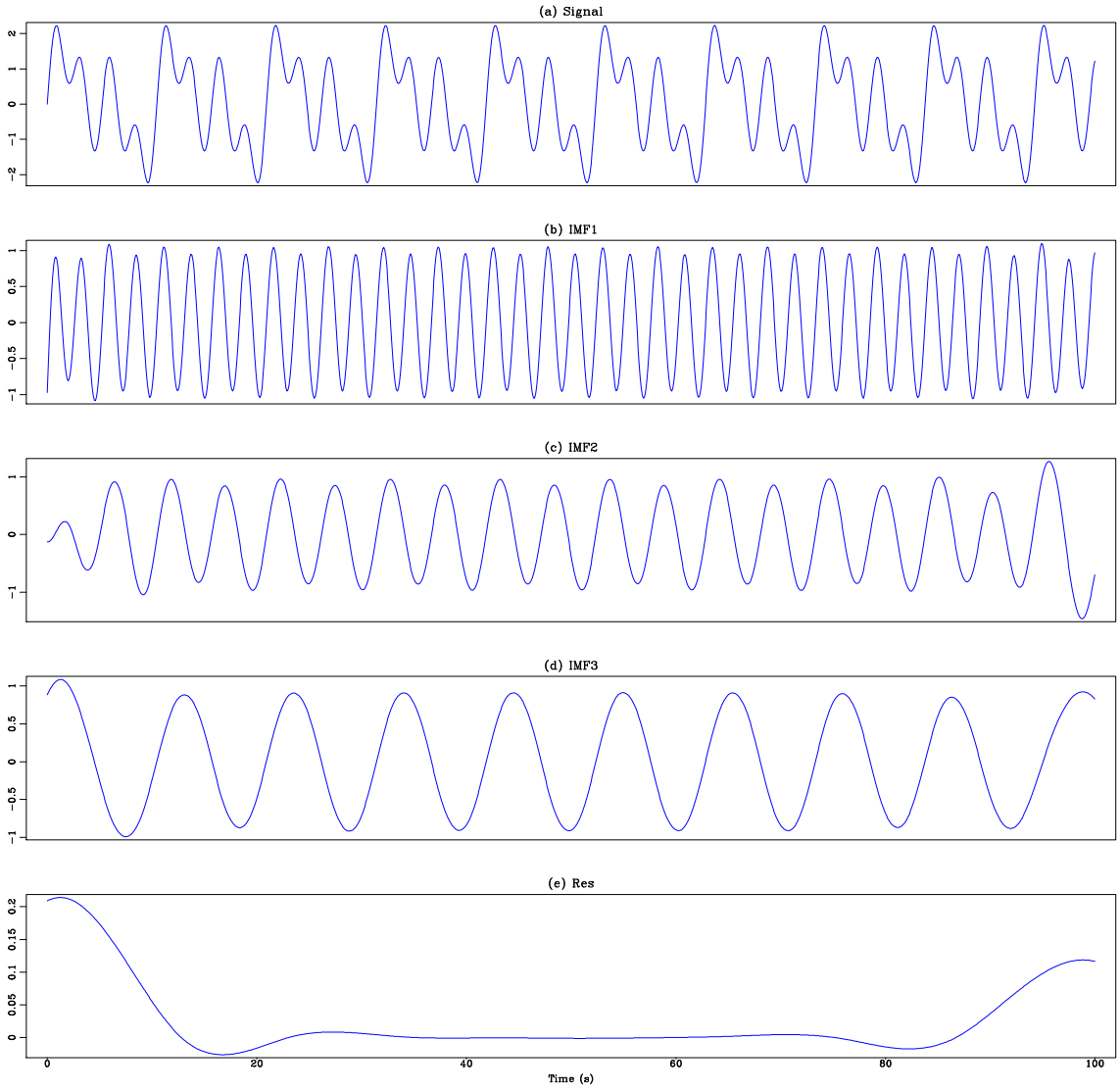


Figure 2: Demonstration of empirical mode decomposition on a synthetic signal. (a) The original signal, (b) first IMF, (c) second IMF, (d) third IMF, (e) residual. emdpf/synsig sigimf

F-X EMPIRICAL MODE DECOMPOSITION PREDICTIVE FILTERING

$f-x$ EMDPF utilizes the property that the first few (generally $1 \sim 3$) IMFs for each frequency slice in the $f-x$ domain contain the high-dip-angle components and random noise. Thus the leaked dipping events can be obtained by applying a predictive filter to these IMFs. Adding the predicted signal to the sum of the remaining IMFs will suppress random noise without harming the effective signals.

$F-x$ EMDPF is a new seismic noise attenuation method which combines the advantages of both $f-x$ predictive filtering and $f-x$ EMD. The detailed algorithmic steps of $f-x$ EMDPF are similar to $f-x$ EMD (2) and are shown below:

1. Select a time window and transform the data to the $f-x$ domain.
2. For every frequency,
   (a) separate real and imaginary parts in the spatial sequence,
   (b) compute IMF1, for the real signal and subtract it to obtain the filtered real signal,
   (c) apply an AR model to IMF1 and add the result to the sum of the remaining IMFs,
   (d) repeat for the imaginary part,
   (e) combine to create the filtered complex signal.
3. Transform data back to the $t-x$ domain.
4. Repeat for the next time window.

It should be emphasized that the number of the filtered IMFs is not limited to one, but is selected according to both the noise level and the distribution of the dip components within the specific seismic data set. If the noise level is high, then a larger number of IMFs should be chosen, because the noise remains not only in the first IMF but also in the second or the third, albeit with decreasing energy. If the dip components are mainly distributed in the high-angle range, then the number of IMFs could be relatively smaller, but when the dip components are distributed in the low- or mid-angle range, we should choose more IMFs in order to ensure that noise is removed whilst still preserving these dipping components.

Generally the number of IMFs for filtering is within the range of $1 \sim 3$. In conventional EMD, the signal is completely decomposed into all the IMFs, along with the remainder. In our proposed algorithm, EMD decomposes a signal into only $1 \sim 3$ components, which correspond to the number of IMFs to be filtered. Compared with the conventional EMD, this uncompleted decomposition algorithm can improve the computation efficiency by about 5 times.

EMD BASED DIP FILTER

In this section, we seek to connect $f-x$ EMDPF with $f-x$ predictive filtering and $f-x$ EMD. We would like to first introduce the so-called EMD based dip filter.

Since the frequency of each IMF decreases according to the order in which it is separated out, by subtracting the first few IMFs of each frequency slice in the $f-x$ domain, we extract the higher wavenumber components, which represent the energy of random noise and high-dip-angle events in seismic sections.

If we divide the set of IMFs into more-detailed zones, we can separate the section into several dip bands. Thus, we reach the definition of the EMD based dip filter:

$$\tilde{u}_i(f,h) = \begin{cases} \epsilon_1 u_i(f,h) & i \in D_1 \\ \epsilon_2 u_i(f,h) & i \in D_2 \\ \vdots & \vdots \\ \epsilon_m u_i(f,h) & i \in D_m \end{cases}, \tag{7}$$

$$\Lambda(f,h) = \sum_{i=1}^{N} \tilde{u}_i(f,h), \tag{8}$$

where $\Lambda(f,h)$ is the filtered data for frequency slice f in the $f-x$ domain. $u_i(f,h)(i=1,2,\cdots,N)$ is the ith separated IMF such that $S(f,h)=\sum_{i=1}^{N} u_i(f,h)$, where $S(f,h)$ is the transformed $f-x$ domain seismic data. $D_i(i=1,2,\cdots,m)$ is the ith of m, the number of dip bands, and ϵ_i is the corresponding weighting coefficient. For a simple high-pass dip filter, we choose $m=2$, $\epsilon_1=1$, $\epsilon_2=0$ and $D_1=\{1,2\}$, $D_2=\{3,4,\cdots,N\}$.

Figure 3 demonstrates how an EMD based dip filter works on a synthetic plane-wave seismic profile containing three events corresponding to three dips. After filtering with high-pass, mid-pass, and low-pass dip filters, respectively, the three plane waves are successfully separated . The parameters we choose in designing these three filters are shown in Table 1.

Type	N	m	ϵ	D
high-pass	10	2	$\epsilon_1=1,\epsilon_2=0$	$D_1=\{1\},D_2=\{2,3,\cdots,10\}$
mid-pass	10	3	$\epsilon_1=0,\epsilon_2=1,\epsilon_3=0$	$D_1=\{1\},D_2=\{2\},D_3=\{3,4,\cdots,10\}$
low-pass	10	2	$\epsilon_1=0,\epsilon_2=1$	$D_1=\{1\},D_2=\{2,3,\cdots,10\}$

Table 1: Parameters in designing high-pass, mid-pass and low-pass dip filters corresponding to Figure 3.

The EMD based dip filter is defined adaptively since the filtering process is data driven. We only need to define the number of IMFs contained in each dip band at the start, a step that is convenient to implement.

If we consider random noise as high-dip-angle components, then $f-x$ EMD denoising (2) is equivalent to applying a high-cut dip filter with the form of equation 7 to the seismic data in order to remove both random noise and ground roll. We can also understand $f-x$ EMDPF from EMD based dip filter. $f-x$ EMDPF first uses an EMD based dip filter to separate the high-dip-angle and low-dip-angle components, where the high-dip-angle components are composed of steeply dipping useful events and noise, and low-dip-angle components are all useful signals. The useful signal in the high-dip-angle components are predicted and subsequently restored. Due to the effects of the EMD based dip filter, a decrease occurs in the number of useful signal components that needs to be predicted. This decrease results in more accurate overall performance when compared with conventional predictive filtering.

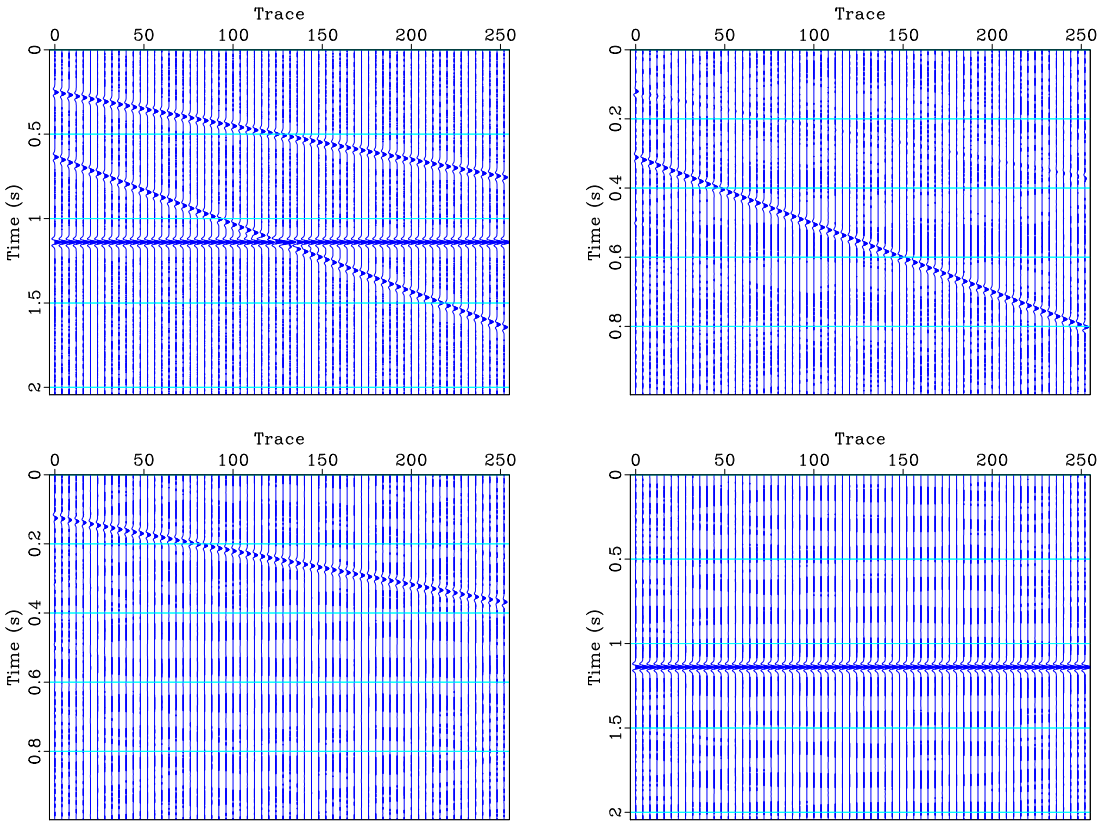


Figure 3: Demonstration of EMD based dip filter. (a) Original synthetic profile, (b) with high-pass dip filter, (c) with mid-pass dip filter, (d) with low-pass dip filter. emdpf/dipfilter plane,planeemd1,planeemd2,planeemd3

EXAMPLES

In this section, we first reuse the previously discussed synthetic data (Figure 1(j)), then show two other synthetic data and one field data example to demonstrate the performance of $f-x$ EMDPF.

Figure 1(k) shows the denoised result after $f-x$ EMDPF. The removed noise section is shown in Figure 1(l). Comparing Figure 1(i) and Figure 1(l), we see that the useful events leaked into Figure 1(i) have been returned to the denoised result (Figure 1(k)), while the noise level stays nearly unchanged.

The second synthetic example is composed of one linear dipping event and two flat events (Figure 4). A 40 Hz Ricker wavelet has been used with a time sample interval of 4 ms. The number of samples for each trace is 501, and the number of traces is 120. Figures 4(d), 4(e), and 4(f) illustrate the comparison of denoised results of the synthetic data using $f-x$ predictive filtering, $f-x$ EMD filtering, and $f-x$ EMDPF, respectively. Figure 4(a) is the noise free data, Figure 4(b) is Gaussian white noise, and Figure 4(c) is the noisy data. The SNR of the noisy data is 2.0 (using the previous definition of SNR). Figures 4(g), 4(h), and 4(i) show the removed noise sections corresponding to $f-x$ predictive filtering, $f-x$ EMD, and $f-x$ EMDPF, respectively. From Figure 4(g), we can see that $f-x$ predictive filtering harms both flat and dipping events to some extent. Although by increasing the length of the predictive step we can decrease the damage done to the signals, the noise suppression is less effective because of the stronger prediction of noise. Also, as seen in Figure 4(d), the $f-x$ predictive filtering introduces some artefacts. In Figure 4(h), we see that $f-x$ EMD tends to harm much of the dip energy but preserves entirely the flat events. Using the same predictive filtering parameters, we see from Figure 4(i) that both flat and dipping signals are hardly affected when $f-x$ EMDPF is applied. In this example, the first IMF is removed for prediction in the process of $f-x$ EMDPF. Figure 5 demonstrates the sensitivity of the $f-x$ EMDPF for increasing the number of filtered IMFs. We can see that, as the number of filtered IMFs increases, the denoising result becomes more similar to $f-x$ predictive filtering; that is, more noise is removed and more obvious artefacts appear in the denoised section. However, for $f-x$ EMDPF, the horizontal events are always totally preserved, which supposes a generally better denoising result than $f-x$ predictive filtering.

The third synthetic example is a benchmark data set from SeismicLab. The central frequency of the Ricker wavelet is 40 Hz and the temporal sampling is 2 ms. The number of time samples is 750 and the number of spatial samples is 50. Figures 6(a), 6(b), and 6(c) denote the clean data, noise section, and noisy section, respectively. The SNR in Figure 6(c) is 2.0. Figures 6(d), 6(e), and 6(f) are the denoised results using $f-x$ predictive filtering, $f-x$ EMD, and $f-x$ EMDPF, respectively. From removed noise sections Figures 6(g), 6(h), and 6(i) we can conclude that $f-x$ predictive filtering harms much useful energy when the number of dip components increases, whereas $f-x$ EMD affects most of the dipping events, and $f-x$ EMDPF preserves the useful energy to the greatest extent while removing the slightly weaker level of noise. In this example, the first IMF is removed for prediction in the process of $f-x$ EMDPF.

The field data is shown in Figure 7. It is a stacked section without migration from the South China Sea. Figure 7(b) is a zoomed portion of Figure 7(a) from 1.5s to 3.0s. The denoised profiles, shown in Figures 8(a), 8(c), and 8(e), demonstrate $f-x$

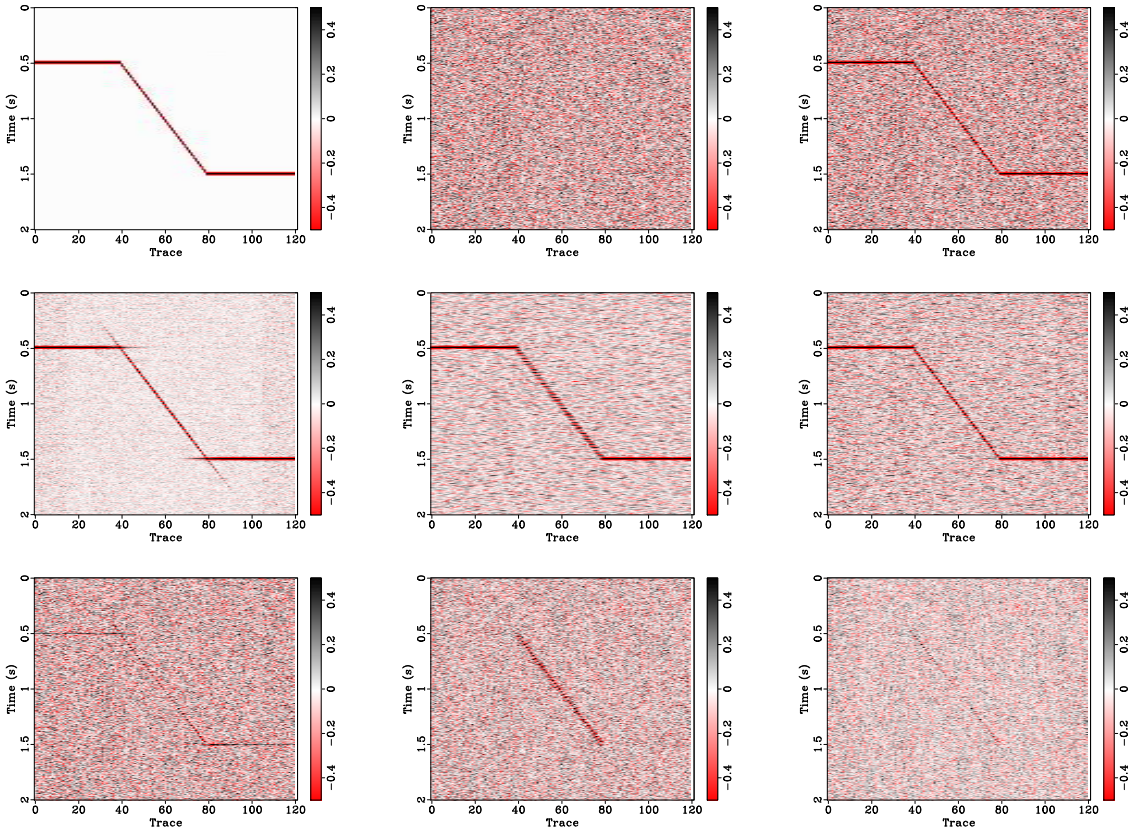


Figure 4: Comparison of denoising effects. (a) Clean data. (b) Gaussian white noise. (c) Noisy data. (d) Denoised result by $f-x$ predictive filtering. (e) Denoised result by $f-x$ EMD. (f) Denoised result by $f-x$ EMDPF. (g) Removed noise section corresponding to (d). (h) Removed noise section corresponding to (e). (i) Removed noise section corresponding to (f).

emdpf/linear syn02-clean,syn02-noise,syn02,syn02-fxdecon,syn02-fxemd,syn02-fxemdpf,syn02-fxdecon-

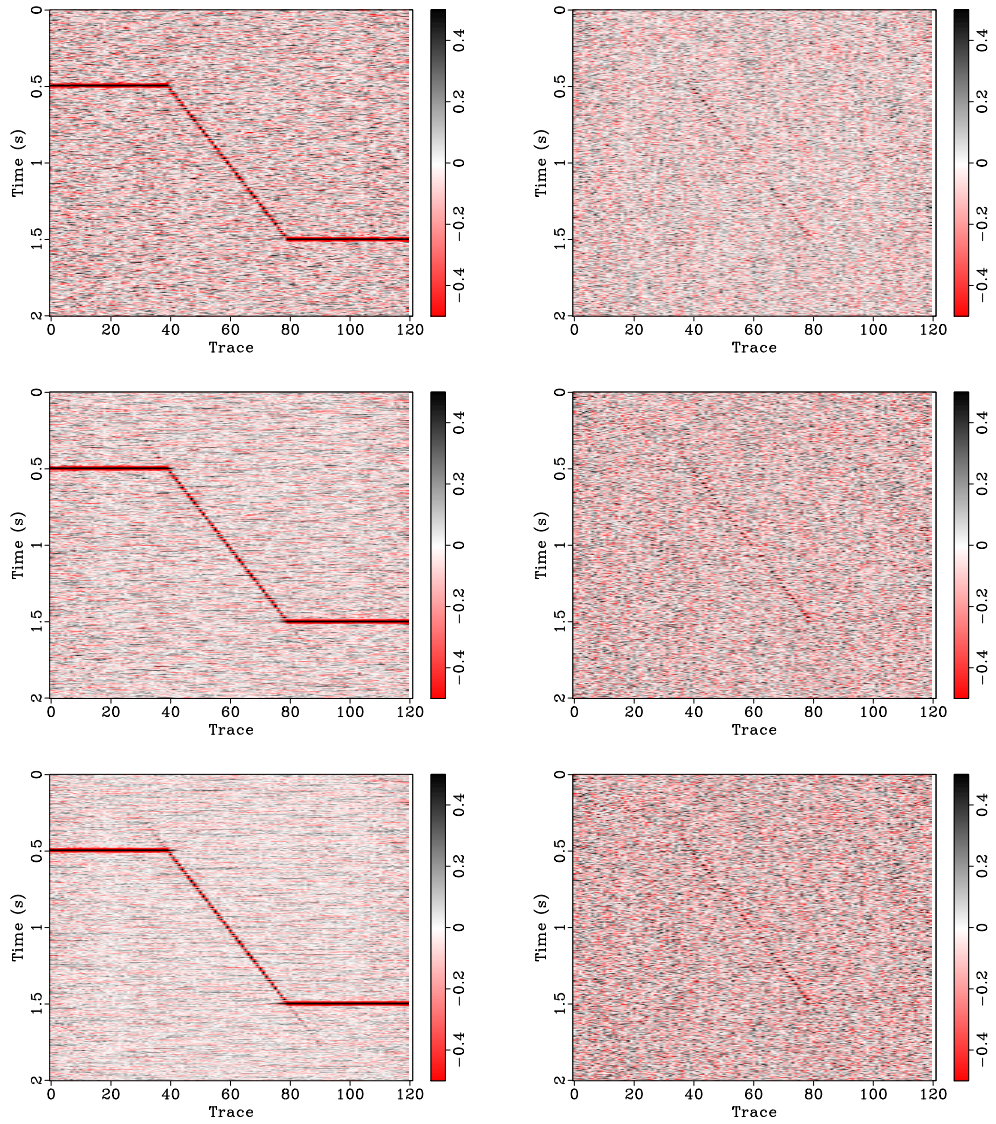


Figure 5: Comparison of denoising effects. (a) $f-x$ EMDPF denoised result with prediction on 1 IMF. (b) Noise section corresponding to (a). (c) $f-x$ EMDPF denoised result with prediction on 2 IMFs. (d) Noise section corresponding to (c). (e) $f-x$ EMDPF denoised result with prediction on 3 IMFs. (f) Noise section corresponding to (e).

emdpf/linear syn02-fxemdpf-1imf,syn02-fxemdpf-1imf-noise,syn02-fxemdpf-2imf,syn02-fxemdpf-2imf-n

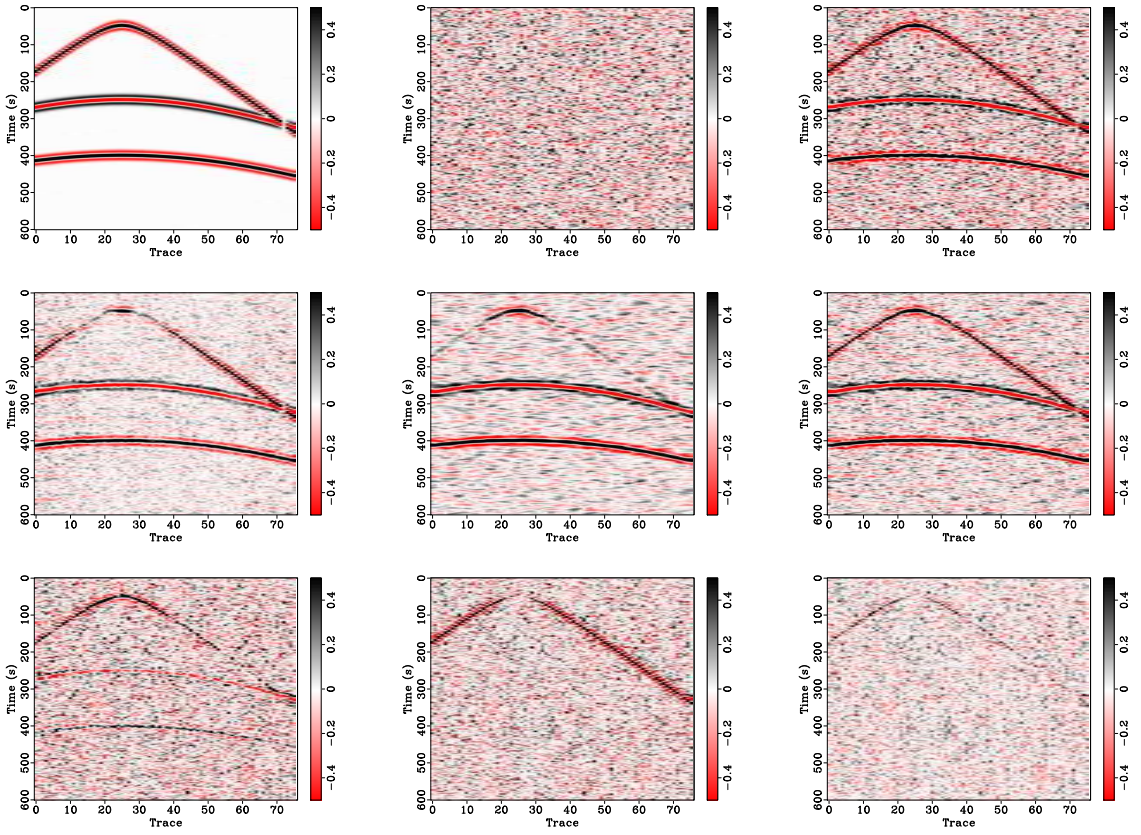


Figure 6: Comparison of denoising effects. (a) Clean data. (b) Gaussian white noise. (c) Noisy data. (d) Denoised result by $f-x$ predictive filtering. (e) Denoised result by $f-x$ EMD. (f) Denoised result by $f-x$ EMDPF. (g) Removed noise section corresponding to (d). (h) Removed noise section corresponding to (e). (i) Removed noise section corresponding to (f).

emdpf/hyper syn03-clean,syn03-noise,syn03,syn03-fxdecon,syn03-fxemd,syn03-fxemdpf,syn03-fxdecon-

predictive filtering, $f-x$ EMD, and $f-x$ EMDPF, respectively. The corresponding noise sections are shown in Figures 8(b), 8(d), and 8(f). From these noise sections, we see clearly that $f-x$ EMD removes many dipping events. Even though we can't see clearly the improvement after applying $f-x$ EMDPF at the scale of Figure 8, the improvement can be identified on the zoomed noise sections shown in Figure 9. The useful energy shown around 1.75s, 2.6s in Figure 9(b) does not exist in the same part of Figure 9(f). From these differences, we conclude that $f-x$ EMDPF is more satisfactory than $f-x$ predictive filtering and $f-x$ EMD, in that it leaves less useful energy in the noise section. In this real data example, we apply the AR model on the first three IMFs for $f-x$ EMDPF. For display reasons, the noise sections have been amplified by 3 times.

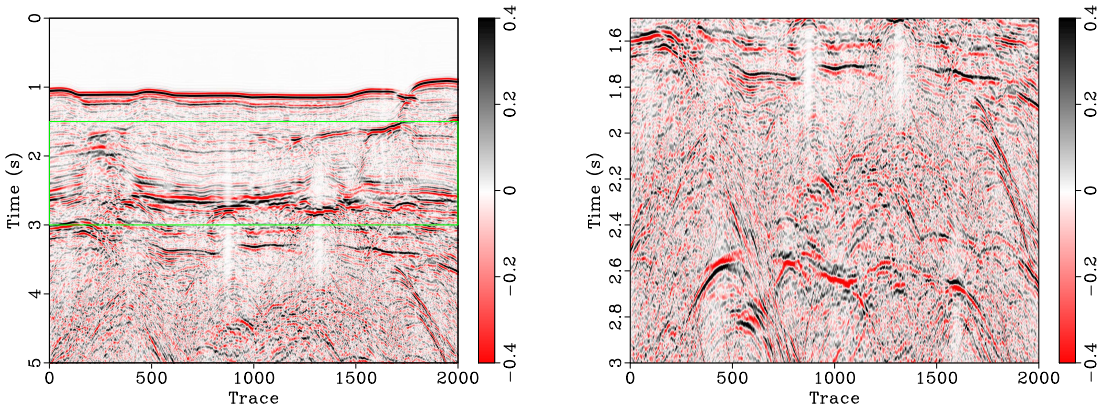


Figure 7: Field data from the South China Sea. (a) Original post-stack pre-migration profile. (b) Temporal zoomed part from 1.5s to 3.0s. emdpf/southsea southsea,southsea-zoom

CONCLUSIONS

We have proposed a new denoising method suitable for complex subsurface structures. We demonstrate that the number of dipping events will affect the denoising performance of $f-x$ predictive filtering. We also give the definition of an EMD based dip filter and ascribe the effectiveness of $f-x$ EMD to applying a high-cut EMD based dip filter to seismic profiles.

By using the AR model to predict the steeply dipping event, $f-x$ EMDPF can deal with complex seismic profiles that conventional $f-x$ EMD can't handle. By applying an EMD based adaptive dip filter in advance, $f-x$ EMDPF can preserve more useful energy as compared with conventional $f-x$ predictive filtering. $f-x$ EMDPF is actually a modification to both $f-x$ predictive filtering and $f-x$ EMD, so it maintains the benefits of being convenient, data driven, whilst combining the dip-selection property of EMD with the power of the AR model used in $f-x$ predictive filtering.

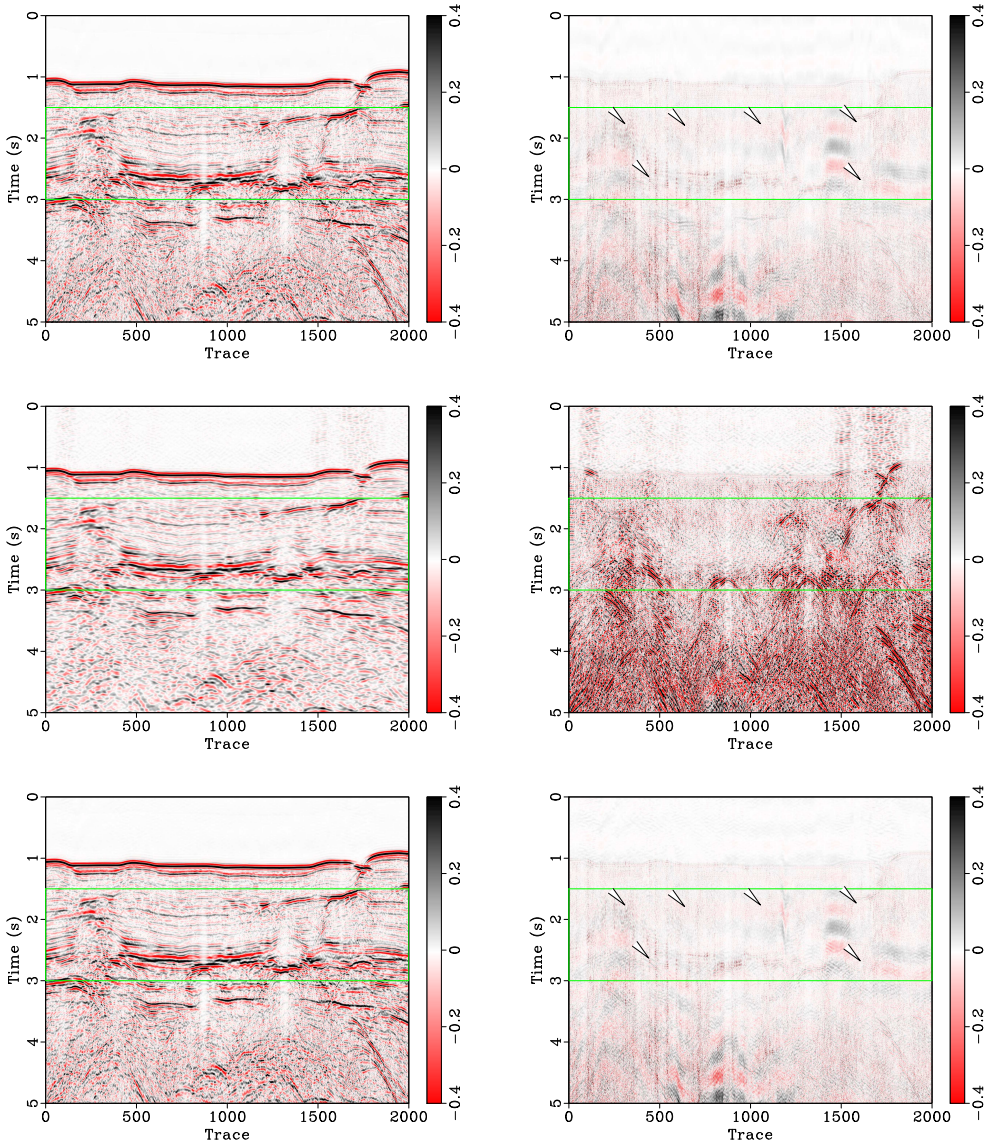


Figure 8: Comparisons between denoised results and corresponding noise sections. (a) Denoised result by $f-x$ predictive filtering. (b) Removed noise section by $f-x$ predictive filtering ($\times 3$). (c) Denoised result by $f-x$ EMD. (d) Removed noise section by $f-x$ EMD ($\times 3$). (e) Denoised result by $f-x$ EMDPF. (d) Removed noise section by $f-x$ EMDPF ($\times 3$).

emdpf/southsea southsea-fxdecon,southsea-fxdecon-noise,southsea-fxemd,southsea-fxemd-noise,southsea-fxemdpf,southsea-fxemdpf-noise

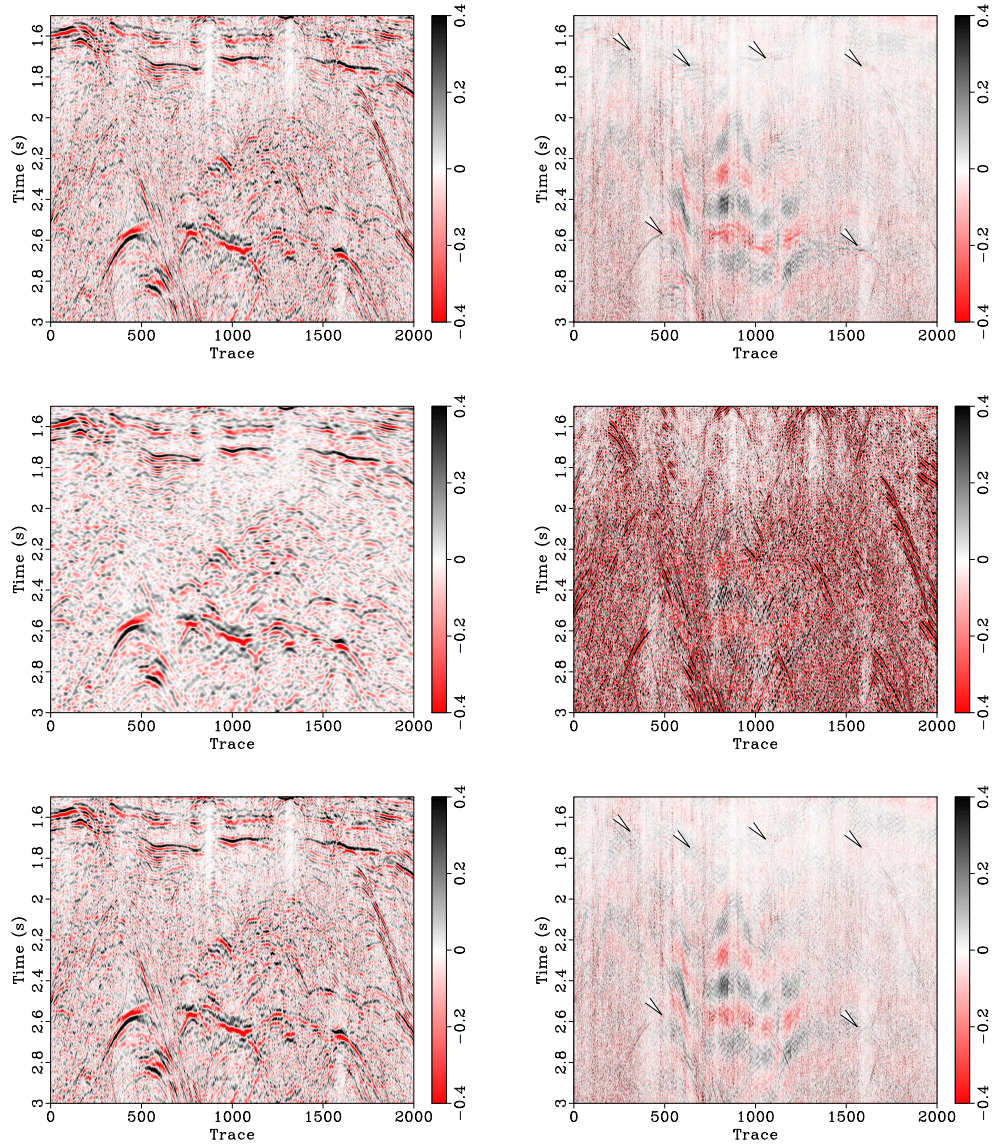


Figure 9: Zoomed part of denoised results and corresponding noise sections. (a) Denoised result by $f-x$ predictive filtering. (b) Removed noise section by $f-x$ predictive filtering ($\times 3$). (c) Denoised result by $f-x$ EMD. (d) Removed noise section by $f-x$ EMD ($\times 3$). (e) Denoised result by $f-x$ EMDPF. (d) Removed noise section by $f-x$ EMDPF ($\times 3$).

emdpf/southsea southsea-fxdecon-zoom,southsea-fxdecon-noise-zoom1,southsea-fxemd-zoom,southsea

Although the incomplete EMD described in this paper can improve computational efficiency, a great deal of time is still required to process the data. Currently, this time requirement is the major drawback of the approach. In addition, continued research is required in order to find an efficient thresholding method in the $f - x$ domain in order to improve the preservation of useful signal.

ACKNOWLEDGEMENTS

We thank Josef Paffenholz, Mauricio Sacchi, Sergey Fomel, and Karl Schleicher for helpful discussions and all the developers of Madagascar and SeismicLab software packages for providing the codes. We also thank the associate editor Danilo Velis and three anonymous reviewers for their constructive suggesions, which helped to improve the paper.

REFERENCES

[1] Bekara, M., and M. van der Baan, 2007, Local singular value decomposition for signal enhancement of seismic data: Geophysics, **72**, V59–V65.

[2] ———, 2009, Random and coherent noise attenuation by empirical mode decomposition: Geophysics, **74**, V89–V98.

[3] Cai, H., Z. He, and D. Huang, 2011, Seismic data denoising based on mixed time-frequency methods: Applied Geophysics, **8**, 319–327.

[4] Canales, L., 1984, Random noise reduction: SEG expanded abstracts:54th Annual international meeting, 525–527.

[5] Fomel, S., and Y. Liu, 2010, Seislet transform and seislet frame: Geophysics, **75**, V25–V38.

[6] Gao, J., J. Mao, W. Chen, and Q. Zheng, 2006, On the denoising method of prestack seismic data in wavelet domain: Chinese J. Geophys, **49**, 1155–1163.

[7] Guo, J., X. Zhou, and H. Yang, 1995, Attenuation of random noise in f-x, y domain: Oil Geophysical Prospecting, **30**, 207–215.

[8] Harris, P., and R. White, 1997, Improving the performance of f-x prediction at low signal-to-noise ratios: Geophysical Prospecting, **45**, 269–302.

[9] Huang, N. E., Z. Shen, S. R. Long, M. C. Wu, H. H. Shih, Q. Zheng, N.-C. Yen, C. C. Tung, and H. H. Liu, 1998, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis: Proceeding of the Royal Society of London Series A, **454**, 903–995.

[10] Kang, Y., C. Yu, W. Jia, and C. Wang, 2003, A study on noise-suppression method in f-x domain: Oil Geophysical Prospecting, **38**, 136–138.

[11] Kopecky, M., 2010, Ensemble empirical mode decomposition: Image data analysis with white-noise reflection: Acta Polytechnica, **50**, 49–56.

[12] Kopsinis, Y., 2009, Development of EMD-based denoising methods inspired by wavelet thresholding: Signal Processing, IEEE Transactions on, **57**, 1351 – 1362.

[13] Liu, C., H. Li, C. Tao, Y. Liu, D. Wang, and C. Gu, 2007, A new fuzzy nesting

multilevel median filter and its application to seismic data processing: Chinese J. Geophysics, **50**, 1534–1542.

[14] Liu, G., X. Chen, J. Du, and J. Song, 2011, Seismic noise attenuation using nonstationary polynomial fitting: Applied Geophysics, **8**, 18–26.

[15] Liu, G., S. Fomel, L. Jin, and X. Chen, 2009a, Stacking seismic data using local correlation: Geophysics, **74**, V43–V48.

[16] Liu, Y., S. Fomel, C. Liu, D. Wang, G. Liu, and X. Feng, 2009b, High-order seislet transform and its application of random noise attenuation: Chinese J. Geophysics, **52**, 2142–2151.

[17] Liu, Z., X. Chen, and J. Li, 2009c, Noncasual spatial prediction filtering based on an ARMA model: Applied Geophysics, **6**, 122–128.

[18] Mao, Y., and P. Que, 2007, Noise suppression and flawdetection of ultrasonic signals via empirical mode decomposition: Rus. J. Nondestruct. Test., **43**, 196–203.

[19] Mayne, W., 1962, Common reflection point horizontal data stacking techniques: Geophysics, **27**, 927–938.

[20] Neelamani, R., A. Baumstein, D. Gillard, M. Hadidi, and W. Soroka, 2008, Coherent and random noise attenuation using the curvelet transform: The Leading Edge, **27**, 240–248.

[21] Sacchi, M. D., 2008, http://seismic-lab.physics.ualberta.ca/: SeismicLab.

[22] Su, G., X. Zhou, and C. Li, 1998, Prediction noise elimination in f-xyz domain: Oil Geophysical Prospecting, **33**, 95–103.

[23] Torres, M. E., M. A. Colominas, G. Schlotthauer, and P. Flandrin, 2011, A complete ensemble empirical mode decomposition with adaptive noise: IEEE International Conference on Acoustics, Speech and Signal Proces- sing (ICASSP), 4144–4147.

[24] Tufts, D., and R. Kumaresan, 1980, Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood: Proceedings of the IEEE, **70**, 975–989.

[25] Wu, Z., and N. E. Huang, 2009, Ensemble empirical mode decomposition: A noise-assisted data analysis method: Advances in Adaptive Data Analysis, **01**, 1–41.

[26] Yilmaz, O., 2001, Seismic data analysis: Processing, inversion, and interpretation of seismic data: Soc. Expl. Geophys.

[27] Zhang, R., and T. Ulrych, 2003, Physical wavelet frame denoising: Geophysics, **68**, 225–231.

[28] Zhong, W., B. Yang, and Z. Zhang, 2006, Research on application of polynomial fitting technique in highly noisy seismic data: Progress in Geophysics, **21**, 184–189.

APPENDIX A: SIFTING ALGORITHM FOR EMPIRICAL MODE DECOMPOSITION

In this appendix, we review the sifting algorithm of empirical mode decomposition (equation 6 in the main paper). For the original signal, we first find the local maxima and minima of the signal. Once identified, fit these local maxima and minima by cubic spline interpolation in turn in order to generate the upper and lower envelopes. Then compute the mean of the upper and lower envelopes m_{11}, the difference between the data and first mean h_{11}.

$$m_{11} = \frac{h_{10}^{+} + h_{10}^{-}}{2}, \tag{A-1}$$

$$h_{11} = h_{10} - m_{11}, \tag{A-2}$$

where h_{ij} denotes the remaining signal after jth sifting for generating the ith IMF, h_{ij}^{+} and h_{ij}^{-} are corresponding upper and lower envelopes, respectively, and m_{ij} is the mean of upper and lower envelopes after jth sifting for generating the ith IMF. Repeating the sifting procedure (A-2) k times, until h_{1k} reach the prerequisites of IMF, these are:

$$h_{1(k-1)} - m_{1k} = h_{1k}. \tag{A-3}$$

The criterion for the sifting process to stop is given by (9) as:

$$0.2 \leq SD = \sum_{t=0}^{T} \left[\frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_{1(k-1)}^2} \right] \leq 0.3, \tag{A-4}$$

where SD denotes the standard deviation. When h_{1k} is considered as an IMF, let $c_1 = h_{1k}$, we separate the first IMF from the original data:

$$d - c_1 = r_1, \tag{A-5}$$

where d is the original signal, c_n denotes the nth IMF, and r_n is the residual after the nth IMF based sifting. Repeating the sifting process from equation A-1 to A-5, changing h_{1j} to h_{ij}, in order to get the following IMFs: $c_2, c_3, \cdots, c_N$. The sifting process can be stopped when the residual r_n, becomes so small that it is less than a predetermined value of substantial consequence, or when r_n becomes a monotonic function from which no more IMF can be extracted.

Finally, we achieved a decomposition of the original data into N modes, and one residual, as shown in equation 6 in the main context.