# Tomographic velocity model estimation using finite-offset kinematic wavefield attributes

*Tilman Kluever*

**email:** *Tilman.Kluever@gpi.uni-karlsruhe.de*

**keywords:** *Tomography, wavefield attributes, velocity model*

**ABSTRACT**

*I present a tomographic inversion scheme that makes use of finite-offset (FO) kinematic wavefront attributes to determine smooth, laterally inhomogeneous, isotropic 2D subsurface velocity models. The kinematic wavefield attributes are associated with rays starting at a common reflection point (CRP) in the subsurface. An optimum model is found iteratively by minimizing the misfit between measured and modeled data in the least-squares sense. The required forward modeled data are obtained during each iteration by dynamic ray tracing. Fréchet derivatives contained in the tomographic matrix are calculated by ray perturbation theory. The tomographic inversion is successfully tested on a synthetic example.*

## INTRODUCTION

A tomographic inversion scheme for the construction of smooth 2D isotropic velocity models for depth imaging based on kinematic wavefield attributes was recently established by Duveneck (2002). The method makes use of the curvature of normal incidence point (NIP) wavefronts, emergence angles of selected zero-offset (ZO) central rays and the corresponding one-way traveltimes. All kinematic wavefield attributes used in this inversion can be extracted in a robust way from seismic prestack data with the Common-Reflection-Surface (CRS) Stack (Mann et al., 1999). An alternative formulation (Duveneck et al., 2003) uses first and second order spatial derivatives of traveltime in place of angles and curvatures.

In this paper the method is extended to the finite-offset (FO) case. Both alternative formulations are explained and their relationship is described.

## KINEMATIC WAVEFIELD ATTRIBUTES

Kinematic wavefield attributes are all kinematic characteristics of a selected wave observable at the surface. The traveltime of a selected wave can be approximated around an observation point on the surface by parabolic and hyperbolic traveltime expansions (Schleicher et al., 1993). They provide the first and second order traveltime derivatives at the observation point. Beside the traveltime $t$ itself this are the wavefield attributes used further in this paper. They can be related to hypothetical experiments yielding the same attributes. In the following only P-waves propagating in isotropic inhomogeneous media are considered. The associated rays are perpendicular to the wavefront. Only primary reflection events are taken into consideration.

In order to introduce the wavefront attributes used in the inversion process I will start with the ZO-case. Then I will extend the ZO case to finite offsets.

### ZO-case

In a ZO-experiment, the central ray associated with a selected wavefront hits the reflecting interface normally. Therefore the point of incidence on the reflector is called normal-incidence-point (NIP). The down-

and up-going ray segments are identical. The kinematic reflection response of a reflector segment around NIP is approximated by the ZO-CRS operator. This operator is expressed by the emergence angle $\beta$ of the central ray and the curvatures of two eigenwaves measured at the surface. The near surface velocity $v$ is assumed to be constant around the observation point. The two eigenwaves are associated with two hypothetical experiments. A point source at NIP provides the so-called NIP wave (Hubral, 1983) emerging on the ground surface with the curvature $K_{NIP}$. An exploding reflector segment around NIP provides the so-called normal wave (Hubral, 1983) emerging at the observation point with the curvature $K_N$. The emergence angle $\beta$ is related to the first-order spatial derivative of the one-way traveltime with respect to the midpoint coordinate $x_m$ as follows:

$$\frac{\partial T}{\partial x_m} = \frac{\sin\beta}{v} \,, \tag{1}$$

where $T = t/2$ denotes the one-way traveltime and $x_m = \frac{1}{2}(x_s + x_G)$ is the location of the midpoint at the measurement surface. The source and receiver locations are denoted by $x_S$ and $x_G$ respectively. The $x$-axis of the coordinate system is assumed to be tangential at the measurement surface in the considered measurement point. The curvature $K_{NIP}$ of the emerging wave due to the point source at NIP is related to the second-order spatial derivative of the one-way traveltime with respect to the half-offset $h$ as follows:

$$\frac{\partial^2 T}{\partial h^2} = \frac{K_{NIP}}{v}\cos\beta^2 \,, \tag{2}$$

where $h$ is defined by $h = \frac{1}{2}(x_G - x_S)$. Please notice that the NIP wave experiment does not include any reflector characteristics. Thus it is possible to place a point source in a subsurface velocity model and to calculate the approximate kinematic traveltime response at the surface. Adjusting the velocity model and the location of the point source provides the same traveltime response actually extracted from the data by means of the CRS-stack. This is the basic idea of the inversion presented by Duveneck (2002).

### Extension to finite-offsets

I consider a point source at the same depth point as in the ZO-case. Let me denote this point by CRP (common-reflection point). Separating the source and receiver location under the condition that the connecting central ray meets Snell's law at the CRP provides all the source/receiver pairs pertaining to the CRP-trajectory. This trajectory connects all points in the three-dimensional prestack data volume pertaining to the same depth point. The two ray segments connecting the CRP with the source location $x_S$ and the receiver location $x_G$ are no longer identical now. One has to distinguish also between the emerging wavefronts at the source and receiver location due to the point source at the CRP. This two wavefronts are the finite-offset equivalents of the NIP wave discussed earlier. The relationship between emergence angles and wavefront curvatures to first and second order spatial derivatives is analogous to the ZO-case. I make the necessary distinction using subscripts $S$ and $G$ (source and geophone). In the following I make use of five kinematic wavefront attributes in the FO-case (see Figure 1): I have to use the measured traveltime $T$ because it is not known how it is distributed on the two ray segments. Its first derivatives with respect to the source and receiver locations are given by $p_S = \frac{\partial T}{\partial x_S}$ and $p_G = \frac{\partial T}{\partial x_G}$ Further I assume the spatial second-order derivatives $M_S = \frac{\partial^2 T}{\partial x^2}$ and $M_G = \frac{\partial^2 T}{\partial x^2}$ associated with the two emerging wavefronts due to a point source at CRP to be known.

## THE TOMOGRAPHIC APPROACH

The wavefront attributes described above are independent of characteristics of the corresponding reflection point because they are related to theoretical experiments using only a point source placed in depth. Thus they allow to neglect all reflector effects in the tomographic approach. Billette and Lambaré (1998) discuss various approaches to tomography including finite-offset ray-pairs. In contrast to me they are not including second-order spatial derivatives. One possible approach discussed by them is to begin at the surface. The ray starting directions are known from the data. Ray-tracing can be performed until the traveltimes along the two ray segments equal the measured traveltime $T$. Using the correct velocity model the two rays should end in the same depth point. It is rather difficult to define a measure of misfit for this crossing

condition. It is more suitable to place a point source in depth and to perform ray tracing until the two rays reach the surface. The forward modeled traveltime $T$ and angles of emergence or alternatively the corresponding first-order traveltime derivatives can be compared to the measured quantities. A measure of misfit can easily be formulated in the least-squares sense. It is ensured that the rays start at the same point, a condition which is also met in reality. Therefore this approach is physically more suitable. One has to define for each ray-pair a starting position and two initial directions at the CRP in the subsurface. The directions can be given as angles to the vertical or as horizontal slowness components $p_x^S$ and $p_x^G$. I make use of slowness components because they are directly involved in the ray-tracing system (see the Appendix).

Additionally to Billette and Lambaré (1998) I make also use of second order traveltime derivatives associated with the waves due to the point source at the CRP, i.e I require the emerging wavefronts to focus at the CRP if propagated back from the surface to depth . Thus I do not use a measure of misfit between first order but between second order traveltime expansions at the measurement surface. For an overview of data and model components view Figure 1.

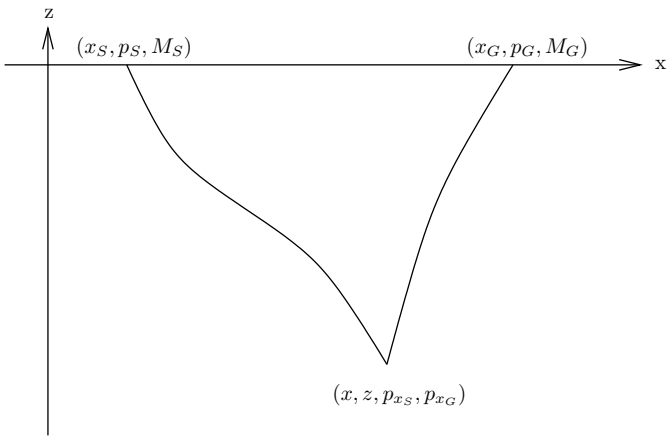



**Figure 1:** Overview over data and model components. The two central rays propagating in an isotropic inhomogeneous medium connect the source and receiver location. They start in a common point, the CRP.

As the ray-starting positions and directions at the CRP in the subsurface are not known a priori, they have to be included as model parameters to be inverted for. The velocity model $v(x, z)$ is represented by two-dimensional B-splines:

$$v(x,z) = \sum_i^{n_x} \sum_j^{n_z} v_{ij} \beta_i(x) \beta_j(z) \ , \tag{3}$$

where $\beta_i(x)$and$\beta_j(z)$ are B-spline basis functions and $n_x$ and $n_z$ are the chosen numbers of knots in the horizontal and vertical direction. During the inversion process smooth third derivatives of the velocity are required. That is why the B-spline basis functions have to be at least of degree four.

The inverse problem can be formulated as follows: The velocity model can be seen to be consistent with the data if all forward modeled quantities due to a point source in depth equal the measured quantities. This leads to the problem of minimizing a merit function $S$ in the least-squares sense:

$$S = \|\mathbf{d}_{mea} - \mathbf{d}_{mod}\|_2^2 = \frac{1}{2} \Delta \mathbf{d}^T(\mathbf{m}) \underline{\mathbf{C}}_D^{-1} \Delta \mathbf{d}(\mathbf{m}) \ , \tag{4}$$

where $\Delta \mathbf{d}$ is defined as the difference between the vectors containing the measured data $\mathbf{d}_{mea}$ and modeled data $\mathbf{d}_{mod}$. The matrix $\underline{\mathbf{C}}_D$ is sometimes called the data covariance matrix, here assumed to be diagonal, which weights the different data components. The modeled data are calculated using a nonlinear forward modeling operator $\mathbf{f}$:

$$\mathbf{d}_{mod} = \mathbf{f}(\mathbf{m}) \ , \tag{5}$$

where $\mathbf{m}$ is a vector containing all model parameters. Because of the nonlinearity of $\mathbf{f}$ the inverse problem has to be solved iteratively. The forward modeling operator is linearized during each iteration. Starting with a first guess model $\mathbf{m}_0$, a sequence of model updates $\Delta\mathbf{m}$ is found, hoping that the process converges to a global minimum. Having found a model $\mathbf{m}^{(n)}$ in the $n$'th iteration, the modeling operator $\mathbf{f}$ is locally linearized by (Duveneck, 2002)

$$\mathbf{f}(\mathbf{m}^{(n)} + \Delta\mathbf{m}^{(n+1)}) = \mathbf{f}(\mathbf{m}^{(n)}) + \underline{\mathbf{F}}^{(n)}\Delta\mathbf{m}^{(n+1)} \; . \tag{6}$$

The matrix $\underline{\mathbf{F}}^{(n)}$ is a matrix containing the Fréchet-derivatives $\frac{\partial(T,M_S,M_G,p_S,p_G,x_S,x_G)}{\partial(x,z,p_x^S,p_x^G)}$. The Fréchet derivatives can be obtained during forward modeling by ray-perturbation theory (Farra and Madariaga, 1987). Their calculation is described in detail in the appendix. A necessary condition for a minimum of $S$ is $\nabla_{\mathbf{m}}S = \mathbf{0}$. Using equation (6) one obtains

$$\nabla_{\mathbf{m}}S \approx -\underline{\mathbf{F}}^T\underline{\mathbf{C}}_D^{-1}(\Delta\mathbf{d}(\mathbf{m}^{(n)}) - \underline{\mathbf{F}}\Delta\mathbf{m}) = \mathbf{0} \; . \tag{7}$$

This leads to a least-squares solution for $\Delta\mathbf{m}$ if the inverse of $\underline{\mathbf{F}}^T\underline{\mathbf{C}}_D^{-1}\underline{\mathbf{F}}$ exists.

In practice, $\underline{\mathbf{F}}$ is usually ill-conditioned because not all model components are sufficiently constrained by the data alone. Additional information is introduced to further constrain the model and to regularize the problem. I am looking for the simplest model explaining the data. Thus, I require the velocity to be smooth. The second spatial derivatives of the velocity $v(x,z)$ give a measure of smoothness. The smoothness criterion is detailed in the appendix. Its application leads to a matrix equation of the following form (Duveneck, 2002):

$$\hat{\underline{\mathbf{F}}}\Delta\mathbf{m} = \Delta\hat{\mathbf{d}} \; , \tag{8}$$

where

$$\hat{\underline{\mathbf{F}}} = \begin{pmatrix} \underline{\mathbf{C}}_D^{-\frac{1}{2}}\underline{\mathbf{F}} \\ [\underline{\mathbf{0}},\underline{\mathbf{B}}] \end{pmatrix} \quad , \quad \Delta\hat{\mathbf{d}} = \begin{pmatrix} \underline{\mathbf{C}}_D^{-\frac{1}{2}}\Delta\mathbf{d}(\mathbf{m}^{(n)}) \\ -[\underline{\mathbf{0}},\underline{\mathbf{B}}]\mathbf{m}^{(n)} \end{pmatrix} \; . \tag{9}$$

The matrix $\hat{\underline{\mathbf{F}}}$ is large for a relatively little amount of data. It is sparse as the Fréchet derivatives relating parameters of one ray-pair to those of another are identically zero. Also each ray-pair is influenced only by a fraction of the total number of velocity knot-point values. This sparsity is taken advantage of using the LSQR algorithm (Paige and Saunders, 1982a,b).

## SYNTHETIC DATA EXAMPLE

To test the inversion algorithm introduced in the previous section I apply it to a synthetic data example. The model to be inverted for is shown in Figure 2. It is described by $8 \times 8$ B-spline coefficients with a horizontal spacing of $500m$ and a vertical spacing of $400m$. 108 data points consisting of $(T, M_S, M_G, p_S, p_G, x_S, x_G)_i$, $i = 1, ..., n_{Data}$, are generated directly by dynamic ray tracing, each ray segment starting at the corresponding CRP with an absolute value of $p = 0.15 \cdot 10^{-3}\frac{s}{m}$ of the horizontal component of the slowness vector. This leads to different starting angles depending on the local velocity at the different CRP's. The results are source-receiver pairs separated by a wide range of finite offsets. The offset varies between $300m$ and $3500m$.

During the inversion process I used no spatially variable weighting of the different regularization terms. The different terms were initially weighted using $\epsilon_{xx} = \epsilon_{xz} = \epsilon_{zz} = 0.005$ and $\epsilon = 1.5 \cdot 10^{-5}$ (see Appendix). The different data components were weighted using the following weighting terms:

$$\sigma_x = 1.0 \quad \sigma_T = 1.0 \quad \sigma_M = 0.9 \quad \sigma_{p_x} = 0.2 \; ,$$

where the different $\sigma$'s are the elements on the main diagonal of $\underline{\mathbf{C}}_D^{\frac{1}{2}}$. To start the inversion process I used a constant velocity medium with $v = 3000\frac{m}{s}$. Before I show the inversion results let me remark that all graphics except the ones showing differences between two models have been clipped using the extremal values of the model shown in Figure 2, i.e. all figures have been clipped against $v_{min} = 1500.0\frac{m}{s}$ and $v_{max} = 3500.0\frac{m}{s}$. As can be seen the model is well reconstructed (see Figure 3). The maximum difference between the inversion result and the model used for forward modeling is in the range of $\pm 110\frac{m}{s}$ as can be seen in Figure 4.

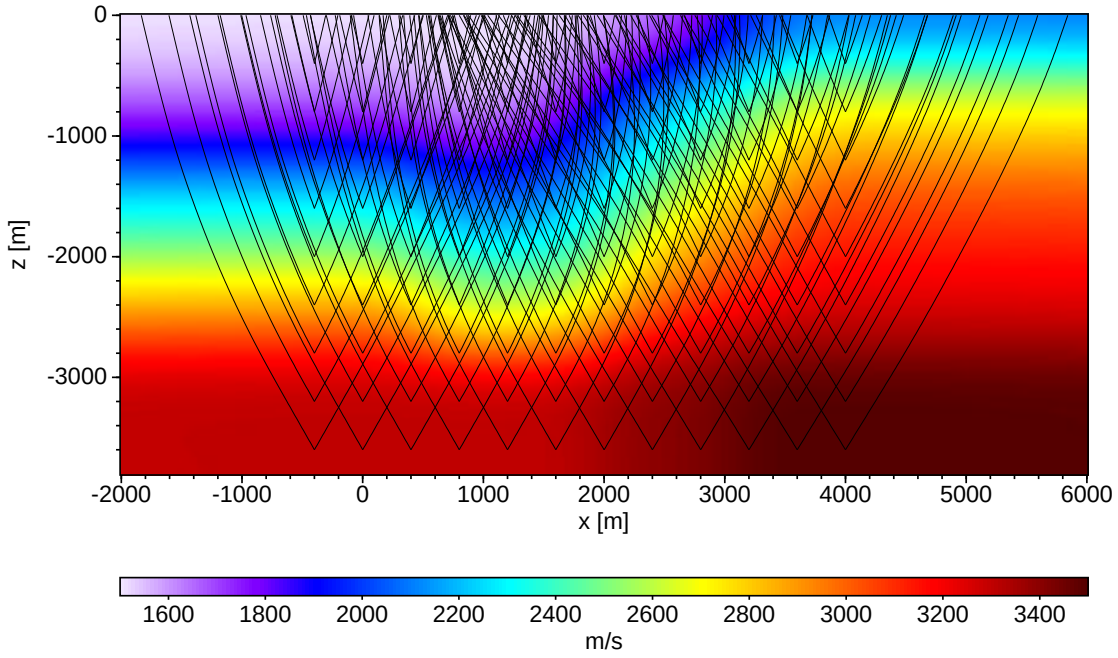



**Figure 2:** Model used for forward modeling of inversion input. Also shown are the forward modeled ray segments.

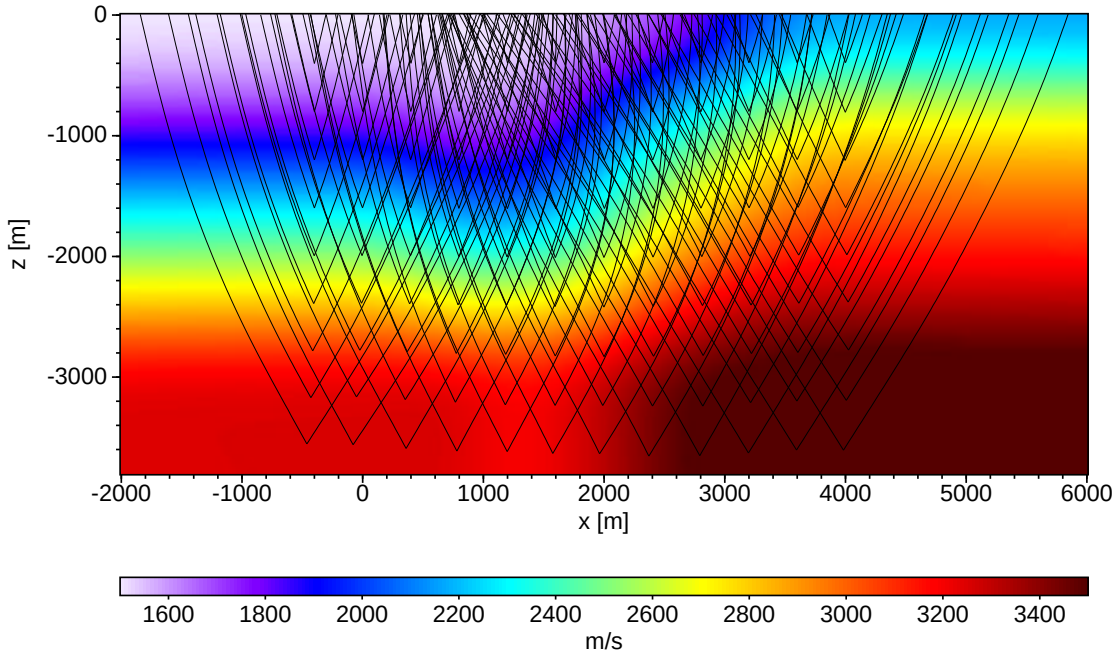



**Figure 3:** Inversion result and ray segments after 12 iterations.

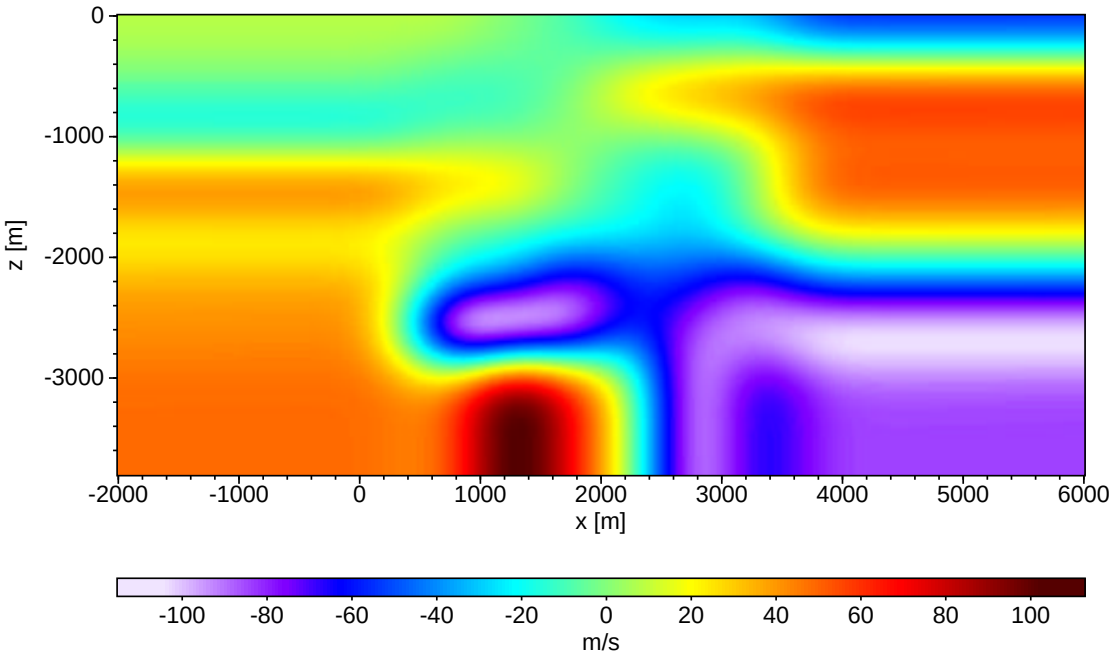
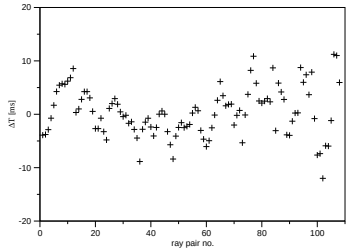
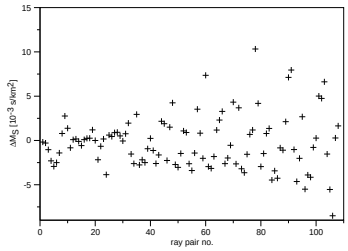


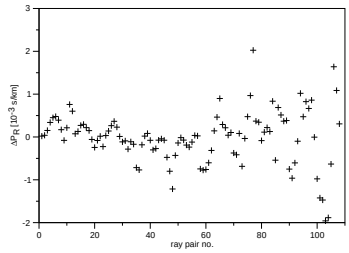

**Figure 4:** Difference between the inversion result and the model shown in Figure 2.



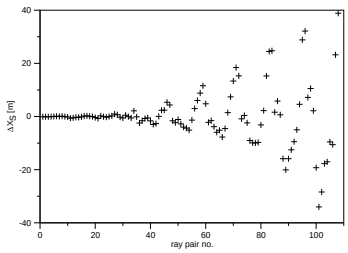

**(a)** Traveltime difference $\Delta T$ versus ray-pair number




**(b)** Difference in the second-order traveltime derivative at the source location versus ray-pair number




**(c)** Difference in the horizontal slowness component at the receiver location versus ray-pair number




**(d)** Difference in the source location versus ray-pair number

**Figure 5:** Comparison of forward modeled data values and those obtained after the inversion process.

Let me now compare the inverted with the forward modeled data. I will show four examples in Figure 5. Each subfigure shows the difference between the forward modeled and inverted data values. To give an impression of how the inversion process progresses I present three intermediate steps in Figure 6.

## CONCLUSIONS

I have presented a tomographic inversion scheme which makes use of finite-offset kinematic wavefield attributes to determine smooth, laterally inhomogeneous subsurface velocity models. The method is a logical extension of the ZO inversion presented by Duveneck (2002). I have tested it successfully with a synthetic example.

In the ZO case kinematic wavefield attributes derived from seismic data using the CRS stack are used as input for the inversion process. In the FO case, attributes derived with the FO CRS stack should be used (Zhang et al., 2001). This requires to extract the second-order traveltime derivatives associated with waves due to a point source at CRP from the FO CRS surface. A way to do this is part of current research.

## ACKNOWLEDGMENTS

This work was kindly supported by the sponsors of the *Wave Inversion Technology (WIT) Consortium*, Karlsruhe, Germany.

## REFERENCES

Billette, F. and Lambaré, G. (1998). Velocity macro-model estimation from seismic reflection data by stereotomography. *Geophys. J. Int.*, 135:671–690.

Bortfeld, R. (1989). Geometrical ray theory: Rays and traveltimes in seismic systems (second-order approximations of the traveltime). *Geophysics*, 54(3):342–349.

Duveneck, E. (2002). Tomographic velocity model inversion with CRS attributes. *Wit-Report 2002*, pages 92–106.

Duveneck, E., Klüver, T., and Mann, J. (2003). Tomographic velocity model estimation with data-derived first and second traveltime derivatives. In *Expanded Abstracts*, volume 8th International Congress of The Brazillian Geophysical Society, Rio de Janeiro. SBGf-Sociedade Brasiliera de Geofísica.

Farra, V. and Madariaga, R. (1987). Seismic waveform modelling in heterogeneous media by ray perturbation theory. *J. Geophys. Res.*, 92(B3):2697–2712.

Hubral, P. (1983). Computing true amplitude reflections in a laterally inhomogeneous earth. *Geophysics*, 48(8):1051–1062.

Mann, J., Jäger, R., Müller, T., Höcht, G., and Hubral, P. (1999). Common-reflection-surface stack - a real data example. *J. Appl. Geoph.*, 42(3,4):301–318.

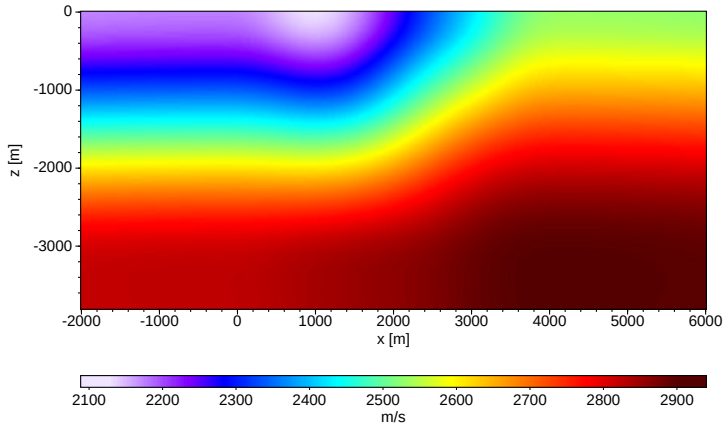
Paige, C. C. and Saunders, M. A. (1982a). Algorithm 583 - LSQR: Sparse linear equations and least squares problems. *ACM Trans. Math. Softw.*, 8(2):195–209.

Paige, C. C. and Saunders, M. A. (1982b). LSQR: An algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Softw.*, 8(1):43–71.

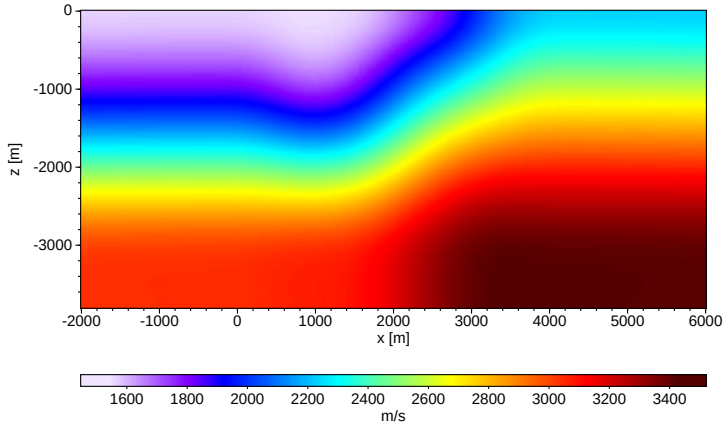
Schleicher, J., Tygel, M., and Hubral, P. (1993). Parabolic and hyperbolic paraxial two-point traveltimes in 3d media. *Geophys. Prosp.*, 41(4):495–513.

Červený, V. (2001). *Seismic ray theory*. Cambridge University Press.

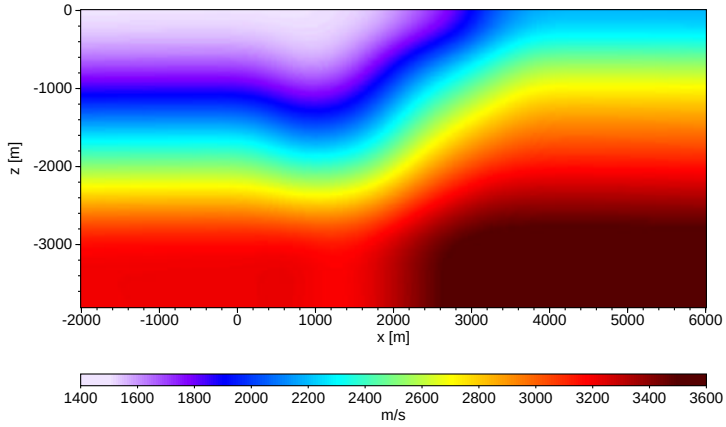
Zhang, Y., Bergler, S., and Hubral, P. (2001). Common-reflection-surface (CRS) stack for common-offset. *Geophys. Prosp.*, 41(4):495–718.




**(a)** Intermediate result after 1 iteration




**(b)** Intermediate result after 4 iterations




**(c)** Intermediate result after 9 iterations

**Figure 6:** Three intermediate results of the inversion process.

## APPENDIX A

## FORWARD MODELING AND CALCULATION OF FRÉCHET DERIVATIVES

The central rays connecting the point source at CRP and the measurement surface are found using standard ray theory with the model parameters of the current inversion iteration as initial conditions. Ray theory is formulated using a Hamiltonian formalism in general Cartesian coordinates. The Hamiltonian function follows directly from the eikonal equation:

$$H(x,z,p_x,p_z) = \frac{1}{v^2(x,z)} - p_x^2 - p_z^2 = 0\ . \tag{10}$$

As I do not consider turning points with respect to the $z$-coordinate, I can solve equation (10) for $p_z$ using $z$ as the independent variable in the ray tracing process. This way I derive the reduced Hamiltonian operator $H^R$:

$$H^R = -p_z = \sqrt{\frac{1}{v^2(x,z)} - p_x^2}\ . \tag{11}$$

The corresponding ray tracing system reads:

$$\frac{dx}{dz} = \frac{\partial H^R}{\partial p_x} \quad , \quad \frac{dp_x}{dz} = -\frac{\partial H^R}{\partial x} \tag{12}$$

Traveltime $T$ is determined by integration of:

$$\frac{dT}{dz} = \frac{1}{v^2(x,z)p_z} \tag{13}$$

Having found the central rays the traveltime T is simply the sum of the traveltimes of the two rays pertaining to one CRP. The first order derivative of T with respect to the $x$-coordinate is given by $p_x$ at the surface.

The second-order spatial derivatives of $T$ for fixed ray starting locations in the subsurface are calculated performing dynamic ray tracing along the known central rays. Perturbations of $x$ and $p_x$ at the ray endpoint may be linearly related to the corresponding properties at the ray starting point with the ray propagator matrix, which will be denoted by $\underline{\mathbf{T}}$. The ray propagator matrix is obtained solving

$$\frac{d}{dz}\begin{pmatrix} \Delta x_0 \\ \Delta p_{x_0} \end{pmatrix} = \underline{\mathbf{S}}\begin{pmatrix} \Delta x_0 \\ \Delta p_{x_0} \end{pmatrix}\ . \tag{14}$$

The matrix $\underline{\mathbf{S}}$ is the so called system matrix containing four elements:

$$\underline{\mathbf{S}} = \begin{pmatrix} \frac{\partial^2 H^R}{\partial x \partial p_x} & \frac{\partial^2 H^R}{\partial p_x^2} \\ \frac{\partial^2 H^R}{\partial x^2} & \frac{\partial^2 H^R}{\partial x \partial p_x^2} \end{pmatrix}\ . \tag{15}$$

All derivatives have to be evaluated on the central ray. The quantities $\Delta x_0$ and $\Delta p_{x_0}$ are perturbations of the initial conditions. The solution of equation (14) relating the perturbations of $x$ and $p_x$ at the ray endpoint to the perturbations of the initial conditions at the ray starting point reads:

$$\begin{pmatrix} \Delta x \\ \Delta p_x \end{pmatrix} = \underline{\mathbf{T}}(z,z_0)\begin{pmatrix} \Delta x_0 \\ \Delta p_{x_0} \end{pmatrix} \tag{16}$$

The quantities $\Delta x$ and $\Delta p_x$ describe the perturbations of $x$ and $p_x$ at the ray end point. The propagator matrix $\underline{\mathbf{T}}$ contains four elements

$$\underline{\mathbf{T}} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \tag{17}$$

and fulfills the initial condition $\underline{\mathbf{T}}(\mathbf{x}_0,\mathbf{x}_0) = \underline{\mathbf{I}}$, where $\underline{\mathbf{I}}$ denotes the identity matrix. It coincides with the surface-to-surface propagator matrix $\underline{\mathbf{T}}$ introduced by Bortfeld (1989) specialized to the case of horizontal

anterior and posterior surfaces. Furthermore the second order derivative of the traveltime $T$ is derived from the surface-to-surface propagator matrix given in equation (17) using (Červený, 2001):

$$\frac{\partial^2 T}{\partial x^2} = M = DB^{-1} \; . \tag{18}$$

Changes in the final slowness and position values due to changes in the velocity model can be calculated using a perturbed reduced Hamiltonian operator (Farra and Madariaga, 1987):

$$H^R = H_0^R + \frac{\partial H_0^R}{\partial v}\Delta v = H_0^R + \Delta H^R \; , \tag{19}$$

where $H_0^R$ denotes the unperturbed Hamiltonian. This results in the following inhomogeneous system at the ray starting position:

$$\frac{d}{dz}\begin{pmatrix} \Delta x_0 \\ \Delta p_{x_0} \end{pmatrix} = \underline{\mathbf{S}}\begin{pmatrix} \Delta x_0 \\ \Delta p_{x_0} \end{pmatrix} + \begin{pmatrix} \frac{\partial \Delta H}{\partial p_x} \\ -\frac{\partial \Delta H}{\partial x} \end{pmatrix} \; . \tag{20}$$

All derivatives have to be evaluated on the unperturbed central ray. The solution of equation (20) can again be found using propagator techniques (Farra and Madariaga, 1987):

$$\begin{pmatrix} \Delta x \\ \Delta p_x \end{pmatrix} = \underline{\mathbf{T}}_0(z, z_0)\begin{pmatrix} \Delta x_0 \\ \Delta p_{x_0} \end{pmatrix} + \int_{z_0}^{z} \underline{\mathbf{T}}_0(z, \tau)\begin{pmatrix} \frac{\partial \Delta H}{\partial p_x} \\ -\frac{\partial \Delta H}{\partial x} \end{pmatrix} d\tau \; . \tag{21}$$

Now I have to look for paraxial rays in the perturbed medium. The paraxial ray tracing system in the perturbed medium is obtained by analysis of the perturbation of the system matrix $\underline{\mathbf{S}}$ (Farra and Madariaga, 1987):

$$\underline{\mathbf{S}} = \underline{\mathbf{S}}_0 + \Delta\underline{\mathbf{S}}_1 + \Delta\underline{\mathbf{S}}_2 \; , \tag{22}$$

where $\Delta\underline{\mathbf{S}}_1$ is a term due to perturbations $\Delta v$ of the velocity, while $\Delta\underline{\mathbf{S}}_2$ is due to perturbations $\Delta x$ and $\Delta p_x$ of the central reference ray:

$$\Delta\underline{\mathbf{S}}_1 = \begin{pmatrix} \frac{\partial^2 \Delta H^R}{\partial x \partial p_x} & \frac{\partial^2 \Delta H^R}{\partial p_x^2} \\ -\frac{\partial^2 \Delta H^R}{\partial x^2} & -\frac{\partial^2 \Delta H^R}{\partial x \partial p_x} \end{pmatrix} \tag{23}$$

$$\Delta\underline{\mathbf{S}}_2 = \left(\Delta x\frac{\partial}{\partial x} + \Delta p_x \frac{\partial}{\partial p_x}\right)\begin{pmatrix} \frac{\partial^2 H_0^R}{\partial x \partial p_x} & \frac{\partial^2 H_0^R}{\partial p_x^2} \\ -\frac{\partial^2 H_0^R}{\partial x^2} & -\frac{\partial^2 H_0^R}{\partial x \partial p_x} \end{pmatrix} \tag{24}$$

The paraxial rays in the perturbed medium are calculated analogous to the ones in the unperturbed medium:

$$\begin{pmatrix} \Delta x \\ \Delta p_x \end{pmatrix} = \underline{\mathbf{T}}(\mathbf{x}, \mathbf{x}_0)\begin{pmatrix} \Delta x_0 \\ \Delta p_{x_0} \end{pmatrix} \; , \tag{25}$$

with $\underline{\mathbf{T}}$ now given by (Farra and Madariaga, 1987)

$$\underline{\mathbf{T}}(z, z_0) = \underline{\mathbf{T}}_0(z, z_0) + \int_{z_0}^{z} \underline{\mathbf{T}}_0(z, \tau)\left(\Delta\underline{\mathbf{S}}_1 + \Delta\underline{\mathbf{S}}_2\right)\underline{\mathbf{T}}_0(\tau, z_0)d\tau \; , \tag{26}$$

where $\underline{\mathbf{T}}_0$ denotes the propagator matrix in the unperturbed medium.

So far no perturbation of the initial $z$-coordinate is considered. Let me define the perturbation $\Delta\underline{\mathbf{T}}$ of $\underline{\mathbf{T}}$ due to a perturbation $\Delta z_0$ of the initial $z$-coordinate:

$$\Delta\underline{\mathbf{T}} = \underline{\mathbf{T}}(z_1, z_0 + \Delta z) - \underline{\mathbf{T}}(z_1, z_0) \; . \tag{27}$$

Using the chain rule (Červený, 2001) this expression can be rewritten in the following form:

$$\Delta\underline{\mathbf{T}} = \underline{\mathbf{T}}(z_1, z_0)\left(\underline{\mathbf{T}}^{-1}(z_0 + \Delta z, z_0) - \underline{\mathbf{I}}\right) \; . \tag{28}$$

Using the fact that

$$\frac{\partial \underline{\mathbf{T}}}{\partial z} = \underline{\mathbf{S}}\,\underline{\mathbf{T}}\,, \tag{29}$$

one can perform a Taylor expansion around $z_0$, yielding a term for $\underline{\mathbf{T}}^{-1}(z_0 + \Delta z, z_0)$. This leads to an expression for the perturbed surface-to-surface propagator matrix.

The perturbed surface-to-surface propagator matrix is a summation of the individual terms (Duveneck, 2002):

$$\underline{\mathbf{T}}(\Delta x, \Delta z, \Delta p_x, \Delta v) = \underline{\mathbf{T}}_0 + \Delta\underline{\mathbf{T}}(\Delta x, \Delta p_x) + \Delta\underline{\mathbf{T}}(\Delta z) + \Delta\underline{\mathbf{T}}(\Delta v)\,. \tag{30}$$

The above presented equations together with

$$\Delta M = B^{-1}\Delta D - DB^{-2}\Delta B$$

suffice to calculate all Fréchet derivatives needed in the inversion process.

## APPENDIX B

## REGULARIZATION OF THE TOMOGRAPHIC MATRIX

Because the tomographic matrix $\underline{\mathbf{F}}$ is in general ill-conditioned the inversion problem has to be regularized by introducing additional constraints on the model parameters. I am looking for a smooth model without any artificial structure. A physically sensible way to obtain such a model is to require the velocity model to have minimum curvature, i.e., minimum second spatial derivatives. In order to be as much as possible independent of the discrete structure of the velocity knot points the minimum curvature condition is applied on the smooth model itself, not on the B-spline coefficients.

The second spatial derivative of the velocity, described by B-splines is ($x$-component):

$$\frac{\partial^2 v(x,z)}{\partial x^2} = \sum_i \sum_j v_{ij} \frac{\partial^2 \beta_i(x)}{\partial x^2} \beta_j(z)\,. \tag{31}$$

The $z$- and mixed components are derived in an analogous fashion. The $L_2$-norm of (31) is given by

$$\left\| \frac{\partial^2 v}{\partial x^2} \right\|_2^2 = \int_x \int_z \left( \frac{\partial^2 v(x,z)}{\partial x^2} \right)^2 dz dx = \mathbf{m}_{(v)}^T \underline{\mathbf{D}}^{xx} \mathbf{m}_{(v)}\,, \tag{32}$$

where $\mathbf{m}_{(v)}$ denotes that part of the model parameter vector containing the B-spline coefficients. Similar expressions are derived for the second derivatives with respect to $z$ and for the mixed derivatives. This leads to the following term included in the merit function $S$:

$$\int_x \int_z \epsilon(x,z) \left[ \epsilon^{xx} \left( \frac{\partial^2 v(x,z)}{\partial x^2} \right)^2 + \epsilon^{xz} \left( \frac{\partial^2 v(x,z)}{\partial x \partial z} \right)^2 + \epsilon^{zz} \left( \frac{\partial^2 v(x,z)}{\partial z^2} \right)^2 + \epsilon v^2(x,z) \right] dz dx = \mathbf{m}_{(v)}^T \underline{\mathbf{D}}'' \mathbf{m}_{(v)} \tag{33}$$

Therein $\epsilon(x,z)$ is a spatially variable weighting function represented by B-splines. It is based on the same knot point locations as the velocity model. The factors $\epsilon^{xx}$, $\epsilon^{xz}$, $\epsilon^{zz}$ and $\epsilon$ are used for normalization and to balance the contributions of the different terms.

Including the term (33), the merit function to minimize is given by (compare to equation (4)):

$$2S(\mathbf{m}) = \Delta\mathbf{d}^T(\mathbf{m}) \underline{\mathbf{C}}_D^{-1} \Delta\mathbf{d}(\mathbf{m}) + \hat{\epsilon}\mathbf{m}_{(v)}^T \underline{\mathbf{D}}'' \mathbf{m}_{(v)}\,, \tag{34}$$

where $\hat{\epsilon}$ is the highest value of $\epsilon^{xx}$, $\epsilon^{xz}$, $\epsilon^{zz}$ and $\epsilon$. The different $\epsilon$ in equation (33) are normalized by $\hat{\epsilon}$. For further calculations I need to find a matrix $\underline{\mathbf{B}}$ with $\underline{\mathbf{B}}^T\underline{\mathbf{B}} = \hat{\epsilon}\underline{\mathbf{D}}''$ (Duveneck, 2002). This is only possible if $\underline{\mathbf{D}}''$ is positive definite. Therefore the term $\epsilon v^2(x,z)$ is included in equation (33). Because there is no physical reason for minimizing the velocity itself, $\epsilon$ has to be chosen much smaller than the other weighting factors.

Assuming the forward modeling operator can be locally approximated around $\mathbf{m}^{(n)}$ by $\mathbf{f}(\mathbf{m}^{(n+1)}) \approx \mathbf{f}(\mathbf{m}^{(n)}) + \underline{\mathbf{F}}\Delta\mathbf{m}^{(n+1)}$ (Duveneck, 2002), where $\mathbf{m}^{(n+1)} = \mathbf{m}^{(n)} + \Delta\mathbf{m}^{(n+1)}$, one obtains

$$\begin{aligned} \nabla_{\mathbf{m}} S &= -\underline{\mathbf{F}}^T \underline{\mathbf{C}}_D^{-1}(\mathbf{d}_{mea} - \mathbf{f}(\mathbf{m})) + \hat{\epsilon}\underline{\mathbf{D}}''\mathbf{m}_{(v)} \\ &\approx -\underline{\mathbf{F}}^T \underline{\mathbf{C}}_D^{-1}\Delta\mathbf{d}(\mathbf{m}^{(n)}) + \underline{\mathbf{F}}^T \underline{\mathbf{C}}_D^{-1}\underline{\mathbf{F}}\Delta\mathbf{m}^{(n+1)} + \hat{\epsilon}\underline{\mathbf{D}}''(\mathbf{m}_{(v)}^{(n)} + \Delta\mathbf{m}^{(n+1)}) \end{aligned} . \tag{35}$$

The necessary condition $\nabla_{\mathbf{m}} S = \mathbf{0}$ for a minimum of the merit function (34) leads to the system which has to be solved (Duveneck, 2002):

$$\hat{\underline{\mathbf{F}}}\Delta\mathbf{m} = \Delta\hat{\mathbf{d}} , \tag{36}$$

with

$$\hat{\underline{\mathbf{F}}} = \begin{pmatrix} \underline{\mathbf{C}}_D^{-\frac{1}{2}}\underline{\mathbf{F}} \\ [\underline{\mathbf{0}}, \underline{\mathbf{B}}] \end{pmatrix} \quad , \quad \Delta\hat{\mathbf{d}} = \begin{pmatrix} \underline{\mathbf{C}}_D^{-\frac{1}{2}}\Delta\mathbf{d}(\mathbf{m}^{(n)}) \\ -[\underline{\mathbf{0}}, \underline{\mathbf{B}}]\mathbf{m}^{(n)} \end{pmatrix} . \tag{37}$$

The searched for model update $\Delta\mathbf{m}$ is the least squares solution to equation (36). The updated model components are given by $\mathbf{m}^{(n+1)} = \mathbf{m}^{(n)} + \lambda\Delta\mathbf{m}^{(n+1)}$, where $0 < \lambda \leq 1$.