Tomographic velocity model inversion with CRS attributes

Eric Duveneck

email: *eric.duveneck@gpi.uni-karlsruhe.de*
keywords: *tomography, velocity model, CRS Stack, attributes*

ABSTRACT

Kinematic information for the construction of velocity models can be extracted in a robust way from seismic prestack data with the Common-Reflection-Surface Stack. This data-driven process results in a number of kinematic wavefield attributes – wavefront curvatures and normal ray emergence angles – that parameterise moveout surfaces in the prestack data, associated with each sample in a simulated zero-offset section. I present a tomographic inversion method that makes use of this kinematic information to determine smooth, laterally heterogeneous subsurface velocity models for depth imaging. The input for the inversion consists of kinematic wavefield attributes picked at a number of locations in the simulated zero-offset section. An optimum model is found in an iterative way by minimisation of the misfit between the picked data and the corresponding modelled values. The required forward modelled quantities are obtained during each iteration by dynamic ray tracing along normal rays pertaining to the input data points. Fréchet derivatives for the tomographic matrix are calculated by ray perturbation theory. The algorithm is tested on 1D and 2D examples and the inversion is successfully applied to a 2D synthetic prestack data set.

INTRODUCTION

The determination of a suitable velocity model is one of the crucial steps for seismic depth imaging in laterally inhomogeneous media. It is important not only for correct positioning of reflection events in the subsurface but also for obtaining an optimally focused image if prestack migration is used. A number of approaches for the construction of velocity models exist, which differ in the criterion used for evaluating the quality of the current model, in the way model updates are determined and in the assumptions made about the velocity model (blocky, layered, smooth, etc.).

Commonly used migration velocity analysis methods are usually based on residual moveout analysis in common image gathers. They make use of the criterion that in the correct velocity model reflector image depths produced by prestack migration should be independent of offset (e.g. Deregowski, 1990; Liu, 1997). Another approach, depth focusing analysis (e.g. Jeannot and Faye, 1986; MacKay and Abma, 1992), employs the fact that during downward continuation in the correct velocity model reflection events should collapse to zero offset at zero traveltime. Both approaches require the repeated application of prestack migration and are therefore expensive in terms of computation time.

An often used tool for the determination of velcity models is reflection tomography (e.g. Bishop et al., 1985; Farra and Madariaga, 1988; Stork, 1992). In reflection tomography rays are traced through a model defined by a velocity distribution and a number of reflectors. Global model updates are computed to minimise the misfit between the modelled traveltimes and the corresponding values picked in the prestack data. The drawback of reflection tomography is the tremendous amount of picking that is necessary to obtain traveltimes from the prestack data and the assumption of continuous reflectors, often across the entire section. Although many attempts to automatise the process have been made, the problem of picking remains.

Recently, Billette and Lambaré (1998) have presented a tomographic velocity inversion method called

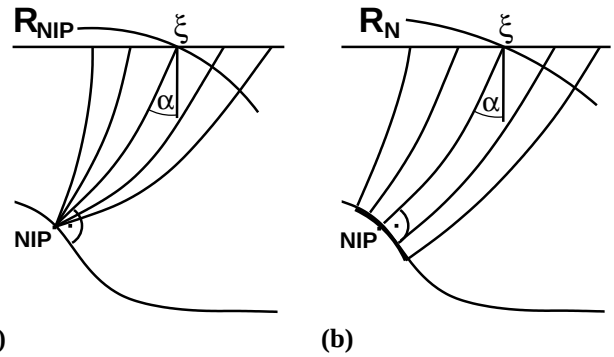


Figure 1: The two eigenwaves associated with R_{NIP} and R_N: (a) the NIP wave and (b) the normal wave.

Stereotomography which makes use of the slopes of locally coherent events in shot and receiver gathers together with traveltimes to obtain a smooth velocity model. An advantage of that method lies in the fact that, as only locally coherent events are considered, no interfaces need to be introduced in the model. Consequently, no picking along continuous events is necessary.

In this paper I present a tomographic velocity inversion method that makes use of kinematic information extracted from the entire prestack data volume in a data-driven way by an application of the Common-Reflection-Surface Stack (e.g. Mann et al., 1999; Jäger et al., 2001). Input data for the velocity inversion may then be picked in the resulting simulated zero-offset section of significantly improved signal-to-noise ratio, thus considerably simplifying the picking. Each picked data point consits of a number of kinematic wavefield attributes that can be used to describe an entire approximate kinematic multi-offset response of a common reflection point. The velocity inversion is performed with a tomographic approach which is in some respects similar to Stereotomography. While in that method source and receiver rays pertaining to a reflection point are considered, I use properties calculated along normal rays by dynamic ray tracing to approximate the kinematic multi-offset response of reflection points.

THE COMMON-REFLECTION-SURFACE STACK AND VELOCITY INVERSION

The recently developed Common-Reflection-Surface Stack (CRS Stack) (e.g. Mann et al., 1999; Jäger et al., 2001) is a stacking technique that makes use of multi-parametric stacking surfaces to obtain optimum simulated zero-offset sections from seismic multi-coverage data in a data-driven way.

One of the advantages of the method lies in the fact that, apart from the simulated zero-offset section, one obtains a number of additional sections containing the values of so-called CRS attributes or kinematic wavefield attributes. These attributes are parameters that determine the shape of the CRS stacking surface. Their optimum values are obtained with a coherence analysis in the prestack data during the CRS Stack.

The CRS attributes – in the 2D case the three attributes are named α, $R_{\rm NIP}$, and $R_{\rm N}$ – can be given a clear physical meaning. While α is the normal ray emergence angle at the surface location ξ, $R_{\rm NIP}$ and $R_{\rm N}$ can be interpreted as radii of curvature of two so-called *eigenwaves* emerging at ξ (Hubral, 1983), associated with two hypothetical experiments. They are obtained from a point-source and an exploding reflector element attached to the normal-incidence point (NIP) on a reflector, respectively (Figure 1). The attribute sections contain physically meaningful information only where sufficiently high coherence values were obtained along the CRS operator. These locations can be easily identified with the help of a coherence section, which is also part of the CRS output. Synthetic examples of CRS Stack results (simulated ZO section and kinematic wavefield attribute sections $R_{\rm NIP}$ and α) are displayed in Figure 6.

The significance of the kinematic wavefield attribute sections lies in the fact that these can be used for a number of different applications (e.g. Bergler et al., 2002), such as the calculation of geometrical spreading factors and approximate projected Fresnel zones, and the separation of reflections from diffractions. In particular, the CRS attributes contain information on the subsurface distribution of seismic velocities. In the following I will briefly discuss their role in velocity model inversion.

The approximate (second order in half-offset) common-reflection-point (CRP) trajectory, i.e. the ap-

proximate kinematic response of a reflection point, lies completely within the CRS surface. For a given point (t_0, ξ) in the zero-offset plane and a given near-surface velocity v_0 the CRP trajectory in time-midpoint-offset space can approximately be described by the CRS attributes R_{NIP} and α alone (Höcht et al., 1999). If $\alpha \neq 0$ it deviates from the CMP trajectory, which is by definition confined to a fixed midpoint value. Restricting the CRS operator to a single midpoint, one finds that the normal moveout velocity v_{NMO}, as it is conventionally defined, can be written in terms of the CRS attributes R_{NIP} and α:

$$v_{\mathrm{NMO}}^2 = \frac{2\, v_0 R_{\mathrm{NIP}}}{t_0 \cos^2(\alpha)} \,. \tag{1}$$

R_{NIP} and α can therefore also be found by stacking-velocity analysis and the determination of local dips in a stacked zero-offset section. Their determination with the CRS Stack should yield much more stable results, though, especially in the presence of a high level of noise in the data.

The link between normal moveout velocity on the one hand and wavefront curvature and emergence angle on the other hand has been recognized by a number of different authors. Shah (1973) used this relation together with wavefront curvature propagation laws in a Dix-type inversion for 2D layered models containing dipping reflectors. Conventional Dix inversion (Dix, 1955) is then just a special case for 1D models and $\alpha = 0$. Hubral and Krey (1980) generalized Dix-type inversion to the case of 3D curved layered models. Recently, Biloti et al. (2002) presented a revised version of that algorithm in 2D based directly on CRS Stack results. They also include vertical velocity gradients within each layer in their inversion.

The concept of "having a NIP wavefront shrink back into its hypothetical source" (Hubral and Krey, 1980) for the determination of interval velocities can, in principle, be generalised to the case of arbitrary smooth velocity distributions. As the NIP wavefront parameters completely describe the approximate kinematic response of a common reflection point (see above), this concept is consistent with the well-known criterion of depth focusing analysis in the determination of migration velocities. It states that a migration velocity model is correct, if seismic reflections, when downward continuation is performed, focus at zero traveltime (e.g. Jeannot and Faye, 1986; MacKay and Abma, 1992).

TOMOGRAPHY WITH CRS ATTRIBUTES

In this section I will introduce a tomographic velocity inversion method for 2D isotropic models, based on the CRS attributes R_{NIP} and α. In contrast to Dix-type inversion, a smooth velocity model description without any discontinuities is used. This allows the continuous reflector assumption to be dropped: independent data points are picked from the CRS Stack section (together with the corresponding R_{NIP} and α sections), each point corresponding to a CRP in the subsurface. The inversion method is based on the above-mentioned criterion for a correct velocity model: in a correct model, all considered NIP waves, when propagated back into the earth along the normal ray, focus at zero traveltime. A first test of the method has been presented by Duveneck and Hubral (2002).

Data and model components

Each point (t_0,ξ) on an event in a CRS Stack section with a sufficiently large value in the corresponding coherence section can be associated with a common reflection point in the subsurface. Its approximate kinematic multi-offset response is defined through the corresponding CRS attributes R_{NIP} and α. Instead of R_{NIP}, I will use

$$M = \frac{1}{v_0\, R_{\mathrm{NIP}}} \,, \tag{2}$$

where v_0 is the near-surface velocity value that has been used during the CRS Stack. This has a number of advantages: firstly, the dependence on the chosen value of v_0 is removed, as R_{NIP} appears in the expression for the CRS operator only in a product with v_0. M is the second spatial derivative of the NIP wave traveltime on the zero-offset ray in the direction normal to the ray at ξ. The second-order approximation in half-offset h of the CRP-trajectory for a given point (t_0, ξ) (Höcht et al., 1999) can be written entirely in terms of M and α. It is independent of the near-surface velocity v_0. Secondly, the quantity M can be easily

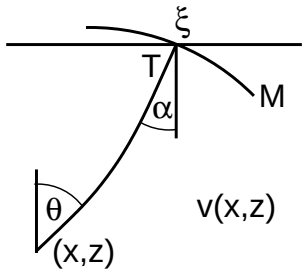


Figure 2: Definition of model and data components for the tomographic inversion.

calculated along a ray with dynamic ray tracing. When the ray tracing is started at the NIP, it is given by

$$M = \frac{P_2}{Q_2}, \tag{3}$$

where P_2 and Q_2 are elements of the ray propagator matrix $\mathbf{\Pi}$ in ray-centered coordinates (e.g. Červený, 2001, see also Appendix A). Thus, the approximate CRP response in the vicinity of a normal ray at any location along the ray can be directly modelled by dynamic ray tracing.

The input for the velocity inversion consists of a number of points picked in the CRS Stack section, defined by their values of t_0 and ξ, together with the associated values of α and M, taken and calculated from the corresponding CRS attribute sections. The ith data point is given by

$$(T, M, \alpha, \xi)_i \ , \tag{4}$$

where the one-way traveltime $T = t_0/2$ is used for convenience and $i = 1 \ldots n_{\mathrm{data}}$, when n_{data} data points are picked.

One might consider implementing the previously mentioned condition for a correct velocity model – the focusing of the NIP wave – directly. This would imply fixing α and ξ at their picked values and propagate the NIP wavefronts, described by the picked values of R_{NIP}, or the corresponding values of M, into the subsurface to check if they focus at $T = 0$ (i.e. $R_{\mathrm{NIP}} = 0$ at $T = 0$). One would then need to find a velocity model that effects this focusing for all considered data points.

On the other hand, all data (T, M, α, ξ) must be expected to be affected by noise or a certain measurement error. An inversion that does not allow for that error can become unstable. This can be taken into account if the dynamic ray tracing is started in the subsurface at the respective CRPs. As the true subsurface positions of the CRPs and the ray starting directions of the corresponding normal rays are initially unknown, these have to be considered as part of the model to be inverted for, along with the velocity distribution. The optimum model is found when the misfit between modelled and measured values of (T, M, α, ξ) is minimised. This is the approach that will be followed here.

In two-dimensional depth models, each CRP is characterised by its location in the subsurface (x, z) and its local dip angle θ, which also gives the direction of the normal ray. The smooth velocity model itself can be described by two-dimensional B-splines:

$$v(x, z) = \sum_{j=1}^{n_x} \sum_{k=1}^{n_z} v_{jk} \ \beta_j(x) \beta_k(z) \ , \tag{5}$$

where n_x and n_z are the chosen numbers of knots in the horizontal and vertical direction, respectively. $\beta_j(x)$ and $\beta_k(z)$ are B-spline basis functions in x and z for given knot sequences (e.g. de Boor, 1978). As shown in Appendix A, smooth third derivatives of the velocity are needed, therefore, B-spline functions of degree 4 will be used. Together with the coordinates and local dips of the reflection points the coefficients v_{jk} in equation (5) define the model. It is given by

$$\begin{matrix} (x, z, \theta)_i \\ v_{jk} \ , \end{matrix} \tag{6}$$

where $i = 1 \ldots n_{\mathrm{data}}$, $j = 1 \ldots n_x$ and $k = 1 \ldots n_z$. Model and data components are illustrated in Fig. 2.

Solution of the inverse problem

The inversion problem described in the previous section can be formulated as follows: find a model vector $\mathbf{m}$ consisting of the model components given in (6), that minimises the misfit between a data vector $\mathbf{d}$ containing the picked data values given in (4) and the corresponding modelled values $\mathbf{d}_{\mathrm{mod}} = \mathbf{f}(\mathbf{m})$. Here the nonlinear operator $\mathbf{f}$ symbolises the dynamic ray tracing in the given model.

If the least-squares norm (e.g. Tarantola, 1987) is used as a measure of misfit, the inverse problem becomes one of minimising a cost function

$$S(\mathbf{m}) = \frac{1}{2} \left\| \mathbf{d} - \mathbf{f}(\mathbf{m}) \right\|_D^2 = \frac{1}{2} \Delta \mathbf{d}^T(\mathbf{m}) \mathbf{C}_D^{-1} \Delta \mathbf{d}(\mathbf{m}), \tag{7}$$

where $\Delta \mathbf{d}(\mathbf{m}) = \mathbf{d} - \mathbf{f}(\mathbf{m})$. The matrix $\mathbf{C}_D$ is sometimes called the data covariance matrix, here assumed to be diagonal, which weights the different data components.

The nonlinear modelling operator $\mathbf{f}$ is locally lineariseable, therefore an iterative application of least-squares minimisation to the linearised problem can be used to find the minimum of the cost function S. 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 the global minimum of S.

Around a given model $\mathbf{m}_n$, corresponding to the nth iteration, the modelling operator can be approximated locally by $\mathbf{f}(\mathbf{m}_n + \Delta \mathbf{m}) \approx \mathbf{f}(\mathbf{m}_n) + \mathbf{F} \Delta \mathbf{m}$, where $\mathbf{F}$ is a matrix containing the Fréchet derivatives of $\mathbf{f}$ at $\mathbf{m}_n$. The Fréchet derivatives can in the present case be obtained during forward modelling by ray perturbation theory (Farra and Madariaga, 1987). Their calculation is detailed in Appendix A. Then

$$\nabla_{\mathbf{m}} S \approx -\mathbf{F}^T \mathbf{C}_D^{-1} \left(\Delta \mathbf{d}(\mathbf{m}_n) - \mathbf{F} \Delta \mathbf{m} \right). \tag{8}$$

Setting $\nabla_{\mathbf{m}} S = \mathbf{0}$ leads to the least-squares solution for $\Delta \mathbf{m}$ if the inverse of $\mathbf{F}^T \mathbf{C}_D^{-1} \mathbf{F}$ exists.

In practice, however, $\mathbf{F}$ is usually ill-conditioned, as not all model components are sufficiently constrained by the data alone. Therefore, a stable inverse cannot be computed. Additional information has to be introduced to further constrain the model parameters and regularise the problem. Constraining the model vector to have minimum length, as is done in damped least squares (e.g. Lines and Treitel, 1984), is not reasonable in the present case. Physically, it makes more sense to require the velocity model to be smooth, as we are looking for the simplest model that explains the data. The smoothness requirement also assures the applicability of dynamic ray tracing along single normal rays to model approximate kinematic CRP responses by determining the required attributes R_{NIP} and α. As shown in Appendix B, application of the smoothness criterion (minimum second spatial derivatives) as an additional constraint, not on the model update, but on the velocity $v(x, z)$ itself, leads to a matrix equation of the following form:

$$\hat{\mathbf{F}} \Delta \mathbf{m} = \Delta \hat{\mathbf{d}}, \tag{9}$$

where

$$\hat{\mathbf{F}} = \begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{F} \\ [\mathbf{0}\,, \mathbf{B}] \end{pmatrix} \quad , \quad \Delta \hat{\mathbf{d}} = \begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \Delta \mathbf{d}(\mathbf{m}_n) \\ -[\mathbf{0}\,, \mathbf{B}]\, \mathbf{m}_n \end{pmatrix}. \tag{10}$$

Here $\mathbf{B}^T \mathbf{B} = \epsilon\, \mathbf{D}''$, and the matrix $\mathbf{D}''$ is explained in Appendix B. The factor ϵ weights the relative contribution of the regularisation to the cost function. Solving equation (9) in the least-squares sense yields the desired model update $\Delta \mathbf{m}$, which minimises the cost function

$$S(\mathbf{m}) = \frac{1}{2} \Delta \mathbf{d}^T(\mathbf{m}) \mathbf{C}_D^{-1} \Delta \mathbf{d}(\mathbf{m}) + \frac{1}{2} \epsilon\, \mathbf{m}_{(v)}^T \mathbf{D}'' \mathbf{m}_{(v)}. \tag{11}$$

If the dimensions of $\hat{\mathbf{F}}$ are not too large the least-squares solution of equation (9) can be found e.g. by singular value decomposition (SVD) (e.g. Lines and Treitel, 1984). For larger matrices this becomes computationally too expensive and we need to apply more efficient methods, that take advantage of the sparsity of $\hat{\mathbf{F}}$ during the solution of (9). Here the LSQR algorithm (Paige and Saunders, 1982a,b) is used. It is an iterative method, based on conjugate-gradients, which solves linear systems of the form of (9) in the least-squares sense without explicitly performing any matrix inversion. The LSQR algorithm is well-known for its favourable numerical properties and is frequently used for tomographic problems in seismology (e.g. Nolet, 1987).

Practical aspects

Before the inversion process can be started, input data points need to be picked in the simulated zero-offset section produced by the CRS Stack. Picks are independent of each other and do not have to follow an event over successive traces. Their lateral separation should be smaller than the horizontal knot interval, though, and they should be well distributed over the entire section. The required values of M and α are then automatically extracted and calculated from the corresponding CRS attribute sections.

An initial velocity model is set up by defining B-spline knot sequences in the horizontal and vertical directions and assigning initial values to the velocity coefficients, such that these describe a 1D velocity distribution with some constant vertical velocity gradient.

For each of the picked data points (T, M, α, ξ), kinematic ray tracing in the downward direction is performed in the initial velocity model to obtain initial coordinates and local dip values (x, z, θ) for the corresponding CRPs. Starting with these values, dynamic ray tracing in the upward direction is performed until the rays reach the measurement surface. Initially, $\Delta\mathbf{d}$ will consist solely of the discrepancy between the measured and modelled values of M.

During ray tracing the Fréchet derivatives are calculated with the help of equations (A1) to (A11) in Appendix A. With these ingredients and a chosen value of ϵ equation (9) can be set up. Its least-squares solution $\Delta\mathbf{d}$ is found in the 1D case by SVD or in the 2D case with the LSQR algorithm. The updated model parameters $\mathbf{m}_{n+1} = \mathbf{m}_n + \lambda\Delta\mathbf{m}$, where $0 < \lambda \leq 1$, are then determined and new modelled data values are calculated by dynamic ray tracing. If the cost function (11) has increased, λ is decreased and the cost function recalculated, otherwise the next iteration is started by calculating Fréchet derivatives in the new model.

In the course of increasing iteration numbers the value of ϵ can be decreased as the model misfit decreases, allowing more and more details to be resolved. The inversion is stopped, if a given maximum number of iterations has been reached, if the data misfit has fallen below a specified value, or if a minimum of the cost function (11) has been reached, i.e. decreasing λ does not lead to a further decrease of the cost function.

SYNTHETIC TESTS

To test the inversion algorithm introduced in the previous section it is firstly applied to two synthetic examples, one in 1D and one in 2D, where the input data have been calculated directly by dynamic ray tracing. They are thus exact, and any limitations of the method due to the assumption that measured CRP-responses can be modelled by dynamic ray tracing along the zero-offset ray do not play a role here. An application to an example where the input data really have been extracted from seismic prestack data with the CRS Stack follows in the subsequent section.

1D test example

As a first simple test I consider the 1D case, i.e. no lateral velocity variation exists and all rays propagate in the vertical direction ($\alpha = 0$). Obviously, this significantly simplifies the forward modelling. Only two different data comonents, M and T, remain:

$$M = \left[\int_{s=s_0}^{s_1} v(s)\, ds\right]^{-1}, \quad T = \int_{s=s_0}^{s_1} \frac{1}{v(s)}\, ds. \tag{12}$$

The Fréchet derivatives (see Appendix A) can be obtained from the following simple expressions (z is defined positive upwards):

$$\begin{aligned} \Delta M &= \left[v(s_0)\Delta z - \int_{s=s_0}^{s_1} \Delta v(s)\, ds\right] \left[\int_{s=s_0}^{s_1} v(s)\, ds\right]^{-2} \\ \Delta T &= \frac{-\Delta z}{v(s_0)} - \int_{s=s_0}^{s_1} \frac{\Delta v(s)}{v^2(s)}\, ds \end{aligned} \tag{13}$$

I apply the 1D tomographic inversion to data that have been calculated in a model with layers of constant velocity (solid line in Fig. 3(a)). The 13 velocity discontinuities act as reflectors and the input data are

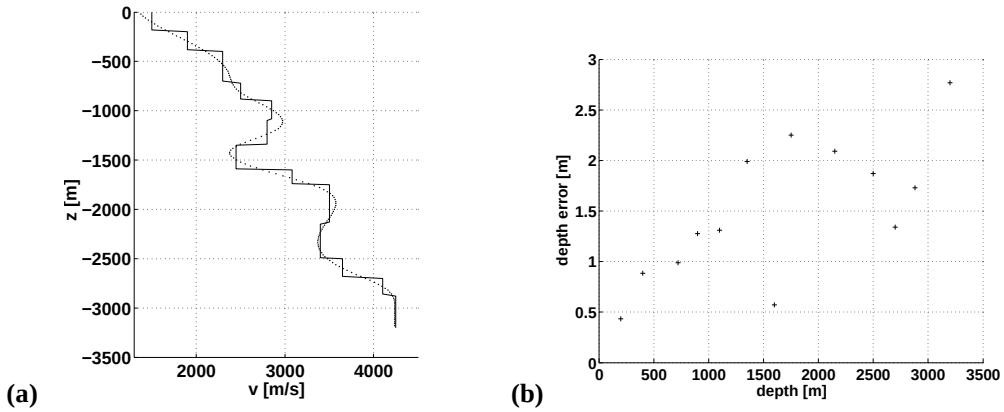
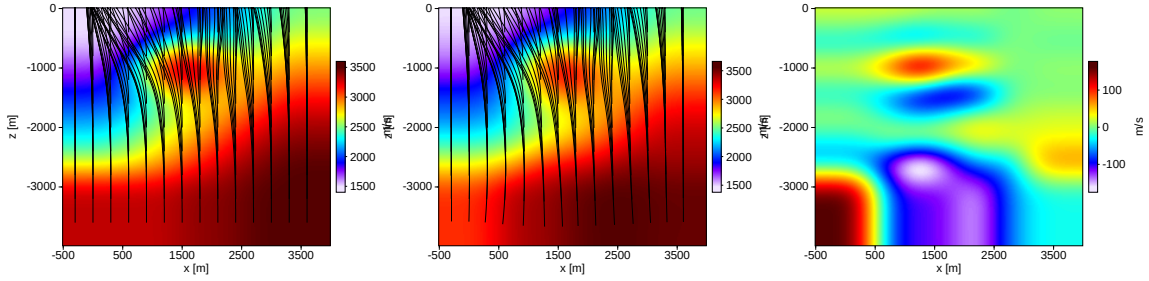


Figure 3: 1D inversion test. (a) The solid line indicates the original layered velocity model, while the dotted line gives the smooth velocity inversion result. See text for details. (b) The absolute value of the depth error between real and inverted reflector depths remains below 3 m. This is a relative depth error of less than 0.1 % for the deepest reflector.



(a) original model **(b)** inverted model **(c)** difference

Figure 4: 2D inversion test. (a) Original smooth velocity model described by 56 B-spline coefficients. For this test 126 data points are modelled directly by dynamic ray tracing along the shown rays. The initial ray direction (vertical) is the same for all rays. (b) Inverted velocity model. The velocity model and all rays are well reconstructed. (c) Difference between the original and the inverted velocity models.

calculated with simple wavefront curvature propagation laws, as given ,e.g., in Hubral and Krey (1980).

The model is described by B-spline coefficients at 15 knot positions with a 200 m spacing in the vertical direction (in 1D, cubic B-splines are used, as no third derivatives are needed). The problem is thus obviously underdetermined, therefore the regularisation term in the cost function (11) is essential and ϵ must be chosen accordingly.

The inversion is started with a velocity at the surface of 1500 m/s and a vertical gradient of 2 s^{-1}. After 12 iterations the velocity distribution given by the dotted line in Fig. 3(a) is obtained (stopping after fewer iterations gives a similar result). It agrees very well with the exact, discontinuous velocity distribution. A comparison of the true and inverted reflector depths (Fig. 3(b)) also shows very good agreement. The maximum depth error amounts to less than 3 m at a depth of more than 3000 m.

2D test example

The ability of the inversion algorithm to handle laterally varying velocity distriubitons is demonstrated with the model given in Fig. 4(a). It is described by 7×8 B-spline coefficients with a horizontal spacing of 500 m and a vertical spacing of 400 m. 126 data points are generated directly by dynamic ray tracing. For this test the ray starting points are distributed regularly in the subsurface with a horizontal interval of 300 m

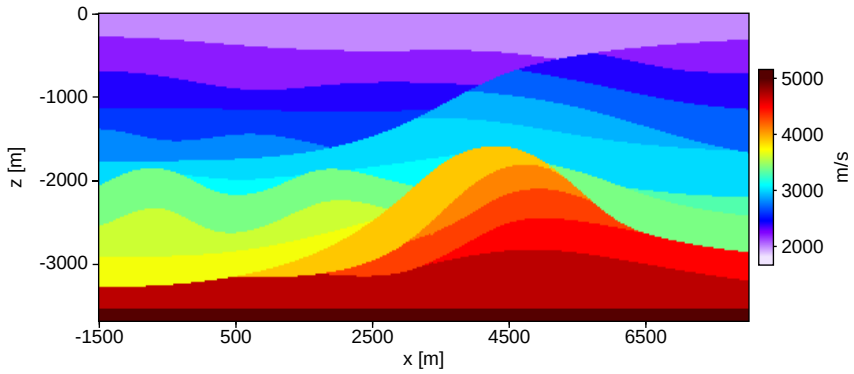


Figure 5: Blocky velocity model used to produce prestack data by ray tracing modelling.

and a vertical interval of 400 m. The initial ray direction is vertical for all rays.

Of course, in a velocity model of this degree of high amplitude short-wavelength lateral variation, the traveltimes associated with a common reflection point in the subsurface can be described by the parameters (T, M, α, ξ) only for small offsets. In general, the moveout curve is non-hyperbolic for larger offsets and CRS Stack results obtained from prestack data in such a model would not be reliable for inversion. Still the model presented here is useful to test the inversion algorithm itself, given perfect input data.

The data associated with the 126 modelled rays are directly used as input for the inversion. A start model with a near-surface velocity of 2000 m/s and a vertical velocity gradient of 0.5 s^{-1} is used. The inversion result after 15 iterations is displayed in Fig. 4(b). Except for some differences in the lower left part of the model, the velocity distribution and ray trajectories have been very well retrieved. Figure 4(c) shows the difference section between the modelled and the inverted velocity distributions.

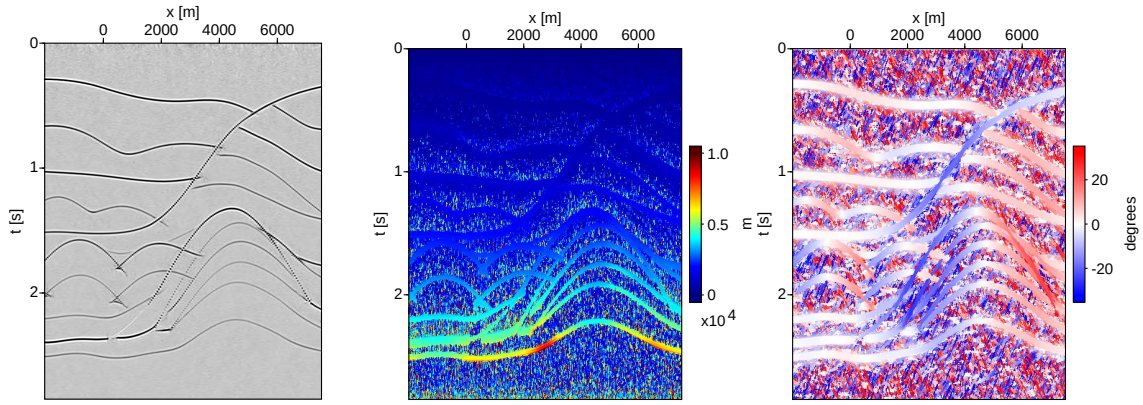
A 2D SYNTHETIC DATA EXAMPLE

Having demonstrated the applicability of the described inversion algorithm in 1D and 2D I will now present a more realistic synthetic data example. Seismic prestack data have been modelled by ray tracing in the blocky velocity model given in Fig. 5 in a marine acquisition geometry. The shot and receiver intervals were 50 m and the maximum offset was 2000 m, which leads to a CMP fold of 20. Also, random noise was added to the data. This multi-coverage seismic dataset served as the starting point for the construction of a smooth velocity model based on the application of the CRS Stack and subsequent tomographic velocity inversion using CRS attributes, as described in the previous sections.

As a first step, the CRS Stack was applied to the multi-coverage dataset, resulting in the sections displayed in Figure 6, along with an R_{N} section and a coherence section (not displayed). 505 data points were then picked in the CRS Stack section, resulting in a set of values (T, M, α, ξ). These data served as the input for the tomographic velocity inversion.

The velocity model is defined by 15 $\times$ 13 B-spline coefficients on a grid with a spacing of 500 m in the horizontal and 300 m in the vertical direction. The start model was chosen to consist of a near-surface velocity of 2000 m/s and a vertical velocity gradient of 2/3 s^{-1}, which is in fact relatively close to the true average vertical velocity gradient (a useful simple start model can also be obtained by performing one inversion iteration, starting with a constant initial velocity, and choosing a very high value of ϵ).

The inversion was stopped after 12 iterations. The result, consisting of the velocity model itself and the reconstructed normal ray trajectories, is displayed in Figures 7(a) and (b). In most parts the reconstructed model resembles a smoothed version of the true discontinuous velocity distribution. Locations where it deviates from that, e.g. in the lower part of the model, especially around x = 3000 m, can be correlated with low ray coverage. The inverted velocity model is kinematically correct, i.e. local reflector elements or dip bars, which can be attached to the end points of reconstructed normal rays, are placed in the correct subsurface positions. This is demonstrated in Figure 7(c), where the inverted dip bars are plotted into the



(a) CRS Stack **(b)** R_{NIP} **(c)** α

Figure 6: 2D inversion test. (a) CRS Stack section obtained from synthetic multicoverage data modelled by ray tracing in the velocity model shown in Fig. 5. A maximum offset of 2000 m has been used and random noise has been added to the prestack data. Diffractions have not been modelled. (b) R_{NIP} section, (c) emergence angle α section.

original velocity model. In particular, the lowermost horizontal reflector is well reconstructed.

In linear inversion problems the effects of errors or noise in the input data on the inversion result can be determined by the calculation of the posterior model covariance matrix (e.g. Tarantola, 1987), given an estimate of the data uncertainties in the data covariance matrix. In nonlinear problems the effect of data errors on the solution is not as easy to determine. To examine the sensitivity of the velocity inversion to noise, Gaussian noise with the following standard deviations was added to the picked input data:

$$\begin{aligned} \sigma_T &= 5 \cdot 10^{-3}\ \mathrm{s} \quad (\sigma_{t_0} = 10^{-2}\ \mathrm{s})\ , \\ \sigma_M &= 10^{-8}\ \mathrm{s/m^2} \quad (\sigma_t \approx 10^{-2}\ \mathrm{s} \quad \text{at 2000 m offset})\ , \\ \sigma_\alpha &= 1^\circ\ , \\ \sigma_\xi &= 10\ \mathrm{m}\ . \end{aligned} \tag{14}$$

The value of σ_M was chosen based on the assumed standard deviation of traveltimes at 2000 m offset and using a parabolic traveltime moveout approximation, which depends linearly on M. The inversion was then performed with a number of different realisations of the Gaussian noise added to the data. The resulting inverted CRP locations (end points of normal rays) for five realisations of noise are displayed together in Fig. 8. As might be expected, the scatter of the inverted CRP positions increases with depth, but the overall reflector structure is well reconstructed. For the examined example, I thus find that the tomographic inversion is stable with respect to noise in the input data, assuming a suitable value for the regularisation weight ϵ has been chosen.

DISCUSSION

I will now briefly discuss some of the advantages and limitations of the tomographic velocity inversion introduced and tested in the previous sections. Firstly, I would like to note that the required input, emergence angles and NIP wave curvatures at different zero-offset locations, is a by-product of the CRS Stack, which has originally been developed for the purpose of producing simulated zero-offset sections from prestack multi-coverage data.

Picking of input data, which is the weak point in conventional reflection tomography, is simplified significantly by a number of features of the inversion method presented here:

- As already pointed out, data points are picked in a simulated zero-offset section, which has a much higher S/N ratio than the original prestack data. Structures can be much better identified in a zero-offset section and multiples, if present, avoided during picking.

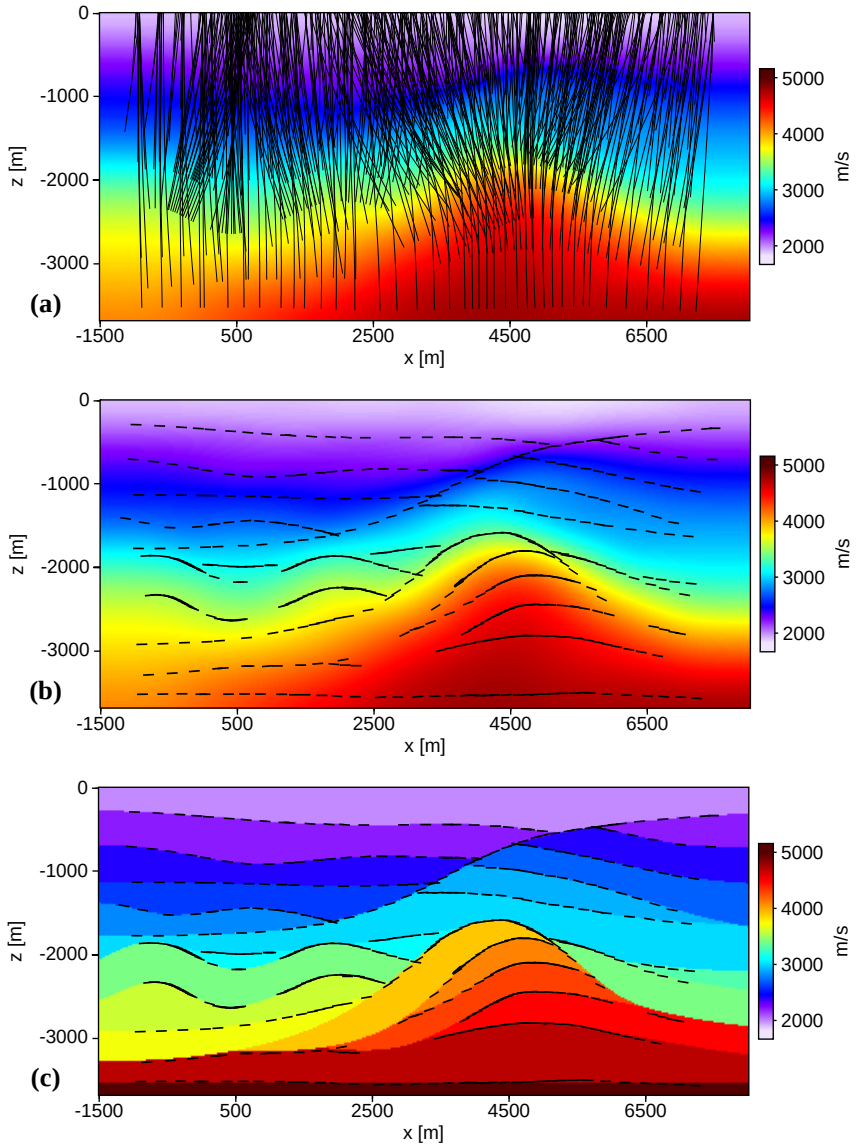


Figure 7: (a) Inverted smooth velocity model and inverted normal ray trajectories. (b) Inverted smooth velocity model and local dip bars calculated from inverted ray endpoints. (c) Inverted dip bars superimposed onto the original velocity model.

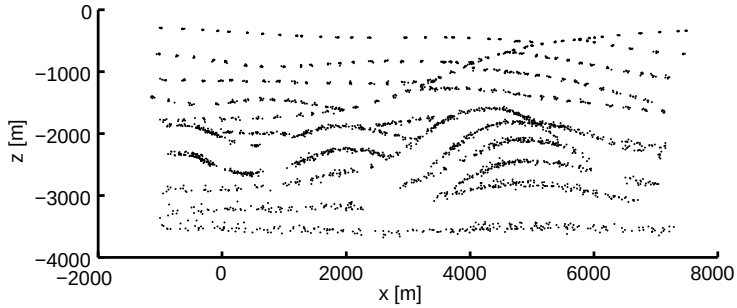


Figure 8: Sensitivity of the inversion to noise. Plot of inverted CRP locations for different noisy input data. Five realisations of Gaussian noise were added to the picked data prior to the inversion. See text for details.

- Each pick already represents the approximate traveltimes of a CRP for different offsets, therefore no picking of data points for different offsets is necessary.

- In a smooth velocity model without discontinuities, as is used here, reflection points do not define the position of velocity block or layer boundaries. Therefore it is not necessary to pick successive samples along a reflector. In fact picked events may lie on a diffraction as well as on a reflection event.

The approximations made during the CRS attribute-based tomographic inversion, namely the description of entire CRP responses with quantities calculated along the normal ray, result in some restrictions for the applicability of the method to complex subsurface velocity distributions. While the method has been designed for the inversion of laterally varying velocities, there is a limit on the minimum wavelength of lateral variation to ensure approximately hyperbolic traveltime moveout within the CRS aperture. This is in accordance with the smoothness criterion imposed on the data by the regularisation during the inversion.

The fulfilment of this requirement can be checked by inspection of a number of CMP gathers in the prestack data. If necessary, the maximum offset used during the CRS Stack should be restricted. A too small aperture, on the other hand, will result in a degraded resolution of the CRS attribute determination, as is well known from conventional stacking velocity analysis.

During the CRS Stack the kinematic wavefield attributes are determined by coherence analysis from a large number of traces with various offsets around the zero-offset trace at ξ. If the subsurface properties are laterally heterogeneous, the optimum determined CRS attributes represent spatial averages and cannot strictly be attributed to the zero-offset ray alone. This may be taken into account in the inversion by using locally averaged values of the higher derivatives of velocity during dynamic ray tracing and the application of ray perturbation theory. Although such a procedure is heuristic, it may broaden the applicability of the inversion to velocity distributions of stronger lateral variation.

The smooth velocity model obtained by the inversion is optimised for depth imaging. The velocities between successive reflection events cannot necessarily be interpreted as interval velocities in geological terms.

CONCLUSIONS

I have presented a new tomographic velocity model inversion method based on kinematic wavefield attributes extracted from the prestack seismic data with the CRS Stack. The method makes use of the CRS attributes R_{NIP} and α, which are suited to approximately describe the kinematic multi-offset response of a common reflection point in the subsurface in the vicinity of the zero-offset ray. The model parametrisation – smooth velocity and reflection point positions and angles corresponding to the input data points – allows to consider each data point independently during picking, without having to follow continuous reflectors. Picking of the input data is performed in a simulated zero-offset section obtained with the CRS Stack.

The inversion algorithm has been tested on 1D and 2D synthetic examples where the input data have been calculated directly by exact forward modelling. The entire inversion procedure, starting with the CRS Stack, followed by picking and the subsequent application of the tomographic inversion has been demonstrated on a 2D synthetic data example involving prestack seismic data modelled in a laterally heterogeneous velocity model. The results are very encouraging, a kinematically correct smooth version of the original velocity model can be reconstructed.

Although the subsurface velocity distribution may not be arbitrarily complex (e.g. complicated salt bodies) the method should be applicable to a wide range of situations arising in seismic exploration. In very complex environments the inversion results may provide a useful start model for detailed migration-based velocity analysis.

A generalisation to the 3D case is possible, the input data for the inversion are then taken from the results of the 3D CRS Stack.

ACKNOWLEDGEMENTS

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

REFERENCES

Bergler, S., Hubral, P., Marchetti, P., Cristini, A., and Cardone, G. (2002). 3D common-reflection-surface stack and kinematic wavefield attributes. *The Leading Edge*, 21(10):1010–1015.

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

Biloti, R., Santos, L. T., and Tygel, M. (2002). Multiparametric traveltime inversion. *Stud. geophys. geod.*, 46:177–192.

Bishop, T., Bube, K., Cutler, R., Langan, R., Love, P., Resnick, J., Shuey, R., Spindler, D., and Wyld, H. (1985). Tomographic determination of velocity and depth in laterally varying media. *Geophysics*, 50(1):903–923.

Červený, V. (2001). *Seismic Ray Theory*. Cambridge University Press.

de Boor, C. (1978). *A practical guide to splines*. Springer-Verlag.

Deregowski, S. M. (1990). Common-offset migrations and velocity analysis. *First Break*, 08(06):224–234.

Dix, C. H. (1955). Seismic velocities from surface measurements. *Geophysics*, 20(1):68–86.

Duveneck, E. and Hubral, P. (2002). Tomographic velocity model inversion using kinematic wavefield attributes. In *72nd Ann. Internat. Mtg*. Soc. Expl. Geophys.

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

Farra, V. and Madariaga, R. (1988). Non-linear reflection tomography. *Geophysical Journal*, 95:135–147.

Höcht, G., de Bazelaire, E., Majer, P., and Hubral, P. (1999). Seismics and optics: hyperbolae and curvatures. *J. Appl. Geoph.*, 42(3,4):261–281.

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

Hubral, P. and Krey, T. (1980). *Interval velocities from seismic reflection traveltime measurements*. Soc. Expl. Geophys.

Jäger, R., Mann, J., Höcht, G., and Hubral, P. (2001). Common-reflection-surface stack: Image and attributes. *Geophysics*, 66(1):97–109.

Jeannot, J. P. and Faye, J. P. (1986). Prestack migration velocities from focusing depth analysis. In *56nd Ann. Internat. Mtg*, pages 438–440. Soc. of Expl. Geophys.

Lines, L. R. and Treitel, S. (1984). Tutorial: A review of least-squares inversion and its application to geophysical problems. *Geophys. Prosp.*, 32:159–186.

Liu, Z. (1997). An analytical approach to migration velocity analysis. *Geophysics*, 62(04):1238–1249.

MacKay, S. and Abma, R. (1992). Imaging and velocity estimation with depth-focusing analysis. *Geophysics*, 57(12):1608–1622.

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.

Nolet (1987). Seismic wave propagation and seismic tomography. In Nolet, G., editor, *Seismic Tomography with Applications in Global Seismology and Exploration Geophysics*. Reidel Publishing.

Nowack, R. L. and Lutter, W. J. (1988). Linearized rays, amplitude and inversion. *PAGEOPH*, 128(1,2):401–421.

Ory, J. and Pratt, R. G. (1995). Are our parameter estimators biased? The significance of finite-difference regularization operators. *Inverse Problems*, 11:397–424.

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.

Shah, P. M. (1973). Use of wavefront curvature to relate seismic data with subsurface parameters. *Geophysics*, 38(5):812–825.

Stork, C. (1992). Reflection tomography in the postmigrated domain. *Geophysics*, 57(5):680–692.

Tarantola, A. (1987). *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation*. Elsevier, Amsterdam.

APPENDIX A

Calculation of the Fréchet derivatives

For the solution of the inverse problem the elements of the matrix $\mathbf{F}$, the Fréchet derivatives $\frac{\partial(T,M,\alpha,\xi)}{\partial(x,z,\theta,v)}$, are needed. For their calculation I make use of ray perturbation theory as described by Farra and Madariaga (1987). As the calculation of M and its derivatives is significantly simplified by the use of ray-centered coordinates (in 2D: q, s) (e.g. Červený, 2001), ray perturbation theory will be applied in these coordinates (s is the arclength along the ray and q is the coordinate normal to the ray direction). The results will then be transformed into cartesian coordinates and angles at the ray starting and end points. Perturbations of the coordinate q and its corresponding slowness component p at the position s_1 along the ray can be related to perturbations Δq and Δp at s_0 and to perturbations of the velocity Δv along the ray by:

$$\begin{pmatrix} \Delta q_1 \\ \Delta p_1 \end{pmatrix} = \mathbf{\Pi}(s_1, s_0) \begin{pmatrix} \Delta q_0 \\ \Delta p_0 \end{pmatrix} + \int_{s_0}^{s_1} \mathbf{\Pi}(s_1, s) \Delta\mathbf{B}(\Delta v, s)\, ds \tag{A1}$$

where $\Delta\mathbf{B} = (0, v_q \Delta v/v^3 - \Delta v_q/v^2)^T$, an index q denotes a partial derivative with respect to q, $\Delta q_1 := \Delta q(s_1)$, etc., and

$$\mathbf{\Pi} = \begin{pmatrix} Q_1 & Q_2 \\ P_1 & P_2 \end{pmatrix} \tag{A2}$$

is the ray propagator matrix. Its elements are solutions of the dynamic ray tracing system for normalised plane wave (Q_1, P_1) and normalised point source (Q_2, P_2) initial conditions, respectively (Červený, 2001):

$$\frac{d}{ds}\mathbf{\Pi} = \begin{pmatrix} 0 & v \\ \frac{1}{v^2} v_{qq} & 0 \end{pmatrix} \mathbf{\Pi}\ . \tag{A3}$$

$\mathbf{\Pi}(s_1, s) = \mathbf{\Pi}(s_1, s_0)\mathbf{\Pi}^{-1}(s, s_0)$, so that (A1) can easily be evaluated during ray tracing. $M = P_2/Q_2$, therefore ΔM is directly related to perturbations of the elements of $\mathbf{\Pi}$:

$$\Delta M = \frac{\Delta P_2}{Q_2} - \frac{P_2}{Q_2^2} \Delta Q_2\ . \tag{A4}$$

These can be linearly related to perturbations $\Delta q, \Delta p$ and Δv along the ray, as shown by Farra and Madariaga (1987):

$$\Delta\mathbf{\Pi} = \begin{pmatrix} \Delta Q_1 & \Delta Q_2 \\ \Delta P_1 & \Delta P_2 \end{pmatrix} = \int_{s_0}^{s_1} \mathbf{\Pi}(s_1, s) \Delta\mathbf{S}(s) \mathbf{\Pi}(s, s_0)\, ds\ , \tag{A5}$$

where $\Delta\mathbf{S}$ is given by $\Delta\mathbf{S} = \Delta\mathbf{S}_1(\Delta v) + \Delta\mathbf{S}_2(\Delta q, \Delta p)$. Evaluated in ray-centered coordinates in 2D, their expressions for $\Delta\mathbf{S}_1$ and $\Delta\mathbf{S}_2$ become (see also: Nowack and Lutter, 1988):

$$\Delta\mathbf{S}_1(\Delta v) = \begin{pmatrix} 0 & \Delta v \\ \frac{2}{v^3}(v_{qq}\Delta v + v_q \Delta v_q - \frac{\Delta v}{v} v_q^2 - \frac{v}{2}\Delta v_{qq}) & 0 \end{pmatrix} \tag{A6}$$

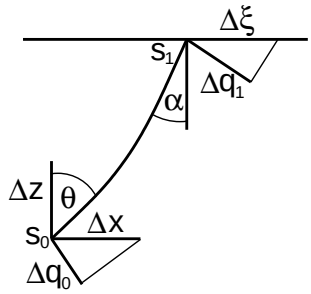


Figure 9: Geometrical relation between perturbations of the ray endpoints in ray-centered coordinates and cartesian coordinates

and

$$\Delta\mathbf{S}_2(\Delta q,\Delta p)=\begin{pmatrix} 2v_q\Delta p & 2v_q\Delta q \\ (\frac{3}{v^3}v_q v_{qq}-\frac{1}{v^2}v_{qqq})\Delta q & -2v_q\Delta p \end{pmatrix}, \tag{A7}$$

where Δq and Δp are related to Δq_0 and Δp_0 through the ray propagator matrix (A2). Obviously, smooth third spatial derivatives of the velocity are needed in (A7). Also required is $\Delta\mathbf{\Pi}(\Delta s_0)$, which can be calculated from the dynamic ray tracing system:

$$\Delta\mathbf{\Pi}(\Delta s_0)=\mathbf{\Pi}\begin{pmatrix} 0 & -v\Delta s_0 \\ -\frac{v_{qq}}{v^2}\Delta s_0 & 0 \end{pmatrix}. \tag{A8}$$

From geometrical considerations (Fig. 9) I find that:

$$\begin{array}{ll} \Delta q_0=\cos(\theta)\Delta x-\sin(\theta)\Delta z & \Delta\xi=\frac{1}{\cos(\alpha)}\Delta q_1 \\ \Delta p_0\approx\frac{1}{v(s_0)}\Delta\theta & \Delta\alpha\approx v(s_1)\Delta p_1 \\ \Delta s_0=\sin(\theta)\Delta x+\cos(\theta)\Delta z\ . & \end{array} \tag{A9}$$

Using the results (A1) to (A8) the quantities $\frac{\partial M}{\partial q_0}$, $\frac{\partial M}{\partial p_0}$, $\frac{\partial M}{\partial s_0}$, $\frac{\partial M}{\partial v}$, $\frac{\partial q_1}{\partial v}$, and $\frac{\partial p_1}{\partial v}$ can be calculated. Together with (A9), these can be used to obtain the required Fréchet derivatives involving M, α and ξ:

$$\begin{array}{lll} \frac{\partial M}{\partial x}=\cos(\theta)\frac{\partial M}{\partial q_0}+\sin(\theta)\frac{\partial M}{\partial s_0} & \frac{\partial\alpha}{\partial x}=v(s_1)\cos(\theta)P_1 & \frac{\partial\xi}{\partial x}=\frac{\cos(\theta)}{\cos(\alpha)}Q_1 \\ \frac{\partial M}{\partial z}=-\sin(\theta)\frac{\partial M}{\partial q_0}+\cos(\theta)\frac{\partial M}{\partial s_0} & \frac{\partial\alpha}{\partial z}=-v(s_1)\sin(\theta)P_1 & \frac{\partial\xi}{\partial z}=-\frac{\sin(\theta)}{\cos(\alpha)}Q_1 \\ \frac{\partial M}{\partial\theta}=\frac{1}{v(s_0)}\frac{\partial M}{\partial p_0} & \frac{\partial\alpha}{\partial\theta}=\frac{v(s_1)}{v(s_0)}P_2 & \frac{\partial\xi}{\partial\theta}=\frac{1}{v(s_0)\cos(\theta)}Q_2 \\ \frac{\partial M}{\partial v}=\text{directly from (A4) - (A6)} & \frac{\partial\alpha}{\partial v}=v(s_1)\frac{\partial p_1}{\partial v} & \frac{\partial\xi}{\partial v}=\frac{1}{\cos(\alpha)}\frac{\partial q_1}{\partial v}\ . \end{array} \tag{A10}$$

The Fréchet derivatives involving T follow from:

$$\Delta T=-\frac{\sin(\theta)}{v(s_0)}\Delta x-\frac{\cos(\theta)}{v(s_0)}\Delta z-\int_{s_0}^{s_1}\frac{\Delta v(s)}{v^2(s)}\,ds\ . \tag{A11}$$

These expressions need to be evaluated for each ray pertaining to the input data and each B-spline knot of the velocity model. For a given B-spline knot (j,k) the velocity perturbation is $\Delta v(x,z)=\beta_j(x)\beta_k(z)$.

APPENDIX B

Regularisation of the tomographic matrix

As the matrix $\mathbf{F}$ is in general ill-conditioned, the inversion problem needs to be regularised by introducing additional constraints on the model parameters. A physically sensible way of doing this is to require the velocity model to have minimum curvature, i.e. minimum second derivatives, as we are looking for a smooth model with no artificial structure. This minimum curvature condition is often implemented with

the use of second-derivative finite-difference operators applied to the model parameters. To ensure that the solution of the inverse problem is as much as possible independent of the model discretisation (i.e. the B-spline knot interval) I impose the minimum curvature condition not on the velocity model parameters (the B-spline coefficients), but directly on the smooth model itself.

In the 1D case the second spatial derivative of the velocity, described by B-splines, is:

$$\frac{\partial^2 v(z)}{\partial z^2} = \sum_j v_j \frac{\partial^2 \beta_j(z)}{\partial z^2} \,. \tag{B1}$$

Its norm can be given by

$$\left\| \frac{\partial^2 v}{\partial z^2} \right\|_2^2 = \int_z \left(\frac{\partial^2 v(z)}{\partial z^2} \right)^2 dz = \int_z \sum_{ij} v_i v_j \frac{\partial^2 \beta_i(z)}{\partial z^2} \frac{\partial^2 \beta_j(z)}{\partial z^2} dz = \mathbf{v}^T \tilde{\mathbf{D}}^{zz} \mathbf{v} \,, \tag{B2}$$

where

$$\tilde{D}_{ij}^{zz} = \int_z \frac{\partial^2 \beta_i(z)}{\partial z^2} \frac{\partial^2 \beta_j(z)}{\partial z^2} dz \,. \tag{B3}$$

If for the regularisation not only second derivatives of v are considered, but also v itself, I obtain:

$$\int_z \epsilon_1 \left(\frac{\partial^2 v(z)}{\partial z^2} \right)^2 + \epsilon_2 \, v^2(z) \, dz = \mathbf{v}^T \tilde{\mathbf{D}}'' \mathbf{v} \tag{B4}$$

with $\tilde{\mathbf{D}}'' = \epsilon_1 \, \tilde{\mathbf{D}}^{zz} + \epsilon_2 \, \tilde{\mathbf{D}}$ and $\tilde{D}_{ij} = \int_z \beta_i(z)\beta_j(z)dz$. (B4) may then be considered a norm squared for v if $\epsilon_2 \neq 0$ (otherwise it is a seminorm). $\epsilon_2 \neq 0$ is necessary, as in the course of further calculations we will need to find a matrix $\tilde{\mathbf{B}}$ with $\tilde{\mathbf{B}}^T \tilde{\mathbf{B}} = \tilde{\mathbf{D}}''$, which is only possible if $\tilde{\mathbf{D}}''$ is positive definite; $\tilde{\mathbf{D}}^{zz}$ alone is not positive definite.

This can be generalised to the 2D case, where the following expression is included in the cost function:

$$\int_x \int_z \epsilon_1 \left(\frac{\partial^2 v(x,z)}{\partial z^2} \right)^2 + \epsilon_2 \left(\frac{\partial^2 v(x,z)}{\partial x^2} \right)^2 + \epsilon_3 \, v^2(x,z) \, dz \, dx = \mathbf{m}_{(v)}^T \mathbf{D}'' \mathbf{m}_{(v)} . \tag{B5}$$

Here $\mathbf{D}'' = \epsilon_1 \, \mathbf{D}^{xx} + \epsilon_2 \, \mathbf{D}^{zz} + \epsilon_3 \, \mathbf{D}$ and the matrices $\mathbf{D}^{xx}$, $\mathbf{D}^{zz}$, and $\mathbf{D}$ contain products of the elements of $\tilde{\mathbf{D}}^{zz}$ and $\tilde{\mathbf{D}}$ given above and of the corresponding matrix $\tilde{\mathbf{D}}^{xx}$. $\mathbf{m}_{(v)}$ is the part of the model parameter vector $\mathbf{m}$ that contains the B-spline coefficients v_{jk}. The factors ϵ_1, ϵ_2, and ϵ_3 are used for normalisation and to balance the contributions of the different terms. The term $\epsilon_3 \, \mathbf{D}$ should be much smaller than those containing derivatives of velocity, as there is no physical reason for minimising the velocity itself.

The minimum curvature constraints are to be applied to the model itself, not to the model update (e.g. Ory and Pratt, 1995). The cost function is then given by:

$$2 \, S(\mathbf{m}) = \Delta \mathbf{d}^T(\mathbf{m}) \mathbf{C}_D^{-1} \Delta \mathbf{d}(\mathbf{m}) + \epsilon \, \mathbf{m}_{(v)}^T \mathbf{D}'' \mathbf{m}_{(v)} \,, \tag{B6}$$

where $\Delta \mathbf{d}(\mathbf{m}) = (\mathbf{d} - \mathbf{f}(\mathbf{m}))$ and $\mathbf{C}_D^{-1}$ is a diagonal matrix effectively weighting the different data components (the data covariance matrix) (Tarantola, 1987), and $\epsilon > 0$. Assume that around the current model $\mathbf{m}_n$ the forward modelling operator $\mathbf{f}(\mathbf{m})$ can be approximated by $\mathbf{f}(\mathbf{m}) \approx \mathbf{f}(\mathbf{m}_n) + \mathbf{F}\Delta\mathbf{m}$, where $\mathbf{m} = \mathbf{m}_n + \Delta\mathbf{m}$ and $\mathbf{F}$ contains the Fréchet derivatives of $\mathbf{f}$ at $\mathbf{m}_n$, we obtain

$$\begin{aligned} \nabla_{\mathbf{m}} S &= -\mathbf{F}^T \mathbf{C}_D^{-1} (\mathbf{d} - \mathbf{f}(\mathbf{m})) + \epsilon \mathbf{D}'' \mathbf{m}_{(v)} \\ &\approx -\mathbf{F}^T \mathbf{C}_D^{-1} \Delta \mathbf{d}(\mathbf{m}_n) + \mathbf{F}^T \mathbf{C}_D^{-1} \mathbf{F} \Delta \mathbf{m} + \epsilon \mathbf{D}'' \left(\mathbf{m}_{(v)\,n} + \Delta \mathbf{m}_{(v)} \right) \\ &\stackrel{!}{=} \mathbf{0} \,. \end{aligned} \tag{B7}$$

It can be easily shown that

$$\nabla_{\mathbf{m}} S = \mathbf{0} \;\; \Leftrightarrow \;\; \hat{\mathbf{F}}^T \hat{\mathbf{F}} \Delta \mathbf{m} = \hat{\mathbf{F}}^T \Delta \hat{\mathbf{d}} \tag{B8}$$

with

$$\hat{\mathbf{F}} = \begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \mathbf{F} \\ [\mathbf{0} \,, \mathbf{B}] \end{pmatrix} \qquad \Delta \hat{\mathbf{d}} = \begin{pmatrix} \mathbf{C}_D^{-\frac{1}{2}} \Delta \mathbf{d}(\mathbf{m}_n) \\ -[\mathbf{0} \,, \mathbf{B}] \, \mathbf{m}_n \end{pmatrix} , \tag{B9}$$

where $\mathbf{B}^T \mathbf{B} = \epsilon \, \mathbf{D}''$. The searched-for model update $\Delta\mathbf{m}$ is therefore the least-squares solution to $\hat{\mathbf{F}}\Delta\mathbf{m} = \Delta\hat{\mathbf{d}}$. The updated model components are then given by $\mathbf{m}_{n+1} = \mathbf{m}_n + \lambda \, \Delta\mathbf{m}$, where $0 < \lambda \leq 1$.