Velocity macro-model estimation from seismic reflection data by stereotomography

Frédéric Billette and Gilles Lambaré
Centre de Recherche en Géophysique, 35 rue Saint Honoré, 77305 Fontainebleau Cédex, France. E-mail: billette@geophy.ensmp.fr

Accepted 1998 May 8. Received 1998 April 28; in original form 1997 December 30

SUMMARY
We introduce a new tomographic method for estimating velocity macro-models from seismic reflection data. In addition to traveltimes picked on locally coherent reflected events, the method requires that the associated local slopes of the events be picked simultaneously in the common-shot and common-receiver trace gathers. The data then consist of a discrete collection of traveltimes, positions and slopes for selected reflected events. Unlike traveltime tomography, picked events are only required to be locally coherent. It is not necessary to follow continuous arrivals all over the trace gathers. Indeed, the method does not require the introduction of interfaces in the model description.

Several approaches of tomography using the slope have already been proposed. We present a unified formulation for slope tomography methods, in which the model is described by the velocity field and a set of ray-segment pairs associated with the reflected/diffracted events. We propose a new robust slope tomography method, which we call ‘stereotomography’. It consists of fitting all observed data (positions, slopes and traveltimes) to data calculated by ray tracing. There are no theoretical limitations in stereotomography for laterally heterogeneous velocity macro-models.

Practically, traveltimes and slopes are picked on local slant stack panels. Ray multipathing can be accounted for since paths are discriminated by their associated slopes. The non-linear inverse problem is iteratively resolved by a local optimization. The Fréchet derivatives are estimated by paraxial ray tracing.

Validation tests on 1-D and 2-D synthetic data are analysed. In the first 1-D example, we study the sensitivity of the method to model parameters (using a singular-value decomposition). The second 1-D example evaluates picking precision and shows that it is sufficient for constraining the velocity field. The last example is a 2-D application in which data are calculated directly by ray tracing. It shows the performance of the method in the presence of strong lateral velocity variations.

Key words: inverse problem, seismic ray theory, seismic reflection, seismic velocities, tomography.

INTRODUCTION
The determination of velocity macro-models is a crucial and unavoidable operation in seismic reflection imaging. The most commonly used method is velocity analysis (Dix 1955). This well-know approach relies on the hypothesis of a laterally homogeneous model (see Yilmaz 1987 for a review), and, although it can be used in gently laterally heterogeneous models, no simple extension to fully heterogeneous models has been proposed until now. In fact, the estimation of velocity macro-models still forms a subject for theoretical research. It is acknowledged that the difficulty comes from: (1) the under-determination of the problem, which implies that the velocity macro-model focusing the reflected events in depth at the correct positions can only be recovered if a priori information, such as sonic logs, is introduced; and (2) the strong non-linearity of the relationship that links the traces to the velocity macro-model (Farra & Madariaga 1988). This second problem has been addressed with various approaches.

First, we must mention reflection tomography (e.g. Bishop et al. 1985; Chiu & Stewart 1987; Farra & Madariaga 1988). In this approach, the model is described as a set of layers with smooth interval velocities and interfaces, and data consist of picked traveltimes for selected events. This so-called ‘blocky’ model is optimized by fitting the calculated traveltimes to the picked traveltimes. The local iterative non-linear optimization
of the velocity field may be made CPU-efficient even in 3-D (Guiziou, Mallet & Madariaga 1996). In reflection tomography, the underdetermination of the problem appears through the velocity–depth ambiguity (Williamson 1990; Stork 1992a; Tieman 1994). Furthermore, in practice, traveltome picking can be a difficult and fastidious operation, since picked events have to be identified all over the traces in the data set, even where the signal-to-noise ratio is very low, and interpreted in terms of particular reflectors in the model. Moreover, developing an efficient and robust ray-tracing algorithm devoted to such an application is difficult, especially in 3-D (Virieux & Farra 1991), and instabilities may arise in the optimization procedure in the case of complex models, e.g. triplets, diffractions (Chapman 1985; Amand & Virieux 1995; Charles 1996).

Alternative approaches have been proposed in order to avoid picking. They rely on the optimization of a coherency function on the traces. As an example, migration velocity analysis (Al-Yahya 1989; Symes & Carazzone 1991; Jin & Madariaga 1993, 1994; Docherty et al. 1997) is based on the assumption that, if the velocity macro-model is correct, each common-offset or common-shot depth migrated profile should provide the same migrated image. The coherency of these profiles can be evaluated and optimized. Although some authors have proposed other strategies, e.g. working directly in the data space rather than on the migrated sections (Landa et al. 1988; Biondi 1992; Plessix, Chavent & De Roeck 1995), migration velocity analysis seems to have established itself as the most common strategy. At the present time, despite this general agreement, the method is still penalized by difficult numerical implementations (for CPU efficiency it is generally based on ray tracing) and by discouraging computer requirements in 3-D. In this context, it appears profitable to try to preserve the advantages of the tomographic approach while improving its robustness in both optimization and picking procedures.

In traveltome tomography, instabilities are associated with singularities in ray tracing, e.g. multipathing, caustics. It is well known that such singularities can be unfolded by considering the ray field, not in the configuration space \((x)\), but in the phase space \((x, p)\) (Chapman 1985; Lambarè, Lucio & Hanyga 1996) where \(p = VT\) denotes the slowness vector. On a common-shot gather or common-receiver gather, the local slope of a reflected event provides a direct estimation of the horizontal component of the slowness vector (see Fig. 1). First, the local slope can be used as extra information in traveltome tomography for unfolding the events in the case of multipathing (Delprat-Jannaud & Lailly 1993; Guiziou et al. 1996). However, the tomographic problem can also be recast more deeply while using the slope, leading to what we call ‘slope tomographic methods’.

In transmission tomography, the use of the polarization vectors (related to the slowness vector) in addition to traveltimes has been developed (Menke 1984; Hu & Menke 1992; Farra & Le Béhat 1995; Yanovskaya 1996) to constrain better the velocity model.

In reflection tomography, the use of slope information also provides many advantages. It was initially proposed by Rieber (1936). Soviet geophysicists recognized the potential of this approach and developed the CDR (controlled directional reception) tomographic method (Riabinikin 1957; Riabinikin et al. 1962). The routine use of CDR for seismic exploration in the USSR intrigued American geophysicists, who went to Moscow to review the merits of the method (Hermont 1979). More recently, at Stanford University, the approach has been re-examined by Sword (1986, 1987). In his approach, the slopes are estimated for a given event in the data on both common-shot gathers and common-receiver gathers. Picking is performed on local slant stack panels and is consequently easier than picking on unstacked trace gathers (as is generally done in traveltime tomography). Picked data consist of a set of shot and receiver positions, associated slopes and two-way travel-times. There is no need to associate a given event with an interface in the model, which can be a smooth velocity field. Sword proposed various misfit functions for his tomographic problem, relying on a misfit in traveltime, position or slope.

![Figure 1](https://via.placeholder.com/150)

**Figure 1.** The slope on a common-shot gather and the slope in ray theory. On a common-shot gather or common-receiver gather, the local slope of the reflected events (left) gives a direct estimation of the horizontal component of the slowness vector (right).
A non-linear iterative local optimization was used. Some validation tests were performed, but the method suffered theoretically and practically from instabilities. In fact, further investigation had to be pursued to improve the method. At Stanford University, an extension of the CDR tomographic method was led by Biondi (1990, 1992) to design a fully automatic velocity estimation method. This no-picking application led to good results, but it was done at the expense of approximations regarding the complexity of the velocity field and an increased CPU time.

We assert that the tomographic CDR method is promising since it exhibits no theoretical limitations for application to complex velocity macro-models and since it involves reasonable computing time. In contrast to Biondi (1992), we believe that if any automatization is to be performed, it should be introduced at the picking stage, as is done in standard velocity analysis. Our work is devoted to the improvement of the CDR method, by avoiding some of the theoretical and practical instabilities of the original approach. This manner of imaging the Earth by looking in two directions at specific angles reminded us of other applications using a stereo view, such as the process by which the relief of the Earth is perceived by combining two photographs of the same landscape shot at different angles. Therefore, we have called our method 'stereotomography'.

In this paper, we recast Sword's original work (Sword 1987) in a global formulation of slope tomography. Stereotomography will be developed in the general frames of paraxial ray theory (Cerveny, Molotch & Psenick 1977; Farra & Madariaga 1987), Hamiltonian formulation of ray theory (Farra & Madariaga 1987; Lambare et al. 1996), and general inverse-problem theory (Tarantola 1987). We propose a new model description and misfit function, which should avoid some of the CDR instabilities during the non-linear local optimization process, and enable an extension to 3-D. Finally, for checking the picking accuracy and sensitivity of the method, we present three validation tests for 1-D and 2-D synthetic models.

**Why should we use the slope?**

There are various ways to answer this question. We propose the following simple scheme. Let us consider a locally coherent event on a common-shot gather and on a common-receiver gather. If it is a primary reflection or diffraction, it can be associated with a reflecting/diffracting point \( \mathbf{X} \). This point is the intersection of the two rays \( \mathbf{S} \rightarrow \mathbf{X} \) and \( \mathbf{X} \rightarrow \mathbf{R} \) (Fig. 1b).

In the trace gathers, we can get the values of the slopes of these two rays at the surface (Fig. 1a). These slopes correspond to the horizontal components of the slowness vectors of the rays at the surface (Fig. 1b).

Now, let us try to retrieve this reflecting/diffracting point \( \mathbf{X} \) starting from the surface. We consider an initial velocity macro-model. In this model, two rays are completely defined by the positions \( \mathbf{S} \) and \( \mathbf{R} \) and the horizontal components \( \mathbf{P}_s \) and \( \mathbf{P}_r \), \( (P_{sx}, P_{sy}, P_{rx}, P_{ry}) \) in 2-D, \( P_{sz}, P_{rx}, P_{ry} \) and \( P_{rz} \) in 3-D) of the associated slowness vectors \( \mathbf{p}_s \) and \( \mathbf{p}_r \). Both rays can be traced down from the surface. The rays are stopped when they reach each other. Did they cross exactly at \( \mathbf{X} \)? If the initial velocity macro-model was the true model, they did; if it was erroneous, they did not. How can we know if the crossing point is the true reflecting/diffracting point \( \mathbf{X} \)? The traveltime provides us with extra information. If the sum of the two-way traveltimes calculated when the two rays cross each other does not fit with the (tomographic) traveltime picked in the data, we can conclude that the velocity macro-model is erroneous. The misfit in traveltime can be linked to the misfit in velocity. This simple configuration shows how using the slopes allows us to constrain the velocity macro-model. It should be pointed out that no assumption has been made concerning the lateral heterogeneity of the velocity field and that no continuous interface has been supposed while introducing \( \mathbf{X} \), which can be any kind of reflector or diffractor.

**Data, models and misfit functions for slope tomography methods**

We have shown that using the slope constrains the velocity without having to introduce interfaces. Now, we shall discuss various slope tomographic methods. For these methods, as was the case in traveltime tomography, the data set is composed of a set of shot and receiver positions, \( \mathbf{S} \) and \( \mathbf{R} \), and traveltimes, \( T_{sr} \), but also the slopes, i.e. horizontal component of the slowness vector, at both receiver and shot locations, \( \mathbf{P}_s \) and \( \mathbf{P}_r \). The data space \( \mathbf{d} \) consists of a set of \( N \) picked values:

\[
d = [(\mathbf{S}, \mathbf{R}, \mathbf{P}_s, \mathbf{P}_r, T_{sr})]_n^{N-1}. \tag{1}
\]

In the exact velocity macro-model, each data pick is associated with a pair of ray segments \( \mathbf{S} \rightarrow \mathbf{X} \) and \( \mathbf{X} \rightarrow \mathbf{R} \). By a ray segment, we mean a truncated part of a ray trajectory, which can be totally defined by its starting or ending point, the initial or final direction and the traveltime. In the exact velocity model, there are boundary conditions for both ray segments. The following conditions are imposed on the two ray segments: (1) they cross each other at their ending (deepest) point; and (2) they fit the data in positions, slopes and two-way traveltime. When the velocity macro-model is erroneous, the two ray segments cannot satisfy all of the boundary conditions simultaneously. Then, at least one of the boundary conditions has to be ‘relaxed’ (become variable). The misfits on the parameters describing the relaxed boundary conditions are
used to constrain the velocity macro-model. We construct an inverse problem in which the model space is described by the velocity and the pairs of ray segments. The dimension of the ray-segment subspace depends on the number of non-relaxed boundary conditions.

In the most general approach, all of the parameters should be involved in the inverse problem, including the parameters describing the velocity macro-model and the boundary conditions describing the ray segments. Then, the model may be described by the velocity macro-model and a set of two ray segments, which do not have to join each other or fit the data (Fig. 2a). Solving the inverse problem would consist of adapting the velocity and ray segments until all of the boundary conditions fit the data (positions and slopes at the surface and two-way traveltime) and join each other at their end points in the subsurface.

Is it necessary to consider this global misfit function? We may decide to relax only a few parameters. For this, many strategies can be undertaken. In our previous simple scheme, boundary conditions were fixed at the surface (ray starting positions and slopes) and at depth (where the two ray segments had to join each other). The only relaxed boundary condition was the two-way traveltime, which was set as the relaxed parameter constraining the velocity macro-model.

For his CDR method, Sword (1987) proposed relaxing a single class of boundary conditions (receiver slope, the emerging position at the surface, the ray-crossing condition, etc.). He provided various model descriptions and misfit functions. For numerical considerations, his final approach was to fix the boundary conditions at the surface (positions and slopes) as well as the two-way traveltimes. Only the ray-crossing condition was relaxed. Then, he considered pairs of rays starting from the surface with initial conditions (positions and slopes) fixed by observed data (Fig. 2b), and propagating down in the

![Figure 2. Boundary conditions for the description of a pair of ray segments. In (a), no boundary condition is fixed: ray segments do not have to join each other in depth or fit the data at the surface. In (b), the surface boundary condition is fixed. The upper extremities of the two way segments have to fit the data at the surface. Their other extremities do not have to fit in depth. In (c), the crossing-point boundary is fixed. The two rays do not have to fit the data at the surface (stereotomography).](image)

a priori velocity macro-model. Each pair of rays was integrated with a constant depth step until the sum of the two one-way traveltimes was equal to the observed two-way travelt ime. If at this stage, the two rays did not join each other, since the velocity macro-model was not yet correct, a velocity perturbation was updated by iterative minimization of the horizontal distance between the last point of each ray, $\Sigma \| (x_{\text{err}}) \|^2$ (Fig. 3).

The main advantage of such an approach was that the model could finally be described simply by the velocity field, because the ray-pair parameters were completely defined by the fixed boundary conditions. Consequently, Sword remained close to the initial goal of the tomographic problem.

However, we claim that relaxing a single class of boundary conditions may lead to instabilities during the minimization scheme. For example, in the case of grazing rays, Sword’s criterion may be in calculable (problems with the depth step); and when rays are propagating in opposite directions, they may never cross and we may not compute the associated two-way time. This problem may often occur when applying the method to complex media, e.g. salt domes. Moreover, fixing the boundary conditions at the surface cannot take data measurement error into account, which cannot be reasonably set to zero, especially for the slopes. This is why another approach (in terms of misfit function), which is more general and robust, had to be introduced.

**Stereotomography**

Our goal is to construct a new method based on the same concept as the CDR method, but which will overcome its limitations. We present an innovative approach to slope tomography based on an original model parametrization and misfit function. Like the CDR method, our data set consists of a set of shot and receiver positions, $S$ and $R$, traveltimes, $T_{sr}$, and slopes at both receiver and shot locations, $P_r$ and $P_s$, picked on locally coherent events (Fig. 4). We suggest relaxing all of the boundary conditions of the ray segments at the surface (position, slope and two-way traveltime) and using a misfit

![Figure 3. Sword’s criterion for checking the velocity field. Sword considers a pair of rays starting from the surface with initial conditions (positions and slopes) fixed by observed data. They are integrated with a constant depth step until the sum of the two one-way traveltimes is equal to the observed one. The velocity perturbation is computed by minimization of $\Sigma \| (x_{\text{err}}) \|^2$.](image)
In stereotomography, these up-going pairs of rays are part of the model, which can no longer be described by the velocity field only. Both parts of the model will have to be updated in order to fit the data. The ray-segment parameters, also recovered by the inversion process, provide information on the distribution of the scattering positions and angles, which can be drawn as dip bars. This technique was used by Sword (1987) to provide migrated images composed of a set of dip bars, but where a 1-D hypothesis was introduced. This a posteriori information can be used to estimate the sampling of the subsurface.

In such a model, we can simulate data to compare to the data picked from trace gathers. The ending point of each up-going ray provides a position and a horizontal slowness, and the sum of the two one-way traveltimes provides a two-way traveltime. This operation has to be done for each datum. Our cost function contains misfits on the traveltimes and slopes, but also on source and receiver positions. In traveltime tomography, misfits on source and receiver positions are considered in the two-point ray tracing, whereas, in a second step, source and receiver positions are fixed for the velocity model optimization. In stereotomography, the same strategy could be implemented, but a joint inversion is expected to be more stable since it avoids the usual instabilities of two-point ray tracing (Hanyaga & Pajchel 1995). The state of our knowledge of data precision will be introduced in terms of measurement errors in the subsection on ‘a priori information’.

**FORWARD AND INVERSE PROBLEMS**

We address the problem of estimating the velocity macro-model in terms of a stochastic inverse problem (Tarantola 1987). The goal of our inverse problem is to find the model that best explains the observed data for a supposed physical relationship $g$. Data $d$ are linked to the model $m$ by a non-linear relationship

$$d = g(m).$$

In this section we develop the forward problem, which consists of the calculation of data by ray tracing and the resolution of the inverse problem by a local iterative optimization, which involves the estimation of Fréchet derivatives using paraxial ray tracing.

**Computation of data by ray tracing**

For a given model, $m$ (eq. 2), we must compute data (source and receiver positions, slopes and two-way traveltimes). In a current velocity model $C_m$, data can be calculated at the endpoints of the up-going ray segments starting from $X_0$, with initial angles $\Theta_0$ and $\Theta_r$, and propagating until the traveltimes $T_{src}$ and $T_{src}$ are both reached (see $d_{cal}$ in Fig. 4). The Hamiltonian formulation is often used to describe ray tracing (Chapman 1985; Farra & Madariaga 1987; Cerveny 1989). Let us introduce the Hamiltonian function (Lambare et al. 1996)

$$H(x, P, t) = \frac{1}{2} [P^2 c^2(x) - 1],$$

where $t$ denotes the time abscissa along the ray trajectory, $x$ the position, and $P$ the slowness vector such that $P = V(x)$ for ray trajectories. We chose to use the traveltime as the integration abscissa such that Fréchet derivatives will be given.
for \( T = \text{constant} \). Then, ray trajectories satisfy the canonical system

\[
\begin{align*}
\frac{\dot{x}}{\partial t} &= \nabla_x H = c^2(x)p, \\
\frac{\dot{p}}{\partial t} &= -\nabla_p H = -p^2c(x)\nabla c(x),
\end{align*}
\]

(5)

with the initial condition \( H(x, p, t) = 0 \) (i.e. \( p^2 = 1/c^2(x) \), Eikonal equation).

### Ray trajectories

\[
y(t) = \begin{pmatrix} x(t) \\ p(t) \end{pmatrix}
\]

can be simply integrated by a numerical approach. In practice, we use a second-order Runge–Kutta method, which directly provides calculated data \( d_{\text{cal}} \) (Fig. 4).

### Computation of Frechet derivatives by paraxial ray tracing

In addition to kinematic ray tracing, ray theory offers a powerful tool with dynamic, or so-called paraxial, ray tracing (Cerveny et al. 1977). It is used for many applications, such as two-point ray tracing, computation of amplitudes, perturbation of ray trajectories with respect to the velocity field (Farra & Madariaga 1987) and preserved amplitude migration (Thiery et al. 1996). In stereotomography, paraxial ray tracing is used to estimate the Frechet derivatives of the data with respect to model parameters.

Paraxial ray tracing gives first-order estimations of the ray trajectory perturbations with respect to initial condition perturbations \( \delta y(t_0) \) and perturbations of the velocity field \( \delta c(x) \). Each kind of perturbation induces a perturbation of the reference Hamiltonian eq. (4). The expression of the first-order perturbations of the ray parameters \( \delta y \) can be expressed along the ray using the propagator matrix method (Aki & Richards 1980) by

\[
\begin{align*}
\delta y(t) &= P(t, t_0) \delta y(t_0) + \int_{t_0}^{t} P(t, t') B(\delta c(x(t'))) dt',
\end{align*}
\]

(6)

where \( B(\delta c(x)) \) is a matrix that depends on the velocity perturbation (Farra & Madariaga 1987; Farra & Le Bégat 1995):

\[
B(\delta c(x)) = \begin{pmatrix} \nabla_x \delta H(\delta c(x)) \\ -\nabla_x \delta H(\delta c(x)) \end{pmatrix}.
\]

(7)

The propagator matrix \( P(t, t') \) is the Jacobian matrix

\[
\frac{\dot{P}(t, t')}{\partial t} = \frac{\delta y(t)}{\delta y(t')}.
\]

(8)

It satisfies the first-order differential system

\[
\frac{\dot{P}}{\partial t} = \begin{pmatrix} \nabla_x \nabla_x H & \nabla_x \nabla_p H \\ -\nabla_x \nabla_p H & -\nabla_p \nabla_p H \end{pmatrix} P
\]

(9)

for the initial condition \( P(t_0, t_0) = \text{Id} \) where \( \text{Id} \) denotes the identity matrix.

In stereotomography, all of the Frechet derivatives required to build the operator can be derived from eqs (5) and (6):

\[
G = \frac{\delta g(m)}{\delta m}.
\]

(10)

The calculations of these Frechet derivatives are detailed in Appendix A.

### Non-linear inversion

Following Tarantola (1987), we introduce a misfit function over the model space, \( S(m) \), and try to minimize it. The most well-known minimization criteria are the least absolute values and the least squares of the misfits. While the first one seems to be well-adapted to geophysical problems (robust in the case of outliers in the data set), the least-squares criterion is currently used more often because it leads to the easiest computations. The probabilistic equivalent is a Gaussian hypothesis, on both the data and the model. In this case, the misfit function is a classical \( L^2 \) norm

\[
S(m) = \frac{1}{2} \{ [g(m) - d_{\text{obs}}]^T C_D^{-1} [g(m) - d_{\text{obs}}] + (m - m_{\text{prior}})^T C_{M}^{-1} (m - m_{\text{prior}}) \},
\]

(11)

where \( C_D \) and \( C_M \) are the covariance matrices in the data space and in the model space respectively, \( m_{\text{prior}} \) is any a priori model, and the superscript \( T \) denotes the adjoint operator.

Several approaches can be proposed for minimizing such a misfit function. When the problem is highly non-linear, one must resort to global optimization methods such as Monte Carlo (Press 1968; Rothman 1985; Jin & Madariaga 1994), simulated annealing (Kirkpatrick, Gelatt & Vecchi 1983; Landau, Beydoun & Tarantola 1989; Jervis, Sen & Stoffa 1996), or genetic algorithms (Jin & Madariaga 1993; Jervis et al. 1996). When the inverse problem is favourable (not too non-linear and no secondary minima), it can be solved by an iterative method, where each iteration step requires the solution of a related linear least-squares problem. To minimize our misfit function, we use local approaches, because global optimization methods are not yet realistic on present computers when applied to a real-sized number of parameters (Sambridge 1990; Jervis et al. 1996) and, consequently, have not yet been applied to 3-D cases. Local approaches involve the gradient of the misfit function, \( \delta S/\delta m \). The Gauss–Newton method is considered to be particularly efficient when the inverse problem is not too non-linear. Each iteration provides the exact solution to the locally linearized problem. This iterative scheme can be expressed as (Tarantola 1987, p. 194)

\[
m_{k+1} = m_k - \left( \frac{\delta^2 S}{\delta m^2}(m_k) \right)^{-1} \frac{\delta S}{\delta m}(m_k),
\]

(12)

where the matrix \( \delta^2 S/\delta m^2 \) is called the Hessian matrix.

### Practical Aspects

After having developed the theoretical aspects of stereotomography, we will discuss some aspects of practical implementation. They concern the model parametrization (smooth vs. blocky), the iterative local inversion scheme, the a priori information in both model and data spaces, and the important problem of data picking (especially for the slopes).

### Local non-linear optimization

In the case of stereotomography, the size of the model grows with the number of picked data, and becomes very large in the case of real data. In the Gauss–Newton approach,
Figure 5. Estimation of slopes on local slant stack panels. On the left-hand side we present a common-shot gather. In the [1.2; 2.2] s time window, we apply a local slant stack, which consists of a slant stack with a Gaussian weighting centred on the $-1200$ m trace. The corresponding local slant stack panel is presented on the right-hand side. The width of a typical event (90 per cent of the maximum value) in the local slant stack panel is evaluated at $2 \times 10^{-5}$ s m$^{-1}$.

Figure 6. First validation test: exact model. The depth–velocity profile is defined by 15 cardinal cubic B-splines with a 200 m knot spacing (right-hand side). The vertical crosses denote the B-spline knot depths. Seventeen data were computed for regularly spaced reflecting/diffracting points (marked with open circles) covering the whole depth profile (left-hand side).

The Hessian matrix becomes huge, sparse and generally ill-conditioned. In practice, the inversion of such a matrix can be a problematic operation from a numerical point of view, but many methods can be used to obtain a numerical solution (Lines & Treitel 1984; van der Sluis & van der Vorst 1987; Spakman & Nolet 1988).

For our first tests on a canonical example, the inversion was led through a singular-value decomposition (SVD) (Lanczos 1956; Jackson 1972) of the Hessian matrix $\delta^2 S/\delta m^2$. This decomposition provides us with an immediate expression of the generalized inverse of the Hessian matrix (Penrose 1955). It also gives access to the eigenvalues and eigenvectors, which allows us to conduct a sensitivity study of the inversion of different classes of parameters (Stork 1992a,b; Farra & Le Bégat 1995; Wang & Pratt 1997). We can also impose the condition number by adding a damping factor (Levenberg 1944; Marquardt 1970), or add a regularization or smoothing operator (Ory & Pratt 1995; Lailly & Sinoquet 1996), which
is equivalent to the introduction of a priori information on the model [see Phillips & Fehler (1991) for a comparative study of the effects of these constraint parameters on a tomographic inversion].

Practically, in real-sized applications, the number of parameters describing the model provides a huge matrix for which SVD becomes prohibitive in terms of computing time. In this case, adaptations of the Gauss–Newton minimization may be more efficient. They can be based, for example, on the fast numerical estimation of the inverse of the Hessian matrix. Numerous kinds of these gradient-type minimizations have been proposed. Among them, the LSQR method (Paige & Saunders 1982) seems to be particularly well-adapted to our problem. This method is based on a conjugate gradient solution of a linear system. It is often used for real-sized tomographic inverse problems since it takes advantage of the structure of large sparse matrices. In stereotomography, our Hessian matrix contains more than 80 per cent of zeros (See Fig. 7 in the first canonical example). Therefore, and because it has proven to be fast and robust in tomographic applications (Spakman & Nolet 1988), we used LSQR in our 2-D application.

A good initial model is necessary for the convergence of the iterative scheme. In our first tests, the initial velocity field was chosen to be homogeneous. In more complex media, a better initial velocity model is preferable. In our 2-D test, we determined an average constant gradient of the velocity. It was obtained through a stereotomographic inversion with one parameter describing the velocity gradient and using the largest traveltimes only.

The ray-pair parameters also need to be initialized. This operation is done from simple geometrical considerations in a homogeneous case (see Fig. 19). The initial position of a reflecting/diffracting point is set to the source–receiver mid-point in x, and to half of the traveltime multiplied by the velocity in z. The initial scattering angles are set to the angles at the surface that are calculated with the slopes picked in the data and the homogeneous velocity. The two one-way travel-times are set to half of the two-way traveltime picked in the data. These initializations lead to pairs of rays that are far from explaining the data, but are corrected as soon as the first iteration has been realized.

Model parametrization

The question of smooth or blocky velocity models for migration is still an open question. Geologists and interpreters, influenced by the stratified aspect of sedimentary rocks, generally recommend blocky models. Until now, methods for
estimating velocity fields have generally provided blocky models [velocity analysis (Dix 1955) or reflection tomography (Bishop et al. 1985; Chiu & Stewart 1987; Farra & Madariaga 1988)]. The development of ray-based migration, currently the only realistic approach for 3-D migration, has changed this view, since ray tracing in smooth velocity models actually has many numerical advantages (Chapman 1985; Lailly & Sinoquet 1996; Lambaré et al. 1996; Lucio, Lambaré & Hanyga 1996; Thierry et al. 1996). Moreover, several studies have shown that using smooth velocity macro-models does not significantly alter imaging quality (Versteeg 1993; Mispel & Hanitzsch 1996). Consequently, there is a need for methods that directly estimate smooth heterogeneous velocity fields, and one of the central benefits of stereotomography is being able to provide such models. Our method could also be used in considering blocky models, but we have not implemented this possibility.

We agree that the smooth velocity models we define should not be viewed as true representations of the subsurface. The next operation in seismic processing, depth imaging, will provide the structurally interpretable image.

In order to describe the velocity fields, we use cardinal cubic B-splines (de Boor 1978) (cubic because the second-order regularity is required for the continuity of paraxial ray tracing). We tested the shape of our misfit function for various parameterizations of the macro-model (velocity, slowness, squared slowness). It appears more parabolic if we describe B-spline weights in terms of velocity, rather than in terms of slowness or squared slowness. This choice seems original in tomographic problems.

As soon as our model is built with different classes of parameters, we must normalize them to keep the values in the same range. This normalization is set to the typical scales of our problem. In our examples, we have used 1000 m s$^{-1}$ for the values of the B-spline knots (in velocity), 1 s for the traveltimes, 0.5 rad for the angles and 1000 m for the positions. These values should be reconsidered for a different-scale application. This normalization is not a priori information on the model, which is not dependent on the units.

**A priori information**

We must consider a priori information in both model and data spaces in order to stabilize our inversion. This may be introduced in terms of covariance matrices on the data space and on the model space.

1. **Covariance on the data $C_D$:** the a priori information on the data consists of measurement error. At this time, we consider a constant value for each class of parameters in the data space. In our examples we used 1 m for the positions (denoting the correct knowledge on source and receiver positions), $2 \times 10^{-5}$ s m$^{-1}$ for the slopes (estimated in the second of the following examples) and 0.004 s for the traveltime (time step in a typical data set). For our forthcoming developments, particularly on real data, we shall assign measurement error to each pick.

2. **Covariance on the model $C_M$:** the a priori information on the model may be difficult to introduce. Without any external source of information, e.g. wells, a priori information may be introduced for numerical reasons simply in order to ensure the condition number of the Hessian matrix. Then we can consider a damping factor that imposes a reasonable conditioning number for the Hessian matrix. A non-constant damping factor can be used. Different approaches have been investigated, including a variable damping factor ensuring a better normalization of Fréchet derivatives (Toomey & Foulger 1989). Other classic a priori information involves the velocity regularization, which smooths the velocity field while attenuating the high-frequency oscillations (see e.g. Lailly & Sinoquet 1996).

**Data picking**

Stereotomography supposes that we are able to determine the traveltime and slopes of locally coherent events in the data set for selected traces. While traveltime picking on a local event is a well-known procedure, this is certainly not the case for the estimation of the local slope. It has to be done around a set of selected traces on a common-shot gather (CSG) or common-receiver gather (CRG).

We recommend the use of local slant stacks. A local slant

![Figure 8. First validation test: spectrum of eigenvalues (top) and corresponding eigenvectors (bottom). The poor conditioning is corrected by a damping factor that is iteratively updated from $1 \times 10^5$ to $1 \times 10^7$. The eigenvectors show that information is contained: in the highest eigenvalues for the traveltimes and positions, in the medium eigenvalues for the positions and the angles, and in the smallest eigenvalues for angles and velocity.](image-url)
stack consists of a slant stack (Schultz & Claerbout 1978) with a Gaussian weighting centred on the considered trace in order to decrease the influence of far events. From any trace in a CSG or a CRG we can obtain a slope–time panel. Both traveltimes and slopes are picked on the local slant stack panels (Fig. 5). Picking in stereotomography appears to be very similar to picking in standard velocity analysis. In 2-D, slopes in a CSG and a CRG must be picked simultaneously.

Picking precision is fundamental for the effective applicability of stereotomography to velocity estimation. Precision of 0.004 s in traveltimes has been estimated. In our second synthetic example, the width of a typical event (90 per cent of the maximum value) in the local slant stack panel was evaluated at $2 \times 10^{-5}$ m$^{-1}$ (see Fig. 12). Validation tests show that this precision is sufficient for constraining the velocity field.

We must note that, since it is based on the hypothesis of primary reflected/diffracted events, our method does not resolve problems linked to other types of arrivals that are not taken into account in our model parametrization, e.g. refracted arrivals, peg-leg multiples. Application to real data is needed to test the influence of such data on the stability of our algorithm. The use of the slope, in addition to other data, should also be studied as a sort-out criterion.

**VALIDATION TESTS**

The goals of the preliminary tests described in this paper are to demonstrate the ability of stereotomography to recover the velocity macro-model as well as its potential applicability to real data.

Our first two tests deal with laterally homogeneous models. Owing to the symmetry, the data set is reduced to sets of offset–traveltime–slope, and data can be picked on a single CMP gather. In this configuration, the number of parameters describing the model is relatively limited and a singular-value decomposition can be done to invert the Hessian matrix. In our first test, data are not picked on local slant stacks but simply calculated by ray tracing. The analysis of eigenvalues and eigenvectors provides us with interesting information about the sensitivity of stereotomography and about the a posteriori coupling of model parameters. The objective of the second test is to evaluate the precision of data picking on an ‘ideal’ synthetic ray + Born CMP gather. Our last test is a fully 2-D synthetic example. The size of the model imposes a non-linear optimization with an iterative LSQR scheme. Once more, data are not picked but computed by ray tracing. This test has been realized to evaluate the potential resolution of

![Figure 9. First validation test: initial and final models. On the top we present the initial ray pairs (left) and the final ray pairs (right). A comparison with Fig. 6 shows that the ray pairs have been perfectly recovered. On the bottom we present the initial velocity profile (dotted line) and the final velocity profile (dashed line), which is very close to the exact one (full line).](image-url)
stereotomography in the case of significant lateral velocity variations.

Sensitivity study of stereotomography

In order to test the optimization procedure only, we computed the data with the same ray-tracing scheme as the one used for our forward problem, and the initial model is described with the same parametrization as the exact model. The depth–velocity profile is defined by 15 cardinal cubic B-splines with a 200 m knot spacing. Seventeen data were computed from regularly spaced reflecting/diffracting points that cover the whole depth profile (Fig. 6).

Values of C₀ were chosen as described in the previous section. Considering the small number of parameters, we used an SVD for the inversion, which gives us access to the eigenvalues and eigenvectors of the 66 × 66 Hessian matrix. It is interesting to see the structure of the Hessian matrix (Fig. 7), since the inverse of the Hessian matrix provides the a posteriori resolution matrix (Farra & Le Bégat 1995). In the Hessian matrix, the submatrix H₁ corresponds to ray-segment parameters [X, Θᵣ, Θₛ, Tᵣ, Tₛ]ₓₙ=₁. Since the events are not coupled, H₁ is a succession of small matrices along the diagonal. The H₂ submatrix corresponds to velocity parameters [Cᵣ]ₓₙ=₁. The idea of inverting the ray-segment parameters (H₁) and the velocity parameters (H₂) separately is an interesting solution from a numerical point of view (Plessix 1996). However, when we pay attention to the H₁ submatrix (H₂ is the transpose of H₁), revealing the coupling between the ray-segment parameters and the velocity parameters, it is clear that these quantities cannot be overlooked. In fact, the two classes of model parameters are strongly coupled, and, with such a parametrization, a joint inversion is unavoidable (Wang & Pratt 1997).

The spectrum of eigenvalues and the corresponding normalized eigenvectors are shown in Fig. 8, which shows that, even if the velocity–depth profile seems properly sampled by reflecting points, the conditioning number is rather bad. Eigenvectors corresponding to strong eigenvalues mainly implicate ray-segment parameters (Fig. 8). The 20 strongest eigenvectors involve principally the traveltime and secondarily the depth starting positions X of ray segments with a rather flat eigenvalue spectrum. The next 31 eigenvectors involve mainly the angles and positions and secondarily the traveltime and the velocity with a significant decay of the eigenvalues. The last 15 eigenvectors implicate principally the velocity parameters and the angles. They correspond to a strong decay of the eigenvalues. Long-wavelength components of the velocity profile are associated with higher eigenvalues than are short-wavelength components. This property is well-known in traveltome tomography and seems to be generalized to stereotomography.

For the iterative non-linear inversion we used a Gauss–Newton scheme. In order to regularize the inversion of the Hessian matrix, we introduced a priori information on the model, Cₘ, which is the identity matrix multiplied by a scalar damping factor. This factor imposes the conditioning number of the Hessian matrix (ratio of the largest over the smallest eigenvalue). There was no added noise, so the conditioning number could be kept rather high during the iterative Gauss–Newton minimization. We chose to increase this conditioning number during iterations, starting from 1 × 10⁵ going up to 1 × 10⁷ (see Fig. 8). The starting model is homogeneous (2000 m s⁻¹) on the same B-spline basis as the exact one. After 20 iterations, the model solution explained all of the data in the ranges given by C₀, and the model was correctly retrieved (Fig. 9) except for the last 400 m. The damping factor slows down the convergence but allows us to converge avoiding numerical instabilities.

Precision of picked data

In order to test the method while using picked data, we computed a CMP gather with ray + Born approximation (Lambare et al. 1992) (Fig. 10). It is corrected for geometrical spreading. The source signature is the second derivative of the Gaussian function S(t) = e⁻¹₈₀.t². The short-wavelength velocity profile comes from a real log, and the reference velocity model is defined by cardinal cubic B-splines with a 200 m knot spacing (Fig. 11). The choice of 200 m as B-spline knot spacing is consistent with the recommendations of Versteeg (1993), who tested the parametrization of velocity macro-models for migration using the Marmousi model and data set. In Fig. 11, we also present the ray segments that fit the data in the exact sense.

© 1998 RAS, GJI 135, 671–690

![Figure 10](image_url)
Figure 11. Second validation test: exact velocity profile and optimal ray pairs. On the right-hand side, we present the velocity profile used for computing the CMP gather (Fig. 10), defined by 22 cardinal cubic B-splines with a 200 m knot spacing. The crosses denote the B-spline knot depths. On the left-hand side, we present the ray pairs best fitting the data for the exact velocity model. They represent the optimal ray pairs we could retrieve in this application.

velocity model. They can be seen as the segments that we are looking to retrieve.

Thirty-three data were picked on seven local slant stack panels. We present one of them (with the 1000 m offset as a reference trace) in Fig. 12. As mentioned before, on this local slant stack panel we estimated the picking precision of slopes and traveltome to $2 \times 10^{-5}$ s$^{-1}$ and 0.004 s respectively. The precision of shot and receiver positions is given by the acquisition report and can be roughly evaluated to 1 m. We notice that the precision of the slope is significantly better than that given by Hu & Menke (1992), which was based on the estimation of the polarization of three-component data. The high-frequency content of seismic exploration data and the associated dense sampling at the surface are the reasons for this improvement. In all of our validation tests, we show that such precision is sufficient to constrain the velocity macro-model in seismic reflection.

The initial model is homogeneous (2000 m s$^{-1}$), described in the same B-spline basis as the exact model. Once more, notice that the precision of the slope is significantly better than that given by Hu & Menke (1992), which was based on the estimation of the polarization of three-component data. The high-frequency content of seismic exploration data and the associated dense sampling at the surface are the reasons for this improvement. In all of our validation tests, we show that such precision is sufficient to constrain the velocity macro-model in seismic reflection.

The initial model is homogeneous (2000 m s$^{-1}$), described in the same B-spline basis as the exact model. Once more, notice that the precision of the slope is significantly better than that given by Hu & Menke (1992), which was based on the estimation of the polarization of three-component data. The high-frequency content of seismic exploration data and the associated dense sampling at the surface are the reasons for this improvement. In all of our validation tests, we show that such precision is sufficient to constrain the velocity macro-model in seismic reflection.
Hessian matrix had to be kept as low as $1 \times 10^5$. Fig. 13 shows the ray segments and the velocity profile obtained after three iterations. The results may be compared to the best-fitting ray segments in the exact model and the exact velocity profile respectively (Fig. 11). We observe that the velocity profile is well-retrieved in the upper part of the model where the density of reflecting/diffracting points is high. In the deeper part of the model, we do not have enough information to converge to the exact model. In fact, when no reflecting/diffracting points are picked, the solution is pulled down to the initial model by the action of the damping factor.

**Applicability to lateral variations of velocity**

The last validation test deals with a synthetic 2-D case. The smooth ‘salt dome’ velocity field is defined by $11 \times 13$ cardinal cubic B-splines with a knot spacing of 500 m in X and 200 m in Z. The heterogeneous velocity field, defined by B-splines, covers a surface limited to $[-1000; 6000]$ m in X and $[400; 3600]$ m in Z respectively. The velocity macro-model is defined by the superposition of a constant gradient of the velocity (the background $v(z) = 1500 \text{ m s}^{-1}$, $z$ being in m) and B-spline perturbations. In our exact model, strong velocity inclusions and dipping structures are introduced by the B-splines (Fig. 14).

Five hundred and fifty ($25 \times 22$) ray pairs were computed. The diffracting/reflecting points were regularly spaced at $250 \times 160$ m covering $[-500; 5500]$ and $[200; 3560]$ in X and Z respectively. The ray pairs are shot in the direction of the surface with a double aperture of 45° (Fig. 15). In Fig. 14, we show a few rays travelling through the high-velocity zone. We can see their significant bending, leading to caustics, created by the strong lateral variations involved in this example.

For $C_D$ we took the same values as in former tests. Owing to the size of the model, we used a non-linear iterative LSQR minimization. Our starting velocity model was a homogeneous background, $v(z) = 1500 \text{ m s}^{-1}$ (Fig. 16), corresponding to the velocity at the surface. Initial diffracting/reflecting points and slopes were estimated from the data using simple geometrical considerations (Fig. 17). In order to improve the initial model, we first inverted the velocity gradient of the background, $a[v(z) = 1500 + a \times z]$. This is what we call the first iteration. A new velocity background was obtained: $v(z) = 1500 + 1.02 \times z \text{ m s}^{-1}$ (Fig. 16). In a second step, it is used as the background for inverting the B-spline components of the velocity $[v(x, z) = 1500 + 1.02 \times z + \delta v_{\text{B-splines}}]$. During the non-linear minimization, we used: a damping factor of $1 \times 10^{-6}$, a maximum of 2000 iterations, a maximum conditioning number of 50000, and $1 \times 10^{-5}$ as an estimate of the relative errors ($atol$ and $btol$) as LSQR parameters (see Paige & Saunders 1982 for more details). The damping...
Figure 14. Third validation test: the 2-D synthetic velocity model. It is defined by $11 \times 13$ cubic cardinal B-splines with a $500 \times 200$ m knot spacing in $X$ and $Z$ respectively. The velocity goes up to $6500$ m $s^{-1}$. We present some of the ray pairs (shown in Fig. 15) to show that the strong lateral variations of the velocity bend the rays significantly, with the possibilities of caustics. The crosses denote the B-spline knot positions.

Figure 15. Third validation test: exact ray pairs. Five hundred and fifty data were computed from $25 \times 22$ reflecting/diffracting points, covering the velocity model perfectly. The ray pairs are shot in the direction of the surface with a double aperture of $45^\circ$. The crosses denote the cardinal cubic B-spline knot positions.

After seven non-linear iterations, the calculated data fitted the observed data in the range set in $C_D$. Fig. 18 shows the iterative velocity models (first, third, fifth and seventh iterations with respect to the velocity gradient background are presented). The recovered ray pairs are plotted in Fig. 19. The final model (velocity and ray pairs) fits well with the exact model, except in the sub-salt area, where slight differences can be noticed (compare Figs 14 and 18, Figs 15 and 19). This limitation of the resolution in the case of deep structures is common in reflection tomography. In Fig. 20, we compare the misfits between the initial and the exact velocity fields, and the misfits between the final and exact velocity fields. We believe that this example demonstrates the ability of stereotomography to deal with lateral velocity variations.

CONCLUSIONS

In this paper, we have presented a new reflection tomographic method, stereotomography, based on the use of the local slope of the reflected events. We have discussed the potential advantages of slope tomographic methods with respect to...
Figure 16. Third validation test: initial and first iteration velocities. The initial velocity is a homogeneous velocity field with a constant value of 1500 m s\(^{-1}\) (water velocity). In the first iteration, we inverted only a constant gradient of the velocity. The gradient value was estimated by considering the 50 largest traveltimes only.

Figure 17. Third validation test: initial ray pairs. We present the 550 initial reflecting/diffracting points and associated ray pairs. It was evaluated with simple geometric considerations in the initial homogenous velocity field (Fig. 16). We notice that it is very far from explaining the data on the surface (compared to last points of the exact rays in Fig. 15), proving that our initial velocity is far from the real one and showing the misfits our algorithm will have to deal with.

© 1998 RAS, GJI 135, 671–690
Figure 18. Third validation test: iterative evolution of the velocity. We present the velocity model for four iterations. Here, our algorithms optimized B-spline perturbations around the first iteration model (Fig. 16). After eight non-linear iterations, the calculated data fitted the observed data in the range set in $C_D$. We can compare the final velocity field with the exact one (Fig. 14).

Figure 19. Third validation test: final ray pairs. We present the 550 final reflecting/diffracting points and associated ray pairs (after eight iterations). Compared to the exact ones (Fig. 15), they are well retrieved, with slight differences in the sub-salt area.

standard velocity analysis and reflection tomography: applicability to laterally heterogeneous media and simplification in terms of data picking (identification of given reflected events all over the data set is not necessary). We have also discussed various possible approaches to slope tomography and proposed stereotomography as the most robust one. With three validation tests, we have demonstrated that precisions of slopes, traveltimes and positions picked on seismic reflection...
Velocity macro-model estimation

Figure 20. Third validation test: misfits on the velocity. On the top we present the misfits between the initial and the exact velocity field, which is what should have been found. On the bottom we present the misfits between the final (after eight iterations) and exact velocity fields, which is what was not found.

The data are sufficient for recovering velocity fields by stereotomography. The first results are very encouraging and we believe that stereotomography is a very promising approach for the recovery of velocity fields from seismic surface data. The fact that it can provide a smooth velocity macro-model at once is, in our opinion, an important advantage for ray-based migration (Thierry et al. 1996). Further demonstrations with an application to real data are needed, with special attention paid to the picking technique. The advantages of picking local events, as done in stereotomography, needs to be developed in terms of practical implementation and model resolution. We believe that the use of many local picked events should constrain the velocity model better than a few continuous events.

ACKNOWLEDGMENTS

This work was partly funded by the European Commission within the JOULE project, Reservoir Oriented Delineation Technology (contract JOF3-CT95-0019). We thank C. H. Sword and Biondo Biondi for encouragement, and Pascal Podvin for fruitful discussions and remarks, revision of this paper and enthusiasm for the approach.

REFERENCES


Guizouli, J.L., Mallet, J.L. & Madariaga, R., 1996. 3-D seismic reflection tomography on top of the GOCAD depth modeler, Geophysics, 61, 1499–1510.


APPENDIX A: CALCULATION OF FRÉCHET DERIVATIVES

In stereotomography, data are \( d = (S, R, P_s, P_r, T_{nu})^{nd} \) and model parameters are \( m = [(X, \theta_s, T_s, T_r, C_m)]^{nd} \) (see Fig. 4). The Fréchet derivatives are the partial derivatives of the data with respect to model parameters \( G = \partial d / \partial m \). Each picked event is independent of the others. Then, most of the Fréchet derivatives are set to zero, except those associated with a single picked event. For a given picked event, the two ray segments are independent except for their common initial point, and the two-way traveltime is defined by \( T_{nu} = T_s + T_r \). Consequently, the Fréchet derivatives

\[
G = \left( \begin{array}{c} J_x \frac{\partial (S, P_s)}{\partial X} \\ J_y \frac{\partial (S, P_s)}{\partial y} \\ J_z \frac{\partial (S, P_s)}{\partial z} \\ J_T \frac{\partial (S, P_s)}{\partial T_s} \\
J_T \frac{\partial (S, P_s)}{\partial T_r} \\
J_m \frac{\partial (S, P_s)}{\partial C_m} \end{array} \right)
\]

for a given picked event are

\[
J_x = \frac{\partial (S, P_s)}{\partial X} = 0 \quad J_y = \frac{\partial (S, P_s)}{\partial y} = 0 \quad J_z = \frac{\partial (S, P_s)}{\partial z} = 0 \\
J_T = \frac{\partial (S, P_s)}{\partial T_s} = 0 \quad J_T = \frac{\partial (S, P_s)}{\partial T_r} = 0 \\
J_m = \frac{\partial (S, P_s)}{\partial C_m} = 1
\]

We estimate the Jacobian matrices, \( J_x, J_y, J_z, J_T \) and \( J_m \) for the source, \( S \), and receiver, \( R \), using the paraxial ray theory as developed in the section on ‘Forward and inverse problems’. The perturbations of the ray parameters \( \delta y \) at the end point of each ray segment depend on the perturbation of the initial ray parameters, \( \delta y_0 \), the velocity field parameters \( \delta C_m \), and the traveltime, \( \delta t \). Since we chose to parametrize ray trajectories with the traveltime, the paraxial approximation at this point can be expressed as (Farra & Madariaga 1987)

\[
\delta y(t) = P(t, t_0)\delta y_0 + \int_{t_0}^{t} P(t, t')B(\delta C_m(x(t')))dt' + \left( \begin{array}{c} \nabla_p H \\ -\nabla_r H \end{array} \right) \delta t,
\]

which can provide all of the Fréchet derivatives we need in stereotomography.

The Fréchet derivatives \( J_T \) and \( J_m \) with respect to the traveltimes are directly provided by the third term of the paraxial approximation (eq. A2)

\[
J_T = I \left( \begin{array}{c} \nabla_p H \\ -\nabla_r H \end{array} \right),
\]

where \( I \) is the submatrix containing the first three lines of the \( 4 \times 4 \) identity matrix. In stereotomography, initial perturbations of ray parameters are decomposed into ray-angle perturbations, \( \delta \theta \), and initial-position perturbations, \( \delta X \), such as

\[
\delta y(t_0) = \left( \begin{array}{c} \delta \theta \\ \delta X \end{array} \right) = \left( \begin{array}{c} 0 \\ -\frac{1}{\alpha(X)}P \nabla c(X) \end{array} \right) \delta X,
\]


where
\[
\hat{\mathbf{p}} = \begin{pmatrix} -p_x \\ p_x \end{pmatrix},
\]
\(I_2\) denotes the \(2 \times 2\) identity matrix, and \(T\) denotes the matrix transposition. Using eq. (A4) and the first term of the paraxial approximation (A2) provides us with the Fréchet derivatives \(J_X\) and \(J_h\) for both source and receiver:

\[
J_X = \textbf{I} \mathbf{P}(t, t_0) \begin{pmatrix} \frac{1}{c(X)} \mathbf{P} \nabla_x \mathbf{c}(\mathbf{x})^T \\ -\frac{1}{c(X)} \mathbf{P} \nabla_x \mathbf{c}(\mathbf{x})^T \end{pmatrix}
\]

(A5)

and

\[
J_h = \textbf{I} \mathbf{P}(t, t_0) \begin{pmatrix} 0 \\ \hat{\mathbf{p}} \end{pmatrix}.
\]

(A6)

The Fréchet derivatives with respect to the velocity \(J_{C_m}\), for both source and receiver, involve the first term of eq. (A2) and the integral part of eq. (A2) such as

\[
J_{C_m} = \mathbf{I} \mathbf{P}(t, t_0) \delta \mathbf{y}_0(\delta C_m) + \int_{t_0}^t \mathbf{P}(t, t') \mathbf{B}(\delta C_m(\mathbf{x}(t'))) \, dt',
\]

(A7)

where \(\delta C_m\) is a unitary perturbation of velocity parameters \(C_m\), and the initial perturbations of ray parameters can be expressed as

\[
\delta \mathbf{y}_0(\delta C_m) = \begin{pmatrix} 0 \\ -\frac{1}{c(x_0)} \mathbf{p}_0 \delta C_m(x_0) \end{pmatrix}.
\]

(A8)

In practice, the propagator matrix, \(\mathbf{P}(t, t_0)\), is integrated along the central ray using eq. (9). Expression (A7) is integrated for each central ray and for each unitary velocity perturbation, i.e. the weight associated to each B-spline function. At each time step along the ray, we use the property of the propagator matrix, \(\mathbf{P}(t, t_0) = \mathbf{P}(t, t_0) \mathbf{P}^{-1}(t', t_0)\), and the explicit expression of the inverse propagator given in Farra & Le Bédat (1995): if

\[
\mathbf{P}(t, t_0) = \begin{pmatrix} \mathbf{P}_{11} & \mathbf{P}_{12} \\ \mathbf{P}_{21} & \mathbf{P}_{22} \end{pmatrix},
\]

then

\[
\mathbf{P}(t_0, t) = \mathbf{P}(t, t_0)^{-1} = \begin{pmatrix} \mathbf{P}_{22}^T & -\mathbf{P}_{12}^T \\ -\mathbf{P}_{21}^T & \mathbf{P}_{11}^T \end{pmatrix}.
\]

(A9)

The total number of operations involved in the integral term is proportional to \((N_{\text{data}} \times N_{\text{time step}} \times N_{C_m})\). Computing time is reasonable. As an indication, for our 2-D application (550 data and 143 B-spline functions), the total computing time for ray tracing and calculation of all of Fréchet derivatives is only a few seconds.