PAPER H

TRAVELTIME INVERSION WITH CONSTRAINTS

Reinaldo J. Michelena
Seismic Tomography Project

ABSTRACT

In this paper, I present a new method for using any given slowness image (prior information) in tomographic traveltime inversion. In contrast with the common practice of using the given image just as starting model, the method I propose uses that image as data points (constraints) simultaneously with the traveltimes. The result is that in the iterative inversion both the misfit in the traveltimes and the misfit between the given image and the updated one are minimized. Minimizing the misfit with the prior model is important when that model contains information from independent data. I show in this paper that this minimization helps to reduce the null space of the problem.

INTRODUCTION

An old problem in geophysics has been the integration of different data sets into the same inverse process with the aim of producing an image consistent with all the information available. It is well known that different methodologies can contribute with complementary insights in the solution of one particular problem, by eliminating the particular ambiguities of each method. For example, Devaney(1984) and Wu and Toksöz (1987) show that in the problem of diffraction tomography the spatial Fourier components of the object to be reconstructed are sampled differently depending upon the geometry used (cross-well, surface seismic or VSP). In exploration with electrical methods is also well known that DC data refers to shallow depths whereas MT data refers to depths of maybe several kilometers (Jupp and Vozoff, 1975). Sonic logs provide accurate depths of the layers boundaries, information that might not be present in surface seismic data.

The inverse problem usually consists on finding some model parameters that minimize the misfit between real and synthetic data. Any other source of information different to the given data set and used in the solution of the problem is called prior information.

A common practice when dealing with data sets that belong to different experiments is to invert them independently. This might be convenient for several reasons. One of them is that we don't have to worry about how to model the coupling between experiments. Another is that we don't have to worry about weights applied to each data set. However, for understanding the different results simultaneously, the interpreter has to introduce his/her own judgement (another form of prior information) to keep from each inversion the most relevant features and discard from them the "absurd" features caused essentially by the limited nature of each data set.

The most common prior information used by the interpreter when analyzing some inversion results is the smoothness of the solution. The interpreter tries to eliminate "noise" in the image and focus his/her attention on the bigger features. This assumption of good behavior of the solution can also be used for inverting the data, and this is the most common way of dealing with the nonuniqueness of the problem. Smoothness is introduced in the problem when we want some filtered version of the unknown to vanish and at the same time it reproduces the data. This is the equivalent of choosing a model norm (Claerbout, 1976). Damped least squares is an example of this technique where the filter used is the identity matrix. In general, the filter is a roughening operator and because of that the output is smooth. A disadvantage of the procedure is that, besides the subjectivity involved in the selection of the filter, the solution can be very sensitive to the type of filter applied.

Prior information is also used when we build an initial model for starting an iterative process of minimization. The model can be obtained from independent information (Lines et al., 1988) or it can be simply a guess. The objective function in the minimization is usually the misfit in the data and the speed of convergency depends not only on how close to the minimum the starting model is, but also on weather the problem is well posed or not. If the initial model contains useful information (result of an independent inversion), it won't be necessarily reproduced in the final result that minimizes the misfit in the data. This is because the inversion does not control the misfit between the calculated and the given model. Tarantola and Vallete (1982) and Tarantola (1984) propose to minimize both the misfit in the data and the misfit between the starting model and the updated one. Both misfits are weighted by the data and model covariance matrices. Their formulation reduces to those given by Franklin (1970) and Jackson (1979) for the linear case. Tarantola and Vallete (1982) show that minimizing the misfit in the model is the right condition to be applied when a nonlinear problem is going to be solved by iteration of linearized problems, in contrast with the common practice of minimizing the norm between successive models. I will show later in the paper that minimizing the misfit between the initial model and the updated one also helps to reduce the the null space of the problem.

The error in the model is calculated through the expression $(\mathbf{m} - \mathbf{m}_0)^T C_{\mathbf{m}}^{-1}(\mathbf{m} - \mathbf{m}_0)$ where $\mathbf{m}$ is the calculated model, $\mathbf{m}_0$ is the starting model, and $C_{\mathbf{m}}$ is the model covariance matrix. Examples of covariances functions used for calculating the model covariance matrix are

$$C_{\mathbf{m}}(\mathbf{m}, \mathbf{m}') = \sigma^2 \exp\{-\frac{\|\mathbf{m} - \mathbf{m}'\|^2}{2L^2}\}$$

or

$$C_{\mathbf{m}}(\mathbf{m},\mathbf{m}') = \sigma^2 \exp\{-\frac{\|\mathbf{m}-\mathbf{m}'\|}{2L}\}.$$

Nolet (1987) points out that $C_{\mathbf{m}}$ is an smoothing operator and then $C_{\mathbf{m}}^{-1}$ is then a roughening operator. This means that in the practice we just have to select the appropriate filter (choose a model norm) and minimize the power of the output (Claerbout, 1976).

The previous techniques introduce the prior information as probability distributions in data and model spaces. Information introduced in this way is called soft bounds. Although these techniques are completely general, they have the major problem that they cannot be applied to large scale problems such as tomography since they involve the inversion of large matrices. The size of the matrices is large because the covariance functions implicitly assume that the model is described as a superposition of small cells (square pixels), and in tomography usually many cells are needed for describing the model correctly.

Several attempts have been made to reduce the number of model parameters in the tomographic inversion. Most of them are based on prior assumptions about the structure of the model. For example, Chiu and Stewart (1987) modeled the earth by continuous curved boundaries (of approximately known position) separating regions of constant velocity. Van Trier (1988) uses two dimensional functions for describing the boundaries of the expected regions in the model and also allows smooth velocity variations within each region. Harlan (1989) describes the model as a superposition of smooth functions (Gaussians) since it is well known that fine details cannot be recovered from traveltimes measurements.

The discretization of the model based on prior knowledge about it certainly helps to reduce the number of model parameters in contrast with the discretization of the model into small cells that assumes no prior knowledge. The problem in the former case is that we can resolve only the features that the parametrization allows to be resolved. Any other information contained in the data is lost. In the second case, we resolve the information contained in the data but at the expense of many model parameters, because no prior information has been used.

From the examples just mentioned we conclude that a problem of major importance in tomographic inversion is the parametrization of the model. Techniques that introduce prior information as probability distributions in data and model space are not practical because a large number of model parameters is needed. Usually the number of model parameters is much larger than the amount of data available. When the number of model parameters is reduced using prior information about the model, some information present in the data might not be used. This problem raises the question about a parametrization of the model that allows the use of prior information in such a way that the information contained in the data is still completely recovered without having to use a large amount of model parameters, such as square pixels. Carrion (1989) and Carrion et al. (1988) developed a technique in which the number of cells can be made arbitrarily large but the size of the problem remains constant (and equal to the number of available rays). Prior constraints in terms of soft and hard bounds can be incorporated.

In this paper, I propose a different approach to solve the problem just mentioned. It is based on image reconstruction in Hilbert spaces. I'll show that a parametrization of the medium with the desired characteristics can be obtained directly from the theoretical expression that describes the generation of the measurements. Prior images have to be converted into data points before using them for constraining the inversion. I'll also show that the concept of consistency between forward modeling and parametrization keeps the number of model parameters equal to the number of data points and the result is an estimate of slowness that reproduces both data and prior images.

With the methodology proposed in this paper it will be possible to combine consistently different prior images or data that belong to different experiments related with slowness measurements. The result should be an improved version of all the previous results obtained by analyzing independently each data set.

CONVENTIONAL CROSS-WELL TOMOGRAPHIC RECONSTRUCTION

The traveltime along a ray path in a medium whose slowness is $S(x, y)$ is calculated with the expression

$$t_m = \int_{l_m} S(x,y) dl_m \qquad m = 1, ..., N, \tag{1}$$

where dl_m is the incremental distance along the ray path l_m. If the variations in the slowness are small, we can consider that the problem is linear (like X-rays tomography). For large variations in slowness, the problem becomes highly nonlinear because the unknown ray path depends on the slowness. This problem is usually solved by a sequence of linearized steps (using Fermat's principle) about some starting model.

The estimation of the slowness $\tilde{S}(x, y)$ usually starts with the discretization of the model in square orthogonal pixels (McMechan, 1983):

$$\tilde{S}(x,y) = \sum_{n=1}^{M} b_n R_n(x,y). \tag{2}$$

where

$$R_i(x,y) = \begin{cases} 1 & \text{if } (x,y) \text{ is in the pixel } i \qquad (i = 1, ..., M) \\ 0 & \text{otherwise} \end{cases}$$

Let's assume for simplicity that the problem is linear. Integrating both sides of Eqn. (2) along the ray l_m we get

$$t_m = \sum_{n=1}^{M} M_{mn} b_n \tag{3}$$

where

$$M_{mn} = \int R_n dl_m. \tag{4}$$

The system of equations (3) is then solved by use of the SVD technique or iterative procedures such as ART, SIRT or conjugate gradients. The coefficients b_n are substituted into (2) to obtain the slowness estimate.

The Problem with the Parametrization

Many different basis functions besides square pixels $R_n(x,y)$ (Eqn. (2)) can be used for discretizing the slowness model. The discretization in square pixels is just one way that is convenient for displaying purposes. Once the system (3) is solved the displaying of the image is immediate after a simple rearrange of the indexes. The selection of the size of the pixels (and consequently the number of them) is also arbitrary, but it is usually based on "resolution" considerations. From Eqn. (3), it is easy to conclude that for reproducing each measurement t_m accurately, the size of the pixels has to be made infinitely small, which means that in the practice we should use a large amount of small pixels. Large amounts of pixels means large systems of equations that are difficult to handle. The common solution to this problem consists simply in discretize the image more coarsely. Thus, a tradeoff between resolution and the computational effort is required.

Discretizing the image coarsely may produce some problems. As I said before, when the problem is nonlinear, it is usually solved as a sequence of linearized steps. The common goal is to minimize the misfit between real and synthetic data. The discretization of the model into square pixels can introduce by itself a misfit different to the one expected when the problem is linearized. This is because the matrix M in Eqn. (4) does not describe the measurements correctly unless many pixels are used. The problems arise because in one hand we try to minimize the misfit between real and calculated data and, in the other hand, the modeling operator M is incorrect. Because of this problem the convergency in the linear case won't be reached in one iteration as expected.

Discretizing the model into a large amount of square pixels means essentially that we don't know anything about it. Because the number of model parameters is much larger than the number of data points, we need to introduce in the problem extra information to resolve all of them. Some information may be introduced with the model covariance matrix but it translates principally into smoothing. Other discretizations might be more convenient in some situations since they can reflect with fewer parameters some prior knowledge about the model (smoothness, structure, etc.). However, most of them fail when used to build a modeling operator that reproduces the measurements correctly.

Therefore, using many or few model parameters can create problems. I will show in the next section that the formalism of reconstruction in Hilbert spaces suggest a discretization of the model where the number of degrees of freedom is same as the number of data points and all the parameters can be resolved without using smoothness conditions. The discretization is suggested by the nature of the measurements.

RECONSTRUCTION FROM PROJECTIONS IN HILBERT SPACES

In this section, I will review the fundamentals of reconstruction in Hilbert spaces. Let the original image $S(x,y)$ that we want to estimate be an element of the Hilbert space H and suppose that the data available about $S(x,y)$ is in the form of inner products with a finite set of functions $\phi_m(x,y) \in H$

$$d_m = < S(x,y), \beta_m(x,y) >= \int_\Omega S(x,y)\beta_m(x,y)dxdy \qquad m = 1, ..., N. \tag{5}$$

This information can be interpreted also as projections of the unknown function $S(x,y)$ in the set of "sampling" functions $\beta_m(x,y)$. It can be shown (Darling et al., 1983, Michelena and Harris, 1990) that the minimum norm estimate $\tilde{S}_1(x,y)$ of $S(x,y)$ is

$$\tilde{S}_1(x,y) = \sum_{n=1}^{N} a_n\beta_n(x,y), \tag{6}$$

where the coefficients a_n are calculated from the system of equations

$$d_m = \sum_{n=1}^{N} a_n < \beta_n(x,y), \beta_m(x,y) > \qquad m = 1, ..., N. \tag{7}$$

The name minimum norm estimate comes from the fact that the function $\tilde{S}_1(x,y)$ minimizes the expression

$$||S(x,y) - \sum_{n=1}^{M} a_n\alpha_n(x,y)||, \tag{8}$$

where $\alpha_n(x,y)$ and M are arbitrary. Although other choices of $\alpha_n(x,y)$ also minimizes (8), the particular choice $\alpha_n(x,y) = \beta_n(x,y)$ is a convenient one since in that case the independent term d_m in Eqn. (7) is form directly by the data points.

The estimate (6) is unique and obviously depends on the definition of the inner product in the Hilbert space.

The inner products $< \beta_n(x,y), \beta_m(x,y) >$ estimate the correlation between two different sampling functions. If they do not overlap, the inner product is zero (orthogonal sampling functions) and we can say that the measurements are uncorrelated. Strictly speaking, the number $< \beta_n(x,y), \beta_m(x,y) >$ is the cross correlation at the origin between the two functions $\beta_n(x,y)$ and $\beta_m(x,y)$. The larger the inner products, the more the overlap between the sampling functions.

The formulation of reconstruction in Hilbert spaces is completely general and can be applied directly any time the measurements are expressed in a way like (5) (the sampling functions $\beta_n(x,y)$ have to be square integrable in the support of the unknown $S(x,y)$). Fourier analysis is an example of reconstruction in Hilbert spaces (Stakgold, 1979). Another example is given by Michelena and Harris (1990). They show that in the problem of traveltime tomography, if the functions $\beta_m(x)$ are interpreted as the beam paths, the traveltimes can be interpreted as projections of the slowness along the beam paths. Another example of the application of this technique is in the problem of diffraction tomography reconstruction with constraints. This is shown in another paper in this volume.

Weighted Inner Products

We can define another Hilbert space H_2 with inner product

$$< S(x,y), \psi_n(x,y) >_{S_0} = \int_\Omega S(x,y)\psi_n(x,y)\frac{1}{S_0(x,y)}, dxdy. \tag{9}$$

where $S_0(x,y)$ is a weighting function that is nonzero in the support of $S(x,y)$, and $\psi_n(x,y)$ is a set of functions to be defined according to the generations of the measurements (5). If we multiply and divide the integrand in (5) by $S_0(x,y)$ we can say that the data can also be described as

$$d_m = < S(x,y), \psi_m(x,y) >_{S_0}, \tag{10}$$

where $\psi_m(x,y) = \beta_m(x,y)S_0(x,y)$.

According to (6) the minimum norm estimate of $S(x,y)$ is

$$\tilde{S}_2(x,y) = \sum_{n=1}^{N} a_n\psi_n(x,y) = S_0(x,y)\sum_{n=1}^{N} a_n\beta_n(x,y). \tag{11}$$

The weighting function $S_0(x,y)$ can be chosen to represent broad expected features of $S(x,y)$. In this way, $\tilde{S}_2(x,y)$ not only reproduces the data but also important features present in $S_0(x,y)$. The coefficients a_n in Eqn. (11) satisfy the system of equations

$$d_m = < \psi_m(x,y), \psi_n(x,y) >_{S_0} a_n \qquad n, m = 1, ..., N \tag{12}$$

where

$$< \psi_m(x,y), \psi_n(x,y) >_{S_0} = \int_\Omega \beta_n(x,y)\beta_m(x,y)S_0(x,y)dxdy. \tag{13}$$

This technique has been applied successfully in the past by Hall et al (1982) to introduce prior information in medical imaging problems and by Michette et al. (1984) in the problem of enhancing the resolution of synthetic seismograms. In the appendix A I show the basic equations that result when this technique is applied to diffraction tomography inversion. The main conclusion of this appendix is that the weighting function contributes to reduce the null space of the problem by increasing the area of coverage of the unknown in the frequency domain.

I am going to illustrate now how this technique can be used in the problem of traveltime tomographic inversion. As I said before, in this problem the traveltimes (d_m in the previous equations) can be interpreted as the projections of the slowness along the beam paths $\phi_n(x,y)$. In the previous equations we have to identify $S(x,y)$ as the real slowness model and $S_0(x,y)$ as a given prior image. In the Fig. 3 shows an example of the application of this technique in the reconstruction of the anomaly of Fig. 1. The prior image used in this case is shown in Fig. 2. The results of the inversion when no prior information is used is shown in Fig. 4, which is taken from Michelena and Harris (1990). They show that the norm of the null space for this reconstruction is $||f_2|| = 2.397$ and the mean absolute error is 2.0 10^{-3}. For the prior image $||f_2|| = 2.188$ and the mean error is 1.6 10^{-3}. When prior information is used (Fig. 3), these two quantities are reduced to 1.582 and 0.8 10^{-3} respectively. As expected, combining both types of information helps to reduce the null space of the problem. The starting model used in both cases is the same. Figures 5 and 6 represent the absolute value of the difference between the original model (Fig. 1) and the inversion with and without prior information (figures 3 and 4 respectively). Note that there are more zeros in Fig. 5 which means that the inversion with constraints reproduces better the values of the original image.

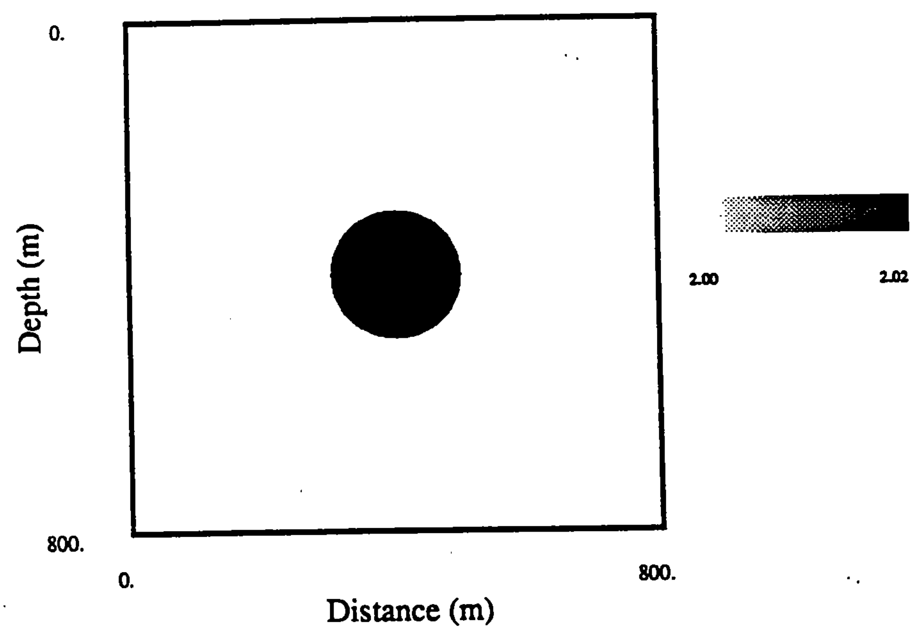

Figure 1: Slowness model. The slowness of the disk of 2.02 and the slowness of the background is 2.00.

The slowness in the prior image might be better constrained in some areas than in others. In this case we want the reconstructed model to reproduce that information. Fig. 3 shows that this technique fails in this case, because it only reproduces the "important" features in the present in the prior image but not the details (that might not be necessarily noise). The reason of this failure is that the term $\sum_{n=1}^{N} a_n\beta_n(x,y)$ in Eqn. (11) cannot be equal to one in limited view problems.

Both sources of information (prior slowness and traveltimes) are used in a different

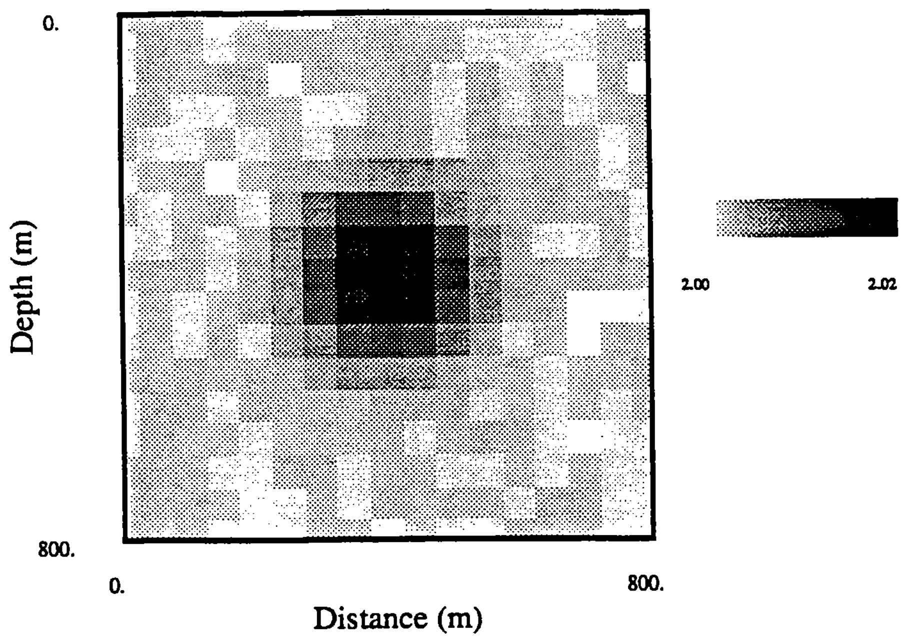

Figure 2: Slowness model used as prior information for the inversion of Fig. 3.

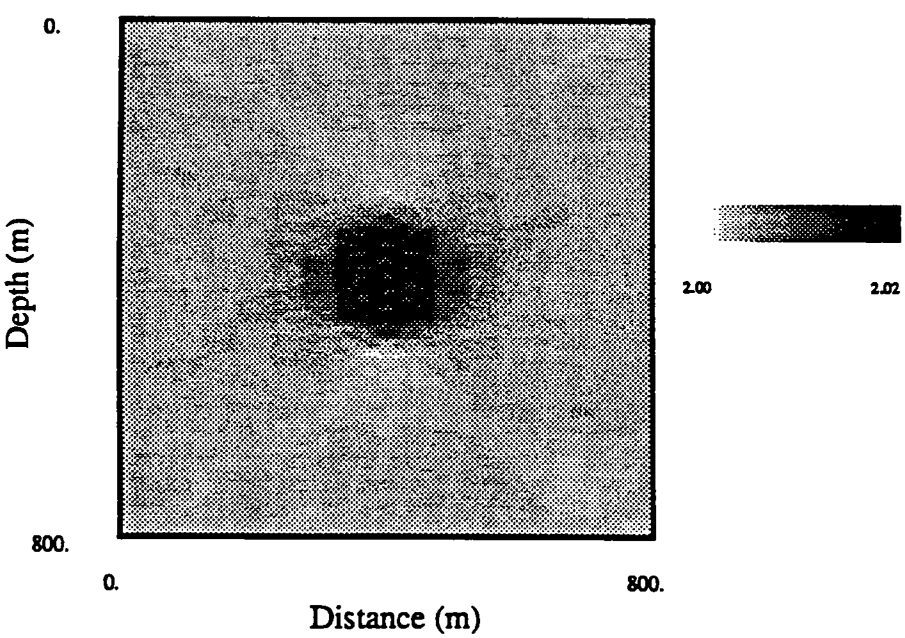

Figure 3: Natural pixels based inversion, using Fig. 2 as prior image.

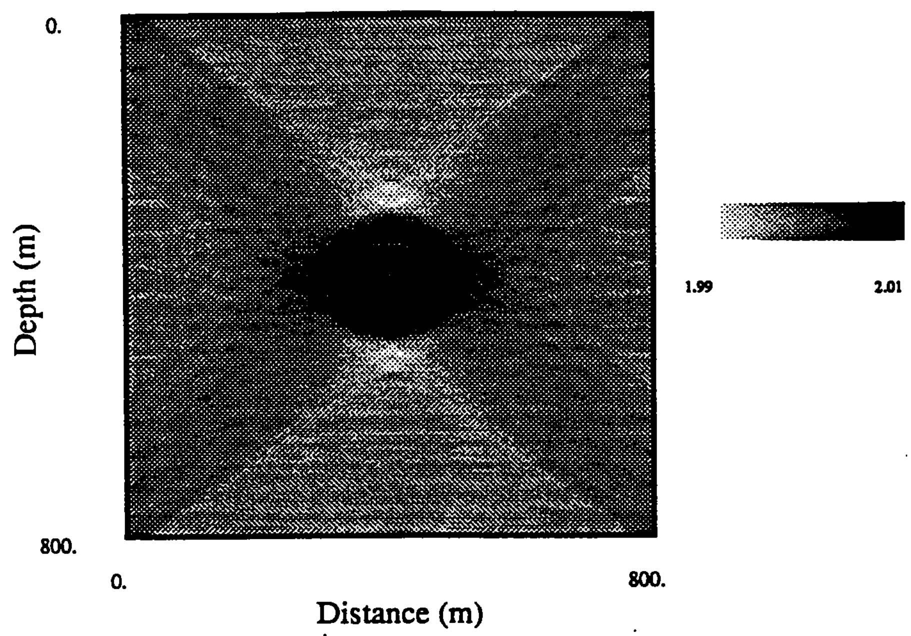

Figure 4: Natural pixels based inversion without using prior information.

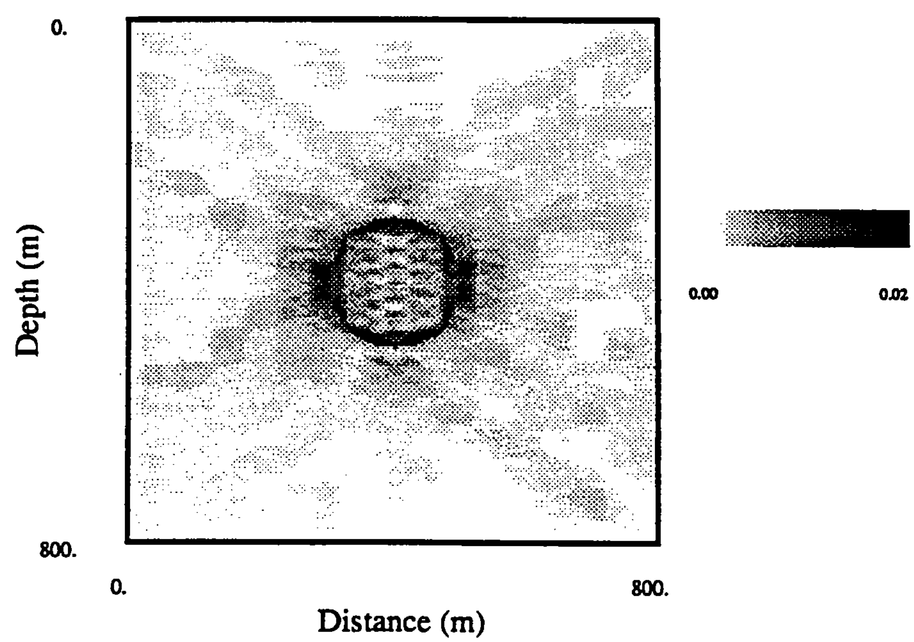

Figure 5: Error in the inversion when prior information is used

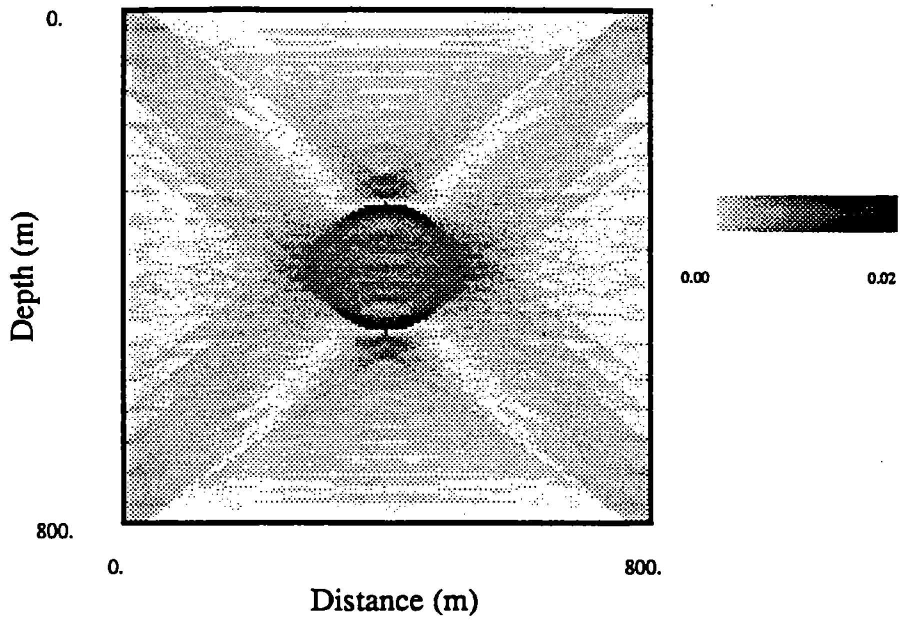

Figure 6: Error in the inversion when no prior information is used

way in this technique. The traveltimes are considered as data points in Eqn. (12), whereas the prior image is a weight the affects the calculation of the matrix coefficients and the final display of the image. If we want the prior model to be reproduced in the reconstruction, it is necessary to consider it as *data* simultaneously with the traveltimes, not as a weighting function. In the next section, I'll study the way of doing this.

Hard Constraints in the Form of Sharp Boundaries

For simplicity, let's assume that the prior image can be decomposed into M *non overlapping* regions $R_n(x,y)$ (layers) of average slowness s_n

$$S_0(x,y) = \sum_{n=N+1}^{N+M} s_n R_n(x,y) \tag{14}$$

where

$$R_i(x,y) = \begin{cases} 1/A_i & \text{if } (x,y) \text{ is in the pixel } i \qquad (i=1,...,N) \\ 0 & \text{otherwise} \end{cases} \tag{15}$$

A_i is the area is the i^{th} region. The reason why the index n in the summation is shifted N positions will be evident later.

Michelena and Harris (1990) show that this is the minimum norm estimate when the measurements are given by

$$s_n = \int_\Omega S(x,y) R_n(x,y) dx dy.$$

Remember that the traveltimes are projections of the slowness along the beam paths. In the same way, the average slowness are the projections of the slowness in the basis set defined by the functions $R_n(x,y)$. Because both kinds of information are just projections, we can treat them simultaneously if we define a new basis set $T_n(x,y)$ in the following way

$$T_n(x,y) \;=\; \begin{cases} \phi_n(x,y) & \text{if n=1,...,N} \\ R_n(x,y) & \text{if n=N+1,...,N+M} \end{cases} \tag{16}$$

The minimum norm estimate of $S(x,y)$ from the projections over the basis set $T_n(x,y)$ is

$$\tilde{S}(x,y) \;=\; \sum_{n=1}^{N+M} c_n T_n(x,y) = \sum_{n=1}^{N} c_n\phi_n(x,y) \;+\; \sum_{n=N+1}^{N+M} c_n R_n(x,y) \tag{17}$$

which becomes after reordering the indexes

$$\tilde{S}(x,y) \;= \sum_{n=1}^{N} a_n\phi_n(x,y) \;+\; \sum_{n=1}^{M} b_n R_n(x,y). \tag{18}$$

This is simply the superposition of the slowness estimate from traveltime measurements and the slowness estimate from average slowness indirect measurements.

The coefficients a_n and b_n are determined simultaneously with the following system of equations

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} <\Phi,\Phi> & <R,\Phi> \\ <R,\Phi>^T & <R,R> \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \tag{19}$$

where

$$<\Phi,\Phi>_{ij} = \int \phi_i(x,y)\phi_j(x,y)dxdy$$

$$<R,\Phi>_{ij} = \int R_i(x,y)\phi_j(x,y)dxdy$$

$$<R,R>_{ij} = 1/A_i\;\delta_{ij}$$

and

$$\mathbf{t} = \begin{pmatrix} t_1 \\ \vdots \\ t_N \end{pmatrix} \quad \mathbf{s} = \begin{pmatrix} s_1 \\ \vdots \\ s_M \end{pmatrix} \quad \mathbf{a} = \begin{pmatrix} a_1 \\ \vdots \\ a_N \end{pmatrix} \quad \mathbf{b} = \begin{pmatrix} b_1 \\ \vdots \\ b_M \end{pmatrix}.$$

The matrix in the system of equations (19) estimates the correlations between the different regions sampled by the basis set $T_n(x,y)$, which is a way of taking into account the redundancy between the different kinds of information. If the prior image is completely contained in the null space of the beam paths, all the terms $<R,\Phi>_{ij}$ vanish and the two sources of information can be inverted independently. This is equivalent to the minimization of the expression

$$\| <\Phi,\Phi>\mathbf{a} - \mathbf{t}\|^2 + \| <R,R>\mathbf{b} - \mathbf{s}\|^2 \tag{20}$$

assuming that **a** and **b** are uncorrelated. In general this is not the case. In the next section I'll study in more detail the implications of treating one given image as a starting independent model in an iterative procedure (sequential inversion) or as constraint in the same procedure (joint inversion).

Sequential Inversion vs. Joint Inversion

Traveltime tomography is essentially a nonlinear problem because the beams paths depend on the slowness model to be determined. The problem is usually solved in a sequence of linearized steps where, starting from an initial model, the forward modeled data is compared with the real measurements. The starting model is modified and the process is repeated until the misfit between the real and calculated data is below a certain level (or until we like the image obtained). This is an example of sequential inversion, since the starting model has not been used as data. Considering the starting model as data means that the projections of the given model in the basis set $\{R_n(x,y)\}$ (average slowness) are treated in the same way as the projections of the slowness in the beam paths (traveltimes).

Let's say that the starting model is $S_0(x,y)$. The correction that has to be made to that model in order to reproduce the real slowness model is

$$\Delta S(x,y) = S(x,y) - S_0(x,y) \tag{21}$$

In the sequential inversion, $\Delta S(x,y)$ is expressed as a linear combination of natural pixels

$$\Delta S(x,y) \;=\; \sum_{n=1}^{N} a_n\phi_n(x,y) + f_{null}(x,y), \tag{22}$$

where the function $f_{null}(x,y)$ describes the null space of the traveltime anomalies

$$\int_\Omega f_{null}(x,y)\phi_n(x,y)dxdy = 0. \tag{23}$$

In other words, $f_{null}(x,y)$ represents regions in the model that produce no traveltime anomalies.

If the starting model is used as data also (joint inversion), ΔS is expanded as follows

$$\Delta S(x,y) \;=\; \sum_{n=1}^{N} a_n\phi_n(x,y) + \sum_{n=1}^{M} b_nR_n(x,y) + f'_{null}(x,y). \tag{24}$$

Note that the null space is different in both formulations ($f_{null}(x,y) \neq f'_{null}(x,y)$). The reason is that if any of the functions $\{R_n(x,y)\}$ cannot be expanded in terms of $\phi_n(x,y)$, we can say that the expression (24) contains terms that belong to the null space $f_{null}(x,y)$ in Eqn. (22). Therefore, the main difference between the two ways of using $S_0(x,y)$ is that using it as data effectively contributes to reduce the null space of the problem. The coefficients a_n and b_n are the solution of the following system of equations

$$\begin{pmatrix} \mathbf{\Delta t} \\ \mathbf{0} \end{pmatrix} = \begin{pmatrix} <\Phi,\Phi> & <R,\Phi> \\ <R,\Phi>^T & <R,R> \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix} \tag{25}$$

The zero in the independent term of the previous equation is because the projections of the perturbation in the regions $R_n(x,y)$ have to be zero. In the linearized inversion based on Eqn. (24) the traveltimes has to be matched while the mismatch with the original one remains always zero. Remember that what remains zero is the *mean* slowness mismatch in each region. The final output is consistent with both sets of projections, average slowness and traveltimes.

If we have no information about the velocities in the regions (layers) $R_n(x,y)$ but we know the boundaries, the Eqn. (24) remains valid, but we have to eliminate the second row in Eqn. (25). The problem is not square any more because of the incomplete information. A typical situation where only the boundaries are known is when the given image is a migrated section.

The reader may wonder about what happen if the prior image is not a simple superposition of independent regions, but a superposition of non orthogonal functions, such as Gaussians. The output of the inversion of independent data may be represented in this form. In other cases we may have images formed as superposition of *many* independent regions (square pixels, for example) and we might want to reduce the number of parameters involved in the description of that image before using it as prior information for inverting another data set. This two problems will be the topic of the next section.

Soft Constraints in the Form of Fuzzy Boundaries

In a previous section, I explained a way of using any image as prior information considering it as a weighting function in the definition of the inner product that describes the generation of the measurements. The main disadvantage of this approach is that, although the null space is reduced, the reconstructed image reproduce only the general features contained in the prior one. Another way of using that first image is as starting model of an iterative procedure. In this case the prior image does not contribute to reduce the null space of the problem. The conclusion of the previous section is that the way to avoid this problem is to consider the prior image also as data. Let's see how we can convert any given image into data, regardless it is expressed as a combination of independent regions or not.

Let's say that we have an image represented as a linear combination of functions $\{C_n(x,y)\}$

$$S_0(x,y) = \sum_{n=1}^{L} c_n C_n(x,y). \tag{26}$$

The representation of $S_0(x,y)$ in another basis set $D_n(x,y)$ is

$$\tilde{S}_0(x,y) = \sum_{n=1}^{K} d_n D_n(x,y) \approx S_0(x,y) \tag{27}$$

Multiplying both sides of Eqn. (27) by $D_m(x,y)$ and integrating, we obtain the system of equations for the coefficients d_n

$$\int S_0(x,y)D_m(x,y)dxdy \;=\; s'_m \;=\; \sum_{n=1}^{K} d_n \int D_n(x,y)D_m(x,y)dxdy \qquad m=1,...,K \tag{28}$$

Therefore, the procedure for representing one given image in a different basis set is the following :

- Compute the projections of the given image in the desired basis set (data generation, left hand side of Eqn. (28)).
- Compute the inner products between all the possible combinations of functions in the new basis set (generation of the matrix).
- Solve the system of equations (28).

It is interesting that is we assume that $S_0(x,y)$ is the real slowness and if $D_m(x,y) = \phi_m(x,y) = (\textit{beam path})_m$, our problem of changing basis turns into the problem of traveltime tomography. The change of basis made in traveltime tomography is from the continuous representation of the real slowness to the discrete basis defined by the beam paths. Of course, a considerable amount of information is lost in that process. This tell us that in any process of changing basis information is lost if the new basis set does not expand the same space expanded by the original one.

If we want to use any given image as constraint in traveltime inversion we have to decide first the basis where we are going to represent it. This selection can be as arbitrary as the selection of the basis function explained at the beginning of the paper. The difference is that now the selected basis set is also used for calculating the data. Let's call this basis set $D_n(x,y)$. Then we have to add one more term to the summation (18) and the estimate of the slowness is now represented as

$$\tilde{S}(x,y) \;=\sum_{n=1}^{N} a_n\phi_n(x,y) \;+\; \sum_{n=1}^{M} a_n R_n(x,y) \;+\; \sum_{n=1}^{K} d_n D_n(x,y). \tag{29}$$

The unknown coefficients a_n, b_n and d_n satisfy

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \\ \mathbf{s}' \end{pmatrix} = \begin{pmatrix} <\Phi,\Phi> & <R,\Phi> & <D,\Phi> \\ <R,\Phi>^T & <R,R> & <D,R> \\ <D,\Phi>^T & <D,R>^T & <D,D> \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{d} \end{pmatrix}. \tag{30}$$

s' represents the data generated from the inner products of the given image with the selected basis set $\{D_n(x,y)\}$.

The slowness estimate in Eqn. (29) is a linear combination of different kinds of information. The first term in the right hand side of (29) is a superposition the beam paths. The term in the center (superposition of orthogonal functions) describes the regions in the model where the average slowness is known ("sharp boundaries") form information independent to the traveltimes. The third term models prior images where no clear boundaries are defined ("fuzzy boundaries"). The last two terms can of course be included into one, but I preferred to keep it that way to make clear the differences between the different kinds of information.

It is important to note that the system of equations (30) has unique solution and therefore, the slowness estimate satisfy all the given projections. That does not mean that the estimate is identical to the prior image. If the mean velocity in one area is given, the model estimated will reproduce that value, but it won't necessarily be homogeneous in that region.

For several reasons we might want the inversion to give more weight to one source of information than another. The next section addresses this topic.

Gradual Mixing of Images

In this section, I propose a way of weighting the traveltimes and the prior images to produce inversions that are influenced only by the traveltimes, only by the prior image or any by combination of both.

Equation (18) tells us that by multiplying the beam paths by a constant α and the regions R_n by a different constant β, we may achieve our goal. The slowness estimate is then

$$\tilde{S}(x,y) \;=\; \alpha \sum_{n=1}^{N} a_n \phi_n(x,y) \;+\; \beta \sum_{n=1}^{M} b_n R_n(x,y). \tag{31}$$

The system of equations for a_n and b_n is

$$\begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \alpha < \Phi, \Phi > & \beta < R, \Phi > \\ \alpha < R, \Phi >^T & \beta < R, R > \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}. \tag{32}$$

We see that for $\beta \;=\; 0$, the inversion for only traveltimes is not reproduced. For $\alpha \;=\; 0$ the inversion for only slowness values is not reproduced either. The solution to this problem is to weight independently the system of equations (32) in the following way

$$\begin{pmatrix} \alpha I_{NN} & 0 \\ 0 & \beta I_{MM} \end{pmatrix} \begin{pmatrix} \mathbf{t} \\ \mathbf{s} \end{pmatrix} = \begin{pmatrix} \alpha I_{NN} & 0 \\ 0 & \beta I_{MM} \end{pmatrix} \begin{pmatrix} \alpha < \Phi, \Phi > & \beta < R, \Phi > \\ \alpha < R, \Phi >^T & \beta < R, R > \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}, \tag{33}$$

where I_{NN} and I_{MM} are the identity matrices of dimensions NxN and MxM respectively.

After multiplying results

$$\begin{pmatrix} \alpha\mathbf{t} \\ \beta\mathbf{s} \end{pmatrix} = \begin{pmatrix} \alpha^2 < \Phi, \Phi > & \alpha\beta < R, \Phi > \\ \alpha\beta < R, \Phi >^T & \beta^2 < R, R > \end{pmatrix} \begin{pmatrix} \mathbf{a} \\ \mathbf{b} \end{pmatrix}. \tag{34}$$

If we constraint $\alpha+\beta \;=\; 1$, we get the desired result, and we can change gradually from inversion influenced only by traveltimes to inversion influenced only by the prior image. The final estimate of the slowness is given by Eqn. (31).

CONCLUSIONS

The theory of reconstruction in Hilbert spaces provides a unified framework for inverting simultaneously traveltimes and prior slowness information about the model. When the prior slowness image is used as a weighting function in the inner product that describes the measurements, the null space of the reconstruction is reduced when compared with the null space of each source of information (traveltimes and prior image). The disadvantage of this procedure is that only the main features in the prior image are reproduced by the the final result. I have shown, theoretically, that if the prior image is converted into data points (like the traveltimes) this problem can be solved and the result should be an image consistent with all the information available.

I plan to test the proposed theory in realistic synthetic examples as well as real data.

ACKNOWLEDGMENTS

I am grateful to INTEVEP, S.A. for supporting my study at Stanford. I am also grateful to Jerry Harris for many useful discussions and important suggestions.

REFERENCES

Carrion, P. M., 1989, Generalized non-linear elastic inversion with constraints in model and data spaces: Geophys. J., **96**, 151-162.

Carrion, P. M., Auyeung, C., and Mersereau, R., 1988, Constrained traveltime tomography: presented at the 58th Ann. Internat. Mtg., Soc. Expl. Geophys.

Chiu, S. K. and Steward, R. R., 1987, Tomographic determination of three-dimensional seismic velocity structure using well logs, vertical seismic profiles and surface seismic data: Geophysics, **52**, 1085-1098.

Claerbout, J. F., 1976, Fundamentals of geophysical data processing: McGraw-Hill

Darling, A. M., Hall, T. J., and Fiddy, M. A., 1983, Stable noniterative object reconstruction from incomplete data using a priori knowledge: J. Opt. Soc. Am., **73**, 1466-1469.

Devaney, A. J., 1984, Geophysical diffraction tomography: IEEE Trans. Geosci. Remote Sensing,**GE-22**, 3-13.

Hall. T, J., Darling, A. M., and Fiddy, M. A., 1982, Image compression and restoration incorporating prior knowledge: Optics Letters, **7**, 467-468.

Harlan, W. S., 1989, Tomographic estimation of seismic velocities from reflected raypaths: presented at the 59th Ann. Internat. Mtg., Soc. Expl. Geophys.

Jackson, D. D., 1979, The use of a priori data to resolve non-uniqueness in linear inversion: Geophys. J. Roy Astr. Soc., **57**, 137-157.

Jupp, D. L. B., and Vozoff, K., 1975, Stable iterative methods for inversion of geophysical data: Geophys. J. Roy. Astr. Soc., **42**, 957-976.

Lines, L. R., Schultz, A. K., and Treitel, S., 1988, Cooperative inversion of geophysical data: Geophysics, **53**, 8-20.

McMechan, G. A., 1983, Seismic tomography in boreholes: Geophys. J. Roy Astr. Soc., **74**, 601-612.

Michelena, R. J., and Harris, J. M., 1990, Tomographic traveltime inversion using natural pixels: paper F in this volume.

Michette, A. G., Darling, A. M., Fiddy, M. A., Ward, W. A., and Franklin, E., 1984, Resolution enhancement of well log and seismic data: Geophys. Prosp., **32**, 564-580.

Nolet, G., 1987, Seismic wave propagation and seismic tomography, *in* Nolet, G., Ed., Seismic Tomography: D. Reidel Publ. Co., 1-23.

Tarantola, A., 1984, Inversion of seismic reflection data in the acoustic approximation: Geophysics, **49**, 1259-1266.

Tarantola, A., and Valette, B., 1982, Generalized nonlinear inverse problems solved using the least squares criterion: Rev. of Gephys. and Space Physics, **20**, 219-232.

Van Trier, J., 1988, Migration velocity analysis using geological constraints: presented at the 58th Ann. Internat. Mtg., Soc. Expl. Geophys.

Wu, R. S., and Toksöz, N., 1987, Diffraction tomography and multisource holography applied to seismic imaging: Geophysics, **52**, 11-25.