SEISMIC RAY TOMOGRAPHY USING L_1 INTEGRAL NORM

Vânia G. de Brito dos Santos[1] and Wilson M. Figueiró[2]

ABSTRACT. Seismic ray tomography methods are usually associated with substantial computer processing time. The reason for this is that at each step of the iterative inversion process defined by the tomographic method the two-point ray tracing problem must be solved for each source-receiver pair. In order to resolve this, an Euclidean norm (L_2 vector norm), commonly used in error functions which are to be minimized in inversion procedures, is substituted by an L_1 integral norm, which enables the estimation of model parameters by minimizing the area between observed and calculated traveltime curves that are interpolated (or adjusted) to the data points. Relatively simple mathematical developments and numerical experiments with two-dimensional compressional seismic wave velocity field models show that L_1 integral norm saves an enormous amount of processing time with no significant loss of accuracy. Occasionally, parameters of the model can be better estimated using L_1 integral norm than the L_2 vector norm that is traditionally utilized in seismic inversion tomography.

Keywords: seismic ray, tomography, polynomial parameterization, seismic velocity field, two-point ray tracing problem, L_2 vector norm, L_1 integral norm.

RESUMO. O alto consumo de tempo em processamento computacional é um problema que, geralmente, está associado aos métodos de tomografia sísmica. Isto ocorre porque, em cada passo do processo iterativo de inversão definido pelo método tomográfico, o problema da conexão de dois pontos, pela curva da trajetória do raio sísmico, deve ser resolvido para cada par fonte-receptor. A fim de reduzir a gravidade deste tipo de problema, a norma Euclideana (norma L_2 de vetor), comumente empregada nas funções de erro a ser minimizado no processo de inversão, é substituída por uma norma L_1 de função. Essa mudança permite estimar parâmetros do modelo através da minimização da área entre curvas de tempos observados e calculados que são interpoladas (ou ajustadas) aos pontos referentes aos dados. Desenvolvimentos matemáticos e experimentos numéricos relativamente simples, com modelos bidimensionais de campos de velocidade sísmica de ondas compressionais, mostram que a norma L_1 de função permite poupar uma enorme quantidade de tempo de processamento sem uma importante perda de precisão. Às vezes, os parâmetros do modelo são estimados de modo mais acurado usando-se a norma L_1 da integral em lugar da norma L_2 de vetor, tradicionalmente usada na inversão tomográfica.

Palavras-chave: raio sísmico, tomografia, parametrização polinomial, campo de velocidade sísmica, problema de conexão de dois pontos por traçado de raio, norma L_2 de vetor, norma L_1 de integral.

[1]Universidade Estadual de Feira de Santana, Departamento de Ciências Exatas, módulo V, Km 3, BR-116, Campus Universitário, s/n, Caixa Postal 252294, 44031-460 Feira de Santana, BA, Brazil. Phone: +55 (75) 3161-8086 – E-mail: vania_goncalves@uol.com.br

[2]Universidade Federal da Bahia, CPGG-UFBA, Centro de Pesquisas em Geofísica e Geologia da Universidade Federal da Bahia, Instituto de Geociências, Rua Barão de Jeremoabo, s/n, 40170-290, Federação, Salvador, BA, Brazil. Phone: +55 (71) 3283-8520/8551; Fax: +55 (71) 3283-8501/3235-2423 – E-mails: figueiro@cpgg.ufba.br; figueiro6@gmail.com

INTRODUCTION

The concept of distance between observed and calculated data has a central role in inverse geophysical problems, where it can be expressed in terms of norms. Over the years certain works have attempted to improve objective functions based on norms, with improvements being witnessed in respect of the following issues: robustness, regularization, stability, accuracy, and processing time reduction. Claerbout & Muir (1973) suggested that some norms are more appropriate than others depending on the application, for example: the L_1-norm is more robust than the L_2-norm. Scales & Gersztenkorn (1988) presented an alternative to the regularization of inverse problems based on the L_1-norm (least-absolute deviation) instead of damped least squares. Bube & Langan (1997) applied iteratively re-weighted least square to a hybrid L_1/L_2 minimization problem that works better for data with outliers than the L_2 algorithm with a small additional computational cost. A combination of the L_2-norm with L_1-norm has been established by the so-called Huber norm that gives us more robust parameter model estimation than the L_2-norm and it retains its smoothness better than the L_1-norm for small residuals (Guitton & Symes, 2003). When wavelets are used to represent models, the L_1-norm is better than L_2 regularization on wavelet decomposition used to represent a model solution of a linearized seismic tomographic problem (Loris et al., 2007). Different norms are compared and used in combination with nonlinear regularization techniques in order to discuss seismic tomography aspects such as: ill-conditioned matrices, model reconstruction, and time-consumption (Loris et al., 2010).

Amongst other issues, seismic ray tomography is faced with the problem of large amounts of processing time being wasted in order to estimate model parameters. The traditional objective functions (Bishop et al., 1985) that use L_2 vector norm, need to have a time calculated at the same receiver where a seismic ray arrives and where a real traveltime is observed. This means that for each source-receiver pair, the two-point ray tracing problem must be solved. The problem is therefore accentuated because, with tomographic work, it is recommended to have a good coverage of the velocity field model, and so it is necessary to have a large number of source-receiver pairs connected by seismic ray trajectories, for which traveltimes are calculated. In addition, even for relatively simple two-dimensional models, the two-point ray tracing problem is an iterative process that can result in prolonged computer processing time. Given that seismic tomography is also an iterative method, the same procedure of connecting source and receiver positions (for each one of the sev-

eral source-receiver pairs) must be repeated to this extent, since this is the supposed number of required iterations for convergence. In order to mitigate against such a worrying problem, we propose a substitution of the traditional L_2 vector norm by an alternative integral L_1-norm that does not demand connection positions by ray trajectories and, consequently, saves a significant amount of processing time.

Seismic Ray Tracing and Traveltime Calculation

If $V(x, z)$ is a two-dimensional velocity field model, T represents traveltime, and ds is an arc length element; then to find the trajectory C that minimizes the functional

$$T(C) = \int_C \frac{ds}{V(x, z)}, \tag{1}$$

requires that the following system of equations be solved (Červený, 1987)

$$\begin{cases} \dfrac{d\mathbf{x}(\tau)}{d\tau} = \mathbf{p}(\tau) \\[2mm] \dfrac{d\mathbf{p}(\tau)}{d\tau} = \dfrac{1}{2}\vec{\nabla}\left(\dfrac{1}{V^2(x,z)}\right), \end{cases} \tag{2}$$

where τ is a ray parameter, $\mathbf{x}(\tau)$ is a vector position of a ray trajectory point at τ and $\mathbf{p}(\tau)$ is the slowness vector that is tangent to the ray trajectory at τ. Numerically, rays are traced by means of the following recursive system

$$\begin{cases} \mathbf{x}(\tau + \Delta\tau) = \mathbf{x}(\tau) + \mathbf{p}(\tau)\Delta\tau \\[2mm] \mathbf{p}(\tau + \Delta\tau) = \mathbf{p}(\tau) + \dfrac{1}{2}\vec{\nabla}\left[\dfrac{1}{V^2(x(\tau),z(\tau))}\right]\Delta\tau, \end{cases} \tag{3}$$

obtained by a Taylor expansion of Equation (2). In this work, the two-point ray tracing problem is solved by the Heun method (Santos & Figueiró, 2006). The traveltime along ray trajectory is calculated numerically by the following expression:

$$\begin{aligned} T(x_{N+1}, z_{N+1}) &= \sum_{i=0}^{N} \Delta T_i \\ &= \sum_{i=0}^{N} \frac{1}{V_i} \cdot \|\mathbf{x}_{i+1} - \mathbf{x}_i\|_2, \end{aligned} \tag{4}$$

where the calculation of the traveltime element, ΔT_i at each knot of the polygonal line, which represents the ray trajectory, is achieved concomitantly with its trace; V_i is the wave velocity at the point $\mathbf{x}_i = (x_i, z_i)$; $N + 1$ is the total number of small straight segments that constitute the ray trajectory up to $\mathbf{x}_{N+1}$; and $\| \|_2$ is the Euclidean norm. At the end of each step in the process of ray trajectory construction, the slowness vector $\mathbf{p}$

must be up-to-date in order to satisfy the eikonal equation in the following way:

$$\|\mathbf{p}\|_2 = \sqrt{p_1^2 + p_2^2} = \frac{1}{V(x, z)} \,, \qquad (5)$$

where $\mathbf{p} = (p_1, p_2)$ with $p_1 = \|\mathbf{p}\|_2 \cdot \cos(\theta)$, $p_2 = \|\mathbf{p}\|_2 \cdot \sin(\theta)$ and θ is the departure angle, meaning the angle between $\mathbf{p}(\mathbf{0})$ and the positive orientation of the X axis, see Figure 1. In other words, the direction of $\mathbf{p}$ is preserved, but its magnitude is altered in order to satisfy Equation (5). Initial conditions necessary to solve the ray tracing equations are: source position $\mathbf{s} = (x = 0, z = 0) = \mathbf{0}$ and the slowness vector, $\mathbf{p}(\mathbf{0})$ given by $\frac{1}{V(0,0)}(\cos(\theta), \sin(\theta))$.
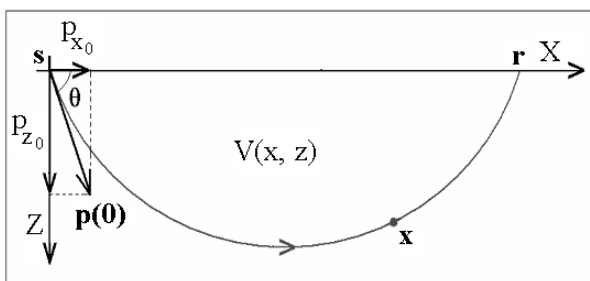
Figure 1 – Illustration of a ray trajectory between the source $\mathbf{s}$ and the receiver $\mathbf{r}$ showing vectors $\mathbf{x}$ and $\mathbf{p}$ (the latter has axis projections p_{x0} and p_{z0}).

In the iterative process performed in order to solve the two-point ray tracing problem that connects source to receiver, the bisection method is used to determine the departure angle θ of the ray trajectory. Figure 2 shows a ray network obtained by employing numerical methods such as those presented here for the seismic velocity target model V_2 used for inversions in this work.
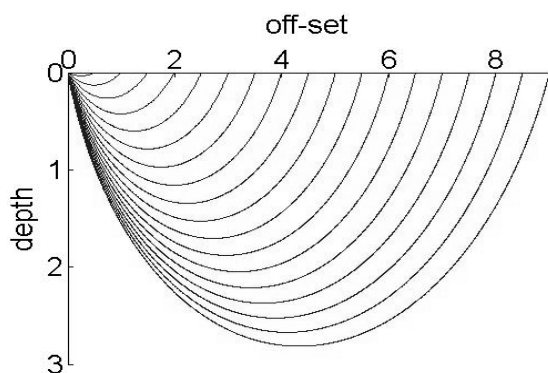
Figure 2 – A ray network on the velocity field model expressed by Equation (29). Within it, the two-point ray tracing problem is solved for each source-receiver pair.

Gauss-Newton Method

The solution of an inverse problem begins with the use of a set of observed data (real or synthetic), and can be divided into three steps: model parameterization, direct modeling, and inverse modeling (Figueiró, 1994). The latter consists of the determination (or estimation) of model parameters that employ the aforementioned data set. The local optimization method called Gauss-Newton is an iterative procedure that enables us to solve a nonlinear problem by means of successive linearization. In fact, the problem is solved on a step-by-step basis by applying, repeatedly, the least squares method.

In the modeling process the input of the system is described by model vector $\mathbf{m}$ whose components are the model parameters. As the problem is linearized, the output, in forward modeling, can be expressed by $\mathbf{Gm}$ where $\mathbf{G}$ is a matrix that describes the wave propagation process that depends on the geometrical structure of the problem. The model $\mathbf{m}$ can be written as:

$$\mathbf{m} = (m_1, m_2, m_3, \ldots, m_m)^T \,. \qquad (6)$$

If the vector $\mathbf{d} = (d_1, d_2, d_3, \ldots, d_n)^T$ contains the observed data (synthetic for the purpose of this work, but it can be real, if available), then in order to solve the inverse problem we need to find the model $\mathbf{m}$ that minimizes the following residual error function:

$$\mathbf{e}(\mathbf{m}) = \mathbf{d} - \mathbf{Gm} \,. \qquad (7)$$

A candidate for a minimum of Equation (7) is a least squares solution for the objective function $\Phi(\mathbf{m})$ written as follows:

$$\Phi(\mathbf{m}) = \mathbf{e}(\mathbf{m})^T \mathbf{e}(\mathbf{m}) = (\mathbf{d} - \mathbf{Gm})^T(\mathbf{d} - \mathbf{Gm}) \,. \qquad (8)$$

The so-called least square solution of Equation (8) is provided by (Menke, 1989):

$$\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d} \,. \qquad (9)$$

And residual error function is given by

$$\Delta\mathbf{d}(\mathbf{m}) = \mathbf{d}_{obs} - \mathbf{d}_{calc}(\mathbf{m}) \,, \qquad (10)$$

where $\mathbf{d}_{obs}$ and $\mathbf{d}_{calc}(\mathbf{m})$ are, respectively, the observed data and the calculated data for the model $\mathbf{m}$. If $\Delta\mathbf{d}$ is nonlinear, it can become linear using its first order Taylor series expansion. It produces:

$$\Delta\mathbf{d}_{ap}(\mathbf{m}) = \Delta\mathbf{d}(\mathbf{m}_0) + \mathbf{S}(\mathbf{m}_0)\Delta\mathbf{m} \,, \qquad (11)$$

where $\Delta\mathbf{d}_{ap}$ is the linear approximation of $\Delta\mathbf{d}$, $\mathbf{m}_0$ is a reference model (not necessarily an initial one, but more appropriately a current model), $\Delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ is a model perturbation, and $\mathbf{S}(\mathbf{m}_k)$ is the sensibility matrix calculated at the current model $\mathbf{m}_k$. The ij-th entry of $\mathbf{S}(\mathbf{m}_k)$ is provided by:

$$S_{ij}(\mathbf{m}_k) = \frac{\partial d_{calc,i}}{\partial \mathbf{m}_j}(\mathbf{m}_k) \,. \qquad (12)$$

The solution of Equation (11) is then:

$$\Delta\mathbf{m}_k = \left[\mathbf{S}(\mathbf{m}_k)^T\mathbf{S}(\mathbf{m}_k)\right]^{-1}\mathbf{S}(\mathbf{m}_k)^T\Delta\mathbf{d}(\mathbf{m}_k)\,. \quad (13)$$

This enables the following iterative procedure:

$$\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta\mathbf{m}_k\,. \quad (14)$$

In this exposed way, the inverse iterative process employing an L_2 vector norm is characterized.

Iterative Process defined by the L_1 Integral Norm

Let $\mathbf{m}_t = (\mu_1, \mu_2, \mu_3, \ldots, \mu_m)^T$ be a target model, where m is its number of parameters. All the parameters of $\mathbf{m}_t$ are fixed and unknown. For $\mathbf{m}_t$ we have n pairs of data (x_i, t_i) where t_i is the synthetic observed traveltime data for the receiver position x_i.

Then, we can construct a polynomial function,

$$t_{obs}(x) = a_0 + a_1x + a_2x^2 + \cdots + a_px^p\,, \quad (15)$$

by means of cubic spline interpolation (Santos, 2009) using pairs $(x_i, t_i), i = 1, 2, \ldots, 18, x_i \in I = [0, x_{\max}]$, obtained by the ray-tracing method. The SPLINE$((x_i, t_i), x)$ MATLAB routine was used to generate a piecewise third degree polynomial $S_3(x)$, i.e. a cubic spline on the interval I. After that, $S_3(x)$ is sampled in the points $x_k = k \cdot 10^{-2}, k = 0, 1, 2, \ldots, 900$, in order to use POLYFIT$(x_k, S_3(x_k), p)$ MATLAB routine, producing a polynomial of degree $p = 8$ by adjustment to the data $(x_k, S_3(x_k))$. Such a choice was motivated by the need to

have a sufficient degree of freedom and not so much oscillation in agreement with the synthetic observed data behavior.

If $\mathbf{m} = (\eta_1, \eta_2, \eta_3, \ldots, \eta_m)^T$ represents a general (or a current) model that can vary freely and produces n pairs $(\tilde{x}_i, \tilde{t}_i)$ of calculated traveltimes, $\tilde{t}_i$ at the arrival position, $\tilde{x}_i$ of a seismic ray; then we can obtain (applying the same procedure used to obtain Equation (15)), a polynomial function

$$t_{calc[\mathbf{m}]}(x) = b_0 + b_1x + b_2x^2 + \cdots + b_qx^q\,, \quad (16)$$

where $[\mathbf{m}]$ is to remind us that the calculated traveltime depends on $\mathbf{m}$. These two kinds of data pairs (observed and calculated) and the two functions $t_{obs}(x)$ and $t_{calc[\mathbf{m}]}(x)$ are represented in Figure 3.

The problem consists in to find a model $\mathbf{m}$ that minimizes the area function

$$
\begin{aligned}
A(\mathbf{m}) &= \|t_{obs}(x) - t_{calc[\mathbf{m}]}(x)\|_{int} \\
&= \int_0^{x_{\max}} \left|t_{obs}(x) - t_{calc[\mathbf{m}]}(x)\right|dx\,,
\end{aligned} \quad (17)
$$

where $A(\mathbf{m})$ is a non-negative scalar functional defined in $\mathbf{R}^m$. It is supposed that the model $\mathbf{m}_{k+1}$ is produced by a perturbation $\Delta\mathbf{m}_k$, i.e., $\mathbf{m}_{k+1} = \mathbf{m}_k + \Delta\mathbf{m}_k$. Then, $A(\mathbf{m}_{k+1}) = A(\mathbf{m}_k + \Delta\mathbf{m}_k)$. Using a Taylor expansion of $A(\mathbf{m}_{k+1})$, we have:

$$
\begin{aligned}
A(\mathbf{m}_{k+1}) &= A(\mathbf{m}_k) + \vec{\nabla}A(\mathbf{m}_k)\Delta\mathbf{m}_k \\
&\quad + \frac{1}{2}\Delta\mathbf{m}_k^T\mathbf{H}(\mathbf{m}_k)\Delta\mathbf{m}_k + \cdots\,,
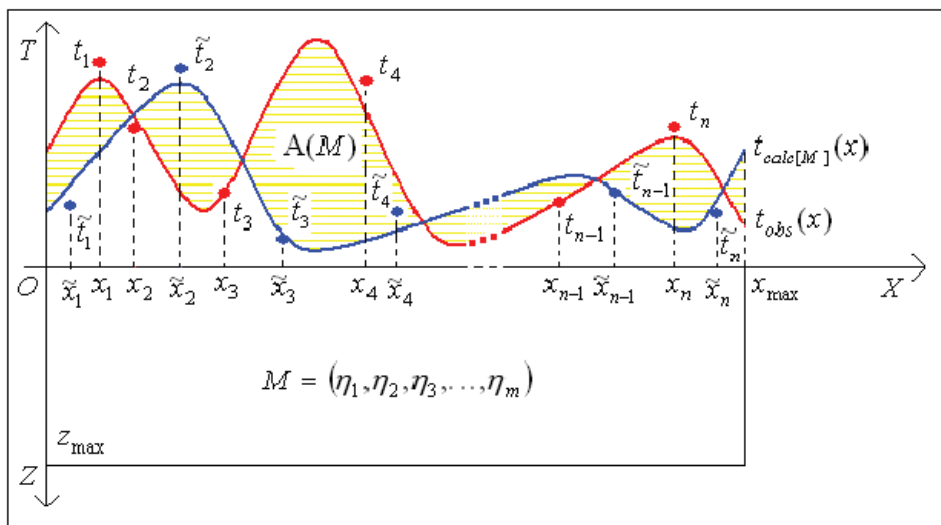\end{aligned} \quad (18)
$$

Figure 3 – Representation of the area (to be minimized employing the L_1 integral norm) between the observed and calculate traveltime curves.

where

$$\vec{\nabla} A(\mathbf{m}_k) = \frac{\partial A}{\partial \eta}(\mathbf{m}_k)$$

$$= \mathbf{g}_k \left[\frac{\partial A}{\partial \eta_1}, \frac{\partial A}{\partial \eta_2}, \ldots, \frac{\partial A}{\partial \eta_m} \right] (\mathbf{m}_k)$$

(19)

and

$$\mathbf{H}(\mathbf{m}_k) = \frac{\partial^2 A}{\partial \eta^2}(\mathbf{m}_k)$$

$$= \begin{bmatrix} \frac{\partial^2 A}{\partial \eta_1^2} & \frac{\partial^2 A}{\partial \eta_1 \partial \eta_2} & \cdots & \frac{\partial^2 A}{\partial \eta_1 \partial \eta_m} \\ \frac{\partial^2 A}{\partial \eta_2 \partial \eta_1} & \frac{\partial^2 A}{\partial \eta_2^2} & \cdots & \frac{\partial^2 A}{\partial \eta_2 \partial \eta_m} \\ \vdots & \vdots & & \vdots \\ \frac{\partial^2 A}{\partial \eta_m \partial \eta_1} & \frac{\partial^2 A}{\partial \eta_m \partial \eta_2} & \cdots & \frac{\partial^2 A}{\partial \eta_m^2} \end{bmatrix} (\mathbf{m}_k)$$

(20)

is the Hessian matrix.

As we expect $A(\mathbf{m}_{k+1})$ to assume a minimum (in fact, the desired value of the minimum we are seeking is zero, as can be seen in (17)), we can make its linear approximation, the two first terms of Equation (18), equal to zero, and so:

$$\Delta \mathbf{m}_k = -(\mathbf{g}_k^T \mathbf{g}_k)^{-1} \mathbf{g}_k^T A(\mathbf{m}_k)$$

(21)

defines an iterative process. The partial derivative $\frac{\partial A}{\partial \eta_1}(\mathbf{m})$ is a real number and is calculated numerically by means of the expression:

$$\frac{A(\ldots, \eta_i + \Delta \eta_i, \ldots) - A(\ldots, \eta_i, \ldots)}{\Delta \eta_i},$$

where $\Delta \eta_i$ is a small perturbation of the parameter η_i. It quantifies the variation of the area between the curves $t_{calc[\mathbf{m}]}(x)$ and $t_{obs}(x)$ when the model parameter η_i is perturbed.

Some Special Comments

To obtain $\frac{\partial A}{\partial \eta_i}(\mathbf{m})$ which is a component of the vector $\mathbf{g}_k$ it is necessary to calculate $A(\mathbf{m})$. This is done by means of the ABS and INT MATLAB routines. However, it is important to say that it is possible to obtain an analytical expression for $A(\mathbf{m})$ and, for the sake of illustration, it is obtained as follows: initially, we have $\Delta T_{[\mathbf{m}]}(x) = t_{obs}(x) - t_{calc[\mathbf{m}]}(x)$. Without loss of generality, we can consider the polynomials (15) and (16) to the same degree, i.e., $p = q$ and $x_{max} > 0$. Let us consider $r_1, r_2, r_3, \ldots, r_u (u \leq p)$ the real roots of $\Delta T_{[\mathbf{m}]}(x)$, where its sign change, and, for the sake of uniformity, $r_0 = 0$

and $r_{u+1} = x_{max}$ (it is possible to choose adjusted polynomials $t_{obs}(x)$ and $t_{calc[\mathbf{m}]}(x)$, such that 0 and x_{max} are not roots of $\Delta T_{[\mathbf{m}]}(x)$).

Then,

$$A(\mathbf{m}) = \int_0^{x_{max}} \left| \Delta T_{[\mathbf{m}]}(x) \right| dx$$

$$= \sum_{i=0}^{u} \int_{r_i}^{r_{i+1}} sgn\left[\Delta T_{[\mathbf{m}]}(x) \right] \Delta T_{[\mathbf{m}]}(x) dx .$$

(22)

This implies:

$$A(\mathbf{m}) = sgn\left[\Delta T_{[\mathbf{m}]}(0) \right]$$

$$\times \sum_{i=0}^{u} (-1)^i \int_{r_i}^{r_{i+1}} \sum_{j=0}^{p} (a_j - b_j) x^j dx$$

(23)

and finally

$$A(\mathbf{m}) = sgn\left[\Delta T_{[\mathbf{m}]}(0) \right]$$

$$\times \sum_{i=0}^{u} (-1)^i \sum_{j=0}^{p} \frac{(a_j - b_j)}{j + 1} \left(r_{i+1}^{j+1} - r_i^{j+1} \right) .$$

(24)

It is interesting to observe that

$$A(\mathbf{m}) = \int_0^{x_{max}} \left| \Delta T_{[\mathbf{m}]}(x) \right| dx$$

$$\leq \sum_{i=0}^{p} \int_0^{x_{max}} |a_i - b_i| x^i dx$$

$$= \sum_{i=0}^{p} \frac{x_{max}^{i+1}}{i + 1} |a_i - b_i|$$

(25)

and if the polynomial coefficients of t_{obs} and $t_{calc[\mathbf{m}]}(x)$ are obtained satisfying the inequality constraint

$$\sum_{j=0}^{7} \sum_{k=j+1}^{8} \frac{x_{max}^{i+k+2}}{(i+1)(k+1)} (a_j - b_j)(a_k - b_k) \geq 0 , \quad (26)$$

we have:

$$\sqrt{\sum_{i=0}^{8}(a_i - b_i)^2} \leq \sqrt{\sum_{i=0}^{8}\left[\frac{x_{\max}^{i+1}}{i+1}(a_i - b_i)\right]^2}$$

$$\leq \sqrt{\sum_{i=0}^{8}\left[\frac{x_{\max}^{i+1}}{i+1}(a_i - b_i)\right]^2} = \left|\sum_{i=0}^{8}\frac{x_{\max}^{i+1}}{i+1}(a_i - b_i)\right| \quad (27)$$

$$= \left|\int_{0}^{x_{\max}}\sum_{i=0}^{8}(a_i - b_i)x^i dx\right| \leq \int_{0}^{x_{\max}}\left|\Delta T_{[\mathbf{m}]}(x)\right|dx$$

$$= A(\mathbf{m}).$$

Equations (25) and (27) shows that $A(\mathbf{m})$ is between the L_2 vector norm and an L_1 vector pondered norm, where the vectors $\mathbf{a} = (a_0, a_1, \ldots, a_8)$ and $\mathbf{b} = (b_0, b_1, \ldots, b_8)$ are composed of the polynomial coefficients of $t_{obs}(x)$ and $t_{calc[\mathbf{m}]}(x)$ with b_i dependent of the model parameters, respectively. Therefore, it represents an intermediary norm that has the robustness and smoothness (near the minimum) that lack to the L_2 vector norm (in case of big residuals) and to the L_1 vector pondered norm (in case of small residuals), respectively. This can help us to understand why good accuracy was achieved in some experiments where it was not expected. In fact, the L_1 integral norm is an L_1-norm with definite derivatives near the minimum and it works like a hybrid L_1/L_2-norm with low computational cost.

As previously mentioned, the formula

$$\sum_{i=0}^{p}\frac{\left|a_i - b_i\right|x_{\max}^{i+1}}{i+1}$$

can be understood as an L_1 vector norm. Despite this, it is not appropriate to improve our understanding of the results. The problem lies in the fact that the vectors $\mathbf{a} = (a_1, a_2, \ldots, a_p)$ and $\mathbf{b} = (b_1, b_2, \ldots, b_p)$ are composed of coefficients of polynomials ("data parameters") and not model parameters, which is not suitable when used as an objective function to be minimized for estimating the parameters of the model $\mathbf{m}$. Moreover, this formula can not be used to calculate $A(\mathbf{m})$, because, in general

$$\left|\sum_{i=1}^{p}(a_i - b_i)x_i\right| \leq \sum_{i=1}^{p}\left|a_i - b_i\right|x^i,$$

seeing that some $(a_i - b_i)x^i$ may be negative, because some

b_i can be larger than a_i and therefore such a situation implies

$$A(\mathbf{m}) \leq \sum_{i=0}^{p}\frac{\left|a_i - b_i\right|x_{\max}^{i+1}}{i+1}.$$

Model Parameterization

Below are three seismic compressional wave velocity field models considered in this work:

$$V_1(x, z) = c_{0,0} + c_{1,0}x + c_{0,1}z, \quad (28)$$

$$V_2(x, z) = c_{0,0} + c_{1,0}x + c_{0,1}z + c_{2,0}x^2 + c_{1,1}xz + c_{0,2}z^2, \quad (29)$$

and

$$V_3(x, z) = c_{0,0} + c_{1,0}x + c_{0,1}z + c_{2,0}x^2 + c_{1,1}xz + c_{0,2}z^2 + c_{3,0}x^3. \quad (30)$$

They are two-dimensional with a horizontal length of 9.0 km and 3.0 km deep. This kind of model representation is able to simulate interfaces by means of strong variations in the velocity field. In order to construct the first target model, the coefficients of Equation (28) are chosen to keep $V_1(x, z)$ within a range of velocities defined by the extreme values 2.0 km/s and 8.0 km/s. To achieve this aim, the coefficient values were found to be: $c_{0,0} = 2.0$ km.s^{-1}, $c_{1,0} = 0.45$ s^{-1}, and $c_{1,0} = 0.66$ s^{-1}. To construct the second target model, with a range of velocities defined by the extreme velocity values 1.0 km/s and 8.0 km/s, the numerical coefficient values used for $V_2(x, z)$ are: $c_{0,0} = 1.0$ km.s^{-1}, $c_{1,0} = 0.098$ s^{-1}, $c_{0,1} = 1.246$ s^{-1}, $c_{2,0} = 0.0285$ km^{-1}.s^{-1}, $c_{1,1} = 0.0015$ km^{-1}.s^{-1}, and $c_{0,2} = 0.0035$ km^{-1}.s^{-1}. Given the same considerations, the numerical value coefficients employed for Equation (30), $V_3(x, z)$ are: $c_{0,0} = 1.0$ km.s^{-1}, $c_{1,0} = 0.058$ s^{-1}, $c_{0,1} = 1.326$ s^{-1}, $c_{2,0} = 0.0195$ km^{-1}.s^{-1}, $c_{1,1} = 0.0016$ km^{-1}.s^{-1}, $c_{0,2} = 0.0055$ km^{-1}.s^{-1}, and $c_{3,0} = 0.0012$ km^{-2}.s^{-1}. The graphical images of these three target models (V_1, V_2 and V_3) are shown in Figures 4, 8 and 12; respectively.

With respect to the application of the proposed method for the resolution of the described inverse problem in this work, it should be said that the geological interpretation of the model is not particularly relevant. Either way, for the above-mentioned coefficients, Equation (28) can be seen to represent a system of sedimentary layers inclined with respect to the horizontal. In an analogous way, the expressions (29) and (30) can represent a similar geological situation with curved interfaces.
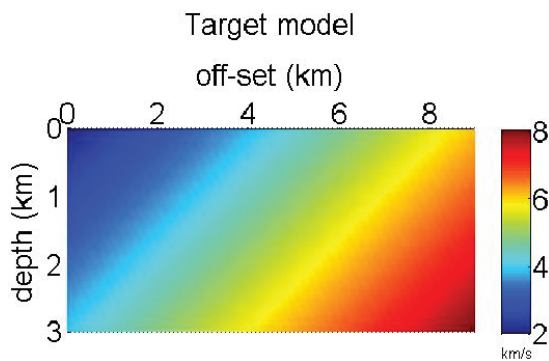
Target model

Figure 4 – Target seismic compressional wave velocity field model, V_1 used in inversion experiments applying both the L_2 vector norm and L_1 integral norm.

RESULTS

For all the situations considered, the source position **s** is kept on the observation surface ($z = 0$) at $x_s = 0.0$ km. Receiver positions, **r** total 18 are distributed on the observation surface in a line obeying the rule $R_i = 0.5 \times i$ km, $i = 1, 2, 3, \ldots, 18$. A seismic ray network for the model provided by Equation (29) is shown in Figure 2. All polynomials interpolated to traveltime data have always degree 8.

For the three models under consideration, the calculated traveltimes are obtained numerically by using Equation (4), and several inversions are performed using the L_2 vector norm. During the inversions it was used a PC with the following configuration: Processor Intel Pentium 4, 3.4 GHz, and RAM 1.0 Gb.

In the case of the model V_1 each iteration of the inversion process takes approximately 1 hour. As the total number of iterations varies between 5 and 9, the total processing time required for convergence is between 5 and 9 hours. This drops dramatically when the L_2 vector norm is substituted by the L_1 integral norm, until it reaches, for the same experiment (with all parameters remaining fixed, except the norm), a total time of calculation is 3 minutes, distributed in 7 iterations, each one a matter of seconds.

With respect to the model V_2 inversions are more unstable and are highly inaccurate in estimating the crossed term $c_{1,1}$. Using L_2 vector norm and the beginning of an initial model in which each parameter has a perturbation of 30%, relative to its correspondent in the target model, the total processing time is 7 hours and 15 minutes, distributed in 7 iterations. However, applying the L_1 integral norm, the same inversion had 20 iterations and a total processing time of 6 minutes. This equates to a 98.85% saving in processing time relative to the inversion using L_2 vector norm.

In the case of model V_1, Figure 5 shows the initial model and its difference in respect of the target model (this difference

is obtained in terms of the least-absolute vector deviation and is approximately 50% when compared to the target model and considering all parameters with the same weight). Employing the L_2 vector norm, a difference of 7% is obtained between the inverted and target models. This difference became 5.21% using the L_1 integral norm. A graphical view of the inverted models and their differences in respect of the target model can be seen in Figures 6 and 7, utilizing the L_2 vector norm and L_1 integral norm, respectively. Results for the considered model can also be seen in Santos & Figueiró (2007).

Initial model

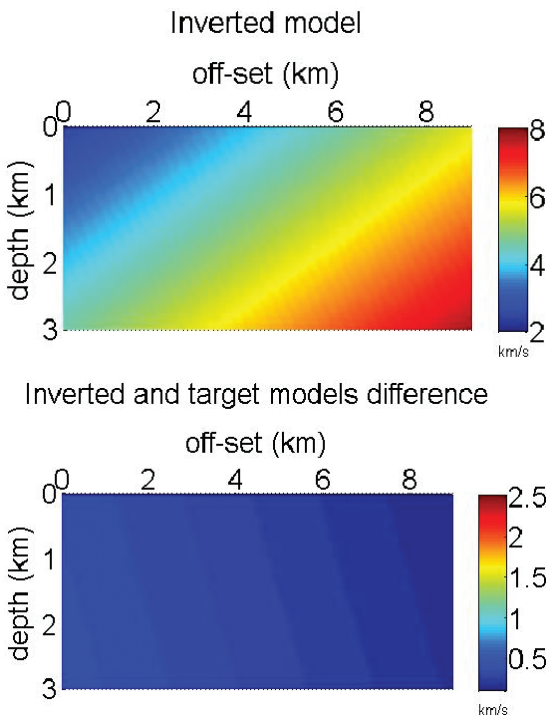## Initial and target models difference
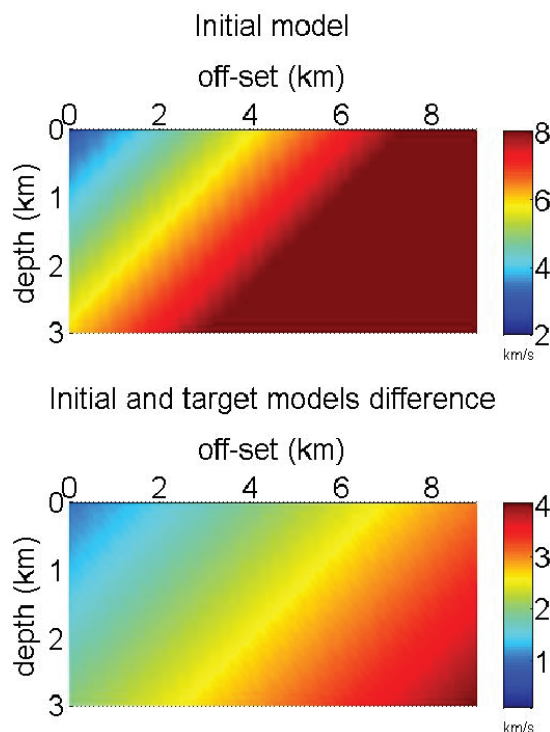
Figure 5 – In this order: initial model, V_1 used in both inversion, and difference between initial and target models. This difference is obtained via least-absolute vector deviation and is 50% when compared to the target model with all parameters being of equal weight.
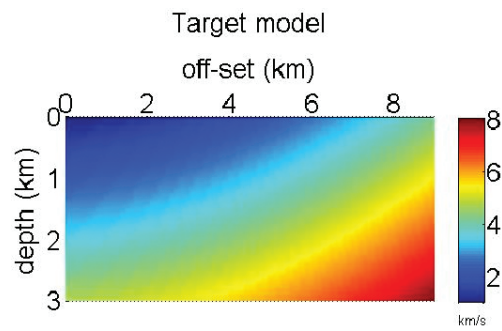
Similarly, an inversion experiment is undertaken involving the model V_2. Figure 9 gives graphical views of the initial model and the difference between initial and target models (calculated using least-absolute vector deviation and being approximately 30% when compared to the target model and taking into consideration all parameters with the same weight). Using the L_2 vector norm, the inverted and target model difference is 19.08% which dropped to 6.47% when the L_1 integral norm was applied. A graphical view of the inverted models and their differences in respect of the target model is showed in Figures 10 and 11, in the case of the L_2 vector norm and L_1 integral norm, respectively.

Inverted model using vector norm

off-set (km)

Inverted and target models difference

off-set (km)
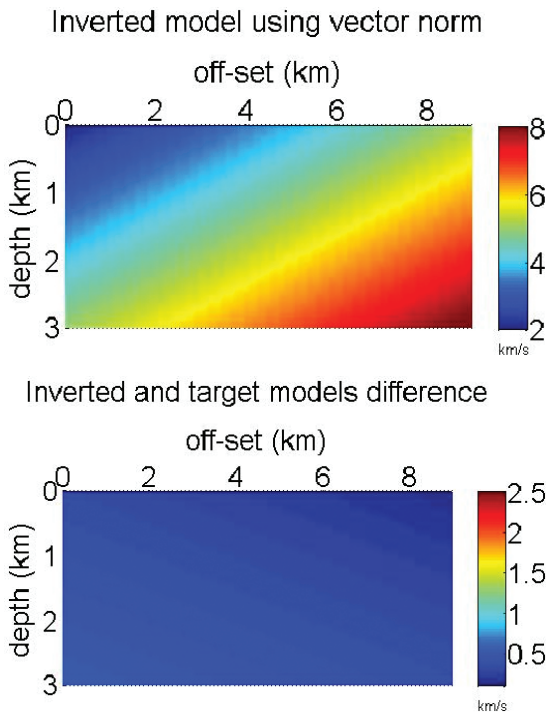
Figure 6 – Inversion of seismic velocity model, V_1 utilizing the L_2 vector norm. In this order: inverted model and difference between inverted and target models. Total processing time: 7 hours (approx.).

Inverted model

off-set (km)

Inverted and target models difference

off-set (km)

Figure 7 – Inversion of seismic velocity model, V_1 using L_1 integral norm. In this order: inverted model and difference between inverted and target models. Total processing time: 3 minutes (approx.).

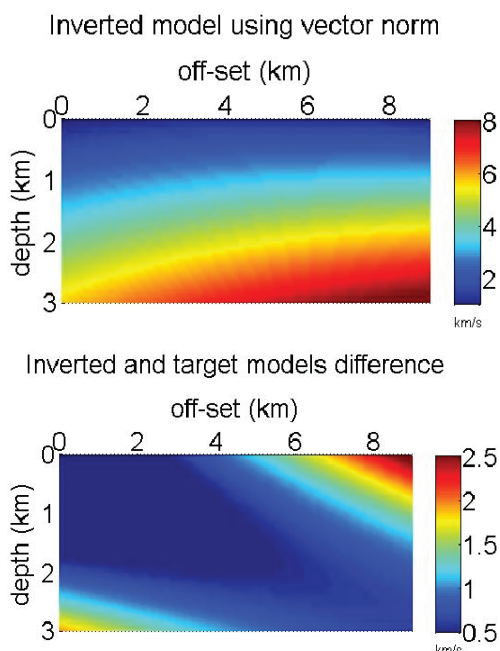## Target model

off-set (km)

Figure 8 – Target seismic compressional wave velocity field model, V_2 employed in inversion experiments using both the L_2 vector norm and L_1 integral norm.

Initial model

off-set (km)

Initial and target models difference

off-set (km)
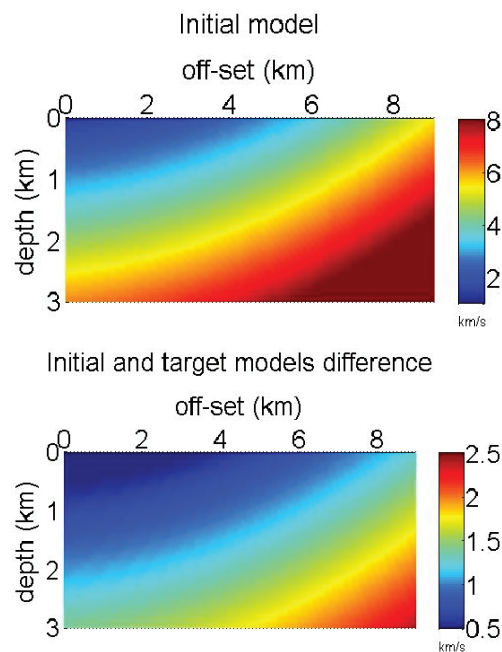
Figure 9 – In this order: initial model V_2 (utilized in both inversions with the two considered norms), and difference between initial and target models. This difference is obtained via least-absolute vector deviation and is 30% when compared to the target model with all parameters being of equal weight.

In addition, the experimental inversion is achieved with the model V_3. Figure 13 gives graphical views of the initial model and of the difference between initial and target models (this difference is calculated by using the least-absolute vector deviation and is approximately 40% when compared to the target model and taking into consideration all parameters with the same weight). Using the L_2 vector norm, the inverted and target model difference is 7.23%, which drops to 6.83% when using the L_1 integral norm. A graphical view of the inverted models and their differences in comparison with the target model is shown in Figure 14 (using the L_2 vector norm) and Figure 15 (using the L_1 integral norm).

Inverted model using vector norm

off-set (km)

Inverted and target models difference

off-set (km)

Figure 10 – Inversion of seismic velocity model V_2 using the L_2 vector norm. In this order: inverted model and difference between inverted and target models. Total processing time: 7 hours and 15 minutes (approx.).

Inverted model using function norm

off-set (km)

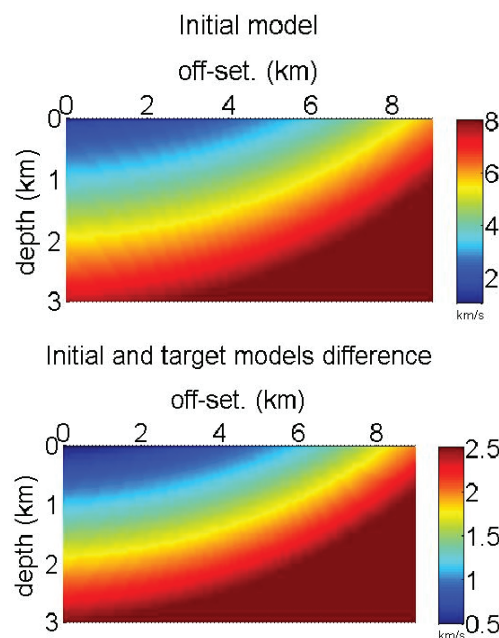## Inverted and target models difference

off-set (km)
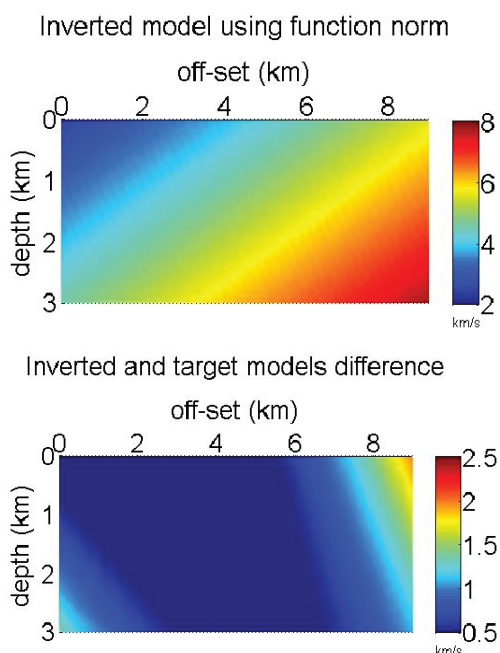
Figure 11 – Inversion of seismic velocity field model V_2 applying the L_1 integral norm. In this order: inverted model and difference between inverted and target models. Total processing time: 5 minutes (approx.).

In this last model, inversions demonstrate high inaccuracy in the estimation of parameters $c_{1,1}$ and $c_{1,2}$, using L_2 vector norm. The same occurs utilizing L_1 integral norm with $c_{2,1}$ and $c_{1,2}$. For each iteration, the processing time using L_2 vector norm is, approximately, 2 hours, however, with the L_1 integral norm, this time was between 12 and 15 seconds. The total time spent in each case, was 9 hours and 50 minutes using L_2 vector norm and 5 minutes and 15 seconds using L_1 integral norm.
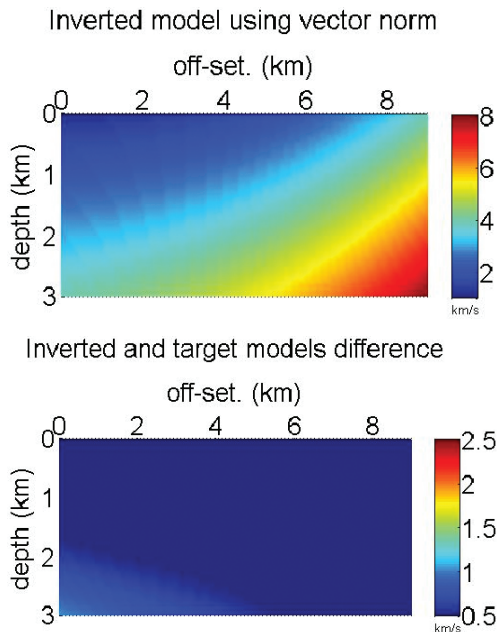
Target model

off-set (km)
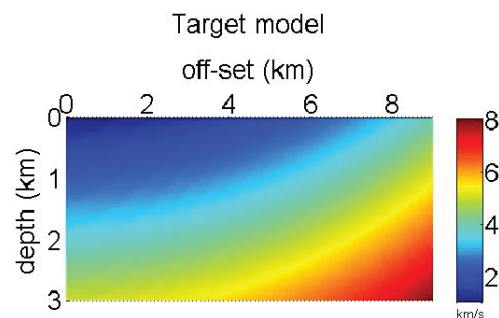
Figure 12 – Target seismic compressional wave velocity field model, V_3 used in inversion experiments employing both the L_2 vector norm and L_1 integral norm.

Initial model

off-set. (km)

Initial and target models difference

off-set. (km)

Figure 13 – In this order: initial model V_3 (used in both inversion: with L_2 vector norm and L_1 integral norm), and difference between initial and target models. This difference is obtained via least-absolute vector deviation and is 40% when compared to the target model with all parameters being of equal weight.

In the case of L_1 integral norm, occasionally, during the iterative process, the matrix $\mathbf{G} = \mathbf{g}_k^T \mathbf{g}_k$ becomes singular. To overcome this problem a regularization method known as Levenberg-Marquardt is carried out, involving the use of the matrix $\mathbf{G} + \lambda \mathbf{I}$ in place of $\mathbf{G}$, where λ is a small positive real number and I is the identity matrix. One way of avoiding numerical artifacts being introduced by regularization is to apply the Newton method. In this case the matrix $\mathbf{G}$ will often not be

singular. Experiments with model V_1 are carried out using the same iterative process defined by the Newton method, employing the L_1 integral norm. The results obtained also presented better accuracy when compared with a similar experiment using the L_2 vector norm, with singular matrices not appearing during the iterative inversion process.
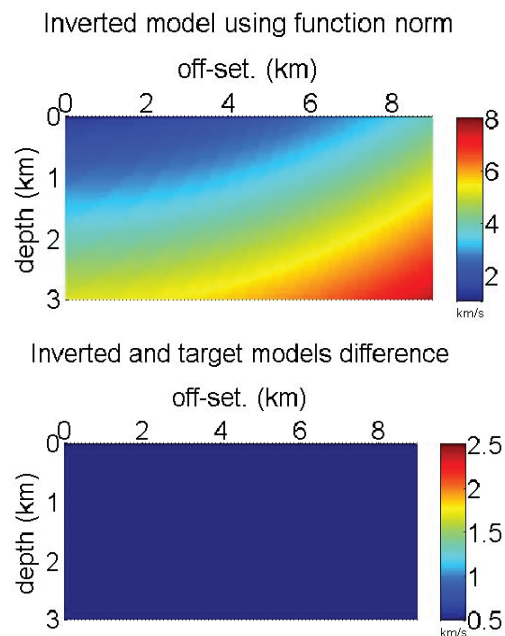
Inverted model using vector norm

Figure 14 – Inversion of seismic velocity model V_3 employing the L_2 vector norm. In this order: inverted model and difference between inverted and target models. Total processing time: 9 hours and 50 minutes (approx.).

DISCUSSIONS

There are scientists who are in agreement that two-point ray tracing is a time-consuming procedure and even dangerous in the vicinity of caustics when using methods like bisection, but disagree about the presence of the stated problems when the method is based on paraxial approximation (Figueiró & Madariaga, 2000). In respect of these restrictions it could be said that the paraxial method is iterative like bisection, yet, depending on the specific situation, it can be as time-consuming as any other method. In addition, the problem with caustics does not disappear when we use the paraxial method. For theoretical, experimental, specific, or intrinsic use, two-point ray tracing, could possibly be a relatively non time-consuming procedure. However, for tomographic use it is undoubtedly an enormous time consumer. At least 70% of processing time is wasted solving the two-point ray tracing problem, if we assume that we do not want to abuse approximation techniques. It should be stated that in tomography the two-point ray tracing problem must be solved

for each source-receiver pair. It is then solved many times over in all iterations of a tomographic process. In any case, we are not proposing the complete abandoning of ray tracing in tomography as rays are inherent in tomography. We are just proposing a better use of rays in tomography. In the same way, this work is not a manifesto against ray tracing as we are simply comparing two kinds of norms: this is the essence of the work. If we use an integral norm, ray tracing must be adapted to be conveniently used by such a norm. If we use a more efficient procedure for ray tracing the inversion will be faster.

Inverted model using function norm

Figure 15 – Inversion of seismic velocity model V_3 using the L_1 integral norm. In this order: inverted model and difference between inverted and target models. Total processing time: 5 minutes and 15 seconds (approx.).

The models utilized cannot be considered unrealistic for the following reasons: they take into consideration a large range of seismic velocities from 1.0 to 8.0 km/s (which, for practical purposes, is almost the total seismic velocity found in real situations); the polynomial parameterization process creates an infinite amount of possibilities to represent the heterogeneities of seismic velocity fields found in a huge variety of real geological situations; our models are a long way from the set of models considered simple; they have strong horizontal and vertical variations (at all points); every continuous function defined in a compact set is the limit of a sequence of polynomial functions (Stone-Weierstrass theorem, see Bartle (1983)); and field discontinuities can be considered as strong variation in order to preserve continuity.

When using integral norm, we benefit from a property of polynomial function, i.e., that polynomials can be easily integrated. Figure 3 has a central role, because it expresses the essence of the work and shows the area that we want to minimize. However, polynomials are being used to represent data and models, so, in this second case, the oscillatory character of polynomials can cause a few problems during ray tracing, and for this reason we tend to avoid very high-degree polynomials. In addition, as the coefficients of polynomials can vary freely, such parameterization can represent a very large family of models: in fact, an infinite set of models. Oscillations of polynomials can be controlled since we can impose a small value for its derivative. Polynomials can represent realistic though not real oscillations.

For the sake of simplicity, and desire to isolate the experiment from other potential problems, discontinuities in the synthetic observed traveltime data are avoided. However, if the presence of a certain kind of problem is required for a more general experiment, calculated data must be considered to also allow discontinuities, in order to have appropriate conditions for comparison and adjustment.

CONCLUSIONS

The use of the L_1 integral norm, instead of the L_2 vector norm, in the seismic ray tomography procedures, produces an impressive saving in processing time without any significant loss of accuracy (in some cases accuracy actually improves). This saving occurs because when the L_1 integral norm is used, it is not necessary to solve the two-point ray tracing problem several times in all iterations of the inversion process. The employed parameterization model is the polynomial, but we believe that this saving of processing time is independent of whatever parameterization is used. Episodic instabilities occur utilizing the L_1 integral norm, but this is not especially problematic, as opposed to other cases that take place using a different norm. For conceptual reasons and because of strong results, we believe this study will work well when applied to more complicated and realistic models. In addition, integrals and, more importantly, areas, have a tendency to be more stable (possibly due to their global character) than vectors and points that have just local properties.

ACKNOWLEDGEMENTS

REFERENCES

BARTLE RG. 1983. Elementos de Análise Real. Editora Campus Ltda. 429 pp.

BISHOP TN, BUBE KP, CUTLER RT, LANGAN RT, LOVE PL, RESNICK JR, SHUEY RT, SPINDLER DA & WYLD HW. 1985. Tomographic determination of velocity and depth in laterally varying media. Geophysics, 50: 903–923.

BUBE KP & LANGAN RT. 1997. Hybrid L_1/L_2 minimization with applications to tomography. Geophysics, 62: 1183–1195.

ČERVENÝ V. 1987. Ray Methods for Three-Dimensional Seismic Modeling. Lecture notes, Continuing Education Course, University of Trondheim, NTH and Mobil Exploration Norway Inc., Trondheim, 830 p.

CLAERBOUT JF & MUIR F. 1973. Robust modeling with erratic data. Geophysics, 38: 826–844.

FIGUEIRÓ WM. 1994. Tomografia de Reflexão no caso de Refletor Curvo. Tese de Doutorado, PPPG-UFBA, Salvador. 150 pp.

FIGUEIRÓ WM & MADARIAGA RI. 2000. A method to avoid arrival caustic points. In: 2000 Technical Program Expanded Abstract, SEG 70th Annual Meeting, Calgary, Canada. CD-ROM.

GUITTON A & SYMES WW. 2003. Robust inversion of seismic data using the Huber norm. Geophysics, 68: 1310–1319.

LORIS I, NOLET G, DAUBECHIES I & DAHLEN FA. 2007. Tomographic inversion using L_1-norm regularization of wavelet coefficients. Geophys. J. Int., 170: 359–370.

LORIS I, DOUMA H, NOLET G, DAUBECHIES I & REGONE C. 2010. Nonlinear regularization techniques for seismic tomography. J. Comp. Phys., 229: 890–905.

MENKE W. 1989. Geophysical Data Analysis: Discrete Inverse Theory. Academic Press, New York, USA. 289 pp.

SANTOS VGB. 2009. Inversão Sísmica Tomográfica usando Norma de Integral de Função. Tese de Doutorado, CPGG-UFBA, Salvador. 91 pp.

SANTOS VGB & FIGUEIRÓ WM. 2006. Inversão sísmica tomográfica usando traçado analítico de raios. In: II Simpósio Brasileiro de Geofísica, Natal, CD-ROM.

SANTOS VGB & FIGUEIRÓ WM. 2007. Seismic ray reflection tomography using integral function norm. In: 2007 Technical Program Expanded Abstract, SEG 77th Annual Meeting, San Antonio, Texas, USA. CD-ROM.

SCALES JA & GERSZTENKORN A. 1988. Robust methods in inverse theory. Inverse Problems, 4: 1071–1091.

NOTES ABOUT THE AUTHORS

Vânia Gonçalves de Brito dos Santos. Graduated (1992) in Mathematics from Catholic University of Salvador. She received an M.Sc. (1999) in Mathematics (differential geometry) and a Ph.D. (2009) in geophysics (seismic tomography) both from Federal University of Bahia. She is Mathematics Professor at the State University of Feira de Santana since 2001. Her research interests are geophysics and geometry.

Wilson Mouzer Figueiró. B.Sc. (1983) in Mathematics from Federal University of Rio de Janeiro (UFRJ). Since 1992, he has been Geophysics Professor at the Geophysics Department of the Geoscience Institute at the Federal University of Bahia (UFBA), where he teaches mathematical geophysics. He obtained his Ph.D. in Geophysics (1994) from UFBA. From 1997 to 1999, he did post-doctoral studies at the Earth-Atmosphere-Ocean Department of the Superior Normal School of Paris (DTAO-ENS), France. Now, he is researcher at the Geophysics and Geology Research Center (CPGG) of the UFBA. His main areas of interest are: geophysical inversion and mathematics applied to geophysics.