BAYESIAN METHODS FOR A PARTICULAR INVERSE PROBLEM: SEISMIC TOMOGRAPHY

ANA K. FERMÍN, JEAN-MICHEL LOUBES AND CARENNE LUDEÑA

Abstract. We propose a stochastic technique to solve the problem inverse of seismic tomography in heterogeneous media with known and unknown reflector depths. This technique combines a Bayesian algorithm with simulated annealing. The advantage of this technique is that it allows to reconstruct a velocity-depth model that minimizes the travel times error, keeping in view the physical restrictions imposed by the properties of the model. Instead of working directly in an infinite dimensional frame, we discretize the problem. Hence, at step n the inverse problem consists to estimate the amplitude values of the unknown velocity at n sites. The main result is that asymptotically, the Bayesian procedure is equivalent to selecting a quasi solution of original inverse problem that maximizes a convex criterion.

1. Introduction

This work deals with slowness field estimation in seismic tomography. One of the main issue in seismology is estimating the exact nature of a soil. It is usually performed by estimating the different velocities at which a sound wave may propagate through the different layers of the medium, each speed corresponding to different soils. For this, the following experiment is conducted. At a fixed point called the *source*, an explosion is generated, while at some observation points called the *receivers*, are measured the time necessary for the sound waves to reach these different points. So we observe travel times of several waves propagating in an heterogeneous field, namely $U \subset \mathbb{R}^2$. These times depend on an unknown function $S^\star : U \to \mathbb{R}$ describing the inverse of the velocity of the wave propagation in a medium U. It is a characteristic of the field to be studied and is called the slowness field. Hence consider the following mathematical model

$$T_i^{obs} = \Phi(S^\star, r_i) + \epsilon_i,\ i = 1, \dots, k. \tag{1}$$

For $i = 1, \dots, k$, T_i^{obs} is the observed travel time, needed for the ray i following the path r_i to travel through the field, measured with observation noise. In the whole paper ϵ_i, $i = 1, \dots, k$ are assumed to be an i.i.d sample of a Gaussian random variable $\mathcal{N}(0, 1)$.

Hence we are facing a nonlinear inverse problem since we aim at building an estimate of the slowness field (or in an equivalent way of the slowness field), observed through its image by a non linear operator Φ.

Seismic Reflexion Tomography has been widely studied in the literature. As it is very well known, seismic tomography problems lead to mathematical models that belong to the family of ill-posed problems. Since the operator Φ is not continuously invertible, solutions of the inverse problem are unstable under data perturbations. Let $\|.\|$ be the quadratic norm over

Key words and phrases. Ill-posed Problem, Seismic Tomography, Large Deviations, Bayesian Estimator.

the set U. So, even in a deterministic framework, finding S such that for a small $\delta > 0$, $\|\Phi(S, r(S)) - \Phi(S^\star, r(S^\star)\| \leq \delta$ does not assure that S and $S^\star$ are close. So, to deal with this issue, regularization methods replace an ill-posed problem by a family of well-posed problems. Their solution, called regularized solutions, are used as approximations of the desired solution of the inverse problem. These methods always involve some parameter measuring the closeness of the regularized and the original (unregularized) inverse problem. Rules (and algorithms) for the choice of these regularization parameters as well as convergence properties of the regularized solutions are central points in the theory of these methods, since they allow to find the right balance between stability and accuracy. We refer to [10] or [22] for general methods to solve ill-posed inverse problems. In the linear case, rates of convergence are given in [6] or [11]. Nonlinear inverse problems are tackled in a restrictive case in [20] or in [18] but seismic tomography is an even more complicated issue as shown later in that paper.

In the framework of this study, the problem is very ill-posed in the sense that the operator is unknown and depends of the function to be estimated. Indeed reflection laws for sound wave imply that the path itself depends on the parameter of interest $S^\star$ and $r = r(S^\star)$. So, standard techniques for solving inverse problem can not be applied. To the authors knowledge few is done in the statistical literature. However the situation we are facing is similar to inverse problems where some generalized moments of an unknown positive measure are observed. To reconstruct the measure, first it is discretized and then estimated on each cell. In their work, Gamboa and Gassiat in [13], [9] or [14] use Bayesian procedures to tackle such issue. More recently Gozlan in [15], extends previous results by adding thiner constraints.

Here, following the same ideas, we introduce a construction called the maximum entropy which relies on a suitable sequence of finite-dimensional discretized inverse problems. This Bayesian procedure is well suited for such difficult ill-posed issued. Indeed the ill-posedness entails for all admissible data either that there is no unique solution or that the solution does not depend continuously on the data. So Bayesian techniques enable us to restrict the space of parameter by adding a constraint corresponding to a suitable a priori. That is the reason why Bayesian methods have gained a strong popularity in seismic tomography since Bayesian inference can naturally give us all the necessary tools to solve real inverse problems.

The slowness function to be reconstructed is a function $S(x, y)$ defined on a compact set $U \subset \mathbb{R}^2$. We divide the region of interest U into a matrix of cells c_i^n, $i = 1, \ldots, n$ and approximate the value of the slowness function in that box by the value at the center of the box, say s_i. Hence our goal in this paper is to estimate the values of $S = (s_1, \ldots, s_n)$ from the observed travel times $T_1^{obs}, \ldots, T_k^{obs}$. To this discretized function is associated a measure on U. Due to the physical nature of the issue, Bayesian techniques define probabilities on the space of models (a priori information) conditioned on the observed data. For example, in a seismic survey we may have a fairly accurate idea of the realist ranges of seismic velocity (or slowness) and after discretization take these initial values and consider the associated measure as a prior. Then we need to be able to build, using a prior for the slowness field, a sample of travel times corresponding to this distribution. It is achieved using a linear approximation of the medium, using a tracing ray algorithm.

Then, we take as an estimator the posterior mean of the distributions which are constrained to provide realistic travel times, realistic with respect to the real observed travel times. As a matter of fact, for a given small constant $\delta > 0$, we will add the constraint that for each simulated travel time $T(S)$ with prior S, we have

$$\|T^{obs} - T(S)\| \leq \delta.$$

Our aim in this article is to estimate seismic velocity in varying media, from wave travel times. The medium is composed of layers with different characteristics. So, we will consider two main cases; the homogeneous case, where to each layer is associated a velocity and the heterogeneous case where velocity is not constant in a layer but may vary. The estimator we define, is similar to the estimator given in [13]. Using their result, we prove consistency in the homogeneous case. In the heterogeneous case, there is no unique minimizer of the previous constraint and hence we use a penalized criterion to select among the different candidates. It is solved computationnaly using a stochastic minimization program, a simulated annealing. Such method is sensitive to initialization points. A good starting point is given by previous Bayesian estimator. The combination of these two algorithms provides a good estimator of the field of slowness in both cases. So, stochastic algorithms provide a constructive method to solve such very ill-posed inverse problems.

The paper falls into 5 main parts. In Section 2, we present the inverse problem and explain how to simulate the propagation of a wave in a discretized environment, given a slowness field. Section 3 is devoted to the construction of the Bayes procedure. In Section 4 we give the asymptotic properties of the estimator, while simulations are given in Section 5. We consider simulation for the three cases: homogeneous and heterogeneous cases with known layer and heterogeneous case with unknown layer. In this last case, we apply our algorithm to estimate both the slowness and the number of layers in the medium.

2. Seismic Travel Time Tomography: presentation of the model

Definition of travel times:

The time needed by a sonic wave to travel from a source to a receiver is called the travel time. The travel time of a seismic wave is the integral of slowness along a ray path connecting the source and the receiver, as follows

$$T = \Phi(S, r) = \int_r S(x)\, dl, \tag{2}$$

where r denotes an arbitrary ray path, $x \in \mathbb{R}^2$ is the position vector and dl is its differential path length. Here the term slowness, S, stands for the inverse of the wave velocity. It is more convenient to write inversion and tomography formulas in terms of wave slowness models, because the pertinent equations are linear in slowness. The calculation of ray travel times between known end points through a given velocity structure is often called the forward problem. It has been widely studied, see for instance the work by [4] or [12].

When more than one ray path exist between a given source and receiver, the path with minimum travel time is the one usually required because first-arrivals are always easier to identify on a seismogram. The equation (2) is formulated by Fermat's principle, which states that of all the paths that join two points A and B in a velocity medium, the true ray paths will be stationary in time. In other words, the correct ray path between two points is one which minimizes $T(S, r)$ with respect to the path r, i.e,

$$T(S) = \inf_r \int_r \Phi(S, r)\, dl = \int_{r^*} S(x)\, dl \tag{3}$$

with r^* the particular path that produces a mimimum travel time.

The difficulty in performing this integration is that, first, the path taken by the seismic energy depends on the velocity structure $r^\star = r^\star(S)$. Moreover, the path needs to be known in order to evaluate the integral. This means that the inverse problem can be very difficult to solve. So, Equation (3) is nonlinear since the integration path depends on the slowness.

There are three basic approaches used in tomography literature, defined as
(i) linear tomography,
(ii) iterative nonlinear tomography,
(iii) fully nonlinear tomography.

In linear tomography, the relationship between travel time residual and velocity perturbation is linearized around a reference model and corrections to the velocity field are made under this assumption. Thus, ray paths are determined only once (through the initial or reference model) and are not retraced. Iterative nonlinear tomography also ignores the path dependence of the velocity correction in the inversion step, but accounts for the nonlinearity of the problem by iteratively applying corrections and retracing rays until, for example, the data are satisfied, or the rate of data fit improvement per iteration satisfies a given tolerance. Fully nonlinear tomography locates a solution without relying on linearization in any way, but is rarely done in practice. Fully nonlinear inversion is required for problems with significant slowness variations across the region of interest.

In this problem we suppose that the ray paths are known a priori, i.e, we suppose that, for a fixed S, there is none dependence of ray on the slowness distributions. However, in our methodology, we will allow ourselves to let S change, according to some constraints and construct a convergent estimator in a Bayesian scheme.

Discretization of the slowness field:

Let c_j^n, $j = 1, \ldots, n$ be a discretization of a 2-D cell model U such that the sequence of discrete measures $P_n := \frac{1}{n}\sum_{j=1}^n \delta_{c_j^n}$ converge weakly to some given probability measure P on U. We assume that P does not allocate any weight at any point of U. Physically, at step n, U is considered as n sites or cells(think is as an image, where a site is a pixel), where to each site c_j^n is associated a slowness s_j. For sake of simplicity, we will write c_j although the cells change when the discretization level increases. Given a discretized model of slowness with n cells $c_1, \ldots, c_n$ and $S = (s_1, \ldots, s_n)$, consider the slowness function

$$\forall M \in U,\ S_n(M) = \frac{1}{n}\sum_{j=1}^n s_j \delta_{c_j}(M),$$

and the corresponding associated measure

$$\nu_n = \frac{1}{n}\sum_{j=1}^n s_j \delta_{c_j}.$$

let l_{ij} be the length of the $i-$th ray path through the cell c_j

$$l_{ij} = \int \mathbf{1}_{r_i \cap c_j} dl.$$

Thus, the equation $T(S) = \Phi(S, r)$ can then be written as

$$T = \sum_{j=1}^{n} s_j l_{ij}, \ \ i = 1, , \ldots, k \tag{4}$$

The discretized model (4) can be rewritten, using matrix notation as

$$T = AS; \tag{5}$$

where S, T y A are defined as follows

$$S = \begin{pmatrix} s_1 \\ s_2 \\ \vdots \\ s_n \end{pmatrix}; \qquad T = \begin{pmatrix} T_1 \\ T_2 \\ \vdots \\ T_k \end{pmatrix}; \qquad A = \begin{pmatrix} l_{11} & l_{12} & \ldots & l_{1n} \\ l_{21} & l_{22} & \ldots & l_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ l_{k1} & l_{k2} & \ldots & l_{kn} \end{pmatrix}. \tag{6}$$

Hence Equation (5) may be viewed as a discretized form of the equation (3). In linear tomography, for a given slowness field we are able to compute A and T. The assumption here is that the ray paths are known a priori, which is justified under a linear approximation that ignores the dependence of the ray paths on the slowness distribution. In this work the ray paths are assumed to be straight lines connecting sources and receivers The principal problem is that the matrix A is often poorly conditioned. Indeed, the matrix $A = (l_{ij})$ is in general very sparse for topographic problems because every ray intersects only a small fraction of the cells. This is in particular the case in three dimensions where the relative number of intersected cells is much smaller than in two dimensions. In Figure 1, is shown an example of discretization with some propagating waves in the medium.

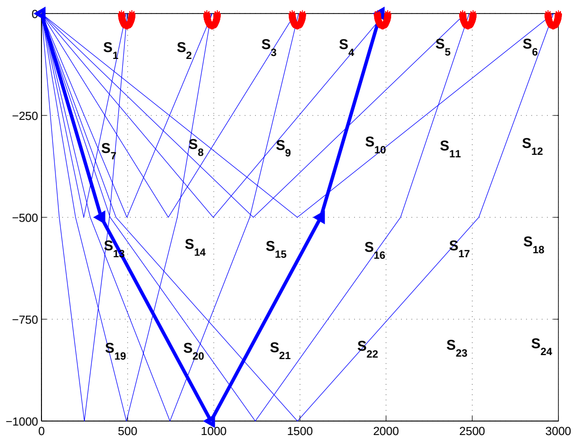

FIGURE 1. Wave in a discretized medium

Our aim is to estimate $S^\star$ by comparing the values of travel times obtained for different S and selecting among them the one closest to the true measurements. Hence, we need to be able to compute the travel times for different values of S. This is the purpose of the tracing ray.

Tracing Rays with Shooting scheme:
The seismic ray tracing problem is based on a "shooting rays" method, see for example [25] or [7]. Shooting methods of ray tracing rely on formulating the ray equation as an initial value problem, where a complete ray path can be determined provided the source coordinates and initial ray direction are known. That is, the shooting angles of the ray at the source point are increased until the ray ends sufficiently close to the receiver. The problem is then solved by shooting rays through the medium from the source and using information from the computed paths to update the initial ray trajectories so that they more accurately target the receivers.

We present the problem of tracing rays in 2D heterogeneous media. Each optimization problem is obtained by applying Fermat's principle to an approximation of the travel time equation from a fixed source to a fixed receiver. We assume a piecewise linear raypath which simplifies of the problem. Figure 2 shows an example of shooting ray diagram. We only drew rays which reach a receiver. Actually, many simulated waves are lost in the sense that they can not be measured by the receivers given in our model. This point is crucial in the simulations since it increases drastically the number of simulations. Moreover, the cells at the bottom of the medium are less visited by the rays than the cells of the first layer. It implies that it is more difficult to estimate velocities far from the surface than the velocities in the first layers.

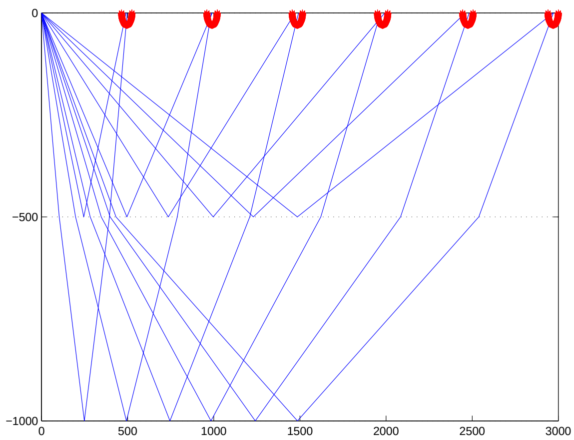

FIGURE 2. Example of shooting ray with two layers

3. Estimation method: statistical model

Consider the following statistical model. We assume that the data are noisy and observe

$$T_i^{obs} = \int_{r_i^\star} S^\star(x)\, dl + \varepsilon_i,\ i = 1, \dots, k, \tag{7}$$

where $T_1^{obs}, T_2^{obs}, \dots, T_k^{obs}$ is a set of observed travel times, from k source-receiver pairs in a medium, $U \subset R^2$, with slowness $S^\star(x)$. Let $r_i^\star$ be the Fermat ray path connecting the

ith source-receiver pair, i.e, the ray follows the trajectory of minimum travel time between the source and the receiver, as explained in Section 2. The inverse problem can be seen as recovering an unobservable signal $S^\star$ on U based on observations T_i^{obs}. Throughout the paper, we shall denote $T^{obs} = (T_i^{obs})_{i=1}^k$ and we shall suppose that the observations noise ε_i are i.i.d. realizations of a certain random variable ε. We will focus on the case where the distribution of ε is assumed to be Gaussian. It is a loss of generality but due to the very ill-posedness of the model (even for $\varepsilon = 0$ the problem is in general highly not invertible and there is no unique solution), it is important to restrict the law of the noise and just focus on the estimation problem (7).

In a deterministic framework, the best possible accuracy, regardless of any discretization and noise corruption is determined by some *a priori smoothness assumption* on the exact solution $S^\star$. The statistical model (7) is formulated as the problem of finding the best-approximate solution of $T_i^\star = \int_{r_i^\star} S^\star(x)\, dl$ in the situation where only perturbed data T^{obs} are available with

$$\|T^\star - T^{obs}\| \leq \delta.$$

Here, δ is called the noise level.

A common approach is to minimize the l^2-norm (squared) of the error

$$\|T(S) - T^{obs}\|^2 = \sum_{i=1}^{k} [(T(S) - T^{obs})_i]^2 \tag{8}$$

where $T(S)$ denote travel times for admissible S, calculated using the technique of the ray shooting tool described previously. Admissible means here that a constraint must be added to this minimization otherwise the corresponding estimator will not be consistent due to ill-posedness of the problem. Here, introducing a Bayesian prior is well suited to model this constraint.

Using ideas arising from the large deviations theory, we propose and study a general Bayesian method to solve the inverse problem of seismic tomography (7). A Bayesian interpretation can be given using the distribution of Gaussian process as priors on the function S, as Wahba shows, for instance in [29]. Other priors are possible in order to take into account the correlations into a medium but we want to be able to apply our procedure in the case where few information is available and so where a Gaussian prior can be a good initial guess.

Hence we construct an estimator using the following scheme.

- We discretize the medium and try to estimate the finite dimensional parameter $S_n = (s_1, \ldots, s_n)$.
- For well chosen slowness fields (i.e depending on a prior), $S_n^{(i)} = (s_1^{(i)}, \ldots, s_n^{(i)})$, $i = 1, \ldots, N$, we compute the approximate travel times $T(S^{(i)})$ using the tracing ray.
- For a given $\delta > 0$, we keep the admissible travel times, satisfying the constraint and define
$$\mathcal{I}_n = \{i = 1, \ldots, N,\ \|T^{obs} - T^{(i)}\| \leq \delta\}.$$

- The Bayesian estimator is then defined as

$$\hat{S}_n^{bay} = \frac{1}{\mathrm{Card}(\mathcal{I}_n)} \sum_{i \in \mathcal{I}_n} S_n^{(i)}. \tag{9}$$

This algorithm computes a Bayesian estimator similar to the one studied by Gamboa and Gassiat in [13]. As n, the discretization level, grows to infinity, this algorithm provides an estimator, converging to the set of minimizers of (8) under the admissibility constraint for a fixed δ, as shown in the next section.

If the medium is homogeneous, then the minimizer is unique, which guarantees the convergence of the estimator, to the true slowness field $S^\star$.

If the medium is heterogeneous, the set of minimizers is non trivial and local minima may exist. Hence, to get consistency of our method, we must select among all the minimizers, the closest to the true parameters. For this, we consider the maximum entropy estimator defined as solution of the minimization problem:

$$\hat{S}_n^{Pen} = \arg \min_{S,\, \|T^{obs} - T(S)\| \le \delta} \left(\|T^{obs} - T(S)\|^2 + \lambda J(S) \right) \tag{10}$$

where λ is a smoothing parameter and J is a suitable functional. Provided an appropriate starting point is chosen, this minimization can be conducted and enables us to construct an estimator of $S^\star$. This initial starting point is given by the Bayes estimator $\hat{S}_n$, close enough to the solution. In the next section, we discuss the choice of the penalty function.

4. Asymptotic behavior of slowness field estimator

Solving any inverse problem requires understanding the uncertainties in the data. We also need methods to incorporate a prior information to eliminate unreasonable models that fit the data. In this section we present two methodologies to solve the inverse problem that can be used to include priori information in the inversion process.

4.1. **Bayesian procedure.**

Maximum entropy on the mean (MEM) is a construction where the inverse problem (7) is approximated by a sequence of finite-dimensional problems, which are obtained by a discretization of the space U. The MEM estimator is then obtained as the limit of the discretized estimators defined for the finite-dimensional problems. The MEM construction gives a natural and practical way of introducing nonlinear but convex constraints considered as prior information. Often, one may want to incorporate some prior information on the slowness model S, for example we may know from previous experiments that S has known support. A way to incorporate this kind of prior information is to consider a probability P on U which heavily weights areas where we suspect S to be concentrated.

To solve the inverse problem, we consider a Bayesian reconstruction method which is used usually in statistics. To compute a Bayesian estimator for the inverse problem (7) we proceed as follows. To each site c_j^n, $j = 1, \ldots, n$ is associated a random variable s_j (s_j is a random slowness). We assume that F_n is the prior distribution of the random vector $S_n = (s_1, ..., s_n)$. In this work we suppose that the unknown slowness S is positive and that S has a density function f (with respect to some known probability P). The density of the prior information is bounded $a(x) < f(x) < b(x)$. These values correspond to bounds given by engineers in petroleum industry. Then, for every j, we choose the support of the distribution F_j of s_j to be included in $[a(c_j^n), b(c_j^n)]$ and if we have no more information define $F_n = \bigotimes_{j=1}^n F_j$, and choose for F_n a uniform probability on U. Alternatively, without any knowledge, we often assume that the prior distribution on the S is Gaussian.

We define at step n the corresponding discrete random measure

$$\nu_n := \frac{1}{n} \sum_{j=1}^{n} s_j \delta_{c_j^n}. \tag{11}$$

We point out that there is a clear correspondence between the measure $\nu_n = \frac{1}{n}\sum_{i=1}^n s_i \delta_{c_i}$ and the slowness $S_n = (s_1, \ldots, s_n)$. Hence to the Bayesian estimator of the slowness field defined in (9) is associated the Bayes estimator of the measure $\hat{\nu}_n^{bay}$. Large deviation theory enable to prove weak convergence of the measure, which implies the convergence of the corresponding slowness field.

Let us now introduce the estimator of the unknown $S^\star$ when noisy data are observed, as

$$\hat{\nu}_n^{bay} = E_{\mathcal{F}_n}(\nu_n \mid \|T^{obs} - T(S_n)\|^2 < \delta^2). \tag{12}$$

where $\mathcal{F}_n$ is the empirical prior distribution of ν_n. That is, at step n we only considerer positive measures supported by c_i^n, $i = 1, \ldots, n$.

For every level n of the discretization of the space U, $\hat{\nu}_n^{bay}$ is a Bayesian estimator with a prior $\mathcal{F}_n$. We then have a sequence of n-dimensional Bayesian problems with a k-dimensional observation. It is important to remark that we are not interested, in this article, in the asymptotic behavior of the estimate when the number of the observations increases nor when the noise level decreases, i.e, we are not interested in the asymptotic on the size of the sample. Our asymptotic will be on the number of sites, namely n.

In order to understand the asymptotic behaviour of the Bayes estimate, we need to introduce the MEM estimate: at stage n, we choose the distribution P_n^{MEM} of S_n using a maximum entropy principle. We refer to [13] or [Dacunha-Castelle and Duflo (1986)] for general references. Set

$$\hat{\nu}_n^{MEM} := \mathbb{E}_{P_n^{MEM}}(\nu_n).$$

So for fixed n, $\hat{\nu_n}^{MEM}$ is the maximum entropy reconstruction of S with reference measure F_n. When n tends to infinity, under some additional assumptions, we can prove that $\hat{\nu}_n^{bay}$ is convergent.

We assume that the following assumptions hold:

Assumption 4.1.

(1) *U is a compact metric space; the set $\mathcal{M}(U)$ of finite Borel measures on U is endowed with the topology of weak convergence,*
(2) *P is a probability measure on U having full support,*
(3) *The inverse problem has at least one solution,*
(4) *Φ is a continuous function on U with linear independents components*

In [13], F. Gamboa and E. Gassiat establish the following theorem, which proves weak consistency of the Bayes estimate and states the equivalence between MEM and the infinite-dimensional Bayesian estimator (defined as the limit of the finite-dimensional Bayesian estimators).

Theorem 4.2. *(Gamboa, Gassiat,* [13] *Theorem 2.3) Let U be a compact metric space, P a probability measure on U and c_i^n, $i = 1, ..., n$ such that $\frac{1}{n}\sum_{i=1}^n \delta_{c_i^n} \to P$. Under Assumptions 4.1, the Bayesian estimator $\hat{\nu}_n^{bay}$, for $\delta > 0$ and n sufficiently large, is well defined. It weakly converges to $\hat{\nu}_\infty^{MEM}$, the unique accumulation point of the sequence $\hat{\nu}_n^{MEM}$.*

Proof. The proof follows directly from [8] and using large deviation principle for ν_n. □

The links between the methodology we propose and the Bayesian measure estimator are obvious. Actually, taking the average of admissible slowness field is equivalent to consider the posterior mean of the distribution, conditionnaly to the rare event defined as admissible travel times. The problem has been linearized since, to compute the travel times from given slowness field, we used the shooting ray approximation defined in Section 2, which, implicitly linearizes the inverse problem around a good approximation for the slowness field. As a result, Theorem 4.2 can be applied and shows that the Bayesian estimator $\hat{S}_n^{bay}$, defined in (9), is convergent to the set of minimizers of the entropy of the distribution of S under the constraint $\|T^{obs} - T(S)\| \leq \delta$. If the field is homogeneous for each layer, then the minimum is unique, which implies the convergence of Bayes estimator towards the unknown slowness field.

If the layers are heterogeneous, hence we lose concavity of the constraint and the set of minimizers is non trivial. That is the reason we try to penalize the selection criterion in order to select the best approximation among all the different solutions.

We point out that this type of Bayes estimator is equivalent to selecting a quasi solution of the original inverse problem that maximizes a criterion. By conditioning by a rare event, i.e only considering the values of S such that $\|T(S) - T^{obs}\| \leq \delta$, we are able to let the corresponding empirical measure to concentrate around the true values. Hence the smaller the δ, the more accurate will be the estimate, as quoted in [15]. On the other hand, the drawback of this method is that if δ is too small, a very large number of simulations with the tracing ray $S_n^{(j)}$, $j = 1, \ldots, N$ is needed to find a sample of admissible travel times. Hence the convergence will be slower. Finding the optimal value is a difficult issue. In the simulations we chose $\delta = 0.01$ which empirically guaranteed a traliteraturede-off between rate of convergence and accuracy of the estimate. Moreover we have to take into account that we use travel times, so we need that the simulated waves reach the receivers, which is not the case for all simulation and so increases also the number of needed simulations.

4.2. **Regularized procedure.**

Heterogeneous layers imply the existence of local maxima, which prevent consistency of Bayes estimator. Hence we aim at choosing the best solution among the admissible solutions. For this, define the following criterion

$$\gamma(S) = \|T^{obs} - T(S)\|^2 + \lambda J(S)$$

where λ is a parameter and J is a suitable functional. In the , many penalty function are investigated depending on the goal of our study. We refer to [24], [28] or [19] for some study of penalized M-estimation. Our construction allows us to clarify the correspondence between Bayesian rules and regularization technique, and justifies the choice of the entropy as a penalty function.

Under the Gaussianity assumption of the observation errors,

$$T^{obs}|S \sim \mathcal{N}\left(T(S), I_k\right).$$

Hence if we are interesting in the maximum a posteriori, we aim at maximizing the logarithm of the posterior distribution with respect to the observations T^{obs}. Define $p(S)$ the prior density of the slowness field S, $\varphi(T^{obs}|S)$ the Gaussian density of the data knowing the travel times and $f(S|T^{obs})$ the posterior density of the slowness field. We obtain

$$\begin{aligned}
\sup_S \log f(S|T^{obs}) &= \sup_S \left(\log \varphi(T^{obs}|S) + \log p(S)\right) \\
&= \inf_S \left(\|T^{obs} - T(S)\|^2 - \log p(S)\right) \\
&= \inf_S \gamma(S),
\end{aligned}$$

if we set $-\log p(S) = \lambda J(S)$. Hence the penalty is the entropy of the prior distribution. Links between maximum entropy estimators and Bayesian estimators are also highlighted in this framework in [14]. In our case, we minimize this functional for all admissible slowness fields, defined as S such that $\|T^{obs} - T(S)\| \leq \delta$ for a fixed $\delta > 0$ and for $T(S)$ calculated by the ray shooting scheme described in Section 2. Hence the estimator is defined as

$$\widehat{S}_{Pen} = \inf_{\{S,\, \|T^{obs} - T(S)\| \leq \delta\}} \gamma(S). \tag{13}$$

There are several ways to conduct this minimization. In this work, the reconstruction is implemented by iterative minimization of the objective function. As with many complex model-based estimations, the objective is non concave, and different initial conditions lead to different reconstruction corresponding to different local maxima. So we use a stochastic algorithm "annealing" for the optimization. This algorithm is well studied in [17] or [16] for example. It belongs to the class of stochastic minimizing algorithms, depending on a parameter often called temperature which may vary in order to avoid local optima. Annealing is designed to seek approximate global maxima to the objective, and thus robustify the problem to initial conditions. Hence finding a good starting point is crucial to get convergence of the algorithm. Here the initial guess is given by the Bayesian estimator S_n^{bay} defined previously. Indeed, convergence results ensure that the Bayesian estimate belongs to the set of minimizers of the objective function and the stochastic algorithm enables us to avoid local minima and find the global minimizer among them. In the next section, we present simulations using this estimation procedure and evaluate its performance.

5. Numerical Results

The principal objective in this article is to estimate velocities and reflector depths in homogeneous and heterogeneous media. We propose a stochastic technique which combines a

Bayesian algorithm with simulated annealing. This combination allows to reconstruct a good model of speeds and depths for both homogeneous and heterogeneous means.

The purpose of this section is to present some numeric results obtained when implementing this combination. Three examples are presented in order to show the abilities of our technique. The first example is the seismic reflexion tomography problem in 2D homogeneous media. Here we show numerically that the solutions (obtained with the mere Bayesian method) converge to a global minimum. Then, in the following example, we give results obtained with the combination of the two methods (Bayesian algorithm and annealing simulated), in heterogeneous media. In these two first examples, we assume that we known the different layers in the medium. Indeed, the velocity does not change too much within one layer in the heterogeneous case, and only the vertical variation are of great importance. As a result, the uniform prior is a good prior in this case. In the last example, we apply our estimation technique to the very general case where both the velocity field and the structure of the medium (i.e the position of the different layers) are unknown. Hence, the problem is far more complicated and the estimation procedure provides the estimated slowness, starting with a non informative prior, a Gaussian prior in this case.

In the numeric results we will present the values of the speeds, since it is but easy to interpret and the slowness field is computed by simply taking the inverse of the velocity field. In this work, it is assumed that reflectors are flat and horizontal and that the layers are also horizontal. Considering different shapes for the layers is more realistic but far beyond the objectives of the estimation procedure. We refer to [4] for more references on this topic.

5.1. **Seismic Tomography in Homogeneous Media.**

We present the synthetic model designed especially to show the ability that has the Bayesian algorithm to solve problems of seismic reflection tomography in homogeneous media. Consider a media that measures 3000 meters of wide and 1000 meters of depth. The medium has two plane layers whose depths are known, namely $z_1 = 500$ and $z_2 = 1000$. In this case, we are only interested in reconstructing the velocity field. In the surface, there is one source, located at the origin, and an arrangement of six equally spaced receivers.

We observe the following travel times with known paths through one or two layers in Table 1.

Table 1. Synthetic observed travel times, T^{obs}

layers/receivers	1	2	3	4	5	6
1	0.5804	0.7336	0.9271	1.1616	1.3834	1.6331
2	0.9251	0.9987	1.1130	1.2454	1.4059	1.5690

In the special case of homogeneous medium, the problem of seismic inversion is linear (since the path of the rays are straight). So, it is not necessary to use an simulated annealing algorithm, since the mean of the solutions obtained with the Bayesian algorithm minimizes the quadratic norm of the travel times error.

Now, we discretize the medium into uniform cells, with a constant velocity in each cell. As the algorithm of Bayesian reconstruction requires of the generation of random speeds $V = 1/S$, then in this example we generate uniform speeds between $v_{\min}$ and $v_{\max}$, being respectively the known minimum speed and maximum of a layer. We will assume that the speeds corresponding to the first layer are between 1800 and 2100 $m.s^{-1}$, and in the second layer between 2500 and 2800 $m.s^{-1}$. Then different speeds are generated uniformly in each layer with the restriction that "to more depth, bigger speed". With such prior we generate rays reaching the receivers and compute the travel times.

In first step, in Figure 3, we present the surface of the quadratic-norm of the travel times error $\varphi(S) = \|T^{obs} - T(S)\|$, corresponding to the models considered in the partition. One can observe that there is a global minimum reached in the point (1919, 4; 2964, 4). It is important to remark that the surface it is not continuous. Indeed, the quadratic norm of the error is not defined for all the slowness models in the chosen range, because these slowness fields do not satisfy the physical properties imposed in the problem. The white bands of the surface of the errors quadratic norm correspond to the areas where the rays are not traced. Now, numerically, we can assure that, for media with homogeneous layers, there is a global minimum (as in Gamboa-Gassiat [13]). In a second step we compute the Bayesian estimator,

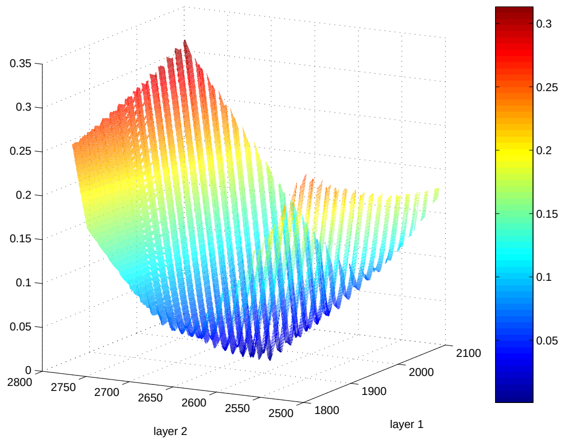

FIGURE 3. l^2 norm of the travel times error

S_n^{bay}, which converges to this minimum. Here we have chosen $\delta = 0.001$ and set $n = 100$. We have generated 400 rays and we observe that 12% of the times we find a solution. Using the average of the admissible distributions for the slowness field, restrained by the posterior condition $\|T^{obs} - T(S)\| \leq \delta$, we obtain the Bayesian estimator. In Table 2, we compare the true velocities, the speed values of the Bayesian estimate, the quadratic error of the velocities and the quadratic error of the travel times.

Table 2 : Results obtained with the Bayesian Method.

Range	true Velocities	Estimates	mean error	travel times error
1800-2100	1919.38032554	1919.93564754	0.00028932358	0.00080078142
2600-2800	2649.39057969	2643.94563385	0.00205516916	0.00547901016

In this table, we can see that the average velocity error is smaller, in the first layer, than 10^{-3} and smaller than 10^{-2} in the second layer. Such error is acceptable for industrial purpose since in both cases the error is smaller than 1%. We point out that the error is greater in the second layer than in the first. A reason for this result could be that fewer rays from the deeper layers are observed, resulting in a slower rate of convergence. Moreover, we can point out that the travel times error is of order $0.8\,m.s^{-1}$ in the first layer and $5.4\,m.s^{-1}$ in the second layer.

As a result, the Bayesian estimator gives very good result and is easy to compute once the tracing ray tool is implemented.

5.2. **Seismic Tomography in Heterogeneous Media with known layers.**

To illustrate some of ideas of the combination a Bayesian algorithm with simulated annealing we present a computational example of reflection tomography in heterogeneous media. The example consists of a simulation designed to demonstrate the effects of the velocities variations in each layer. The simulated region contains flat reflectors (unknowns) and lateral-vertical velocity changes in the three layers. In this case the rays are curves of stationary travel time. It is important to remark that the velocities (or slowness) are constant in each cell and thus the ray paths result straight in each cell.

We consider a medium of 3000 meters of length and 500 meters of depths with two layers. At the surface are located as in the previous example one source and six equally spaced receivers. The observed travel times are given in Table 3.

Table 3 : Observed travel times T^{obs}

layers/receivers	1	2	3	4	5	6
1	0.5953	0.7500	0.9502	1.1802	1.4303	1.6787
2	.8973	.9212	1.1202	1.5020	1.9231	2.1027

As previously, we take a uniform prior to generate admissible rays. The bounds are $(1800, 1850)$ for the first layer and $(1900, 1950)$ for the second layer. We display in Figure 4 the quadratic norm of travel times errors. As expected, there are numerous local minima. Hence the objective function is non convex for heterogeneous media. Hence, first we compute the Bayesian

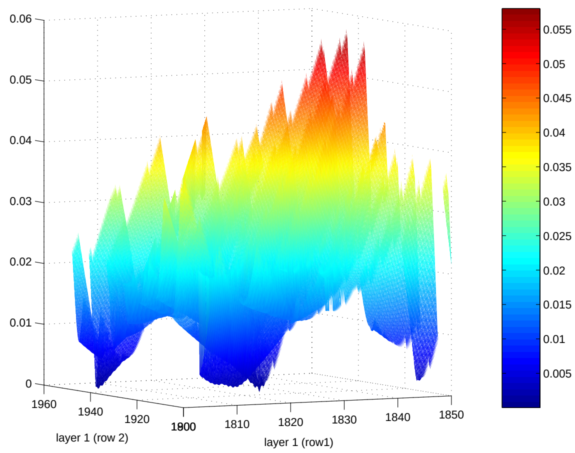

FIGURE 4. slowness mean error

estimator using as prior the uniform distribution, which enables to find admissible travel times, i.e satisfying the condition

$$\|T^{obs} - T(S)\| < 0.01. \tag{14}$$

In order to select among the sets of these realizations, we solve the penalized minimization program with annealing. We needed 400 iterations of the algorithm as starting point the Bayesian estimator. The convergence of the algorithm is achieved after 17.69 seconds. In Table 5 we show the quadratic error of the estimates and the quadratic error of the travel times.

We point out that the l^2-norm of the errors is of order 10^{-4}, so the algorithm computes close estimators of the true discretized velocities.

Table 5 : Velocity field error and travel times error in heterogeneous media

Rank	Quadratic error	Travel times errors
1800-1850	0.00100143	7.71192605e-004
1900-1950	0.00076514	7.71192605e-004

We also point out that the solutions are good approximations of the velocities but that the rays corresponding to these estimations are admissible. It shows that the estimates are good approximations in the sense that the estimated travel times are close to the true travel times.

5.3. **Heterogeneous media with unknown layers.**

Consider an heterogeneous medium with horizontal and vertical variations of velocities. This example is very complicated since we aim at finding estimators of the velocity field which provides admissible travel times and which are close to the true values of the unknown velocity field. In this case, we need to generate a larger amount of models for the rays and then choose among them the closest to the observations. A large amount of simulated slowness field with the corresponding travel times is needed.

The true medium is made of three layers corresponding to major difference between the velocities. Each layer is made of two sub-layers with small velocities variations. The velocities range from 1800 and 2100 $m.s^{-1}$ in the first layer, from 2600 and 2900 $m.s^{-1}$ for the second layer and from 3400 and 3700 $m.s^{-1}$ in the third layer. Within one layer there are small horizontal variations, ranging from -30 to 30. We will consider a discretization with 10 columns, so the parameter of interest is of size 60. We generate slowness field with Gaussian priors.

We present in Table 6 the results obtained with the two algorithms: the Bayesian procedure taken as a starting point of the annealing algorithm. Namely, we consider $n = 100$, the discretization level and run the algorithm with $k = 1000$ ray samples giving admissible travel times with the corresponding priors. Then we obtain the different values for the cells and compute the estimate in the 60 real cells by averaging the values in the cells of the discretization. So, we present the quadratic error of the velocities for each layer. We also estimate the position of the layer by finding the depth where there is a significative change in the slowness field, i.e when the horizontal average of the velocities of a line change. We get 3 major changes at position p to be compared with the true position of the layers.

Table 6 : Estimation of depth and velocities in an unknown heterogeneous media

layers/receivers	true layer position	estimated position	average l^2 error
1	500	508	0.0058
2	1000	1006.3	0.0062
3	1500	1512	0.0076

In the following figure 5, we show the image of the estimation error of the velocities for each true cell of the discretized medium. We can show that, for all cells, the error is less than 0.02. Nevertheless, there are large differences between the cells. It can be explained by the fact that there are not the same number of rays going through each cell, hence the estimator does not converge at the same rate for all cells and is not as accurate uniformly in the medium. Figure 5 shows the velocities error in a 2D heterogeneous media with three layers. Most of the

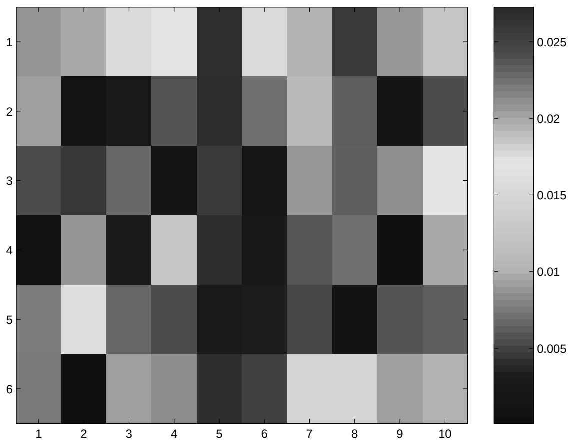

Figure 5. velocity mean error

velocity structure can be well determined. However, certain velocity features are determined relatively poorly.

We developed a robust algorithm able to solve difficult tomography problems accurately

The Bayesian algorithm provides solutions in homogeneous media, and the perceptual errors of the velocities are always smaller than 1%. In heterogeneous media, solutions obtained with the Bayesian algorithm are not always solutions, but the perceptual errors of the velocities are smaller than 2% in some of the cells over the mesh.

The Bayesian algorithm yields good initial estimates when the positions of the reflectors are not known.

The combination of both algorithms allows us to reconstruct a model (in heterogeneous media) that minimizes the l2-norm of the error plus a penalty term, incorporating the restrictions imposed by the physical properties of the pattern.

6. Conclusions

Seismic tomography is a difficult issue since it implies solving a severely ill-posed problem with an implicit unknown operator. Bayesian techniques enable to find an approximated solution that undergoes the physical conditions of the slowness field. For homogeneous medium, we prove the consistency of a one step Bayesian estimator, while for heteregoneous medium, we use this estimator as a starting point for an annealing algorithm and construct a two-step estimator. If consistency is not proved theoretically in this case, yet simulations are very efficient and the results we obtain are very encouraging. Hence, this Bayesian procedure gives rise to a fast algorithm and provides a more robust alternative to determinist algorithms.

Changing the prior may improve the results on a practical point of view. For instance, introducing dependency between cells in the prior distribution, according to seismic models, as cited in [23] or [21], could lead to faster convergence as shown in a forthcoming work.

Letting δ decreases to zero could also improve the asymptotic behaviour of the Bayesian estimator. However, when conditioning by very rare events, it becomes more difficult to simulate rays satisfying to the condition. As a result far more simulations are needed in able to get a consistent estimator.

Thanks

This Research was partially supported by Projects Agenda Petróleo-Venezuela, and Université Paul Sabatier Toulouse-France. The authors would like to thank Fabrice Gamboa for helpful discussions.

References

[1] Barron A. et al., (1999). "Risk Bounds for Model Selection via Penalization ". Probab. Theory and Related Fields. 113, pp. 467-493.

[2] J. G. Berryman, *Lectures on Nonlinear Inversion and Tomography*, University of California, 1991.

[3] J. M. Borwein and A. S. Lewis, *Partially-Finite Programming in L_1 and the Existence of Maximum Entropy Estimates*, Siam J. Optimization, Vol. 3, No. 2, pp. 248-267, May 1993.

[4] K. Bube, *Regularization Methods for Travel Time Tomography*, University of Whashington, Department of Mathematics, 2001.

[5] Bube K., *Tomographic Determination of Velocity and Depth in Laterally Varying Media*, Geophysics Vol. 50, No. 6, June 1985.

[6] L. Cavalier, G. K. Golubev, D. Picard, and A. B. Tsybakov. Oracle inequalities for inverse problems. *Ann. Statist.*, 30(3):843–874, 2002. Dedicated to the memory of Lucien Le Cam.

[7] J. F. Claerbout, *Imaging the Earth's Interior*, Blakwell Scientific Publications, 1985.

[8] D. Dacunha-Castelle and F. Gamboa. *Maximum d'entropie et problème des moments*. Annales de l'IHP Proba., 26:567-596, 1990.

[9] Csiszár, Imre and Gamboa, Fabrice and Gassiat, Elisabeth, *.MEM pixel correlated solutions for generalized moment and interpolation problems*, IEEE Trans. Inf. Theory, vol 45, number 7, 2253-2270,1999.

[10] Heinz W. Engl, Martin Hanke, and Andreas Neubauer. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.

[11] Heinz W. Engl. Regularization methods for solving inverse problems. In *ICIAM 99 (Edinburgh)*, pages 47–62. Oxford Univ. Press, Oxford, 2000.

[12] D. V. Fitterman, *Fundamentals of Seismic Tomography*, Geophysical Monograph Series, Number 6.

[13] F. Gamboa and E. Gassiat , *Bayesian Methods and Maximum Entropy for Ill-Posed Inverse Problems*, The Annals of Statistics, 1997, Vol. 25, N. 1, 328-350.

[14] Gamboa, F., *New Bayesian methods for ill posed problems.*, Stat. Decis. 17, number 4, 315-337, 1999.

[15] N. Gozlan. *Conditional principles for random weigted measures.* To appear in Esaim-PS, 2005.

[16] X. Guyon , *Random Fields on a Network*, Springer-Verlag, 1995.

[17] I.-T. Hsiao, A. Rangarajan and G. Gindi, *Bayesian Image Reconstruction for Transmission Tomography Using Deterministic annealing*, 2001.

[18] Barbara Kaltenbacher. Regularization by projection with a posteriori discretization level choice for linear and nonlinear ill-posed problems. *Inverse Probl.*, 16(5):1523–1539, 2000.

[19] Jean-Michel Loubes and Sara A. van de Geer. Adaptive estimation using thresholding type penalties. *Statistica Neerlandica*, 56:1–26, 2002.

[20] J-M. Loubes and C. Ludena. *Complexity penalty for nonlinear inverse problems*. Prepublications of University Paris Sud, 2004.

[21] F. Renard and P. Lailly. *How to take correlations of travel time information into account in reflection tomography*, preprint, 2004.

[22] Bernard A. Mair and Frits H. Ruymgaart. Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.*, 56(5):1424–1444, 1996.

[23] J.A. Scales and L. Tenorio, *Prior Information and Uncertainty in Inverse Problems*, 2000.

[24] B. W. Silverman. Some aspects of the spline smoothing approach to nonparametric regression curve fitting. *J. Roy. Statist. Soc. Ser. B*, 47(1):1–52, 1985. With discussion.

[25] R. Snieder and J. Trampert , *Inverse Problems in Geophisics*, Dept. of Geophisics, Utrecht University, The Netherlands, email snieder@geo.uu.nl.

[26] S. R. Taylor, A. A. Velasco and others, *Bayesian Tomography Applied to Seismic Event Identification Problems*, 23^{rd} Seismic Research Review, 2001.

[27] Tikhonov, A.,and Arsenin, V., (1977).“Solutions of Ill-Posed Problems ”. Wiley, New York.

[28] Sara A. van de Geer. *Applications of empirical process theory*. Cambridge University Press, Cambridge, 2000.

[29] Wahba, G., *Spline models for observational data*, Philadelphia SIAM, 1984.

[30] J.F. Yao, *On Constrained Simulation and Optimisation by Metropolis Chains*, 1999.

Escuela de Matemáticas, Facultad de Ciencias, UCV, Av. Los Ilustres, Los Chaguaramos, Código Postal 1020-A, Caracas - Venezuela. Telf.: (58)212-6051481.

Département de mathématiques, Equipe de probabilités et statistique, Université Montpellier II, 34000 Montpellier, France

Departamento de Matemáticas, IVIC, Carretera Panamericana KM. 11, Aptdo. 21827, Código Postal 1020-A, Caracas - Venezuela. Telf.: (58)212-5041412.

E-mail address: afermin@euler.ciens.ucv.ve
E-mail address: Jean-Michel.Loubes@math.univ-montp2.fr
E-mail address: cludena@ivic.ve