Recovery of the coefficients of the elastodynamics equation using two statistical estimators

Samih Zein[1], Nabil Nassif[2], Jocelyne Erhel[3], and Édouard Canot[4]

[1] IRISA-INIRA Rennes, France samih.zein@irisa.fr
[2] Amerian university of beirut, Lebanon nn12@aub.edu.lb
[3] IRISA-INIRA Rennes, France jocelyne.erhel@irisa.fr
[4] IRISA-INIRA Rennes, France edouard.canot@irisa.fr

Summary. In this paper, we are interested in the estimation of the mechanical parameters of a solid that are nonlinearly associated with the solution of a marine geosciences problem governed by a system of partial differential equations. Such estimation requires studying an "inverse wave propagation problem" consisting in the determination of the properties of solid elastic medium in contact with a fluid medium. The two-dimensional model being used is based on measuring the variation of the pressure in the fluid while propagating a seismic wave. Two stochastic methods, Markov Chain Monte Carlo (MCMC) with an accelerated version and Stochastic Perturbation Simultaneous Approximation (SPSA), are implemented and compared with respect to cost and accuracy.

Key words: Continuous Model, Inverse Problem, Bayesian Inference Model, MCMC, SPSA.

1 Introduction

Marine geosciences study genesis and dynamics of processes taking place at the ocean-subsoil interface and the neighbouring solid sublayers. Such studies provide a deeper knowledge of the impact of marine processes resulting from petroleum industries, on the environment and the natural resources. To improve our knowledge of these processes, it is necessary to have equipments capable of recognizing the geological properties of the subsoil marine. This is made by seismic campaigns that send punctual waves, then measure the reflection of these waves on each geological layer.

1.1 The Forward Problem: a Two-Dimensional Model and its Numerical Solution

Our study starts with a two-dimensional mathematical model that consists in finding (p, v_f, v_s and σ) that verify the following system of partial differential equations and

over an infinite domain constituted by a solid medium (Ω_s) and a fluid medium (Ω_f) and separated by an interface (Γ):

$$\begin{cases} \dfrac{\partial p}{\partial t} + c_f^2 \rho_f \ div\ v_f = 0\ (\Omega_f) & (1.1) \\ \rho_f \dfrac{\partial v_f}{\partial t} + \nabla p = 0\ (\Omega_f) & (1.2) \\ A\dfrac{\partial \sigma}{\partial t} - \epsilon(v_s) = 0\ (\Omega_s) & (1.3) \\ \rho_s \dfrac{\partial v_s}{\partial t} - \ div\ \sigma = 0\ (\Omega_s) & (1.4) \\ v_s.n = v_f.n\ (\Gamma) & (1.5) \\ \sigma.n = -p.n\ (\Gamma) & (1.6) \end{cases} \tag{1}$$

This system is constituted by two separate schemes coupled at the interface. Equations (1.1 , 1.2) are the acoustic wave equations in the fluid, equations (1.3 , 1.4) are the elastodynamics equations in the solid, and they are coupled with equations (1.5 , 1.6).
The unknowns are: p: the perssure field in the fluid medium, v_f: the velocity field in the fluid, σ: the stress tensor field in the solid and v_s: the velocity field in the solid. The physical parameters are: c_f: the wave propagation speed in the fluid, ρ_f: the fluid density, ρ_s: the solid density, and A: the inverse of the elasticity tensor which is function of λ and μ, the Lame's coefficients.

This model has been derived in [Diaz05], and the code which performs the computations was developed by team ONDES at INRIA Rocquencourt.

1.2 The Bayesian Model and the Inverse Problem

In our model of the inverse problem, we will be using a given set of data $y := \{y_{ij} = y(x_i, t_j)\}$ representing measures of the pressure at a set of points $\{x_i\}$ in the fluid and at instints t_j to recover the mechanical properties of the solid $\theta = (\rho, \lambda, \mu)$ which are the coefficients of equations (1.3 , 1.4) of the system (1).
Specifically, let $y = u + \epsilon$ where u is the numerical solution and $\epsilon := \{\epsilon_{ij}\}$ is a set of independent, identically distributed random variables, each following a Gaussian law: $N(0, s^2)$; s^2 is the fixed variance of the variables, taken to be a small percentage of the minimum of $\{u_{ij}\}$.
Estimation of θ requires $p(\theta|y)$ the probability distribution of θ given y. It can be deduced from Bayesian formula $p(\theta|y) = \frac{1}{p(y)} p(y|\theta)p(\theta)$ In what follows let $D_\theta = \prod_i [\theta_{i,min}, \theta_{i,max}]$ be the domain of acceptable values of θ. Use of (**??**) is coupled with the following assumptions:

- The prior probability density ($p(\theta)$):

$$p(\theta) \propto \begin{cases} 1 \text{ if } \theta \in D_\theta \\ 0 \text{ elsewhere} \end{cases} \tag{2}$$

- The likelihood probability density ($p(y|\theta)$) is:

$$p(y|\theta) \propto \exp\Big(\frac{-1}{2} \sum_i \Big(\frac{y_i - u(x_i, T, \theta)}{s}\Big)^2\Big) \tag{3}$$

The previous equations lead to the expression of the posterior probability density:

$$p(\theta|y) \propto \begin{cases} \exp\left(\frac{-1}{2}\sum_i \left(\frac{y_i - u(x_i,T,\theta)}{s}\right)^2\right) \text{ if } \theta \in D_\theta \\ 0 \qquad\qquad\qquad\qquad\qquad\qquad \text{elsewhere} \end{cases} \tag{4}$$

On these basis, we implement the following estimators:

1. The first seeks the expectation of θ given y:

$$E(\theta|y) = \int \theta \; p(\theta|y) \; d\theta$$

   This estimator is known to be optimal in the sense that it minimizes the prior probability weighted average of the mean square error. It will be computed using the MCMC method.
2. The second computes the maximum of the posterior probability $p(\theta|y)$:

$$\theta^* = \arg \max_{\theta \in D_\theta} \; p(\theta|y)$$

   where arg provides the (unique) value of θ that optimizes the given objective function ($p(\theta|y)$). This estimator is also known to be asymptotically unbiased and efficient. It will be computed using the SPSA method.

2 Markov Chain Monte Carlo (MCMC)

An approximation of $E[\theta|y]$ using a Monte Carlo mothod is done through draws of n independent, identically distributed (i.i.d.) samples of θ, say θ_k, $k = 1, \ldots, n$, following a uniform distribution. The average of $\theta_k p(\theta_k|y)$ is then computed, leading to the estimate: $E[\theta|y] = \frac{1}{n}\sum_{k=1}^{n} \theta_k p(\theta_k|y) + O(1/\sqrt{n})$.

In a simple Monte Carlo integration scheme, points are sampled uniformly, waisting considerable effort in sampling the tails of $p(\theta|y)$. Techniques for overcoming this problem act to increase the density of points in regions of interest and hence improve the overall efficiency. To compensate this draw-back Monte Carlo integration is combined with a Markov chain process that can produce a sequence of *dependent* samples having $p(\theta|y)$ as a limiting distribution (see [Robert96, Nicholls707SC]). Thus, most of the drawn samples fall in the region of interest and no computing samples from outside this region would be used and waisted. This is the essence of the Metropolis-Hasting algorithm (see [Robert96, Nicholls707SC]). In what follows, we introduce a slightly modified version of this algorithm that reduces significantly the cost of the computations.

2.1 An Accelerated Version of MCMC

The Metropolis-Hastings algorithm generates a sequence of samples from the probability distribution of one or more variables (see [Robert96, Nicholls707SC]). Such sequence is then used to compute the expected value integral.
The main idea for reducing the cost of computations is done through generating candidates that will likely be accepted. Thus we avoid computations for the candidates

that are likely to be refused by not evaluating $p(C|y)$. The acceleration consists in taking a proposal distribution $q^*(.|\theta_k)$ based on $p^*(.|y)$ a linear interpolation of $p(.|y)$ and a random walk proposal distribution $q(.|\theta_k)$. The actual proposal distribution is taken as:

$$q^*(C|\theta_k) = \alpha_{pred}(C, \theta_k)q(C|\theta_k) + (1 - r(\theta_k))\delta_{\theta_k}(C) \tag{5}$$

where $r(\theta) = \int \alpha(C, \theta)q(C|\theta)\ dC$ and $\delta_{\theta_k}(C)$ is the Dirac distribution not null at θ_k ("e.g". see [Robert96]). The construction of this interpolation needs knowing some points of $p(.|y)$. These points are obtained by running first the standard M-H algorithm. The modification of the standard algorithm is in the insertion of an intermediate step between the generation and the acceptance steps of the standard M-H:

1. At θ_k generate a proposal C from $q(\cdot|\theta_k)$.
2. With probability

$$\alpha_{pred}(C, \theta_k) = min\Big\{\frac{p^*(C|y)}{p^*(\theta_k|y)}, 1\Big\} \tag{6}$$

promote C to be a candidate to the standard M-H algorithm. Otherwise, pose $\theta_{k+1} = \theta_k$.
3. With probability

$$\alpha(C, \theta) = min\Big\{\frac{q^*(\theta_k|C)p(C|y)}{q^*(C|\theta_k)p(\theta_k|y)}, 1\Big\} \tag{7}$$

accept $\theta_{k+1} = C$; Otherwise reject C, $\theta_{k+1} = \theta_k$

2.2 Results with MCMC

With this method, only the case of a homogeneous solid medium is considered with only three parameters ($\lambda,\ \mu,\ \rho_s$) to be estimated. The wave source is placed at 50 m above the solid-fluid interface. A standard choice for the transmitted signal is the first order derivative of a Gaussian. Its frequency is 100 Hz and its amplitude is 1000 Pa. The main results are for the accelerated version of MCMC shown in table 1 where we have the exact values of the parameters used to simulate the inverse problem input data, a confidence interval for our parameter estimations and the relative error between exact and computed values. Validation of the results is done through four distinct tests.

Table 1. Results with 19000 samples of the Markov chain with the accelerated version (6000 evaluations of $p(C|y)$) and a pressure error $< 6\%$

θ	**Exact Value** (SI)	**Confidence Interval**	% **of error**
λ	11.5×10^9	$10.9\times10^9 \pm 2.6\%$	5.2%
μ	6×10^9	$6.5\times10^9 \pm 2\%$	8%
ρ	1850	$1867 \pm 0.15\%$	0.9%

1. *Convergence of our estimate to $E[\theta|y]$:* We want to know if the number of samples is sufficient to have a correct estimation of our parameters. The plot of the average $E[\theta|y] = \frac{1}{n}\sum_{k=1}^{n}\theta_k$ with respect to the first n samples of the chain is shown below (**Fig.** 1). One must see that this average becomes constant after a certain value of n (see [Robert96]) so it can be considered that the convergence is reached. One can notice in our case that this property is true for $n > 15000$ samples .
2. *Sampling from $p(\theta|y)$*: We need to check if the samples follow $p(\theta|y)$, the limiting distribution of the Markov Chain. One must expect the same evaluation for the parameters at the convergence with two different initial points (see [Robert96]). Two Markov chains of length 19000 samples are run and the results are compared in table 2. The two results are not the same but they are close enough to be able to consider that a length of 19000 samples for the Markov chain is sufficient and the samples are quite distributed according to $p(C|y)$.

Table 2. Parameter estimation with two different starting points and a pressure error $< 6\%$ (19000 samples)

θ	**Confidence Interval with θ_{min}**	**Confidence Interval with θ_{max}**
λ	$11.1\times10^9 \pm 2.8\%$	$10.8\times10^9 \pm 2.7\%$
μ	$5.82\times10^9 \pm 2.5\%$	$6.14\times10^9 \pm 2.6\%$
ρ	$1827 \pm 0.21\%$	$1911 \pm 0.4\%$

3. *Uniqueness of the solution:* the plot of the frequency of the states taken by the chain is shown in **Fig.** 2 from which one can verify that the posterior probability has one mode which means that the inverse problem has a unique solution.
4. *Accuracy of our estimation:* The estimate of the posterior mean of parameters θ is the average of n correlated samples from a Markov chain, its variance is calculated as if the samples are independent and multiplied by the integrated autocovariance time: $var(\overline{\theta}_{MC}) = \tau\frac{var(\theta)}{n}$ (see [Nicholls01]). One can estimate τ with this formula: $\overline{\tau} = 1 + 2\sum_{s=1}^{M}\rho(s)$ where $\overline{\rho}(s)$ is the autocovariance function given by $C(s) = cov(\theta_k, \theta_{k+s})$ and $\overline{\rho}(s) = C(s)/C(0)$ and which is shown in fig.4, and M is the smallest integer such that $\overline{\rho}(M) = 0$ (see [Nicholls01, Nicholls707SC]).

The accelerated MCMC version will also require a large number of simulations of the system (1): 6000 siulations for a chain of a 19000 samples length with the accelerated version instead of 19000 simulations for the standard MCMC. Thus, we have preferred another approach, the simultaneous perturbation stochastic approximation method which will require only hundreds of simulations.

3 Simultaneous Perturbation Stochastic Approximation (SPSA)

In the SPSA method, the estimator for θ is obtained by minimizing the loss function $L(\theta)$, where $L(\theta) = \log\ p(\theta|y)$.

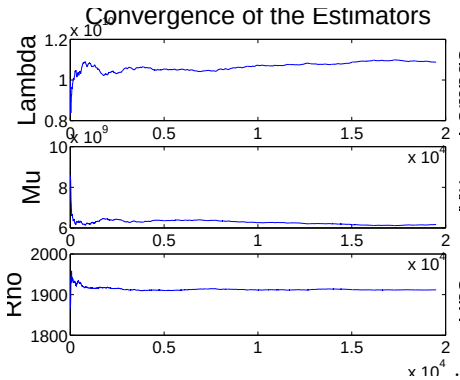

Fig. 1. Convergence of the estimators

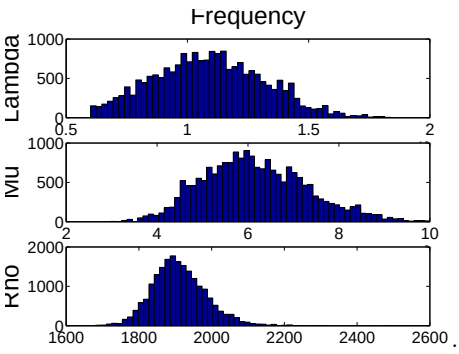

Fig. 2. Mode of the posterior distribution

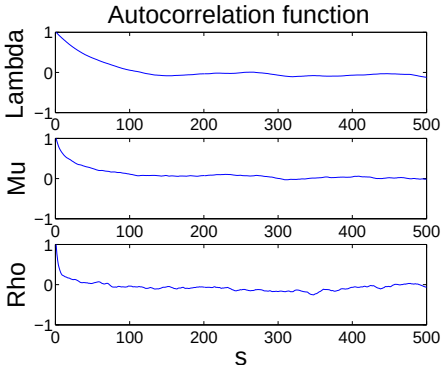

Fig. 3. Autocorrelation function of each parameter

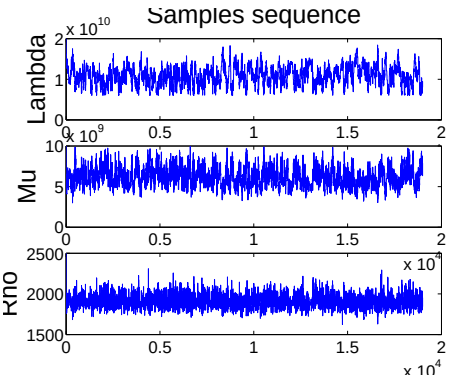

Fig. 4. Samples sequence generated by the Markov chain

However, given that the expression of the gradient of $L(.)$ is not available, we compute approximations for $\partial L/\partial\theta$ using measurements of $L(\theta)$ through simulations of the direct model (1).

3.1 Unconstrainted Optimization

The SPSA algorithm is of the form:

$$\hat{\theta}_{k+1} = \hat{\theta}_k - a_k \hat{g}_k(\hat{\theta}_k) \tag{8}$$

where $\hat{g}_k(\hat{\theta}_k)$ is an estimation of the true gradient $g(\theta) \equiv \partial L/\partial\theta$. In the SPSA method, one uses a special approximation for the gradient that requires only two evaluations of $L(\theta)$. All the variables are perturbed at the same time by drawing two random points centered at $\hat{\theta}_k$ to form the gradient approximation:

$$\hat{g}_k(\hat{\theta}_k) = \frac{L(\hat{\theta}_k + c_k\Delta_k) - L(\hat{\theta}_k - c_k\Delta_k)}{2c_k}[\Delta_{k1}^{-1}, \Delta_{k2}^{-1}, \ldots, \Delta_{kp}^{-1}]^T \tag{9}$$

where Δ_k is the perturbation random variable vector of mean zero, a_k and c_k are two positive sequences (see [Spall03]).

3.2 Implementation of SPSA

The SPSA algorithm is sumarized in the following steps (see [Spall03]):

1. **Initialization step**: With the SPSA, the sequences a_k and c_k are of the form $ak = a/(A+k)^\alpha$ and $ck = c/k^\gamma$. Set counter index k=1. Pick values for the non-negative coefficients a, c, A, α, and γ. The choice of the gain sequences (a_k and c_k) is critical to the performance of SPSA.
2. **Generation of the simultaneous perturbation vector**: Generate a p-dimensional random perturbation vector Δ_k, where each of the p components of Δ_k are independently generated from a Bernoulli distribution with probability of 1/2 for each outcome.
3. **Loss function evaluations**: Obtain two measurements of the loss function L(.) based on the simultaneous perturbation around the current θ_k: $L(\theta_k + c_k\Delta_k)$ and $L(\theta_k - c_k\Delta_k)$ with the c_k and Δ_k from Steps 1 and 2.
4. **Gradient approximation:** Generate the simultaneous perturbation approximation to the unknown gradient $g(\theta_k)$ from (9).
5. **Updating θ estimate:** Use (8), the standard SA form, to update θ_k to a new value θ_{k+1}.
6. **Iteration or termination:** Return to Step 2 with k+1 replacing k. Terminate the algorithm if there is little change in several successive iterates or the maximum allowable number of iterations has been reached.

3.3 Results with SPSA

The configuration for the solid and the seimic source is considered the same as above. The results are shown in table 3. This algorithm demands only 600 iterations while MCMC required 6000 simulations.

Table 3. Pressure measures with error $< 6\%$ (300 interations)

θ	**Exact Values** (SI)	**Confidence Intervals**	**Errors**
λ	11.5×10^9	$12.2\times10^9 \pm 6.12\%$	6.6%
μ	6×10^9	$5.4\times10^9 \pm 7.2\%$	9.5%
ρ	1850	$1868 \pm 0.83\%$	1%

4 Concluding Remarks

In this paper we describe two methods to determine the density and the elasticity of a homogeneous solid medium from the variation of the pressure in the fluid due to the transmission of a seismic wave.

Two statitical methods are considered. MCMC gives an estimation with an acceptable error but its very costly in computations. SPSA, gives approximately the same accuracy and has the advantage to be much less expensive in computations. This method will be considered in the case of a solid with multiple layers.

References

[Diaz05] J. Diaz and P. Joly: Robust high order non conforming finite element formulation for time domain fluid-structure interaction. Journal of computational acoustics, vol. 13 (2005).

[Fox05] J. Andre's Christen and C. Fox: MCMC using approximation, Journal of Computational and Graphical statistics (2005).

[Nicholls707SC] S.M. Tan, C. Fox and G. Nicholls: Physics 707SC course materials at Auckland university

[Niederreiter92] Harold Niederreiter: Random number generation and quasi-Monte Carlo methods. SIAM (1992).

[Nicholls01] S. Meyer, N. Christensen and G. Nicholls: Fast Bayesian reconstruction of chaotic dynamical systems via extended Kalman filtering. Physical Review, 2001

[Robert96] C. Robert and G. Casella: Monte carlo statistical methods. Springer (1996)

[Spall03] J.C. Spall: Introduction to Stochastic Search and Optimization: Estimation, Simulation, and Control, Wiley, Hoboken, NJ (2003)