Some Examples of Inverse Problems in Geophysics

Marc Lavielle
UFR de Mathématiques et Informatique, Université Paris V
45 rue des Saints-Pères, 75006 Paris, France
e-mail: lavielle@math-info.univ-paris5.fr

David Marquez
Université de Marne-la-Vallée, France
e-mail: marquez@math.univ-mlv.fr

ABSTRACT

Most of the problems arising in geophysics consist in recovering some characteristics of the earth (reflectivity, resistivity, ...), from some measurements usually made at the top of the earth (seismic data, electromagnetic data, ...). The observations can be seen as the output data of a noisy (linear or non linear) system, that can be known, or unknown. When the system is unknown, it can be estimated, together with the noise variance, by using a stochastic version of EM algorithm.

When the system is identified, some adaptations of MCMC techniques allow one to estimate the a posteriori distribution of the input sequence, according to the prior distribution. This prior can take into account the regularity of the input sequence, as well as the presence of abrupt changes in this sequence.

1 Introduction

In many physical systems, two kinds of data are present: a vector $\boldsymbol{y} = (y_i,\ 1 \leq i \leq m)$ of observed data, and a vector $\boldsymbol{z} = (z_i,\ 1 \leq i \leq n)$ of unobserved data. We consider here that this physical system can be modeled by using a parametric model of the form

$$\boldsymbol{y} = H(\boldsymbol{z};\theta) + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{\varepsilon}$ is an additive random noise, and where θ is a parameter that takes its values in a given subset Θ of $\mathbb{R}^d$.

The identification of the model consists in estimating the parameter θ, while the recovery of the unknown input signal consists in estimating the sequence $\boldsymbol{z}$.

The EM algorithm is an iterative and deterministic method for computing the maximum likelihood estimate of θ. The EM algorithm maximizes the observed likelihood $g(\boldsymbol{y};\theta)$ by iteratively maximizing conditional expectations of the complete log-likelihood $l(\boldsymbol{y},\boldsymbol{z};\theta)$. Then, an appealing property of EM is its numerical stability: each iteration increases the complete data likelihood and the convergence is almost always to a local maximum. Unfortunately, the conditional expectation of $l(\boldsymbol{y},\boldsymbol{z};\theta)$ cannot be computed in closed-form in most of interesting models.

Then, SAEM (for Stochastic Approximation of EM) algorithm proposed by Lavielle *et al.* [10] uses a stochastic approximation procedure in order to estimate the conditional expectation of the complete log-likelihood. The k'th step of SAEM may thus be summarized as follows:

- *Simulation* : generate a realisation $\boldsymbol{z}_k$ of the missing data vector under $k(\boldsymbol{z}|\boldsymbol{y};\theta_k)$, the a posteriori density of the complete data given the observations and the current fit θ_k of the parameter vector.
- *Stochastic approximation* : compute the current approximation $Q_k(\theta)$ of the conditional expectation of the log-likelihood $l(\boldsymbol{y},\boldsymbol{z};\theta)$ according to

$$Q_k(\theta) = Q_{k-1}(\theta) + \gamma_k(l(\boldsymbol{y},\boldsymbol{z}_k;\theta) - Q_{k-1}(\theta)) \tag{2}$$

  where (γ_k) is a sequence of positive stepsizes.
- *Maximization*: maximize $Q_k(\theta)$ in the feasible set Θ, *i.e.* find $\theta_{k+1} \in \Theta$ such that:

$$Q_k(\theta_{k+1}) \geq Q_k(\theta) \quad \forall \theta \in \Theta.$$

The stepsize allows the tuning of the stochastic excitation fed into the algorithm, avoiding the convergence towards spurious stationary points (e.g. saddle points or local minima). Indeed, under some general hypothesis, SAEM algorithm converges towards a local maximum of the observed likelihood.

When the parameter θ has been estimated, different estimators of the input sequence $\boldsymbol{z}$ can be proposed. The MAP (Maximum a Posteriori) maximizes the joint *a posteriori* distribution $k(\boldsymbol{z}|\boldsymbol{y};\theta)$. The MPM (Maximum Posterior Mode) maximizes the marginal distributions $k(z_i|\boldsymbol{y};\theta)$, for $1 \leq i \leq n$. Another Bayesian estimate of z_i is the mean of $k(z_i|\boldsymbol{y};\theta)$, instead of its mode. When the joint distribution $k(\boldsymbol{z}|\boldsymbol{y};\theta)$, or its local characteristics $k(z_i|\boldsymbol{y};\theta)$, cannot be computed in closed-form, they are estimated by using a MCMC method.

2 Seismic deconvolution

In seismic deconvolution, the observed data is a set of seismic traces and the non observed data series $\boldsymbol{z}$ is the sequence of reflection coefficients of the earth. In a first approximation, a seismic trace is the convolution of the reflectivity sequence with a *wavelet* $\boldsymbol{a}$ sent from the top of the earth. A random measurement noise $\boldsymbol{\varepsilon}$ is also present:

$$y_t = \sum_{l=0}^{L} a(l) z_{t-l} + \sigma \varepsilon_t . \tag{3}$$

We assume that the input reflectivity sequence is an independent and identically distributed sequence of random variables distributed according to a mixture of two zero-mean Gaussian distributions with different variances:

$$z_i \sim \lambda \mathcal{N}(0, \sigma_1^2) + (1-\lambda)\mathcal{N}(0, \sigma_2^2) \tag{4}$$

This model is standard in deconvolution of seismic traces (see [8, 11]). Indeed, the main reflectors, which indicate separation between the layers have a variance bigger than the secondary reflectors, which indicate small variations inside layers.

Assuming that $\boldsymbol{\varepsilon}$ is a Gaussian white noise with unit variance, the parameters to be estimated are the filter $\boldsymbol{a} = (a(0), \ldots, a(L))^t$, the variance of the additive noise σ^2, and the prior distribution π of $\boldsymbol{z}$, that is $(\lambda, \sigma_1^2, \sigma_2^2)$.

Here, some maxima of the likelihood are made up of filters having different phases and SAEM can converge towards a local maximum, that is, a filter with a wrong phase. A simulated annealing procedure improve convergence towards a global maximum of g. Then, we can hope to recover the correct phase if the number of data is large enough.

A synthetic example is shown Figure 1. In this example, the prior distribution π is assumed to be known. The original filter and the initialization are displayed in Figure 1–a (the initialization is a spike at $t = 5$ while the filter is symmetric around $t = 10$). With this initial guess, SAEM converges to a wrong solution, that is a local maximum of the likelihood (Figure 1–b). On the other hand, the simulated annealing version of SAEM allows a correct recovery of the system phase (Figure 1–c). It is interesting to remark that these two estimated filters essentially differ from their phase (Figure 1–d). See [9] for a complete description of this procedure.

The *Simulation step* of SAEM is performed by using a MCMC algorithm. Indeed, several sampling schemes can be used for generating the input signals with the a posteriori distribution, when the prior distribution is a Gaussian mixture [1, 2, 5, 6].

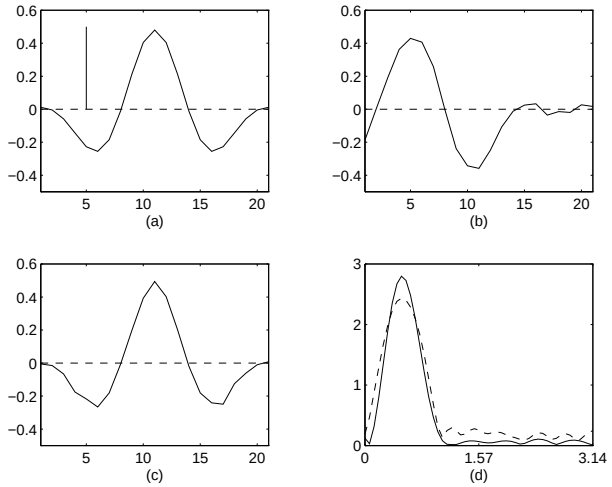

Figure 1: *(a) The coefficients* $(a(l))$ *of the original filter of length* $L = 20$ *and the initialization (a spike at* $t = 5$*), (b) estimation obtained with SAEM, (c) estimation obtained with the simulated annealing version of SAEM, (d) comparison of the modulus of the transfer functions of the filters displayed in (b) (dashed line) and (c) (solid line).*

3 Electromagnetic imaging

The electromagnetic inverse problem consists in estimating the distribution of rock resistivity from electromagnetic measurements on the surface of the earth. We consider the example of magnetotelluric soundings limited to the 1-D approximation. In this case, the electrical resistivity will only vary with the depth, therefore, the 1-D model could be described as a digitized medium of n homogeneous thin layers whose thickness increases with depth [7]. The missing data is the vector $\boldsymbol{z} = (z_i, 1 \leq i \leq n)$ of resistivity of the layers. The observed magnetotelluric data set is a vector $\boldsymbol{y} = (y_j,\ 1 \leq j \leq m)$, corresponding to different frequencies $(f_1, \ldots, f_m)$:

$$y_j = H(\boldsymbol{z}; \lambda_j)(1 + \sigma \varepsilon_j). \tag{5}$$

In this example of application, the *forward problem* is solved, *i.e.*, the function H in (5) is known. Furthermore, the variance of y_j depends on the frequency λ_j.

In the framework of the 1-D electromagnetic inverse problem, the known physical properties of geological environment and results from other earth sciences enable us to deduce the maximum and minimum values of possible resistivities on each layer. Then, a discretization of this set of values reduces to L the number of different possible values for the resistivity. In other words, z_i takes its value in a known finite set $E = \{\rho_1, \ldots, \rho_L\}$.

3.1 The estimation of the noise variance

We consider first that $\boldsymbol{z}$ is a sequence of i.i.d. random variables, uniformly distributed on E: $P(z_i = \rho_\ell) = 1/L$ for any $1 \leq \ell \leq L$ and any $1 \leq i \leq n$.

Since H is known, the identification of the model reduces to the estimation of σ^2. In this example, it is easy to see that the *Stochastic Approximation* step (2) and the *Maximization* step of SAEM reduce to the following recursion:

$$\sigma^2_{k+1} = \sigma^2_k + \gamma_k (S^2_k - \sigma^2_k) \tag{6}$$

where S^2_k is the empirical estimate of σ^2 computed with $\boldsymbol{y}$ and with the simulated sequence $\boldsymbol{z}_k$:

$$S^2_k = \frac{1}{m} \sum_{j=1}^{m} \left| \frac{y_j - H(\boldsymbol{z}_k; \lambda_j)}{H(\boldsymbol{z}_k; \lambda_j)} \right|^2 . \tag{7}$$

Thus, if we set, for example, $\gamma_k = 1/k$, then σ^2_{k+1} is the empirical mean of the sequence $(S^2_1, S^2_2, \ldots, S^2_k)$.

On the other hand, the SEM algorithm proposed by Celeux and Diebolt [4], is obtained with $\gamma_k = 1$. In this case, $\sigma^2_{k+1} = S^2_k$.

We display in Figure 2 a comparison of the SEM algorithm with the SAEM procedure, with $\gamma_k = 1/k$. We clearly see with this example that the decreasing sequence of stepsizes (γ_k) allows to reduce the randomness of the estimation, and to converge almost surely to the maximum likelihood estimate of σ^2 (see the detail of the two trajectories shown in Figure 2-b).

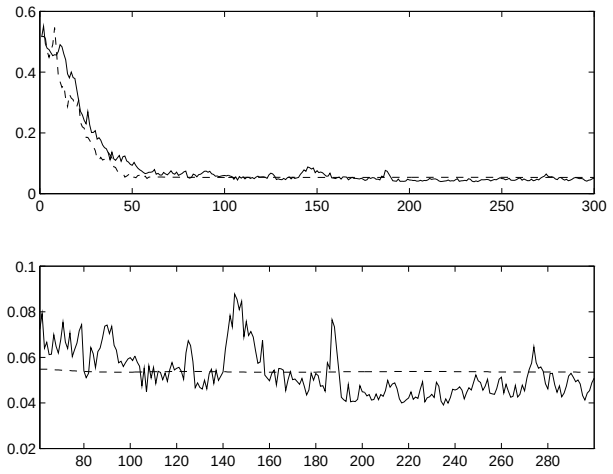

Figure 2: *Comparison of SEM and SAEM.* —— *: the trajectory* (σ^2_k) *obtained with SEM* $(\gamma_k = 1)$*,* −−− *: the trajectory* (σ^2_k) *obtained with SAEM* $(\gamma_k = 1/k)$*.*

3.2 The estimation of the resistivity sequence

We estimate in this section the *a posteriori* distribution of the reflectivity sequence. We shall not consider here that the reflectivity sequence is i.i.d. Indeed, the reflectivity is known to be a smooth function, with abrupt changes at some unknown instants. Then, we introduce a new vector $\boldsymbol{q} = (q_i, 1 \leq i \leq n)$, with $q_i = 1$ if there is a change at i, and $q_i = 0$ elsewhere. On the other hand, the prior distribution π must favors minimal variation between the resistivity values of consecutive layers beetween two abrupt changes. The prior distribution π of $(\boldsymbol{z}, \boldsymbol{q})$ is then defined as follows:

i) The sequence $\boldsymbol{q}$ is i.i.d. with

$$P(q_i = 1) = 1 - P(q_i = 0) = p \ .$$

Here, p is the prior probability to have a change at any instant i.

ii) Assume that $q_{i_1 - 1} = q_{i_2} = 1$ and $q_i = 0$ for any $i_1 \leq i < i_2$. Then,

$$\pi(z_{i_1}, \ldots, z_{i_2}) = C_\lambda \nu(z_{i_1}) \prod_{i=i_1+1}^{i_2} \exp\left\{ -\lambda \left(\log \frac{z_i}{z_{i-1}} \right)^2 \right\} .$$

The constant C_λ is a normalizing factor, the parameter λ controls the smoothness of the reflectivity function, and the initial law ν is choosen such that the marginal distributions function of the z_i's are identical.

According to this prior distribution π, we can define the *a posteriori* distributions of q_i and z_i:

$P(q_i = 1 | \boldsymbol{y})$ is the probability to have a change at depth i, conditionnally to the observations $\boldsymbol{y}$.

$P(z_i = \rho_\ell | \boldsymbol{y})$ is the probability that the resistivity at depth i is equal to ρ_ℓ, conditionnally to the observations $\boldsymbol{y}$.

These two a posteriori distributions can be estimated for all $1 \leq i \leq n$, by using the Gibbs Sampler, together with a Rao-Blackwellisation scheme, in order to reduce the variance of estimation [3].

A numerical experiment is proposed in Figure 3. In this example, $\sigma^2 = 0.05$. We consider the problem of the estimation of $(\boldsymbol{q}, \boldsymbol{z})$, obtained with the smoothing parameter $\lambda = 20$ and with $p = 0.01$. The true resistivity sequence is displayed in Figure 2-a together with the estimation of $(P(z_i = \rho_\ell | \boldsymbol{y}),\ 1 \leq i \leq n,\ 1 \leq \ell \leq L)$.

We display also in Figure 2-a the estimation of the posterior means $(E(z_i | \boldsymbol{y}),\ 1 \leq i \leq n)$. This sequence gives a very accurate estimation of the true reflectivity sequence.

The estimation of $(P(q_i = 1|\boldsymbol{y}),\ 1 \leq i \leq n)$ is displayed in Figure 2-b. We can see that the mode of the conditional distribution $P(z_i = \rho_\ell|\boldsymbol{y})$ provides a very good estimate of the resistivity z_i. On the other hand, the original changes at depth 1.5km, 1km, 3km and 3.5km are well detected.

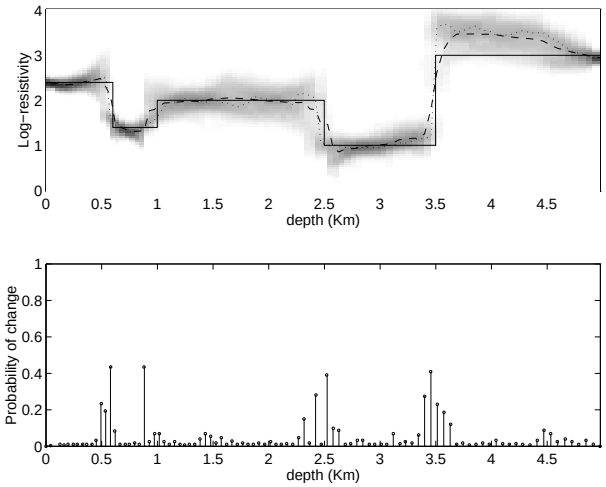

Figure 3: *The a posteriori distribution of (q_i, z_i), estimated by using the Gibbs Sampler procedure. We display the matrix $(P(z_i = \rho_\ell|y),\ 1 \leq i \leq n,\ 1 \leq \ell \leq L)$ in Figure 3-a ; each value of this matrix is represented with a gray level: a dark cell correspond to a high probability, while a white cell corresponds to a low probability. The true resistivity sequence is in solid line and the posterior mean is in dotted line. The vector $(P(q_i = 1|y),\ 1 \leq i \leq n)$ is displayed in Figure 3-b.*

References

[1] O. Cappé, A. Doucet, M. Lavielle, and E. Moulines. Methods for blind maximum-likelihood linear system identification. *(submitted)*, 1997.

[2] C.K. Carter and R. Kohn. Markov chain Monte Carlo in conditionnally Gaussian state space models. *Biometrika*, 83(3):589–601, 1996.

[3] G. Casella and C.P. Robert. Rao-blackwellisation of sampling schemes. *Biometrika*, 83:81–94, 1996.

[4] G. Celeux and J. Diebolt. A stochastic approximation type EM algorithm for the mixture problem. *Stoch. and Stoch. Reports*, 41:119–134, 1992.

[5] P. DeJong and N. Shepard. The simulation smoother for time series models. *Biometrika*, 82(2):339–350, 1995.

[6] A. Doucet and P. Duvaut. Bayesian estimation of state space models applied to deconvolution of Bernoulli-Gaussian processes. *Signal Processing*, 1997.

[7] H. Gandis, M. Minvielle, and M. Roussignol. Bayesian inversion with markov chains: the magnetotelluric 1-D case. *(submitted)*, 1997.

[8] M. Lavielle. 2-D Bayesian deconvolution. *Geophysics*, 56(12):2008–2018, 1991.

[9] M. Lavielle and E. Moulines. A simulated annealing version of the EM algorithm for non-Gaussian deconvolution. *Statistics and Computation*, 7(4):229–236, 1997.

[10] M. Lavielle, E. Moulines, and B. Delyon. Convergence of a stochastic approximation version of the EM algorithm. *The Annals of Stat. (accepted for publication)*, 1997.

[11] J.F. Mendel. *Optimal seismic deconvolution.* Academic Press Inc., 1983.