EM PARAMETER ESTIMATION FOR A PIECEWISE AR

Marc Fayolle and Jérôme Idier

Laboratoire des Signaux et Systèmes (CNRS–ESE–UPS)
Plateau de Moulon, 91192 Gif-sur-Yvette Cedex, France
fayolle@lss.supelec.fr, idier@lss.supelec.fr

ABSTRACT

We design a model meant to be the equivalent of Blake's weak string but in the probabilistic framework. Independent line sites delimit piecewise stationary Gaussian autoregressives AR(1) corrupted with Gaussian white noise. Thanks to the Bayesian interpretation, we define the joint probability which in turn yields the likelihood. We demonstrate how to make its computation possible in cubic time. This calculation allows the set of parameters to be tested but not estimated due to the complex form of the criterion. Yet the computations done so far provide the materials for an iterative maximization. Indeed, the Expectation Maximization algorithm happens to match the features of this model and is also easily calculable. When the likelihood is known, the cost of one step of the latter algorithm is negligible in comparison with the previous calculations.

1. INTRODUCTION

Unidimensional piecewise continuous data are often represented by models including boundaries between homogeneous data called "breakpoints". Blake [1] devised a dynamic programming algorithm which computes the exact global minimum of the energy of one of these models known as the weak string. In this respect, it allows optimal supervised segmentation, but the adequate tuning of parameters is left to empirical supervision.

From the Bayesian viewpoint, minimizing the energy is equivalent to finding the "maximum a posteriori" (MAP) estimate. On the other hand, the Bayesian interpretation gives rise to new possibilities of probabilistic handling of the weak string. One of the most fruitful is the definition of the likelihood of the parameters, which has no equivalent in the deterministic framework. In the following, our goal is to estimate the parameters by a tractable maximum likelihood method. Computation of the true maximum likelihood estimator is impossible due to the complex form of the criterion. This leads us to an iterative method with increasing likelihood called Expectation Maximization (EM) [2] [3].

The paper is organized as follow. In Section 2, we begin with some definitions and notations to formulate our model. Then, in the third section, we address the computation of the likelihood. We give an optimized procedure whose intermediate calculations will be useful later. The EM procedure can be found in Section 4. It uses the by-products of the preceding section. In the fifth section, the examples used as benchmarks are found in Blake [1], they intend to show how the unsupervised segmentation works.

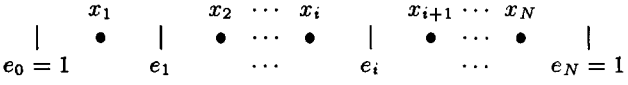
This work was supported by the Direction de la Recherche et des Études Techniques (DRET).

2. PROBABILISTIC FRAMEWORK

2.1. Definition of the model

Consider an unobserved real valued random vector, $\boldsymbol{X}$ which is defined on a set of *pixel sites* indexed by integers $1, \ldots, N$. The observed part of the model is a sequence $\boldsymbol{Y}$ following the simple observation model: $Y_i = X_i + \varepsilon_i$, where $\varepsilon = (\varepsilon_i)$ is a zero-mean Gaussian white sequence of constant variance σ_b^2. Vectors $\boldsymbol{X}$ and ε are assumed independent.

Between pairs of adjacent pixel sites, we introduce a dual sequence of sites constituting a vector $\boldsymbol{E}$ of *breakpoints*, *i.e.*, 0-1 binary variables When $E_i = 1$, the breakpoint is said to be active, which corresponds to a rupture in the vector $\boldsymbol{X}$ between pixel sites i and $i+1$. Finally it is useful to introduce additional breakpoints E_0 and E_N, which are assumed to be active.



Now we define the *energy* of the underlying process $(\boldsymbol{X}, \boldsymbol{E})$

$$E = \alpha \textstyle\sum_{i=0}^{N} e_i + \frac{1}{2\sigma_\epsilon^2} \sum_{i=1}^{N-1} (\beta x_i - x_{i+1})^2 \bar{e}_i + \frac{1}{2\sigma_x^2} \sum_{i=0}^{N-1} x_{i+1}^2 e_i,$$

where $\bar{e} \triangleq 1 - e$ and the parameters $\alpha, \beta, \sigma_\epsilon, \sigma_x$ are real and positive. In accordance with the Gibbsian meaning of energy, it is equivalent to the definition of a Bayesian prior likelihood given by

$$p(\boldsymbol{x} \,|\, \boldsymbol{e}) P(\boldsymbol{e}) \propto \exp(-E). \tag{1}$$

In order to get the exact energetic formulation of the weak string, β and σ_x should tend to 1 and ∞ respectively. Instead we impose a stationarity condition: $(1-\beta^2)\sigma_x^2 = \sigma_\epsilon^2$. According to the observation model, the energy of the whole set $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{e})$ is:

$$\begin{aligned} F &= E + \frac{1}{2\sigma_b^2} \textstyle\sum_{i=1}^{N} (y_i - x_i)^2 \\ &= \alpha \mathbf{1}^{\mathrm{t}} \boldsymbol{e} + \boldsymbol{x}^{\mathrm{t}} \left(\frac{1}{2\sigma_\epsilon^2} \boldsymbol{P}^{\mathrm{t}} \boldsymbol{Q} \boldsymbol{P} + \frac{1}{2\sigma_x^2} \boldsymbol{H} \right) \boldsymbol{x} + \frac{\|\boldsymbol{x} - \boldsymbol{y}\|^2}{2\sigma_b^2} \end{aligned} \tag{2}$$

In order to obtain such a compact form for F, let $\mathbf{1} = (1,\dots,1)^{\mathrm{t}}$ and matrices $\boldsymbol{P} = \boldsymbol{P}(N), \boldsymbol{Q} = \boldsymbol{Q}(\boldsymbol{e}), \boldsymbol{H} = \boldsymbol{H}(\boldsymbol{e})$, of size $(N-1)\times N$, $(N-1)\times(N-1)$, $N\times N$, according to

$$\boldsymbol{P}(N) \triangleq \begin{pmatrix} \beta & -1 & & (0) \\ & \beta & -1 & \\ & & \ddots & \ddots \\ (0) & & \beta & -1 \end{pmatrix},$$

$\boldsymbol{Q}(\boldsymbol{e}) \triangleq \mathrm{Diag}\,\{\bar{e}_1,\dots,\bar{e}_{N-1}\}$, $\boldsymbol{H}(\boldsymbol{e}) \triangleq \mathrm{Diag}\,\{e_0,\dots,e_{N-1}\}$.

2.2. Block partitioning

From (1,2), it is easy to identify the inverse of the prior covariance matrix of $\boldsymbol{X}$ knowing $\boldsymbol{E} = \boldsymbol{e}$ as:

$$(\boldsymbol{\Gamma}^{x|e})^{-1} = \sigma_\epsilon^{-2}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{Q}\boldsymbol{P} + \sigma_x^{-2}\boldsymbol{H},$$

which is clearly block-diagonal, with block limits corresponding to the active breakpoints. More generally, the active breakpoints induce a partition of pixel sites into *blocks*, according to the first of the following definitions.

Definition 1 *For any sequence (z_k), let us denote $z_{i,j} = (z_i,\dots,z_j)^{\mathrm{t}}$. Given a set of breakpoints $\boldsymbol{e}$, a* block *is an interval $[i,j]$ of sites such that $e_{i,j} = (1,0,\dots,0,1)$. If $[i,j]$ is the kth block, it will be convenient to denote $e_{i,j} = e_k$ alternatively. Such a notion of block will be used in contrast with the basic term* interval, *where the values of breakpoints are not specified. Let us introduce the following compact notations: $\pi(z_{i,j}) \triangleq p\big([i,j]$ is a block$\big)$, and for any random sequence $(Z_k), k = 1,\dots,N$,*

$$\pi(\boldsymbol{z}_{i,j}) \triangleq p\big(\boldsymbol{z}_{i,j} \,|\, [i,j] \text{ is a block}\big)$$

stand for the conditional probability density of $\boldsymbol{Z}_{i,j}$ w.r.t. "$[i,j]$ is a block", whereas $p(\boldsymbol{z}_{i,j})$ is the probability density of $\boldsymbol{Z}_{i,j}$. Note that both notations are implicit as regards the random vector itself.

2.3. The line process

We shall make use of another equivalent representation of the line process. Instead of giving the value of the vector of breakpoints $\boldsymbol{E}$, we express it with a sequence (L_k). L_k is the location of the kth active breakpoint, $k \geq 1$:

$$L_k = \inf_i \left\{ i \in [1,N], \textstyle\sum_{j=1}^{i} e_j = k \right\} \qquad (3)$$

Since $\boldsymbol{E}$ follows a Bernoulli law, L_k is ruled by an exponential law.

3. COMPUTATION OF THE LIKELIHOOD

3.1. Summation upon $\boldsymbol{e}$

Our purpose was to define and calculate the joint probability, $p(\boldsymbol{x},\boldsymbol{y}\,|\,\boldsymbol{e})P(\boldsymbol{e})$, according to the Bayesian method. It is possible to deduce the marginal law of $\boldsymbol{y}$ from a summation over all the realizations of $\boldsymbol{e}$. As the line process is a Bernoulli process, it can be any of the 2^{N-1} realizations which means that 2^{N-1} summations will have to be done. But when we use the equivalent model based on the arrival time we manage to do it with a polynomial complexity at a much lower computational cost. Let $\boldsymbol{y}$ be an interval that may not necessarily consist of the entire signal. To compute its likelihood, consider the first breakpoint L say. Then we can write:

$$p(\boldsymbol{y}) = \pi_{0,N}\pi(\boldsymbol{y}_{0,N}) + \textstyle\sum_{L=1}^{N-1}\pi(\boldsymbol{y}_{0,L})p(\boldsymbol{y}_{L,N})\pi_{0,L}, \qquad (4)$$

From the above formula, we can find a recursion giving the likelihood of an interval when you know the same for all its sub-intervals and, the laws $\pi(\boldsymbol{y}_{0,L})$ of all the blocks starting at the same point.

The formulae subsequently given are classical to the Gaussian case. Till the end of this part, the calculations are limited to a single block. Therefore, we shall not explicitly refer to its boundary or its dependence to the line process:

$$\boldsymbol{\Gamma}_k^y = \sigma_b^2\boldsymbol{I}_k + \boldsymbol{\Gamma}_k^x, \quad (\boldsymbol{\Gamma}_k^{x|y})^{-1} = (\boldsymbol{\Gamma}_k^x)^{-1} + \sigma_b^{-2}\boldsymbol{I}_k.$$

$$p(\boldsymbol{y}\,|\,\boldsymbol{e}) = \exp -\frac{1}{2}\textstyle\sum_k \left(\boldsymbol{y}_k^{\mathrm{t}}(\boldsymbol{\Gamma}_k^y)^{-1}\boldsymbol{y}_k - \log\det 2\pi\boldsymbol{\Gamma}_k^y\right) \qquad (5)$$

$$P(\boldsymbol{e}) = (1+e^{-\alpha})^{1-N}\exp -\alpha\textstyle\sum_{k=1}^{N-1} e_k.$$

Matrices $\boldsymbol{\Gamma}_k^y$ and $\boldsymbol{\Gamma}_k^{x|y}$ are respectively the covariance of $\boldsymbol{y}_k$ knowing $\boldsymbol{e}$, and of $\boldsymbol{X}_k$ knowing $\boldsymbol{y}_k$ and $\boldsymbol{e}$.

3.2. Computation of the likelihood of a block

The probability of a block $\pi(\boldsymbol{y}_{i,j})$ can be found in a constant number of operations when $\pi(\boldsymbol{y}_{i,j-1})$ is known. This is a classical issue optimally solved by a Kalman smoother [4]. Here, we rather use a Gram-Schmidt structure that is not more expensive. By doing so, useful intermediate results that could not be obtained by the Kalman procedure are available for subsequent purposes.

In order to evaluate the probability in (5), we need to compute the quantities, $\boldsymbol{y}_{i,j}^{\mathrm{t}}\boldsymbol{\Gamma}^y(j-i)^{-1}\boldsymbol{y}_{i,j}$ and $\det\boldsymbol{\Gamma}^y(n)$. Let for $n = 1,\dots,N$:

$$\begin{aligned} c_1 &= (1+\sigma_\epsilon^2\sigma_b^{-2})^{-1/2}, \\ c_n &= (1+\beta^2+\sigma_\epsilon^2\sigma_b^{-2}-c_{n-1}^2\beta^2)^{-1/2}, \\ \tilde{c}_1 &= (1-\beta^2+\sigma_\epsilon^2\sigma_b^{-2})^{-1/2}, \\ \tilde{c}_n &= (1+\sigma_\epsilon^2\sigma_b^{-2}-c_{n-1}^2\beta^2)^{-1/2}, \end{aligned}$$

and for $i = 0,\dots,N-1$ and $j = i+1,\dots,N$:

$$\begin{aligned} u_{i,j} &= (y_j+\beta c_{j-i-1}u_{i,j-1})c_{j-i}, & u_{i,i+1} &= y_{i+1}c_1, \\ \tilde{u}_{i,j} &= (y_j+\beta c_{j-i-1}u_{i,j-1})\tilde{c}_{j-i}, & \tilde{u}_{i,i+1} &= y_{i+1}\tilde{c}_1, \\ z_{i,j} &= \sigma_b^{-2}y_j^2 - \sigma_\epsilon^2\sigma_b^{-4}u_{i,j}^2 + z_{i,j-1}, & z_{i,i} &= 0, \end{aligned}$$

then we have:

$$\boldsymbol{y}_{i,j}^{\mathrm{t}}\boldsymbol{\Gamma}^y(j-i)^{-1}\boldsymbol{y}_{i,j} = \sigma_b^{-2}y_j^2 - \sigma_\epsilon^2\sigma_b^{-4}\tilde{u}_{i,j}^2 + z_{i,j-1}.$$

In addition, the second term follows from:

$$\det\boldsymbol{\Gamma}^y(n)^{-1} = \sigma_b^{-2n}(1-\beta^2)\tilde{c}_n^2\textstyle\prod_{i<n}c_i.$$

When the latter algorithm has been performed, the probabilities of all the blocks are known in order N^2 operations.

It only remains to use the recursion based on (4) which is of order N^3 operations. But this critical step is made of simple calculations. Indeed the inmost loop only contains an addition. As a result, we have an efficient algorithm for the likelihood computation.

4. PARAMETERS ESTIMATION

In the preceding part, we addressed the computation of the likelihood as it enables us to find the proper parameters, but its expression is too unwieldy to be maximized with respect to them easily. The alternative is to use an iterative method. The Expectation-Maximization method (EM) appears to fit this issue. Actually, it is feasible for $(\alpha, \sigma_\epsilon, \sigma_b)$ but not for β.

To formulate the problem, let the vector of parameters we want to estimate be $\boldsymbol{\theta} = (\alpha, \sigma_\epsilon^2, \sigma_b^2)$. Then we construct $(\boldsymbol{\theta}_n)$ a sequence whose aim is to converge to it. Let:

$$Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) = E\left[\log p(\boldsymbol{L}, \boldsymbol{y}, \boldsymbol{x}\,;\, \tilde{\boldsymbol{\theta}}) \,|\, \boldsymbol{y}\,;\, \boldsymbol{\theta}\right] p(\boldsymbol{y}\,;\, \boldsymbol{\theta}) \qquad (6)$$

Here, we multiply the usual function found in [2] by $p(\boldsymbol{y}\,;\,\boldsymbol{\theta})$ to have simpler formulations. Since it does not depend on $\tilde{\boldsymbol{\theta}}$, this will not change the algorithm. The scheme of the fixed-point method is to find: $\boldsymbol{\theta}_{n+1} = \arg\max_{\boldsymbol{\theta}} Q(\boldsymbol{\theta}_n, \boldsymbol{\theta})$. In some cases $(\boldsymbol{\theta}_n)$ converges to the solution of the optimization problem. But generally, we cannot assure that this will be observed.

In general, evaluation of the function Q is a difficult task, if not impossible. In the present case, we have the following formula:

$$\begin{aligned} Q(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}) \;=\; & \textstyle\sum_{0\le i<j\le N} \pi(\boldsymbol{y}_{i,j})\pi_{i,j} p(\boldsymbol{y}_{0,i}) p(\boldsymbol{y}_{j,N}) \qquad (7) \\ & \times \int_{\boldsymbol{x}_{i,j}} \tilde{\phi}(\boldsymbol{y}_{i,j}, \boldsymbol{x}_{i,j}, i, j)\pi(\boldsymbol{x}_{i,j} \,|\, \boldsymbol{y}_{i,j})\, \mathrm{d}\boldsymbol{x}_{i,j} \end{aligned}$$

where $\tilde{\phi}$ is a non unique set of Gibbsian potentials for $p(\boldsymbol{L}, \boldsymbol{x}, \boldsymbol{y}\,;\,\tilde{\boldsymbol{\theta}})$ verifying:

$$\textstyle\sum_k \tilde{\phi}(\boldsymbol{y}_k, \boldsymbol{x}_k, l_k, l_{k+1}) = \log p(\boldsymbol{L}, \boldsymbol{x}, \boldsymbol{y}, \tilde{\boldsymbol{\theta}}) \qquad (8)$$

and $\pi(\boldsymbol{x}_{i,j} \,|\, \boldsymbol{y}_{i,j})$ is the Gaussian distribution of $\boldsymbol{x}_{i,j}$ knowing $\boldsymbol{y}_{i,j}$ on the block $[i,j]$. $\pi(\boldsymbol{y}_{i,j})$ is the Gaussian distribution on the block $[i,j]$. $p(\boldsymbol{y}_{i,j})$ is the distribution of $\boldsymbol{y}$ on the interval $[i,j]$. These distributions are all under $\boldsymbol{\theta}$. One of the good potentials that we have considered is:

$$\begin{aligned} & \phi(\boldsymbol{x}_k, \boldsymbol{y}_k, i, j) = \\ & \frac{1}{2\sigma_\epsilon^2}\boldsymbol{x}_k^{\mathrm{t}}\boldsymbol{P}^{\mathrm{t}}\boldsymbol{P}\boldsymbol{x}_k + \frac{1}{2\sigma_b^2}\left\|\boldsymbol{x}_k - \boldsymbol{y}_k\right\|^2 + \frac{1}{2\sigma_x^2}\boldsymbol{x}_k^{\mathrm{t}}\boldsymbol{J}\boldsymbol{x}_k + \log\frac{\sigma_x}{\sigma_\epsilon} \\ & +\alpha - 1_{j=N}\left[\alpha - N\log(\sigma_\epsilon\sigma_b) + (N-1)\log(1+e^{-\alpha})\right]. \end{aligned}$$

All the calculations done so far correspond to the E-step of the algorithm. We now have the materials required for the optimization.

4.1. Estimation of α

In the following three parts, we address the resolution of the M-step. Each time, we solve for a different parameter.

We begin with $\hat{\alpha}$. Taking the derivative of Q, equating it to 0, and splitting the summations, it is easy to find:

$$\hat{\alpha} = \log\frac{(N-1)\sum_{0\le i<j<N} p(\boldsymbol{y}_{0,i})\pi(\boldsymbol{y}_{i,N})\pi_{i,N}}{\sum_{0\le i<j<N} p(\boldsymbol{y}_{0,i})\pi(\boldsymbol{y}_{i,j})\pi_{i,j}p(\boldsymbol{y}_{j,N}) - 1}. \qquad (9)$$

4.2. Estimation of σ_b

The derivation with respect to σ_b^2 yields a similar result:

$$\begin{aligned} \hat{\sigma}_b^2 = & \left[N\textstyle\sum_{i<N} p(\boldsymbol{y}_{0,i})\pi(\boldsymbol{y}_{i,N})\pi_{i,N}\right]^{-1}\Big[\left\|\hat{\boldsymbol{x}}_{i,j} - \boldsymbol{y}_{i,j}\right\|^2 \\ & + \operatorname{tr}\boldsymbol{\Gamma}^{x|y}(j-i)\Big]\Big[\textstyle\sum_{i<j} p(\boldsymbol{y}_{0,i})\pi(\boldsymbol{y}_{i,j})p(\boldsymbol{y}_{j,N})\pi_{i,j}\Big]. \qquad (10) \end{aligned}$$

None of the terms in the brackets are known from a previous calculation. However, we can estimate them from the by-products of the likelihood computation. We expand the first one to obtain:

$$\begin{aligned} \left\|\hat{\boldsymbol{x}}_{i,j} - \boldsymbol{y}_{i,j}\right\|^2 \;=\; & \left\|\boldsymbol{y}_{i,j}\right\|^2 - 2\sigma_b^{-2}\boldsymbol{y}_{i,j}^{\mathrm{t}}\boldsymbol{\Gamma}^{x|y}(j-i)\boldsymbol{y}_{i,j} \\ & +\sigma_b^{-4}\left\|\boldsymbol{\Gamma}^{x|y}(j-i)\boldsymbol{y}_{i,j}\right\|^2. \end{aligned}$$

Only the last term is to be calculated. This is the purpose of the set of equations stated underneath.

$$\begin{aligned} J_{i,i+1} &= \beta\, c_2\, d_1^2\, y_{i+1}, \\ J_{i,j-1} &= \beta c_{j-i+1} c_{j-i}(J_{i,j-1} + d_{j-i}u_{i,j}), \\ \tilde{J}_{i,j-1} &= \beta \tilde{c}_{j-i+1} c_{j-i}(J_{i,j-1} + d_{j-i}u_{i,j}), \\ I_{i,i+1} &= y_{i+1}^2 d_1^2, \\ I_{i,j-1} &= I_{i,j-1} + 2J_{i,j-1}u_{i,j} + u_{i,j}^2 d_{j-i}, \\ \tilde{I}_{i,i+1} &= I_{i,j-1} + 2\tilde{J}_{i,j-1}\tilde{u}_{i,j} + \tilde{u}_{i,j}^2 \tilde{d}_{j-i}, \end{aligned}$$

$$\text{then} \quad \left\|\boldsymbol{\Gamma}^{x|y}(1)\, y_{i,i+1}\right\|^2 = \left[\frac{\sigma_b^2\sigma_\epsilon^2 y_{i+1}}{(1-\beta^2)\sigma_b^2 + \sigma_\epsilon^2}\right]^2,$$

$$\left\|\boldsymbol{\Gamma}^{x|y}(j-i)\, y_{i,j}\right\|^2 = \sigma_\epsilon^4 \tilde{I}_{i,i+1} \quad \text{if } j > i+1.$$

Similarly, the trace of the matrix is obtained by recursions using sequences calculated during the likelihood issue:

$$\begin{aligned} d_1 &= (1 + \sigma_\epsilon^2\sigma_b^{-2})^{-1}, \\ d_n &= c_n^2(1 + c_{n-1}^2\beta^2 d_{n-1}), \\ \tilde{d}_n &= \tilde{c}_n^2(1 + c_{n-1}^2\beta^2 d_{n-1}), \\ s_n &= s_{n-1} + d_n, \end{aligned}$$

$$\text{then} \quad \operatorname{tr}\boldsymbol{\Gamma}^{x|y}(1) = \frac{\sigma_b^2\sigma_\epsilon^2}{(1-\beta^2)\sigma_b^2 + \sigma_\epsilon^2},$$

$$\operatorname{tr}\boldsymbol{\Gamma}^{x|y}(n) = \sigma_\epsilon^2(S_{n-1} + \tilde{d}_n) \quad \text{if } n > 1.$$

4.3. Estimation of σ_ϵ

Once again, we apply the same scheme for the solution and no new recursion is necessary to compute derive the last reestimation formula.

$$\hat{\sigma}_\epsilon^2 = \frac{\sum_{i<j} p(\boldsymbol{y}_{0,i})\pi(\boldsymbol{y}_{i,j})p(\boldsymbol{y}_{j,N})\pi_{i,j}\sigma_\epsilon^2 A(\boldsymbol{y}_{i,j})}{N\sum_{i<N} p(\boldsymbol{y}_{0,i})\pi(\boldsymbol{y}_{i,N})\pi_{i,N}} \qquad (11)$$

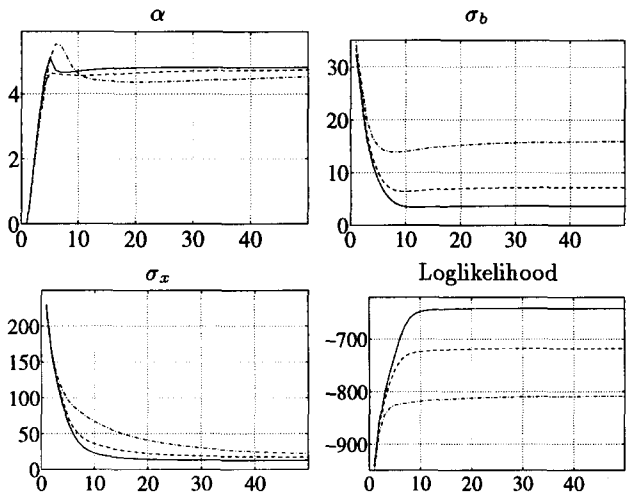


Figure 1: Behavior of the EM algorithm on the antisymmetric step of [1] with added Gaussian noise,
(—) $\sigma = 4$, (– –) $\sigma = 8$, (–·) $\sigma = 16$.

where $A(\boldsymbol{y}_{i,j}) = -\sigma_b^{-2} \operatorname{tr} \boldsymbol{\Gamma}^{x|y}(j-i) + j - i$
$-\sigma_b^{-4} \boldsymbol{y}_{i,j}^{\mathrm{t}} \boldsymbol{\Gamma}^{x|y}(j-i) \boldsymbol{y}_{i,j} - \sigma_b^{-6} \left\| \boldsymbol{\Gamma}^{x|y}(j-i)\, \boldsymbol{y}_{i,j} \right\|^2 .$

4.4. The algorithm

The estimation formulae (9,10,11) are reminiscent of the classical EM formulae. Besides, they resemble each other apart from one factor. In particular, the denominator involves always the same sum. Therefore, it is efficient to conduct the three estimations at the same time. When the likelihood is already computed, the EM improvement of the parameters is of order n^2 operations. Thus it is negligible in comparison with the previous computations. All this is made possible by using the optimized procedure instead of brutally inverting the matrices at each step.

5. EXPERIMENTAL RESULTS

Our model is based on the weak string defined by Blake. As an attempt to stick to his study, and because the data are available, we performed our algorithm on the curves he gave. The antisymmetric step fits our purposes of testing because it was corrupted with different realizations of noise. As is depicted in Fig. 1. The treatment required few iterations. We verified that the estimation of the parameters α and σ_x describing the underlying process was not affected by a change in the noise variance. This is logical since this process does not depend on σ_b; but was not clear from the expression of the re-estimation formulae. Besides, the estimated variance of the noise matches with what was expected according to [1].

Finally, there remains one parameter, β which cannot be evaluated. Table 1 describes the behavior of the estimated parameters with respect to β. As it tends to 1, they converge to limit values. In this respect, it allows us to extend some of the properties of the model to this limit case.

Table 1: Range of values of the estimated parameters when β is close to 1. The original signal is a step found in [1] with an added Gaussian noise with $\sigma_b = 4$ for the first three lines, 8 the second three ones, and 16.

β	0.9	0.99	0.9999	$1-10^{-6}$	$1-10^{-10}$
α	4.767	4.814	4.755	4.705	4.702
σ_b	1.025	3.635	3.914	4.025	4.030
σ_x	14.39	13.07	27.81	32.45	31.94
α	4.618	4.731	4.694	4.673	4.672
σ_b	5.278	7.187	7.616	7.676	7.678
σ_x	17.87	17.27	29.30	32.46	32.46
α	4.338	4.600	4.650	4.653	4.653
σ_b	14.09	15.99	16.11	16.10	16.10
σ_x	22.51	19.13	28.14	30.66	30.70

For instance a segmentation is feasible no matter how close to an unstable process the model is. We want to stress that in the cases at hand, we always found a good edge detection with estimated parameters. On the contrary, when they are chosen at random, the results often prove to be aberrant.

6. CONCLUSION

Addressing the weak string defined by Blake in a deterministic framework as a Markov process is possible due to the Bayesian interpretation. This leads us to a model made of Gaussian AR(1) bounded by random breakpoints. In this respect, the Expectation Maximization, EM algorithm gives good results and particularly matches the characteristics of the model. Indeed, its computation is almost straightforward when the likelihood is known. We gave optimized procedures for performing both the calculation of the likelihood and the EM re-estimation of the parameters. Other studies in the same field, proved that this approach enables us to define new segmentation estimators which give an alternative to Blakes's dynamic programming algorithm.

7. REFERENCES

[1] A. Blake, "Comparison of the efficiency of deterministic and stochastic algorithms for visual reconstruction", *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-11, no. 1, pp. 2–12, January 1989.

[2] A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm", *J. Roy. Statist. Soc. B*, vol. 39, pp. 1–38, 1977.

[3] L. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains", *Annals of Mathematical Statistics*, vol. 41, no. 1, pp. 164–171, 1970.

[4] R. Kalman, "A new approach to linear filtering and prediction problems", *J. Basic Engng.*, vol. 82-D, pp. 35–45, 1960.