MAP AND LS ESTIMATION OF ABRUPT CHANGES IN MULTIPLICATIVE NOISE USING DYNAMIC PROGRAMMING

Martial COULON and Jean-Yves TOURNERET
INP ENSEEIHT / TéSA, 2 rue Camichel BP 7122
31071 TOULOUSE Cedex 7, FRANCE
e-mail: coulon@len7.enseeiht.fr, tournere@len7.enseeiht.fr

ABSTRACT

This paper addresses the problem of estimating abrupt changes corrupted by multiplicative noise. Two estimators are considered depending on the noise assumptions: the Maximum A Posteriori (MAP) estimator and the Least-Squares (LS) estimator. Both estimators are computed using dynamic programming. This study is then applied to edge detection in SAR images.

1 PROBLEM FORMULATION

Edge detection is a challenging issue in Synthetic Aperture Radar (SAR) image analysis. Indeed, because of the multiplicative noise, known as speckle, most standard edge detectors such as gradient-based detectors are inefficient in SAR images. A new edge detector for SAR images, denoted ROEWA detector, was recently studied in [2]. The ROEWA detector is a line-by-line and column-by-column detector, which consists of computing the ratio of exponentially weighted averages (ROEWA) on opposite sides of the central pixel in the horizontal and vertical directions. Thus, the observed data are filtered by an optimal filter minimizing an appropriate mean square error [2]. This paper studies line-by-line and column-by-column SAR image changepoint estimators which minimize Bayesian or LS cost functions.

A line of the SAR image intensity can be modelled by:

$$y_n = b_n m_n + e_n; \qquad n = 1; \ldots; N, \tag{1}$$

where $b_n; e_n; m_n$ and y_n are the multiplicative noise, the additive noise, the uncorrupted and corrupted line of the SAR image, respectively. Eq. (1) assumes that the transfer function of the SAR system does not vary significantly over the bandwith of interest as in [2]. In SAR image processing, the time series $b_n; e_n$ and m_n can be defined as follows:

- the uncorrupted line of the SAR image can be written:

$$m_n = A_i; \quad n \in \left[l^\star_{i-1}; l^\star_i \right]; \quad i = 1; \ldots; K^\star; \tag{2}$$

where $K^\star$ is the number of steps, and A_i is the amplitude of the ith step. Denote T as the sampling period. The actual change locations are $t_i = l_i T + \tau$, with $0 < \tau < T$. The normalized segment lengths are assumed to be bounded, i.e. [5]:

$$\exists \alpha > 0 \text{ such that } \frac{l_i - l_{i-1}}{N} > \alpha; \quad i = 1; \ldots; K^\star \tag{3}$$

- the marginal distribution of the multiplicative speckle noise b_n is a Gamma distribution with parameter L (when L images or looks are averaged), with probability density function (pdf):

$$f(b) = \frac{L^L b^{L-1}}{\Gamma(L)} \exp(-Lb), \qquad b \geq 0 \tag{4}$$

The mean and variance of this distribution are $E[b_n] = 1$ and $Var[b_n] = 1/L$, which shows the speckle noise reduction due to the averaging of L independent images,

- the additive measurement noise e_n is usually modelled by a zero-mean possibly colored Gaussian sequence which is independent of the multiplicative noise b_n.

This paper addresses the problem of estimating the changepoints $l^\star_i$ from an observed SAR image or from a line of this SAR image i.e. y_n, $n = 1; \ldots; N$. Unfortunately, because of the speckle noise coloration, the joint distributions of the SAR image or SAR line intensities have not a tractable closed form expression, even if the additive noise e_n can be neglected with respect to $b_n m_n$ [4]. This paper studies two SAR image edge detectors. The first detector assumes that the multiplicative noise is an i.i.d. sequence and that the additive noise can be neglected. Both assumptions are realistic in some practical applications, as emphasized in [2]. The Maximum A Posteriori (MAP) estimator is then shown to be well suited for the detection. The second detector is based on the Least-Squares (LS) criterion. The LS detector can be performed without prior knowledge on the signal and noise distributions. Consequently, it is well suited to SAR image intensities corrupted by colored multiplicative noise and additive noise.

2 MAP ESTIMATION

When the additive noise e_n can be neglected, the observed signal can be written:

$$y_n = b_n m_n \qquad n = 1, \ldots, N,$$

where the unknown parameters are the changepoint locations $l^* = \left(l_1^*, \ldots, l_{K^*-1}^*\right)^t$, the changepoint amplitudes $A = (A_1, \ldots, A_K)^t$ and K^*. The parameter a posteriori pdf is used in the Bayesian formalism. This a posteriori pdf is the product of the likelihood function conditioned on the parameters and the parameter priors. The likelihood function of the observed data $y = (y_1, \ldots, y_N)^t$ (where t denotes transposition), conditioned on the changepoint locations $l = (l_1, \ldots, l_{K-1})^t$ and amplitudes $A = (A_1, \ldots, A_K)^t$ is defined by:

$$\begin{aligned} f(y|K;A;l) &= \prod_{k=1}^{K} \prod_{i=l_{k-1}+1}^{l_k} \left(\frac{L}{A_k}\right)^L \frac{y_i^{L-1}}{(L-1)!} \exp\left(-\frac{Ly_i}{A_k}\right) \\ &= \left(\frac{L^L}{(L-1)!}\right)^N \prod_{i=1}^{N} y_i^{L-1} \prod_{k=1}^{K} \frac{1}{A_k^{Ln_k}} \exp\left(-\sum_{k=1}^{K} \frac{LS_k}{A_k}\right) \end{aligned} \tag{5}$$

with $n_k = l_k - l_{k-1}$ and $S_k = \sum_{i=l_{k-1}+1}^{l_k} y_i$. The following reparametrization was used successfully in [5], [6]:

$$\begin{cases} r_j = 1 & \text{if there is a changepoint at lag } j, \\ r_j = 0 & \text{otherwise,} \end{cases}$$

with $j = 1, \ldots, N-1$. Conventionally, $r_N = 1$ such that the number of changepoints is equal to the number of steps denoted $K(r) = \sum_{j=1}^{N} r_j$. In a Bayesian framework, $\theta = (r, A(r))$ is estimated from the posterior distribution $f(\theta|y)$ by minimizing the mean of an appropriate cost function ([7], p. 55). The likelihood function of y can then be rewritten:

$$f(y|\theta) \propto \exp\left[-L \sum_{k=1}^{K(r)} \left\{\frac{S_k(r)}{A_k(r)} + n_k(r)\log(A_k(r))\right\}\right] \tag{6}$$

where "$\propto$" means "proportional to", $r = (r_1, \ldots, r_{N-1})^t$, $A(r) = \left(A_1(r), \ldots, A_{K(r)}(r)\right)^t$ and $\theta = (r, A(r))$. This paper focuses on the MAP estimator of θ defined by:

$$\hat{\theta}_{MAP} = \arg\max_{\theta=(r;A)} f(\theta|y).$$

The choice of parameter priors in Bayesian inference is a fundamental problem. In this paper, we propose to use the following priors for the changepoint detection problem:

1) Independent Bernoulli priors are chosen for the changepoint locations:

$$f(r|\lambda) = \lambda^{K(r)-1}(1-\lambda)^{N-K(r)}, \qquad r \in \{0,1\}^{N-1}. \tag{7}$$

The parameter $\lambda \in]0;1[$ is the Bernoulli parameter which represents the a priori probability to have a changepoint at a given position,

2) Independent uniform improper priors are chosen for the step amplitudes:

$$f(A(r)|r) = \prod_{i=1}^{K(r)} I_{]0;+\infty[}(A_i(r)), \tag{8}$$

where $I_A(.)$ is the indicator of the set A ($I_A(x) = 1$ if $x \in A$ and $I_A(x) = 0$ if $x \notin A$). This non-informative prior density expresses ignorance about the value of the parameter vector $A(r)$.

Straighforward computations (using Bayes' theorem) show that the MAP estimator of θ reduces to the minimization of

$$U_{MAP}(r) = \beta K(r) + \sum_{k=1}^{K(r)} Ln_k \log\frac{S_k}{n_k}, \tag{9}$$

with respect to (w.r.t.) r, or equivalently to the minimization of

$$U_{MAP}(K;l) = \beta K + \sum_{k=1}^{K} Ln_k \log\frac{S_k}{n_k}, \tag{10}$$

w.r.t. $(K;l)$. In eq.'s (9) and (10), $\beta = \log\frac{1-\lambda}{\lambda}$. The changepoint amplitude MAP estimations are then defined by $\hat{A}_k = \frac{S_k}{n_k}$, i.e. by the means of the observed process y_n on $]l_{k-1}; l_k]$. Note that the parameter β is a decreasing function of the a priori probability to have a changepoint at a given position λ. Consequently, the smaller β, the higher the number of changes. Unfortunately, a closed-form expression of the MAP estimator cannot be obtained. Thus, numerical algorithms have to be considered to solve the minization problem.

3 LS ESTIMATION

When the changepoint number $K^* - 1$ is known, the Least-Squares (LS) algorithm consists of determining the changepoint locations $(l_i)_{i=1,\ldots,K^*-1}$ by minimizing the following criterion w.r.t. l [6]

$$U_{LS}(l) = \sum_{i=1}^{K^*} \sum_{n=l_{i-1}+1}^{l_i} \left(y_n - \frac{S_i}{n_i}\right)^2. \tag{11}$$

Provided that b_n and e_n have specific structures (for instance, b_n and e_n can be stable ARMA sequences), the LS solution $\hat{l}$ minimizing the criterion (11) converges in probability to the true value l^*, when $N \to \infty$ [5].

When the changepoint number is unknown, the previous criterion (11) has to be penalized. The Penalized Least Squares (PLS) problem consists of minimizing w.r.t. $(K;l)$ the following criterion

$$U_{PLS}(K;l) \triangleq \sum_{i=1}^{K} \sum_{n=l_{i-1}+1}^{l_i} \left(y_n - \frac{S_i}{n_i}\right)^2 + \gamma(N)K. \tag{12}$$

In order to ensure the convergence in probability of the estimates $(\hat{K};\hat{l})$ to $(K^*;l^*)$, it is assumed, along with eq. (3), that the penalizing parameter $\gamma(N)$ is a positive real sequence such that

$$\lim_{N\to\infty}\gamma(N)=0 \text{ and } \lim_{N\to\infty} N^{2-h}\gamma(N)=+\infty;$$

where $h\in[1;2[$ depends on the multiplicative and additive noise structures [5] (in particular, it can be proved that $h=1$ when b_n and e_n are ARMA sequences [1]). The choice of $\gamma(N)$ depends on the resolution considered to fit the signal. Indeed, the smaller $\gamma(N)$, the larger the number of changes ($\gamma(N)$ in (12) is very similar to β in (9)). In the extreme case $\gamma(N)=0$, a changepoint is detected at each lag n.

4 DYNAMIC PROGRAMMING

Eq. (10) and (12) show that the MAP and LS estimators require the optimization of an energy function which is defined on a finite but huge set. For large number of samples N and number of changes K^*, enumeration cannot be achieved because of high computational cost. Consequently, numerical methods have to be considered. The LS and MAP energy functions were minimized successfully in [6] using Monte-Carlo Markov Chains (MCMC) methods. MCMC methods consist of generating samples by running a Markov chain whose target distribution is the energy function. These samples can then be used to estimate the global minimum of the energy function. However, these stochastic techniques only provide approximated solutions and are time demanding. Instead, this paper proposes to compute the MAP and LS estimators using Dynamic Programming (DP) [3]. DP is a deterministic algorithm which yields the exact solution of a discrete minimization problem without evaluating all combinations. Denote $\Delta_i(l_{i-1}+1;l_i)$ as the contribution of the ith segment in the MAP or LS criterion, i.e. $\Delta_i(l_{i-1}+1;l_i) = L n_i \ln\frac{S_i}{n_i} + \beta$ for the MAP estimator, and $\Delta_i(l_{i-1}+1;l_i) = \sum_{n=l_{i-1}+1}^{l_i}\left(y_n - \frac{S_i}{n_i}\right)^2 + \gamma$ for the PLS estimator. The energy function to be optimized can be written as:

$$U(K;l) = \sum_{k=1} \Delta_i(l_{i-1}+1;l_i),$$

where $U(K;l)$ stands for $U_{MAP}(K;l); U_{LS}(l)$ or $U_{PLS}(K;l)$. Define

$$I_k(L) \triangleq \min_{\substack{0<l_1<\dots<l_{k-1}<L\\ l_0=0,\ l_k=L}} \sum_{i=1} \Delta_i(l_{i-1}+1;l_i)\ , \quad L\in[1;N]; \qquad (13)$$

such that $I_{K^*}(N)$ is the minimum of $U(K;l)$. It is straightforward to prove the following recursion:

$$I_k(L) = \min_{l_{k-1}\in[k-1;L[}\left(I_{k-1}(l_{k-1}) + \Delta_k(l_{k-1}+1;L)\right) \qquad (14)$$

for $L\in[1;N]$. Consequently, the optimal number of segments is

$$\hat{K} = \arg\min_{k\in[1;N]} I_k(N)\ . \qquad (15)$$

Eq.'s (14) and (15) show that the minimization of $U(K;l)$ w.r.t. $(K;l)$ can be performed using the following algorithm:

1. $k=1$;
   - for $L=1;\dots;N$, compute $I_1(L)=\Delta_1(1;L)$
2. for $k=2;\dots;N$,
   - for $L=k;\dots;N$, compute

   $$\begin{aligned} I_k(L) &= \min_{l_{k-1}\in[k-1;L[}\left(I_{k-1}(l_{k-1}) + \Delta_k(l_{k-1}+1;L)\right) \\ \underline{l}_{k-1}(L) &= \arg\min_{l_{k-1}\in[k-1;L[}\left(I_{k-1}(l_{k-1}) + \Delta_k(l_{k-1}+1;L)\right) \end{aligned}$$

3. computation of $\hat{K} = \arg\min_{k\in[1;N]} I_k(N)$
4. computation of $\left(\hat{l}_k\right)_{k=1;\dots;\hat{K}}$:
   - for $k=\hat{K}$,

     $\hat{l}_{\hat{K}} = N$,
   - for $k=\hat{K}-1;\dots;1$,

     $\hat{l}_k = \underline{l}_k\left(\hat{l}_{k+1}\right)$

It is important to note that the computational cost of DP grows linearly with the number of steps (whereas enumeration computation grows exponentially). In practice, an upper bound K_{max} for K is often available. In this case, it is sufficient to compute $\hat{K} = \arg\min_{k\in[1;K_{max}]} I_k(N)$, which makes the algorithm much faster. Fig. 1 (resp. 2) presents results obtained for the MAP (resp. LS) estimator with different values of parameter β (resp. γ). As explained above, the higher β (resp. γ), the lower the number of changes. Similar results were obtained in [6] using MCMC instead of DP. However, DP provides a significant gain in the computational cost (the DP algorithm is about 150 times faster than the MCMC algorithm).

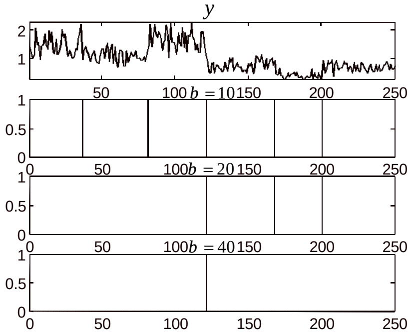

Fig.1: MAP estimation for different values of the penalizing parameter β.

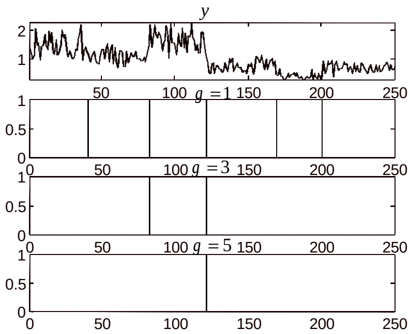

Fig. 2: LS estimation for different values of the penalizing parameter γ.

5 APPLICATION

The MAP or LS estimation can be applied to line-by-line or column-by-column changepoint detection in SAR imagery. Denote $\hat{\epsilon}_{r,i}$ (resp. $\hat{\epsilon}_{c,j}$) as the estimated vector obtained with the ith row (resp., jth column). An edge strength map C can be defined by setting $C_{i,j} = \sqrt{\left(\hat{\epsilon}_{l,i}(j)\right)^2 + \left(\hat{\epsilon}_{c,j}(i)\right)^2}$, $i, j = 1, ..., N$ [2]. In order to obtain skeleton edges, this map has to be smoothed by an appropriate filter. Significant edges are then extracted, thined, and closed using an appropriate modified watershed algorithm [2]. Fig. 3 presents the result obtained with the MAP estimator on a real SAR image, which represents an agricultural scene near Bourges in France. This figure shows that significant edges are detected. However, over and under-segmentation may appear and the position of edges is not always very accurate. These problems can be overcome using appropriate repositioning algorithms.

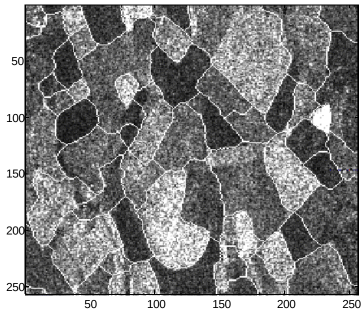

Fig. 3: MAP Edge Detection of an Agricultural Scene. © ESA-ERS-1 data-1993, Distribution Spot Image.

6 CONCLUSION

This paper considered MAP and LS estimators for abrupt change detection in multiplicative noise. Both estimators were computed using dynamic programming. DP provided the exact solution of minimization problems with reduced computational cost compared to MCMC algorithms. The signal resolution was adjusted by the a priori probability to have a change at a given position. The estimation of this hyperparameter is currently under investigation. The changepoint detection was applied successfully to SAR images. Moreover, the use of repositioning technics is also under study.

References

[1] M. Coulon, Contribution à la Détection de Modèles Paramétriques en Présence de Bruit Additif et Multiplicatif, Ph.-D. Dissertation (in French), INPT, n°1569, Toulouse, France, 1999.

[2] R. Fjørtoft, A. Lopès, P. Marthon and E. Cubero-Castan, "An Optimal Multiedge Detector for SAR Image Segmentation," IEEE Trans. Geosci. and Remote Sensing, vol. 36, n° 3, pp. 793-802, May 1998.

[3] S. M. Kay, Modern Spectral Estimation: Theory and Application, Prentice Hall, Englewood Cliffs, 1988.

[4] P.A. Kelly, H. Derin, K.D. Hartt, "Adaptative Segmentation of Speckled Images using Hierarchical Random Field Model," IEEE Trans. Acoustic, Speech, and Signal Processing, vol.36, No: 10, pp. 1628-1641, Oct. 1988.

[5] M. Lavielle and E. Moulines, "Least Squares estimation of an unknown number of shifts in a time series," to appear in J. Time Series Analysis, 1999.

[6] J.Y. Tourneret, M. Coulon and M. Doisy, "Least-Squares Estimation of Multiple Abrupt Changes Contaminated by Multiplicative Noise using MCMC", Proc. of HOS'99, Caesarea, Israel, July 1999.

[7] H. L. Van Trees, Detection, Estimation and Modulation Theory, Part I, New York: Wiley, 1968.