IONIZING RADIATION DETECTION USING JUMP MARKOV LINEAR SYSTEMS

Luc Eglin, Eric Barat, Thierry Montagu, Thomas Dautremer, Jean-Christophe Trama

CEA Saclay, France

ABSTRACT

The systems commonly used to detect photons and estimate their energies are usually irrelevant for high flux. Hidden Markov Model and Jump Markov Linear Systems (JMLS) provide a framework which allows us to get an optimal estimate for stochastic processes, whose occurrences are randomly distributed according to time of detection, length and magnitude. It is perfectly adapted to the spectrometry issue. We use the Maximum a Posteriori (MAP) criterion to estimate the state vector. In the high signal-to-noise ratio (SNR) case, the system can be simplified as a Kalman smoother set up in an on-line version in our lab. An extension in low SNR case is proposed.

Keywords: photon detection, energy estimation, hidden Markov model, Jump Markov Linear Systems.

1. INTRODUCTION

This paper deals with the detection and the energy estimation of photons. The aim is to propose software to process numerical data corresponding to the output signal of a detector (HpGe, Bolometer …). The classical approach is to consider this problem as the optimization of the signal-to-noise ratio. The adaptive filter uses this approach [1]. Let us briefly describe this technique before explaining our approach. One can adopt a general form to describe the observed signal y_t at time t in the presence of photon signal s_t :

$$y_t = s_t + n_t$$

where n_t is a gaussian noise, with spectral density N_ω, and:

$$y_t = n_t$$

if the photon signal is absent. By assuming that s_t is known, it can be shown that the adaptive filter, which optimizes the SNR for a given time T, is chosen by taking the linear filter:

$$\kappa_\omega = Ge^{-i\omega T}\frac{S_\omega^*}{N_\omega}$$

where G is a positive constant, and S_ω^* is the conjugate of the spectral density of the signal. The output of the filter is compared to a threshold to determine if there is detection during time T. So, we must "inverse" the shape of the signal to be detected. In practice, the response must be causal, and this condition involves an approximation of the filter: we have to truncate the ideal response. The approximation is good if the truncated part is short. For that, the time of decision T must be as high as possible. This compromise is not always easy to make. It is the first reason why we choose another approach. From a more general point of view, the problem is that, due to the fact that the time of detection is unknown, the SNR approach is not optimal.

The original approach developed at CEA Saclay is different: we express the problem as an optimal statistical detection based on the observation y_t [2]. The criterion is not the optimization of the SNR, but the maximum likelihood (or MAP) one. In this case, one has to choose between two hypotheses: $\mathbf{H_0}$ if there is only noise, $\mathbf{H_1}$ if there is an impulsion:

$$\begin{cases} \mathbf{H_0}: & y_t = n_t & \text{if only noise} \\ \mathbf{H}_1: & y_t = s_t + n_t & \text{otherwise} \end{cases}$$

$\mathbf{H_1}$ can be viewed as an impulsion occurrence with unknown energy, unknown shape and unknown time of detection. Generalized Likelihood Ratio (GLR) test adopts this approach [3, 4]. The detection and the estimation of the unknown parameters with GLR test requires a model of the physical signal (what we will suppose in the rest of the paper). For instance, we can adopt a state-space model (discrete time):

$$\begin{cases} x_{k+1} = Fx_k + v_k + \Psi(k,\tau_0) \\ y_k = Hx_k + w_k \end{cases}$$

The matrix F contains the parameters of the physical signal and noise (for instance AR parameters if the noise is modeled by an ARMA process). x_k is the state vector, y_k the observation, v_k (w_k respectively) is a dynamic (observation) white noise, H is the observation matrix. $\Psi(k,\tau_0)$ defines the dynamic shape of an impulsion detected at time τ_0. It is not necessary to know the impulsion shape. One only imposes a finite length for the impulsion. The time of detection is assumed to be unknown, which is more realistic than in the previous approach. The algorithms obtained with this approach,

even if not optimal, provide more rigorous results in practice. However, we must admit that these methods have some characteristics which decrease according to the photon flux. We lack a framework allowing us to obtain optimal estimators for stochastic processes which are subject to multiple jumps. These jumps are randomly distributed in time, in length and in magnitude.

2. NEW APPROACH FOR SPECTROMETRY

The time of detection of photons is a stochastic law. This stochastic behavior can be taken into account by considering that the different states of the system (absence or presence of a photon) are represented by a HMM [5]. Let r_k be a Markov chain:

$$\Pr\{r_k = i_k \,|\, r_{k-1} = i_{k-1}, r_{k-2} = i_{k-2}, \ldots, r_1 = i_1\} = \Pr\{r_k = i_k \,|\, r_{k-1} = i_{k-1}\}$$

We consider that the problem can be represented by a state-space model, whose parameters depend on a Markov chain. Such a model is called Jump Markov Linear System [6]. An example of such a system can be represented as:

$$\begin{cases} x_{k+1} = F(r_{k+1})x_k + B(r_{k+1})v_{k+1} \\ y_k = H(r_k)x_k + D(r_k)w_k \end{cases}$$

where:

$$\begin{cases} r_k & \text{unknown Markov chain;} \\ x_k & \text{state of the JMLS, } F \text{ state matrix;} \\ v_k, w_k & \text{gaussian white noise, } B \text{ (resp. } D\text{) dynamic (observation)} \\ & \text{noise matrix;} \\ y_k & \text{noisy data, } H \text{ observation matrix;} \end{cases}$$

With this model, the jump notion does not use a dynamic profile, but makes the system dependant on the finite-state discrete-time Markov chain r_k. Thus the system allows us to take into account model jumps whose occurrences are random. For instance, we can take:

$$\begin{cases} r_k = 0 \text{ if only noise} \\ r_k = 1 \text{ if detection} \end{cases}$$

Now we can use interesting estimates, for instance Maximum a Posteriori (MAP) estimate of x_k and/or r_k:

$$\begin{cases} \hat{\boldsymbol{r}}^{\text{MAP}} = \arg\max_{\boldsymbol{r}} f(\boldsymbol{r}/\boldsymbol{y}) & \text{with } \boldsymbol{r} = \{r_1, \ldots, r_N\} \text{ and } \boldsymbol{y} = \{y_1, \ldots, y_N\} \\ \hat{\boldsymbol{x}}^{\text{MAP}} = \arg\max_{\boldsymbol{x}} f(\boldsymbol{x}/\boldsymbol{y}) & \text{with } \boldsymbol{x} = \{x_1, \ldots, x_N\} \\ (\hat{\boldsymbol{r}}, \hat{\boldsymbol{x}}^{\text{MAP}}) = \arg\max_{(\boldsymbol{r},\boldsymbol{x})} f(\boldsymbol{x}, \boldsymbol{r}/\boldsymbol{y}) \end{cases}$$

Where $f(.\,|\,.)$ is the conditional probability density, and N the number of data points. $\hat{x}$ is the estimate of the system state. It is composed of the estimate of the physical signal (without noise), whose integral gives the energy estimate of the photon (if $r_k = 1$). There are several techniques to optimize the density function. The best results are undoubtedly obtained by the Markov Chain Monte Carlo techniques. But these methods do not yet have reasonable computational time (in our applications, we have to process at least one million impulsions to get a spectrum). It is the reason why we have chosen a deterministic method: the so-called EM (Expectation Maximization) algorithm. The theoretical basis of this approach is taken from [6]. We now present an on-line version, applied to detections with high SNR. We will present the low SNR version in section 4.

3. ON-LINE SYSTEM

If we do not seek detections close to the noise level, we can simplify the previously written system [7]. For that, we consider that detection involves a significant change of the dynamic noise. So only dynamic noise matrix depends on the Markov chain:

$$\begin{cases} x_{k+1} = Fx_k + B(r_{k+1})v_{k+1} \\ y_k = Hx_k + D.w_k \end{cases}$$

$$\text{where } B(r_k) = \begin{cases} 0 \text{ if only noise} \\ B_1 >> 1 \text{ otherwise} \end{cases}$$

In [6], the optimization of conditional probability density of $\boldsymbol{x}$ leads to an iterative algorithm composed of two steps: 1. a fixed-interval Kalman smoother which computes the conditional expectations $\mathrm{E}\{x_k|\boldsymbol{y}, \boldsymbol{R}^{(l)}\}$ and $\mathrm{E}\{x_{k-m}\, x_{k-n}|\boldsymbol{y}, \boldsymbol{r}^{(l)}\}$, where $\boldsymbol{r}^{(l)}$ is the Markov chain estimate at iteration l (Expectation step), 2. a Viterbi algorithm to compute $\boldsymbol{r}^{(l+1)}$ (Maximization step). In the high SNR assumption, we can cancel the Viterbi algorithm step. In our approach, we have only one forward and one backward process, what it corresponds to a Kalman smoother. The detection decision is taken during the forward process. Indeed we use a Hard Rejecter to test if there is a photon:

$$\begin{cases} r_k = 0 \text{ if } |\varepsilon_k| < \eta\sqrt{P_{k|k-1}} \\ r_k = 1 \text{ if } |\varepsilon_k| \geq \eta\sqrt{P_{k|k-1}} \end{cases}$$

ε_k is the innovation, $P_{k|k-1}$ the prediction error covariance, and η a pre-defined constant (typically equal to 3).

Thus an iteration of the forward process is:

$$
\begin{aligned}
&x_{k+1/k} = Fx_{k/k} \\
&\varepsilon_{k+1} = y_{k+1} - Hx_{k+1/k} \\
&P' = FP_{k/k}F^* + B(r_k)B(r_k)^* \\
&V'_{k+1} = H'P'H'^* \\
&\text{if } (\varepsilon_{k+1})^2 \geq \eta^2 V'_{k+1} \quad \text{then} \\
&\quad r_{k+1} = 1 \\
&\text{otherwise} \\
&\quad r_{k+1} = 0 \\
&\text{end if} \\
&P_{k+1/k} = FP_{k/k}.F^* + B(r_{k+1})B(r_{k+1})^* \\
&V_k = HP_{k+1/k}H^* \\
&K_{k+1} = P_{k+1/k}H^*V_k^{-1} \\
&x_{k+1/k+1} = x_{k+1/k} + K_{k+1}\,\varepsilon_{k+1} \\
&P_{k+1/k+1} = (I - K_{k+1}H)\,P_{k+1/k} \\
&k \leftarrow k+1
\end{aligned}
$$

where $x_{k|k-1}$ is the prediction, I the identity matrix, K_{k+1} the Kalman gain, $x_{k|k}$ the state estimate, and $P_{k|k}$ the estimate error covariance. The backward process is a standard Kalman smoother applied to estimate the conditional expectations of the state vector. A stable on-line version has been developed and applied on real data issued from HpGe detectors (patent pending).

4. EXTENSION IN LOW SNR CASE

The JMLS approach has been extended to the low SNR case. State-space model and algorithms are modified to process such data, because the previous assumption (detection involves a significant change of the dynamic noise) in this case is irrelevant. With bolometric detectors, it corresponds to photons with energy around several tens eV. We have tested our algorithm for the detection of photons with energies of 20 eV-10 keV. In order to allow us to detect photons with such energies, we have to determine a model which depends on the energy. Our idea is to incorporate a vector $\gamma(r_k)$ in the dynamic equation of the state-space model, commonly called the input (or control) matrix. This term depends on the Markov chain, such as:

$$
\begin{cases}
r_k = 0 & \text{if only noise} \\
r_k = 1 & \text{lowest energie impulsions} \\
r_k = 2 & \text{second lowest energies impulsions} \\
\quad\vdots & \qquad\vdots \\
r_k = N_m & \text{highest energie impulsions}
\end{cases}
$$

Each Markov chain corresponds on an energy channel. We choose typically from several to several tens energy channels N_m. $\gamma(r_k)$ represents the mean slope of the state vector x_k :

$$
\begin{cases}
\gamma(r_k) = 0 & \text{if } r_k = 0 \\
\gamma(r_k) > \gamma(r_{k-1}) & \text{for } r_k = 1 \ldots N_m
\end{cases}
$$

$\gamma(r_k)$ is associated with the impulsion part of the state vector, so the command is composed by $\gamma(r_k)$ and zeros. Then the state-space model becomes:

$$
\begin{cases}
x_{k+1} = F(r_{k+1})x_k + \begin{pmatrix} \gamma(r_{k+1}) \\ \mathbf{0}_{n_x-1,1} \end{pmatrix} + B(r_{k+1})v_{k+1} \\
y_k = Hx_k + D.w_k
\end{cases}
$$

where n_x is the length of the state-vector, and $\mathbf{0}_{r,c}$ the zero matrix with dimension $r \times c$, and $D << 1$. Here, we make the matrix F dependant on the finite-state discrete-time Markov chain r_k. For convenience, we will assume that $F(r_k)=F$ for $r_k = 0 \ldots N_m$, which means that there is only one possible shape of the impulsions. Let us assume that impulsions s_k are exponential, the noise n_k is an AR-1 process, and there is a bias b_k in the data:

$$
\begin{cases}
s_{k+1} = \alpha s_k + \gamma(r_{k+1}) + \sigma_s(r_{k+1})e_{k+1} \\
n_{k+1} = a_1 n_k + \sigma_n \varepsilon_{k+1} \\
b_{k+1} = b_k + \sigma_b u_{k+1}
\end{cases}
$$

where e_k, ε_k and u_k are normalized white gaussian noises, σ_s, σ_n and σ_b the associated standard deviations. $\sigma_s(r_{k+1})$ is chosen to allow us to get a continuous energy distribution. It is easy to extend this model to a more general one: s_k a sum of exponential functions and n_k an ARMA process. α-coefficients can be determined by fitting a pulse with high SNR, ARMA-coefficients and σ_n from an innovation algorithm, and the standard deviation of the bias by taking the histogram mode of the values taken by the K first data samples. With our hypotheses, the system can be written as:

$$
\begin{cases}
x_{k+1} = \begin{pmatrix} s_{k+1} \\ b_{k+1} \\ n_{k+1} \end{pmatrix} = \overbrace{\begin{pmatrix} \alpha & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & a_1 \end{pmatrix}}^{F} \overbrace{\begin{pmatrix} s_k \\ b_k \\ n_k \end{pmatrix}}^{x_k} + \overbrace{\begin{pmatrix} \gamma(r_{k+1}) \\ 0 \\ 0 \end{pmatrix}}^{\text{"Control"}} \\
\qquad\qquad + \begin{pmatrix} \sigma_s(r_{k+1})e_{k+1} \\ \sigma_b u_{k+1} \\ \sigma_n \varepsilon_{k+1} \end{pmatrix} \\
y_k = \underbrace{\begin{bmatrix} 1 & 1 & 1 \end{bmatrix}}_{H} x_k + D.w_k
\end{cases}
$$

In order to allow us to detect events with magnitude around noise level, it is important to choose accurately the values of $\gamma(r_k)$ and $\sigma_s(r_{k+1})$. At first, we have to fix the mean of the lowest energy distribution. Let us assume that we want to process pulses whose minimum magnitude is around:

$$\mathrm{Mag}_{\min} = 3\sigma_n$$

The value $\gamma(r_k = 1)$ associated with the lowest energy mode is estimated by dividing this magnitude by the length N_p of the impulsion:

$$\gamma(r_k = 1) = \mathrm{Mag}_{\min} / N_p$$

The other values of $\gamma(r_k)$ are either regularly or more and more spaced. For instance, in the simplest case:

$$\gamma(r_k > 1) = r_k \times \gamma(r_k = 1)$$

$\sigma_d(r_k)$ are dependant on $\gamma(r_k)$. We obtain good results by taking simply:

$$\sigma_d(r_k) = \gamma(r_k) - \gamma(r_{k-1}) \text{ for } r_k > 1$$

$\sigma_d(r_k)$ for $r_k = 0$ is dependant on the algorithm iteration (according our approach: it is not always true). This algorithm is a standard MAP-EM. A Kalman filter, then a Kalman smoother are applied by using the augmented version of the model, as described in [6]. The Markov chain is then estimated from a Viterbi algorithm, by using the smoothed estimation of x previously obtained. This algorithm is iterative and stops when the Markov chain is no longer updated ($r_k^{(l)} = r_k^{(l-1)}$ where l is the iteration counter). Moreover, the solution is dependant on the initialization. We have chosen a robust and simple solution for initialization conditions:

$$\begin{cases} r_k^{(l=1)} = 0 \quad \forall k = 1 \ldots N \\ \sigma_s^{(l=1)}(r_k = 0) = \sigma_s(r_k = 1) \\ \sigma_s^{(l>1)}(r_k = 0) = 1\mathrm{e}-6 \end{cases}$$

So we suppose at first there is no detection. In order to allow the system to leave this state, the standard deviation associated to the noise mode is equal to the lowest energetic one. Then it is taken equal to a very low value (not zero for analytic convenience). Finally, the Viterbi step requires the estimations of the probabilities p_{r_k} and p_{r_k, r_m} (respectively, the probability to be in the r_k mode and the transition probability from r_k to r_m). We assume that we have a good approximation of the source activity and the pulse lengths, which allows a simple computation of these values. However, there are some methods to solve this problem, for instance based on Baum-Welch approach.

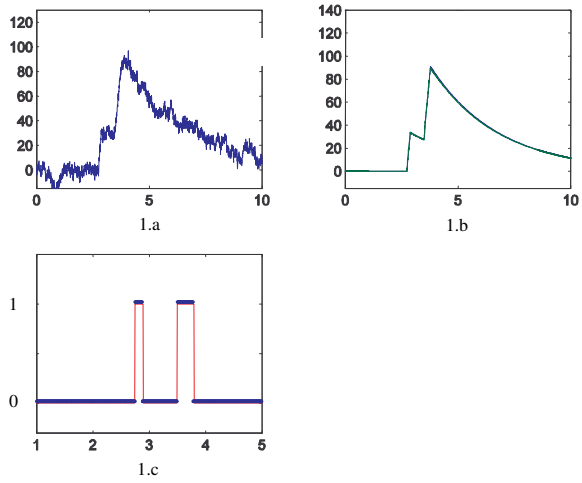

Figure 1: a: simulated data, b: ideal and estimated signal, c: ideal and estimated Markov chain. x-axis: time (ms), y-axis: arbitrary units

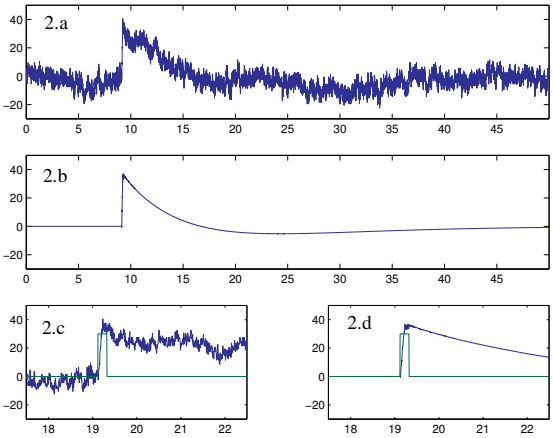

Figure 2: a: real data (*Kirchhoff-Institut für Physik, Universität Heidelberg*). b: estimated signal, c: estimated chain and data d: estimated chain and signal. x-axis: time (ms), y-axis: arbitrary units

5. CONCLUSION

The JMLS approach has been tested from simulated bolometric signal, whose characteristics (noise, impulsion magnitude and time constants) are determined from real data. The first tests on real data are promising. The low SNR corresponds in this case to energies of 20-100 eV. One iteration of the algorithm takes around 1 sec ($N = 2^{16}$), while it requires between one and three iterations. We have analyzed results according to different scenarios. Figure 1 shows a pile-up example: two photons are detected, with two different lengths (energies: 30 then 50 eV). We see here that the estimated Markov chain is very close to the ideal one, and signal is very well restored too.

Table 1 shows quantitative results. We do not observe false alarms with such data. Finally, figure 2 shows a result obtained from real data. The photon energy is around 70 eV. We are currently continuing to validate this approach on real data.

Energy (eV)	Detection rate/100 experiments	Energy error[2] (%)	Length error (%)
6500 (K)[1]	100	0.2	0.0
770 (L)[1]	100	1.0	0.7
100	100	1.1	8.1
75 (M)[1]	100	3.7	12
55	96	9	17
30	73	15	29
20	46	21	45

Table 1: quantitative results obtained with simulated data. [1] K, L and M electron captures energies are given according to ^{55}Fe characteristics. [2] Energy error is determined by comparing the true area to estimated area under the signal during an impulsion.

6. REFERENCES

[1] S. Haykin, “Adaptive Filter Theory,” 3rd edition, Prentice-Hall, 1995.

[2] E. Barat, T. Dautremer, “Du filtre adapté vers une detection optimale en spectrométrie γ”, CEA intern. report (french), 2001.

[3] M. Basseville and I. Nikiforov, “Detection of Abrupt Changes: Theory and Applications,” Prentice Hall, Englewood Cliffs, NJ, 1993.

[4] A.S. Willsky and H.L. Jones, “A generalized likelihood ratio approach to the detection and estimation of jumps in linear systems,” *IEEE Trans. on Automatic Control*, 21, pp: 108-112, February, 1976.

[5] L. R. Rabiner, "A tutorial on hidden Markov models and selected applications in speech recognition," *Proc. of IEEE*, vol. 77, no. 2, pp. 257-286, February 1989.

[6] A. Logothetis, V. Krishnamurthy, “Expectation Maximization Algorithms for MAP Estimation of Jump Markov Linear Systems*,” IEEE Transactions Signal Processing*, Vol.47, No.8, pp.2139-2156, August 1999.

[7] T. Dautremer, E. Barat, “L’expérience Adonis, l’algorithme au coeur des DSP”, CEA internal report (french), 2001.