RESTORING HIDDEN NON STATIONARY PROCESS USING TRIPLET PARTIALLY MARKOV CHAIN WITH LONG MEMORY NOISE

Wojciech Pieczynski and Pierre Lanchantin

GET/INT, CITI Department, CNRS UMR 5157
9, rue Charles Fourier, 91000 Evry, France

ABSTRACT

The hidden Markov chains (HMC), which are widely used in different data restoration problems, have recently been generalised to pairwise partially Markov chains (PPMC), in which the distribution of the observed chain conditional on the hidden one is of any form. In particular, long-memory noise cases can be dealt with. The aim of this paper is to propose a parameter estimation method and to show, via experiments, that unsupervised PPMC based image segmentation can perform better, when the noise is a long-memory one, than the classical HMC based methods.

1. INTRODUCTION

Let $X=(X_n)_{n=1}^N$, $Y=(Y_n)_{n=1}^N$ be two stochastic processes, where X is hidden and Y is observable. Each X_n takes its values in $\Omega=\{\omega_1,\dots,\omega_K\}$ and each Y_n takes its values in R. The problem of estimating X from Y, which occurs in numerous applications, can be solved with Bayesian methods once one has chosen the accurate distribution for $Z=(X,Y)$. The Hidden Markov Chains (HMC) model is the simplest and most well known model. This model has been extended to Pairwise Markov Chains (PMC [4, 8]) and then to Triplet Markov Chains (TMC [7]). The PMC and TMC models, on their hand, have then been extended to Pairwise Partially Markov Chains (PPMC [9]) and Triplet Partially Markov Chains (TPMC [9]), in which the distribution of the noise – in other words, the distribution of Y conditionally on X which will be denoted by $p(y|x)$ – is not necessarily a Markov chain (MC) [9]. One possible application, which we deal with in this paper, is to consider a "long-memory" noise, which occurs in numerous situations [1, 2, 3] and which can not be taken into account via classical Markov models.

The aim of this paper is to show, via some experiments, the existence of situations in which PPMC are of interest with respect to PMC. In particular, we are interested in the long-memory noise in which the correlations in $p(y|x)$ decrease in a "slow" manner. Moreover, we show how the non stationarity of the hidden process can be modelled in the same way as in [6] by a third random process and the use of TPMC.

The organisation of the paper is the following. In the next Section we briefly recall the TMC and PMC models and we develop the PPMC and TPMC models in which the noise is of a long memory kind. We recall how it works and, in particular, how the Bayesian MPM method enables us to recover the hidden process from the observed one. Then we show how to take the non stationarity of the hidden process into account via a TPMC. The third section is devoted to the Gaussian case, which makes possible explicit calculations of interest. Section 4 is devoted to two experiments in which we compare the recent HMC with long-memory noise (HMC-LMN) model, which is a particular PPMC, with the classical HMC model in supervised and unsupervised ways.

2. PAIRWISE PARTIALLY MARKOV CHAINS

2.1 Pairwise and Triplet Markov chains

Let $X=(X_n)_{n=1}^N$, $Y=(Y_n)_{n=1}^N$ be two stochastic processes, where $X=x$ is unobservable and has to be estimated from the observation $Y=y$. The stochastic interactions between the hidden and the observable processes are then given by the distribution $p(x,y)$ of $Z=(X,Y)$. When $p(x,y)=p(z)$ is simple enough, many Bayesian methods are available. In particular, HMC with independent noise (HMC-IN), whose distribution is given by (1), have been widely used and studied.

$$p(z)=p(x_1)p(y_1|x_1)\prod_{n=1}^{N-1}p(x_{n+1}|x_n)p(y_{n+1}|x_{n+1}) \qquad (1)$$

We see that in HMC-IN X is a MC with $p(x) = p(x_1)\prod_{n=1}^{N-1} p(x_{n+1}|x_n)$ and "independent noise" $p(y|x) = \prod_{n=1}^{N} p(y_n|x_n)$ means that the random variables $(Y_n)_{n=1}^N$ are independent conditionally on X. The simplicity of $p(y|x)$ is often difficult to justify. To improve the latter, the HMC-IN model has been generalized into the PMC model, in which one directly assumes the Markovianity of Z, so that:

$$p(z) = p(z_1)\prod_{n=1}^{N-1} p(z_{n+1}|z_n) \tag{2}$$

PMC is strictly more general than HMC; in fact, X is no longer necessarily an MC. However, X remains a MC conditionally on Y, and this property enables the development of analogous Bayesian restoration algorithms.

PMC has further been extended to TMC. Roughly speaking, in the TMC model the distribution of Z is a marginal distribution of the $T = (X, U, Y)$ assumed to be an MC, and $U = (U_n)_{n=1}^N$ is an auxiliary process, which can have a physical signification or not. When the random variables $(U_n)_{n=1}^N$ are not too complex (for example, finite with not too rich set of values), it appears that $X = (X_n)_{n=1}^N$ can still be estimated from $Y = (Y_n)_{n=1}^N$ by Bayesian methods that are analogous to those used in the classical HMC model. In particular, $p(x_n|y)$ is computable, which makes the estimation of x by the Maximum Posterior Mode (MPM) possible. Let us notice that TMC generalizes some classical models in the sense that none of the chains X, U, Y, $V = (X, U)$, $Z = (X, Y)$ or (U, Y) needs to be an MC [11]. The wider generality of PMC with respect to HMC and of TMC with respect to PMC, can also be seen through the expression of $p(y|x)$. In an HMC-IN, $p(y|x)$ is often (too) simple for certain applications. Since in a PMC, $p(y|x)$ is an MC, it is much richer, and in a TMC, $p(y|x)$ is the marginal distribution of the MC $p(u, y|x)$ and therefore even richer than an MC. These increasingly complex models are likely to meet the growing need for a better modelling of the noise in many applications such as, for example, image processing.

2.2 Pairwise and Triplet Partially Markov chains

Recently, PMC have been extended to Pairwise Partially Markov Chains (PPMC), where the distribution of the noise $p(y|x)$ is not necessarily a MC. The pairwise chain $Z = (X, Y)$ is a PPMC if its distribution verifies :

$$\begin{aligned} p(z_{n+1}|z^n) &= p(z_{n+1}|z_n, y^{n-1}) \\ &= p(x_{n+1}|z_n, y^{n-1})\,p(y_{n+1}|z_n, x_{n+1}, y^{n-1}) \end{aligned} \tag{3}$$

where $z^n = (z_i)_{i=1}^n$, and ditto for x^n and y^n. We again find the classical HMC for $p(x_{n+1}|z_n, y^{n-1}) = p(x_{n+1}|x_n)$ and $p(y_{n+1}|z_n, x_{n+1}, y^{n-1}) = p(y_{n+1}|x_{n+1})$, and we again find the classical PMC for $p(x_{n+1}|z_n, y^{n-1}) = p(x_{n+1}|z_n)$ and $p(y_{n+1}|z_n, x_{n+1}, y^{n-1}) = p(y_{n+1}|z_n, x_{n+1})$. Otherwise, the distribution of $Z = (X, Y)$ can be written:

$$p(z) = p(z_1)\prod_{n=1}^{N-1} p(z_{n+1}|z_n, y^{n-1}) \tag{4}$$

This enables one to show classically that $p(x|y)$ is a MC, with transitions given by $p(x_{n+1}|x_n, y)$ $= p(z_{n+1}|z_n, y^{n-1})\beta_{n+1}(x_{n+1})/\beta_n(x_n)$, with $\beta^n(x_n)$ calculable by the following « backward » recursions

$$\begin{cases} \beta^N(x_N) = 1 \\ \beta^n(x_n) = \sum_{x_{n+1}} p(z_{n+1}|z_n, y^{n-1})\beta^{n+1}(x_{n+1}), 1 \le n \le N-1 \end{cases} \tag{5}$$

Therefore, these transitions are calculable once the transitions $p(z_{n+1}|z_n, y^{n-1})$ given by (3) are calculable for every $1 \le n \le N-1$.

By introducing an auxiliary process $U = (U_n)_{n=1}^N$, PPMC can be extended to TPMC in the same manner as PMC can be extended to TMC. The triplet process $T = (X, U, Y)$ will be called "Triplet Partially Markov Chain" if for each $1 \le n \le N-1$, $p(t_{n+1}|t^n) = p(t_{n+1}|v_n, y^n)$. PPMC is then a particular case of TPMC in which $X = U$ (this means that $V = X$ and that there is no latent process)

3. GAUSSIAN PPMC

Let us briefly recall the so-called « Gaussian » PPMC model proposed in [10] (in French). The crux is that in Gaussian PPMC, in which $p(y|x)$ are Gaussian, the transitions $p(x_{n+1}|x_n, y)$ are calculable. More precisely, we need to calculate the transitions $p(z_{n+1}|z_n, y^{n-1})$ given in (3) for every $1 \le n \le N-1$. Let us consider a particular case in which $p(x_{n+1}|z_n, y^{n-1}) = p(x_{n+1}|x_n)$ and $p(y|x)$ are Gaussian. The transitions $p(y_{n+1}|z_n, x_{n+1}, y^{n-1})$ are then also Gaussian and can be recursively calculated using the following classical property (P) :

Property (P):

Let $W = (W_n)_{n=1}^N$ be a real Gaussian chain with, for each $1 \le n \le N$, $M^n = (M_i)_{i=1}^n$ the mean vector and $\Gamma^n = (\gamma_{kl})_{k \le n, l \le n}$ the covariance matrix of $W^n = (W_i)_{i=1}^n$. It is then possible to calculate, for each n, the Gaussian density $p(y^n)$ corresponding to M^n, Γ^n. Classically, one uses the fact that $p(y^n) = p(y^{n-1})p(y_n|y^{n-1})$, where $p(y_n|y^{n-1})$ is Gaussian with mean $M_n + (A^n)^T(\Gamma^{n-1})^{-1}(y^{n-1} - M^{n-1})$ and variance $\gamma_{nn} - (A^n)^T(\Gamma^{n-1})^{-1}A^n$, where $A^n = ((\gamma_{i,n})_{i=1}^{n-1})^T$ (the matrix $(\Gamma^{n-1})^{-1}$ in $p(y_n|y^{n-1})$ is given by the Gaussian density $p(y^{n-1})$).

The idea given in [10] is to apply this property k^2 times (remember that k is the number of possible values of each x_n). More precisely, for each (x_n, x_{n+1}) the transition $p(y_{n+1}|z_n, x_{n+1}, y^{n-1})$ is the transition corresponding to (x_n, x_{n+1}) and it is calculated using the property above.

3.1 Bayesian segmentation using PPMC

To estimate $X = x$ from $Y = y$ by Maximum Posterior Mode (MPM), it is necessary to compute $p(x_n|y)$ for $1 \le n \le N$. It can be done in the following way:

1) Compute $p(z_{n+1}|z^n)$ transitions according to the property (P) by k^2 forward recursions;
2) Compute $\beta^n(x_n)$ by backward recursions and deduce $p(x_{n+1}|x_n, y)$ and $p(x_1|y)$;
3) Compute $p(x_n|y)$ for $1 \le n \le N$ by the following forward recursions: $p(x_{n+1}|y) = \sum_{x_n \in \Omega} p(x_{n+1}|x_n, y)p(x_n|y)$.

We see that the points 2) and 3) are classical and used in HMC, while the point 1) is new and is due to the "partially" Markov aspect of the model.

3.2 Long Memory noise with non stationary hidden process

We are now interested in the long memory noise, in which the correlations of $p(y|x)$ decrease in a "slow" manner so that $\lim_{\tau \to \infty} \tau^\alpha \rho(\tau) = c_\rho$, where $\alpha \in]0,1[$, $c_\rho > 0$ is a constant, and $\rho(\tau)$ is an autocorrelation function. These processes are useful in numerous complex systems [1, 3] and, in particular, in telecommunication networks [2]. As the sequence $\Gamma^n = (\gamma_{kl})_{k \le n, l \le n}$ of covariance matrices considered above is of any kind, it suffices to take them of the form $\gamma_{kl} = \gamma_{nn}\rho(|k-l|) = \gamma_{nn}|k-l|^{-\alpha}$.

When, in such processes Y, there exists a hidden "switching" process X (in other words, when the given long memory process Y is not stationary), we can consider that Y is a noisy version of X, with a long-memory noise. Then the corresponding PPMC enables us to estimate X from Y, using some Bayesian methods like "Maximum a Posteriori" (MAP) or MPM.

Also, we showed in [6] that when (X, Y) is the classical HMC-IN with non stationary X, this non stationarity can be modelled by a third random chain U and the use of the TMC (X, U, Y) enables us to improve the results obtained with the HMC (X, Y). The same idea can be apply considering long memory noise by using the TPMC model.

4. EXPERIMENTS

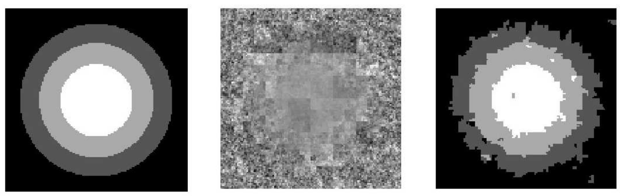
4.1 Supervised restoration of a process with long memory noise

In the first experiment, we consider the following case of a noisy class image segmentation. There are four classes in the hidden image, and it is considered as a realization of a Markov chain X. The mono-dimensional process X is obtained from the bi-dimensional set of pixels using a Hilbert-Peano scan, as already used in [5]. Moreover, we assume that $p(y_{n+1}|x_n, x_{n+1}, y^n) = p(y_{n+1}|x_{n+1}, y^n)$ and $p(x_{n+1}|z_n, y^{n-1}) = p(x_{n+1}|x_n)$.

The four Gaussian distributions (we have four distributions instead of sixteen because of the particular case $p(y_{n+1}|z_n, x_{n+1}, y^{n-1}) = p(y_{n+1}|x_{n+1}, y^n)$) of $W = (W_n)_{n=1}^N$ will also assumed to be stationary, with all means null and all variances equal to one. Thus the correlations are the only discriminating parameters. All the four autocorrelations have the following form: $\rho(\tau) = |\tau + 1|^{-\alpha}$

where $\tau = |j-i|$, $\alpha_{\omega_1} = 0.99$, $\alpha_{\omega_2} = 0.3$, $\alpha_{\omega_3} = 0.05$, $\alpha_{\omega_4} = 0.01$ (and are therefore "long memory" autocorrelations). The model parameters are then the distribution $p(x_1, x_2)$ on $\Omega^2 = \{\omega_1, \omega_2\}^2$, and $\alpha > 0$. The noisy version of (a) with long memory noise is presented on (b) and the MPM restoration is presented on (c).

The model is then a particular case of PPMC that will be denoted by "HMC with long memory noise" (HMC-LMN) in the following.



(a) (b) (c)

Figure 1

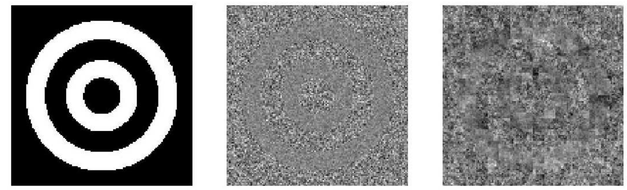
(a) a four classes image, (b) its noisy version (same means and same variances for all four classes), and (c) the Bayesian MPM segmentation result considering the HMC -LMN model. (a) and (b) are converted into mono-dimensional chains via Hilbert-Peano scan (see [5]).

The misclassified pixels' ratio is equal to 6.9%. We can notice that the noise is rather strong and the human eye can hardly distinguish anything in the image (b).

4.2 Unsupervised restoration of a process with long memory noise

In the second experiment, we restore, in an unsupervised way, a two-class image with 2 different noises. In the same way as in the first experiment, we assume that X is a stationary Markov chain after conversion of the image (a) using an Hilbert-Peano scan, and we assume that $p(y_{n+1} | z_n, x_{n+1}, y^{n-1}) = p(y_{n+1} | x_{n+1}, y^n)$. In both cases, the two Gaussian distributions of $W = (W_N)_{n=1}^{N}$ are stationary, with all means null.

The first noisy image presented in the figure 2 (b), is obtained by using an independent noise with variances respectively equal to 1 and 4 for the two classes. The second noisy image presented in (c) is obtained by using a long memory noise, with both variances equal to 1 and $\alpha_{\omega_1} = 0.99$, $\alpha_{\omega_2} = 0.2$.



(a) (b) (c)

Figure 2

(a) is a two classes image, (b) the noisy version with independent noise, and (c) the noisy version with long memory noise

The parameters are estimated by an Expectation-Maximisation (EM) algorithm considering the classical HMC-IN model and the HMC-LMN model, with means set to zero in both cases. The estimations are given in Table I and Table II and the hidden processes restored by MPM are given in Figure 3.

The MPM restoration results concerning the independent noise are quite similar when it comes to the HMC-IN (a) and the HMC-LMN (b) models, with a misclassified pixels' ratio equal to 5.3%. This confirms the fact that HMC-IN is a particular case of the HMC-LMN model. Also, we notice the high estimate values of $\alpha_{\omega_1} = 6.31$, $\alpha_{\omega_2} = 5.09$, which highlights the short range memory nature of the independent noise.

Concerning the long memory noise, we can see that the HMC-IN model (d) is unable to take account of the long range correlation. The wrong estimates imply a poor MPM restoration with a misclassified pixels' ratio equal to 27.6%. On the other hand, the HMC-LMN gives better estimates which results in a good MPM restoration (c) of the hidden process, with a misclassified pixels' ratio equal to 6.5%.

$p(x_1, x_2)$	HMC-IN	HMN-LMN
IN	$\begin{pmatrix} 0.69 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$	$\begin{pmatrix} 0.69 & 0.00 \\ 0.00 & 0.30 \end{pmatrix}$
LMN	$\begin{pmatrix} 0.61 & 0.02 \\ 0.02 & 0.34 \end{pmatrix}$	$\begin{pmatrix} 0.67 & 0.00 \\ 0.00 & 0.32 \end{pmatrix}$

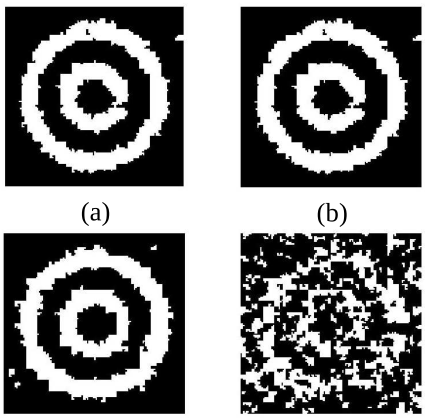
Table I

$p(x_1, x_2)$ EM Estimates of the two models (HMC-IN and HMC-LMN) for both noises (IN and LMN). The "real" value, it is to say the value estimated from the class image, is $\begin{pmatrix} 0.66 & 0.02 \\ 0.02 & 0.30 \end{pmatrix}$.

		HMC-IN		HMC-LMN	
		ω_1	ω_2	ω_1	ω_2
IN	σ^2	3.93	0.97	3.93	0.97
	α	-	-	6.31	5.09
LMN	σ^2	1.25	0.33	1.02	0.70
	α	-	-	0.97	0.32

Table II

σ^2 and α (EM Estimates) of the two models (HMC-IN and HMC-LMN) for both noises (IN and LMN)



(a) (b)

(c) (d)

Figure 3

First line: MPM restoration results for the independent noise considering (a) the HMC-IN model (error ratio=5.3%) and (b) the HMC-LMN model (error ratio=5.3%). Second line: MPM restoration result considering (c) the HMC-IN model (error ratio=27.6%) and (d) the HMC-LMN model (error ratio=6.5%).

4. CONCLUSIONS

The aim of this paper was to show, via experiments, that the recently introduced Pairwise and Triplet "Partially" Markov chains enable one to deal with long-memory noise hidden Markov chains. In fact, estimating all the parameters, by the classical "Expectation-Maximization" (EM) method, we gave two series of results, showing that these recent models based supervised and unsupervised segmentation methods can significantly improve the classical hidden Markov chains based ones. Although we have not presented any results concerning Triplet "Partially" Markov chains, passing from pairwise models to triplet one is quite straightforward and does not pose any particular problem.

5. REFERENCES

[1] J. Beran and M. S. Taqqu, Statistics for Long-Memory processes. *Monographs on Statistics and Applied Probability*, Chapman and Hall, New York, 1994.

[2] O. Cappé, E. Moulines, J. C. Pesquet, A. Petropulu, and X. Yang, Longe-range dependence and heavy-tail modeling for teletraffic data, *IEEE Signal Processing Magazine*, Vol. 19, No. 3, pp. 14-27, 2002.

[3] F. Chapeau-Blondeau, M. Guglielmi; Modèles de signaux à longue dépendance statistique, in *Signaux Aléatoires : Modélisation, Estimation, Détection*. Chap. 4, M. Guglielmi, éd., Traité IC2, Hermes (Paris) 2004.

[4] S. Derrode and W. Pieczynski, Signal and Image Segmentation using Pairwise Markov Chains, *IEEE Trans. on Signal Processing*, Vol. 52, No. 9, pp. 2477-2489, 2004.

[5] N. Giordana and W. Pieczynski, Estimation of Generalized Multisensor Hidden Markov Chains and Unsupervised Image Segmentation, *IEEE Trans. on PAMI*, Vol. 19, No. 5, pp. 465-475, 1997.

[6] P. Lanchantin and W. Pieczynski, Unsupervised non stationary image segmentation using triplet Markov chains, *Advanced Concepts for Intelligent Vision Systems (ACVIS 04)*, Aug. 31-Sept. 3, Brussels, Belgium, 2004.

[7] W. Pieczynski, C. Hulard, and T. Veit, Triplet Markov Chains in hidden signal restoration, *SPIE's International Symposium on Remote Sensing*, September 22-27, Crete, Greece, 2002.

[8] W. Pieczynski, Pairwise Markov chains, *IEEE Trans. on PAMI*, Vol. 25, No. 5, pp. 634-639, 2003.

[9] W. Pieczynski, Triplet Partially Markov Chains and Tree, *2nd International Symposium on Image/Video Communications over fixed and mobile networks (ISIVC'04)*, Brest, France, July 2004.

[10] W. Pieczynski, Copules gaussiennes dans les chaînes triplet partiellement de Markov, *Comptes Rendus de l'Académie des Sciences - Mathématiques*, to appear, 2005.

[11] W. Pieczynski and F. Desbouvries, On triplet Markov chains, International Symposium on Applied Stochastic Models and Data Analysis, (ASMDA 2005), Brest, France, May 2005.