Unsupervised Restoration of Hidden Nonstationary Markov Chains Using Evidential Priors

Pierre Lanchantin and Wojciech Pieczynski

***Abstract*—This paper addresses the problem of unsupervised Bayesian hidden Markov chain restoration. When the hidden chain is stationary, the classical "Hidden Markov Chain" (HMC) model is quite efficient, and associated unsupervised Bayesian restoration methods using the "Expectation–Maximization" (EM) algorithm work well. When the hidden chain is non stationary, on the other hand, the unsupervised restoration results using the HMC model can be poor, due to a bad match between the real and estimated models. The novelty of this paper is to offer a more appropriate model for hidden nonstationary Markov chains, via the theory of evidence. Using recent results relating to Triplet Markov Chains (TMCs), we show, via simulations, that the classical restoration results can be improved by the use of the theory of evidence and Dempster–Shafer fusion. The latter improvement is performed in an entirely unsupervised way using an original parameter estimation method. Some application examples to unsupervised image segmentation are also provided.**

***Index Terms*—Bayesian restoration, Dempster–Shafer fusion, expectation–maximization algorithm, Hidden Markov chains, nonstationary Markov chain restoration, parameter estimation, theory of evidence.**

I. Introduction

THE hidden Markov chain (HMC) model is widely used for various problems, including signal and image processing, economical prediction, and health sciences. In such a model, the unobservable—or hidden—signal $x = (x_1, \dots, x_N)$, with each x_n in a finite set $\Omega = \{1, \dots, K\}$, is assumed to be a realization of a Markov chain $X = (X_1, \dots, X_N)$. The observed signal $y = (y_1, \dots, y_N)$ is assumed to be a realization of a stochastic process $Y = (Y_1, \dots, Y_N)$. The links between X and Y are then modeled by the following joint distribution: $p(x, y) = p(x_1)p(x_2 \mid x_1) \dots p(x_N \mid x_{N-1})p(y_1 \mid x_1) \dots p(y_N \mid x_N)$. It allows one to recover the hidden data $X = x$ from the observed data $Y = y$ in different Bayesian ways, which are very efficient and widely used ([1], [2]). When $p(x, y)$ is based on unknown parameters $\theta \in \Theta$, the latter can be estimated from $Y = y$, considering the stationary HMC model, by methods like "Expectation–Maximization" (EM) [3]. Thus, when the HMC is effectively stationary, parameters are well estimated, and the restoration based on the estimated parameters, which is called "unsupervised restoration," still works well. However, when the HMC is not stationary, the estimation necessarily gives incorrect results, which can imply poor restoration of $X = x$. The aim of this paper is to describe how to improve the quality of such poor unsupervised nonstationary HMC restoration by using the theory of evidence [4]–[8]. Our approach is based on the two following points:

1) The posterior distribution $p(x \mid y)$ of an HMC, which is needed to Bayesian restoration, can be seen as the Dempster–Shafer fusion (DS fusion) of the prior Markov distribution $p(x) = p(x_1)p(x_2 \mid x_1) \dots p(x_N \mid x_{N-1})$ of X with a probability $q^y \propto p(y_1 \mid x_1) \dots p(y_N \mid x_N)$ defined on Ω^N by $Y = y$.
2) When $p(x)$ is incompletely known, it can be replaced by so-called "belief function," which is obtained from $p(x)$, whose aim is to model the lack of precise knowledge of $p(x)$. It can be fused with q^y using DS fusion. The result of the latter fusion is a probability distribution on Ω^N, and although it is not necessarily a Markov distribution, it can be used to perform Bayesian restorations. Indeed, the latter feasibility is due to the fact that the fused distribution is a triplet Markov chain [9], [10].

We provide different simulation studies, showing that such an introduction of an appropriate belief function can improve the results obtained in the classical unsupervised Bayesian restoration. The important point is that the latter appropriate belief function is found in an entirely unsupervised way using an original parameter estimation method.

Let us mention some existing methods of dealing with the nonstationarity of $p(x)$, which call on the introduction of the "time duration function" [11]. Parameters can then be estimated by a Monte Carlo Markov Chain (MCMC) method, as proposed in [12]. Then, further generalizations to hidden semi-Markov chains have been proposed (see [13] with the related references). Therefore, the method we propose, which is based on the theory of evidence, is a quite different alternative one.

The paper is organized in the following way: The next section is devoted to a brief description of the HMC model. In Section III, we introduce the main theory of evidence tools used in this paper. In this section, we also deal with a very simple case of DS fusion, outside any Markov model, to show how DS fusion can improve the Bayesian classification based on incorrect parameters. Pairwise and Triplet Markov chains are described in Section IV. The hidden "evidential" Markov chain (HEMC) model is addressed in the Section V, and different simulation results are described in Section VI. Finally, Section VII contains our conclusions and perspectives.

Manuscript received December 10, 2003; revised September 15, 2004. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. G. Tong Zhou.

The authors are with GET/INT Département CITI, UMR-CNRS 5157, Institut National des Télécommunications, 91011 Evry, France (e-mail: Pierre.Lanchantin@int-evry.fr; Wojciech.Pieczynski@int-evry.fr).

Digital Object Identifier 10.1109/TSP.2005.851131

II. Unsupervised Restoration of Stationary HMC With Independent Noise

Let $X = (X_1, \ldots, X_N)$ and $Y = (Y_1, \ldots, Y_N)$ be two stochastic processes. X is hidden (each X_n takes its values in a finite set $\Omega = \{1, \ldots, K\}$), and Y is observed (each Y_n takes its values in the set of real numbers $\mathbb{R}$). The problem is then to estimate $X = x$ from $Y = y$. The distribution of (X, Y) is assumed to be given by

$$p(x,y) = p(x_1)p(x_2 \mid x_1) \ldots p(x_N \mid x_{N-1}) \times p(y_1 \mid x_1)p(y_2 \mid x_2) \ldots p(y_N \mid x_N). \quad (1)$$

Such a model is called a "hidden" Markov chain with "independent noise" (because the hidden process X is a Markov chain, and the random variables $(Y_1, \ldots, Y_N)$ are independent conditionally on X). Hereafter, it will be denoted by HMC-IN.

Let us consider, for each $1 \leq n \leq N$, the "forward" probability $\alpha(x_n) = p(x_n, y_1, \ldots, y_n)$ and the "backward" probability $\beta(x_n) = p(y_{n+1}, \ldots, y_N \mid x_n)$. These probabilities can then be recursively calculated for $1 \leq n \leq N-1$ by

$$\left| \begin{array}{l} \alpha(x_1) = p(x_1)p(y_1 \mid x_1) \\ \alpha(x_{n+1}) = \sum_{x_n \in \Omega} \alpha(x_n) p(x_{n+1} \mid x_n) p(y_{n+1} \mid x_{n+1}) \\ \beta(x_N) = 1 \\ \beta(x_n) = \sum_{x_{n+1} \in \Omega} \beta(x_{n+1}) p(x_{n+1} \mid x_n) p(y_{n+1} \mid x_{n+1}) \end{array} \right. \quad (2)$$

Further, one can show that for each $1 \leq n \leq N$, the marginal posterior distribution of the hidden state can be calculated by

$$\gamma(x_n) = p(x_n \mid y_1, \ldots, y_N) \propto \alpha(x_n)\beta(x_n) \quad (3)$$

and the joint posterior distribution of two successives hidden states by

$$\xi(x_n, x_{n+1}) = p(x_n, x_{n+1} \mid y_1, \ldots, y_N) \propto \alpha(x_n) p(x_{n+1} \mid x_n) \times p(y_{n+1} \mid x_{n+1}) \beta(x_{n+1}). \quad (4)$$

Therefore, one can easily calculate $\gamma(x_n)$, which makes the use of Bayesian Maximum Posterior Mode (MPM) restoration $\hat{s}_{\mathrm{MPM}}(y_1, \ldots, y_N) = (\hat{x}_1, \ldots, \hat{x}_N)$ possible, and is given by $\hat{x}_n = \arg\max_{x_n \in \Omega} \gamma(x_n)$. Another Bayesian restoration method is the maximum *a posteriori* (MAP), which is sometimes also referred to as the "Viterbi algorithm" [2]. The calculus of MAP is also possible, and it is widely used.

Let us consider the case of stationary and Gaussian HMC-IN model. Thus, $p(x_n, x_{n+1})$ do not depend on $1 \leq n \leq N-1$, just as $p(y_n \mid x_n)$ and the K distributions $p(y_n \mid x_n = 1), \ldots, p(y_n \mid x_n = K)$ on $\mathbb{R}$ are Gaussian. Finally, the distribution $p(x,y)$ verifying (1) is defined by K^2 parameters $p_{ij} = p(x_1 = i, x_2 = j)$, which is a probability on Ω^2, K means $\mu_1, \ldots, \mu_K$, and K variances $\sigma_1^2, \ldots, \sigma_K^2$ of the K Gaussian densities above. The estimation of all these parameters with the iterative EM method runs as follows.

i) Take an initial value $\theta^0 = (p_{ij}^0, \mu_k^0, (\sigma_k^0)^2), 1 \leq i,j \leq K, 1 \leq k \leq K$.

ii) For each $q \geq 1$, θ^{q+1} is calculated from θ^q and y in two steps:
   - Step E) Use θ^q and y to calculate $\alpha^q(x_n)$and $\beta^q(x_n)$; then, deduce $\gamma^q(x_n)$ and $\xi^q(x_n, x_{n+1})$.
   - Step M) Calculate $\theta^{q+1} = (p_{ij}^{q+1}, \mu_k^{q+1}, (\sigma_k^{q+1})^2)$ with

$$p_{ij}^{q+1} = \frac{1}{N-1} \sum_{n=1}^{N-1} \xi^q(x_n = i, x_{n+1} = j) \quad (5)$$

$$\mu_k^{q+1} = \frac{1}{N} \sum_{n=1}^{N} \gamma^q(x_n = k) y_n \quad (6)$$

$$\left(\sigma_k^{q+1}\right) = \frac{1}{N} \sum_{n=1}^{N} \gamma^q(x_n = k) \left(y_n - \mu_k^{q+1}\right)^2. \quad (7)$$

As mentioned in the Introduction, associating EM with MPM or MAP in the stationary HMC-IN model provides a very efficient unsupervised restoration method, which is at the origin of the very success of the HMC-IN model.

Now, let us consider a nonstationary and Gaussian HMC-IN (X, Y) in which $p(y_1 \mid x_1), \ldots, p(y_N \mid x_N)$ remain Gaussian and equal, but $p(x_n, x_{n+1})$ do depend on $1 \leq n \leq N-1$. Therefore, the EM estimates above give an unique probability on Ω^2, which will necessarily vary, possibly in a significant manner, from the different $p(x_n, x_{n+1})$. In other words, the estimate gives a stationary Markov distribution $p^\star(x) = p^\star(x_1)p^\star(x_2 \mid x_1) \ldots p^\star(x_N \mid x_{N-1})$, which is necessarily different from the real nonstationary distribution $p(x) = p(x_1)p(x_2 \mid x_1) \ldots p(x_N \mid x_{N-1})$. We are going to show how to improve the results of the MPM based on $p^\star(x)$ by using the theory of evidence.

III. Dempster–Shafer Theory of Evidence

This section is devoted to the basic notions of the theory of evidence that will be needed to the purpose of this paper. Let us consider a finite set of classes, which is called, in the theory of evidence context, "the frame of discernment" $\Omega = \{\omega_1, \ldots, \omega_K\}$, which is made up of K exclusives hypotheses. Further, let us consider its power set $P(\Omega) = \{A_1, \ldots, A_{2^K}\}$, which represents the set of all subsets of Ω.

By considering single hypotheses and compound hypotheses, as well, the Dempster–Shafer theory of evidence provides a way to represent missing information or ignorance using the so-called "mass function" m. The latter is a function from $P(\Omega)$ to $\mathbb{R}^+$, verifying (8).

$$\begin{cases} m(\emptyset) = 0 \\ \sum_{A \in P(\Omega)} m(A) = 1. \end{cases} \quad (8)$$

We see that when m is null outside singletons, it can be assimilated to a probability, and thus, it appears as a kind of "generalization" of the latter. The mass functions verifying (8) are the only ones we will use in this paper; however, let us mention two other possible representations of the uncertainty linked with the mass functions: the degree of plausibility (Pls) given by Pls $(A) = \sum_{A \cap B \neq \emptyset} m(B)$ and the degree of belief (Bel), given by Bel $(A) = \sum_{B \subset A} m(B)$. In fact, Pls, and Bel can be defined in an axiomatic way, and each of m, Pls, and Bel defines the two others. A probability function P illustrates a particular case in which Pls = Bel.

Example 3.1: Let us consider a family of possible probability distributions $(p_\theta)_{\theta\in\Theta}$ on $\Omega = \{\omega_1, \omega_2\}$ and the following "lower" probability $\breve{p}(\omega_n) = \inf_{\theta\in\Theta} p_\theta(\omega_n)$. Then, m defined by $m(\{\omega_1\}) = \breve{p}(\omega_1)$, $m(\{\omega_2\}) = \breve{p}(\omega_2)$, and $m(\{\omega_1, \omega_2\}) = 1 - \breve{p}(\omega_1) - \breve{p}(\omega_2)$ is a mass function, where $m(\{\omega_1, \omega_2\})$ can be interpreted as a quantity that models the lack of information about the exact probability on $\Omega = \{\omega_1, \omega_2\}$.

Example 3.2: Let us consider the problem of satellite or airborne optical image segmentation into two classes: $\Omega = \{\omega_1, \omega_2\}$ "forest" and "water". The prior knowledge is modeled by m_1, which simply is a probability on Ω. For an observation y_n, which is a real number, on a pixel n, we have three possibilities: "forest," "water," or "clouds." The possible presence of clouds can then be modeled by a probability measure m_2 on $P(\Omega) = \{\emptyset, \omega_1, \omega_2, \Omega\}$, which is a mass function defined by $m_2(\omega_1) \propto p(y_n|\omega_1)$, $m_2(\omega_2) \propto p(y_n|\omega_2)$, and $m_2(\Omega) \propto p(y_n|\Omega)$. Therefore, $m_2(\Omega)$ models the ignorance attached with the fact that one cannot see through clouds. The DS fusion (see (9) below) of m_1 with m_2 then gives a probability on Ω, which generalizes the posterior probability (this is found again when the clouds disappear) and can be used to perform some Bayesian classification. An example of such a situation with visual presentation is dealt with in [14].

Having defined the evidence functions m_1 and m_2 of two information sources on the same frame of discernment $\Omega = \{\omega_1, \dots, \omega_K\}$, one can combine them using the Dempster–Shafer fusion (DS fusion):

$$m(A) = (m_1 \oplus m_2)(A) \propto \sum_{B_1 \cap B_2} m_1(B_1) m_2(B_2). \quad (9)$$

We will now establish a link between the previous section and the current one. We will call a mass function m "probabilistic" when it, being null outside singletons, defines a probability, and we will call it an "evidential" mass function when it is not a probabilistic one. Then, we have the following classical result.

Proposition 3.1: The result of DS fusion of a probabilistic mass function with an evidential mass function is a probabilistic one.

With an HMC-IN defined by (1), let us consider the Markov distribution $p(x) = p(x_1)p(x_2 \mid x_1) \dots p(x_N \mid x_{N-1})$ as being a probabilistic mass function m_1, and let the second probabilistic mass function m_2 be defined by $m_2(x) \propto p(y \mid x) = p(y_1 \mid x_1) \dots p(y_N \mid x_N)$ (the mass function m_2 depends on y, which is fixed). Therefore, m_1 models the prior information, and m_2 models the information provided by the observation y. The interesting result is that the posterior distribution $p(x \mid y)$ of X is the DS fusion of m_1 and m_2:

$$(m_1 \oplus m_2)(x) = \frac{p(x)p(y \mid x)}{\sum_{x' \in \Omega^N} p(x')p(y \mid x')} = p(x \mid y). \quad (10)$$

The aim of this paper is, as mentioned earlier, to improve the unsupervised restoration of a nonstationary HMC-IN by substituting the estimated stationary prior distribution of the hidden Markov chain by some evidential mass function in an unsupervised manner. To do so, we are going to briefly present the pairwise and triplet Markov chains. In fact, when m_1 is an evidential mass function, the fusion (10) may no longer be a Markov chain, and if it is not, the standard Bayesian restoration methods cannot be applied directly. Before we start, let us remark that we do not present a rigorous mathematical proof that such a substitution must improve the Bayesian restoration of hidden Markov chains; however, some calculus given in Example 3.3 is possible in a case of simple independent variables.

Example 3.3: Let us consider $\Omega = \{\omega_1, \omega_2\}$, and a sequence $Z_1 = (X_1, Y_1), \dots, Z_N = (X_N, Y_N)$ of random variables, with each X_n taking its values in Ω and each Y_n taking its values in $[0, 1]$. Let us assume that $p_n = P(X_n = \omega_1)$ depends on n, but the two densities $f_1(y) = 2(1-y), f_2(y) = 2y$ of the distributions $p(y_n \mid x_n = \omega_1)$ and $p(y_n \mid x_n = \omega_2)$ do not depend on n. When using the true parameter p_n, the Bayesian restoration d corresponding to the classical "0–1" loss function is $d(y) = \omega_1$ if $y \le p_n$, and $d(y) = \omega_2$ if $y \ge p_n$, which gives the error probability $\mathrm{ERR}[p_n] = p_n(1-p_n)$. When using a false r instead of p_n, the error probability becomes $\mathrm{ERR}[p_n, r] = p_n(1-p_n) + (p_n - r)^2$. Finally, according to what will be done in the Markov context below, we replace the false $r = P[X_n = \omega_1], 1 - r = P[X_n = \omega_2]$ with a "weakened" mass function $m(\{\omega_1\}) = r - t, m(\{\omega_2\}) = 1 - r - t, m(\{\omega_1, \omega_2\}) = 2t$. The DS fusion $m \oplus q$, where $q(\omega_1) \propto f_1(y)$ and $q(\omega_2) \propto f_2(y)$, is then a probability, which generalizes the classical posterior probability. Using the latter $m \oplus q$ probability to perform the restoration (that is to say, putting $d'(y) = \omega_1$ if $(m \oplus q)(\omega_1) \ge (m \oplus q)(\omega_2)$, and $d'(y) = \omega_2$ if $(m \oplus q)(\omega_1) \le (m \oplus q)(\omega_2)$) gives $d'(y) = \omega_1$ if $y \ge (r+t)/(1+2t)$, and $d'(y) = \omega_2$ if $y \le (r+t)/(1+2t)$, which leads to the error probability $\mathrm{ERR}[p_n, r, t] = p_n(1-p_n) + (p_n - (r+t)/(1+2t))^2$. Therefore, the problem is to know whether a $t > 0$ does exist in such a way that $\mathrm{ERR}[p_n, r, t] < \mathrm{ERR}[p_n, r]$; in other words, is it possible to decrease the error probability by introducing a mass function m when having a false r? The response is positive in the following context. As $p_1, \dots, p_N$ are not known and can vary with N, let us assume that they do vary with N and that they are realizations of a random variable W, with $E[W] = 1/2$. The problem is then to see whether the expectation of $\mathrm{ERR}[W, r, t]$ decreases when using the mass function m instead of r or, in other words, when t starts from 0. A classical calculus leads to $(d/dt)E[\mathrm{ERR}(W, r, t)](0) = -(1-2r)^2$, which shows that the "mean" error decreases, and thus, for an N large enough, the error also decreases when using m instead of r.

IV. Pairwise and Triplet Markov Chains

A. *Pairwise Markov Chains*

The HMC-IN model has recently been generalized to the so-called "pairwise Markov chain" model, in which the pairwise process $Z = (Z_1, \dots, Z_N)$, with $Z_1 = (X_1, Y_1), \dots, Z_N = (X_N, Y_N)$ is directly assumed to be a Markov chain [15]. Its distribution is then written

$$p(z) = p(z_1)p(z_2 \mid z_1) \dots p(z_N \mid z_{N-1}). \quad (11)$$

One can then see that HMC-IN are PMC (with $p(z_1) = p(x_1)p(y_1 \mid x_1)$ and $p(x_{n+1}, y_{n+1} \mid x_n, y_n) = p(x_{n+1} \mid x_n)p(y_{n+1} \mid x_{n+1})$), but, as shown in [15], PMC are not necessarily HMC-IN. More precisely, we have three models: HMC-IN, HMC in which both Z and X are Markov chains, and PMC. Then, HMC-IN are HMC and HMC are PMC, but the converse propositions do not hold. Further, in the stationary PMC, we have the following result [10].

Proposition 4.1: Let Z be a PMC verifying the following.

a) $p(z_n, z_{n+1})$ does not depend on $1 \leq n \leq N-1$.
b) $p(z_n, z_{n+1}) = p(z_{n+1}, z_n)$.

Then, the three following conditions hold.

i) Z is an HMC.
ii) For each $2 \leq n \leq N, p(y_n \mid x_n, x_{n-1}) = p(y_n \mid x_n)$.
iii) For each $1 \leq n \leq N, p(y_n \mid x) = p(y_n \mid x_n)$ are equivalent.

These results provide pleasant intuitive ideas about the respective generalities of the three models. In particular, given that $p(x_{n+1}, y_{n+1} \mid x_n, y_n) = p(x_{n+1} \mid x_n, y_n)p(y_{n+1} \mid x_{n+1}, x_n, y_n)$, the PMC Z is an HMC if and only if $p(x_{n+1} \mid x_n, y_n) = p(x_{n+1} \mid x_n)$, and it is a HMC-IN if and only if it is an HMC with $p(y_{n+1} \mid x_{n+1}, x_n, y_n) = p(y_{n+1} \mid x_{n+1})$.

However, like HMC-IN, PMC can be used to estimate $X = x$ from $Y = y$ by different Bayesian methods in unsupervised manner, and the different results shown in [16] are quite encouraging. In fact, considering the same "forward" probability, $\alpha(x_n) = p(x_n, y_1, \ldots, y_n)$, and the following "backward" probability $\beta(x_n) = p(y_{n+1}, \ldots, y_N \mid x_n, y_n)$ (which generalize to the PMC model the classical ones valid in the HMC-IN model), we have something analogous to (2) recursions. More precisely, $\alpha(x_1) = p(x_1)p(y_1 \mid x_1)$, and $\alpha(x_{n+1}) = \sum_{x_n \in \Omega} \alpha(x_n)p(z_{n+1} \mid z_n)$ for $1 \leq n \leq N-1; \beta(x_N) = 1$, and $\beta(x_n) = \sum_{x_{n+1} \in \Omega} \beta(x_{n+1})p(z_{n+1} \mid z_n)$, for $1 \leq n \leq N-1$. Furthermore, generalizing (3) and (4), we have $p(x_n \mid y_1, \ldots, y_N) \propto \alpha(x_n)\beta(x_n)$ for each $1 \leq n \leq N$, and $p(x_n, x_{n+1} \mid y_1, \ldots, y_N) \propto \alpha(x_n)p(z_{n+1} \mid z_n)\beta(x_{n+1})$ for each $1 \leq n \leq N-1$. Therefore, as in the HMC-IN context, $p(x_n \mid y_1, \ldots, y_N)$ and $p(x_n, x_{n+1} \mid y_1, \ldots, y_N)$ are calculable, which allows one to propose different unsupervised Bayesian restoration methods.

B. Triplet Markov Chains

In triplet Markov chains (TMCs), a third process $U = (U_1, \ldots, U_N)$ is introduced, with each U_n taking its values in a finite set $\Lambda = \{\lambda_1, \ldots, \lambda_m\}$. Then, let us put $T_n = (X_n, U_n, Y_n), Z_n = (X_n, Y_n), V_n = (X_n, U_n)$, and T, Z, V as the corresponding processes. Assuming that T is Markovian, the process (V, Y) is a PMC, and we can formulate the distribution of (V_n, Y) as $p(v_n, y) \propto \alpha(v_n)\beta(v_n)$, with $\alpha(v_n)$ and $\beta(v_n)$ calculable. This means that the distributions $p(x_n, y) = \sum_{u_n \in \Lambda} p(x_n, u_n, y) = \sum_{u_n \in \Lambda} p(v_n, y)$ are also computable, giving us $p(x_n \mid y)$. Finally, although the distribution of (X, Y) is not necessarily a Markov one, the marginal distributions $p(x_n \mid y)$ are computable, which in particular enables the use of the Bayesian MPM restoration method. Of course, TMC are of interest if they generalize PMC. In the case of stationary TMC, we have the following result, whose proof is analogous to the proof of the Proposition 4.1, with U instead of Y and T instead of Z.

Proposition 4.2: Let T be a TMC verifying the following.

a) $p(t_n, t_{n+1})$ does not depend on $1 \leq n \leq N-1$.
b) $p(t_n, t_{n+1}) = p(t_{n+1}, t_n)$.

Then, we have the three following conditions.

i) Z is a Markov chain (T is a PMC).
ii) For each $2 \leq n \leq N$, $p(u_n \mid z_n, z_{n-1}) = p(u_n \mid z_n)$.
iii) For each $1 \leq n \leq N$, $p(u_n \mid z) = p(u_n \mid z_n)$ are equivalent.

V. Unsupervised Restoration Using Evidential Priors

This section is devoted to the problem mentioned at the end of Section II. We propose a specific mass function to replace the distribution $p^\star$, show that the DS fusion result is a TMC, and propose an original method of parameters estimation of the new model.

A. Hidden Evidential Markov Chains

Definition 5.1: Every mass function defined on $[P(\Omega^N)]$ and verifying the following:

i) m vanishes outside $[P(\Omega)]^N$; and
ii) m is of the "Markovian" form $m(u_1, u_2, \ldots, u_N) = m(u_1)m(u_2 \mid u_1)\ldots m(u_N \mid u_{N-1})$, with $(u_1, u_2, \ldots, u_N) \in [P(\Omega)]^N$

will be called "Evidential Markov Chain" (EMC).

We see how an EMC m generalizes Markov chain $p(x)$, which can be seen as a particular m being nonzero only on $(u_1, u_2, \ldots, u_N)$ such that each $u_1, u_2, \ldots, u_N$ is a singleton. An example of an EMC is specified in Remark 5.1 below.

The idea behind this paper, which will now be developed, is to consider a stationary EMC to replace the incorrect stationary Markov chain $p^\star(x)$ estimated with EM in the nonstationary case (see the end of Section II). The probabilistic mass function m_2 defined from the observations $(y_1, \ldots, y_N)$ by $m_2(x_1, \ldots, x_N) \propto p(y \mid x) = p(y_1 \mid x_1)\ldots p(y_N \mid x_N)$ can be fused with an EMC m_1, and we can say, according to (10), that when $m_1 = p^\star$, the fusion result is $p(x \mid y)$. The difficulty is that when m_1 is no longer probabilistic, the fusion $m_1 \oplus m_2$ is no longer a Markov chain. However, $m_1 \oplus m_2$ is a marginal distribution of a TMC and can therefore be used to restore the hidden signal. More precisely, we have the following result.

Proposition 5.1: Let m_1 be an EMC, and let $m_2(x_1, \ldots, x_N) \propto p(y \mid x) = p(y_1 \mid x_1)\ldots p(y_N \mid x_N)$ be the probabilistic mass function defined by the observations $y = (y_1, \ldots, y_N)$. Let $T = (X, U, Y)$ be a TMC, with each U_n taking its values in $\Lambda = P(\Omega)$, whose distribution is defined by

$$p(t_1, \ldots, t_N) \propto q_1(t_1, t_2)q(t_2, t_3)\ldots q_{N-1}(t_{N-1}, t_N) \quad (12)$$

$q_1(t_1, t_2) = 1_{[x_1 \in u_1]}m(u_1)p(y_1 \mid x_1)1_{[x_2 \in u_2]}m(u_2 \mid u_1) p(y_2 \mid x_2)$, and $q_n(t_n, t_{n+1}) = \mathbb{1}_{[x_{n+1} \in u_{n+1}]}m(u_{n+1} \mid u_n) p(y_{n+1} \mid x_{n+1})$ for $2 \leq n \leq N-1$. Then, $m = m_1 \oplus m_2$ is the conditional distribution $p(x \mid y)$ defined by $p(x, y)$, where $p(x, y)$ is the marginal distribution of the TMC (12). As a consequence, $p(x_n, u_n \mid y)$ are calculable, and thus, $p(x_n \mid y)$ are as well.

The proof is based on the fact that the sum of $m(t_1, \ldots, t_N)$ over $(u_1, \ldots, u_N) \in [P(\Omega)]^N$ is, on one hand, the marginal

$p(z_1, \dots, z_N)$ distribution of $p(t_1, \dots, t_N)$ and, on the other hand, the DS fusion $m = m_1 \oplus m_2$ defined by (9). Further, $p(t_1, \dots, t_N)$ defined with (12) is necessarily a Markov chain.

Let us note that the transitions $p(t_{n+1} \mid t_n)$ could be calculated from $q_1(t_1, t_2) \dots q_{N-1}(t_{N-1}, t_N)$ by "backward" recursions: Putting $\beta'(t_N) = 1$ and $\beta'(t_n) = \int q_n(t_n, t_{n+1}) \beta'(t_{n+1}) dt_{n+1}$, for $1 \leq n \leq N-1$, we have $p(t_{n+1} \mid t_n) = \beta'(t_{n+1})/\beta'(t_n)$. However, it is important for practical applications to calculate the transitions $p(v_{n+1} \mid v_n, y)$, which are calculated in the strictly same manner, considering y as a constant in (12). Therefore, again, we find that with $m = m_1 \oplus m_2$ being the marginal distribution of a TMC, it is possible, according to the previous section, to estimate the hidden signal $X = x$ from the observations $Y = y$.

Definition 5.2: The model given by an EMC m_1 and a probabilistic mass function m_2 given from the observations $(y_1, \dots, y_N)$ by $m_2(x_1, \dots, x_N) \propto p(y \mid x) = p(y_1 \mid x_1) \dots p(y_N \mid x_N)$ will be called "hidden evidential Markov chain with independent noise" (HEMC-IN).

B. Learning HEMC

We propose, in this subsection, an original parameter estimation method, derived from the EM algorithm, to estimate the parameters of HEMC-IN from the only observations $(y_1, \dots, y_N)$. We consider a stationary HEMC-IN $T = (V, Y)$, with $V = (X, U)$, which means that $p(t_n, t_{n+1})$ does not depend on n. According to (12), we notice that when considering that $V = (X, U)$ is the hidden process in $T = (V, Y)$, the latter is a classical HMC-IN. Although this is a particular HMC-IN in the sense that $p(y_n \mid x_n, u_n) = p(y_n \mid x_n)$, we may view the use of the EM method, which is widely used in the HMC-IN case.

We will now look at the Gaussian case, where $p(y_n \mid x_n)$ are Gaussian. Considering K classes $\Omega = \{\omega_1, \dots, \omega_K\}$, we have to estimate, according to (12), the following parameters: K means $\mu_1, \dots, \mu_K$, and K variances $\sigma_1^2, \dots, \sigma_K^2$ of the K Gaussian densities $p(y_n \mid x_n = 1), \dots, p(y_n \mid x_n = K)$, and $(2^K - 1) \times (2^K - 1)$ parameters $m_{ij} = m(u_1 = i, u_2 = j)$, which is a mass function on $[P(\Omega)]^2$. The EM method adapted to HEMC-IN runs as follows.

i) Consider an initial value $\theta^0 = (m_{ij}^0, \mu_k^0, (\sigma_k^0)^2)$ for $1 \leq i, j \leq 2^K - 1$, $1 \leq k \leq K$.

ii) For each $q \in \mathbb{N}^*$, calculate θ^{q+1} from $(y_1, \dots, y_N)$ and θ^q in two steps.

- Step E: Calculate $\alpha^q(v_n)$, and $\beta^q(v_n)$, and then $\gamma^q(v_n)$ and $\xi^q(v_n, v_{n+1})$.
- Step M: Calculate $\theta^{q+1} = (m_{ij}^{q+1}, \mu_k^{q+1}, (\sigma_k^{q+1})^2)$ with

$$m_{ij}^{q+1} = \frac{1}{A} \sum_{n=1}^{N-1} \sum_{(x_n, x_{n+1}) \in \Omega^2} \xi^q(x_n, u_n = i, x_{n+1}, u_{n+1} = j) \tag{13}$$

$$\mu_k^{q+1} = \frac{1}{N} \sum_{n=1}^{N} \sum_{u_n \in P(\Omega)} \gamma^q(x_n = k, u_n) y_n \tag{14}$$

$$\left(\sigma_k^{q+1}\right)^2 = \frac{1}{N} \sum_{n=1}^{N} \sum_{u_n \in P(\Omega)} \gamma^q(x_n = k, u_n) \left(y_n - \mu_k^{q+1}\right)^2 \tag{15}$$

with $A = (N-1)\#i\,\#j$, where $\#i$ is the cardinal of i. Notice that when the HEMC-IN (12) becomes a classical HMC-IN, that is to say when $m_{ij} = p(u_1 = i, u_2 = j)$ are null, except for i and j singletons, the formulas (13)–(15) give the formulas of the classical EM.

C. Unsupervised Restoration of Poorly Modeled Hidden Signal

Finally, when the distribution of the hidden signal X is poorly known and when the noise is Gaussian, we propose the following method to recover it from the observed signal $Y = y$.

i) Model the prior knowledge on X by an EMC.

ii) Estimate the parameters by the EM method above.

iii) Use the Bayesian MPM method given by $\hat{s}_{\mathrm{MPM}}(y_1, \dots, y_N) = (\hat{x}_1, \dots, \hat{x}_N)$ with $\hat{x}_n = \arg\max_{x_n \in \Omega} (\sum_{u_n \in P(\Omega)} \gamma(v_n))$.

Some numerical results of different experiments we performed are presented in the next section.

Remark 5.1: We considered the case of HMC-IN, but more general HMC, in which the random variables $Y_1, \dots, Y_N$ that are not independent conditionally on X can still be viewed. In fact, every PMC can be written

$$\begin{aligned} p(z) &= \frac{p(z_1, z_2) \dots p(z_{N-1}, z_N)}{p(z_1) \dots p(z_{N-1})} \\ &= \underbrace{\frac{p(x_1, x_2) \dots p(x_{N-1}, x_N)}{p(x_1) \dots p(x_{N-1})}}_{a(x)} \\ &\quad \times \underbrace{\frac{p(y_1, y_2 \mid x_1, x_2) \dots p(y_{N-1}, y_N \mid x_{N-1}, x_N)}{p(y_2 \mid z_2) \dots p(y_{N-1} \mid z_{N-1})}}_{b(x,y)} \end{aligned} \tag{16}$$

and is an HMC if and only if $a(x)$ is the distribution of X [15]. In the latter case, we can take $m_2(x_1, \dots x_N) \propto p(y \mid x) = b(x, y)$ instead of $m_2(x_1, \dots, x_N) \propto p(y \mid x) = p(y_1 \mid x_1) \dots p(y_N \mid x_N)$, and an analogous result to the one specified in Proposition 5.1 holds. However, since m_2 is no longer a simple product but a Markov chain, the parameter estimation method should be adapted. Note that by applying this model to the situation of Example 3.2, considered with correlated noise, we obtain a simple example of an EMC defined on $[P(\Omega)]^N$ by $b(x, y)$ (with the class "clouds" assimilated to $\Omega = \{\omega_1, \omega_2\}$).

Remark 5.2: Let us note that there are different kinds of nonstationarities in HMC. In fact, they can be due to the nonstationarity of $p(x)$ to the nonstationarity of $p(y \mid x)$, or both of them. Therefore, this paper deals with models with nonstationary $p(x)$; however, the nonstationarity of $p(y \mid x)$ could be treated in analogous manner. To do so, the probability $m_2(x_1, \dots, x_N) \propto p(y \mid x) = p(y_1 \mid x_1) \dots p(y_N \mid x_N)$ would be replaced by a mass function $m_2(u_1, \dots, u_N) = m_2(u_1) \dots m_2(u_N)$, in which $m_2(u_n)$ is obtained in some way, preferably in an unsupervised way, from the probability $q(x_n) \propto p(y_n \mid x_n)$. Another interesting way of dealing with the nonstationarity of $p(y \mid x)$, which is proposed and successfully applied in speech processing in [17], [18], is by replacing the time-varying mean of the Gaussian $p(y_n \mid x_n = \omega_j)$ by a mean of the form $\sum_{m=0}^{M} B_j(m) f_m(n-\tau_j)$, where the functions $f_0, \dots, f_M$ are given, and the time-shift parameter τ_j is such that $n-\tau_j$ represents the sojourn time in state ω_j (therefore, $x_n = \cdots = x_{n-\tau_j} = \omega_j$, and $x_{n-\tau_j-1} \neq \omega_j$). Thus, we may say that in such models, the nonstationarity of $p(y \mid x)$ is managed by increasing the order of Markovianity; in fact, we see that $p(y \mid x_1, \dots, x_n) = p(y_n \mid x_{n'}, \dots, x_n)$ for some $n' < n$. As in the present paper, the authors propose the all parameter estimation method and the resulting unsupervised restoration.

VI. Unsupervised Restoration of Nonstationary Signals Using HEMC-IN

This section is dedicated to some applications of the unsupervised restoration method above the following problem. (X, Y) is a classical Gaussian HMC-IN, where $p(y_n \mid x_n)$ is independent of n. The difficulty is that the hidden signal X, which is a Markov chain, can be strongly nonstationary, and its distribution is not known. The problem is then to estimate X from $Y = y$. The classical unsupervised MPM restoration method would consist of considering that (X, Y) is a stationary HMC-IN, estimating the parameters with the classical EM, and applying MPM based on the estimates obtained. We compare the results so obtained with the results obtained with the MPM based on an HEMC-IN, which is also in an unsupervised manner, as described in the previous section. We also show the existence of different situations in which the new method proposed in this paper significantly improves the results obtained with the classical one.

We present two series of experiments. In the first one, nonstationary simulated HMC-IN are considered, and different unsupervised restoration results with both classical and new methods are given. In the second one, we consider the problem of image segmentation. A nonstationary noisy class image is segmented by both evidential and classical methods, where the bidimensional set of pixels are previously transformed into a mono-dimensional set via the Hilbert–Peano scan, as previously used in [19].

A. Hidden Nonstationary Markov Chains

We consider an HMC-IN (X, Y), verifying (1), with $\Omega = \{1, 2\}$, $n = 1024$ and the two following matrices

$$M_1 = \begin{bmatrix} 0.98 & 0.02 \\ 0.02 & 0.98 \end{bmatrix}, \quad M_2 = \begin{bmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{bmatrix}.$$

The Markov chain $X = (X_1, \dots, X_N)$ is nonstationary in the following way. Given the two transition matrices M_1 and M_2, for $s = 1, 2, \dots, X^i = (X_{(i-1)s+1}, X_{(i-1)s+2}, \dots, X_{is})$, the nonstationary chain $X = (X_1, \dots, X_N)$ then verifies the following.

1) The distribution of X_1 is (0.5,0.5).
2) M_1 is the transition matrix in $X^2, X^4, \dots$.
3) M_2 is the transition matrix in $X^3, X^5, \dots$.

A realization $X = x$ is simulated, and $Y = y$ is sampled according to $p(y_1 \mid x_1), \dots, p(y_N \mid x_N)$, where $p(y_n \mid x_n = 1)$ is Gaussian with mean 0 and variance 1, and $p(y_n \mid x_n = 2)$ is Gaussian with mean 2 and variance 1. The realization $X = x$ is then estimated by the Bayesian MPM method from $Y = y$ in three different ways.

1) The first restoration is obtained using the real parameters of the hidden nonstationary chain. This is a reference one, and according to the general theory, the error rate $\tau_{\text{min}}^{\text{MPM}}$, which is the rate of wrongly classified observations, is minimal.
2) The second restoration is obtained using the parameters estimated with the EM algorithm, considering that (X, Y) is an HMC-IN (the hidden chain is assumed stationary). The error rate is denoted by $\tau_{\text{HMC-EM}}^{\text{MPM}}$.
3) The third restoration is obtained using the parameters estimated with the EM algorithm, considering that (X, Y) is an HEMC-IN. The error rate is denoted by $\tau_{\text{HEMC-EM}}^{\text{MPM}}$ (the HEMC-IN-EM is initialized by the results of HMC-IN-EM).

TABLE I
Error Ratios Corresponding to M_1, M_2 and Different s

s	τ_{min}^{MPM}	τ_{HMC-EM}^{MPM}	$\tau_{HEMC-EM}^{MPM}$
2	7.8	11.0	11.0
4	7.8	11.3	11.1
8	8.0	12.2	11.3
16	8.1	12.7	10.8
32	8.3	14.7	10.4
64	8.7	14.7	9.9
128	8.4	15.4	9.4
256	8.2	16.5	8.7
512	8.2	17.5	8.7

TABLE II
Error Ratios Corresponding to M_1, M_3 and Different s

s	τ_{min}^{MPM}	τ_{HMC-EM}^{MPM}	$\tau_{HEMC-EM}^{MPM}$
2	8.1	13.3	13.3
4	8.0	13.3	12.7
8	8.0	14.3	11.8
16	8.8	16.9	11.3
32	8.5	19.1	11.1
64	8.6	20.0	9.7
128	8.7	22.0	9.2
256	8.7	22.9	9.0
512	8.6	22.8	8.8

We present two series of results: The first one, corresponding to M_2, is given in Table I, and the second one, corresponding to the same M_1 and $M_3 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}$ instead of M_2, is given in Table II. According to these results, we see that $\tau_{\text{HEMC-EM}}^{\text{MPM}}$ is always inferior to $\tau_{\text{HMC-EM}}^{\text{MPM}}$, and for some s, the gain of efficiency can be quite striking.

B. Noisy Nonstationary Images

Our example here is a 128×128 size image containing two classes, which is presented in Fig. 1. It is then noised with Gaussian independent noise, with variance 1 and means 0 and 2,

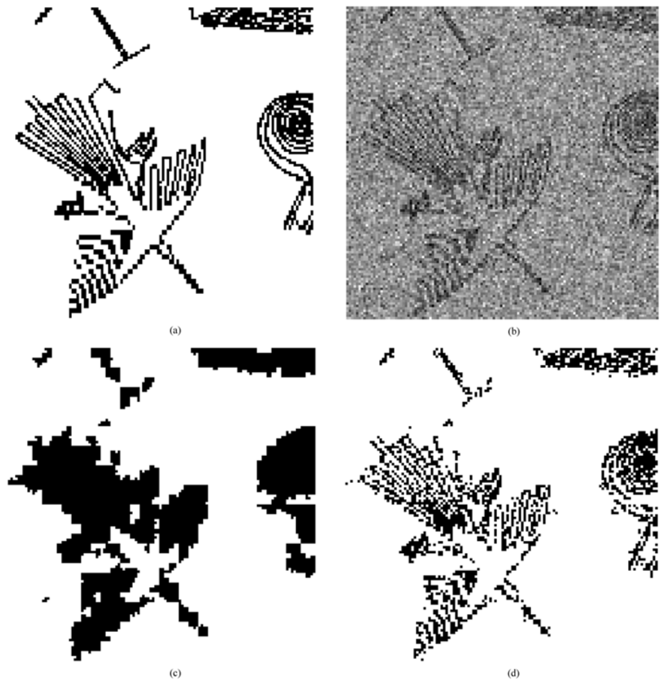

Fig. 1. Restorations using HMC-EM-MPM and HEMC-IN-EM-MPM of a two-class image noised by $\mathcal{N}(0,1)$ and $\mathcal{N}(2,1)$. (a) Initial image $X = x$. (b) Noised image $Y = y$. (c) Restoration result considering HMC-IN model: error rate $= 14.62\%$. (d) Restoration result considering the HEMC-IN model: error rate $= 5.05\%$.

TABLE III
ESTIMATES WITH EM IN HMC AND HEMC-IN CONTEXTS. REAL NOISE GAUSSIAN DENSITIES ARE $\mathcal{N}(0,1)$ AND $\mathcal{N}(2,1)$

HMC-IN			$\mathcal{N}(0,1)$	$\mathcal{N}(2,1)$
$p(\omega_1)$	0.2789	μ	-0.002	1.063
$p(\omega_2)$	0.7213	σ	1.0150	2.075
$p(x_1, x_2)$		ω_1	ω_2	
ω_1		0.2730	0.0015	
ω_2		0.0056	0.7156	
HMC-IN estimated parameters by EM				
HEMC-IN			$\mathcal{N}(0,1)$	$\mathcal{N}(2,1)$
$m(\{\omega_1\})$	0.0234	μ	-0.0065	2.0815
$m(\{\omega_2\})$	0.7119	σ	1.0128	0.9252
$m(\{\omega_1, \omega_2\})$	0.2648			
$m(u_1, u_2)$		$\{\omega_1\}$	$\{\omega_2\}$	$\{\omega_1, \omega_2\}$
$\{\omega_1\}$		0.0205	0.0000	0.0028
$\{\omega_2\}$		0.0000	0.7074	0.0045
$\{\omega_1 \omega_2\}$		0.0028	0.0045	0.2573
HEMC-IN estimated parameters by EM				

resulting in a "Noisy image" that is also presented in Fig. 1. The bidimensional set of pixels is converted, via a Hilbert–Peano scan, into a mono-dimensional set $1, 2, \ldots, N = 128 \times 128$, as described in [19]. The results presented in Fig. 1 clearly show that the unsupervised HMC-IN-EM-based MPM is unable to find details in the wings and tail of the bird, whereas the unsupervised HEMC-IN-EM-based MPM is able to find the details. The estimates of different parameters by both the HMC-IN-EM and HEMC-IN-EM methods are presented in Table III. We can see that HEMC-IN-EM is more efficient in the noise parameter estimation. An another interesting point is that HEMC-IN-EM attributes the probability 0.2573, which is rather strong, to $\{\omega_1, \omega_2\} \times \{\omega_1, \omega_2\}$, which could be seen, roughly speaking, as an inadequacy measurement of the stationary HMC-IN, which is given by HMC-IN-EM, to the data. In fact, when the hidden data suit a stationary HMC-IN, HEMC-IN-EM gives the latter model, and thus, the probabilities attributed to $\{\omega_1\} \times \{\omega_1, \omega_2\}$, $\{\omega_2\} \times \{\omega_1, \omega_2\}$, and $\{\omega_1, \omega_2\} \times \{\omega_1, \omega_2\}$ tend toward zero. Notice that outside its interest in image processing, the results of this subsection show a kind of robustness, which is quite useful when considering real data. In fact, transforming the set of pixels into a mono-dimensional set, the realizations $X = x$ obtained from different real images, like in Fig. 1, are not necessarily Markov chain realizations. Therefore, the results obtained seem to show that the whole modeling and associated unsupervised restoration method proposed in the paper presents a good statistical robustness.

VII. CONCLUSIONS AND PERSPECTIVES

In this paper, we have provided an original unsupervised method for restoring nonstationary hidden Markov chains, with potential applications to various problems. The main contribution was to tackle the lack of stationarity using the theory of evidence. More precisely, the prior nonstationary distribution of the hidden Markov chain was replaced with some particular mass function, which was found in an unsupervised manner. Bayesian restoration techniques are then rendered applicable, according to a recent result that asserts that the Dempster–Shafer fusion of such evidential priors, with the probability provided by the observations, is a triplet Markov chain. Simulations attest to the favorable behavior of the method.

As for perspectives, we may envisage extensions to more complex noise processes than the Gaussian ones considered here. The EM, or Iterative Conditional Estimation (ICE) [20] based methods considered in [19] could then be adapted to the nonstationary case considered in this paper. Other perspectives could concern the use of spatially correlated noise or still the use of fully Bayesian methods, in which one considers a prior knowledge on parameters.

APPENDIX

$$\begin{aligned} \mathrm{ERR}(p_n) &= (1-p_n)\int_0^{p_n} f_2(y)dy + p_n \int_{p_n}^1 f_1(y)dy \\ &= p_n(1-p_n) \end{aligned} \quad (17)$$

$$\begin{aligned} \mathrm{ERR}(p_n, r) &= (1-p_n)\int_0^{r} f_2(y)dy + p_n \int_{r}^1 f_1(y)dy \\ &= p_n(1-p_n) + (p_n - r)^2. \end{aligned} \quad (18)$$

Following the DS fusion rule, we have

$$\begin{aligned} (m \oplus q)(\omega_1) &\propto (m(\{\omega_1\}) + m(\{\omega_1, \omega_2\})q(\omega_1) \\ &\propto (m(\{\omega_1\}) + m(\{\omega_1, \omega_2\})f_1(y) \\ &\propto 2(r+t)(1-y) \\ (m \oplus q)(\omega_2) &\propto (m(\{\omega_2\}) + m(\{\omega_1, \omega_2\})q(\omega_2) \\ &\propto (m(\{\omega_2\}) + m(\{\omega_1, \omega_2\})f_2(y) \\ &\propto 2(1-r+t)y. \end{aligned}$$

Therefore, $(m \oplus q)(\omega_1) \geq (m \oplus q)(\omega_2)$ means $(r+t)(1-y) \geq (1-r+t)y$ or $y \leq (r+t)/(1+2t)$. The calculus of $\mathrm{ERR}[p_n, r, t]$ is analogous to the calculus of $\mathrm{ERR}[p_n, r]$, with $(r+t/1+2t)$ instead of r. According to (18), the calculus gives

$$\mathrm{ERR}[p_n, r, t] = p_n(1-p_n) + \left(p_n - \frac{r+t}{1+2t}\right)^2.$$

Assuming that

$$\frac{d}{dt}(E[\mathrm{ERR}(p_n, r, t)]) = E\left[\frac{d}{dt}(\mathrm{ERR}(p_n, r, t))\right]$$

we have

$$\frac{d}{dt}[\mathrm{ERR}(p_n, r, t)] = -2\left(p_n - \frac{r+t}{1+2t}\right)\frac{1-2r}{(1+2t)^2}$$

and thus

$$\frac{d}{dt}[\mathrm{ERR}(p_n, r, t)](0) = -2(p_n - r)(1-2r).$$

Taking the expectation gives

$$\begin{aligned} E\left[\frac{d}{dt}[\mathrm{ERR}(W, r, t)](0)\right] &= -2(E[W] - r)(1-2r) \\ &= -2\left(\frac{1}{2} - r\right)(1-2r) \\ &= -(1-2r)^2. \end{aligned}$$

REFERENCES

[1] L. E. Baum, T. Petrie, G. Soules, and N. Weiss, "A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains," *Ann, Math. Statist.*, vol. 41, pp. 164–171, 1970.

[2] G. D. Fornay, "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–277, 1973.

[3] G. McLachlan and T. Krishnan, *EM Algorithm and Extensions*. New York: Wiley, 1997.

[4] T. Denoeux, "Reasoning with imprecise belief structures," *Int. J. Approx. Reasoning*, vol. 20, no. 1, pp. 79–111, 1999.

[5] G. Shafer, *A Mathematical Theory of Evidence*. Princeton, NJ: Princeton Univ. Press, 1976.

[6] P. Smets and R. Kennes, "The transferable belief model," *Artificial Intell.*, vol. 66, no. 2, pp. 191–234, 1994.

[7] R. R. Yager, J. Kacprzyk, and M. Fedrizzi, *Advances in the Dempster–Shafer Theory of Evidence*. New York: Wiley, 1994.

[8] L. Fouque, A. Appriou, and W. Pieczynski, "An evidential Markovian model for data fusion and unsupervised image classification," in *Proc. Third Int. Conf. Inf. Fusion*, vol. 1, Paris, France, Jul. 10th–13th, 2000, pp. TuB4-25–TuB4-31.

[9] W. Pieczynski, "Chaînes de Markov triplet—Triplet Markov chains," *Comptes Rendus de l'Académie des Sciences—Mathématiques*, no. I 335, pp. 275–278, 2002.

[10] ——, "Triplet Markov chains and theory of evidence," *Int. J. Approx. Reasoning*, 2003, submitted for publication.

[11] B. Sin and J. H. Kim, "Nonstationary hidden Markov model," *Signal Process.*, vol. 46, pp. 31–46, 1995.

[12] P. M. Djuric and J.-H. Chun, "An MCMC sampling approach to estimation of nonstationary hidden Markov models," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1113–1123, May 2002.

[13] S.-Z. You and H. Kobayashi, "A hidden semi-Markov model with missing data and multiple observation sequences for mobility tracking," *Signal Process.*, vol. 83, pp. 235–250, 2003.

[14] A. Bendjebbour, Y. Delignon, L. Fouque, V. Samson, and W. Pieczynski, "Multisensor images segmentation using Dempster-Shafer fusion in Markov fields context," *IEEE Trans. Geosci. Remote Sens.*, vol. 39, no. 8, pp. 1789–1798, Aug. 2001.

[15] W. Pieczynski, "Pairwise Markov chains," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 25, no. 5, pp. 634–639, May 2003.

[16] S. Derrode and W. Pieczynski, "Signal and image segmentation using pairwise Markov chains," *IEEE Trans. Signal Process.*, vol. 52, no. 9, pp. 2477–2489, Sep. 2004.

[17] L. Deng, M. Aksmanovic, D. Sun, and C. F. J. Wu, "Speech recognition using hidden Markov models with polynomial regression functions as nonstationary states," *IEEE Trans. Speech Audio Process.*, vol. 2, no. 4, pp. 507–520, Dec. 1994.

[18] H. Sameti and L. Deng, "Nonstationary-state hidden Markov model representation of speech signals for speech enhancement," *Signal Process.*, vol. 82, pp. 205–227, 2002.

[19] N. Giordana and W. Pieczynski, "Estimation of generalized multisensor hidden Markov chains and unsupervised image segmentation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 5, pp. 465–475, May 1997.

[20] J.-P. Delmas, "An equivalence of the EM and ICE algorithm for exponential family," *IEEE Trans. Signal Process.*, vol. 45, no. 10, pp. 2613–2615, Oct. 1997.

Pierre Lanchantin was born in Paris, France, in 1976. He received the M.Sc. degree in electrical engineering from the Institut National des Télécommunications (INT), Evry, France, in 2002. Since October 2002, he has been pursuing the Ph.D. degree with the Communications, Image and Information Processing (CITI) Department, INT.

His research interests include statistical signal and image processing using hidden Markov models.

Wojciech Pieczynski received the Doctorat d'Etat degree in mathematical statistics from the Université Pierre et Marie Curie, Paris, France, in 1986.

He is currently Professor with the Communications, Image, and Information Processing (CITI) Department, Institut National des Télécommunications (INT), Evry, France. His research interests include mathematical statistics, stochastic processes, and statistical image processing.