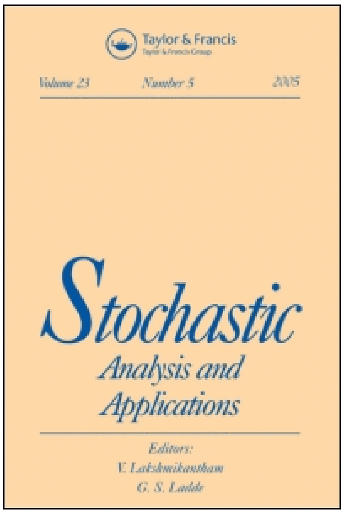

Stochastic Analysis and Applications
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713597300

State Estimation Schemes for Independent Component Coupled Hidden Markov Models

W. P. Malcolm[a]; Novi Quadrianto[b]; Lakhdar Aggoun[c]
[a] Mathematical Sciences Institute, Australian National University, Canberra, Australia [b] National Information Communication Technology Centre of Australia (NICTA), Canberra, Australia [c] Department of Mathematics and Statistics, Sultan Qaboos University, Sultanate of Oman

Online publication date: 12 April 2010

To cite this Article Malcolm, W. P. , Quadrianto, Novi and Aggoun, Lakhdar(2010) 'State Estimation Schemes for Independent Component Coupled Hidden Markov Models', Stochastic Analysis and Applications, 28: 3, 430 — 446
To link to this Article: DOI: 10.1080/07362991003708481
URL: http://dx.doi.org/10.1080/07362991003708481

Stochastic Analysis and Applications, 28: 430–446, 2010

ISSN 0736-2994 print/1532-9356 online
DOI: 10.1080/07362991003708481

State Estimation Schemes for Independent Component Coupled Hidden Markov Models

W. P. MALCOLM,[1] NOVI QUADRIANTO,[2] AND LAKHDAR AGGOUN[3]

[1]Mathematical Sciences Institute, Australian National University, Canberra, Australia
[2]National Information Communication Technology Centre of Australia (NICTA), Canberra, Australia
[3]Department of Mathematics and Statistics, Sultan Qaboos University, Sultanate of Oman

Conventional Hidden Markov models generally consist of a Markov chain observed through a linear map corrupted by additive noise. This general class of model has enjoyed a huge and diverse range of applications, for example, speech processing, biomedical signal processing and more recently quantitative finance. However, a lesser known extension of this general class of model is the so-called Factorial Hidden Markov Model (FHMM). FHMMs also have diverse applications, notably in machine learning, artificial intelligence and speech recognition [13, 17]. FHMMs extend the usual class of HMMs, by supposing the partially observed state process is a finite collection of distinct Markov chains, either statistically independent or dependent. There is also considerable current activity in applying collections of partially observed Markov chains to complex action recognition problems, see, for example, [6].

In this article we consider the Maximum Likelihood (ML) parameter estimation problem for FHMMs. Much of the extant literature concerning this problem presents parameter estimation schemes based on full data log-likelihood EM algorithms. This approach can be slow to converge and often imposes heavy demands on computer memory. The latter point is particularly relevant for the class of FHMMs where state space dimensions are relatively large.

The contribution in this article is to develop new recursive formulae for a filter-based EM algorithm that can be implemented online. Our new formulae are equivalent ML estimators, however, these formulae are purely recursive and so, significantly reduce numerical complexity and memory requirements. A computer simulation is included to demonstrate the performance of our results.

Received August 7, 2008; Accepted July 17, 2009

W. P. M. and N. Q. would like to gratefully acknowledge Professor Alex Smola for supporting this work. W. P. M. would also like to thank Zineb Aggoun for many helpful technical discussions.

Address correspondence to W. P. Malcolm, Mathematical Sciences Institute, Australian National University, ACT 0200 Canberra, Australia; E-mail: paul.malcolm@anu.edu.au

Keywords Change of measure techniques; EM algorithms; Factorial hidden Markov models; Filtering; Martingales; M-ary detection; Parameter estimation; Smoothing.

Mathematics Subject Classification 60G35; 60J22.

1. Introduction

Hidden Markov models (HMMs) have been extensively researched and used for the past several decades [10, 18]. However, a common feature of much of the HMM literature is its emphasis on models with a single partially observed Markov chain. In this work, we consider a collections of Markov chains, where the set of values of each of the chains is taken together as a *meta state*. There are also numerous examples of models constructed from a hidden collection of Markov chains applied in speech processing and certain studies of music (see [5, 13, 17, 21] and the references therein).

Our core task in this article, however, is state estimation. There are two main motivations for this work, one is state estimation, (or filtering), in its own right, the second is to develop a smoother (all smoothers are functions of their underlying filters), to facilitate parameter estimation through schemes such as the expectation maximization (EM) algorithm. In this first article, we restrict our attention to the state estimation problem. The model class we consider is the so-called "Factorial Hidden Markov Model" (FHMM). The output of these FHMM models is an observation process which depends, in some coupled way, on the all the states of the individual and independent Markov chains. Our filter for this model will, at each discrete time, estimate the conditional probability density of the hidden state collection given an observation history. To simplify our calculations we will represent our Markov chains on a canonical basis of indicator functions. The technique we use to compute conditional expectations is the method of reference probability.

2. Models with Independent Hidden State Dynamics

The representation we use for our Markov chains identifies the state space of the chain on a canonical basis of indicator functions.

2.1. *Basic Markov Chains*

Consider a discrete time finite state, time-homogeneous Markov chain $X^\ell = \{X_k^\ell, 0 \leq k\}$, which can assume one of $N^\ell \in \mathbb{N}$ distinct states. Here the superscript ℓ labels a particular Markov chain within a collection of $\{1, 2, \ldots, M\}$ Markov chains. Without loss of generality we can identify the state space of X^ℓ with the canonical basis of unit vectors

$$\mathscr{S}^\ell \triangleq \{e_1^\ell, e_2^\ell, \ldots, e_{N^\ell}^\ell\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^{N^\ell}. \tag{1}$$

Suppose X^ℓ is defined on the probability space $(\Omega, \mathcal{F}, P)$ and write

$$\pi^\ell_{\{j,i\}} \triangleq P(X^\ell_{k+1} = \boldsymbol{e}^\ell_j \mid X^\ell_k = \boldsymbol{e}^\ell_i) \tag{2}$$

$$= P(X^\ell_1 = \boldsymbol{e}^\ell_j | X^\ell_0 = \boldsymbol{e}^\ell_i). \tag{3}$$

To denote the matrix of transition probabilities for the process X^ℓ, we write

$$\Pi^\ell = \left[\pi^\ell_{\{j,i\}}\right]_{\substack{1 \le j \le N^\ell \\ 1 \le i \le N^\ell}}. \tag{4}$$

Write $\mathcal{F}^\ell_{0,k} \triangleq \sigma\{X^\ell_0, X^\ell_1, \dots, X^\ell_k\}$. Then the dynamics for the process X^ℓ have the from:

$$X^\ell_{k+1} = \Pi^\ell X^\ell_k + V^\ell_{k+1}. \tag{5}$$

Here, V^ℓ is an increment of the $(P, \mathcal{F}^\ell_{0,k})$-martingale $\varphi^\ell_k = \sum_{r=1}^k V^\ell_r$.

2.2. *Observation Model*

Notation: We denote the determinant of the matrix B by $|B|$, further, a $d \times d$ identity matrix is written as $I_{d\times d}$. The regular inverse of the matrix B is written as $\mathrm{inv}(B)$.

The collection of state processes values at time k, that is the collection $\{X^1_k, \dots, X^M_k\}$, are not observed directly, instead this collection is partially observed through a vector-valued Gaussian random variable in $\mathbb{R}^{d\times 1}$, with density

$$P(Y_k \in dY \mid \{X^1_k, \dots, X^M_k\}) = \frac{(2\pi)^{-d/2}}{\sqrt{|C|}} \exp\left\{-\frac{1}{2}\left(Y - \sum_{\ell=1}^M W^\ell X^\ell_k\right)' \mathrm{inv}(C) \left(Y - \sum_{\ell=1}^M W^\ell X^\ell_k\right)\right\} dY. \tag{6}$$

Here, $Y \in \mathbb{R}^{d\times 1}$, W^ℓ is a $d \times N^\ell$ matrix, and $C \in \mathbb{R}^{d\times d}_+$ is a covariance matrix. This model is the standard in the literature (see, e.g., [13] and the references therein). In the sequel, we shall write $\widetilde{C}$ for a Cholesky Factor of the Cholesky decomposition of the covariance matrix C, that is, $C = \widetilde{C}\widetilde{C}'$.

Remark 2.1. It should be noted here, that a stochastic process generated by equation (6) is not Gaussian, except in a special degenerate case. That is, there is no joint Gaussian density that will describe a realization $\{Y_1, \dots, Y_k\}$. It is always true, that at each time k, Y_k is a Gaussian random variable. However, if the weight matrices W^ℓ are chosen so that the the sums $\sum_{\ell=1}^M W^\ell X^\ell_k$ degenerate to a single mean, then the output process will be a Gaussian stochastic process.

In the models of practical interest, we consider the weight matrices W^ℓ as being chosen so that the sums $\sum_{\ell=1}^M W^\ell X^\ell_k$ generate $K \in \mathbb{N}$ distinct means. In these scenarios the the stochastic process generated will be non-Gaussian, with a density composed, (roughly), from K sub-densities, each of which is individually Gaussian. The particular sub-density in effect at any time k will depend on the states of the Markov chains.

Notation: We introduce the filtrations $\mathbb{F}_{0,k} = \{\mathcal{F}_{0,k}\}$, $\mathbb{Y}_{0,k} = \{\mathcal{Y}_{0,k}\}$ and $\mathbb{G}_{0,k} = \{\mathcal{G}_{0,k}\}$, where

$$\mathcal{F}_{0,k} = \sigma\{X_0^1, \dots, X_k^1, X_0^2, \dots, X_k^2, \dots, X_0^M, \dots, X_k^M\}, \tag{7}$$

$$\mathcal{Y}_{0,k} = \sigma\{Y_1, Y_2, \dots, Y_k\}, \tag{8}$$

$$\mathcal{G}_{0,k} = \sigma\{X_0^1, \dots, X_k^1, X_0^2, \dots, X_k^2, \dots, X_0^M, \dots, X_k^M, Y_1, Y_2, \dots, Y_k\}. \tag{9}$$

Remark 2.2. The Gaussian probability density given at (6) explains the term "coupled" HMM. In this particular formulation, the M Markov processes are completely independent, however, the term coupling arises from the form of the process mean in the Gaussian observation model. These means are weighted linear combinations of each state value in the collection $\{X_k^1, \dots, X_k^M\}$. Consequently each observation will depend, (in a coupled manner), on each of the $1, \dots, M$ Markov processes through the mean process of the density given at (6).

3. Reference Probability

In filtering and parameter estimation, the central mathematical objects are conditional expectations. Often these expectations can be complex and difficult to evaluate. The aim of the reference probability technique, in this context, is to choose a reference probability measure that simplifies the basic calculations, that is, conditional expectations.

Initially we suppose all processes are defined on an ideal probability space $(\Omega, \mathcal{F}, P)$.

Definition 1. Suppose $\xi \in \mathbb{R}^{d\times 1}$, write

$$\Phi(\xi) \triangleq \frac{1}{(2\pi)^{d/2}} \exp\left\{-\frac{1}{2}\xi'\xi\right\}. \tag{10}$$

Under a new probability measure $P^\dagger$, defined below, the dynamics for each of the M Markov chains remains unchanged, while the observation process Y is now a vector-valued iid Gaussian process with zero mean and $I_{d\times d}$ covariance, that is under $P^\dagger$:

(1) $\{Y_k\} \sim \Phi(Y)$ for all k,
(2) for each $\ell \in \{1, 2, \dots, M\}$, the processes $\{X^\ell\}$ have dynamics $X_{k+1}^\ell = \Pi^\ell X_k^\ell + V_{k+1}^\ell$.

Definition 2. Suppose a scalar-valued process Λ is defined by

$$\Lambda_{0,k} \triangleq \prod_{n=0}^{k} \lambda_n, \tag{11}$$

where for $n = 1, 2, \dots,$

$$\lambda_n \triangleq \frac{\Phi\big(\mathrm{inv}(\widetilde{C})(Y_n - \sum_{\ell=1}^M W^\ell X_n^\ell)\big)}{|\widetilde{C}|\Phi(Y_n)} \tag{12}$$

and at time zero

$$\lambda_0 \triangleq 1. \tag{13}$$

Remark 3.1. Equation (12) suggests, roughly, an indication of the intuition upon which the reference probability method is based. The quantity λ_n is a quotient of two probability density functions. The numerator in this quantity is the density of the observations corresponding to the real world, while the denominator is the chosen density, under which we would like to do our simplified calculations.

We define the "real world" probability measure P by setting

$$\frac{dP}{dP^\dagger}\bigg|_{\mathscr{G}_{0,k}} \triangleq \Lambda_{0,k}. \tag{14}$$

It is standard to show that the stochastic process Λ, is an $(P^\dagger, \mathscr{G})$-martingale.

Lemma 1. *Write*

$$U_k \triangleq \mathrm{inv}(\widetilde{C})\left(Y_k - \sum_{\ell=1}^{M} W^\ell X_k^\ell\right). \tag{15}$$

Under the probability measure P, the process U is iid and normally distributed as $N(\boldsymbol{0}, I_{d\times d})$.

Lemma 2. *Under the reference probability measure $P^\dagger$, the dynamics for each of the $1, 2, \ldots, M$ Markov chains remain unchanged.*

Lemmata 1 and 2 are stated without proof.

In the sequel, we will make use of an abstract form of Bayes' Rule. Suppose that $\{\gamma\}$ is a $\mathscr{G}$-adapted process and that we wish to estimate $E[\gamma_k \mid \mathscr{Y}_{0,k}]$. Using a form of Bayes' rule [9],

$$E[\gamma_k \mid \mathscr{Y}_{0,k}] = \frac{E^\dagger[\Lambda_{0,k}\gamma_k \mid \mathscr{Y}_{0,k}]}{E^\dagger[\Lambda_{0,k} \mid \mathscr{Y}_{0,k}]}. \tag{16}$$

Here, $E^\dagger[\cdot]$ denotes expectation under a reference measure $P^\dagger$.

4. State Estimation

4.1. *Recursive Filter Dynamics*

Recall that for each of our $1, 2, \ldots, M$ Markov processes, the dimensions of the corresponding state spaces are, respectively, $\{N^1, N^2, \ldots, N^M\}$. To label a particular state in any of these M processes we use an index i_ℓ. For example, if considering the ℓth Markov process, the range of this index would be $1 \le i_\ell \le N^\ell$. What we are interested to compute, at each time denoted by the index k, are the estimated conditional probabilities corresponding to joint events of the form

$$\begin{aligned} A_k(i_1, i_2, \ldots, i_M) \triangleq \{\omega \mid X_k^1(\omega) = \boldsymbol{e}_{i_1}^1\} \cap \{\omega \mid X_k^2(\omega) = \boldsymbol{e}_{i_2}^2\} \\ \cap \cdots \cap \{\omega \mid X_k^M(\omega) = \boldsymbol{e}_{i_M}^M\}. \end{aligned} \tag{17}$$

These events are each uniquely labeled by a set of indices $\{i_1, i_2, \ldots, i_M\}$, where

$$\{i_1, i_2, \ldots, i_M\} \in \Gamma \triangleq \{1, 2, \ldots, N^1\} \times \{1, 2, \ldots, N^2\} \times \cdots \times \{1, 2, \ldots, N^M\}. \quad (18)$$

Since our M Markov processes are independent, we need only consider products of indicator functions, that is,

$$\begin{aligned} \mathbf{1}_{\{\{\omega \mid X_k^1(\omega)=e_{i_1}^1\}\cap\cdots\{\omega \mid X_k^M(\omega)=e_{i_M}^M\}\}} &= \prod_{\ell=1}^{M} \mathbf{1}_{\{\omega \mid X_k^\ell(\omega)=e_{i_\ell}^\ell\}} \\ &= \prod_{\ell=1}^{M} \langle X_k^\ell, e_{i_\ell}^\ell \rangle. \end{aligned} \quad (19)$$

Definition 3. Given the information $\mathcal{Y}_{0,k}$, the un-normalized conditional probability for the event $A_k(i_1, i_2, \ldots, i_M)$, denoted by $q_k(i_1, i_2, \ldots, i_M)$, is defined by the conditional expectation

$$q_k(i_1, \ldots, i_M) \triangleq E^\dagger\left[\Lambda_{0,k} \prod_{\ell=1}^{M} \langle X_k^\ell, e_{i_\ell}^\ell \rangle \mid \mathcal{Y}_{0,k}\right]. \quad (20)$$

The conditional expectation in (20) is a numerator in the abstract form of Bayes' rule given at (16), however, the denominator is readily computed from this numerator, which we now show.

First, note that for any of the M state processes

$$\sum_{i_\ell=1}^{N^\ell} \langle X_k^\ell, e_{i_\ell}^\ell \rangle = 1. \quad (21)$$

It follows that

$$\begin{aligned} \sum_{i_1=1}^{N^1} \cdots \sum_{i_M=1}^{N^M} q_k(i_1, \ldots, i_M) &= E^\dagger\left[\Lambda_{0,k} \sum_{i_1=1}^{N^1} \cdots \sum_{i_{M-1}=1} \sum_{i_M=1}^{N^M} \left\{\prod_{\ell=1}^{M} \langle X_k^\ell, e_{i_\ell}^\ell \rangle\right\} \middle| \mathcal{Y}_{0,k}\right] \\ &= \sum_{\{i_1,\ldots,i_M\}\in\Gamma} q_k(i_1, \ldots, i_M) \\ &= E^\dagger[\Lambda_{0,k} \mid \mathcal{Y}_{0,k}]. \end{aligned} \quad (22)$$

Therefore, the corresponding normalized conditional probability for the event $A_k(i_1, \ldots, i_M)$, is given by

$$\begin{aligned} p_k(i_1, \ldots, i_M) &= E\left[1_{\{\{\omega \mid X_k^1(\omega)=e_{i_1}^1\}\cap\cdots\{\omega \mid X_k^M(\omega)=e_{i_M}^M\}\}} \mid \mathcal{Y}_{0,k}\right] \\ &= E\left[\prod_{\ell=1}^{M} \langle X_k^\ell, e_{i_\ell}^\ell \rangle \mid \mathcal{Y}_{0,k}\right] \\ &= \frac{q_k(i_1, \ldots, i_M)}{\sum_{\{j_1,\ldots,j_M\}\in\Gamma} q_k(j_1, \ldots, j_M)}. \end{aligned} \quad (23)$$

Theorem 1 (State Estimation Filter). *Write*

$$\Psi(Y_k, i_1, \dots, i_M) \triangleq \frac{\Phi\big(\mathrm{inv}(\widetilde{C})(Y_k - \sum_{\ell=1}^{M} W^\ell \boldsymbol{e}_{i_\ell}^\ell)\big)}{|\widetilde{C}|\Phi(Y_k)}. \tag{24}$$

The un-normalized conditional density for the event $A_k(i_1, \dots, i_M)$ is computed by the recursion,

$$\begin{aligned} q_k\big(i_1, \dots, i_M\big) &= E^\dagger\Big[\Lambda_{0,k}\mathbf{1}_{\{X_k^1 = e_{i_1}^1, X_k^2 = e_{i_2}^2, \dots, X_k^M = e_{i_M}^M\}} \mid \mathcal{Y}_{0,k}\Big] \in \mathbb{R}_+ \\ &= \Psi(Y_k, i_1, \dots, i_M)\sum_{j_1}^{N^1}\cdots\sum_{j_M}^{N^M}\left[\prod_{\ell=1}^{M}\pi_{\{i_\ell, j_\ell\}}^\ell\right] q_{k-1}(j_1, \dots, j_M). \end{aligned} \tag{25}$$

Proof of Theorem 1. Since the Radon–Nikodym derivative $\Lambda_{0,k}$ is a product, we first note that

$$\begin{aligned} q_k(i_1, \dots, i_M) &= E^\dagger\left[\Lambda_{0,k}\prod_{\ell=1}^{M}\langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle \mid \mathcal{Y}_{0,k}\right] = E^\dagger\left[\Lambda_{0,k-1}\lambda_k\prod_{\ell=1}^{M}\langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle \mid \mathcal{Y}_{0,k}\right] \\ &= E^\dagger\left[\Lambda_{0,k-1}\frac{\Phi\big(\mathrm{inv}(\widetilde{C})(Y_k - \sum_{\ell=1}^{M} W^\ell X_k^\ell)\big)}{|\widetilde{C}|\Phi(Y_k)}\prod_{\ell=1}^{M}\langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle \mid \mathcal{Y}_{0,k}\right] \\ &= \Psi(Y_k, i_1, \dots, i_M)E^\dagger\left[\Lambda_{0,k-1}\prod_{\ell=1}^{M}\langle X_k, \boldsymbol{e}_{i_\ell}^\ell\rangle \mid \mathcal{Y}_{0,k}\right]. \end{aligned} \tag{26}$$

Recalling the Markov process dynamics given at (5), we write

$$\prod_{\ell=1}^{M}\langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle = \prod_{\ell=1}^{M}\langle \Pi^\ell X_{k-1}^\ell + V_k^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle. \tag{27}$$

Since we are conditioning the product term at (27) and noting that V is a martingale increment, then the only terms we need consider in the expansion of the right hand side of (27), are those in the product $\prod_{\ell=1}^{M}\langle \Pi^\ell X_{k-1}^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle$.

Further, since $\forall k \in \mathbb{N}$,

$$\sum_{j_1}^{N^1}\cdots\sum_{j_M}^{N^M}\left\{\prod_{\ell=1}^{M}\langle X_k^\ell, \boldsymbol{e}_{j_\ell}^\ell\rangle\right\} = 1, \tag{28}$$

we see that

$$\begin{aligned} &q_k(i_1, \dots, i_M) \\ &\quad = \Psi(Y_k, i_1, \dots, i_M)E^\dagger\left[\Lambda_{0,k-1}\prod_{\ell=1}^{M}\langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle \mid \mathcal{Y}_{0,k-1}\right] \\ &\quad = \Psi(Y_k, i_1, \dots, i_M)E^\dagger\left[\Lambda_{0,k-1}\prod_{\ell=1}^{M}\langle \Pi^\ell X_{k-1}^\ell, \boldsymbol{e}_{i_\ell}^\ell\rangle\sum_{j_1}^{N^1}\cdots\sum_{j_M}^{N^M}\left\{\prod_{\ell=1}^{M}\langle X_{k-1}^\ell, \boldsymbol{e}_{j_\ell}^\ell\rangle\right\} \Bigg| \mathcal{Y}_{0,k-1}\right] \end{aligned}$$

$$= \Psi(Y_k, i_1, \ldots, i_M) \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} E^{\dagger}\left[\Lambda_{0,k-1} \prod_{\ell=1}^{M} \langle \Pi^{\ell} X_{k-1}^{\ell}, e_{i_\ell}^{\ell}\rangle \prod_{\ell=1}^{M} \langle X_{k-1}^{\ell}, e_{j_\ell}^{\ell}\rangle \mid \mathcal{Y}_{0,k-1}\right]$$

$$= \Psi(Y_k, i_1, \ldots, i_M) \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \left\{ \prod_{\ell=1}^{M} \langle \Pi^{\ell} e_{j_\ell}^{\ell}, e_{i_\ell}^{\ell}\rangle \right\} E^{\dagger}\left[\Lambda_{0,k-1} \prod_{\ell=1}^{M} \langle X_{k-1}^{\ell}, e_{j_\ell}^{\ell}\rangle \mid \mathcal{Y}_{0,k-1}\right]$$

$$= \Psi(Y_k, i_1, \ldots, i_M) \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \left[\prod_{\ell=1}^{M} \pi_{\{i_\ell, j_\ell\}}^{\ell} \right] q_{k-1}(j_1, \ldots, j_M). \tag{29}$$

□

4.2. *Fixed Interval Smoother Dynamics*

In this section, we compute dynamics to estimate the smoother probabilities. We do this first for the so-called fixed-interval smoother, subsequently one may specialize our general formula to either a fixed lag smoother or a fixed point smoother.

Suppose $0 \leq k \leq T$ and we are given the information $\mathcal{Y}_{0,T}$. In smoothing, we are concerned with estimating the probability of an event $A(i^1, \ldots, i^M)$, given $\mathcal{Y}_{0,T}$. Again from Bayes' Theorem

$$E\left[\prod_{\ell=1}^{M} \langle X_k^{\ell}, e_{i_\ell}^{\ell}\rangle \mid \mathcal{Y}_{0,T}\right] = \frac{E^{\dagger}\left[\Lambda_{0,T} \prod_{\ell=1}^{M} \langle X_k^{\ell}, e_{i_\ell}^{\ell}\rangle \mid \mathcal{Y}_{0,T}\right]}{E^{\dagger}[\Lambda_{0,T} \mid \mathcal{Y}_{0,T}]}. \tag{30}$$

Write

$$\Lambda_{k+1,T} = \prod_{n=k+1}^{T} \lambda_n. \tag{31}$$

Now

$$E^{\dagger}\left[\Lambda_{0,T} \prod_{\ell=1}^{M} \langle X_k^{\ell}, e_{i_\ell}^{\ell}\rangle \mid \mathcal{Y}_{0,T}\right] = E^{\dagger}\left[\Lambda_{0,k} \prod_{\ell=1}^{M} \langle X_k^{\ell}, e_{i_\ell}^{\ell}\rangle E^{\dagger}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_{0,T} \vee \mathcal{F}_{0,k}\right] \mid \mathcal{Y}_{0,T}\right]. \tag{32}$$

Since our M state processes are Markov processes,

$$E^{\dagger}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_{0,T} \vee \mathcal{F}_{0,k}\right] = E^{\dagger}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_{0,T} \vee \sigma\{X_k^1, X_k^2, \ldots, X_k^M\}\right]. \tag{33}$$

Write

$$v_{k,T}(i_1, \ldots, i_M) \triangleq E^{\dagger}\left[\Lambda_{k+1,T} \mid \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = e_{i_1}^1, X_k^2 = e_{i_2}^2, \ldots, X_k^M = e_{i_M}^M\}\right]. \tag{34}$$

Remark 4.1. The process v is not an un-normalized probability. This process, at time index $k+1$, contains the future information in the observations $\{Y_{k+1}, \ldots, Y_T\}$ and is combined with q_k in a product form to construct the un-normalized smoother probability.

Lemma 3. *The process v, corresponding to an event $A_k(i_1, \ldots, i_M)$, for each index $k \in \{0, 1, \ldots, T-1\}$, satisfies the backward dynamics*

$$v_{k,T}(i_1, \ldots, i_M) = \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Psi(Y_{k+1}, j_1, \ldots, j_M) \left[\prod_{\ell=1}^{M} \pi_{\{j_\ell, i_\ell\}}^{\ell}\right] v_{k+1,T}(j_1, \ldots, j_M). \tag{35}$$

At the boundary, that is, the time index T, we have

$$v_{T,T}(i_1, \ldots, i_M) = 1, \quad \forall\{i_1, \ldots, i_M\} \in \Gamma. \tag{36}$$

Proof of Lemma 3.

$$\begin{aligned}
v_{k,T}(i_1, \ldots, i_M) &= E^{\dagger}\big[\Lambda_{k+1,T} \mid \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\}\big] \\
&= E^{\dagger}\big[\Lambda_{k+2,T}\lambda_{k+1} \mid \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\}\big] \\
&= E^{\dagger}\Bigg[\Lambda_{k+2,T} \frac{\Phi\big(\mathrm{inv}(\widetilde{C})(Y_{k+1} - \sum_{\ell=1}^{M} W^{\ell} X_{k+1}^{\ell})\big)}{|\widetilde{C}|\Phi(Y_{k+1})} \,\Bigg|\, \\
&\qquad \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\}\Bigg] \\
&= E^{\dagger}\Bigg[\Lambda_{k+2,T} \frac{\Phi\big(\mathrm{inv}(\widetilde{C})(Y_{k+1} - \sum_{\ell=1}^{M} W^{\ell} X_{k+1}^{\ell})\big)}{|\widetilde{C}|\Phi(Y_{k+1})} \\
&\qquad \times \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Bigg[\prod_{\ell=1}^{M} \big\langle X_{k+1}^{\ell}, \boldsymbol{e}_{j_\ell}^{\ell}\big\rangle\Bigg] \,\Bigg|\, \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\}\Bigg] \\
&= \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} E^{\dagger}\Bigg[\Lambda_{k+2,T} \frac{\Phi\big(\mathrm{inv}(\widetilde{C})(Y_{k+1} - \sum_{\ell=1}^{M} W^{\ell} X_{k+1}^{\ell})\big)}{|\widetilde{C}|\Phi(Y_{k+1})} \\
&\qquad \times \prod_{\ell=1}^{M} \big\langle X_{k+1}^{\ell}, \boldsymbol{e}_{j_\ell}^{\ell}\big\rangle \mid \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\}\Bigg] \\
&= \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Psi(Y_{k+1}, j_1, \ldots, j_M) E^{\dagger} \\
&\qquad \times \Bigg[\Lambda_{k+2,T} \prod_{\ell=1}^{M} \big\langle X_{k+1}^{\ell}, \boldsymbol{e}_{j_\ell}^{\ell}\big\rangle \mid \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\}\Bigg].
\end{aligned} \tag{37}$$

To complete this proof we apply the rule of repeated conditioning, noting that

$$\begin{aligned}
\mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\} &\subset \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\} \\
&\quad \vee \sigma\{X_{k+1}^1 = \boldsymbol{e}_{j_1}^1, \ldots, X_{k+1}^M = \boldsymbol{e}_{j_M}^M\},
\end{aligned} \tag{38}$$

consequently we see that,

$$\begin{aligned}
v_{k,T}(i_1, \ldots, i_M) = \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Psi\big(Y_{k+1}, j_1, \ldots, j_M\big) E^{\dagger}\Bigg[E^{\dagger}\Bigg[\Lambda_{k+2,T}\Bigg[\prod_{\ell=1}^{M} \big\langle X_{k+1}^{\ell}, \boldsymbol{e}_{j_\ell}^{\ell}\big\rangle\Bigg] \,\Bigg|\, \\
\mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\} \\
\vee \sigma\{X_{k+1}^1 = \boldsymbol{e}_{j_1}^1, \ldots, X_{k+1}^M = \boldsymbol{e}_{j_M}^M\}\Bigg] \,\Bigg|\, \\
\mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\}\Bigg]
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Psi(Y_{k+1}, j_1, \ldots, j_M) E^\dagger \Bigg[\Bigg[\prod_{\ell=1}^{M} \langle X_{k+1}^\ell, \boldsymbol{e}_{j_\ell}^\ell \rangle \Bigg] \\
&\quad \times E^\dagger \Bigg[\Lambda_{k+2,T} \,\Big|\, \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\} \\
&\qquad \vee \sigma\{X_{k+1}^1 = \boldsymbol{e}_{j_1}^1, \ldots, X_{k+1}^M = \boldsymbol{e}_{j_M}^M\} \Bigg] \Bigg| \\
&\qquad \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\} \Bigg] \\
&= \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Psi\big(Y_{k+1}, j_1, \ldots, j_M\big) E^\dagger \Bigg[\Bigg[\prod_{\ell=1}^{M} \langle \Pi^\ell X_k^\ell + V_{k+1}^\ell, \boldsymbol{e}_{j_\ell}^\ell \rangle \Bigg] \\
&\quad \times v_{k+1,T}(j_1, \ldots, j_M) \,\big|\, \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\} \Bigg] \\
&= \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Psi\big(Y_{k+1}, j_1, \ldots, j_M\big) \Bigg[\prod_{\ell=1}^{M} \langle A^\ell \boldsymbol{e}_{i_\ell}^\ell, \boldsymbol{e}_{j_\ell}^\ell \rangle \Bigg] \\
&\quad \times E^\dagger \Big[v_{k+1,T}(j_1, \ldots, j_M) \,\big|\, \mathcal{Y}_{0,T} \vee \sigma\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\} \Big] \\
&= \sum_{j_1}^{N^1} \cdots \sum_{j_M}^{N^M} \Psi\big(Y_{k+1}, j_1, \ldots, j_M\big) \Bigg[\prod_{\ell=1}^{M} \pi_{\{j_\ell, i_\ell\}}^\ell \Bigg] v_{k+1,T}\big(j_1, \ldots, j_M\big). \qquad (39)
\end{aligned}
$$

□

Remark 4.2. The backward recursion at (35), while similar in form to the forward recursion at (25), is in fact quite different. The process v is in effect a type of adjoint process corresponding to its q process.

Theorem 2 (Smoother)**.** *The normalized smoother probabilities for the event $A_k\big(i_1, \ldots, i_M\big)$ are computed by*

$$
\begin{aligned}
p_k\big(A(i_1, \ldots, i_M) \mid \mathcal{Y}_{0,T}\big) &\triangleq E\Bigg[\prod_{\ell=1}^{M} \langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell \rangle \mid \mathcal{Y}_{0,T} \Bigg] \\
&= \frac{q_k\big(i_1, \ldots, i_M\big) v_{k,T}\big(i_1, \ldots, i_M\big)}{\sum_{\{j_1, \ldots, j_M\} \in \Gamma} q_k\big(j_1, \ldots, j_M\big) v_{k,T}\big(j_1, \ldots, j_M\big)}. \qquad (40)
\end{aligned}
$$

Proof of Theorem 2. To establish Theorem 2, we compute the numerator and denominator of the Bayes' rule give at (30), respectively, there quantities are,

$$
\begin{aligned}
&E^\dagger \Bigg[\Lambda_{0,T} \prod_{\ell=1}^{M} \langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell \rangle \mid \mathcal{Y}_{0,T} \Bigg] \\
&= E^\dagger \Bigg[\Lambda_{0,k} \Lambda_{k+1,T} \prod_{\ell=1}^{M} \langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell \rangle \mid \mathcal{Y}_{0,T} \Bigg] \\
&= E^\dagger \Bigg[E^\dagger \Bigg[\Lambda_{0,k} \Lambda_{k+1,T} \prod_{\ell=1}^{M} \langle X_k^\ell, \boldsymbol{e}_{i_\ell}^\ell \rangle \mid \mathcal{Y}_{0,T} \vee \big\{X_k^1 = \boldsymbol{e}_{i_1}^1, \ldots, X_k^M = \boldsymbol{e}_{i_M}^M\big\} \Bigg] \Bigg| \mathcal{Y}_{0,T} \Bigg]
\end{aligned}
$$

$$
\begin{aligned}
&= E^{\dagger}\left[\Lambda_{0,k}\prod_{\ell=1}^{M}\langle X_k^{\ell}, \boldsymbol{e}_{i_\ell}^{\ell}\rangle E^{\dagger}\left[\Lambda_{k+1,T}\mid \mathcal{Y}_{0,T}\vee\{X_k^1=\boldsymbol{e}_{i_1}^1,\ldots,X_k^M=\boldsymbol{e}_{i_M}^M\}\right]\,\middle|\,\mathcal{Y}_{0,T}\right]\\
&= E^{\dagger}\left[\Lambda_{0,k}\prod_{\ell=1}^{M}\langle X_k^{\ell}, \boldsymbol{e}_{i_\ell}^{\ell}\rangle \mid \mathcal{Y}_{0,T}\right]v_{k,T}(i_1,\ldots,i_M)\\
&= E^{\dagger}\left[\Lambda_{0,k}\prod_{\ell=1}^{M}\langle X_k^{\ell}, \boldsymbol{e}_{i_\ell}^{\ell}\rangle \mid \mathcal{Y}_{0,k}\right]v_{k,T}(i_1,\ldots,i_M)\\
&= q_k(i_1,\ldots,i_M)v_{k,T}(i_1,\ldots,i_M)
\end{aligned}
\tag{41}
$$

and

$$
\begin{aligned}
\sum_{\{j_1,\ldots,j_M\}\in\Gamma} q_k(j_1,\ldots,j_M)v_{k,T}(j_1,\ldots,j_M) &= \sum_{\{j_1,\ldots,j_M\}\in\Gamma} E^{\dagger}\left[\Lambda_{0,T}\prod_{\ell=1}^{M}\langle X_k^{\ell}, \boldsymbol{e}_{j_\ell}\rangle \mid \mathcal{Y}_{0,T}\right]\\
&= E^{\dagger}\left[\Lambda_{0,T}\sum_{\{j_1,\ldots,j_M\}\in\Gamma}\left[\prod_{\ell=1}^{M}\langle X_k^{\ell}, \boldsymbol{e}_{j_\ell}\rangle\right]\,\middle|\,\mathcal{Y}_{0,T}\right]\\
&= E^{\dagger}\left[\Lambda_{0,T}\mid \mathcal{Y}_{0,T}\right].
\end{aligned}
\tag{42}
$$

□

Remark 4.3. Theorem 2 is the most general form of a smoother, here a so-called fixed interval smoother. This smoother can readily be specialized to either a fixed-point smoother or a fixed-lag smoother.

5. *M*-Ary Detection Schemes

In this section, we suppose that a finite collection of candidate model parameters sets is known in advance. That is, we suppose that under model hypothesis H_j, the specific parameters of the stochastic system defined by dynamics (5) and (6), is the set,

$$
H_j \triangleq \{\Pi_{H_j}^1, \Pi_{H_j}^2, \ldots, \Pi_{H_j}^M, C_{H_j}, W_{H_j}^1, W_{H_j}^2, \ldots, W_{H_j}^M\}. \tag{43}
$$

Further, we suppose that K candidate model hypotheses are proposed, where $K \in \mathbb{N}$ and $K \geq 2$, that is, our collection of candidate model hypotheses is $\{H_1, H_2, \ldots, H_K\}$. What we would like to do, is develop an online recursion to estimate/compute the conditional probabilities corresponding each of these K probabilities, that is, the probabilities

$$
p_k^j \triangleq E[H = H_j \mid \mathcal{Y}_{0,k}]. \tag{44}
$$

We suppose the, (unknown), state of K-valued simple random variable α, indicates which candidate model is in effect. The simple random variable α is defined on a canonical basis of unit vectors, that is,

$$
\alpha \in \mathcal{S} \triangleq \{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_K\} = \left\{\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix}, \begin{bmatrix}0\\1\\\vdots\\0\end{bmatrix}, \ldots, \begin{bmatrix}0\\0\\\vdots\\1\end{bmatrix}\right\} \subseteq \mathbb{R}^K. \tag{45}
$$

The state of α has the following implication,

$$H = H_j \Leftrightarrow \alpha = \boldsymbol{f}_j. \tag{46}$$

Theorem 3 (M-Ary Detector Scheme). *Write*

$$r_k^{H_j} \triangleq E^{\dagger}[\langle \alpha, \boldsymbol{f}_j \rangle \Lambda_k \mid \mathcal{Y}_{0,k}]. \tag{47}$$

Here, the quantity Λ_k and the sigma algebra $\mathcal{Y}_{0,k}$, are as defined above.
Write

$$\Psi^{H_j}(Y_k, i_1, \ldots, i_L) \triangleq \frac{\Phi\Big(\mathrm{inv}(\widetilde{C}_{H_j})(Y_k - \sum_{\ell=1}^{L} W_{H_j}^{\ell} e_{i_\ell}^{\ell})\Big)}{|\widetilde{C}_{H_j}|\Phi(Y_k)}. \tag{48}$$

The un-normalized probability $r_k^{H_j}$ is an estimate of the relative likelihood of model hypothesis H_j explaining the observation information $\mathcal{Y}_{0,k}$. The scalar quantity $r_k^{H_j}$ is computed from the following recursion,

$$r_k^{H_j} = r_{k-1}^{H_j} \sum_{i_1=1}^{N^1} \sum_{i_2=1}^{N^2} \cdots \sum_{i_L=1}^{N^L} \Psi^{H_j}(Y_k, i_1, \ldots, i_L) \frac{q_k^{H_j}(i_1, \ldots, i_L)}{\sum_{\{j_1,\ldots,j_L\}\in\Gamma} q_k^{H_j}(j_1, \ldots, j_L)} \in \mathbb{R}^1. \tag{49}$$

The corresponding normalized probability, that is $p\big(\alpha = \boldsymbol{f}_j \mid \mathcal{Y}_{0,k}\big)$, is computed as follows

$$p(\alpha = \boldsymbol{f}_j \mid \mathcal{Y}_{0,k}) = \frac{r_k^{H_j}}{\sum_{\ell=1}^{K} r_k^{H_\ell}}. \tag{50}$$

Proof of Theorem 3. We first recall that the random process Λ can be factored at any time k, therefore,

$$\begin{aligned} r_k^{H_j} &= E^{\dagger}\big[\langle \alpha, \boldsymbol{f}_j \rangle \Lambda_k \mid \mathcal{Y}_{0,k}\big] \\ &= E^{\dagger}\big[\langle \alpha, \boldsymbol{f}_j \rangle \lambda_k \Lambda_{k-1} \mid \mathcal{Y}_{0,k}\big]. \end{aligned} \tag{51}$$

Now, since we are considering a collection of candidate models, our Radon–Nikodym derivative (in the context of the reference probability technique), is in effect a convex combination of individual Radon–Nikodym derivatives. Consequently

$$\lambda_k = \sum_{i=1}^{K} \langle \alpha, \boldsymbol{f}_j \rangle \frac{\Phi\Big(\mathrm{inv}(\widetilde{C}_{H_i})(Y_k - \sum_{\ell=1}^{L} W_{H_i}^{\ell} X_k^{\ell})\Big)}{|\widetilde{C}_{H_i}|\Phi(Y_k)}. \tag{52}$$

Since α can be in one and only one state, we see that

$$r_k^{H_j} = E^{\dagger}\left[\langle \alpha, \boldsymbol{f}_j \rangle \frac{\Phi(\mathrm{inv}(\widetilde{C}_{H_j})(Y_k - \sum_{\ell=1}^{L} W_{H_j}^{\ell} X_k^{\ell}))}{|\widetilde{C}_{H_j}|\Phi(Y_k)} \Lambda_{k-1} \mid \mathcal{Y}_{0,k}\right]$$

$$= E\left[\langle \alpha, \boldsymbol{f}_j \rangle \frac{\Phi(\mathrm{inv}(\widetilde{C}_{H_j})(Y_k - \sum_{\ell=1}^{L} W_{H_j}^{\ell} X_k^{\ell}))}{|\widetilde{C}_{H_j}|\Phi(Y_k)} \,\middle|\, \mathcal{Y}_{0,k}\right] E^{\dagger}\left[\Lambda_{k-1} \mid \mathcal{Y}_{0,k}\right]$$

$$= E\left[\frac{\Phi\big(\mathrm{inv}(\widetilde{C}_{H_j})(Y_k - \sum_{\ell=1}^{L} W_{H_j}^{\ell} X_k^{\ell})\big)}{|\widetilde{C}_{H_j}|\Phi\big(Y_k\big)} \alpha = \boldsymbol{f}_j \,\&\, \mathcal{Y}_{0,k}\right]$$

$$\times E\left[\langle \alpha, \boldsymbol{f}_j \rangle \mid \mathcal{Y}_{0,k-1}\right] E^{\dagger}\left[\Lambda_{k-1} \mid \mathcal{Y}_{0,k-1}\right]$$

$$= r_{k-1}^{H_j} E\left[\frac{\Phi(\mathrm{inv}(\widetilde{C}_{H_j})(Y_k - \sum_{\ell=1}^{L} W_{H_j}^{\ell} X_k^{\ell}))}{|\widetilde{C}_{H_j}|\Phi(Y_k)} \,\middle|\, \alpha = \boldsymbol{f}_j \,\&\, \mathcal{Y}_{0,k}\right]$$

$$= r_{k-1}^{H_j} \sum_{i_1=1}^{N^1} \sum_{i_2=1}^{N^2} \cdots \sum_{i_L=1}^{N^L} \Psi^{H_j}(Y_k, i_1, \ldots, i_L)\, p_k(i_1, \ldots, i_L \mid \alpha = \boldsymbol{f}_j \,\&\, \mathcal{Y}_{0,k})$$

$$= r_{k-1}^{H_j} \sum_{i_1=1}^{N^1} \sum_{i_2=1}^{N^2} \cdots \sum_{i_L=1}^{N^L} \Psi^{H_j}(Y_k, i_1, \ldots, i_L) \frac{q_k^{H_j}(i_1, \ldots, i_L)}{\sum_{\{j_1,\ldots,j_L\}\in\Gamma} q_k^{H_j}(j_1, \ldots, j_L)}. \tag{53}$$

□

6. Simulation Example

For a simulation study, we consider the performance of the State Estimation Filter given in Theorem 1. What we would like to demonstrate, is the performance of the state estimation filter in a scenario of two hidden Markov chains and further, test this performance against a criteria analogous to decreasing Signal to Noise Ratio (SNR), as would be done in any more *standard* filtering context. However, since our observation process is a multivariate non-Gaussian density, there is no standard measure of SNR, as the observation process, in our context, is not corrupted by an additive noise process.

In our simulation we consider two Hidden Markov state processes, all realisations were 300 sample points in length. Each of these processes are time-homogeneous and each are two-state Markov chains. The corresponding transition matrices are,

$$\Pi^1 \triangleq \begin{bmatrix} 0.7 & 0.3 \\ 0.4 & 0.6 \end{bmatrix}, \tag{54}$$

$$\Pi^2 \triangleq \begin{bmatrix} 0.8 & 0.2 \\ 0.1 & 0.9 \end{bmatrix}. \tag{55}$$

The observation model weights are

$$W^1 \triangleq \begin{bmatrix} 1 & -4 \\ 2 & 2 \end{bmatrix}, \tag{56}$$

$$W^2 \triangleq \begin{bmatrix} 1 & 5 \\ 1 & -3 \end{bmatrix}. \tag{57}$$

The covariance matrix for the multivariate Gaussian observations was set at $C = 0.8I$, here I is the 2×2 identity matrix. Using the above parameters, we generate an observation process sampled from four Gaussian densities with a

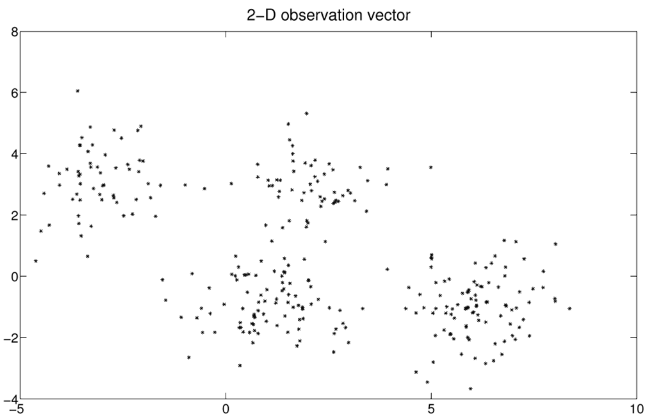

Figure 1. A small sampling of the output observation process. This plots indicates that the probability density of the observation process is clearly not Gaussian, however, it also suggests the existence of four sub-densities with distinct means on the (X, Y) plane. Each of these means are determined by the Weight matrices W^{ℓ}.

common covariance. These Gaussian densities are distinct by virtue of their particular vector-valued means only, which in turn are determined by the states of the two hidden Markov chains. These four sub-densities have means: $(2, 3)'$, $(6, -1)'$, $(-3, 3)'$ and $(1, -1)'$. Figure 1 shows a typical sampling in the two dimensional plane of the density means. In Figure 2 we show the state estimation filter performance for the scenario described above. The plotted state estimates were computed as soft decision conditional mean estimates.

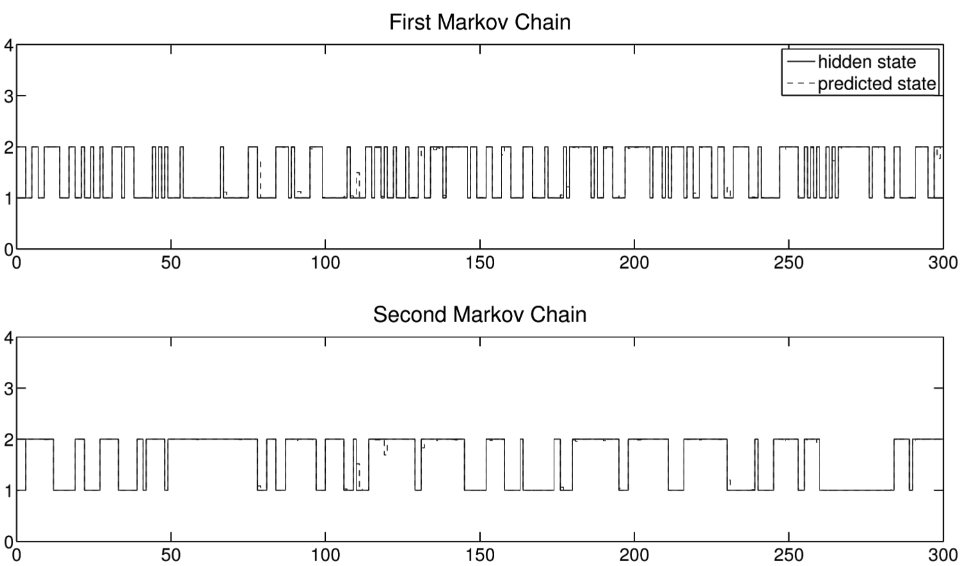

Figure 2. The two (true) partially observed Markov chains with their conditional mean estimates overlayed.

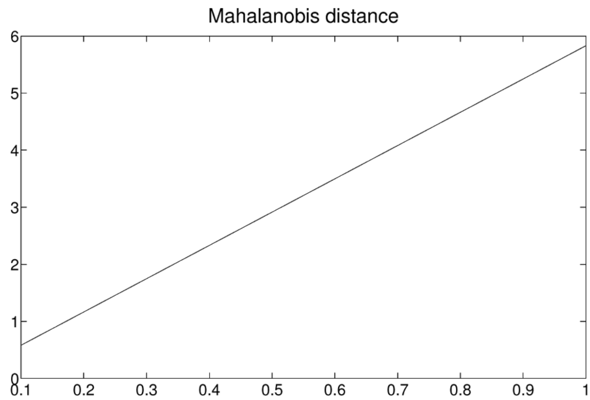

Figure 3. The Mahalanobis distance as a function of the scaled covariance matrix C as described in the text.

A natural question to consider in the specific simulation scenario we consider here, particularly in the absence of a measure of SNR, is: "what features of this model might challenge the estimator performance?" We consider the common covariance matrix C. If this quantity is increased, then the four Gaussian sub-densities generating the observation process will begin to overlap and hence challenge the state estimation scheme. In order to quantify the distance separating our four sub-densities, as a function of C, we compute the minimum Mahalanobis distance between all sub-density pairs, which we denote as $D(C)$, defined by the minimization

$$D(C) \stackrel{\Delta}{=} \min \left\{ D^M(g_1, g_2), D^M(g_1, g_3), D^M(g_1, g_4), \ldots, D^M(g_4, g_3) \right\}. \tag{58}$$

Here $D^M(g_p, g_m)$ is the Mahalanobis distance between the two Gaussian densities g_p and g_m. A calculation of $D(C)$ as a function of β, where $C = \frac{1}{\beta^2} I$, is shown in Figure 3. The corresponding performance of our state estimation filter against the same set of C values, is shown in Figure 4. In Figure 4 the dependent variable is

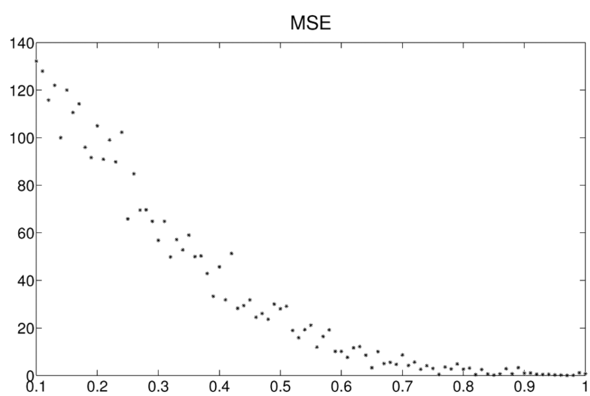

Figure 4. The MSE plotted against a scaling factor decreasing the covariance, consequently separating the (Gaussian) sub-densities in the observation stochastic process.

the Mean Square Error of the estimated state processes and the true state processes. This measure or error was defined by the sample average

$$MSE\big(X^1, X^2, \widehat{X}^1, \widehat{X}^2\big) \triangleq E\big((X^1 - \widehat{X}^1)^2 + (X^2 - \widehat{X}^2)^2\big). \tag{59}$$

As might be expected, the state estimator performance exponentially improves as the variance is linearly decreased.

7. Conclusion

In this work, new state estimation schemes have been developed for a collection of Markov Chains observed through a multivariate non-Gaussian density, whose stochastic means are determined by the states of the hidden Markov chains. A term used in the literature for this class of model is "Factorial Hidden Markov Model." The state estimation schemes developed were: a recursive state estimation filter, a general smoother and an M-ary detection scheme. Future work will investigate a filter-based and smoother-based expectation maximization algorithm for online parameter estimation using the filters and smoothers developed in this article. Future work will also include the extension of the model class to collections of hidden Markov chains that are statistically dependent.

References

1. Aggoun, L., and Elliott R.J. 2004. *Measure Theory and Filtering: Introduction with Applications*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, UK.
2. Bach, F., and Jordan, M. 2005. Discriminative training of hidden Markov models for multiple pitch tracking. International Conference on Acoustics, Speech and Signal processing.
4. Baum, L.E., and Petrie, T. 1966. Statistical inference for probabilistic functions of finite state Markov chains. *Annals of the Institute of Statistical Mathematics* 37:1554–1563.
4. Birney, L. 2001. Hidden Markov models in biological sequence analysis. *IBM Journal of Research and Development* 45:449–454.
5. Brand, M. 1997. Coupled Hidden Markov Models for Modeling Interacting Processes MIT Media Lab Perceptual Computing/Learning and Common Sense Techincal Report 405 (Revised), June.
6. Brand, M., Oliver, N., and Pentland A. 1997. Coupled hidden Markov models for complex action recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, San Juan, Puerto Rico.
7. Dempster, A.P., Laird, N.M., and Rubin, D.B. 1977. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B* 39:1–38.
8. Durbin, R., Eddy, S.R., and Krogh, A. 1998. *Biological Sequence Analysis*. Cambridge University Press, Cambridge, UK.
9. Elliott, R.J. 1982. *Stochastic Calculus and Its Applications*, Springer-Verlag, New York.
10. Elliott, R.J., Aggoun, L., and Moore, J.P. 1995. *Hidden Markov Models: Estimation and Control*. Springer-Verlag, New York.
11. Elliott, R.J., and Malcolm, W.P. 2000. Reproducing Gaussian densities and linear Gaussian detection. *Systems and Control Letters* 40:133–138.
12. Elliott, R.J., and Malcolm, W.P. 2006. Data-recursive smoother formulae for partially observed discrete-time Markov chains. *Stochastic Analysis and Applications* 24:579–597.
13. Ghahramani, Z., and Jordan, M. 1997. Factorial hidden Markov models. *Machine Learning* 29:245–273.

14. Husmeier, D. 2005. *Phylogenetic Factorial Hidden Markov Models*. ISBM, Detroit, MI.
15. Hyvaerinen, A., Karhunen, J., and Oja, E. 2001. *Independent Component Analysis*. John Wiley and Sons Inc., New York.
16. Koski, T. 2002. *Hidden Markov Models for Bioinformatics*. Springer Verlag, New York.
17. Logan, B., and Moreno, P.J. 1997. Factorial Hidden Markov Models for Speech Recognition: Preliminary Experiments. Cambridge Research Laboratory, Technical Report CRL 97/7, September.
18. MacDonald, I.L., and Zucchini, W. 1997. *Hidden Markov and Other Model for Discrete-Valued Time Series*. Chapman and Hall, Boca Raton, FL.
19. Virtanen, T. 2006. Speech recognition using factorial hidden Markov models for separation in the feature space, Proceedings of Interspeech, Pittsburgh, PA.
20. Wu, C.F.J. 1983. On the convergence properties of the EM algorithm. *Annals of Statistics* 11:95–103.
21. Zhong, S., and Ghosh, J. 2001. Coupled Hidden Markov Models. Tech. Report, June, The University of Texas at Austin Texas.