Robust M-ary Detection Filters and Smoothers for Continuous-Time Jump Markov Systems

Robert J. Elliott and W. P. Malcolm

Abstract—In this paper, we consider a dynamic M-ary detection problem when Markov chains are observed through a Wiener process. These systems are fully specified by a candidate set of parameters, whose elements are, a rate matrix for the Markov chain and a parameter for the observation model. Further, we suppose these parameter sets can switch according to the state of an unobserved Markov chain and thereby produce an observation process generated by time varying (jump stochastic) parameter sets. Given such an observation process and a specified collection of models, we estimate the probabilities of each model parameter set explaining the observation. By defining a new augmented state process, then applying the method of reference probability, we compute matrix-valued dynamics, whose solutions estimate joint probabilities for all combinations of candidate model parameter sets and values taken by the indirectly observed state process. These matrix-valued dynamics satisfy a stochastic integral equation with a Wiener process integrator. Using the gauge transformation techniques introduced by Clark and a pointwise matrix product, we compute robust matrix-valued dynamics for the joint probabilities on the augmented state space. In these new dynamics, the observation Wiener process appears as a parameter matrix in a linear ordinary differential equation, rather than an integrator in a stochastic integral equation. It is shown that these robust dynamics, when discretised, enjoy a determinsitic upper bound which ensures nonnegative probabilities for any observation sample path. In contrast, no such upper bounds can be computed for Taylor expansion approximations, such as the Euler–Maryauana and Milstein schemes. Finally, by exploiting a duality between causal and anticausal robust detector dynamics, we develop an algorithm to compute smoothed mode probability estimates without stochastic integrations. A computer simulation demonstrating performance is included.

Index Terms—Jump Markov systems, M-ary detection, martingales, reference probability.

I. INTRODUCTION

ONE WELL-KNOWN application for estimation with jump Markov systems is that of target tracking. Suppose we wish to estimate the state of a moving object. One can typically write down a single specific form of motion dynamics for a moving object. However, in most tracking applications the moving object will execute several distinct types of motion: for example, constant velocity, or a turning maneuver. In some scenarios one can assume the form of a law concerning the switching between candidate sets of dynamics. In these scenarios, a jump Markov system arises quite naturally. Here,

the task of estimation is to jointly detect which motion dynamics are in effect and estimate the state. The generic tracking problem just described is one of many applications for jump Markov systems.

For linear Gauss–Markov systems, the default jump Markov system algorithm is the so called interacting multiple model algorithm (IMM) introduced by Blom [2]. The IMM provides a scheme to compute both M-ary detection probabilities and state estimates. Recently two new algorithms have been proposed for jump Markov systems in [1] and [14]. Both of these new schemes, and the IMM, are for discrete time jump Markov systems. In this paper, we consider continuous time jump Markov systems, where a Markov chain is observed though a Wiener process. The primary emphasis in our work is upon robust dynamics to compute M-ary detection probabilities. However, once estimates of these probabilities are obtained, one can readily compute state estimates by a variety of different means.

The robust M-ary detector dynamics we compute are easily discretized in time to construct an approximate equation, not unlike the Duncan–Mortenson–Zakai equation, whose solution gives the unnormalized probability of a candidate model explaining an observation. The importance of these dynamics is that the unnormalized probabilities can then be used to obtain state estimates without stochastic integrations. By exploiting a duality we compute the corresponding anticausal (backward) dynamics, or the "dual" dynamics for the forward robust M-ary detector. Using both the forward and backward schemes we present a general algorithm to compute smoothed estimates of the M-ary detector probabilities. This algorithm can readily be specialised into the particular smoothing schemes known as: fixed point, fixed interval and fixed lag smoothers.

This paper is organized as follows. In Section II, the dynamics for the state process and the observation process are defined, also our reference probability measure and Radon–Nikodym derivative are described. In Section III, we briefly recall the seminal results of [3], in the context of a Wonham filter. In Section IV, there are three subsections. In Section IV-A, we compute a standard M-ary detection scheme for a fixed systems; in Section IV-B, we compute M-ary detection dynamics for a jump Markov system, these dynamics include stochastic integrals against the observation sample path; and in Section IV-C, we compute the corresponding robust M-ary detector dynamics, which do not include stochastic integration. In Section V, we compute a discrete-time recursion for the robust M-ary detection dynamics and compute a deterministic upper bound for the maximum grid step, ensuring a measure of filter stability. In Section VI, we compute smoother dynamics

for M-ary detection. Finally, in Section VII, we present a simulation, demonstrating the the performance of the detection dynamics given in Section V.

II. SIGNAL MODELS

Initially, we suppose that all processes are defined on the measurable space $(\Omega, \mathcal{F})$ with probability measure P.

A. State Process Dynamics

Suppose a state process $X = \{X_t, 0 \leq t\}$ is a finite state time-homogeneous Markov chain evolving in continuous time. It was shown in [5], that such processes admit convenient canonical representations. Without loss of generality, we can take the state space of X as $\mathcal{L} = \{e_1, e_2, \ldots, e_n\} \subseteq \mathbb{R}^n$, where e_i denotes a column vector in $\mathbb{R}^n$ with unity in the ith position and zero elsewhere. The dynamics for this process are

$$X_t = X_0 + \int_0^t A X_u ds + M_t \qquad (2.1)$$

where M_t is a $(P, \sigma\{X_u, 0 \leq u \leq t\})$-martingale and $A \in \mathbb{R}^{n \times n}$ is a rate matrix.

B. Observation Process Dynamics

We suppose that the process X is not observed directly, rather, we observe a scalar valued process

$$y_t = \int_0^t \langle X_u, g \rangle du + \widetilde{W}_t. \qquad (2.2)$$

Here, $\widetilde{W}$ is a standard Wiener process and $g = (\langle g, e_1 \rangle, \ldots, \langle g, e_n \rangle)' \in \mathbb{R}^n$, is a vector of the so called drift coefficients, or levels for the Markov chain. Our results can easily be extended to vector observations y. For our σ-fields, we write

$$\mathcal{F}_t = \sigma\{X_u; 0 \leq u \leq t\} \qquad (2.3)$$
$$\mathcal{G}_t = \sigma\{y_u, X_u; 0 \leq u \leq t\}. \qquad (2.4)$$

C. Reference Probability

We have under the "real world" probability P, dynamics of the form

$$P \begin{cases} dX_t = AX_t dt + dM_t \\ dy_t = \langle X_t, g \rangle dt + d\widetilde{W}_t. \end{cases} \qquad (2.5)$$

Suppose, however, that $P^\dagger$ is a new(reference) measure, under which X remains a Markov chain with dynamics (2.1) and the observation process y is a standard Brownian motion independent of X.

Denote by Λ_t

$$\Lambda_t = \exp\left\{ \int_0^t \langle X_u, g \rangle dy_u - \frac{1}{2} \int_0^t \langle X_u, g \rangle^2 du \right\}. \qquad (2.6)$$

Then, Λ has dynamics

$$\Lambda_t = 1 + \int_0^t \Lambda_u \langle X_u, g \rangle dy_u. \qquad (2.7)$$

Further

$$\frac{dP}{dP^\dagger}\bigg|_{\mathcal{G}_t} \triangleq \Lambda_t. \qquad (2.8)$$

In the sequel, we will carry out our calculations under the reference measure $P^\dagger$ for which our processes have dynamics

$$P^\dagger \begin{cases} dX_t = AX_t dt + dM_t \\ dy_t = dW_t. \end{cases} \qquad (2.9)$$

III. ROBUST STATE ESTIMATION FILTERS

In this section, we briefly recall the so called robust state estimation filter presented in [3], and give a precise meaning to the term "robust" in our context. Suppose a state process has dynamics given by (2.1) and an observation process has dynamics given by (2.2). Write

$$q_t = E^\dagger[\Lambda_t X_t \mid \mathcal{Y}_t]. \qquad (3.10)$$

The expectation operator $E^\dagger[\cdot]$ denotes an expectation under the reference measure $P^\dagger$. For a system with state dynamics (2.1) and observation dynamics (2.2), the state estimation filter [19] has dynamics

$$q_t = q_0 + \int_0^t A q_u du + \int_0^t \text{diag}\{\langle g, e_i \rangle\} q_u dy_u. \qquad (3.11)$$

Remark 1: The state estimation filter given at (3.11) is a continuous time filter and to obtain q it must be discretized in time. One common technique to perform such a discretization is the Euler–Maruyama approximation (see [11]). However, the stochastic integral in (3.11) almost surely has sample paths of unbounded variation. Any observation path y will have unbounded variation.

Definition 1: Define a matrix-valued stochastic process $\Phi \in \mathbb{R}^{n \times n}$, where

$$\Phi_t = \text{diag}\{\phi_t^1, \phi_t^2, \ldots, \phi_t^n\} \qquad (3.12)$$

with, $\phi_t^i = \exp(\langle g, e_i \rangle y_t - (1/2)\langle g, e_i \rangle^2 t) i = 1, 2 \ldots, n$.

It was shown in [3], that the transformed process $\bar{q}_t \triangleq \Phi_t^{-1} q_t$, satisfies the linear ordinary differential equation

$$\frac{d\bar{q}_t}{dt} = \Phi_t^{-1} A \Phi_t \bar{q}_t \quad \bar{q}_0 = q_o \in \mathbb{R}^n. \qquad (3.13)$$

The fundamental importance of this result is that it provides a means by which filtered estimates of the state process can be obtained without recourse to stochastic integration. Further, in [3] and [10] it was shown the quantity

$$\pi(X_t) \triangleq \frac{\Phi_t \bar{q}_t}{\langle \Phi_t \bar{q}_t, \mathbf{1} \rangle} \qquad (3.14)$$

defines a locally Lipschitz continuous version of the expectation $E[X_t \mid \mathcal{Y}_t]$. In this paper, the term "robust" is used to describe quantities (such as the expectation at (3.14)), which have Lipschitz continuous dependence upon the observation sample path. The value of computing filters with this property is immediately apparent when discretising filter dynamics in time.

IV. M-ARY DETECTION FILTERS

In this section, we first establish a basic detection filter for a fixed dynamical system. Subsequently, we develop detector dynamics for scenarios where model parameter sets switch according to a known law.

A. M-ary Detectors for a Fixed Dynamical System

Consider a system with state dynamics (2.1) and observation dynamics (2.2). To denote a single candidate set of parameters for a system of this type, we write H_j, where

$$H_j = \{A_{H_j}, \boldsymbol{g}_{H_j}\}. \tag{4.15}$$

Suppose one has a "reasonable" set of m candidate hypotheses $\{H_1, H_2, \ldots, H_m\}$, and that one has an observation process believed to have been generated by one and only one of the models H_j. The M-ary detection problem here is to compute the probabilities $P(H_j \mid \mathcal{Y}_t)$ for $j = 1, 2, \ldots, m$. To solve this problem we use a vector-valued simple random variable which acts like an m-state "switch," assumed random, yet fixed in one and only one position. Our simple random variable α takes values in a state–space $\mathcal{S}$, where

$$\alpha \in \mathcal{S} = \{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_m\} = \left\{ \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \ldots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\} \subseteq \mathbb{R}^m. \tag{4.16}$$

The meaning of each value taken by α is, $\alpha = \boldsymbol{f}_i \Leftrightarrow H_j$. Further, α will not appear in our final dynamics; its role in our calculations is simply that of an auxiliary variable.

Using α we rewrite the Radon–Nikodym derivative defined at (2.7) as

$$\Lambda_t = 1 + \sum_{j=1}^{m} \int_0^t \Lambda_u \langle \alpha, \boldsymbol{f}_j \rangle \langle X_u, \boldsymbol{g}_{H_j} \rangle dy_u. \tag{4.17}$$

Remark 2: To interpret equation (4.17), one reads it as both general and specific: general, in the sense that all models are included by the sum over $\boldsymbol{f}_j$ and specific, in the sense that one and only one equation is in effect according to the state of α.

Theorem 1: Write

$$q_t^j \triangleq E^\dagger[\Lambda_t \langle \alpha, \boldsymbol{f}_j \rangle \mid \mathcal{Y}_t]. \tag{4.18}$$

The unnormalized probability process q_t^j, satisfies the stochastic integral equation

$$q_t^j = q_0^j + \int_0^t \langle \widehat{X_u, \boldsymbol{g}_{H_j}} \rangle q_u^j dy_u \tag{4.19}$$

where $\langle \widehat{X_t, \boldsymbol{g}_{H_j}} \rangle = E[\langle X_t, \boldsymbol{g}_{H_j} \rangle \mid \mathcal{Y}_t]$ is evaluated under the probability measure P, given that the hypothesis H_j holds.

Proof of Theorem 1: We first note the semimartingale form of the process $\Lambda_t \langle \alpha, \boldsymbol{f}_j \rangle$

$$\Lambda_t \langle \alpha, \boldsymbol{f}_j \rangle = \langle \alpha, \boldsymbol{f}_j \rangle + \int_0^t \Lambda_u \langle \alpha, \boldsymbol{f}_j \rangle \langle X_u, \boldsymbol{g}_{H_j} \rangle dy_u. \tag{4.20}$$

Conditioning both sides of (4.20) on $\mathcal{Y}_t$, under the reference measure $P^\dagger$, we get

$$q_t^j = q_0^j + \int_0^t E^\dagger[\Lambda_u \langle \alpha, \boldsymbol{f}_j \rangle \langle X_u, \boldsymbol{g}_{H_j} \rangle \mid \mathcal{Y}_t] dy_u. \tag{4.21}$$

Since the expectation in the integrand of equation (4.21) is taken under the measure $P^\dagger$, and under this measure the observation in the model (2.9) is an independent increment process, then we may as well condition this integrand on $\mathcal{Y}_u$, rather than on $\mathcal{Y}_t$.

We first recall a version of Bayes' rule (see [7])

$$\begin{aligned} P(H_j \mid \mathcal{Y}_t) &= E[\langle \alpha, \boldsymbol{f}_j \rangle \mid \mathcal{Y}_t] \\ &= \frac{E^\dagger[\Lambda_t \langle \alpha, \boldsymbol{f}_j \rangle \mid \mathcal{Y}_t]}{E^\dagger[\Lambda_t \mid \mathcal{Y}_t]}. \end{aligned} \tag{4.22}$$

By using this version of Bayes rule and the rules of conditional probabilities, we see that

$$\begin{aligned} E^\dagger &\left[\Lambda_u \langle \alpha \boldsymbol{f}_j \rangle \langle X_u \boldsymbol{g}_{H_j} \rangle \mid \mathcal{Y}_u \right] \\ &= E[\langle \alpha, \boldsymbol{f}_j \rangle \langle X_u, \boldsymbol{g}_{H_j} \rangle \mid \mathcal{Y}_u] E^\dagger[\Lambda_u \mid \mathcal{Y}_u] \\ &= E[\langle X_u, \boldsymbol{g}_{H_j} \rangle \mid \alpha = \boldsymbol{f}_j, \mathcal{Y}_u] \\ &\quad \times E[\langle \alpha, \boldsymbol{f}_j \rangle \mid \mathcal{Y}_u] E^\dagger[\Lambda_u \mid \mathcal{Y}_u] \\ &= E[\langle X_u, \boldsymbol{g}_{H_j} \rangle \mid \alpha = \boldsymbol{f}_j, \mathcal{Y}_u] \\ &\quad \times \frac{E^\dagger[\Lambda_u \langle \alpha, \boldsymbol{f}_j \rangle \mid \mathcal{Y}_u]}{E^\dagger[\Lambda_u \mid \mathcal{Y}_u]} E^\dagger[\Lambda_u \mid \mathcal{Y}_u] \\ &= E[\langle X_u, \boldsymbol{g}_{H_j} \rangle \mid \alpha = \boldsymbol{f}_j, \mathcal{Y}_u] q_u^j \\ &= \langle \widehat{X_u, \boldsymbol{g}_{H_j}} \rangle q_u^j. \end{aligned} \tag{4.23}$$

$\square$

To compute the corresponding normalized probabilities, $p_t^j = P(\alpha = \boldsymbol{f}_j \mid \mathcal{Y}_t)$, we simply normalize by taking

$$p_t^j = \frac{q_t^j}{\sum_{\ell=1}^{m} q_t^\ell}. \tag{4.24}$$

Remark 3: The dynamics for the M-ary detection given previously include a stochastic integration. It is difficult to eliminate this stochastic integration when the problem is framed as before. However, by considering an augmented state-space model, including the state process and the auxiliary variable α, one can compute detection filters which require no stochastic integrations.

B. M-ary Detectors for Jump Markov Systems

We now suppose that α is itself a stochastic process with the previously defined state space $\mathcal{S}$. The dynamics for the process $\alpha = \{\alpha_u; 0 \le u \le t\}$ have the form

$$\alpha_t = \alpha_0 + \int_0^t B\alpha_u du + V_t. \tag{4.25}$$

Here, $B \in \mathbb{R}^{m \times m}$ is a rate matrix and V is a $(P, \sigma\{\alpha_u, 0 \le u \le t\})$-martingale.

To facilitate computing detector dynamics for a jump Markov system, we consider a new augmented state variable, taking values in the product space $\mathcal{L} \times \mathcal{S}$.

Write

$$\mathcal{G}_t \triangleq \sigma\{\alpha_u, X_u, y_u, 0 \leq u \leq t\}. \tag{4.26}$$

Definition 2: Define a matrix-valued process $Z = \{Z_u, 0 \leq u \leq t\}$, where

$$
\begin{aligned}
Z_t &\triangleq \alpha_t X_t' \\
&= \begin{bmatrix}
\langle \alpha_t, \boldsymbol{f}_1 \rangle \langle X_t, \boldsymbol{e}_1 \rangle & \cdots & \langle \alpha_t, \boldsymbol{f}_1 \rangle \langle X_t, \boldsymbol{e}_n \rangle \\
\langle \alpha_t, \boldsymbol{f}_2 \rangle \langle X_t, \boldsymbol{e}_1 \rangle & \cdots & \langle \alpha_t, \boldsymbol{f}_2 \rangle \langle X_t, \boldsymbol{e}_n \rangle \\
\vdots & \ddots & \vdots \\
\langle \alpha_t, \boldsymbol{f}_m \rangle \langle X_t, \boldsymbol{e}_1 \rangle & \cdots & \langle \alpha_t, \boldsymbol{f}_m \rangle \langle X_t, \boldsymbol{e}_n \rangle
\end{bmatrix} \\
&\subseteq \mathbb{R}^{m \times n}.
\end{aligned} \tag{4.27}
$$

Here, the process α satisfies the dynamics (4.25) and the process X satisfies the dynamics (2.1). The state–space for the process Z is the finite collection of matrices $\boldsymbol{F}_{(j,i)} = \boldsymbol{f}_j \boldsymbol{e}_i'$

$$
\begin{aligned}
Z \in \mathcal{M} &= \{\boldsymbol{F}_{(j,i)}\} \\
&= \left\{
\begin{bmatrix} 1 & 0 & \cdots & 0 \\ 0 & \cdots & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \ldots,
\begin{bmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & & \ddots & \vdots \\ 1 & 0 & \cdots & 0 \end{bmatrix}, \right. \\
&\qquad\qquad\qquad \vdots \\
&\left. \begin{bmatrix} 0 & \cdots & 0 & 1 \\ \vdots & \cdots & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 \end{bmatrix}, \ldots,
\begin{bmatrix} 0 & \cdots & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ \vdots & & \cdots & 0 \\ 0 & \cdots & 0 & 1 \end{bmatrix}
\right\}.
\end{aligned} \tag{4.28}
$$

Remark 4: The process Z takes values on a canonical basis of matrix-valued indicator functions, each of which jointly indicates a particular model hypothesis, and a particular value taken by the state process. For example, any one of the states of $Z \in \mathcal{M}$, at a certain time t, will indicate one and only one event of the form

$$
\begin{aligned}
&\mathbf{1}_{\{\omega | \alpha_t(\omega) = \boldsymbol{f}_j\} \cap \{\omega | X_t(\omega) = \boldsymbol{e}_i\}} \\
&\qquad\qquad \mathbf{1}_{\{\omega | \alpha_t(\omega) = \boldsymbol{f}_j\}} \mathbf{1}_{\{\omega | X_t(\omega) = \boldsymbol{e}_i\}}
\end{aligned} \tag{4.29}
$$

where the indicator functions are taken over all pairs (j,i), with $j \in \{1, 2, \ldots, m\}$ and $i \in \{1, 2, \ldots, n\}$.

Using the process α, our state dynamics, observer dynamics and Radon–Nikodym derivative now read, respectively

$$X_t = X_0 + \sum_{j=1}^{m} \int_0^t \langle \alpha_u, \boldsymbol{f}_j \rangle A_{H_j} X_u du + M_t \tag{4.30}$$

$$y_t = \sum_{j=1}^{m} \int_0^t \langle \alpha_u, \boldsymbol{f}_j \rangle \langle X_u, \boldsymbol{g}_{H_j} \rangle du + W_t \tag{4.31}$$

$$\Lambda_t = 1 + \sum_{j=1}^{m} \int_0^t \langle \alpha_u, \boldsymbol{f}_j \rangle \Lambda_u \langle X_u, \boldsymbol{g}_{H_j} \rangle dy_u. \tag{4.32}$$

The dynamics at (4.30)–(4.32) are again *general*, in the sense that all candidate models are included, but specific, in the sense

that the process α can take one and only one state value **at** any time t.

Theorem 2: Write

$$\widehat{Z}_t \triangleq E^\dagger[\Lambda_t Z_t \mid \mathcal{Y}_t] = E^\dagger[\Lambda_t \alpha_t X_t' \mid \mathcal{Y}_t] \in \mathbb{R}^{m \times n}. \tag{4.33}$$

Define

$$
G = \begin{bmatrix}
\langle \boldsymbol{g}_{H_1}, \boldsymbol{e}_1 \rangle & \langle \boldsymbol{g}_{H_1}, \boldsymbol{e}_2 \rangle & \cdots & \langle \boldsymbol{g}_{H_1}, \boldsymbol{e}_n \rangle \\
\langle \boldsymbol{g}_{H_2}, \boldsymbol{e}_1 \rangle & \langle \boldsymbol{g}_{H_2}, \boldsymbol{e}_2 \rangle & \cdots & \langle \boldsymbol{g}_{H_2}, \boldsymbol{e}_n \rangle \\
\vdots & \vdots & \ddots & \vdots \\
\langle \boldsymbol{g}_{H_m}, \boldsymbol{e}_1 \rangle & \langle \boldsymbol{g}_{H_m}, \boldsymbol{e}_2 \rangle & \cdots & \langle \boldsymbol{g}_{H_m}, \boldsymbol{e}_n \rangle
\end{bmatrix} \in \mathbb{R}^{m \times n}. \tag{4.34}
$$

The process $\widehat{Z}$, defined by (4.33), satisfies the dynamics

$$
\widehat{Z}_t = \widehat{Z}_0 + \int_0^t B \widehat{Z}_u du \\
+ \sum_{j=1}^{m} \sum_{i=1}^{n} \int_0^t (\boldsymbol{f}_j' \widehat{Z}_u \boldsymbol{e}_i) \boldsymbol{F}_{(j,i)} A_{H_j}' du + \int_0^t G \odot \widehat{Z}_u dy_u. \tag{4.35}
$$

The symbol $\odot$ in equation (4.35) denotes a pointwise matrix product, where for two matrices of the same dimensions, the pointwise product is

$$A \odot B = [a_{(i,j)} b_{(i,j)}]_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}. \tag{4.36}$$

Note that the binary operation $\odot$, defined previously, is commutative

$$A \odot B = B \odot A. \tag{4.37}$$

Write

$$\mathbf{1}_m = (1, 1, \ldots, 1)' \in \mathbb{R}^m \tag{4.38}$$

$$\mathbf{1}_n = (1, 1, \ldots, 1)' \in \mathbb{R}^n. \tag{4.39}$$

Recalling the numerator in Bayes' rule (4.22), we note that

$$
\begin{aligned}
\mathbf{1}_m' E^\dagger[\Lambda_t \alpha_t X_t' \mid \mathcal{Y}_t] \mathbf{1}_n &= E^\dagger[\Lambda_t \mathbf{1}_m' \alpha_t X_t' \mathbf{1}_n \mid \mathcal{Y}_t] \\
&= E^\dagger[\Lambda_t \mid \mathcal{Y}_t]. \tag{4.40}
\end{aligned}
$$

So, by computing the numerator in Bayes' rule, we can readily compute the normalizing denominator $E^\dagger[\Lambda_t \mid \mathcal{Y}_t]$. The matrix quantity $\widehat{Z}$, defined at (4.33), is an unnormalized probability, so, the corresponding normalized probability for $\widehat{Z}_t$ is computed by

$$P(Z_t \mid \mathcal{Y}_t) = E[Z_t \mid \mathcal{Y}_t] = \frac{\widehat{Z}_t}{\mathbf{1}_m' \widehat{Z}_t \mathbf{1}_n}. \tag{4.41}$$

To recover the normalized M-ary detection probabilities from the quantity $\widehat{Z}_t$, one computes

$$
\left\{ \frac{\widehat{Z}_t}{\mathbf{1}_m' \widehat{Z}_t \mathbf{1}_n} \right\} \mathbf{1}_n = \begin{bmatrix}
P(\alpha_t = \boldsymbol{f}_1 \mid \mathcal{Y}_t) \\
P(\alpha_t = \boldsymbol{f}_2 \mid \mathcal{Y}_t) \\
\vdots \\
P(\alpha_t = \boldsymbol{f}_m \mid \mathcal{Y}_t)
\end{bmatrix}. \tag{4.42}
$$

Proof of Theorem 2: Applying the product rule we see that

$$
\begin{aligned}
\alpha_t X_t' &= \alpha_0 X_0' + \int_0^t B \alpha_u X_u' du + \int_0^t dV_u X_u' \\
&+ \int_0^t \alpha_u dM_u' + \int_0^t \alpha_u \sum_{j=1}^{m} \langle \alpha_u, \boldsymbol{f}_j \rangle X_t' A_{H_j}' du. \tag{4.43}
\end{aligned}
$$

Similarly

$$\Lambda_t \alpha_t X'_t = \alpha_0 X'_0 + \int_0^t \Lambda_u B \alpha_u X'_u du$$

$$+ \int_0^t \Lambda_u dV_u X'_u + \int_0^t \Lambda_u \alpha_u dM'_u$$

$$+ \sum_{j=1}^m \sum_{i=1}^n \int_0^t \Lambda_u (\boldsymbol{f}'_j \alpha_u)(X'_u \boldsymbol{e}_i) \boldsymbol{f}_j \boldsymbol{e}'_i A'_{H_j} du$$

$$+ \sum_{j=1}^m \int_0^t \Lambda_u \langle \alpha_u, \boldsymbol{f}_j \rangle \langle X_u, \boldsymbol{g}_{H_j} \rangle \alpha_u X'_u dy_u. \tag{4.44}$$

The last term in (4.44) can be simplified by noting

$$\langle X_t, \boldsymbol{g}_{H_j} \rangle = \sum_{i=1}^n \langle X_u, \boldsymbol{e}_i \rangle \langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle \tag{4.45}$$

so

$$\sum_{j=1}^m \int_0^t \Lambda_u \langle \alpha_u, \boldsymbol{f}_j \rangle \langle X_u, g H_j \rangle \alpha_u X'_u dy_u$$

$$= \sum_{j=1}^m \sum_{i=1}^n \int_0^t \Lambda_u \langle \alpha_u, \boldsymbol{f}_j \rangle \langle X_u, \boldsymbol{e}_i \rangle \langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle \alpha_u X'_u dy_u$$

$$= \int_0^t \Lambda_u G \odot (\alpha_u X'_u) dy_u. \tag{4.46}$$

Using this simplification and rewriting (4.44) in terms of the process Z, we get

$$\Lambda_t Z_t = Z_0 + \int_0^t \Lambda_u B Z_u du$$

$$+ \int_0^t \Lambda_u dV_u X'_u + \int_0^t \Lambda_u \alpha_u dM'_u$$

$$+ \sum_{j=1}^m \sum_{i=1}^n \int_0^t \Lambda_u (\boldsymbol{f}'_j \alpha_u X'_u \boldsymbol{e}_i) \boldsymbol{F}_{(j,i)} A'_{H_j} du$$

$$+ \int_0^t \Lambda_u G \odot (\alpha_u X'_u) du. \tag{4.47}$$

Finally, conditioning both sides of (4.47) on $\mathcal{Y}_t$, under the reference measure $P^\dagger$ and using a version of Fubini's theorem, we get

$$\widehat{Z}_t = \widehat{Z}_0 + \int_0^t B\widehat{Z}_u du + \sum_{j=1}^m \sum_{i=1}^n \int_0^t (\boldsymbol{f}'_j \widehat{Z}_u \boldsymbol{e}_i) \boldsymbol{F}_{(j,i)} A'_{H_j} du$$

$$+ \int_0^t G \odot \widehat{Z}_u dy_u. \tag{4.48}$$

$\square$

Corollary 1: Suppose that $\forall \ H_j \ \in \ \{H_1, H_2, \ldots, H_m\}$, $A_{H_j} = A$. Then

$$\widehat{Z}_t = \widehat{Z}_0 + \int_0^t B\widehat{Z}_u du + \int_0^t \widehat{Z}_u A' du + \int_0^t G \odot \widehat{Z}_u dy_u. \tag{4.49}$$

Corollary 2: Suppose that $\forall \ H_j \ \in \ \{H_1, H_2, \ldots, H_m\}$, $\boldsymbol{g}_{H_j} = \boldsymbol{g}$. Then, the matrix G, in the dynamics at (4.35), has the form

$$G = \begin{bmatrix} \langle \boldsymbol{g}, \boldsymbol{e}_1 \rangle & \langle \boldsymbol{g}, \boldsymbol{e}_2 \rangle & \ldots & \langle \boldsymbol{g}, \boldsymbol{e}_n \rangle \\ \langle \boldsymbol{g}, \boldsymbol{e}_1 \rangle & \langle \boldsymbol{g}, \boldsymbol{e}_2 \rangle & \ldots & \langle \boldsymbol{g}, \boldsymbol{e}_n \rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle \boldsymbol{g}, \boldsymbol{e}_1 \rangle & \langle \boldsymbol{g}, \boldsymbol{e}_2 \rangle & \ldots & \langle \boldsymbol{g}, \boldsymbol{e}_n \rangle \end{bmatrix} \in \mathbb{R}^{m \times n}. \tag{4.50}$$

C. Robust M-Ary Detectors for Jump Markov Systems

Definition 3: Define a matrix-valued process Φ

$$\Phi_t = \left[\phi_t^{(j,i)} \right]_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} \in \mathbb{R}^{m \times n} \tag{4.51}$$

where

$$\phi_t^{(j,i)} = \exp(\langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle y_t - \frac{1}{2} \langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle^2 t)$$

$$\forall \ (j,i), j = 1, \ldots, m, \qquad i = 1, \ldots, n. \tag{4.52}$$

Each of the elements in Φ_t is an exponential function and, therefore, each of these elements have well defined reciprocals.

Definition 4: The inverse of the matrix Φ_t has the form

$$\Phi_t^{-1} \triangleq \left[\frac{1}{\phi_t^{(j,i)}} \right]_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}} \in \mathbb{R}^{m \times n}. \tag{4.53}$$

Theorem 3: Write

$$\overline{Z}_t \triangleq \Phi_t^{-1} \odot \widehat{Z}_t. \tag{4.54}$$

The process $\overline{Z}_t$ satisfies the linear ordinary differential equation

$$\frac{d\overline{Z}_t}{dt} = \Phi_t^{-1} \odot B(\Phi_t \odot \overline{Z}_t)$$

$$+ \Phi_t^{-1} \odot \left\{ \sum_{j=1}^m \sum_{i=1}^n \boldsymbol{f}'_j (\Phi_t \odot \overline{Z}_t) \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A'_{H_j} \right\}. \tag{4.55}$$

Proof of Theorem 3: To establish Theorem 3, we apply the product rule to the process $\Phi_t^{-1} \odot \widehat{Z}_t$. First, we compute the dynamics for the process Φ_t^{-1}. Each element of the matrix Φ_t^{-1} has the form

$$\frac{1}{\phi_t^{(j,i)}} = \exp\left(-\langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle y_t + \frac{1}{2} \langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle^2 t \right). \tag{4.56}$$

Using the Itô rule, we get

$$d\left(\frac{1}{\phi_t^{(i,j)}} \right) = \langle \boldsymbol{g}_{H_t}, \boldsymbol{e}_i \rangle^2 \frac{1}{\phi_t^{(i,j)}} dt - \langle \boldsymbol{g}_{h_j}, \boldsymbol{e}_i \rangle \frac{1}{\phi_t^{(i,j)}} dy_t. \tag{4.57}$$

It follows from (4.57), that the matrix Φ_t^{-1} satisfies the stochastic integral equation

$$\Phi_t^{-1} = \Phi_0^{-1} + \int_0^t G \odot G \odot \Phi_u^{-1} du - \int_0^t G \odot \Phi_u^{-1} dy_u. \tag{4.58}$$

Using the dynamics (4.58) and (4.55), we see that

$$
\begin{aligned}
d(\Phi_t^{-1} \odot \widehat{Z}_t) &= d\Phi_t^{-1} \odot \widehat{Z}_t + \Phi_t^{-1} \odot d\widehat{Z}_t + d[\Phi_t^{-1}, \widehat{Z}_t] \\
&= (G \odot G \odot \Phi_t^{-1} dt - G \odot \Phi_t^{-1} dy_t) \odot \widehat{Z}_t \\
&\quad + \Phi_t^{-1} \\
&\quad \odot \left\{ B\widehat{Z}_t dt + \sum_{j=1}^{m} \sum_{i=1}^{n} (\boldsymbol{f}_j' \widehat{Z}_t \boldsymbol{e}_i) \boldsymbol{F}_{(j,i)} A_{H_j}' dt \right. \\
&\qquad \left. + G \odot \widehat{Z}_t dy_t \right\} - G \odot \Phi_t^{-1} \odot G \odot \widehat{Z}_t dt \\
&= \Phi_t^{-1} \odot B(\Phi_t \odot \overline{Z}_t) dt + \Phi_t^{-1} \\
&\quad \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}_j'(\Phi_t \odot \overline{Z}_t) \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j} \right\} dt.
\end{aligned}
$$
(4.59)

$\square$

Corollary 3: Suppose the rate matrix B, in the dynamics given at (4.25), is a zero matrix in $\mathbb{R}^{m \times m}$. This condition corresponds to a scenario where the α process is a degenerate stochastic process, here a simple random variable. The robust dynamics for this scenario are

$$
\frac{d\overline{Z}_t}{dt} = \Phi_t^{-1} \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}_j'(\Phi_t \odot \overline{Z}_t) \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}' \right\}.
$$
(4.60)

To recover the normalized M-ary detection probabilities in this scenario, one computes

$$
\left\{ \frac{\Phi_t \odot \overline{Z}_t}{\boldsymbol{1}_m'(\Phi_t \odot \overline{Z}_t) \boldsymbol{1}_n} \right\} \boldsymbol{1}_n = \begin{bmatrix} P(\alpha = \boldsymbol{f}_1 \mid \mathcal{Y}_t) \\ P(\alpha = \boldsymbol{f}_2 \mid \mathcal{Y}_t) \\ \vdots \\ P(\alpha = \boldsymbol{f}_m \mid \mathcal{Y}_t) \end{bmatrix}.
$$
(4.61)

Theorem 4: For the process X, the quantity

$$
\pi(\alpha_t X_t') \triangleq \frac{\Phi_t \odot \overline{Z}_t}{\boldsymbol{1}_m'(\Phi_t \odot \overline{Z}_t) \boldsymbol{1}_n}.
$$
(4.62)

defines a locally Lipschitz continuous version of the expectation $E[\alpha_t X_t' \mid \mathcal{Y}_t]$.

Theorem 4 is stated here without proof. Similar proofs are given in [3] and [10].

V. DISCRETE-TIME DYNAMICS

A. Discrete-Time Filter

To compute a practical form of the filter given by Theorem 3, we choose a regular time partition to develop suboptimal discrete-time dynamics for the process $\widehat{Z}$.

Consider a regular partition where the interval $[0, T]$ is discretized by

$$
0 = t_1 < \cdots t_k < \cdots < t_{K-1} < t_K = T.
$$
(5.63)

Here, $t_k = k\Delta$, where $\Delta = T/K$ for some $K \in \mathbb{N}$. Using the partition just described, we discretize the robust dynamics given in Theorem 3. First, we rewrite the ODE given at (4.55) in its corresponding integral equation form

$$
\begin{aligned}
\overline{Z}_{t_k} &= \overline{Z}_{t_{k-1}} + \int_{t_{k-1}}^{t_k} \Phi_u^{-1} \odot B(\Phi_u \odot \overline{Z}_u) du + \int_{t_{k-1}}^{t_k} \Phi_u^{-1} \\
&\quad \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}_j(\Phi_u \odot \overline{Z}_u) \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}' \right\} du \\
&\approx \overline{Z}_{t_{k-1}} + \Phi_{t_{k-1}}^{-1} \odot B(\Phi_{t_{k-1}} \odot \overline{Z}_{t_{k-1}}) \Delta + \Phi_{t_{k-1}}^{-1} \\
&\quad \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}_j(\Phi_{t_{k-1}} \odot \overline{Z}_{t_{k-1}}) \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}' \right\} \Delta.
\end{aligned}
$$
(5.64)

Recalling $\overline{Z}_{t_k} \triangleq \Phi_{t_k}^{-1} \odot \widehat{Z}_{t_k}$, we compute discrete dynamics for $\widehat{Z}_{t_k}$, by premultiplying Φ_{t_k} on both sides of the approximation at (5.64)

$$
\begin{aligned}
\widehat{Z}_{t_k} &= \Phi_{t_k} \odot \overline{Z}_{t_{k-1}} + \Phi_{t_k} \odot \Phi_{t_{k-1}}^{-1} \odot B(\Phi_{t_{k-1}} \odot \overline{Z}_{t_{k-1}}) \Delta \\
&\quad + \Phi_{t_t} \odot \Phi_{t_{k-1}}^{-1} \\
&\quad \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}_j'(\Phi_{t_{k-1}} \odot \overline{Z}_{t_{k-1}}) \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}' \right\} \Delta \\
&= \Phi_{t_k} \odot \Phi_{t_{k-1}}^{-1} \odot \widehat{Z}_{t_{k-1}} + \Phi_{t_k} \odot \Phi_{t_{k-1}}^{-1} \odot (B\widehat{Z}_{t_{k-1}}) \Delta \\
&\quad + \Phi_{t_t} \odot \Phi_{t_{k-1}}^{-1} \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}_j \widehat{Z}_{t_{k-1}} \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}' \right\} \Delta.
\end{aligned}
$$
(5.65)

To simplify notation, for any time dependent quantity ξ_t, we write $\xi_k = \xi_{t_k}$.

Write

$$
\begin{aligned}
\Psi_{k,k-1} &\triangleq \Phi_k \odot \Phi_{k-1}^{-1} \\
&= \left[\exp(\langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle (y_k - y_{k-1}) \right. \\
&\qquad \left. - \frac{1}{2} \langle \boldsymbol{g}_{H_j}, \boldsymbol{e}_i \rangle^2 \Delta) \right]_{\substack{i=1,\ldots,n \\ j=1,\ldots,m}}.
\end{aligned}
$$
(5.66)

Finally

$$
\begin{aligned}
\widehat{Z}_k &= \Psi_{k,k-1} \odot \widehat{Z}_{k-1} + \Delta \Psi_{k,k-1} \odot (B\widehat{Z}_{k-1}) \\
&\quad + \Delta \Psi_{k,k-1} \odot \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}_j' \widehat{Z}_{k-1} \boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}'.
\end{aligned}
$$
(5.67)

B. Discretization Limits

For a partition of the interval $[0, T]$, not necessarily regular, we write

$$
\Pi_K^\ell \triangleq \left\{ 0 = t_0^\ell, t_1^\ell, \ldots, t_K^\ell = T \right\}.
$$
(5.68)

Each partition Π_K^ℓ is strictly increasing, that is, $t_0^\ell < t_1^\ell < \cdots < t_K^\ell$.

Definition 5: A numerical implementation of dynamics at (5.67), is said to be stable on the partition Π_K^ℓ, if for each $i \in \{1, 2, \ldots, n\}$, for each $j \in \{1, 2, \ldots, m\}$ and for each $k \in \{1, 2, \ldots, K\}$, the following inequality holds:

$$\boldsymbol{f}_j' \widehat{Z}_k \boldsymbol{e}_i \geq 0. \tag{5.69}$$

Theorem 5: The robust time-discretised dynamics at (5.67) are stable on the partition Π_K^ℓ, provided the following inequality is satisfied:

$$\|\Pi_K^\ell\| \triangleq \max_k \left(t_k^\ell - t_{k-1}^\ell\right)$$
$$\leq \frac{1}{\max_i \{|b_{(i,i)}|\} + \max_{(j,i)} \left\{\left|a_{(i,i)}^{H_j}\right|\right\}}. \tag{5.70}$$

Proof of Theorem 5: Consider an element of the matrix $\widehat{Z}_k$, at row j and column i and write

$$\widehat{z}_k^{(j,i)} \triangleq \boldsymbol{f}_j' \widehat{Z}_k \boldsymbol{e}_i. \tag{5.71}$$

Similarly, we write

$$\psi_{k,k-1}^{(j,i)} \triangleq \boldsymbol{f}_j' \Psi_{k,k-1} \boldsymbol{e}_i. \tag{5.72}$$

The scalar quantity $\widehat{z}_k^{(j,i)}$ corresponds to an approximation, in discrete-time (at time t_k), of the unnormalized probability $E^\dagger[\langle \alpha_t, \boldsymbol{f}_j \rangle \langle X_t, \boldsymbol{e}_i \rangle \mid \mathcal{Y}_{t_k}]$. Recalling the dynamics at (5.67), we see that

$$\boldsymbol{f}_j' \widehat{Z}_k \boldsymbol{e}_i = \widehat{z}_k^{(j,i)} = \boldsymbol{f}_j' \left(\Psi_{k,k-1} \odot \widehat{Z}_{k|k-1}\right) \boldsymbol{e}_i$$
$$+ \boldsymbol{f}_j' \left(\Delta \Psi_{k,k-1} \odot \left(B \widehat{Z}_{k-1}\right)\right) \boldsymbol{e}_i$$
$$+ \boldsymbol{f}_j' \left(\Delta \Psi_{k,k-1} \odot \sum_{j=1}^m \sum_{i=1}^n \boldsymbol{f}_j' \widehat{Z}_{k-1} \boldsymbol{e}_i F_{(j,i)} A'_{H_j}\right) \boldsymbol{e}_i$$
$$= \psi_{k,k-1}^{(j,i)} \widehat{z}_{k-1}^{(j,i)}$$
$$+ \Delta^{(j,i)} \psi_{k,k-1}^{(j,i)} \left(-|b_{(j,j)}| \widehat{z}_{k-1}^{(j,i)} + \sum_{k \neq j}^m b_{j,k} \widehat{z}_{k-1}^{(j,i)}\right)$$
$$+ \Delta^{(j,i)} \psi_{k,k-1}^{(j,i)} \left(-\left|a_{(i,i)}^{H_j}\right| \widehat{z}_{k-1}^{(j,i)} + \sum_{k \neq i}^m a_{k,i}^{H_j} \widehat{z}_{k-1}^{(j,i)}\right). \tag{5.73}$$

In oder to satisfy the stability criterion given at Definition 5, the left-hand side of the recursion at (5.73) must remain positive, that is

$$\psi_{k,k-1}^{(j,i)} \widehat{z}_{k-1}^{(j,i)} - \Delta^{(j,i)} \psi_{k,k-1}^{(j,i)} |b_{(j,j)}| \widehat{z}_{k-1}^{(j,i)}$$
$$+ \Delta^{(j,i)} \psi_{k,k-1}^{(j,i)} \sum_{k \neq j}^m b_{j,k} \widehat{z}_{k-1}^{(j,i)}$$
$$- \Delta^{(j,i)} \psi_{k,k-1}^{(j,i)} \left|a_{(i,i)}^{H_j}\right| \widehat{z}_{k-1}^{(j,i)}$$
$$+ \Delta^{(j,i)} \psi_{k,k-1}^{(j,i)} \sum_{k \neq i}^m a_{k,i}^{H_j} \widehat{z}_{k-1}^{(j,i)} \geq 0. \tag{5.74}$$

Simplifying this inequality, we get

$$\left(1 - \Delta^{(j,i)} \left(|b_{(i,i)}| + \left|a_{(i,i)}^{H_j}\right|\right)\right) \widehat{z}_{k-1}^{(j,i)}$$
$$+ \Delta^{(j,i)} \sum_{k \neq j}^m b_{j,k} \widehat{z}_{k-1}^{(j,i)} + \Delta^{(j,i)} \sum_{k \neq i}^m a_{k,i}^{H_j} \widehat{z}_{k-1}^{(j,i)} \geq 0. \tag{5.75}$$

Since the off-diagonal elements of the two rate matrices B and A^{H_j}, are always positive, then the terms including these elements in (5.75) are both positive. So, to ensure that the inequality at (5.75) is satisfied, all we need do and indeed all we "can do," is choose $\Delta^{(j,i)}$, such that the quantity $\left(1 - \Delta^{(j,i)} \left(|b_{(i,i)}| + \left|a_{(i,i)}^{H_j}\right|\right)\right)$ is nonnegative, that is

$$\Delta^{(j,i)} \leq \frac{1}{|b_{(i,i)}| + \left|a_{(i,i)}^{H_j}\right|}. \tag{5.76}$$

The corresponding global upper limit (over all (j,i) pairs) is, therefore,

$$\Delta \leq L \triangleq \min_{(j,i)} \left\{\Delta^{(j,i)}\right\}$$
$$= \frac{1}{\max_i \{|b_{(i,i)}|\} + \max_{(j,i)} \left\{\left|a_{(i,i)}^{H_j}\right|\right\}}. \tag{5.77}$$

$\square$

Remark 5: It is interesting to note that the upper bound for time discretization, given by Theorem 5, does not depend on the drift vectors $\boldsymbol{g}_{H_1}, \ldots, \boldsymbol{g}_{H_m}$ and only upon the the diagonal elements of the rate matrices B and $A_{H_1}, \ldots, A_{H_M}$.

Remark 6: To emphasise to value of Theorem 5, consider a similar calculation, such as the calculation given to prove this Theorem, but for an alternative time discretization of the stochastic integral equation at (4.35), for example, the Euler–Maryauana scheme, or the Milstein scheme. Omitting the details, what one derives from such a calculation, is an upper limit which depends upon the difference $y_k - y_{k-1}$. What this means, first, is that such a limit is clearly stochastic, and secondly, it could be negative in sign and therefore meaningless. In contrast, using the robust discretization at (5.67), one has a deterministic upper limit which does not depend upon any sample path, so the stability criterion in Definition 5 can be guaranteed.

VI. SMOOTHERS FOR JUMP MARKOV SYSTEMS

In this section, we use a novel idea to compute smoother dynamics without recourse to backward stochastic integrals. This idea was first introduced for finite state process models in [12] and [17], and subsequently, for continuous-state process models in [9].

To compute smoothed estimates of the M-ary detector probabilities, we wish to evaluate the expectation

$$E[\alpha_t X_t' \mid \mathcal{Y}_T], \qquad 0 \leq t \leq T. \tag{6.78}$$

Notation: The Radon–Nikodym derivatives $\Lambda_{0,t}$ and $\Lambda_{t,T}$, are defined, respectively, by

$$\Lambda_{0,t} = \exp\left\{\int_0^t \langle X_u, \boldsymbol{g}\rangle dy_u - \frac{1}{2}\int_0^t \langle X_u, \boldsymbol{g}\rangle^2 du\right\} \quad (6.79)$$

$$\Lambda_{t,T} = \exp\left\{\int_t^T \langle X_u, \boldsymbol{g}\rangle dy_u - \frac{1}{2}\int_t^T \langle X_u, \boldsymbol{g}\rangle^2 du\right\}. \tag{6.80}$$

Note that $\Lambda_{0,T} = \Lambda_{0,t}\Lambda_{t,T}$, for $0 \le t \le T$.

For two matrices A and B of the same dimension $m \times n$, we define an inner product by

$$\langle\langle A, B\rangle\rangle \triangleq \sum_{j=1}^m \sum_{i=1}^n a_{i,j} b_{i,j}. \tag{6.81}$$

Theorem 6: Suppose an observation $y = \{y_u, 0 \le u \le t\}$ is generated by the dynamics at (2.1) and (2.2). Further, suppose that a known collection of candidate parameters for this model, that is, $\{H_1, \ldots, H_m\}$, where $H_j = \{A_{Hj}, \boldsymbol{g}_{H_j}\}$, can switch according to a Markov process α, whose dynamics are given at (4.25). For any t, $0 \le t \le T$, the conditional probability $p(\alpha_t \mid \mathcal{Y}_T) \in \mathbb{R}^m$, is given by

$$p(\alpha_t \mid \mathcal{Y}_T) = \begin{bmatrix} p(\alpha_t = \boldsymbol{f}_1 \mid \mathcal{Y}_T) \\ p(\alpha_t = \boldsymbol{f}_2 \mid \mathcal{Y}_T) \\ \vdots \\ p(\alpha_t = \boldsymbol{f}_m \mid \mathcal{Y}_T) \end{bmatrix}$$

$$= \left\{\frac{\sum_{j=1}^m \sum_{i=1}^n \widehat{z}_t^{(j,i)} \widehat{v}_t^{(j,i)} \boldsymbol{F}_{(j,i)}}{\boldsymbol{1}_n' \left\{\sum_{j=1}^m \sum_{i=1}^n \widehat{z}_t^{(j,i)} \widehat{v}_t^{(j,i)} \boldsymbol{F}_{(j,i)}\right\} \boldsymbol{1}_m}\right\} \boldsymbol{1}_m \tag{6.82}$$

where $\overline{Z}$ satisfies the forward linear ordinary differential equation

$$\frac{d\overline{Z}_t}{dt} = \Phi_t^{-1} \odot B(\Phi_t \odot \overline{Z}_t)$$
$$+ \Phi_t^{-1} \odot \left\{\sum_{j=1}^m \sum_{i=1}^n \boldsymbol{f}_j'(\Phi_t \odot \overline{Z}_t)\boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}'\right\}$$

with

$$\overline{Z}_0 = \Phi_0^{-1} \odot \widehat{Z}_0 = \widehat{Z}_0$$

and the process $\overline{V}$ satisfies the linear ordinary differential equation

$$\frac{d\overline{V}_t}{dt} = -\Phi_t \odot B^*(\Phi_t^{-1} \odot \overline{V}_t)$$
$$- \Phi_t \odot \left\{\sum_{j=1}^m \sum_{i=1}^n \boldsymbol{f}_j'(\Phi_t^{-1} \odot \overline{V}_t)\boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j}\right\}$$

with

$$\overline{V}_T = \Phi_T \odot \widehat{V}_T = \Phi_T \odot \boldsymbol{1}.$$

Remark 7: The dynamics for $\overline{Z}$ and $\overline{V}$, both depend on the observation process y, only through the matrices Φ and Φ^{-1}. Consequently, no stochastic integration is required to compute the smoothed probability $p(\alpha_t \mid \mathcal{Y}_T)$.

Proof of Theorem 6: Write

$$\widehat{Z}_{t,T} \triangleq E^\dagger[\Lambda_{0,T}\alpha_t X_t' \mid \mathcal{Y}_T] \qquad \forall\, t, 0 \le t \le T. \tag{6.83}$$

Using Bayes' rule

$$E[\alpha_t X_t' \mid \mathcal{Y}_T] = \frac{E^\dagger[\Lambda_{0,T}\alpha_t X_t' \mid \mathcal{Y}_T]}{E^\dagger[\Lambda_{0,T} \mid \mathcal{Y}_T]}. \tag{6.84}$$

Using repeated conditioning, we see that

$$\widehat{Z}_{t,T} \triangleq E^\dagger[\Lambda_{0,T}\alpha_t X_t' \mid \mathcal{Y}_T]$$
$$= E^\dagger[\Lambda_{0,t}\alpha_t X_t' \Lambda_{t,T} \mid \mathcal{Y}_T]$$
$$= E^\dagger[\Lambda_{0,t}\alpha_t X_t' E^\dagger[\Lambda_{t,T} \mid \mathcal{Y}_T \vee \mathcal{F}_t] \mid \mathcal{Y}_T]. \tag{6.85}$$

Here, $\mathcal{F}_t = \sigma\{\alpha_u, X_u', 0 \le u \le t\}$.

Define

$$\widehat{v}_{t,T}^{(j,i)} \triangleq E^\dagger\left[\Lambda_{t,T} \mid \mathcal{Y}_T \text{ and } \{\alpha_t = \boldsymbol{f}_j, X_t = \boldsymbol{e}_i\}\right] \tag{6.86}$$

and write

$$\widehat{V}_t \triangleq \left[\widehat{v}_t^{(j,i)}\right]_{\substack{j=1,\ldots,m \\ i=1,\ldots,n}} \in \mathbb{R}^{m \times n}. \tag{6.87}$$

Then

$$\boldsymbol{f}_j' E^\dagger[\Lambda_{0,T}\alpha_t X_t' \mid \mathcal{Y}_T]\boldsymbol{e}_i$$
$$= E^\dagger\Big[\Lambda_{0,t}\boldsymbol{f}_j'\alpha_t X_t'\boldsymbol{e}_i E^\dagger$$
$$\qquad \cdot [\Lambda_{t,T} \mid \mathcal{Y}_T \,\&\, \alpha_t = \boldsymbol{f}_j, X_t = \boldsymbol{e}_i] \mid \mathcal{Y}_T\Big]$$
$$= E^\dagger[\Lambda_{0,t}\langle \boldsymbol{f}_j, \alpha_t\rangle\langle X_t, \boldsymbol{e}_i\rangle \mid \mathcal{Y}_T]\,\widehat{v}_{t,T}^{(j,i)}$$
$$= \boldsymbol{f}_j' E^\dagger[\Lambda_{0,t}\alpha_t X_t' \mid \mathcal{Y}_T]\boldsymbol{e}_i\widehat{v}_{t,T}^{(j,i)}$$
$$= \boldsymbol{f}_j\widehat{Z}_t\boldsymbol{e}_i\widehat{v}_{t,T}^{(j,i)} = \widehat{z}_t^{(j,i)}\widehat{v}_{t,T}^{(j,i)}. \tag{6.88}$$

Therefore

$$E^\dagger[\Lambda_{0,T}\alpha_t X_t' \mid \mathcal{Y}_T] = \sum_{j=1}^m \sum_{i=1}^n \widehat{z}_t^{(j,i)}\widehat{v}_t^{(i,j)} \boldsymbol{F}_{(j,i)}. \tag{6.89}$$

Note that

$$\boldsymbol{1}_m\left\{\sum_{j=1}^m \sum_{i=1}^n \widehat{z}_t^{(i,j)}\widehat{v}_t^{(i,j)} \boldsymbol{F}_{(j,i)}\right\}\boldsymbol{1}_n = \sum_{j=1}^m \sum_{i=1}^n \widehat{z}_t^{(i,j)}\widehat{v}_t^{(i,j)}$$
$$= E^\dagger[\Lambda_{0,T} \mid \mathcal{Y}_T]$$
$$= \langle\langle \widehat{Z}_t, \widehat{V}_t\rangle\rangle. \tag{6.90}$$

It follows from the calculation at (6.90), that the quantity $\langle\langle \widehat{Z}_t, \widehat{V}_t\rangle\rangle$ is invariant with respect to time t, so

$$\frac{d}{dt}\langle\langle \widehat{Z}_t, \widehat{V}_t\rangle\rangle = 0. \tag{6.91}$$

Remark 8: The process $\widehat{V}$ is not a matrix of unnormalized probabilities, rather a quantity that incorporates the extra information obtained from the observations collected from T *back to* t.
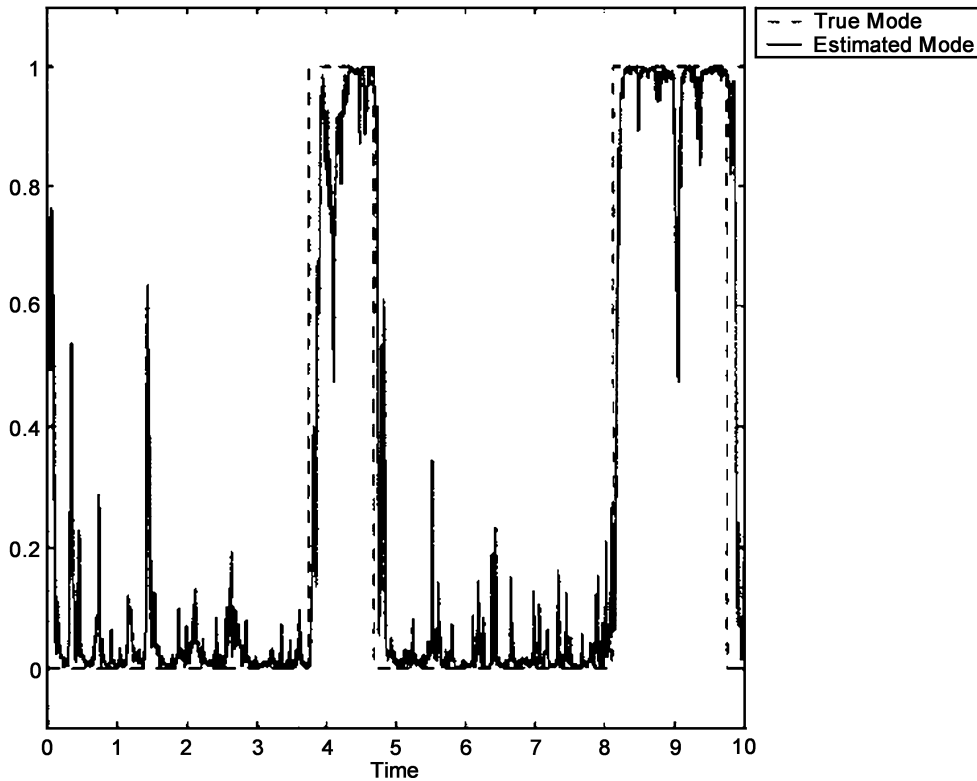
Fig. 1.   True model process, $H_1 = 0$ and $H_2 = 1$, and the estimated mode probability $E[\langle \alpha_t, \boldsymbol{f}_2 \rangle \mid \mathcal{Y}_t]$.

Recall the process $\overline{Z}_t \triangleq \Phi_t^{-1} \odot \widehat{Z}_t$. Define the process

$$\overline{V}_t \triangleq \Phi_t \odot \widehat{V}_t. \tag{6.92}$$

We wish to establish dynamics for $\overline{V}$. By duality, we have

$$\langle\langle \overline{Z}_t, \overline{V}_t \rangle\rangle = \langle\langle \widehat{Z}_t, \widehat{V}_t \rangle\rangle \qquad \forall\, t \in [0, T]. \tag{6.93}$$

Given that the quantity $\langle\langle \widehat{Z}_t, \widehat{V}_t \rangle\rangle$ is invariant to time, we see that

$$\frac{d}{dt}\langle\langle \widehat{Z}_t, \widehat{V}_t \rangle\rangle$$

$$= 0 = \frac{d}{dt}\langle\langle \overline{Z}_t, \overline{V}_t \rangle\rangle$$

$$= \left\langle\left\langle \frac{d\overline{Z}_t}{dt}, \overline{V}_t \right\rangle\right\rangle + \left\langle\left\langle \overline{Z}_t, \frac{d\overline{V}_t}{dt} \right\rangle\right\rangle$$

$$= \left\langle\left\langle \Phi_t^{-1} \odot B(\Phi_t \odot \overline{Z}_t) + \Phi_t^{-1} \right.\right.$$

$$\left.\left. \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}'_j(\Phi_t \odot \overline{Z}_t)\boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A'_{H_j} \right\}, \overline{V}_t \right\rangle\right\rangle$$

$$+ \left\langle\left\langle \overline{Z}_t, \frac{d\overline{V}_t}{dt} \right\rangle\right\rangle$$

$$= \left\langle\left\langle \overline{Z}_t, \Phi_t \odot B^*(\Phi_t^{-1} \odot \overline{V}_t) + \Phi_t \right.\right.$$

$$\left.\left. \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}'_j(\Phi_t^{-1} \odot \overline{V}_t)\boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j} \right\} \right\rangle\right\rangle$$

$$+ \left\langle\left\langle \overline{Z}_t, \frac{d\overline{V}_t}{dt} \right\rangle\right\rangle. \tag{6.94}$$

Therefore

$$\frac{d\overline{V}_t}{dt} = -\Phi_t \odot B^*(\Phi_t^{-1} \odot \overline{V}_t)$$

$$-\Phi_t \odot \left\{ \sum_{j=1}^{m} \sum_{i=1}^{n} \boldsymbol{f}'_j(\Phi_t^{-1} \odot \overline{V}_t)\boldsymbol{e}_i \boldsymbol{F}_{(j,i)} A_{H_j} \right\}. \tag{6.95}$$

$\square$

Remark 9: The algorithm presented in Theorem 6 is quite general and can be used to construct any of the standard smoothing schemes, referred to in the literature as: Fixed point smoothers, fixed lag smoothers and fixed interval smoothers. For example, to compute state estimates for the jump Markov systems studied here, one might construct a fixed lag smoothing scheme and thereby use smoothed/improved model probability estimates to construct state estimates.

Remark 10: The discretization limits ensuring stability, computed in Section V-B, also apply to the corresponding discretization of the dynamics at (6.95).

Remark 11: The contributions of this article are considered, in part, for scenarios with Markov modulated Poisson observations in [13].

VII. EXAMPLE

Our simulation considers a scenario in which two Wonham filter model-parameter sets, jump only in their respective sets of drift coefficients, that is

$$H_1 \triangleq \left\{ A = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix} \quad \boldsymbol{g}_{H_1} = (-5, 0)' \right\} \tag{7.96}$$

$$H_2 \triangleq \left\{ A = \begin{bmatrix} -0.5 & 0.5 \\ 0.5 & -0.5 \end{bmatrix}, \boldsymbol{g}_{H_2} = (7, 10)' \right\}. \quad (7.97)$$

The models H_1 and H_2, jump according to the rate matrix

$$B \triangleq \begin{bmatrix} -0.5 & 0.5 \\ 0.67 & -0.33 \end{bmatrix}. \quad (7.98)$$

Further, this type of scenario corresponds to the dynamics given in Corollary 1. An observation sample path was generated according to the dynamics at (4.31). This path was generated on the time interval $[0, 10]$, with $\Delta = 2^{-7}$. In Fig. 1, we plot an estimated model probability against the true model state, here we show just one of the estimated model probabilities, namely $E[\langle \alpha_t, \boldsymbol{f}_2 \rangle \mid \mathcal{Y}_t]$.

REFERENCES

[1] N. Bergman, A. Doucet, and N. Gordon, "Optimal estimation and the cramér-rao bounds for partial nongaussian state space models," *Ann. Inst. Statist. Math.*, vol. 53, no. 1, pp. 97–112, 2001.

[2] H. A. P. Blom, "An efficient filter for abruptly changing systems," presented at the IEEE Conf. Decision Control, Las Vegas, NV, Dec. 1984.

[3] J. M. C. Clark, "The design of robust approximations to the stochastic differential equations for nonlinear filtering," in *Communications Systems and Random Process Theory*, J. K. Skwirzynski, Ed. Alphen aan den Rijn, The Netherlands: Sijthoff and Noorhoff, 1978, pp. 721–734.

[4] R. J. Elliott, *Stochastic Calculus and Its Applications*. New York: Springer-Verlag, 1982.

[5] ——, "New finite dimensional filters and smoothers for noisily observed markov chains," *IEEE Trans. Inform. Theory*, vol. 39, pp. 265–271, Jan. 1993.

[6] ——, "Exact adaptive filters for markov chains observed in gaussian noise," *Automatica*, vol. 30, no. 9, pp. 1399–1408, Sept. 1994.

[7] R. J. Elliott, L. Aggoun, and J. B. Moore, *Hidden Markov Models Estimation and Control*. New York: Springer-Verlag, 1995, Applications of Mathematics.

[8] R. J. Elliott and W. P. Malcolm, "Reproducing gaussian densities and linear gaussian detection," *Syst. Control Lett.*, no. 40, pp. 133–138, 2000.

[9] ——, "Robust smoother dynamics for poisson processes driven by an itô diffusion," in *IEEE Conf. Decision Control*, Orlando, FL, Dec. 2001.

[10] M. R. James, V. Krishnamurthy, and F. L. Gland, "Time discretization for continuous time filters and smoothers for hmm parameter estimation," *IEEE Trans. Inform. Theory*, vol. 42, pp. 593–605, Mar. 1998.

[11] P. E. Kloeden and E. Platen, *Numerical Solutions for Stochastic Differential Equations*. New York: Springer-Verlag, 1995, Applications of Mathematics.

[12] W. P. Malcolm and R. J. Elliott, "A general smoothing equation for poisson observations," presented at the IEEE Conf. Decision Control, Phoenix, AZ, Dec. 1999.

[13] ——, "Robust detection filters for jump markov systems with doubly stochastic poisson process models," presented at the IEEE Conf. Information Decision Control, Adelaide, South Australia, Feb. 2002.

[14] D. D. Sworder, J. Boyd, and R. J. Elliott, "Modal estimation in hybrid systems," *J. Math. Anal. Applicat.*, no. 245, pp. 225–247, 2000.

[15] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems*. New York: Springer-Verlag, 1985.

[16] W. M. Wonham, "Some applications of stochastic differential equations to nonlinear filtering," *SIAM J. Control Optim.*, no. 2, 1965.

[17] W. P. Malcolm, "Robust filtering and estimation with Poisson observations," Ph.D. dissertation, The Australian National University, Canberra, Australia, 1999.

[18] W. P. Malcolm, R. J. Elliott, and J. van der Hock, "On the numerical stability of time-discretised state estimation via Clark transformations," in *Proc. 42nd IEEE Conf. Decision Control*, vol. 2, Dec. 2003, pp. 1406–1412.

[19] R. J. Elliott, F. Dufour, and D. D. Sworder, "Exact hybrid filters in discrete time," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 1807–1810, Dec. 1996.

Robert J. Elliott received the Bachelors and Masters degrees from Oxford University, Oxford, U.K., and the Ph.D. and D.Sc. degrees from Cambridge University, Cambridge, U.K., in 1961, 1964, 1965, and 1983, respectively.

Currently, he is the Royal Bank Professor of Finance at the University of Calgary, Calgary, AB, Canada, where he is also an Adjunct Professor in both the Department of Mathematics and the Department of Electrical Engineering. He has authored six books and over 300 papers. His work in recent years has investigated stochastic processes in engineering and finance.

W. P. Malcolm received the B.S. degree from the Royal Melbourne Institute of Technology, Melbourne, Australia, the M.S. degree from the Flinders University of South Australia, and the Ph.D. degree from the Australian National University, Canberra, Australia, in 1989, 1995, and 1990, respectively.

He is currently a postdoctoral student with Professor Robert Elliott at the Haskayne School of Business, at the University of Calgary, Calgary, AB, Canada. His research interests include filtering estimation and stochastic control, in particular, filtering and control with point process observations and marked point process observations.