State Estimation in Stochastic Hybrid Systems with Sparse Observations

Eugenio Cinquemani and Mario Micheli

Abstract

In this paper we study the problem of state estimation for a class of sampled-measurement stochastic hybrid systems, where the *continuous* state x satisfies a linear stochastic differential equation, and noisy measurements y are taken at assigned discrete-time instants. The parameters of both the state and measurement equation depend on the *discrete* state q of a continuous-time finite Markov chain. Even in the fault detection setting we consider – at most one transition for q is admissible – the switch may occur *between* two observations, whence it turns out that the optimal estimates cannot be expressed in parametric form and time integrations are unavoidable, so that the known estimation techniques cannot be applied. We derive and implement an algorithm for the estimation of the states x, q and of the discrete-state switching time that is convenient for both recursive update and the eventual numerical quadrature. Numerical simulations are illustrated.

Index Terms

Jump Markov Linear Systems, Bayesian Estimation, Stochastic Hybrid Systems, Kalman Filtering, Fault Detection.

The work of Eugenio Cinquemani was supported in part by the European Community through the RECSYS project of the V Framework Program.

Eugenio Cinquemani is with the Department of Information Engineering, University of Padova, via Gradenigo 6/B, 35131 Padova, Italy (e-mail: fivehand@dei.unipd.it).

Mario Micheli is with the Division of Applied Mathematics, Box F, Brown University, Providence, RI 02912, USA (e-mail: mariom@dam.brown.edu).

I. Introduction

Over the last three decades, a substantial effort has been dedicated to the study of the so-called *Jump Markov Linear Systems* (JMLS). These are linear Gaussian systems switching among a finite number of linear modes indexed by a *discrete state* q. In general, measurements y of the *continuous state* x also depend on q. Switching follows the laws of a Markov chain independent of the initial state $x(0)$ and of the system inputs. Typical linear system estimation problems carry over to JMLS, such as continuous state filtering and prediction. In addition, being the trajectory of q the outcome of an unobserved stochastic process, estimation of q from the available measurements is also a concern. However, for any fixed trajectory of q, the system is linear time-variant. It follows that the statistics of x and y are mixtures of Gaussian distributions, and their optimal (Bayesian) estimates are found by *averaging of conditional Kalman filters*; this also leads to an optimal estimation of state q.

Most of the literature deals with discrete-time JMLS, for which the complexity of the optimal Bayesian estimates of x and q is exponential in time. Thus, a big effort has been devoted to derive effective *suboptimal* algorithms of reasonable (bounded) complexity, usually by eliciting a fixed number of "most likely" discrete state trajectories and obtaining approximate estimates by suitable averaging. In the Generalized Pseudo Bayes approach of Ackerson and Fu [1], approximation is achieved by fitting a Gaussian distribution to the actual distribution of the state. Tugnait [18] uses a Detection-Estimation strategy: at each step, the most probable mode $q(t+1)$ is detected out of the possible transitions of $q(t)$ according to a new measurement of the system's state. Blom & Bar-Shalom [2] keep track of a fixed number of trajectories $q(t)$ by "merging" those that prove "undistinguishable", and pruning the unlikely ones. Sequential Monte Carlo methods are considered in Doucet *et al.* [8], [9], among others, which *explore* at random the space of all possible discrete trajectories based on a convenient generating distribution. Further tecniques are illustrated in the works by Costa [6], [7], Chen & Liu [3], Elliott *et al.* [10], Germani *et al.* [12], and Logothethis & Krishnamurthy [16]. Less attention has been dedicated to the continuous-time counterpart. The interested reader may consult the work by Hibey & Charalambous [13], Hu *et al.* [14], Miller & Runggaldier [17], and Zhang [19].

Both the discrete-time and the continuous-time models are unsatisfactory whenever continuous-time information (estimates) about the system state needs to be drawn from measurements *sampled* at a rate comparable to that of the system dynamics. In the present paper we introduce a model where the continuous state x evolves in continuous time according to a linear stochastic differential equation, whereas noisy measurements are acquired at given discrete time instants. The parameters of both the state equation and the measurement equation depend on a discrete state q which evolves in time as a *continuous-time Markov chain*. Since the discrete-state switch occurs almost surely *between* two successive measurements, ordinary JMLS estimating techniques such as those listed above *cannot* be applied. In fact, since the optimal state estimates will rely on a *continuous* mixture of Gaussian densities the estimation problem cannot be solved in a parametric manner and numerical integrations over time intervals are *not* avoidable. In this work we restrict our attention to a typical fault detection setting where all but one discrete state are absorbing. We formulate a recursive estimation scheme that updates at each step a finite number of parameters and allows to isolate out of the recursion the integral approximation, which is performed at the very end of the computation. The use of such scheme is suitable for the analysis of processes that are subject to sudden changes but observations are relatively sparse. In medical applications, for example, measurements such as blood samples cannot be taken too frequently, however the evolution of a disease or the effect of a therapy need to be closely monitored.

II. Problem Formulation

Let $\mathcal{T} = \{t_k\}_{k\in\mathbb{N}_0}$ (where $\mathbb{N}_0 = \{0, 1, 2, \ldots\}$) be an arbitrary deterministic sequence, with $t_k < t_{k+1}$ and $t_k \to \infty$. Consider a finite *state space* $\mathcal{Q} = \{0, 1, 2, \ldots, N-1\}$ and let q denote its generic element. Assume that we are given matrix functions: $F : \mathcal{Q} \to \mathbb{R}^{n\times n}$, $G : \mathcal{Q} \to \mathbb{R}^{n\times m}$, $H : \mathcal{Q} \to \mathbb{R}^{p\times n}$, and $K : \mathcal{Q} \to \mathbb{R}^{p\times r}$, which assign to each value $q \in \mathcal{Q}$ a four-tuple of matrices (F_q, G_q, H_q, K_q).

Consider the following dynamical model:

$$\begin{cases} \dot{x}(t) &= F_{q(t)}x(t) + G_{q(t)}u(t) \\ y_k &= H_{q(t_k)}x(t_k) + K_{q(t_k)}v_k \end{cases}, \quad t \in \mathbb{R},\ t_k \in \mathcal{T}, \tag{1}$$

where $x : \mathbb{R} \to \mathbb{R}^n$, $y : \mathbb{N}_0 \to \mathbb{R}^p$, are stochastic processes. The equations in (1) are called, respectively, the *state* and *measurement equations*. Continuous-time noise $u(t)$, $t \in \mathbb{R}$, and discrete-time noise v_k, $k \in \mathbb{N}_0$, are zero-mean, normalized, white and Gaussian. We assume that $\{u(t)\}_{t\in\mathbb{R}}$, $\{v_k\}_{k\in\mathbb{N}_0}$ and the initial condition $x(t_0) \sim \mathcal{N}\big(\mu_0, \Sigma_0\big)$ are mutually independent; μ_0 and Σ_0 are given. Furthermore, we shall assume that $q(t)$, $t \in \mathbb{R}$, is a continuous-time, homogeneous *Markov chain* (independent of $\{u(t)\}$, $\{v_k\}$ and x_0) with assigned transition probabilities $T_{i,j}(\delta) \triangleq \mathbb{P}[q(t+\delta) = j \,|\, q(t) = i]$, $i, j \in \mathcal{Q}$, and initial probabilities $p_q \triangleq \mathbb{P}[q(t_0) = q]$, $q \in \mathcal{Q}$.

Our problem is the following: given measurements up to time t_k, that is $y^k \triangleq \{y_0, \ldots, y_k\}$, we wish to compute the least-squares estimate of the *continuous* state $x(t_\ell)$:

$$\hat{x}^a_{\ell|k} \triangleq \arg \min_{g(y^k): g\in\mathcal{M}} \mathbb{E}\left[||g(y^k) - x(t_\ell)||^2\right] = \mathbb{E}\big[x(t_\ell) \,\big|\, y^k\big], \tag{2}$$

(where $\mathcal{M}$ is the set of measurable functions $g : \mathbb{R}^{(k+1)p} \to \mathbb{R}^n$) and at the same time the *a posteriori* probability distribution of the *discrete* state:

$$p_{\ell|k}(q) \triangleq \mathbb{P}\big[q(t_\ell) = q \,\big|\, y^k\big]\,. \tag{3}$$

We will mostly restrict our attention to the cases $\ell = k$ (filtering) and $\ell = k+1$ (prediction). Estimate $\hat{x}^a$ is, as we shall see, a weighed *average* of a *continuum* of Bayesian estimates for different linear stochastic systems —hence the superscript "*a*".

According to our model the discrete state switches among different values in $\mathcal{Q}$ *between* two successive measurements with probability one —in principle, even more than once: this makes exact estimation a formidable task. In the present paper, however, we shall limit ourselves to a *fault detection* setting: i.e. we shall assume that nonzero states in $\mathcal{Q}$ are *absorbing*. This induces a constraint on the structure of the $N \times N$ transition probability matrix $T(\delta) = [T_{i,j}(\delta)]$, which

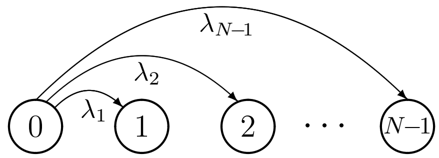

Fig. 1. Graphical representation of Markov Process (4).

takes the form:

$$T(\delta) = \left[\begin{array}{c|ccc} e^{-\lambda\delta} & \cdots & \frac{\lambda_j}{\lambda}\left(1 - e^{-\lambda\delta}\right) & \cdots \\ \hline 0 & & I_{N-1} & \end{array}\right] \tag{4}$$

where $\{\lambda_1, \ldots, \lambda_{N-1}\}$ is a given set of positive parameters, $\lambda \triangleq \sum_{i=1}^{N-1} \lambda_i$, and I_{N-1} is the $(N-1)$-dimensional identity matrix. See Figure 1 for the graphical representation of such Markov process.

Given the restriction above, trajectory $q(t)$ is *characterized* by the *switching time* $t^\star$ (i.e. the time at which the event takes place) and the *final discrete state* $q^\star$. The joint probability distribution of $t^\star$ and $q^\star$ is given by $F_{t^\star,q^\star}(t,q) \triangleq \mathbb{P}[t^\star \leq t, q^\star = q] = \frac{\lambda_q}{\lambda}(1 - e^{-\lambda t})p_0 + p_q$ for $q = 1, \ldots, N-1$ and $t \geq 0$; it is zero for $q = 0$ and is undefined for $t < 0$. If we introduce the further simplifying hypothesis that $p_0 = 1$, i.e. that the discrete state q is initially zero, then the switching time becomes exponentially distributed, $t^\star \sim \mathcal{E}(\lambda)$ and random variables $t^\star$ and $q^\star$ are in fact *independent*. The assumption $p_0 = 1$ will be maintained throughout the rest of the paper. In addition to the problem of computing the state estimates (2) and (3), we shall consider the problem of computing the least-squares estimate of the switching time:

$$\hat{t}_k^\star \triangleq \arg \min_{h(y^k):h\in\mathcal{M}} \mathbb{E}\left[||h(y^k) - t^\star||^2\right] = \mathbb{E}[t^\star|y^k]. \tag{5}$$

Given the continuous nature of random variable $t^\star$, this problem is peculiar of the sampled-measurement model we are considering and has no equivalent in the context of discrete-time jump Markov systems. One key ingredient that we shall need to solve our problem will be

$$f_{t^\star,q^\star|y^k}(t,q) \triangleq \frac{\mathrm{d}}{\mathrm{d}t}\,\mathbb{P}\big[\,t^\star \leq t,\, q^\star = q \,\big|\, y^k\,\big]. \tag{6}$$

In fact one may easily compute $p_{\ell|k}(\cdot)$ and $\hat{t}_k^\star$ from the above density. Moreover, we will see in the following section that estimates (2) can be computed as the weighted average of linear Bayesian estimates, where the weight is given precisely by density (6). We shall denote this quantity simply by $f(t^\star, q^\star|y^k)$. Similarly, we shall write $f(t^\star, q^\star)$ to denote $\frac{\mathrm{d}}{\mathrm{d}t}F_{t^\star,q^\star}(t,q) = \lambda_q e^{-\lambda t}$.

III. The conditioned Kalman filtering approach

Consider the computation of $\hat{x}^a_{\ell|k}$. Application of the total probability law to equation (2) yields

$$\hat{x}^a_{\ell|k} = \sum_{q^\star=1}^{N-1} \int_0^{+\infty} \hat{x}_{\ell|k}(t^\star, q^\star) f(t^\star, q^\star|y^k)\mathrm{d}t^\star, \tag{7}$$

with $\hat{x}_{\ell|k}(t^\star, q^\star) \triangleq \mathbb{E}[x(t_\ell)|y^k, t^\star, q^\star]$. By Bayes'rule, $f(t^\star, q^\star|y^k) = f(y^k|t^\star, q^\star)f(t^\star, q^\star)/f(y^k)$, with

$$f(y^k) = \sum_{q^\star=1}^{N-1} \int_0^{+\infty} f(y^k|t^\star, q^\star) f(t^\star, q^\star)\mathrm{d}t^\star. \tag{8}$$

In turn, $f(y^k|t^\star, q^\star)$ may be computed by the recursion $f(y^{k+1}|t^\star, q^\star) = f(y_{k+1}|y^k, t^\star, q^\star)f(y^k|t^\star, q^\star)$, initialized by $f(y^0|t^\star, q^\star) = f(y_0)$, which is purely Gaussian. The computation of $\hat{t}_k^\star$ and of $p_{\ell|k}(q)$ also follows from that of $f(t^\star, q^\star|y^k)$ by way of suitable integrations. Therefore, computing estimates of the state as well as of the switching time amounts to computing $\hat{x}_{\ell|k}(t^\star, q^\star)$ and $f(y_{k+1}|y^k, t^\star, q^\star)$.

Consider the model obtained from (1) by *fixing* the trajectory $q(\cdot)$ according to given values of $t^\star$ and $q^\star$. According to this model, which will be called the *conditioned* model, the optimal estimator of $x(t_\ell)$ given measurements y^k is precisely $\hat{x}_{\ell|k}(t^\star, q^\star)$. Since the conditioned model is linear and Gaussian, $\hat{x}_{\ell|k}(t^\star, q^\star)$ may be computed by a Kalman recursion matched to the specific system parameters [15]. In addition,

$$f(y_{k+1}|y^k, t^\star, q^\star) = \mathcal{N}(\hat{y}_{k+1|k}(t^\star, q^\star), \Lambda_{k+1}(t^\star, q^\star)), \tag{9}$$

where $\hat{y}_{k+1|k}(t^\star, q^\star)$ is the Kalman predictor for the model conditioned on $t^\star$ and $q^\star$ and $\Lambda_{k+1}(t^\star, q^\star)$ is the corresponding innovations variance.

Due to the continuous nature of random variable $t^\star$, integrations are *not* avoidable. In general, they cannot be solved analytically. Therefore, a key issue is to express the conditioned estimates $\hat{x}_{\ell|k}(t^\star, q^\star)$, $\hat{y}_{k+1|k}(t^\star, q^\star)$ and $\Lambda_{k+1}(t^\star, q^\star)$ in a way convenient for both update and numerical integration. Specifically, we wish to carry out the integrations by an adaptive quadrature approach [11], such as Simpson's method. The idea is to evaluate the integrand on a grid of points which is adapted to the shape of the function by an iterative procedure. Based on this strategy, arbitrarily low integration errors may usually be achieved at the cost of an increasing number of iterations. In the present context, adaptation implies evaluating $\hat{x}_{\ell|k}(t^\star, q^\star)$, $\hat{y}_{k+1|k}(t^\star, q^\star)$ and $\Lambda_{k+1}(t^\star, q^\star)$ on a set of values of $t^\star$ which depends on the data and is different at every k. In this case, the use of standard Kalman filtering is largely inefficient, because a different matched filter needs to be run for every different value of $t^\star$. Instead, parametric expressions of the conditioned estimates should be used, allowing for both *explicit* evaluation at arbitrary values of $t^\star$ and recursive parameter update. Formulae as such will be derived in the next sections. In order to achieve this, we shall interpret $\hat{x}_{\ell|k}(t^\star, q^\star)$, $\hat{y}_{k+1|k}(t^\star, q^\star)$ and $\Lambda_{k+1}(t^\star, q^\star)$ in terms of the Kalman filter designed for a discrete-time version of the conditioned model. The properties of the discrete-time conditioned system will be used to rewrite the Kalman recursions so to make the dependence on $t^\star$ explicit. Algorithms for the computation of $\hat{x}^a_{\ell|k}$, $p_{\ell|k}$ and $f(t^\star|y^k)$ will follow.

IV. Explicit expressions for the conditioned Kalman filter

Fix $q^\star$, an index h and a value of $t^\star \in (t_h, t_{h+1})$. For $x_k \triangleq x(t_k)$, let us write the conditioned system in the form of a discrete-time system:

$$\left\{ \begin{array}{rcl} x_{k+1} & = & A_k(t^\star, q^\star)\, x_k \;+\; u_k \\ y_k & = & C_k(t^\star, q^\star)\, x_k \;+\; D_k(t^\star, q^\star)\, v_k \end{array} \right. \tag{10}$$

with $\mathrm{Var}(u_k) = Q_k(t^\star, q^\star)$, $\{u_k\}$ white and independent of $\{v_k\}$ and x_0. We wish to choose the parameters of (10) so to preserve the joint statistical description of the measurements and the state of (1) at sample times. Integrating the state equation of (1) along the trajectory $q(t)$ fixed by $t^\star$ and $q^\star$, one gets the following [4]. For $k < h$ (i.e. $t^\star > t_{k+1}$):

$$A_k(t^\star, q^\star) = \widehat{A}_{k,0}, \qquad Q_k(t^\star, q^\star) = \widehat{Q}_{k,0},$$
$$C_k(t^\star, q^\star) = H_0, \qquad D_k(t^\star, q^\star) = K_0;$$

for $k > h$ (i.e. $t^\star < t_k$):

$$A_k(t^\star, q^\star) = \widehat{A}_{k,q^\star}, \qquad Q_k(t^\star, q^\star) = \widehat{Q}_{k,q^\star},$$
$$C_k(t^\star, q^\star) = H_{q^\star}, \qquad D_k(t^\star, q^\star) = K_{q^\star};$$

for $k = h$ (i.e. $t_k < t^\star < t_{k+1}$):

$$A_k(t^\star, q^\star) = \widetilde{A}_k(t^\star, q^\star)\widetilde{A}_k(t^\star, 0), \qquad Q_k(t^\star, q^\star) = \widetilde{A}_k(t^\star, q^\star)\widetilde{Q}_k(t^\star, 0)\widetilde{A}_k(t^\star, q^\star) + \widetilde{Q}_k(t^\star, q^\star),$$
$$C_k(t^\star, q^\star) = H_0, \qquad D_k(t^\star, q^\star) = K_0,$$

with $\widehat{A}_{k,q} \triangleq e^{F_q(t_{k+1}-t_k)}$, while $\widetilde{A}_k(t^\star, q) \triangleq e^{F_0(t^\star - t_k)}$ for $q = 0$ and $\widetilde{A}_k(t^\star, q) \triangleq e^{F_q(t_{k+1}-t^\star)}$ for $q \neq 0$. If the spectrum of F_q is *non-mixing* (i.e. the spectra of F_q and $-F_q$ are disjoint), then $\widehat{Q}_k(t^\star, q) = J_q - \widehat{A}_{k,q} J_q \widehat{A}_{k,q}^T$ and $\widetilde{Q}_k(t^\star, q) = J_q - \widetilde{A}_k(t^\star, q) J_q \widetilde{A}_k^T(t^\star, q)$, where J_q is the *unique* (symmetric) solution of Lyapunov equation $F_q J_q + J_q F_q^T = -G_q G_q^T$. Otherwise, $\widehat{Q}_k(t^\star, q)$ and $\widetilde{Q}_k(t^\star, q)$ must be computed by numerical integration; this instance is rather pathological an will not be discussed here, see instead [4]. The conditioned estimates of section III may now be expressed in terms of the following discrete-time Kalman equations in which, in principle, all quantities depend on h, $t^\star$ and $q^\star$:

$$\begin{aligned} \hat{x}_{k|k} &= \hat{x}_{k|k-1} + P_{k|k-1} C_k^T \Lambda_k^{-1} [y_k - C_k \hat{x}_{k|k-1}] \\ P_{k|k} &= P_{k|k-1} - P_{k|k-1} C_k^T \Lambda_k^{-1} C_k P_{k|k-1} \end{aligned} \tag{11}$$

$$\begin{aligned} \hat{x}_{k+1|k} &= A_k \hat{x}_{k|k} \\ P_{k+1|k} &= A_k P_{k|k} A_k^T + Q_k \end{aligned} \tag{12}$$

with $\Lambda_k = C_k P_{k|k-1} C_k^T + D_k D_k^T$ and $\hat{y}_{k|k-1} = C_k \hat{x}_{k|k-1}$. Initializations are as follows: $\hat{x}_{0|-1} = x_0$ and $P_{0|-1} = P_0$. Note that the above equations are *exact*, in the sense that no approximation has been introduced in the discretization process. However, one notes that A_k and Q_k are *constant* w.r.t. $t^\star$ for $k \neq h$. Similarly, C_k, D_k never depend on $t^\star$. On the contrary, for $k = h$, A_k and Q_k depend on the specific value of $t^\star$. Therefore, for $k < h$, the conditioned estimates do not depend on $t^\star$; for $k = h$, the value of $t^\star$ plays a role in the time update (12) only; and finally, for $k > h$, their expressions depend on $t^\star$ *only through the values of* $\hat{x}_{h+1|h}$ and $P_{h+1|h}$. This analysis translates to the following statements. Let superscript "^{0}" denote the estimates computed by running equations (11)÷(12) with parameters $A_k = \widehat{A}_{k,0}$, $Q_k = \widehat{Q}_{k,0}$, $C_k = H_0$, $D_k = K_0$, $\forall k$ (i.e. as if q never switched). These quantities do not depend on h, $t^\star$ or $q^\star$.

Proposition 1: Assume that $t^\star \in (t_h, t_{h+1})$. Then it holds that

$$\hat{x}_{k|k}(t^\star, q^\star) = \hat{x}^0_{k|k}, \qquad P_{k|k}(t^\star, q^\star) = P^0_{k|k}, \qquad k \leq h,$$
$$\hat{x}_{k+1|k}(t^\star, q^\star) = \hat{x}^0_{k+1|k}, \qquad P_{k+1|k}(t^\star, q^\star) = P^0_{k+1|k}, \qquad k < h.$$

As a consequence, for $k < h$, $\Lambda_{k+1}(t^\star, q^\star) = \Lambda^0_{k+1}$ and $\hat{y}_{k+1|k}(t^\star, q^\star) = \hat{y}^0_{k+1|k} = H_0 \hat{x}^0_{k+1|k}$. Next, denote with a superscript "†" the estimates obtained by running equations (11)÷(12) for $k \geq h+1$ with parameters $A_k = \widehat{A}_{k,q^\star}$, $Q_k = \widehat{Q}_{k,q^\star}$, $C_k = H_{q^\star}$, $D_k = K_{q^\star}$, initialized by $\hat{x}_{h+1|h} = 0$ and $P_{h+1|h} = 0$. Let $\Phi^\dagger_k \triangleq \widehat{A}_{k,q^\star}(I - P^\dagger_{k|k-1} H^T_{q^\star} (\Lambda^\dagger_k)^{-1} H_{q^\star})$ be the associated predictor transition matrix [15]. These quantities depend on h and $q^\star$, but not on $t^\star$. Also define $\Delta_{q^\star} \triangleq H^T_{q^\star}(K_{q^\star} K^T_{q^\star})^{-1} H_{q^\star}$.

Proposition 2: Assume that $t^\star \in (t_h, t_{h+1})$. For $k \geq h+1$, it holds that:

$$\hat{x}_{k|k}(t^\star, q^\star) = \hat{x}^\dagger_{k|k} + \widehat{A}^{-1}_{k,q^\star} U_k \{I + S_k P_{h+1|h}(t^\star, q^\star)\}^{-T} \{\hat{x}_{h+1|h}(t^\star, q^\star) + P_{h+1|h}(t^\star, q^\star) M_k\}, \tag{13}$$

$$P_{k|k}(t^\star, q^\star) = P^\dagger_{k|k} + \widehat{A}^{-1}_{k,q^\star} U_k \{I + S_k P_{h+1|h}(t^\star, q^\star)\}^{-T} P_{h+1|h}(t^\star, q^\star) U^T_k \widehat{A}^{-T}_{k,q^\star}, \tag{14}$$

$$\hat{x}_{k+1|k}(t^\star, q^\star) = \hat{x}^\dagger_{k+1|k} + U_k \{I + S_k P_{h+1|h}(t^\star, q^\star)\}^{-T} \{\hat{x}_{h+1|h}(t^\star, q^\star) + P_{h+1|h}(t^\star, q^\star) M_k\}, \tag{15}$$

$$P_{k+1|k}(t^\star, q^\star) = P^\dagger_{k+1|k} + U_k \{I + S_k P_{h+1|h}(t^\star, q^\star)\}^{-T} P_{h+1|h}(t^\star, q^\star) U^T_k, \tag{16}$$

where U_k, S_k and M_k obey the recursions $U_k = \Phi^\dagger_k U_{k-1}$, $S_k = U^T_{k-1}(I + \Delta_{q^\star} P^\dagger_{k|k-1})^{-1} \Delta_{q^\star} U_{k-1} + S_{k-1}$ and $M_k = U^T_{k-1}(H^T_{q^\star}(K_{q^\star} K^T_{q^\star})^{-1} y_k - (I + \Delta_{q^\star} P^\dagger_{k|k-1})^{-1} \Delta_{q^\star} \widetilde{M}_k) + M_{k-1}$, initialized by $U_h = I$, $S_h = 0$ and $M_h = 0$. In turn, $\widetilde{M}_k = \Phi^\dagger_{k-1} \widetilde{M}_{k-1} + P^\dagger_{k|k-1} H^T_{q^\star}(K_{q^\star} K^T_{q^\star})^{-1} y_k$, with $\widetilde{M}_h = 0$.

Proposition 2 relies on algebraic results discussed in [5]. In (13)÷(14), the inversion of matrix $\widehat{A}_{k,q^\star}$ may be avoided thanks to the equality $\widehat{A}_{k,q^\star}^{-1} U_k = (I - P_{k|k-1}^\dagger C_k^T (\Lambda_k^\dagger)^{-1} C_k) U_{k-1}$. Quantities U_k, M_k, S_k depend on h and $q^\star$ but not on $t^\star$. Therefore equations (13)÷(16) provide formulae for the conditioned estimates in terms of $\hat{x}_{h+1|h}(t^\star, q^\star)$, $P_{h+1|h}(t^\star, q^\star)$. Explicit formulae in terms of $t^\star$ are then obtained by means of relations (12) evaluated at $k = h$, namely:

$$\begin{aligned}
\hat{x}_{h+1|h}(t^\star, q^\star) &= \widetilde{A}_k(t^\star, q^\star) \widetilde{A}_k(t^\star, 0) \hat{x}^0_{h|h}, \\
P_{h+1|h}(t^\star, q^\star) &= \widetilde{A}_k(t^\star, q^\star) \big\{ \widetilde{A}_k(t^\star, 0) P^0_{h|h} \widetilde{A}_k(t^\star, 0)^T + \widetilde{Q}_k(t^\star, 0) \big\} \widetilde{A}_k(t^\star, q^\star)^T + \widetilde{Q}_k(t^\star, q^\star).
\end{aligned}$$

Finally, for every $q^\star$, every h and all $t^\star \in (t_h, t_{h+1})$, an explicit expression for $f(y_{k+1}|y^k, t^\star, q^\star)$ follows from those of $\hat{y}_{k+1|k}(t^\star, q^\star)$ and $\Lambda_{k+1}(t^\star, q^\star)$ by means of equation (9). In particular, by Proposition 1, it is constant w.r.t. to $t^\star$ for $h \geq k+1$. Hence $f(y_{k+1}|y^k, t^\star, q^\star) = f^0(y_{k+1}|y^k)$ for $t^\star > t_{k+1}$.

V. Computation of the estimates

For the sake of clarity, let us begin with the computation of the normalization factor (8). Fix $k \in \mathbb{N}_0$. The recursive computation of $f(y^k|t^\star, q^\star)$ yields $f(y^k|t^\star, q^\star) = \prod_{i=0}^{k-1} f(y_{i+1}|y^i, t^\star, q^\star) \cdot f(y_0|t^\star, q^\star)$, where $f(y_0|t^\star, q^\star) = f(y_0)$ does not depend on $t^\star$ and $q^\star$. Following on the previous section, for every $h = 0, \ldots, k-1$, consider the piece of integration where $t^\star \in (t_h, t_{h+1})$. For $i = 0, \ldots, h-1$, the term $f(y_{i+1}|y^i, t^\star, q^\star)$ is constant w.r.t. $t^\star$ and $q^\star$, being determined by $\hat{x}^0_{i+1|i}$ and $P^0_{i+1|i}$. For $i = h, \ldots, k-1$, it is expressed in terms of the parameters $\hat{x}^\dagger_{i+1|i}(h, q^\star)$, $P^\dagger_{i+1|i}(h, q^\star)$, $S_i(h, q^\star)$, $M_i(h, q^\star)$ and $U_i(h, q^\star)$. Next consider the piece of integration where $t^\star \in (t_k, +\infty)$. All terms $f(y_{i+1}|y^i, t^\star, q^\star)$, $i = 0, \ldots, k-1$ are determined by $\hat{x}^0_{i+1|i}$ and $P^0_{i+1|i}$, i.e. they do not depend on $t^\star$ and $q^\star$. Therefore, we shall rewrite the integration in (8) as

$$\sum_{h=0}^{k-1} \left\{ f^0_h \cdot \int_{t_h}^{t_{h+1}} \Big(\prod_{i=h}^{k-1} f(y_{i+1}|y^i, t^\star, q^\star) \Big) f(t^\star, q^\star) \mathrm{d}t^\star \right\} + f^0_k \cdot \int_{t_k}^{+\infty} f(t^\star, q^\star) \mathrm{d}t^\star, \tag{17}$$

where, for a generic index j, $f^0_j \triangleq \prod_{i=0}^{j-1} f^0(y_{i+1}|y^i) f(y_0)$. The rightmost integral has a simple analytic solution. To solve the integrations over (t_h, t_{h+1}), $h = 0, \ldots, k-1$, a routine is set up which takes fixed values of $t^\star \in (t_h, t_{h+1})$ and parameters $\hat{x}^\dagger_{i+1|i}(h,q^\star)$, $P^\dagger_{i+1|i}(h,q^\star)$, $S_i(h,q^\star)$, $M_i(h,q^\star)$ and $U_i(h,q^\star)$ as inputs and evaluates the integrand based on the explicit formulae developed above. This routine shall be invoked repeatedly by an adaptive quadrature algorithm so to achieve a desired accuracy. At step k, the overall computation requires the knowledge of $\hat{x}^\dagger_{i+1|i}(h,q^\star)$, $P^\dagger_{i+1|i}(h,q^\star)$, $S_i(h,q^\star)$, $M_i(h,q^\star)$ and $U_i(h,q^\star)$, for $i = h,\ldots,k-1$ and $h = 0,\ldots,k-1$; and of f^0_j, for $j = 0,\ldots,k$. For increasing values of k, these parameters are updated by Algorithm 1 below, which uses the formulae of section IV and the recursion $f^0_k = f^0(y_k|y^{k-1})f^0_{k-1}$.

Algorithm 1 Iterative parameter update

{Initialization}
$\hat{x}^0_{0|-1} \Leftarrow \mu_0$
$P^0_{0|-1} \Leftarrow P_0$
$f^0_{-1} \Leftarrow 1$
$k \Leftarrow 0$
loop {Iterations}
  compute $f^0(y_k|y^{k-1})$ from $\hat{x}^0_{k|k-1}$, $P^0_{k|k-1}$
  compute f^0_k from f^0_{k-1}, $f^0(y_k|y^{k-1})$
  {Measurement update}
  for $h = 0,\ldots,k-1$ **and all** $q^\star$ **do**

compute :	from :
$\hat{x}^\dagger_{k\|k}(h,q^\star)$	$\hat{x}^\dagger_{k\|k-1}(h,q^\star), P^\dagger_{k\|k-1}(h,q^\star)$
$P^\dagger_{k\|k}(h,q^\star)$	$P^\dagger_{k\|k-1}(h,q^\star)$
$U_k(h,q^\star)$	$U_{k-1}(h,q^\star)$
$S_k(h,q^\star)$	$S_{k-1}(h,q^\star), U_{k-1}(h,q^\star), \ldots \ldots P^\dagger_{k\|k-1}(h,q^\star)$
$\widetilde{M}_k(h,q^\star)$	$\widetilde{M}_{k-1}(h,q^\star), P^\dagger_{k\|k-1}(h,q^\star)$
$M_k(h,q^\star)$	$M_{k-1}(h,q^\star), \widetilde{M}_{k-1}(h,q^\star), \ldots \ldots U_{k-1}(h,q^\star), P^\dagger_{k\|k-1}(h,q^\star)$

  end for
  compute $\hat{x}^0_{k|k}$ from $\hat{x}^0_{k|k-1}$, $P^0_{k|k-1}$
  compute $P^0_{k|k}$ from $P^0_{k|k-1}$
  {Time update}
  for all $h = 0,\ldots,k-1$ **and all** $q^\star$ **do**
    compute $\hat{x}^\dagger_{k+1|k}(h,q^\star)$ from $\hat{x}^\dagger_{k|k}(h,q^\star)$
    compute $P^\dagger_{k+1|k}(h,q^\star)$ from $P^\dagger_{k|k}(h,q^\star)$
  end for
  compute $\hat{x}^0_{k+1|k}$ from $\hat{x}^0_{k|k}$
  compute $P^0_{k+1|k}$ from $P^0_{k|k}$
  {Initialization of the next step}
  for all $q^\star$ **do**
    $\hat{x}^\dagger_{k+1|k}(k,q^\star) \Leftarrow 0$
    $P^\dagger_{k+1|k}(k,q^\star) \Leftarrow 0$
    $U_k(k,q^\star) \Leftarrow I$
    $S_k(k,q^\star) \Leftarrow 0$
    $\widetilde{M}_k(k,q^\star) \Leftarrow 0$
    $M_k(k,q^\star) \Leftarrow 0$
  end for
  $k \Leftarrow k+1$
end loop

The computation of $\hat{x}^a_{\ell|k}$, $p_{\ell|k}$ and $\mathbb{E}[t^\star|y^k]$ is set up in a similar way. For instance, for $\ell = k$, the integration in (7) is recast as the following formula, which must be normalized by $f(y^k)$:

$$\sum_{h=0}^{k-1}\left\{f^0_h \cdot \int_{t_h}^{t_{h+1}} \hat{x}_{k|k}(t^\star,q^\star)\Big(\prod_{i=h}^{k-1} f(y_{i+1}|y^i,t^\star,q^\star)\Big)f(t^\star,q^\star)\mathrm{d}t^\star\right\} + \hat{x}^0_{k|k}f^0_k \cdot \int_{t_k}^{+\infty} f(t^\star,q^\star)\mathrm{d}t^\star.$$

VI. NUMERICAL EXAMPLE

We will now show numerical results concerning a specific example. We will compute probability $p_{k|k}(\cdot)$ and estimates $\hat{t}_k^\star$ and $\hat{x}^a_{k|k}$, as functions of index k. Let $\mathcal{Q} = \{0, 1, 2\}$. Consider system (1) with $t_k \triangleq k \cdot T$, $T = 0.5$, and parameters $\mu_0 = 0$, $\Sigma_0 = 0.1 \cdot I_{2\times 2}$. We chose 4-tuples (F_q, G_q, H_q, K_q) to be

$$\left(\begin{bmatrix} -0.4 & 0.6 \\ c_q & -0.5 \end{bmatrix}, \begin{bmatrix} 0.2 & 0.1 \\ 0.1 & 0.2 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0.1 & 0 \\ 0 & 0.1 \end{bmatrix} \right)$$

where $c_0 = 0$, $c_1 = 1$, $c_2 = -2$. That is, only the state evolution matrix changes with q. This modifies the character of the continuous-time system from stable ($q = 0$, stable node) to unstable ($q = 1$, saddle) or oscillatory ($q = 2$, stable focus), according to the different spectra $\sigma(F_0) = \{-0.4, -0.5\}$, $\sigma(F_1) = \{-1.22, 0.32\}$, $\sigma(F_2) = \{-0.45 \pm i\,1.09\}$. In this setting y is a noisy version of the state x. As usual we assume that $p_0 = 1$; switching intensities are fixed to $\lambda_1 = 0.06$, $\lambda_2 = 0.08$. With this choice, $\mathbb{E}[t^\star] = 1/\lambda \simeq 7.14$, $\mathbb{P}[q^\star = 1] \simeq 0.43$, $\mathbb{P}[q^\star = 2] \simeq 0.57$. We started off the simulations from $x(0) = 0$. We then randomly generated x and y up to time $k_{\text{max}} \cdot T$, with $k_{\text{max}} = 30$, for a jump of $q(t)$ occurring at time $\bar{t} = 5.25 < \mathbb{E}[t^\star]$; we considered both $\bar{q} = 1$ and $\bar{q} = 2$ as final discrete states ($\bar{t}$ and $\bar{q}$ are the *sample values* of random variables $t^\star$ and $q^\star$). The values assumed by $t^\star$ (i.e. $\bar{t}$) and $q^\star$ (i.e. $\bar{q} = 1$ and $\bar{q} = 2$, in turn) have been chosen manually by the programmer; note also the relative sparseness of measurements. Computations were carried out in MATLAB. Function `quad` – implementing Simpson's adaptive quadrature – was used to solve numerical integrations.

Figure 2 shows, for different values of k, the *a posteriori* density $f(t^\star|y^k)$ for the cases $\bar{q} = 1$ (left) and $\bar{q} = 2$ (right). The evolution from the exponential prior to a density roughly concentrated around the true switching instant may be observed. The evolution of estimate $\hat{t}_k^\star$ and posterior probability $p_{k|k}(q)$ are reported in Figure 3 for $\bar{q} = 1$ and 2. One may note that, even before the switch happens, $p_{k|k}(0)$ adjusts to values that are less than one, due to fluctuations of the state x around zero. Also, for $\bar{q} = 1$, $\hat{t}_k^\star$ grows in a quasi-linear fashion: this is due to the memoryless nature of random variable $t^\star$. In fact one may prove the following bound [4]:

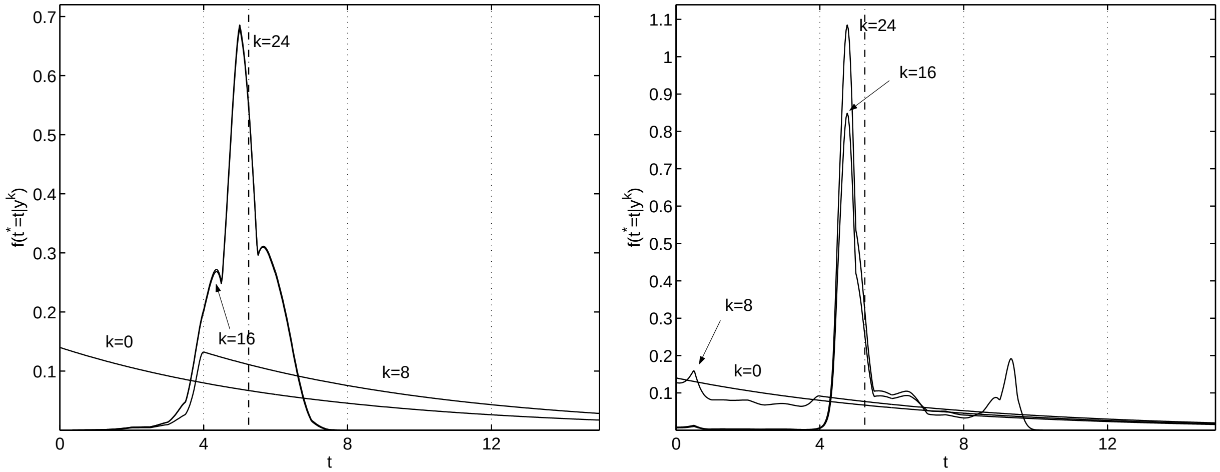

Fig. 2. Density function $f(t^\star|y^k)$ plotted for $k = 0, 8, 16$ and 24. Left: $\bar{q} = 1$; Right: $\bar{q} = 2$. Dash-dotted lines mark the *actual* switching time $\bar{t} = 5.25$.

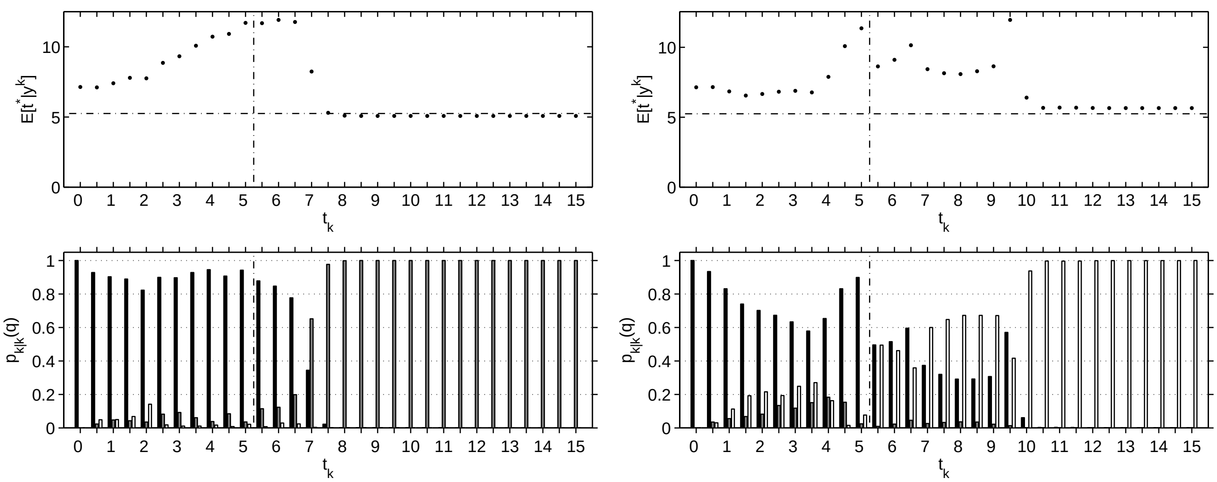

Fig. 3. Evolution of the estimate $\hat{t}_k^\star$ (above) and of the probability function $p_{k|k}(q)$ (below; left bar: $p_{k|k}(0)$, center bar: $p_{k|k}(1)$; right bar: $p_{k|k}(2)$). Left: $\bar{q} = 1$; Right: $\bar{q} = 2$. Dash-dotted lines mark the *actual* switching time $\bar{t} = 5.25$.

$p_{k|k}(0)/\lambda + p_{k|k}(0) \cdot t_k \leq \hat{t}_k^\star \leq p_{k|k}(0)/\lambda + t_k$. Whence the larger the value of $p_{k|k}(0)$ is, the tighter the bounds will be, as in the case $\bar{q} = 1$. After the switch occurs a few measurements suffice to both detect the new discrete state $\bar{q}$ and estimate the switching time $\bar{t}$. However, comparing the plots suggests that is is relatively easier to detect a switch from $q = 0$ (stable) to $\bar{q} = 1$ (unstable) rather than to $\bar{q} = 1$ (stable), which keeps state x close to zero. In general, the more "different" the modes are, the quicker the algorithm is to detect the switch.

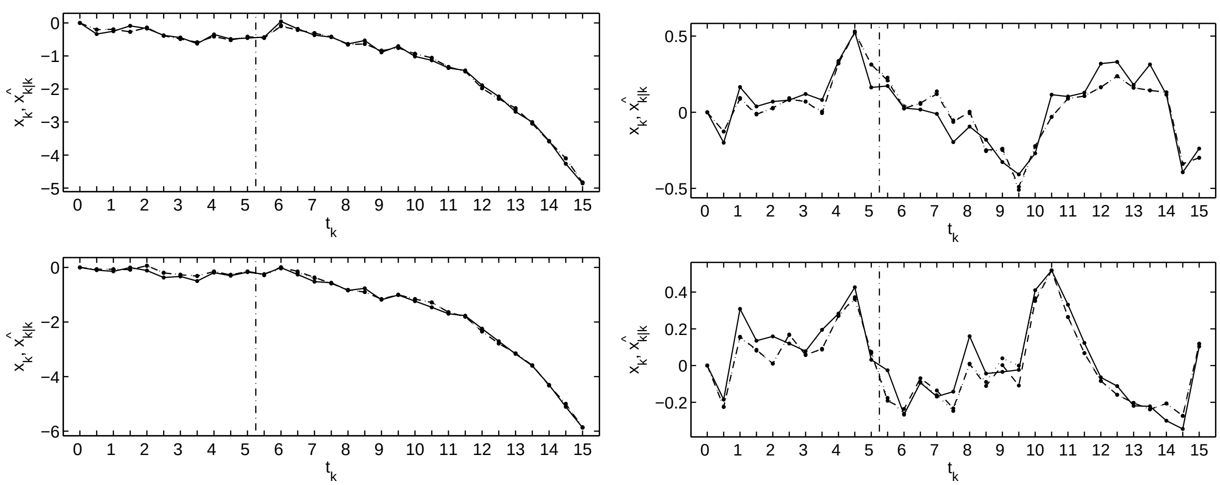

Fig. 4. Evolution of the state x_k (solid line) and its estimates $\hat{x}_{k|k}(\bar{t})$ (dotted line) and $\hat{x}^a_{k|k}$; first and second component of state and relevant estimates are considered above and below, in the order. Left: $\bar{q} = 1$; Right: $\bar{q} = 2$. Dash-dotted lines mark the *actual* switching time $\bar{t} = 5.25$.

Plots of the estimates of the continuous state x are drawn in Figure 4. The optimal estimates $\hat{x}^a_{k|k}$ are compared with the true values x_k and with the best estimates $\hat{x}_{k|k}(\bar{t}, \bar{q})$ one could produce in case the switching event were known in advance. In both the cases $\bar{q} = 1$ and $\bar{q} = 2$, estimates $\hat{x}^a_{k|k}$ follow the benchmark $\hat{x}_{k|k}(\bar{t}, \bar{q})$ quite accurately, even in the "transient" between the actual switching instant and the time when $p_{k|k}(q)$ clearly singles out the final value of the discrete state.

VII. Conclusions

In this paper we introduced a switching state-space model in which the continuous state x satisfies a linear stochastic differential equation, noisy measurements are taken at known sample times and the parameters of the whole system change in time according to a continuous-time Markov chain q of known statistics. Such model accounts for switches *between* measurements, whence classical estimation techniques for JMLS cannot be applied.

We focused on a fault detection setting and solved the problem of the recursive Bayesian estimation of the joint state (x, q) at sampling times t_k from the collection of measurements y^k by optimal averaging of the conditional state estimates associated to every possible switching event. Using the same tools, we also solved the problem of computing optimal Bayesian estimates of the switching time $t^\star$. Due to the continuous-time nature of $t^\star$ the overall system is non-Gaussian, whence the optimal estimates cannot be expressed in a simple parametric form and numerical integrations over time intervals are unavoidable. One major challenge is to express the key quantities $f(t^\star, q^\star|y^k)$, $f(x_k|t^\star, q^\star, y^k)$ in a form that is convenient for both numerical quadrature and recursive update. We achieved this by a formal discretization of the system *conditioned* on the switching event. The estimation algorithms we proposed update a finite number of parameters by way of exact matrix iterations, whereas integral approximations are isolated out of the recursion and are performed at the very end of the overall calculation. This prevents accumulation of errors and yields an accurate computation of the estimates.

REFERENCES

[1] G. A. Ackerson and K. S. Fu. On state estimation in switching environments. *IEEE Transactions on Automatic Control*, 15(1):10–17, 1970.

[2] H. A. P. Blom and Y. Bar-Shalom. The interacting multiple model algorithm for systems with Markovian switching coefficients. *IEEE Transactions on Automatic Control*, 33(8):780–783, Aug. 1988.

[3] R. Chen and J. S. Liu. Mixture Kalman Filters. *Journal of the Royal Statistical Society – Series B*, 62:493–508, 2000.

[4] E. Cinquemani. *Estimation Problems in Jump Markov Systems*. Ph.D. Thesis, University of Padova, Italy, 2005.

[5] E. Cinquemani. A superposition principle for the Kalman filter. *Systems and Control Letters*, 55(1):38–44, Jan. 2006.

[6] O. L. V. Costa. Linear minimum mean square error estimation for discrete-time Markovian jump linear systems. *IEEE Transactions on Automatic Control*, 39(8):1685–1689, Aug. 1994.

[7] O. L. V. Costa and S. Guerra. Stationary filter for linear minimum mean square error estimator of discrete-time Markovian jump systems. *IEEE Transactions on Automatic Control*, 47(8):1351–1356, Aug. 2003.

[8] A. Doucet, N. J. Gordon, and V. Krishnamurthy. Particle filters for state estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(3):613–624, 2001.

[9] A. Doucet, A. Logothetis, and V. Krishnamurthy. Stochastic sampling algorithms for state estimation of Jump Markov linear systems. *IEEE Transactions on Automatic Control*, 45(2):188–201, 2000.

[10] R. J. Elliott, F. Dufour, and D. D. Sworder. Exact hybrid filters in discrete time. *IEEE Transactions on Automatic Control*, 41(12):1807–1810, Dec. 1996.

[11] W. Gander and W. Gautschi. Adaptive quadrature – revisited. *BIT Numerical Mathematics*, 40:84–101, 2000.

[12] A. Germani, C. Manes, and P. Palumbo. Polynomial filtering for stochastic systems with Markovian switching coefficients. In *Proceedings of the 42nd IEEE Conference on Decision and Control (CDC 2003)*, pages 1392–1397, Maui, Hawaii, Dec. 2003.

[13] J. L. Hibey and C. D. Charalambous. Conditional densities for continuous-time nonlinear hybrid systems with application to fault detection. *IEEE Transactions on Automatic Control*, 44(11):2164–2169, Nov. 1999.

[14] J. Hu, J. Lygeros, and S. S. Sastry. Towards a Theory of Stochastic Hybrid Systems. In *Third International Workshop on Hybrid Systems: Computation and Control*, Pittsburgh, PA, 2000. Springer Verlag Lecture Notes on Computer Science, vol. 1790.

[15] A. H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, London, 1970.

[16] A. Logothetis and V. Krishnamurthy. Expectation maximization algorithms for map estimation of jump Markov linear systems. *IEEE Transactions on Signal Processing*, 47(8):2139–2156, Aug. 1999.

[17] B. M. Miller and W. J. Runggaldier. Kalman filtering for linear systems with coefficients driven by a hiddem Markov jump process. *Systems and Control Letters*, 31:93–102, 1997.

[18] J. K. Tugnait. Adaptive estimation and identification for discrete systems with Markov jump parameters. *IEEE Transactions on Automatic Control*, 27(5):1054–1065, Oct. 1982.

[19] Q. Zhang. Hybrid filtering for linear systems with non-Gaussian disturbances. *IEEE Transactions on Automatic Control*, 45(1):50–61, Jan. 2000.