State Estimation and Prediction in a Class of Stochastic Hybrid Systems

Eugenio Cinquemani* Mario Micheli†* Giorgio Picci*

*Dipartimento di Ingegneria dell'Informazione, Università di Padova, Padova, Italy
†Division of Applied Mathematics, Brown University, Providence, RI 02912, USA

{eugenio.cinquemani,mario.micheli,giorgio.picci}@unipd.it

Abstract

We consider a dynamical system whose state equation evolves continuously in time according to a linear stochastic differential equation; the parameters of such SDE depend on a discrete variable that follows the laws of a continuous-time Markov process. Noisy measurements of the continuous state are made available at discrete deterministic times, by a static linear equation whose parameters depend, again, on the discrete state. Therefore the discrete state may switch between different values *between* successive measures. We solve the problem of estimating both the continuous and the discrete state, given the measurements up to a certain time, in an on-line manner. Models like the one we analyze arise naturally in industrial applications such as fault detection.

1 Introduction

In the recent past there has been a proliferation of papers on the topic of state estimation of Jump Markov Linear Systems, often referred to as *discrete-time* Stochastic Hybrid Systems in the electrical engineering community. To name a few, see Tugnait [21][22], Bar-Shalom [2], Elliott *et al.* [9], Murphy [20], Logothetis & Krishnamurthy [18], Chen & Liu [3], Lerner *et al.* [17], Koutsoukos *et al.* [15][16], Doucet *et al.* [7][8], Hofbaur & Williams [12], Costa *et al.* [5][6], Germani *et al.* [10]. Other authors have worked on the filtering problem for stochastic hybrid systems governed by *continuous-time* equations in both the state dynamics *and* the measurement process. See for instance Miller & Runggaldier [19], Hibey & Charalambous [11], Hu *et al.* [13], and Zhang [23].

In the present paper we study a model where the continuous state x evolves in time described by a linear stochastic *differential equation*, and noisy measurements are acquired at fixed deterministic *time instants* $\{t_k\}$. 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*. The goal is estimating the pair (x, q), given the available measurements. Note that the discrete state may switch (in principle, even more than once) *between* two different measurements. Such a model arises naturally in applications where the switching rate of the discrete state is high relatively to the frequency of measurements. In this paper we will formulate the problem mostly restricting our attention to the fault detection setting, for which the model is particularly well-suited.

The paper is organized as follows. In section 2 we introduce the general *continuous*-time dynamics, *discrete*-time measurement *stochastic hybrid model* and formalize the state estimation problems of our concern. We then focus on the special case of fault detection and introduce the switching time $t^\star$ as an alternative characterization of $q(t)$. In Section 3 we derive a statistically equivalent model based on system discretization along fixed trajectories of q, which we call *conditioned system*. Based on this result, the interpretation of the estimation of x in terms of averaging of *conditional Kalman filters* – i.e. ordinary Kalman filters conditioned on the switching time – is discussed in section 4. The following section presents methods for the efficient computation of the *a posteriori* density of $t^\star$, also showing its intimate connection with the computation of the *a posteriori* density of x for *arbitrary values* of $t^\star$. The latter problem is studied in section 6 and reduced once again to conditional Kalman filtering. It is then solved by way of original algorithms of minimum complexity for the recursive update of the conditional Kalman filter thought of as a function of $t^\star$. Final comments and perspectives of our work are reported in Section 7.

For reasons of space, all proofs will be omitted. We refer the reader to [4] for details.

2 Problem formulation

Let[1] $\mathcal{T} = \{t_k\}_{k\in\mathbb{N}_0}$ be a *deterministic sequence* such that, for all k,

$$0 = t_0 < t_1 < \ldots < t_k < t_{k+1} < \ldots$$

and $t_k \to \infty$ as $k \to \infty$. 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 4-tuple of matrices (F_q, G_q, H_q, K_q).

Consider the following dynamical model:[2]

$$\left\{\begin{array}{rclcl} \dot{x}(t) & = & F_{q(t)}x(t) & + & G_{q(t)}u(t) \\ y_k & = & H_{q(t_k)}x(t_k) & + & K_{q(t_k)}v_k \end{array}\right. , \qquad t \in \mathbb{R},\;\; t_k \in \mathcal{T}, \tag{1}$$

where $x : \mathbb{R} \to \mathbb{R}^n$, $y : \mathbb{N}_0 \to \mathbb{R}^p$, are stochastic processes. In the above linear model two different white, zero-mean, normalized Gaussian stationary noise inputs appear: the continuous-time noise $u(t)$, $t \in \mathbb{R}$ and the discrete-time noise v_k indexed by $k \in \mathbb{N}_0$. We assume that $\{u(t)\}_{t\in\mathbb{R}}$, $\{v_k\}_{k\in\mathbb{N}_0}$ and initial condition $x(t_0) \sim \mathcal{N}\big(\mu_0, \Sigma_0\big)$ are mutually independent. Furthermore, we shall assume that $q(t), t \in \mathbb{R}$ is a continuous-time, homogeneous *Markov process* (independent of inputs $\{u(t)\}$, $\{v_k\}$ and random variable x_0) with assigned transition probabilities $T_{ij}(\Delta) \triangleq P[q(t+\Delta) = j \mid q(t) = i]$ (independently of $x(t)$); the initial probabilities $p_i \triangleq \mathbb{P}[q(t_0) = i]$, $i \in \mathcal{Q}$ are also assigned.

The process $q(t)$ switches in time between different states in $\mathcal{Q}$ (and the time interval between two subsequent jumps is a memoryless random variable), thus changing the parameters of both the state equation (which is a linear stochastic differential equation) and the (static) measurement equation. 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 "best" estimate for the joint state (x, q). More precisely, for $j, k \in \mathbb{N}_0$ we wish to compute the least squares estimate of the *continuous* state $x(t_j)$:

$$\hat{x}_{j|k} \triangleq \arg\min_{z\in\mathbb{R}^n} \mathbb{E}\left[||z - x(t_j)||^2 \,\middle|\, y^k\right] = \mathbb{E}\big[x(t_j) \,\big|\, y^k\big], \tag{2}$$

[1] The notation $\mathbb{N}_0$ stands for $\mathbb{N} \cup \{0\}$.

[2] We will refer to the first equation in (1) as *state equation*, and the second one as *measurement equation*.

and the *a posteriori* probability distribution of the *discrete* state:

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

We will mostly restrict our attention to the cases $j = k$ (filtering) and $j = k+1$ (prediction).

According to our model the discrete state can switch between different values in $\mathcal{Q}$ *between* two successive measurements —in principle, even more than once between the *same* two measurements: this makes the exact computation of the above estimates a formidable task. In order to simplify our problem, in the present paper we shall limit ourselves to a *fault detection* setting. That is, we will assume that $\mathcal{Q} = \{0, 1\}$ and that state $q = 1$ is *absorbing*: in other words, the transition probability matrix $[T_{ij}(\Delta)]$ is given by

$$T(\Delta) = \begin{bmatrix} e^{-\lambda\Delta} & 1 - e^{-\lambda\Delta} \\ 0 & 1 \end{bmatrix}$$

for some given parameter λ. Therefore there can be *at most* one switching time (from state 0 to state 1) which we will indicate with $t^\star$; its probability distribution function, for $t \geq 0$, is

$$F_{t^\star}(t) = (1 - e^{-\lambda t})p_0 + p_1$$

and is undefined for $t < 0$. In particular, when $p_1 = 0$ we have that $t^\star \sim \mathcal{E}(\lambda)$. Since there is only one switching time $t^\star$, we shall compute the *a posteriori* probability density

$$f(t^\star \,|\, y^k) \quad \text{for } t^\star > 0,$$

from which probabilities (3) follow immediately. For the clarity of the exposition, the assumption $p_1 = 0$ will be maintained throughout the paper.

This setting can be generalized to a broader class of Markov chains. For instance, a straightforward extension is to consider Markov chains having $N - 1$ absorbing states out of an arbitrary number of states N. However, this will form the object of future studies.

3 The conditioned system

Notwithstanding the stochastic nature of the switching time, one may fix the value of $t^\star$ and study the system associated to the corresponding trajectory $q(t)$. In this way, all parameters of (1) are determined, and a standard linear time-varying Gaussian system is obtained. It is common knowledge that such a system can be discretized, *i.e.* a *discrete-time*, time-varying linear Gaussian system can be associated to it so to preserve the joint statistical description of the *sampled state*

$$x_k \triangleq x(t_k)$$

and the measurements y_k. Precisely, we may introduce the *conditioned system*

$$\begin{cases} x_{k+1} &= A_k(t^\star)x_k &+ & u_k \\ y_k &= C_k(t^\star)x_k &+ & D_k(t^\star)v_k \end{cases}, \tag{4}$$

$$u_k \sim \mathcal{N}(0, Q_k(t^\star)),$$

with $\{u_k\}$ white and independent of $\{v_k\}$ and x_0, where the parameters $A_k(t^\star)$, $Q_k(t^\star)$, $C_k(t^\star)$, $D_k(t^\star)$ can be determined from those of the original system (1) *and* the value of $t^\star$ so to guarantee the desired statistical equivalence. This will be done in the next section.

Of course, (4) is a state-space representation of the random variables x_k and y_k *conditioned on* $t^\star$. Moreover, for changing values of $t^\star$, (4) describes a family of models corresponding to the different possible realizations of $t^\star$.

3.1 Computation of the conditioned system parameters

In this section we will assume that $t^\star$ takes values in a certain interval (t_h, t_{h+1}), with t_h, $t_{h+1} \in \mathcal{T}$. The interval is assumed to be open without loss of generality.

Proposition 1. *Assume that F_q and $-F_q$ have disjoint spectra, $q = 0, 1$. Then:*

1. *the Lyapunov equation*
$$F_q J_q + J_q F_q^T = -G_q G_q^T$$
admits a unique solution in J_q, $q = 0, 1$;

2. *the parameters of the conditioned system are given by:*

Case $k \neq h$ (i.e. $t^\star \notin (t_k, t_{k+1})$):

$$\begin{array}{rclrcl} A_k(t^\star) & = & e^{F_q(t_{k+1}-t_k)} & C_k(t^\star) & = & H_q \\ Q_k(t^\star) & = & J_q - A_k(t^\star) J_q A_k^T(t^\star) & D_k(t^\star) & = & K_q \end{array}$$

where $q = 0$ if $k < h$ and $q = 1$ if $k > h$;

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

$$\begin{array}{rclrcl} A_k(t^\star) & = & A_{k,1}(t^\star) A_{k,0}(t^\star) & C_k(t^\star) & = & H_0 \\ Q_k(t^\star) & = & -A_k(t^\star) S_{k,0}(t^\star) A_k^T(t^\star) + S_{k,1}(t^\star) & D_k(t^\star) & = & K_0 \end{array}$$

where, for $i = 0, 1$,

$$\begin{array}{rcl} A_{k,i}(t^\star) & = & e^{F_i(-1)^{i+1}(t_{k+i}-t^\star)}, \\ S_{k,i}(t^\star) & = & J_i - A_{k,i}^{-1}(t^\star) J_i A_{k,i}^{-T}(t^\star). \end{array}$$

Remark 1. The assumption on the spectrum of F_q only plays a role in the existence and uniqueness of the solution of the Lyapunov equation of point 1. (see [1], pp. 203-204). For arbitrary matrices F_0 and F_1, the computation of Q_k, $S_{k,0}$ and $S_{k,1}$ (where J_0 and J_1 appear) can still be accomplished although in a less elegant form.

Remark 2. For $k \neq h$, $A_k(t^\star)$ and $Q_k(t^\star)$ do not depend on the *specific* value of $t^\star$. In fact, they depend on $t^\star$ *only through* h. The same clearly holds for $C_k(t^\star)$, $D_k(t^\star)$, for *any* k.

Remark 3. At this stage, all parameters of the conditioned system are expressed in terms of explicit functions of $t^\star$. Notice that J_i, $i = 0, 1$ may be computed *offline* with arbitrary precision using standard numerical tecniques. Parameters A_k, Q_k, $k \neq h$ can be computed offline as well.

4 The filtering problem as averaging of Kalman filters

Let us take a deeper look at the estimation problems we stated in section 2. For any index j, consider the computation of $\hat{x}_{j|k}$. Applying the Law of Total Probability we write

$$f(x_j|y^k) = \int_0^{+\infty} f(x_j|t^\star, y^k) f(t^\star|y^k) \mathrm{d}t^\star. \tag{5}$$

We recognize $f(x_j|t^\star, y^k)$ to be the *a posteriori* density of the state x_j given y^k of the conditioned system (4).

In the light of the discussion of section 3, *for any fixed value of* $t^\star$ it must hold that

$$f(x_j|t^\star, y^k) \sim \mathcal{N}(\hat{x}_{j|k}(t^\star), P_{j|k}(t^\star)), \tag{6}$$

where mean and variance may be interpreted as the minimum error variance estimate of x_j given y^k and the estimation error covariance matrix for the conditioned system (see for instance [14]). In particular,

$$\hat{x}_{k|k}(t^\star) \tag{7}$$

$$\hat{x}_{k+1|k}(t^\star) \tag{8}$$

are the *conditional Kalman filter* and the *conditional Kalman predictor* for the corresponding conditioned system, whereas

$$P_{k|k}(t^\star) \tag{9}$$

$$P_{k+1|k}(t^\star) \tag{10}$$

are the relative covariance matrices. Of course, these may be computed by an obvious *conditional Kalman recursion*, which we report for later convenience:

Measurement update:

$$\begin{array}{rcl} L_k(t^\star) & = & P_{k|k-1}(t^\star) C_k^T(t^\star) [C_k(t^\star) P_{k|k-1}(t^\star) C_k^T(t^\star) + D(t^\star) D^T(t^\star)]^{-1} \\ \hat{x}_{k|k}(t^\star) & = & \hat{x}_{k|k-1}(t^\star) + L_k(t^\star)[y_k - C_k(t^\star)\hat{x}_{k|k-1}(t^\star)] \\ P_{k|k}(t^\star) & = & P_{k|k-1}(t^\star) - L_k(t^\star) C_k(t^\star) P_{k|k-1}(t^\star) \end{array} \tag{11}$$

Time update:

$$\begin{array}{rcl} \hat{x}_{k+1|k}(t^\star) & = & A_k(t^\star)\hat{x}_{k|k}(t^\star) \\ P_{k+1|k}(t^\star) & = & A_k(t^\star) P_{k|k}(t^\star) A_k^T(t^\star) + Q_k(t^\star) \end{array} \tag{12}$$

By equation (5), estimate (2) is therefore equal to the conditional average

$$\hat{x}_{j|k} = \int_0^{+\infty} \hat{x}_{j|k}(t^\star) f(t^\star|y^k) \mathrm{d}t^\star. \tag{13}$$

Hence, for $j = k$ (or $k + 1$), we have a natural interpretation of $\hat{x}_{j|k}$ as *averaging of Kalman filters (or predictors)*. Note that (5) is a *weighted* average of Gaussian densities parameterized by $t^\star$. What is obtained in general is not at all Gaussian, hence there is no hope to compute $\hat{x}_{j|k}$ in a linear recursive manner [14].

It is now evident that the *a posteriori* density $f(t^\star|y^k)$ plays a major role in the estimation (2). In fact, it is intimately related to the computation of (8), as it will be clear in the next section. Hence, with the computation of integral (13) in mind, the attention shifts to deriving *explicit* expressions for $f(t^\star|y^k)$ (section 5) and (6) (section 6), with special regard to filtering ($j = k$) and prediction ($j = k + 1$).

5 Switching time estimation

In this section we shall present a technique for the computation of the conditional probability density $f(t^\star|y^k)$, i.e. the *a posteriori* statistical description of the switching time (given the data up to time t_k). The knowledge of $f(t^\star|y^k)$ obviously has importance *per se*, since it allows to compute probabilities such as $\mathbb{P}[q(t_j) = 1 \,|\, y^k] = \mathbb{P}[t^\star < t_j \,|\, y^k]$, which is part of the solution to the state estimation problem (more precisely, when $j > k$, $j = k$, or $j < k$ we are dealing respectively with prediction, filtering, or smoothing). But on the other hand, as discussed in the previous paragraph, the above density plays a fundamental role in computing the estimate of continuous state $x(t_j)$ as well.

We will obtain $f(t^\star|y^k)$ by first computing the likelihood function $f(y^k|t^\star)$ and then applying Bayes'rule. Two different methods for the computation of $f(y^k|t^\star)$ are presented, both making use of the results of section 3. Note incidentally that the parameters of the conditioned system (4) are functions of the random variable $t^\star$, hence they are random themselves; however, for the time being, fix a particular value of $t^\star$. Keeping this in mind, sometimes we shall drop the $t^\star$ from our notation.

Direct computation of $f(y^k|t^\star)$. Define the following vectors and matrices:

$$\widetilde{\mu}_k \triangleq \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ \mu_0 \end{bmatrix} \in \mathbb{R}^{(k+1)n}, \qquad \widetilde{\Sigma}_k(t^\star) \triangleq \begin{bmatrix} Q_{k-1}(t^\star) & 0 & \cdots & 0 & 0 \\ 0 & Q_{k-2}(t^\star) & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & Q_0(t^\star) & 0 \\ 0 & 0 & \cdots & 0 & \Sigma_0 \end{bmatrix},$$

$$\Theta_k(t^\star) \triangleq \begin{bmatrix} I & A_{k-1} & A_{k-1}A_{k-2} & \cdots & A_{k-1}A_{k-2}\ldots A_2A_1 & A_{k-1}A_{k-2}\ldots A_1A_0 \\ 0 & I & A_{k-2} & \cdots & A_{k-2}A_{k-3}\ldots A_2A_1 & A_{k-2}A_{k-3}\ldots A_1A_0 \\ 0 & 0 & I & \cdots & A_{k-3}A_{k-4}\ldots A_2A_1 & A_{k-3}A_{k-4}\ldots A_1A_0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & A_1 & A_1A_0 \\ 0 & 0 & 0 & \cdots & I & A_0 \\ 0 & 0 & 0 & \cdots & 0 & I \end{bmatrix}$$

(note that al the A_i's are functions of $t^\star$),

$$\Xi_k \triangleq \left[\; A_{k-1} \mid A_{k-1}A_{k-2} \mid \cdots \mid A_{k-1}A_{k-2}\ldots A_2A_1 \mid A_{k-1}A_{k-2}\ldots A_1A_0 \;\right],$$

$$\Upsilon_k(t^\star) \triangleq \mathrm{diag}\{C_k(t^\star), \ldots, C_0(t^\star)\} \qquad \text{and} \qquad \Lambda_k(t^\star) \triangleq \mathrm{diag}\{D_k(t^\star), \ldots, D_0(t^\star)\}.$$

The following proposition holds:

Proposition 2. *Let $y^k \triangleq [y_k^T, y_{k-1}^T, \ldots, y_0^T]^T \in \mathbb{R}^{(k+1)\times p}$ be the vector of all measurements up to time t_k. Then y^k conditioned on $t^\star$ has the following multivariate Gaussian density:*

$$f(y^k|\, t^\star) \sim \mathcal{N}\left(\mu_{y^k}(t^\star), \Sigma_{y^k}(t^\star)\right),$$

where

$$\begin{aligned} \mu_{y^k}(t^\star) &\triangleq \Upsilon_k(t^\star)\Theta_k(t^\star)\,\widetilde{\mu}_k\,, \\ \Sigma_{y^k}(t^\star) &\triangleq \Upsilon_k(t^\star)\Theta_k(t^\star)\,\widetilde{\Sigma}_k(t^\star)\,\Theta_k^T(t^\star)\Upsilon_k^T(t^\star) + \Lambda_k(t^\star)\Lambda_k^T(t^\star)\,. \end{aligned}$$

The above quantities may be computed by the following iteration on k:

1. *each n-dimensional subvector of $\mu_{y^k}(t^\star)$ is obtained just by left-multiplying the one below it by $C_k(t^\star)A_k(t^\star)$;*

2. *$\Sigma_{y^k}(t^\star)$ can be obtained by adding n rows and columns to $\Sigma_{y^{k-1}}(t^\star)$ as follows:*

$$\Sigma_{y^k} = \begin{bmatrix} \Phi_k & \Psi_k \\ \Psi_k^T & \Sigma_{y^{k-1}} \end{bmatrix}$$

where matrices Φ_k and Ψ_k are given by:

$$\begin{aligned} \Phi_k &= C_k(\Xi_k \widetilde{\Sigma}_{k-1} \Xi_k^T + Q_k) C_k^T + D_k D_k^T, \\ \Psi_k &= C_k \Xi_k \Sigma_{k-1} \Theta_{k-1}^T \Upsilon_{k-1}^T. \end{aligned}$$

An iterative formulation for the computation of $f(y^k|t^\star)$. The above computation may look somewhat cumbersome: however it only requires the computation of the conditioned system's parameters, which we performed in section 3.

Note that we may write density $f(y^k|t^\star)$ simply as follows:

$$f(y^k|\, t^\star) = f(y_k|\, t^\star, y^{k-1})\, f(y^{k-1}|\, t^\star); \tag{14}$$

this formula provides an iterative method for computing $f(y^k|t^\star)$. Since y^k is a *given* vector of data, for a fixed value of $t^\star$ we have that $f(y^{k-1}|t^\star)$ is just a number that we carry on from the previous computation. Such number has to be multiplied by $f(y_k|\, t^\star, y^{k-1})$, whose value is easily obtainable from $f(x_k|\, t^\star, y^{k-1})$. The latter quantity plays a fundamental role in the estimation of continuous state x, as we saw in section 4; in section 6 we will show a precise technique for computing it. However, if one is just interested in the posterior density of $t^\star$ and not in the estimation of the continuous state, the formulation given by Proposition 2 may be sufficient. Otherwise, once $f(x_k|\, t^\star, y^{k-1})$ is known, the application of (14) is more appropriate.

Application of Bayes' rule. The posterior density of $t^\star$ is given by:

$$f(t^\star|y^k) = \frac{f(y^k|t^\star) f(t^\star)}{\int_0^\infty f(y^k|t^\star) f(t^\star)\, \mathrm{d}t^\star} \tag{15}$$

where $f(t^\star) = \lambda e^{-\lambda t^\star}$ for $t^\star > 0$. In fact, for $t^\star > t_k$ density $f(y^k|t^\star)$ is independent of the specific value assumed by $t^\star$ (compare the initial discussion of section 6). Therefore, we have that the denominator of (15) is given by:

$$\int_0^{t_k} f(y^k|t^\star) f(t^\star)\, \mathrm{d}t^\star + f(y^k|t^\star > t_k)\, \mathbb{P}[t^\star > t_k]. \tag{16}$$

In principle, the above integration requires computing $f(y^k|t^\star)$ for infinite values of $t^\star$. In practice, by Proposition 2 and the results of section 3, likelihood $f(y^k|t^\star)$ may be efficiently evaluated at any $t^\star$ in the *finite interval* $(0, t_k)$. Hence, quadrature methods apply successfully.

Remark. When combined with the direct method for the computation of $f(y^k|\, t^\star)$, the computation of (15) does not make use of conditional Kalman filtering. However, whenever computation of $f(x_{k+1}|, t^\star, y^k)$ is carried out for the estimation of x (see sections 4 and 6), the iterative method (14), which takes advantage of the computation of the latter density, should be preferred. We will come back to this at the end of the next section.

6 Conditional Kalman filtering

Following section 4, for any fixed value of $t^\star$ one may think of computing the a posteriori densities

$$f(x_k|t^\star, y^k) \tag{17}$$
$$f(x_{k+1}|t^\star, y^k) \tag{18}$$

at once by simply running the conditional Kalman recursion associated to $t^\star$. In principle, the procedure solves the problem of computing (17) and (18) for *any* value of $t^\star$. In practice, however, it cannot deal with the computation of integrals such as (5) and (16) (see also (14) and related comments), where (17) and (18) need to be known *for all* $t^\star$, or at least for a relatively large set of values.

It turns out that the dependence on $t^\star$ can be singled out by suitably rearranging the computation of (11) and (12). Indeed, fix $h \in \mathbb{N}_0$ and let $t^\star$ assume *any value* in the interval (t_h, t_{h+1}). We note the following:

(i) (7), (9) and (8), (10) are *independent of the specific* $t^\star$ for $k \leq h$ and $k < h$, respectively;

(ii) for $k \geq h+1$, (7)÷(10) depend on $t^\star$ *only through their new initial conditions* $\hat{x}_{h+1|h}(t^\star)$, $P_{h+1|h}(t^\star)$.

Indeed, the parameters of the conditioned system are constant (w.r.t. $t^\star$) before t_h (when $q(t) \equiv 0$) and after t_{h+1} (when $q(t) \equiv 1$). Therefore, for *any* $t^\star \in (t_h, t_{h+1})$, (11) and (12) *evolve independently of* $t^\star$ before t_h and after t_{h+1}, whereas the role of $t^\star$ is concentrated in the time update at step $k = h$. Based on these two key remarks, the rest of the section will be devoted to deriving an explicit representation of densities (17), (18).

Again, let $h \in \mathbb{N}_0$ and $t^\star \in (t_h, t_{h+1})$. The first result is just a formalization of (i).

Proposition 3. *It holds that:*

$$\begin{array}{rclrcll} \hat{x}_{k|k}(t^\star) & = & \hat{x}_{k|k}(\infty) & P_{k|k}(t^\star) & = & P_{k|k}(\infty), & k \leq h, \\ \hat{x}_{k+1|k}(t^\star) & = & \hat{x}_{k+1|k}(\infty) & P_{k+1|k}(t^\star) & = & P_{k+1|k}(\infty), & k < h. \end{array}$$

The next result states how $t^\star$ affects (12) at step $k = h$.

Proposition 4. *In the same hypotheses of Proposition 1,*

$$\begin{array}{rcl} \hat{x}_{h+1|h}(t^\star) & = & A_h(t^\star)\hat{x}_{h|h}(\infty), \\ P_{h+1|h}(t^\star) & = & A_h(t^\star)(P_{h|h}(\infty) - J_0)A_h^T(t^\star) + A_{h,1}(t^\star)(J_0 - J_1)A_{h,1}^T(t^\star) + J_1. \end{array}$$

In essence, the above expresses the new initial conditions for the recursion steps $k \geq h+1$ as explicit functions of $t^\star$. Recall that $A_h(t^\star)$ is a known matrix exponential. We need now to show how (7) and (8) depend on $t^\star$ for $k \geq h+1$.

Proposition 5. *Assume C_k full row rank. For $k > h$, define the recursions*

$$\begin{aligned}
\Pi_k &= \begin{bmatrix} A_k^{-T} & A_k^{-T}\Delta_k \\ Q_k A_k^{-T} & A_k + Q_k A_k^{-T}\Delta_k \end{bmatrix} \Pi_{k-1}, & \Pi_h &= I, \\
N_k &= [A_k^T \Pi_k^{1,1} - \Pi_{k-1}^{1,1}]^T C_k^\dagger y_k + N_{k-1}, & N_h &= 0, \\
M_k &= [A_k^T \Pi_k^{1,2} - \Pi_{k-1}^{1,2}]^T C_k^\dagger y_k + M_{k-1}, & M_h &= 0,
\end{aligned}$$

with $C_k^\dagger = C_k^T(C_k C_k^T)^{-1}$, $\Delta_k = C_k^T(D_k D_k^T)^{-1}C_k$, and[3]

$$\Gamma_k = \begin{bmatrix} I & \Delta_k \\ A_k^T Q_k A_k^{-T} & A_k^T A_k + A_k^T Q_k A_k^{-T}\Delta_k \end{bmatrix} \Pi_{k-1}.$$

Then, for $k \geq h+1$,

$$\begin{aligned}
\hat{x}_{k|k}(t^\star) &= [\Gamma_k^{1,1} + \Gamma_k^{1,2} P_{h+1|h}(t^\star)]^{-T} \cdot [\hat{x}_{h+1|h}(t^\star) + N_k + P_{h+1|h}(t^\star) M_k] \\
P_{k|k}(t^\star) &= -A_k^{-1} Q_k A_k^{-T} + [\Gamma_k^{1,1} + \Gamma_k^{1,2} P_{h+1|h}(t^\star)]^{-T} \cdot [\Gamma_k^{2,1} + \Gamma_k^{2,2} P_{h+1|h}(t^\star)]^T \\
\hat{x}_{k+1|k}(t^\star) &= [\Pi_k^{1,1} + \Pi_k^{1,2} P_{h+1|h}(t^\star)]^{-T} \cdot [\hat{x}_{h+1|h}(t^\star) + N_k + P_{h+1|h}(t^\star) M_k] \\
P_{k+1|k}(t^\star) &= [\Pi_k^{1,1} + \Pi_k^{1,2} P_{h+1|h}(t^\star)]^{-T} \cdot [\Pi_k^{2,1} + \Pi_k^{2,2} P_{h+1|h}(t^\star)]^T
\end{aligned}$$

where superscript $^{(i,j)}$ indicates the (i,j)-th matrix block.

Remark. Observe that Π_k, N_k, M_k and Γ_k do not depend on the specific value of $t^\star$. In fact, they appear as constants in the expressions of $\hat{x}_{k|k}(t^\star)$, $P_{k|k}(t^\star)$ and $\hat{x}_{k+1|k}(t^\star)$, $P_{k+1|k}(t^\star)$. Hence, the latter depend on $t^\star$ only through $\hat{x}_{h+1|h}(t^\star)$, $P_{h+1|h}(t^\star)$. Also observe that the time update step of the above proposition holds trivially for $k = h$ too.

In practice, knowledge of (17) and (18) is required for (almost) every $t^\star$. For any index k, it may be obtained by considering the restriction of $\hat{x}_{j|k}(t^\star)$ and $P_{j|k}(t^\star)$, $j = k, k+1$ to each of the $k+2$ intervals

$$(t_0, t_1), \ldots, (t_h, t_{h+1}), \ldots, (t_k, t_{k+1}), (t_{k+1}, +\infty),$$

and applying the results of Propositions 3 and 5 to form (6) piecewise. All the procedure needs to compute is matrices $\Pi_k(h)$, $N_k(h)$, $M_k(h)$ and $\Gamma_k(h)$ for each $h \leq k$ (restrictions of $t^\star$ to (t_h, t_{h+1}), Proposition 5), plus a standard Kalman recursion up to step k (restriction of $t^\star$ to $(t_{k+1}, +\infty)$, Proposition 3). Moreover, the whole scheme can be put in recursive form as follows:

Initialization: set $\hat{x}_{0|-1} = \mu_0$, $P_{0|-1} = \Sigma_0$;

Iteration ($k \geq 0$): as measurement y_k arrives,

1. for $h = 0, \ldots, k-1$ compute $\Gamma_k(h)$ from $\Pi_{k-1}(h)$; compute $\hat{x}_{k|k}(\infty)$, $P_{k|k}(\infty)$ from $\hat{x}_{k|k-1}(\infty)$, $P_{k|k-1}(\infty)$;

[3]It is in fact $\Pi_k = \Pi_k(h)$, $N_k = N_k(h)$, $M_k = M_k(h)$, $\Gamma_k = \Gamma_k(h)$. For notational conciseness, the dependence on h of these and other quantities is not reported here.

2. for $h = 0, \ldots, k-1$, compute $\Pi_k(h)$, $N_k(h)$, $M_k(h)$ from $\Pi_{k-1}(h)$, $N_{k-1}(h)$, $M_{k-1}(h)$; set $\Pi_k(k) = I$, $N_k(k) = 0$, $M_k(k) = 0$; compute $\hat{x}_{k+1|k}(\infty)$, $P_{k+1|k}(\infty)$ from $\hat{x}_{k|k}(\infty)$, $P_{k|k}(\infty)$.

Of course, the initialization step gives the parameters that are needed to represent $f(x_0|\cdot, y^{-1})$, whereas points 1 and 2 of the iteration step yield the parameters to represent $f(x_k|\cdot, y^k)$ and $f(x_{k+1}|\cdot, y^k)$, respectively. *With this scheme, a complete, explicit representation of* (17) *and* (18) *in terms of the parameter* $t^\star$ *is computed with* $\mathcal{O}(k^2)$ *complexity.*

6.1 Application to switching time estimation

Based on expressions (14) and (15) of section 5 we get the following result.

Proposition 6. *The* a posteriori *density* $f(t^\star|y^k)$ *can be computed as follows:*

$$f(t^\star|y^k) = \frac{\prod_{j=0}^{k} f(y_j|t^\star, y^{j-1}) f(t^\star)}{(\ldots)},$$

where

$$\begin{aligned}(\ldots) \;&=\; \sum_{h=0}^{k-1} \Big[\prod_{j=0}^{h} f(y_j|t^\star > t_h, y^{j-1}) \cdot \int_{t_h}^{t_{h+1}} \prod_{j=h+1}^{k} f(y_j|t^\star, y^{j-1}) f(t^\star)\, dt^\star \Big] \\ &\quad + \prod_{j=0}^{k} f(y_j|t^\star > t_k, y^{j-1}) \mathbb{P}[t^\star > t_k].\end{aligned}$$

In fact, all the terms $f(y_j|t^\star, y^{j-1})$ may be trivially deduced from the corresponding densities $f(x_j|t^\star, y^{j-1})$. By considering their restriction to the relevant interval of integration, one may apply the algorithm presented above and suitable numerical quadrature so to obtain an efficient evaluation of all integrals, i.e. of the normalization factor. Similarly, this representation of $f(t^\star|y^k)$ is extremely well suited for a piecewise computation of integral (13).

7 Conclusions

In this paper we have presented a new method for estimating the state (x, q) of a class of stochastic hybrid systems, where the continuous state evolves according to a linear SDE, the discrete state is a continuous-time Markov chain, while noisy measurements of the continuous state are discrete in time.

For a *given* trajectory of the discrete state $q(t)$ the problem is solvable by applying ordinary Kalman filtering to the corresponding time-varying discrete-time dynamical system, sampled in correspondence of the measurement times. In order to solve our problem, however, we must average these Kalman filters against the *a posteriori* distribution of the discrete state switching time. This averaging operation eliminates the Gaussian nature of the estimate, which cannot therefore be described in a parametric way. However, we managed to formulate an algorithm that is exact *up to* the averaging operation. In other words, it involves *exact* and efficient computation of parameters until the very last moment, that is when integrals (5) or (16) have to be computed. Note, for example, that by following this procedure any approximation (due to the numerical computation of integral (5)) that is introduced for the calculation of $f(x_k|y^k)$ does *not* influence the degree of approximation of $f(x_\ell|y^\ell)$ for $\ell > k$, since the latter density is not computed directly from the former.

We believe that the class of stochastic hybrid systems that we consider is a natural one for many applications where discrete state jumps may occur at a rate that is higher than the frequency of measurements. We are currently considering to extend our algorithm to models that are more complex than the one we studied: i.e., instead of having an absorbing state (or more absorbing states), describing the evolution of the discrete state by more general Markov models allowing for multiple switches between two consecutive measurements.

8 Acknowledgements

This work was supported in part by the European Community through the project RECSYS of the V Framework Program. Mario Micheli's research was partially supported by the *Fondazione "Ing. Aldo Gini"* (Padova, Italy). We wish to thank Professor Augusto Ferrante for his precious advice.

References

[1] R. Bhatia. *Matrix Analysis.* Springer-Verlag New York Inc., New York, USA, 1997.

[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, M. Micheli, and G. Picci. State estimation for a class of stochastic hybrid systems. Journal paper, in preparation.

[5] 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.

[6] 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.

[7] A. Doucet and C. Andrieu. Iterative algorithms for state estimation of Jump Markov linear systems. *IEEE Transactions on Signal Processing*, 49(6):1216–1227, June 2001.

[8] 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.

[9] 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.

[10] 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.

[11] 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.

[12] M. W. Hofbaur and B. C. Williams. Mode estimation of probabilistic hybrid systems. In *Hybrid Systems: Computation and Control (HSCC 2002)*, Lecture Notes on Computer Sciences. Springer Verlag, 2002.

[13] 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.

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

[15] X. Koutsoukos, J. Kurien, and F. Zhao. Monitoring and diagnosis of hybrid systems using particle filtering methods. In *Proceedings of the Fifteenth International Symposium on the Mathematical Theory of Networks and Systems (MTNS '02)*, University of Notre Dame, South Bend, Indiana, Aug. 2002.

[16] X. Koutsoukos, J. Kurien, and F. Zhao. Estimation of distributed hybrid systems using particle filtering methods. In *Hybrid Systems: Computation and Control (HSCC 2003)*. Springer Verlag Lecture Notes on Computer Science, vol. 2623, Pittsburgh, PA, 2003.

[17] U. Lerner, R. Parr, D. Koller, and G. Biswas. Bayesian fault detection and diagnosis in dynamic systems. In *Proceedings of the 17th National Conference on Artificial Intelligence (AAAI)*, pages 531–537, Austin, Texas, July 2000.

[18] 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.

[19] 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.

[20] K. P. Murphy. Switching Kalman filters. Report 98-10, Compaq Cambridge Research Laboratory, 1998.

[21] 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.

[22] J. K. Tugnait. Detection and estimation for abruptly changing systems. *Automatica*, 18(5):607–615, 1982.

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