Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

Maximum-likelihood Kalman Filtering for Switching Discrete-time Linear Systems

A. Alessandri, M. Baglietto, and G. Battistelli

Abstract— **State estimation is addressed for a class of discrete-time systems that may switch among different modes taken from a finite set. The system and measurement equations of each mode are assumed to be linear and perfectly known, but the current mode of the system is unknown. Moreover, we assume that independently normally distributed noises affect the dynamics and the measurements. First, relying on a well-established notion of mode observability developed "ad hoc" for switching systems, an approach to system mode estimation based on a maximum likelihood criterion is proposed. Second, such mode estimator is embedded in a Kalman filtering framework to estimate the continuous state. Under the assumption of mode observability, stability properties in terms of boundedness of the mean square estimation error are proved for the resulting filter. Simulation results that show the effectiveness of the proposed filter are reported.**

I. Introduction

The problem of state estimation for switching systems has been tackled according to various approaches. The main difference among such methodologies concerns the knowledge about the system mode, as well as the computational complexity.

If a Markov chain is used to model the evolution of the discrete dynamics, it is difficult in general to derive the optimal estimator analytically, even under the assumption that the system is linear. Thus, suboptimal solution are searched that can be computationally tractable (see, e.g., [1] and the references therein). Among the various techniques developed up to now, the most popular ones refer to the so-called interacting multiple model (IMM) approach [2].

The issues arisen in designing estimators for switching systems have motivated a number of research activities in various directions. We shall focus on the methodologies that stem from the idea of basing on the properties of observability of the discrete state (i.e., the system mode) to devise the estimation technique. The problem of finding conditions that ensure to distinguish between two different discrete states for unforced noise-free switching linear systems was first addressed in [3], [4]. More recent advances on this topic have been developed in [5], where arbitrary switching sequences were considered. Such results have been extended to comply with the presence of bounded disturbances that corrupt the dynamics and the measures (see [6]).

After defining a suitable notion of mode observability, one can figure out a method to estimate the continuous and discrete states. For linear dynamic systems where only the measurement equations may switch, in [7] an observer is proposed that results from the on-line solution of a nonlinear least-squares problem. The problem of state estimation for noise-free piecewise affine linear systems using a moving-horizon estimation method is addressed in [8], where the switching is modelled via the so-called mixed logical dynamical (MLD) representation. Such methodology has been extended to account for the presence of disturbances in [9]. An approach to estimation for switching linear discrete-time systems affected by bounded disturbances is presented in [6], where the continuous state and the mode of the system are estimated by minimizing a least-squares cost function according to a moving-horizon strategy.

In this paper, the estimation of the mode is addressed in the presence of disturbances affecting both the dynamics and the measures. More specifically, such disturbances are supposed to be zero-mean Gaussian random noises. Under this assumption, in Section II, a method is proposed to estimate the system modes based on the maximum likelihood with respect to a batch of measurements collected over a moving horizon. Such an approach can be interpreted as a "statistical correction" of the minimum distance criterion of [10], [11]. In Section III, a new filtering technique is proposed that consists in a Kalman filter based on the nominal model of the switching system, where the discrete state is estimated using the above approach. Under the unique assumption of mode observability, the uniform boundedness of the mean square estimation error is proved. Simulation results to evaluate the performance of the proposed approach are presented in Section IV. The proofs are omitted due to space constraints, the interested reader is referred to [12].

Before concluding this section, let us introduce some notations and basic definitions. Given a generic vector v, $\|v\|$ denotes the Euclidean norm of v and, given a positive definite matrix P, $\|v\|_P$ denotes the weighted norm of v, $\|v\|_P \triangleq (v^\top P v)^{1/2}$. Given a generic sequence $\{z_t;\, t = 0, 1, \ldots\}$ and two time instants $t_1 \le t_2$, we define $\mathbf{z}_{t_1,t_2} \triangleq \mathrm{col}\,(z_{t_1}, z_{t_1+1}, \ldots, z_{t_2})$. Finally, $\mathbb{E}(\cdot)$ and $\mathbb{P}(\cdot)$ denote the expectation and, respectively, the probability operators.

A. Alessandri is with the Department of Production Engineering, Thermoenergetics, and Mathematical Models, DIPTEM–University of Genoa, P.le Kennedy Pad. D, 16129 Genova, Italy, alessandri@diptem.unige.it

M. Baglietto is with the Department of Communications, Computer and System Sciences, DIST–University of Genoa, Via Opera Pia 13, 16145 Genova, Italy, mbaglietto@dist.unige.it

G. Battistelli is with the Dipartimento di Sistemi e Informatica, Università degli Studi di Firenze, Via di S. Marta 3, 50139 Firenze, Italy, battistelli@dsi.unifi.it

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

II. Maximum likelihood mode estimation

Let us consider a class of switching discrete-time linear systems described by

$$\begin{aligned} x_{t+1} &= A(\lambda_t)\, x_t + w_t \\ y_t &= C(\lambda_t)\, x_t + v_t \end{aligned} \tag{1}$$

where $t = 0, 1, \ldots$ is the time instant, $x_t \in \mathbb{R}^n$ is the continuous state vector (the initial continuous state x_0 is unknown), $\lambda_t \in \mathcal{L} \triangleq \{1, 2, \ldots, L\}$ is the system mode or discrete state, $w_t \in \mathbb{R}^n$ is the system noise vector, $y_t \in \mathbb{R}^m$ is the vector of the measurements, and $v_t \in \mathbb{R}^m$ is the measurement noise vector. $A(\lambda)$ and $C(\lambda)$, $\lambda \in \mathcal{L}$, are $n \times n$ and $m \times n$ matrices, respectively.

It is supposed that the system noise vectors w_t and the measurement noise vectors v_t, for $t = 0, 1, \ldots$, are normally distributed independent random variables, i.e.,

$$p(w_t) \sim N(0, Q(\lambda_t))\,, \quad p(v_t) \sim N(0, R(\lambda_t))\,,$$

for $t = 0, 1, \ldots$. Where $Q(\lambda)$, $\lambda \in \mathcal{L}$ are symmetric positive semidefinite $n \times n$ matrices and $R(\lambda)$, $\lambda \in \mathcal{L}$ are symmetric positive definite $m \times m$ matrices. We assume that no a-priori probabilistic information is available, neither on the initial continuous state x_0 nor on the switching sequence λ_t for $t = 0, 1, \ldots$. Note that the considered model is quite different from the one usually referred as jump Markov system, since here the evolution of the discrete state is *not* supposed to be governed by a hidden-state Markov chain.

In this section, a moving-horizon maximum likelihood criterion is proposed for the estimation of the discrete state of system (1). More specifically, given the noisy observations sequence $\mathbf{y}_{t-N,t}$ over a time interval $[t-N, t]$, our goal consists in obtaining a "reliable" estimate of the *switching sequence* $\boldsymbol{\lambda}_{t-N,t}$ (or at least a portion of it). Since system (1) is time-invariant with respect to the extended state (x_t, λ_t), in the following, for the sake of simplicity and without loss of generality, we shall always consider the interval $[0, N]$.

If the evolution of the discrete state is completely unpredictable, the switching sequence $\boldsymbol{\lambda}_{0,N}$ can assume any value in the set $\mathcal{L}^{N+1}$. However, in many practical cases, the a-priori knowledge of the system may allow one to consider a restricted set of "admissible" switching patterns. Think, for example, of the case in which the discrete state is slowly varying, i.e., there exists a minimum number τ of steps between one switch and the following one. Of course, such a-priori knowledge may make the task of estimating the discrete state from the measurements $\mathbf{y}_{0,N}$ considerably simpler. As a consequence, instead of considering all the possible switching sequences belonging to $\mathcal{L}^{N+1}$, we shall consider a restricted set $\mathcal{A} \subseteq \mathcal{L}^{N+1}$ of all the *admissible switching sequences*, i.e., of all the switching sequences consistent with the a-priori knowledge of the evolution of the discrete state.

Some preliminary definitions are now needed. Let us consider a generic switching sequence $\boldsymbol{\lambda} \triangleq \mathrm{col}\left(\lambda^{(0)}, \ldots, \lambda^{(N)}\right)$ and define the matrices $F(\boldsymbol{\lambda})$ and $H(\boldsymbol{\lambda})$ as at the top of the next page. Then the observations sequence $\mathbf{y}_{0,N}$ can be written as

$$\mathbf{y}_{0,N} = F(\boldsymbol{\lambda}_{0,N})\, x_0 + H(\boldsymbol{\lambda}_{0,N})\, \mathbf{w}_{0,N-1} + \mathbf{v}_{0,N}\,. \tag{2}$$

A. Mode Observability in the Absence of Noises

For the sake of clarity, let us first recall some results on the observability of the discrete state in the absence of noises. Towards this end, let us consider the noise-free system

$$\begin{aligned} x_{t+1} &= A(\lambda_t)\, x_t \\ y_t &= C(\lambda_t)\, x_t\,. \end{aligned} \tag{3}$$

In this case, since the observations sequence can be expressed as $\mathbf{y}_{0,N} = F(\boldsymbol{\lambda}_{0,N})\, x_0$, the set $\mathcal{S}(\boldsymbol{\lambda})$ of all the possible vectors of observations in the interval $[0, N]$ associated with a switching sequence $\boldsymbol{\lambda} \in \mathcal{A}$ corresponds to the linear subspace

$$\mathcal{S}(\boldsymbol{\lambda}) \triangleq \left\{ \mathbf{y} \in \mathbb{R}^{m(N+1)} : \mathbf{y} = F(\boldsymbol{\lambda})x\,,\ x \in \mathbb{R}^n \right\}.$$

The following notion of *distinguishability* between two switching patterns in the noise-free case can be introduced.

Definition 1: For system (3), two switching sequences $\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in \mathcal{A}$ with $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}'$ are said to be *distinguishable* if $F(\boldsymbol{\lambda})\, x \neq F(\boldsymbol{\lambda}')\, x'$ for all $x, x' \in \mathbb{R}^n$ with $x \neq 0$ or $x' \neq 0$.

As shown in [3], the *joint observability matrix* $\left[\begin{array}{cc} F(\boldsymbol{\lambda}) & F(\boldsymbol{\lambda}') \end{array}\right]$ plays a key role in determining the distinguishability of two switching sequences $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$. More specifically, the following lemma holds.

Lemma 1: Let us consider two generic switching sequences $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}' \in \mathcal{A}$. Then $\boldsymbol{\lambda}$ is distinguishable from $\boldsymbol{\lambda}'$ if and only if $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ are *jointly observable*, i.e., $\mathrm{rank}\left(\left[\begin{array}{cc} F(\boldsymbol{\lambda}) & F(\boldsymbol{\lambda}') \end{array}\right]\right) = 2n\,.$ □

In the light of Lemma 1, if the joint-observability condition were satisfied for every couple of switching sequences $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}' \in \mathcal{A}$, then it would be possible to uniquely determine the switching sequence $\boldsymbol{\lambda}_{0,N}$ on the basis of the observations sequence $\mathbf{y}_{0,N}$, provided that the initial continuous state x_0 is not null. Unfortunately, as shown in [6], unless the number of measurements available at each time step is at least equal to the number of continuous state variables (i.e., $m \geq n$), in general it is not possible to satisfy the joint observability condition for all $\boldsymbol{\lambda} \neq \boldsymbol{\lambda}' \in \mathcal{A}$ (this happens because it is not possible to detect switches that occur in the last or in the first instants of an observations window). As a consequence, even in the absence of noises, it is not possible to uniquely determine the whole switching pattern $\boldsymbol{\lambda}_{0,N}$.

In order to overcome such a drawback, following the lines of [6], we shall look for two integers, α and ω, with $\alpha, \omega \geq 0$ and $\alpha + \omega \leq N$, such that it is possible to uniquely determine the discrete state λ_t in the restricted interval $[\alpha, N-\omega]$ on the basis of the observations sequence $\mathbf{y}_{0,N}$. Towards this end, given a switching sequence $\boldsymbol{\lambda}$ in the interval $[0, N]$, let us denote as $r^{\alpha,\omega}(\boldsymbol{\lambda})$ the restriction of $\boldsymbol{\lambda}$ to the interval $[\alpha, N-\omega]$. Thus, the following notion

$$F(\boldsymbol{\lambda}) \triangleq \begin{bmatrix} C(\lambda^{(0)}) \\ C(\lambda^{(1)})A(\lambda^{(0)}) \\ \vdots \\ C(\lambda^{(N)})\prod_{i=1}^{N} A(\lambda^{(N-i)}) \end{bmatrix} \quad H(\boldsymbol{\lambda}) \triangleq \begin{bmatrix} 0 & 0 & \cdots & 0 \\ C(\lambda^{(1)}) & 0 & \cdots & 0 \\ C(\lambda^{(2)})A(\lambda^{(1)}) & C(\lambda^{(2)}) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ C(\lambda^{(N)})\prod_{i=1}^{N-1} A(\lambda^{(N-i)}) & C(\lambda^{(N)})\prod_{i=1}^{N-2} A(\lambda^{(N-i)}) & \cdots & C(\lambda^{(N)}) \end{bmatrix}$$

of mode observability in the restricted interval $[\alpha, N-\omega]$ can be introduced.

Definition 2: System (3) is said to be (α,ω)-*mode observable* in $N+1$ steps if, for every couple $\boldsymbol{\lambda}, \boldsymbol{\lambda}' \in \mathcal{A}$ such that $r^{\alpha,\omega}(\boldsymbol{\lambda}) \neq r^{\alpha,\omega}(\boldsymbol{\lambda}')$, $\boldsymbol{\lambda}$ is distinguishable from $\boldsymbol{\lambda}'$ (or, equivalently, $\boldsymbol{\lambda}$ and $\boldsymbol{\lambda}'$ are jointly observable).

According to Definition 2, if system (3) is (α,ω)-mode observable, then different switching sequences in the interval $[\alpha, N-\omega]$ generate different observations sequence in the interval $[0, N]$, provided that the initial continuous state is not null. As a consequence, the switching sequence $\boldsymbol{\lambda}_{\alpha,N-\omega}$ can be determined uniquely from the observations sequence $\mathbf{y}_{0,N}$. Note that there could be more than one switching sequence $\boldsymbol{\lambda}$ such that $\mathbf{y}_{0,N} \in \mathcal{S}(\boldsymbol{\lambda})$; however, they would correspond to the same switching sequence in the restricted interval $[\alpha, N-\omega]$.

B. Mode Estimation in the Presence of Gaussian Noises

With these observability results in mind, let us now focus on the noisy system (1). As the noise vectors are assumed to be random variables with known probability density functions, a natural criterion to derive an estimate of the switching sequence $\boldsymbol{\lambda}_{0,N}$ consists in following a *maximum likelihood* approach.

Towards this end, let us denote by $p(\mathbf{y}|x,\boldsymbol{\lambda})$ the probability density function that specifies the probability of observing a sequence $\mathbf{y}$ in the interval $[0,N]$ given that the initial continuous state is x and the switching sequence is $\boldsymbol{\lambda}$. Then, the optimal estimate $\hat{\boldsymbol{\lambda}}_{0,N}$ according to the maximum likelihood criterion can be obtained by maximizing the probability $p(\mathbf{y}_{0,N}|x,\boldsymbol{\lambda})$ with respect to $\boldsymbol{\lambda}$. Since also the initial continuous state is unknown, the probability $p(\mathbf{y}_{0,N}|x,\boldsymbol{\lambda})$ has to be maximized also with respect to x. Thus, in practice, we shall address the following nested maximization problem

$$\hat{\boldsymbol{\lambda}}_{0,N} \in \arg\max_{\boldsymbol{\lambda}\in\mathcal{A}} \left[\max_{x\in\mathbb{R}^n} p(\mathbf{y}_{0,N}|x,\boldsymbol{\lambda}) \right]. \tag{4}$$

In order to explicitly solve the maximizations in (4), first note that, in the considered framework, given an initial continuous state x and a switching sequence $\boldsymbol{\lambda} = \mathrm{col}(\lambda^{(0)},\ldots,\lambda^{N})$, the observations sequence $\mathbf{y}$ turns out to be a normally distributed random variable with mean $F(\boldsymbol{\lambda})x$ and covariance matrix

$$\begin{aligned}\Sigma(\boldsymbol{\lambda}) \ \triangleq\ & H(\boldsymbol{\lambda})\,\mathrm{diag}[Q(\lambda^{(0)}),\ldots,Q(\lambda^{(N-1)})]H(\boldsymbol{\lambda})^\top \\ & + \mathrm{diag}[R(\lambda^{(0)}),\ldots,R(\lambda^{(N)})].\end{aligned}$$

Therefore, each probability $p(\mathbf{y}_{0,N}|x,\boldsymbol{\lambda})$ can be written as

$$\begin{aligned} p(\mathbf{y}_{0,N}|x,\boldsymbol{\lambda}) = & \left[(2\pi)^{(N+1)m}\,|\Sigma(\boldsymbol{\lambda})|\right]^{-1/2} \\ & \times \exp\left(-\frac{1}{2}\|\mathbf{y}_{0,N} - F(\boldsymbol{\lambda})x\|^2_{\Sigma(\boldsymbol{\lambda})^{-1}}\right) \end{aligned} \tag{5}$$

where, given a generic matrix M, $|M|$ denotes its determinant. By exploiting (5), with a little algebra one may conclude that the solution of the maximization problem in (4) is equivalent to the solution of the following minimization problem

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_{0,N} \ \in\ & \arg\min_{\boldsymbol{\lambda}\in\mathcal{A}} \Big[\log|\Sigma(\boldsymbol{\lambda})| \\ & + \min_{x\in\mathbb{R}^n} \|\mathbf{y}_{0,N} - F(\boldsymbol{\lambda})x\|^2_{\Sigma(\boldsymbol{\lambda})^{-1}} \Big]. \end{aligned} \tag{6}$$

As to the innermost minimization in (6), it is immediate to verify that the vector $\hat{x}(\boldsymbol{\lambda})$ that minimizes the quadratic form $\|\mathbf{y}_{0,N} - F(\boldsymbol{\lambda})x\|^2_{\Sigma(\boldsymbol{\lambda})^{-1}}$ with respect to x is given by

$$\hat{x}(\boldsymbol{\lambda}) = \left[F(\boldsymbol{\lambda})^\top\Sigma(\boldsymbol{\lambda})^{-1}F(\boldsymbol{\lambda})\right]^{-1} F(\boldsymbol{\lambda})^\top\Sigma(\boldsymbol{\lambda})^{-1}\mathbf{y}_{0,N}.$$

Thus, the maximum likelihood estimate $\hat{\boldsymbol{\lambda}}_{0,N}$ can be obtained as

$$\begin{aligned} \hat{\boldsymbol{\lambda}}_{0,N} \ \in\ & \arg\min_{\boldsymbol{\lambda}\in\mathcal{A}} \Big[\log|\Sigma(\boldsymbol{\lambda})| \\ & + \|[I - \tilde{P}(\boldsymbol{\lambda})]\mathbf{y}_{0,N}\|^2_{\Sigma(\boldsymbol{\lambda})^{-1}} \Big] \end{aligned} \tag{7}$$

where $\tilde{P}(\boldsymbol{\lambda})$ is the matrix of the projection on $\mathcal{S}(\boldsymbol{\lambda})$ according to the weighted norm $\|\cdot\|_{\Sigma(\boldsymbol{\lambda})^{-1}}$, i.e.,

$$\tilde{P}(\boldsymbol{\lambda}) \triangleq F(\boldsymbol{\lambda})\left[F(\boldsymbol{\lambda})^\top\Sigma(\boldsymbol{\lambda})^{-1}F(\boldsymbol{\lambda})\right]^{-1} F(\boldsymbol{\lambda})^\top\Sigma(\boldsymbol{\lambda})^{-1}.$$

In the following of the paper, for the sake of compactness, we shall use the definition

$$\tilde{d}(\mathbf{y}_{0,N},\boldsymbol{\lambda}) \triangleq \|[I - \tilde{P}(\boldsymbol{\lambda})]\mathbf{y}_{0,N}\|_{\Sigma(\boldsymbol{\lambda})^{-1}}.$$

Remark 1: It is worth noting that the quantity $\tilde{d}(\mathbf{y}_{0,N},\boldsymbol{\lambda})$ represents the distance of the observations sequence $\mathbf{y}_{0,N}$ from the linear subspace $\mathcal{S}(\boldsymbol{\lambda})$ according to the weighted norm $\|\cdot\|_{\Sigma(\boldsymbol{\lambda})^{-1}}$. With this respect, the maximum likelihood criterion (7) can be seen as a "statistical correction" of the minimum distance criterion of [10], [11] that takes into account the probabilistic knowledge on the noise vectors. Note that the two criteria turn out to be coincident whenever $\Sigma(\boldsymbol{\lambda}) = \sigma I$, $\forall \boldsymbol{\lambda} \in \mathcal{L}^{N+1}$.

Since in this case the system and measurement noise vectors are not bounded in norm, it is not possible to guarantee a-priori that the maximum likelihood criterion leads to the exact determination of the switching sequence in the restricted interval $[\alpha, N-\omega]$ as instead it happened in the framework of [10] (see Theorem 1). Nevertheless one may expect that, under mode observability assumptions, a probabilistic equivalent of such a result can be developed.

Towards this end, let us denote by $\mathbb{P}(\boldsymbol{\lambda} \succ \boldsymbol{\lambda}' | x_0, \boldsymbol{\lambda}_{0,N})$ the probability that $\boldsymbol{\lambda}$ is *more likely* than $\boldsymbol{\lambda}'$ (i.e., preferable over $\boldsymbol{\lambda}'$ according to the maximum likelihood criterion) given that the true initial continuous state is x_0 and the true switching sequence is $\boldsymbol{\lambda}_{0,N}$. Of course such probability corresponds to the probability that

$$\log|\Sigma(\boldsymbol{\lambda})| + \tilde{d}(\mathbf{y}_{0,N}, \boldsymbol{\lambda}) < \log|\Sigma(\boldsymbol{\lambda}')| + \tilde{d}(\mathbf{y}_{0,N}, \boldsymbol{\lambda}) \,.$$

Then the following theorem can be stated.

Theorem 1: Let us consider a switching sequence $\boldsymbol{\lambda}' \in \mathcal{A}$ (with $\boldsymbol{\lambda}' \neq \boldsymbol{\lambda}_{0,N}$) such that $\boldsymbol{\lambda}_{0,N}$ and $\boldsymbol{\lambda}'$ are jointly observable. Then suitable scalars $k > 0$ and $h > 0$ exist such that the following inequality holds

$$\begin{aligned} &\mathbb{P}(\boldsymbol{\lambda}_{0,N} \succ \boldsymbol{\lambda}' | x_0, \boldsymbol{\lambda}_{0,N}) \\ &\qquad \geq \gamma_{(N+1)m}\left(\frac{k}{h}\|x_0\|^2 + \frac{1}{h}\log\frac{|\Sigma(\boldsymbol{\lambda}')|}{|\Sigma(\boldsymbol{\lambda}_{0,N})|}\right) \end{aligned} \tag{8}$$

where $\gamma_{(N+1)m}(\cdot)$ is the cumulative distribution function of a χ^2-distribution with $(N+1)m$ degrees of freedom.
As a consequence,

$$\lim_{\|x_0\|\to+\infty} \mathbb{P}(\boldsymbol{\lambda}_{0,N} \succ \boldsymbol{\lambda}' | x_0, \boldsymbol{\lambda}_{0,N}) = 1 \,.$$

□

Theorem 1 ensures that the more the initial continuous state x_0 is "far from the origin" the more probable is to prefer the true switching sequence $\boldsymbol{\lambda}_{0,N}$ over another sequence sequence $\boldsymbol{\lambda}'$ that is distinguishable from $\boldsymbol{\lambda}_{0,N}$ according to Definition 1. In other words, Theorem 1 ensures that, if the noise-free system (3) is (α,ω)-mode observable in $N+1$ steps, then the greater is the initial continuous state, the more reliable is the maximum likelihood estimate of the switching sequence in the restricted interval $[\alpha, N-\omega]$.

III. Combining Maximum Likelihood Mode Estimation with Kalman Filtering

In this section, the previous results are applied to the development of a recursive scheme for the estimation of both the discrete and the continuous state.

A. State Estimation Scheme

If the discrete state λ_t were available, system (1) could be seen as a time-varying linear system corrupted by Gaussian noises. Then, at each time t, the optimal estimate $\hat{x}_t$ of the continuous state x_t could be computed by means of the Kalman filter recursion

$$\begin{aligned} K_t &= \bar{P}_t C(\lambda_t)^\top \left[R(\lambda_t) + C(\lambda_t)\bar{P}_t C(\lambda_t)^\top\right]^{-1} \\ \hat{x}_t &= \bar{x}_t + K_t\left[y_t - C(\lambda_t)\bar{x}_t\right] \\ P_t &= \left[I - K_t C(\lambda_t)\right]\bar{P}_t \\ \bar{x}_{t+1} &= A(\lambda_t)\hat{x}_t \\ \bar{P}_{t+1} &= A(\lambda_t)P_t A(\lambda_t)^\top \ + \ Q(\lambda_t) \end{aligned} \tag{9}$$

where P_t and K_t are the estimation error covariance at time t and the Kalman gain at time t, respectively.

In the considered framework, since the switching mode λ_t is not known exactly, the estimation of the continuous state turns out to be much more difficult. In this connection, a first very simple idea would consist in: computing at any time t an estimate $\hat{\lambda}_t$ of λ_t on the basis of the observations up to time t (e.g, according to a maximum likelihood criterion as in Section II); applying a Kalman filter recursion wherein the estimate $\hat{\lambda}_t$ is used instead of the true value λ_t.

Unfortunately, as pointed out in the previous section, a certain delay ω is unavoidable in order to obtain a reliable estimate of the discrete state λ_t. However, it is possible to devise a modified estimation scheme that exploits the observability results of Section II. Specifically, let $N = \alpha+\omega$. Further, let us denote by $\hat{\lambda}_{i,t}$ and $\hat{x}_{i,t}$ the estimates (made at time t) of λ_i and x_i, respectively. Then one can proceed as follows:

(i) at any time instant $t = N, N+1, \ldots,$ compute the estimates $\hat{\lambda}_{t-\omega|t}, \ldots, \hat{\lambda}_{t|t}$ of the system mode in the restricted interval $[t-\omega, t]$ on the basis of the measurements collected in the observation window $[t-N, t]$;

(ii) compute the estimates $\hat{x}_{t-\omega|t}, \ldots, \hat{x}_{t|t}$ of the continuous state in the restricted interval $[t-\omega, t]$ by propagating the Kalman filter recursion corresponding to the estimated discrete states.

Let us first consider step (i). In accordance with the maximum-likelihood criterion proposed in Section II, at every time instant $t = N, N+1, \ldots,$ the estimate $\hat{\boldsymbol{\lambda}}_{t-N,t|t}$ is obtained by addressing the minimization

$$\hat{\boldsymbol{\lambda}}_{t-N,t|t} \in \arg\min_{\boldsymbol{\lambda}\in\mathcal{A}} \left\{\log|\Sigma(\boldsymbol{\lambda})| + \tilde{d}(\mathbf{y}_{t-N,t}, \boldsymbol{\lambda})\right\} . \tag{10}$$

Then, only the estimates in the restricted interval $[t-\omega, t]$ are retained and exploited in step (ii). Note that, in order to take into account the possibility of an a-priori knowledge on the discrete state, the estimate $\hat{\boldsymbol{\lambda}}_{t-N,t|t}$ is constrained to belong to the set $\mathcal{A}$ of all the admissible switching sequences, i.e., the set of all the switching sequences in the observation window $[t-N, t]$ consistent with the a-priori knowledge of the evolution of the discrete state.

For what concerns step (ii), at any time $t = N, N+1, \ldots,$ the objective is to find estimates of the continuous state vectors $x_{t-\omega}, \ldots, x_t$ on the basis of the measurements collected in the observations window $[t-\omega, t]$, of a "prediction" $\bar{x}_{t-\omega|t}$, and of the estimates $\hat{\lambda}_{t-\omega|t}, \ldots, \hat{\lambda}_{t|t}$ obtained in step (i). As we have assumed the disturbances to be Gaussian random variables, a natural criterion to derive the estimates consists in resorting to the Kalman filter recursion. For reasons that will be clarified in Section III-B, the poles of the original system are scaled by a factor $\gamma > 1$ to take into account possible errors in the estimation of the switching sequence. This amounts to designing the Kalman gains with the state matrices $\gamma A(\hat{\lambda}_{i|t})$ for $i = t-\omega, \ldots, t$. Note that the scale factor γ is customarily adopted for Kalman filter design so as to enforce a certain convergence rate [13]. In Section III-B, sufficient conditions will be given on the scalar γ to ensure the stability of the overall estimation scheme.

As to the propagation of the estimation procedure from time t to time $t+1$, since the estimates $\hat{\lambda}_{t-\omega+1|t}, \ldots, \hat{\lambda}_{t|t}$ are not reliable, only $\hat{x}_{t-\omega|t}$ and $\hat{\lambda}_{t-\omega|t}$ have to be retained. These estimates enable one to find the prediction $\bar{x}_{t-\omega+1|t+1}$ through the use of the noise-free state equation, i.e.,

$$\bar{x}_{t-\omega+1|t+1} = A(\hat{\lambda}_{t-\omega|t})\hat{x}_{t-\omega|t}\,.$$

When the observation y_{t+1} becomes available, one can refer to the new observations vector $\mathbf{y}_{t-N+1,t+1}$ and generate the new estimates $\hat{\lambda}_{t-\omega+1|t+1}, \ldots, \hat{\lambda}_{t+1|t+1}$ and $\hat{x}_{t-\omega+1|t+1}, \ldots, \hat{x}_{t+1|t+1}$. The same mechanism is applied at each time step $t = N, N+1, \ldots$.

Summing up, the following iterative estimation procedure has to be applied at any time instant $t = N, N+1, \ldots$.

Procedure 1:

1) Given the observations vector $\mathbf{y}_{t-N,t}$, compute the maximum likelihood estimate $\hat{\boldsymbol{\lambda}}_{t-N,t|t}$ as in (10).
2) Given the maximum likelihood estimate $\hat{\boldsymbol{\lambda}}_{t-\omega,t|t}$, the observations vector $\mathbf{y}_{t-\omega,t}$, the prediction $\bar{x}_{t-\omega|t}$ and the covariance matrix $\bar{P}_{t-\omega|t}$, compute the estimates $\hat{x}_{t-\omega|t}, \ldots, \hat{x}_{t|t}$ by applying the Kalman filter recursion

$$\begin{aligned}
K_{i|t} &= \bar{P}_{i|t} C(\hat{\lambda}_{i|t})^\top \\
&\quad \times \left[R(\hat{\lambda}_{i|t}) + C(\hat{\lambda}_{i|t}) \bar{P}_{i|t} C(\hat{\lambda}_{i|t})^\top \right]^{-1} \\
\hat{x}_{i|t} &= \bar{x}_{i|t} + K_{i|t} \left[y_i - C(\hat{\lambda}_{i|t}) \bar{x}_{i|t} \right] \\
P_{i|t} &= \left[I - K_{i|t} C(\hat{\lambda}_{i|t}) \right] \bar{P}_{i|t} \\
\bar{x}_{i+1|t} &= A(\hat{\lambda}_{i|t})\, \hat{x}_{i|t} \\
\bar{P}_{i+1|t} &= \gamma^2 A(\hat{\lambda}_{i|t}) P_{i|t} A(\hat{\lambda}_{i|t})^\top \ + \ Q(\hat{\lambda}_{i|t})
\end{aligned} \tag{11}$$

for $i = t-\omega, \ldots, t$.
3) Set $\bar{x}_{t-\omega+1|t+1} = \bar{x}_{t-\omega+1|t}$ and $\bar{P}_{t-\omega+1|t+1} = \bar{P}_{t-\omega+1|t}$.

The procedure is initialized at time $t = N$ with some a-priori estimate $\bar{x}_{N-\omega|N}$ and some a-priori covariance $\bar{P}_{N-\omega|N}$.

It is important to note that the form of the set $\mathcal{A}$ plays a central role in the possibility of computing the minimum in step 1) in a reasonable time. In fact, if the cardinality of the set $\mathcal{A}$ grows very rapidly with the size N of the observations window or with the number L of possible discrete states, such a computation may become too time-demanding (this happens, for example, when the system can switch arbitrarily at every time step). Such issues can be avoided if the a-priori knowledge on the evolution of the discrete state leads to a considerable reduction of the number of admissible switching patterns. This is the case, for example, when the size $N+1$ of the observation window is smaller than the minimum admissible number of steps between one switch and the following one. In fact, under such an assumption, the cardinality of the set $\mathcal{A}$ is $L[(L-1)N+1]$ (see [6]).

B. Stability Analysis

In order to prove the stability of the proposed estimation scheme, it is convenient to rewrite the system equations (1) as

$$\begin{aligned}
x_{t-\omega+1} &= A(\hat{\lambda}_{t-\omega|t})\, x_{t-\omega} + \xi_{t-\omega} \\
y_{t-\omega} &= C(\hat{\lambda}_{t-\omega|t})\, x_{t-\omega} + \eta_{t-\omega}
\end{aligned} \tag{12}$$

for $t = N, N+1, \ldots$ where

$$\begin{aligned}
\xi_{t-\omega} &\triangleq \left[A(\lambda_{t-\omega}) - A(\hat{\lambda}_{t-\omega|t}) \right] x_{t-\omega} + w_{t-\omega} \\
\eta_{t-\omega} &\triangleq \left[C(\lambda_{t-\omega}) - C(\hat{\lambda}_{t-\omega|t}) \right] x_{t-\omega} + v_{t-\omega}\,.
\end{aligned}$$

The vectors $\xi_{t-\omega}$ and $\eta_{t-\omega}$ can be seen as fictitious (or virtual) noises that account for the true noises $w_{t-\omega}$ and $v_{t-\omega}$, respectively, as well as for the possible mismatch between the true discrete state $\lambda_{t-\omega}$ and its estimate $\hat{\lambda}_{t-\omega|t}$.

Since the estimate $\hat{\lambda}_{t-\omega|t}$ is obtained according to the maximum likelihood criterion (10), one can exploit Theorem 1 and state the following result.

Lemma 2: Suppose that system (1) is (α, ω)-mode observable in $N+1$ steps (see Definition 2). Then there exist two positive constants ρ_ξ and ρ_η such that

$$\mathbb{E}\left(\|\xi_{t-\omega}\|^2 \right) \le \rho_\xi\,, \quad \mathbb{E}\left(\|\eta_{t-\omega}\|^2 \right) \le \rho_\eta$$

for $t = N, N+1, \ldots$.

Further, if system (1) is noise-free, i.e., $w_t = 0$ and $v_t = 0$ for $t = 0, 1, \ldots$, then

$$\xi_{t-\omega} = 0\,, \quad \eta_{t-\omega} = 0$$

for $t = N, N+1, \ldots$. □

In words, Lemma 2 ensures that under mode-observability the virtual noises are uniformly bounded in the mean-square sense regardless of the continuous state trajectory. This state of affairs can be understood by noting that, as shown in Theorem 1, the greater the continuous state $x_{t-\omega}$, the more reliable the estimate $\hat{\lambda}_{t-\omega|t}$. In particular, in the limit for $\|x_{t-\omega}\| \to +\infty$ one has that $\hat{\lambda}_{t-\omega|t} = \lambda_{t-\omega}$ with probability one.

Let us now consider the sequence of estimates $\hat{x}_{t-\omega|t}$ for $t = N, N+1, \ldots$. It is immediate to see that, according to

Procedure 1, such estimates obey the Kalman filter recursion

$$
\begin{aligned}
K_{t-\omega|t} &= \bar{P}_{t-\omega|t} C(\hat{\lambda}_{t-\omega|t})^\top \\
&\quad \times \left[R(\hat{\lambda}_{t-\omega|t}) + C(\hat{\lambda}_{t-\omega|t}) \bar{P}_{t-\omega|t} C(\hat{\lambda}_{t-\omega|t})^\top \right]^{-1} \\
\hat{x}_{t-\omega|t} &= \bar{x}_{t-\omega|t} + K_{t-\omega|t} \\
&\quad \times \left[y_{t-\omega} - C(\hat{\lambda}_{t-\omega|t}) \bar{x}_{t-\omega|t} \right] \\
P_{t-\omega|t} &= \left[I - K_{t-\omega|t} C(\hat{\lambda}_{t-\omega|t}) \right] \bar{P}_{t-\omega|t} \\
\bar{x}_{t-\omega+1|t+1} &= A(\hat{\lambda}_{t-\omega|t}) \, \hat{x}_{t-\omega|t} \\
\bar{P}_{t-\omega+1|t+1} &= \gamma^2 A(\hat{\lambda}_{t-\omega|t}) P_{t-\omega|t} A(\hat{\lambda}_{t-\omega|t})^\top \\
&\quad + Q(\hat{\lambda}_{t-\omega|t})
\end{aligned}
\tag{13}
$$

for $t = N, N+1, \ldots$. Then, taking into account (12) and (13), the dynamics of the estimation error $e_{t-\omega|t} \triangleq x_{t-\omega} - \hat{x}_{t-\omega|t}$ can be written as

$$
e_{t-\omega+1|t+1} = \Phi_{t-\omega} \, e_{t-\omega|t} + \varepsilon_{t-\omega} \tag{14}
$$

for $t = N, N+1, \ldots$ where

$$
\Phi_{t-\omega} \triangleq \left[I - K_{t-\omega+1|t+1} C(\hat{\lambda}_{t-\omega+1|t+1}) \right] A(\hat{\lambda}_{t-\omega|t})
$$

and

$$
\begin{aligned}
\varepsilon_{t-\omega} \triangleq & \left[I - K_{t-\omega+1|t+1} C(\hat{\lambda}_{t-\omega+1|t+1}) \right] A(\hat{\lambda}_{t-\omega|t}) \, \xi_{t-\omega} \\
& - K_{t-\omega+1|t+1} \, \eta_{t-\omega+1} \, .
\end{aligned}
$$

Note that Lemma 2 ensures that also $\varepsilon_{t-\omega}$ is uniformly bounded in mean square by some positive constant ρ_ε, i.e.,

$$
\mathbb{E}\left(\|\varepsilon_{t-\omega}\|^2 \right) \leq \rho_\varepsilon
$$

for $t = N, N+1, \ldots$.

In what follows, some basic properties of the covariance matrices $P_{t-\omega|t}$ and of the estimation error state matrices $\Phi_{t-\omega}$ are summarized that descend directly from well-known results about Kalman filtering for time-varying and switching systems (see, for instance, [14] and [15]).

Lemma 3: Suppose that system (1) is uniformly observable with respect to the continuous state x_t, i.e., there exists an integer N_o such that for any switching sequence $\boldsymbol{\lambda}_{0,N_o-1}$ the observability matrix $F(\boldsymbol{\lambda}_{0,N_o-1})$ has full rank. Then the following results hold:

(a) there exist two positive constants a_1 and a_2 such that for $t = N, N+1, \ldots$

$$
a_1 I \leq P_{t-\omega|t} \leq a_2 I
$$

(b) for $t = N, N+1, \ldots$

$$
\gamma^2 \, \Phi_{t-\omega}^\top P_{t-\omega+1|t+1} \Phi_{t-\omega} - P_{t-\omega|t} \leq 0 \, .
$$

□

It is important to remark that Lemma 3 indicates that the time-varying quadratic function $V_{t-\omega}(e) \triangleq e^\top P_{t-\omega|t} \, e$ is a Lyapunov function for the noise-free estimation error dynamics or equivalently that the free response of system (14) converges exponentially to zero (the convergence rate being γ^2).

It is also worth noting that mode-observability, as defined in Section II, implies that also uniform observability with respect to the continuous state holds: the rank condition on the joint observability matrices $\left[\, F(\boldsymbol{\lambda}) \quad F(\boldsymbol{\lambda}') \, \right]$ can be satisfied only if each matrix $F(\boldsymbol{\lambda})$ has full rank. Hence, in the light of Lemmas 2 and 3, one can state the following stability result.

Theorem 2: Suppose that system (1) is (α, ω)-mode observable in $N+1$ steps. Then the following results hold:

(a) the estimation error is bounded in mean square as

$$
\mathbb{E}\left(\|e_{t-\omega|t}\|^2 \right) \leq \zeta_{t-\omega}^2
$$

for $t = N, N+1, \ldots$ The sequence $\{\zeta_{t-\omega}\}$ is defined recursively as

$$
\begin{aligned}
\zeta_{N-\omega} &= \left[\frac{a_2}{a_1} \, \mathbb{E}\left(\|e_{N-\omega|N}\|^2 \right) \right]^{1/2} \\
\zeta_{t-\omega+1} &= \frac{1}{\gamma} \, \zeta_{t-\omega} + \left(\frac{a_2}{a_1} \, \rho_\varepsilon \right)^{1/2} , \\
& \qquad t = N, N+1, \ldots
\end{aligned}
$$

(b) if $\gamma > 1$, then the sequence $\{\zeta_t\}$ converges exponentially to the asymptotic value

$$
\zeta_\infty \triangleq \left(\frac{a_2}{a_1} \, \rho_\varepsilon \right)^{1/2} \frac{\gamma}{\gamma - 1} \, ;
$$

(c) if, in addition, system (1) is noise-free, i.e., $w_t = 0$ and $v_t = 0$ for $t = 0, 1, \ldots$, then

$$
\lim_{t \to +\infty} e_{t-\omega|t} = 0 \, .
$$

□

Theorem 2, points (a) and (b), shows that, under mild assumptions, the estimation error $e_{t-\omega|t}$ is asymptotically bounded in mean square by the quantity ζ_∞. Further, according to point (c), when the system and measurement noises are identically zero, it turns out that the proposed estimator is an asymptotic observer for the continuous state as its estimation error converges to zero.

IV. NUMERICAL RESULTS

In this section, a simulation example is given to illustrate the effectiveness of the proposed approach to state estimation for switching linear systems in the presence of Gaussian noises. Let us consider the discretized equations of an undamped oscillator that may switch between two different oscillation frequencies

$$
\begin{aligned}
A(1) &= \begin{bmatrix} \cos(\omega_1 \Delta) & -\omega_1 \sin(\omega_1 \Delta) \\ \frac{1}{\omega_1} \sin(\omega_1 \Delta) & \cos(\omega_1 \Delta) \end{bmatrix} , \\
A(2) &= \begin{bmatrix} \cos(\omega_2 \Delta) & -\omega_2 \sin(\omega_1 \Delta) \\ \frac{1}{\omega_2} \sin(\omega_2 \Delta) & \cos(\omega_2 \Delta) \end{bmatrix} , \\
C(1) &= C(2) = \begin{bmatrix} 0 & 1 \end{bmatrix}
\end{aligned}
\tag{15}
$$

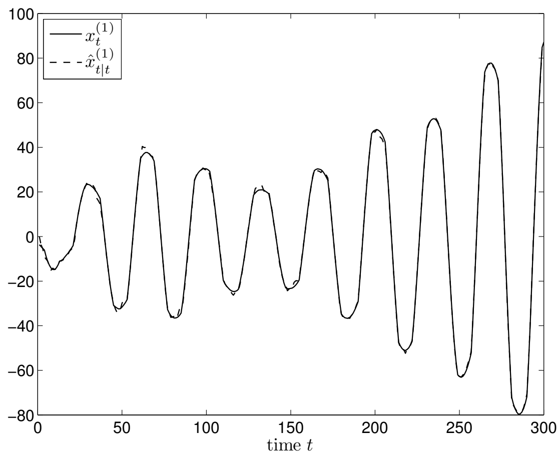

Fig. 1. True values and estimates obtained with the MLSKF of the first component of the state for a randomly chosen simulation with $p = 10$, $q = 0.1$, $r = 0.1$.

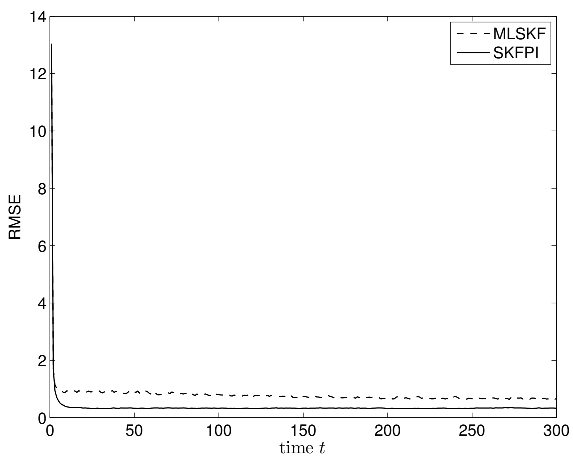

Fig. 2. RMSEs of the considered filters for $p = 10$, $q = 0.1$, $r = 0.1$.

where $\omega_1 = 1$, $\omega_2 = 2$, and the sampling time Δ is equal to 0.1. Note that both the system matrices have the two eigenvalues on the unit circle. However, due to the switching nature of the system, the trajectory of the continuous state may show a divergent behavior even in the absence of noises. It is supposed that the considered system has a minimum dwell time (i.e., the minimum number of steps between a switch and the next one) equal to 7. Moreover, x_0, w_t and v_t, $t = 0, 1, \ldots$, are supposed to be normally distributed independent random variables with zero mean and covariance $\bar{P}_0 = p^2 I$, $Q = q^2 I$, and $R = r^2 I$, respectively.

It is immediate to verify that the considered system is (α, ω)-*mode observable* in $\alpha + \omega + 1$ steps with $\omega = 2$ and $\alpha \in \{1, \ldots, 4\}$. Then the estimation scheme described in Procedure 1 leads to an estimation error bounded in mean square (see Theorem 2).

In the following, for the sake of brevity, we shall refer to the estimator obtained by repeatedly applying Procedure 1 as the *Maximum-Likelihood Switching Kalman Filter* (MLSKF). In order to evaluate the ability of the proposed estimation scheme to deal with unknown switches in the discrete state, the proposed filter is compared with the Kalman filter (9) obtained exploiting the exact knowledge of the discrete state instead of estimating it. Such an estimator will be called the *Switching Kalman Filter with Perfect Information* (SKFPI).

In order to compare the performance of the considered filters, 1000 Monte Carlo simulations have been carried out by randomly varying the noise realizations, the initial continuous state, and the switching sequence. In Fig. 1, the behavior of the true values and the estimates of the first component of the state is shown for a randomly chosen simulation (the second component exhibits a similar behavior).In Fig. 2, the plots of the *Root Mean Square Errors* (RMSEs) for the considered filters are shown. Note that the MLSKF shows just a little decay of performance with respect to the ideal case (perfect knowledge of the switching sequence).

REFERENCES

[1] R. Elliott, F. Dufour, and W. Malcolm, "State and mode estimation for discrete-time jump Markov systems," *SIAM J. on Control and Optimization*, vol. 44, no. 3, pp. 1081–1104, 2005.

[2] Y. Bar-Shalom and X. Li, *Multitarget-Multisensor Tracking: Principles and Techniques*. YBS Publishing, Storrs-CT, 1995.

[3] R. Vidal, A. Chiuso, and S. Soatto, "Observability and identifiability of jump linear systems," in *Proc. of the 41st IEEE Conference on Decision and Control*, Las Vegas, Nevada, 2002, pp. 3614–3619.

[4] R. Vidal, A. Chiuso, S. Soatto, and S. Sastry, "Observability of linear hybrid systems," in *Hybrid Systems: Computation and Control, Lecture Notes in Computer Science 2623*, F. Wiedijk, O. Maler, and A. Pnueli, Eds. Springer, 2003, pp. 523–539.

[5] M. Babaali and M. Egerstedt, "Observability of switched linear systems," in *Hybrid Systems: Computation and Control*, ser. Lecture Notes in Computer Science, R. Alur and G. J. Pappas, Eds. Springer, 2004, pp. 48–63.

[6] A. Alessandri, M. Baglietto, and G. Battistelli, "Receding-horizon estimation for switching discrete-time linear systems," *IEEE Trans. on Automatic Control*, vol. 50, no. 11, pp. 1736–1748, 2005.

[7] M. Babaali, M. Egerstedt, and E. Kamen, "A direct algebraic approach to observer design under switching measurement equations," *IEEE Trans. on Automatic Control*, vol. 49, no. 11, pp. 2044 – 2049, 2004.

[8] G. Ferrari-Trecate, D. Mignone, and M. Morari, "Moving horizon estimation for hybrid systems," *IEEE Trans. on Automatic Control*, vol. 47, no. 10, pp. 1663–1676, 2002.

[9] L. Pina and M. Botto, "Simultaneous state and input estimation of hybrid systems with unknown inputs," *Automatica*, vol. 42, no. 5, pp. 755–762, 2006.

[10] A. Alessandri, M. Baglietto, and G. Battistelli, "Minimum-distance receding-horizon state estimation for switching discrete-time linear systems," in *Assessment and Future Directions in Nonlinear Model Predictive Control*, ser. Lecture Notes in Computer Science, R. Findeisen, F. Allgower, and L. Biegler, Eds. Springer-Verlag, 2007, pp. 347–358.

[11] ——, "Luenberger observers for switching discrete-time linear systems," *International Journal of Control*, vol. 80, no. 12, pp. 1931–1943, 2007.

[12] ——, "Maximum-likelihood Kalman filtering for switching discrete-time linear systems," available on line at http://www.dist.unige.it/mbaglietto/mlkf.pdf.

[13] B. Anderson and J. Moore, *Optimal control: linear quadratic methods*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1990.

[14] ——, "Detectability and stabilizability of time-varying discrete-time linear systems," *SIAM Journal on Control and Optimization*, vol. 19, no. 1, pp. 20–32, 1981.

[15] G. Böker and J. Lunze, "Stability and performance of switching Kalman filters," *Int. Journal of Control*, vol. 75, pp. 1269–1281, 2002.