



**HAL**  
open science

# A maximum likelihood estimator for switching linear systems with unknown inputs

Luc Meyer, Dalil Ichalal, Vincent Vigneron

► **To cite this version:**

Luc Meyer, Dalil Ichalal, Vincent Vigneron. A maximum likelihood estimator for switching linear systems with unknown inputs. *Automatica*, 2019, 108, pp.108490. 10.1016/j.automatica.2019.07.003 . hal-02184492

**HAL Id: hal-02184492**

**<https://hal.science/hal-02184492>**

Submitted on 11 Jun 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# A maximum likelihood estimator for switching linear systems with Unknown Inputs <sup>★</sup>

Luc Meyer<sup>a</sup>, Dalil Ichalal<sup>b</sup>, Vincent Vigneron<sup>b</sup>

<sup>a</sup>ONERA, DTIS, Université Paris Saclay, 6 Chemin de la Vauve aux Granges, 91120 Palaiseau, France

<sup>b</sup>IBISC, Univ Evry Université Paris Saclay, 40, rue de Pelvoux, 91020 Courcouronnes, France

---

## Abstract

This paper investigates the construction of an estimator for switching linear systems with unknown inputs (UI) and Gaussian noises, in the case of unknown switching sequence. First, a maximum-likelihood switching sequence estimator is constructed. Then, both state and UI estimators are proposed. An example illustrates the theoretical contributions.

*Key words:* State Estimation, Minimum Variance Estimation, Unknown Input, Switching Systems, Discrete-Time System

---

## 1 Introduction

A switching system is a dynamical system with both continuous and discrete states. Several approaches exist in order to estimate the discrete-state of such a system. If the switching sequence is known, different observers have been proposed for systems without [4] or with Unknown Inputs (UI) [6]. If the switching sequence is unknown but its dynamic follows a known pattern (as a particular Markov model), then several estimators have also been proposed [7][5]. Without any assumption made on the discrete state dynamic, [1] and [2] have proposed state estimation for discrete-time linear switching systems affected by bounded noises. For continuous-time systems with bounded perturbations, a sliding-mode based observer is giving in [9]; in particular, this observer also deals with UI. In [3], a linear switching system affected by Gaussian noises in both state and measurement equations has been considered, but without the presence of UI.

The present paper focuses on discrete-time linear switching systems affected by both Gaussian noises and UI in both state and measurement equations. Its main contribution is to provide both state and UI estimators for such systems. It is the first paper dealing with linear switching systems affected by UI in the case in which the matrix associated

with the UI is not assumed to be constant. One of the main difficulty is to decouple the UI from the state estimation error despite of that fact. Another difficulty is to estimate the UI in such a case. The paper is organized as follows. In section 2, the problem is stated. In section 3, a method used for a maximum likelihood estimation of the switching parameter is presented. In section 4, a practical algorithm for estimating both the continuous state and the UI and based on the switching parameter estimation is given. Finally, in section 5, an example illustrating the approach is given.

The following notations are used in the paper. For any matrix  $A \in \mathcal{M}_{n,p}(\mathbb{R})$ ,  $A^T$  denotes its transpose, and  $A^\dagger$  its pseudo-inverse. For any square matrix  $A \in \mathcal{M}_{n,n}(\mathbb{R})$ ,  $tr(A)$  denotes its trace,  $det(A)$  denotes its determinant,  $\lambda(A)$  denotes the vector of its eigenvalues, and  $\lambda_{min}(A)$  (resp.  $\lambda_{max}(A)$ ) denotes its minimum (resp. maximum) singular value. For any vector  $v$ ,  $\|v\|$  denotes its euclidean norm. For any symmetric positive matrix  $M$ , let set the norm  $\|\cdot\|_M$  such that, for any vector  $v$  of appropriate dimension,  $\|v\|_M = \sqrt{v^T M v}$ . If  $x$  is a stochastic vector, its expectancy is denoted by  $E[x]$ . For any vector sequence  $(v_k)_{k=1,\dots}$  and any integers  $\alpha \geq 0$  and  $\omega \geq 0$ , let set  $v_{k-\alpha:k+\omega} = \begin{bmatrix} v_{k-\alpha}^T & v_{k-\alpha+1}^T & \dots & v_{k+\omega}^T \end{bmatrix}^T$ .

## 2 Problem Statement

Let consider the following switching system:

$$\begin{cases} x_{k+1} &= A_{\lambda_k} x_k + D_{\lambda_k} d_k + F_{\lambda_k} w_k \\ y_k &= C_{\lambda_k} x_k + E_{\lambda_k} d_k + v_k \end{cases} \quad (1)$$

---

<sup>★</sup> This paper was not presented at any IFAC meeting. Corresponding author L. Meyer

*Email addresses:* luc.meyer@onera.fr (Luc Meyer),  
dalil.ichalal@univ-evry.fr (Dalil Ichalal),  
vincent.vigneron@univ-evry.fr (Vincent Vigneron).

where  $x_k \in \mathbb{R}^{n_x}$  is the continuous state,  $y_k \in \mathbb{R}^{n_y}$  is the output,  $d_k \in \mathbb{R}^{n_d}$  is the Unknown Input (UI),  $\lambda_k \in \Lambda := \{1, 2, \dots, n_\Lambda\}$  is the *unknown* switching parameter (also called discrete mode or discrete state). Matrices  $A_{\lambda_k}$ ,  $D_{\lambda_k}$ ,  $F_{\lambda_k}$ ,  $C_{\lambda_k}$  and  $E_{\lambda_k}$  are switching matrices with appropriate dimensions. The perturbations  $w_k$  and  $v_k$  are independent Gaussian noises with zero mean and covariance matrices denoted by  $W_{\lambda_k}$  and  $V_{\lambda_k}$  respectively. The *unknown* initial state follows a Gaussian distribution (independent of  $w_k$  and  $v_k$ ) with *known* mean  $\hat{x}_0$  and *known* covariance matrix  $P_0$ . Let assume that for any  $\lambda \in \Lambda$ , the matrix  $V_\lambda$  is non singular, i.e.  $\det(V_\lambda) \neq 0$  (it is a widely used assumption for Kalman filtering).

In the whole paper, let us consider two integers  $\alpha \geq 0$  and  $\omega \geq 0$ , used as time indexes: for any time step  $k \geq \alpha$ , the measures obtained between time steps  $k - \alpha$  and  $k + \omega$  are used for the estimation of the unknown switching parameter  $\lambda_k$ . For any  $k \geq \alpha$ , let  $\lambda_{k-\alpha:k+\omega}$  be the switching sequence from time step  $k - \alpha$  to time step  $k + \omega$ . Then, the observation sequence  $y_{k-\alpha:k+\omega}$  gives:

$$y_{k-\alpha:k+\omega} = \mathcal{A}_{\lambda_{k-\alpha:k+\omega}} x_{k-\alpha} + \mathcal{D}_{\lambda_{k-\alpha:k+\omega}} d_{k-\alpha:k+\omega} + \mathcal{F}_{\lambda_{k-\alpha:k+\omega}} w_{k-\alpha:k+\omega} + v_{k-\alpha:k+\omega} \quad (2)$$

with matrices  $\mathcal{A}_{\lambda_{k-\alpha:k+\omega}}$ ,  $\mathcal{D}_{\lambda_{k-\alpha:k+\omega}}$  and  $\mathcal{F}_{\lambda_{k-\alpha:k+\omega}}$  given in (3). The following assumption is needed for removing the UI from equation (2).

**Assumption 1** For any switching sequence  $\pi \in \Lambda^{\alpha+\omega+1}$ , the rank  $r_d(\pi)$  of  $\mathcal{D}_\pi$  satisfies  $r_d(\pi) < (\alpha + \omega + 1)n_y$ .

For all  $\pi \in \Lambda^{\alpha+\omega+1}$ , there exists a full row rank matrix  $H_\pi \in \mathcal{M}_{(\alpha+\omega+1)n_y - r_d(\pi), (\alpha+\omega+1)n_y}(\mathbb{R})$  (with rank equal to  $r_h(\pi) = (\alpha + \omega + 1)n_y - r_d(\pi)$ ) such that  $H_\pi \mathcal{D}_\pi = 0$ . Assumption 1 ensures that  $r_h(\pi) > 0$ . Then, if  $\pi = \lambda_{k-\alpha:k+\omega}$ , when left-multiplying equation (2) by  $H_{\lambda_{k-\alpha:k+\omega}}$ , it comes:

$$H_{\lambda_{k-\alpha:k+\omega}} y_{k-\alpha:k+\omega} = H_{\lambda_{k-\alpha:k+\omega}} \mathcal{A}_{\lambda_{k-\alpha:k+\omega}} x_{k-\alpha} + H_{\lambda_{k-\alpha:k+\omega}} \mathcal{F}_{\lambda_{k-\alpha:k+\omega}} w_{k-\alpha:k+\omega} + H_{\lambda_{k-\alpha:k+\omega}} v_{k-\alpha:k+\omega}$$

equation in which the unknown input does not appear anymore.

### 3 Maximum-likelihood switching sequence estimation

This section deals with the switching sequence estimation problem. Let us start by introducing a definition of observability adapted to the considered switching system.

**Definition 1**  $\pi \in \Lambda^{\alpha+\omega+1}$  and  $\pi' \in \Lambda^{\alpha+\omega+1}$  are saying to be jointly observable if:

$$\text{rank} \left( \begin{bmatrix} H_{\pi'} \mathcal{A}_{\pi'} & H_{\pi'} \mathcal{A}_{\pi'} & H_{\pi'} \mathcal{D}_{\pi'} \\ H_{\pi} \mathcal{A}_{\pi} & H_{\pi} \mathcal{A}_{\pi} & H_{\pi} \mathcal{D}_{\pi} \end{bmatrix} \right) = 2n_x + (\alpha + \omega + 1)n_d \quad (8)$$

Definition 1 does not depend on the matrix  $H_{\pi'}$  chosen under Assumption 1, as this matrix is uniquely defined modulo a left-multiplication by a non singular square matrix.

Let us consider a time step  $k \geq \alpha$ , and let consider the following maximization problem in the presence of a UI:

$$\hat{\lambda}_{k-\alpha:k+\omega} \in \underset{\pi \in \Lambda^{\alpha+\omega+1}}{\text{argmax}} \left[ \max_{x \in \mathbb{R}^{n_x}} l(x, \pi | H_\pi y_{k-\alpha:k+\omega}) \right] \quad (9)$$

where  $l(x, \pi | H_\pi y_{k-\alpha:k+\omega}) = p(H_\pi y_{k-\alpha:k+\omega} | x, \pi)$  is the likelihood function to maximize.

Due to the left-multiplication of  $y_{k-\alpha:k+\omega}$  by  $H_\pi$  in (9), if  $\pi = \lambda_{k-\alpha:k+\omega}$  is the true switching sequence, then the likelihood  $l(x, \lambda_{k-\alpha:k+\omega} | H_{\lambda_{k-\alpha:k+\omega}} y_{k-\alpha:k+\omega})$  is not a function of the unknown input (see equation (8)). Moreover, the vector  $H_\pi y_{k-\alpha:k+\omega}$  is normally distributed with mean  $H_\pi \mathcal{A}_{\lambda_{k-\alpha:k+\omega}} x_{k-\alpha}$  and covariance matrix  $H_\pi \tilde{\Sigma}_{\lambda_{k-\alpha:k+\omega}} H_\pi^T$  (where for any  $\pi' \in \Lambda^{\alpha+\omega+1}$ ,  $\tilde{\Sigma}_{\pi'} = \mathcal{F}_{\pi'} W_{\pi'} \mathcal{F}_{\pi'}^T + V_{\pi'}$ ).

Then, following the same line as in [3], the maximum likelihood estimate can be obtained as:

$$\hat{\lambda}_{k-\alpha:k+\omega} \in \underset{\pi \in \Lambda^{\alpha+\omega+1}}{\text{argmin}} \left[ \log |\Sigma_\pi| + \delta(H_\pi y_{k-\alpha:k+\omega}, \pi) \right] \quad (10)$$

where  $\Sigma_\pi = H_\pi \tilde{\Sigma}_\pi H_\pi^T$ , and for any vector  $v \in \mathbb{R}^{r_h(\pi)}$ ,  $\delta(v, \pi) = \|(I - \Delta_\pi)v\|_{\Sigma_\pi^{-1}}^2$ , with  $\Delta_\pi$  the projection matrix on the linear span of  $H_\pi \mathcal{A}_\pi$ .

Given that the true continuous state is  $x_{k-\omega}$  at time step  $k - \omega$  and the true switching sequence is  $\lambda_{k-\alpha:k+\omega}$  between time steps  $k - \alpha$  and  $k + \omega$ , the probability that  $\lambda_{k-\alpha:k+\omega}$  is more likely to be chosen by (10) than a discrete sequence  $\pi' \in \Lambda^{\alpha+\omega+1}$  is denoted by  $\mathbb{P}(\lambda_{k-\alpha:k+\omega} \succ \pi' | x_{k-\alpha}, \lambda_{k-\alpha:k+\omega})$ . The following theorem gives a lower bound of such a probability.

**Theorem 1** Let assumption 1 hold. Let be  $k \geq \alpha$ . Let  $\pi' \in \Lambda^{\alpha+\omega+1}$  be a switching-sequence. Then, the following inequality holds:

$$\mathbb{P}(\lambda_{k-\alpha:k+\omega} \succ \pi' | x_{k-\alpha}, \lambda_{k-\alpha:k+\omega}) \geq \gamma_r \left( \frac{k(\lambda_{k-\alpha:k+\omega}, \pi')}{h(\lambda_{k-\alpha:k+\omega}, \pi')} \right) \times \left( \|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k+\omega}\|^2 \right) + \frac{1}{h(\lambda_{k-\alpha:k+\omega}, \pi')} \log \left| \frac{|\Sigma_{\pi'}|}{|\Sigma_{\lambda_{k-\alpha:k+\omega}}|} \right| \quad (11)$$

where  $\gamma_r(\cdot)$  is the cumulative distribution function of a  $\chi^2$  random variable with  $r = (\alpha + \omega + 1)n_y$  degrees of freedom. Moreover for any switching sequences  $\pi, \pi' \in \Lambda^{\alpha+\omega+1}$ , the coefficients  $h(\pi, \pi')$  and  $k(\pi, \pi')$  are given by equations (4), where  $\tilde{\Sigma}_\pi = \mathcal{F}_\pi W_\pi \mathcal{F}_\pi^T + V_\pi$ . Besides, if  $\pi$  and  $\pi'$  are jointly observable,  $k(\pi, \pi') > 0$ .

**PROOF.** In order to simplify the notations,  $\lambda_{k-\omega:k+\omega}$  is denoted by  $\pi$  in the following proof, but the reader has to keep

$$\mathcal{A}_{\lambda_{k-\alpha+\omega}} = \begin{bmatrix} C_{\lambda_{k-\alpha}} \\ C_{\lambda_{k-\alpha+1}} A_{\lambda_{k-\alpha}} \\ \dots \\ C_{\lambda_{k+\omega}} A_{\lambda_{k+\omega-1}} \dots A_{\lambda_{k-\alpha+1}} A_{\lambda_{k-\alpha}} \end{bmatrix}, \mathcal{D}_{\lambda_{k-\alpha+k+\omega}} = \begin{bmatrix} E_{\lambda_{k-\alpha}} & & 0_{n_y, n_d} & \dots & 0_{n_y, n_d} \\ C_{\lambda_{k-\alpha+1}} D_{\lambda_{k-\alpha}} & & E_{\lambda_{k-\alpha+1}} & \dots & 0_{n_y, n_d} \\ \dots & & \dots & \dots & \dots \\ C_{\lambda_{k+\omega}} A_{\lambda_{k+\omega-1}} \dots A_{\lambda_{k-\alpha+1}} D_{\lambda_{k-\alpha}} & C_{\lambda_{k+\omega}} A_{\lambda_{k+\omega-1}} \dots A_{\lambda_{k-\alpha+2}} D_{\lambda_{k-\alpha+1}} & \dots & \dots & E_{\lambda_{k+\omega}} \end{bmatrix}, \quad (3)$$

$$\mathcal{F}_{\lambda_{k-\alpha:k+\omega}} = \begin{bmatrix} 0_{n_y, n_y} & & & & \\ C_{\lambda_{k-\alpha+1}} F_{\lambda_{k-\alpha}} & & & & \\ \dots & & & & \\ C_{\lambda_{k+\omega}} A_{\lambda_{k+\omega-1}} \dots A_{\lambda_{k-\alpha+1}} F_{\lambda_{k-\alpha}} & C_{\lambda_{k+\omega}} A_{\lambda_{k+\omega-1}} \dots A_{\lambda_{k-\alpha+2}} F_{\lambda_{k-\alpha+1}} & \dots & C_{\lambda_{k+\omega}} F_{\lambda_{k+\omega-1}} & 0_{n_y, n_y} \end{bmatrix}$$

$$\begin{cases} k(\pi, \pi') &= \frac{1}{2} \lambda_{\min} \left\{ \left[ \mathcal{A}_{\pi'} \mathcal{D}_{\pi'} \right]^T H_{\pi'}^T (I - \Delta_{\pi'})^T \Sigma_{\pi'}^{-1} (I - \Delta_{\pi'}) H_{\pi'} \left[ \mathcal{A}_{\pi'} \mathcal{D}_{\pi'} \right] \right\} \\ h(\pi, \pi') &= \lambda_{\max} \left\{ H_{\pi}^T (I - \Delta_{\pi})^T \Sigma_{\pi}^{-1} (I - \Delta_{\pi}) H_{\pi} + H_{\pi'}^T (I - \Delta_{\pi'})^T \Sigma_{\pi'}^{-1} (I - \Delta_{\pi'}) H_{\pi'} \right\} \lambda_{\max} \{ \tilde{\Sigma}_{\pi} \} \end{cases} \quad (4)$$

$$\mathbb{P}(\pi_{k-\alpha:k+\omega} \succ \pi' | x_{k-\alpha}, \pi_{k-\alpha:k+\omega}) \geq \mathbb{P} \left( z^T \tilde{\Sigma}^{-1} z \leq \frac{k(\pi, \pi')}{h(\pi, \pi')} (||x_{k-\alpha}||^2 + ||d_{k-\alpha:k+\omega}||^2) + \frac{1}{h(\pi, \pi')} \log \frac{|\Sigma_{\pi'}|}{|\Sigma_{\pi}|} \right), \quad (5)$$

$$\Phi_{\lambda_{k-\alpha:k-1}} = \begin{bmatrix} A_{\lambda_{k-1}} \dots A_{\lambda_{k-\alpha}} & A_{\lambda_{k-1}} \dots A_{\lambda_{k-\alpha+1}} D_{k-\alpha} & \dots & A_{\lambda_{k-1}} D_{\lambda_{k-2}} & D_{\lambda_{k-1}} \end{bmatrix}, \Psi_{\lambda_{k-\alpha:k-1}} = \begin{bmatrix} A_{\lambda_{k-1}} \dots A_{\lambda_{k-\alpha+1}} F_{k-\alpha} & \dots & A_{\lambda_{k-1}} F_{\lambda_{k-2}} & F_{\lambda_{k-1}} \end{bmatrix} \quad (6)$$

$$\begin{aligned} & [(A_{\lambda_k} - A_{\hat{\lambda}_k}) - K_{k+1} (C_{\lambda_{k+1}} A_{\lambda_k} - C_{\hat{\lambda}_{k+1}} A_{\hat{\lambda}_k})] x_k = [(A_{\lambda_k} - A_{\hat{\lambda}_k}) - K_{k+1} (C_{\lambda_{k+1}} A_{\lambda_k} - C_{\hat{\lambda}_{k+1}} A_{\hat{\lambda}_k})] \Phi_{\lambda_{k-\alpha:k-1}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k-1} \end{bmatrix} \\ & + [(A_{\lambda_k} - A_{\hat{\lambda}_k}) - K_{k+1} (C_{\lambda_{k+1}} A_{\lambda_k} - C_{\hat{\lambda}_{k+1}} A_{\hat{\lambda}_k})] \Psi_{\lambda_{k-\alpha:k-1}} w_{k-\alpha:k-1} \end{aligned} \quad (7)$$

in mind that it is the actual discrete sequence. The following steps are similar to [3] in the structure, however, the technical difficulties raised by the presence of the UI makes the proof very different. Using equation (10), the probability  $\mathbb{P}(\pi \succ \pi' | x_{k-\alpha}, \lambda_{k-\alpha:k+\omega})$  is equal to the probability that the following inequality holds:

$$\log |\Sigma_{\pi}| + \delta(H_{\pi} y_{k-\alpha:k+\omega}, \pi) < \log |\Sigma_{\pi'}| + \delta(H_{\pi'} y_{k-\alpha:k+\omega}, \pi') \quad (12)$$

Let set  $z = \mathcal{F}_{\pi} w_{k-\alpha:k+\omega} + v_{k-\alpha:k+\omega}$ . Then  $z$  is normally distributed with zero mean and covariance matrix equal to  $\tilde{\Sigma}_{\pi}$ . On one hand, we have:

$$\delta(H_{\pi'} y_{k-\alpha:k+\omega}, \pi') = \delta(H_{\pi'} \mathcal{A}_{\pi} x_{k-\alpha} + H_{\pi'} \mathcal{D}_{\pi} d_{k-\alpha:k+\omega} + H_{\pi'} z, \pi'). \quad (13)$$

Besides:

$$\begin{aligned} & \delta(H_{\pi'} \mathcal{A}_{\pi} x_{k-\alpha} + H_{\pi'} \mathcal{D}_{\pi} d_{k-\alpha:k+\omega}, \pi') \\ &= \delta(H_{\pi'} \mathcal{A}_{\pi} x_{k-\alpha} + H_{\pi'} \mathcal{D}_{\pi} d_{k-\alpha:k+\omega} + H_{\pi'} z - H_{\pi'} z, \pi') \\ &\leq 2\delta(H_{\pi'} y_{k-\alpha:k+\omega}, \pi') + 2\delta(H_{\pi'} z, \pi'), \end{aligned} \quad (14)$$

the last inequality coming from one of the property of the function  $\delta(\cdot, \pi)$ :  $\delta(u - v, \pi) \leq \delta(u, \pi) + \delta(v, \pi)$ . Thus:

$$\begin{aligned} & \delta(H_{\pi'} y_{k-\alpha:k+\omega}, \pi') \\ &\geq \frac{1}{2} \delta(H_{\pi'} \mathcal{A}_{\pi} x_{k-\alpha} + H_{\pi'} \mathcal{D}_{\pi} d_{k-\alpha:k+\omega}, \pi') - \delta(H_{\pi'} z, \pi') \\ &\geq k(\pi, \pi') (||x_{k-\alpha}||^2 + ||d_{k-\alpha:k+\omega}||^2) - \delta(H_{\pi'} z, \pi'), \end{aligned} \quad (15)$$

where  $k(\pi, \pi')$  is given by equation (4). Note, that under the joint observability of  $\pi$  and  $\pi'$ , it comes that  $(I - \Delta_{\pi'}) H_{\pi'} \begin{bmatrix} \mathcal{A}_{\pi} \\ \mathcal{D}_{\pi} \end{bmatrix}$  is full column rank. Then, it follows that  $k(\pi, \pi') > 0$ . On the other hand, we have:

$$\begin{aligned} \delta(H_{\pi} y_{k-\alpha:k+\omega}, \pi) &= \delta(H_{\pi} \mathcal{A}_{\pi} x_{k-\alpha} + H_{\pi} \mathcal{D}_{\pi} d_{k-\alpha:k+\omega} + H_{\pi} z, \pi) \\ &= \delta(H_{\pi} z, \pi) \end{aligned} \quad (16)$$

as  $H_{\pi} \mathcal{D}_{\pi} d_{k-\alpha:k+\omega} = 0$ , and  $(I - \Delta_{\pi}) H_{\pi} \mathcal{A}_{\pi} = 0$ . Then it comes that:

$$\begin{aligned} \delta(H_{\pi} z, \pi) + \delta(H_{\pi'} z, \pi') &= ||(I - \Delta_{\pi}) H_{\pi} z||_{\Sigma_{\pi}^{-1}}^2 \\ &+ ||(I - \Delta_{\pi'}) H_{\pi'} z||_{\Sigma_{\pi'}^{-1}}^2 \leq h(\pi, \pi') ||z||_{\tilde{\Sigma}_{\pi}^{-1}}^2 \end{aligned} \quad (17)$$

where  $\tilde{\Sigma}_{\pi} = \mathcal{F}_{\pi} W_{\pi} \mathcal{F}_{\pi}^T + V_{\pi}$  and  $h(\pi, \pi')$  is given by equation (4). Finally, by combining (15), (16) and (17) it comes that

$$\begin{aligned} & \delta(H_{\pi} y_{k-\alpha:k+\omega}, \pi) - \delta(H_{\pi'} y_{k-\alpha:k+\omega}, \pi') \\ &\leq h(\pi, \pi') ||z||_{\tilde{\Sigma}_{\pi}^{-1}}^2 - k(\pi, \pi') (||x_{k-\alpha}||^2 + ||d_{k-\alpha:k+\omega}||^2) \end{aligned} \quad (18)$$

and thus a sufficient condition for equation (12) to hold is:

$$h(\pi, \pi') ||z||_{\tilde{\Sigma}_{\pi}^{-1}}^2 - k(\pi, \pi') (||x_{k-\alpha}||^2 + ||d_{k-\alpha:k+\omega}||^2) \leq \log \frac{|\Sigma_{\pi'}|}{|\Sigma_{\pi}|} \quad (19)$$

which leads to equation (5), and recalling that  $z$  is normally distributed with zero mean and covariance matrix

$\tilde{\Sigma}_\pi$ , it comes that  $z^T \tilde{\Sigma}_\pi^{-1} z$  is a  $\chi^2$  random variable with  $\dim(z) = (\alpha + \omega + 1)n_y$  degrees of freedom.

#### 4 A maximum-likelihood estimator

The aim of this section is to construct a linear estimator based on the maximum likelihood switching sequence estimation developed in section 3. In order to do it, it is necessary (and also sufficient as it will be seen in Theorem 2) to decouple the UI from the state estimation error (otherwise the behavior of this error would depend on the UI behavior), whether the estimated switching parameter is the true one or not.

**Assumption 2** *The following rank condition holds:*

$$\text{rank} \left( \begin{bmatrix} \mathbf{C} & \mathbf{E} \\ \mathbf{DI} & \mathbf{0} \end{bmatrix} \right) = \text{rank} \left( \begin{bmatrix} \mathbf{C} & \mathbf{E} \end{bmatrix} \right), \quad (20)$$

where  $\mathbf{I} = [I_{n_\Lambda n_d} \dots I_{n_\Lambda n_d}]$  ( $n_\Lambda$  times the matrix  $I_{n_\Lambda n_d}$ ),  $\mathbf{D} = [D_1 \dots D_{n_\Lambda}]$ ,  $\mathbf{C} = [C_1 \mathbf{D} \dots C_{n_\Lambda} \mathbf{D}]$  and  $\mathbf{E} = [E_1 \dots E_{n_\Lambda}]$ .

Note that this assumption differs from Assumption 1, as the former one is used in order to make the information of the output *not affected* by the UI, and thus to estimate *at best* the switching sequence, whereas assumption 2 is used in order to decouple the UI from the state estimation error for any choice of  $\hat{\lambda}_k$  in  $\Lambda$ . Under assumption 2, there exists a matrix  $K$  such that:

$$K \begin{bmatrix} \mathbf{C} & \mathbf{E} \end{bmatrix} = \begin{bmatrix} \mathbf{DI} & \mathbf{0} \end{bmatrix}, \quad (21)$$

with  $K$  given by:

$$K = G + XH, \quad (22)$$

with  $G = \begin{bmatrix} \mathbf{DI} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{C} & \mathbf{E} \end{bmatrix}^\dagger$  and  $H = I_{n_y, n_y} - \begin{bmatrix} \mathbf{C} & \mathbf{E} \end{bmatrix} \begin{bmatrix} \mathbf{C} & \mathbf{E} \end{bmatrix}^\dagger$ , and  $X \in \mathcal{M}_{n_x, n_y}(\mathbb{R})$  is any matrix (this means that for any matrix  $X$  of appropriate dimensions, equation (21) holds with  $K$  given by (22)). In fact,  $X$  is chosen differently at each time step  $k \geq \alpha$ . Thus, from now, it will be denoted by  $X_k$  and  $K$  will be denoted by  $K_k = G + X_k H$ . It is worth noting that, under Assumption 2, and according to what precedes the following equation holds for any  $\lambda, \lambda' \in \Lambda$ , and any choice of  $X_k$ :

$$K_k \mathcal{D}_{\lambda, \lambda'} = \begin{bmatrix} D_\lambda & \mathbf{0} \end{bmatrix}, \quad (23)$$

where  $\mathcal{D}_{\lambda, \lambda'} = \begin{bmatrix} C_{\lambda'} D_\lambda & E_{\lambda'} \end{bmatrix}$ . The following assumption is needed for the UI estimation.

**Assumption 3** *The following rank condition holds:*

$$\text{rank} \left( \begin{bmatrix} \mathbf{C} & \mathbf{E} \\ \mathbf{I}_{n_\Lambda^2, n_d} & \mathbf{0} \end{bmatrix} \right) = \text{rank} \left( \begin{bmatrix} \mathbf{C} & \mathbf{E} \end{bmatrix} \right) \quad (24)$$

where  $\mathbf{I}_{n_\Lambda^2, n_d} = [I_{n_d} \dots I_{n_d}]$  ( $n_\Lambda^2$  times the matrix  $I_{n_d}$ ).

Following the same lines as previously, there exists a matrix  $L_k$  such that for any  $\lambda, \lambda' \in \Lambda$ :  $L_k \begin{bmatrix} C_{\lambda'} D_\lambda & E_{\lambda'} \end{bmatrix} = \begin{bmatrix} I_{n_d} & \mathbf{0}_{n_d} \end{bmatrix}$ ,

where  $L_k = J + Z_k H$  with  $J = \begin{bmatrix} I_{n_d} & \mathbf{0}_{n_d} \end{bmatrix} \begin{bmatrix} C_{\lambda'} D_\lambda & E_{\lambda'} \end{bmatrix}^\dagger$ , and  $Z_k$  is any matrix ( $Z_k$  will be chosen later). Saying that, the proposed state and UI estimator is the following for any  $k \geq \alpha$ :

$$\begin{cases} \hat{\lambda}_{k-\alpha:k+\omega} = \underset{\pi \in \Lambda^{\alpha+\omega}}{\text{argmin}} [\log |\Sigma_\pi| + \delta(H_\pi y_{k-\alpha:k+\omega}, \pi)] \\ X_k = [(I_{n_x} - GC_{\hat{\lambda}_k}) \tilde{P}_k C_{\hat{\lambda}_k}^T \\ \quad + GV_{\hat{\lambda}_k}] H^T [HC_{\hat{\lambda}_k} \tilde{P}_k C_{\hat{\lambda}_k}^T H^T + HV_{\hat{\lambda}_k} H^T]^{-1} \\ Z_k = J(C_{\hat{\lambda}_k} \tilde{P}_k C_{\hat{\lambda}_k}^T) H^T [(H(C_{\hat{\lambda}_k} \tilde{P}_k C_{\hat{\lambda}_k}^T) H^T)^{-1} \\ K_k = G + X_k H, \quad L_k = J + Z_k H \\ \hat{x}_k = \tilde{x}_k + K_k (y_k - C_{\hat{\lambda}_k} \tilde{x}_k), \quad \hat{d}_{k-1} = L_k (y_k - C_{\hat{\lambda}_k} \tilde{x}_k) \\ P_k = (I_{n_x} - K_k C_{\hat{\lambda}_k}) \tilde{P}_k (I_{n_x} - K_k C_{\hat{\lambda}_k})^T + K_k V_{\hat{\lambda}_k} K_k^T \\ \tilde{x}_{k+1} = A_{\hat{\lambda}_k} \hat{x}_k, \quad \tilde{P}_{k+1} = \gamma^2 A_{\hat{\lambda}_k} P_k A_{\hat{\lambda}_k}^T + F_{\hat{\lambda}_k} W_{\hat{\lambda}_k} F_{\hat{\lambda}_k}^T \end{cases} \quad (25)$$

**Remark 1** *The global structure of algorithm (25) comes from the structure of the Kalman filter in the presence of a UI [8]. It is adapted using what precedes in order to take into account the case of an unknown switching sequence. Moreover, the case of the UI estimation is also added here. The matrix  $X_k$  (resp.  $Z_k$ ) is constructed in order to minimize the trace of the covariance matrix  $P_k$  (resp.  $P_{d_k} = E[e_{d_k}^T e_{d_k}]$ ), under the constraint of the unknown switching sequence.*

In equations (25),  $\gamma$  is a scale factor chosen strictly greater than 1. The aim of such a factor is to enforce a certain convergence rate (more on this in the stability analysis given in [3]).

**Assumption 4** *For any switching sequences  $\pi, \pi' \in \Lambda^{\alpha+\omega+1}$ , such that  $\pi[\alpha+1] \neq \pi'[\alpha+1]$ , the switching sequences  $\pi$  and  $\pi'$  are assumed to be jointly observable.*

**Theorem 2** *Let assumptions 1, 2, 3 and 4 hold. Then, equations (25) define both state and UI estimators for system (1) with estimation errors  $e_k = x_k - \hat{x}_k$  and  $e_{d_k} = Ld_k - \hat{d}_k$  having the following dynamics:  $e_{k+1} = (I - K_{k+1} C_{\hat{\lambda}_{k+1}}) A_{\hat{\lambda}_k} e_k + \xi_k$  and  $e_{d_k} = -L_{k+1} C_{\hat{\lambda}_{k+1}} A_{\hat{\lambda}_k} e_k + \zeta_k$ , where  $\xi_k$  and  $\zeta_k$  are bounded in mean square, i.e. there exist positive constants  $\rho_\xi$  and  $\rho_\zeta$  such that for all  $k \geq 0$ ,  $\|\xi_k\|^2 \leq \rho_\xi$  and  $\|\zeta_k\|^2 \leq \rho_\zeta$ .*

**PROOF.** Due to space limitation, only the proof on the state estimation error  $e_k = x_k - \hat{x}_k$  convergence is established, the one of the UI estimation error  $e_{d_k}$  follows the same line. Using equations (25) and (23) the dynamics of the state estimation error is:  $e_{k+1} = (I - K_{k+1} C_{\hat{\lambda}_{k+1}}) A_{\hat{\lambda}_k} e_k + \xi_k$ , where  $\xi_k = [(A_{\lambda_k} - A_{\hat{\lambda}_k}) - K_{k+1} (C_{\lambda_{k+1}} A_{\lambda_k} - C_{\hat{\lambda}_{k+1}} A_{\hat{\lambda}_k})] x_k + (I - K_{k+1} C_{\hat{\lambda}_{k+1}}) F_{\lambda_k} w_k - K_{k+1} v_{k+1}$ . Let prove the boundedness in mean-square of  $\xi_k$ . Its second and third terms

are clearly bounded. Let be  $\lambda_{k-\alpha:k-1}$  the true discrete state sequence between  $k - \alpha$  and  $k - 1$ . It comes:

$$x_k = \Phi_{\lambda_{k-\alpha:k-1}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k-1} \end{bmatrix} + \Psi_{\lambda_{k-\alpha:k-1}} w_{k-\alpha:k-1}, \quad \text{where}$$

$\Phi_{\lambda_{k-\alpha:k-1}}$  and  $\Psi_{\lambda_{k-\alpha:k-1}}$  are given by equation (6). Thus, the first term of  $\xi_k$  can be written as in equation (7). Here again, the second term is clearly bounded in mean square. Following the same lines as in [3], let prove the bounded-

ness in mean-square of  $\Phi_{\lambda_{k-\alpha:k}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k} \end{bmatrix}$ . Given the state  $x_{k-\alpha}$  and the unknown input sequence  $d_{k-\alpha:k+\omega}$ , the mean

square error of  $\Phi_{\lambda_{k-\alpha:k-1}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k-1} \end{bmatrix}$  is given by:

$$\begin{aligned} E & \left[ \left\| \Phi_{\lambda_{k-\alpha:k-1}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k-1} \end{bmatrix} \right\|^2 | x_{k-\alpha}, d_{k-\alpha:k+\omega}, \pi \right] \\ & = \sum_{\lambda' \in \Lambda, \lambda' \neq \lambda_k} \mathbb{P} \left( \hat{\lambda}_k = \lambda' | x_{k-\alpha}, d_{k-\alpha:k+\omega}, \pi \right) \left\| \Phi_{\lambda_{k-\alpha:k-1}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k-1} \end{bmatrix} \right\|^2. \end{aligned} \quad (26)$$

On one hand, it comes:  $\left\| \Phi_{\lambda_{k-\alpha:k-1}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k-1} \end{bmatrix} \right\|^2 \leq \bar{\Phi} (\|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k+\omega}\|^2)$ ,

where  $\bar{\Phi} = \max_{\pi_\alpha \in \Lambda^\alpha} \lambda_{\max} \{ \Phi_{\pi_\alpha}^T \Phi_{\pi_\alpha} \}$ . On the other hand, it comes:

$$\begin{aligned} & \mathbb{P} \left( \hat{\lambda}_k = \lambda' | x_{k-\alpha}, d_{k-\alpha:k+\omega}, \pi \right) \\ & = \sum_{\pi'_\alpha \in \Lambda^\alpha, \pi'_\omega \in \Lambda^\omega} \mathbb{P} \left( \hat{\lambda}_{k-\alpha:k+\omega} = \pi'_\alpha \otimes \lambda' \otimes \pi_\omega | x_{k-\alpha}, d_{k-\alpha:k+\omega}, \pi \right) \\ & \leq \text{card}(\Lambda)^{\alpha+\omega} \left[ 1 - \gamma_{(\alpha+\omega+1)n_y} (\kappa (\|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k+\omega}\|^2) + \nu) \right] \end{aligned} \quad (27)$$

the last inequality coming from theorem 1, where  $\kappa = \min_{\pi, \pi' \in \Lambda^{\alpha+\omega+1}} \frac{k(\pi, \pi')}{h(\pi, \pi')}$  and  $\nu = \min_{\pi, \pi' \in \Lambda^{\alpha+\omega+1}} \frac{1}{h(\pi, \pi')} \log \frac{|\Sigma_\pi|}{|\Sigma_{\pi'}|}$ . Note that thanks to assumption 4 and theorem 1, the constant  $\kappa$  is positive. Then, using both previous items, it comes that the quantity

$$E \left[ \left\| \Phi_{\lambda_{k-\alpha:k-1}} \begin{bmatrix} x_{k-\alpha} \\ d_{k-\alpha:k-1} \end{bmatrix} \right\|^2 | x_{k-\alpha}, d_{k-\alpha:k+\omega}, \pi \right] \quad (28)$$

is smaller than:

$$\begin{aligned} & \text{card}(\Lambda)^{\alpha+\omega} \bar{\Phi} (\|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k+\omega}\|^2) \\ & \times \left[ 1 - \gamma_{(\alpha+\omega+1)n_y} (\kappa (\|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k+\omega}\|^2) + \nu) \right]. \end{aligned} \quad (29)$$

Then, recalling that  $\gamma_n(s)$  (with any positive integer  $n$ ) converges exponentially to 1 as  $s$  tends to infinity, we can conclude that (29) converges to 0 as  $\|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k}\|^2$  tends to infinity ( $\kappa$  being positive). Finally, noting that (29) depends continuously on  $\|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k+\omega}\|^2$  on the interval  $[0; \infty)$ , it comes that (29) can be upper bounded by some constant independent from  $\|x_{k-\alpha}\|^2 + \|d_{k-\alpha:k+\omega}\|^2$ .

**Remark 2** Note that in order to apply equations (25), a delay equal to  $\omega$  is needed. In order to avoid such a delay, it is possible to use the following equations for  $i = k+1 \dots k+\omega$  (so that  $x_{k+\omega}$  is estimated using  $y_{k-\alpha:k+\omega}$ ):

$$\begin{cases} X_i = [(I - GC_{\hat{\lambda}_i}) \tilde{P}_i + GV_k] C_{\hat{\lambda}_i}^T H^T [HC_{\hat{\lambda}_i} \tilde{P}_i C_{\hat{\lambda}_i}^T H^T + HV_i H^T]^{-1} \\ Z_i = J(C_{\hat{\lambda}_i} \tilde{P}_i C_{\hat{\lambda}_i}^T) H^T [(H(C_{\hat{\lambda}_i} \tilde{P}_i C_{\hat{\lambda}_i}^T) H^T)^{-1} \\ K_i = G + X_i H, \quad L_i = J + Z_i H \\ \hat{x}_i = \tilde{x}_i + K_i (y_i - C_{\hat{\lambda}_i} \tilde{x}_i), \quad \hat{d}_{i-1} = L_i (y_i - C_{\hat{\lambda}_i} \tilde{x}_i) \\ P_i = (I - K_i C_{\hat{\lambda}_i}) \tilde{P}_i (I - K_i C_{\hat{\lambda}_i})^T + K_i V_i K_i^T \\ \tilde{x}_{i+1} = A_{\hat{\lambda}_i} \tilde{x}_i, \quad \tilde{P}_{i+1} = \gamma^2 A_{\hat{\lambda}_i} P_i A_{\hat{\lambda}_i}^T + F_{\hat{\lambda}_i} W_i F_{\hat{\lambda}_i}^T \end{cases} \quad (30)$$

**Remark 3** The stability results stated in [3] remain true by considering  $X_k$  instead of  $K_k$  (as in [3]) for the gain matrix, and  $(I - GC_{\hat{\lambda}_{k+1}} - X_k HC_{\hat{\lambda}_{k+1}}) A_{\hat{\lambda}_k}$  instead of  $(I - K_k C_{\hat{\lambda}_{k+1}}) A_{\hat{\lambda}_k}$  (as in [3]) for the transition matrix of the state error vector, as well as some minor adaptations. They are not detailed here for a lack of space.

## 5 Illustrative Example

Let consider the following academic example (adapted from [3]) of the form (1) where  $\lambda \in \Lambda = \{1; 2\}$ ,  $\Delta = 0.1$ ,

$$\begin{aligned} A_1 & = \begin{bmatrix} \cos(\omega_1 \Delta) & -\omega_1 \sin(\omega_1 \Delta) \\ \frac{1}{\omega_1} \sin(\omega_1 \Delta) & \cos(\omega_1 \Delta) \end{bmatrix}, \quad \text{with } \omega_1 = 0.5, \\ A_2 & = \begin{bmatrix} \cos(\omega_2 \Delta) & -\omega_2 \sin(\omega_2 \Delta) \\ \frac{1}{\omega_2} \sin(\omega_2 \Delta) & \cos(\omega_2 \Delta) \end{bmatrix}, \quad \text{with } \omega_2 = 1, D_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \\ D_2 & = \begin{bmatrix} 0 \\ 2 \end{bmatrix}, F_1 = F_2 = I_2, C = \begin{bmatrix} 0 & 1 \\ 2 & 1 \\ 1 & 0 \end{bmatrix}, E = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, w_k \text{ and } v_k \end{aligned}$$

are Gaussian noises with covariance matrices  $W = 0.01I_{n_x}$  and  $V = 0.01I_{n_y}$  respectively, the unknown input is set to  $d_k = 3 \sin(\pi/100k)$  for  $k \geq 0$ , the simulation is launched for 400 time steps, with  $\gamma = 1.1$ . By setting  $\alpha = 3$  and  $\omega = 2$ , we can check that assumptions 1, 2 and 4 are satisfied. The Delayed Maximum Likelihood Estimator with UI (DMLEUI) of equations (25) is compared with its Undelayed version presented in remark 2 ('UMLEUI'), and with the Switching Kalman Filter with Perfect Information ('SKFPI'), i.e. the same filter applied with the true switching parameter at each time step. Table 1 gives the performance of each estimator in terms of the Root Mean Square Error (RMSE), defined for

the  $i$ th component by  $RMSE(x[i]) = \sqrt{\frac{1}{400} \sum_{k=0}^{399} (x_k[i] - \hat{x}_k[i])^2}$ .

It is clear that the SKFPI is only slightly better to the DMLEUI. This is due to the possible error made by estimating the switching sequence in the MLEUI estimator. Besides, this error increases when there are less measurements available as in the case of the UMLEUI, which explains the slightly worse performance of this former.

Table 1  
RMSE (average on 100 simulations)

RMSE on...	SKFPI	DMLEUI	UMLEUI
$x[1]$	0.1000	0.1000	0.7011
$x[2]$	5.1225	5.1689	5.2434

## 6 Conclusions

The main contribution of this paper is the development of an observer for switching systems affected by Gaussian noises and Unknown Inputs (UI) with unknown switching sequence. The main contribution holds in the way of dealing with the UI in order to decouple it from the state estimation error. First, a method that estimates the switching parameter with maximum likelihood is proposed despite the presence of a UI. Then, this method is used in both state and UI estimators.

## References

- [1] Angelo Alessandri, Marco Baglietto, and Giorgio Battistelli. Receding-horizon estimation for switching discrete-time linear systems. *IEEE Transactions on Automatic Control*, 50(11):1736–1748, 2005.
- [2] Angelo Alessandri, Marco Baglietto, and Giorgio Battistelli. Luenberger observers for switching discrete-time linear systems. *International Journal of Control*, 80(12):1931–1943, 2007.
- [3] Angelo Alessandri, Marco Baglietto, and Giorgio Battistelli. A maximum-likelihood kalman filter for switching discrete-time linear systems. *Automatica*, 46(11):1870–1876, 2010.
- [4] Angelo Alessandri and Paolo Coletta. Switching observers for continuous-time and discrete-time linear systems. In *Proceedings of the 2001 American Control Conference (Cat. No. 01CH37148)*, volume 3, pages 2516–2521. IEEE, 2001.
- [5] Yaakov Bar-Shalom, Subhash Challa, and Henk AP Blom. Imm estimator versus optimal estimator for hybrid systems. *IEEE Transactions on Aerospace and Electronic Systems*, 41(3):986–991, 2005.
- [6] Francisco Javier Bejarano, Alessandro Pisano, and Elio Usai. Finite-time converging jump observer for linear switched systems with unknown inputs. *IFAC Proceedings Volumes*, 42(17):352–357, 2009.
- [7] Oswaldo Luiz 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, 2002.
- [8] Mohamed Darouach, Michel Zasadzinski, and Mohamed Boutayeb. Extension of minimum variance estimation for systems with unknown inputs. *Automatica*, 39(5):867–876, 2003.
- [9] Héctor Ríos, Jorge Davila, and Leonid Fridman. High-order sliding mode observers for nonlinear autonomous switched systems with unknown inputs. *Journal of the Franklin Institute*, 349(10):2975–3002, 2012.