



**HAL**  
open science

## Identification récursive de systèmes à commutations

Laurent Bako, Stéphane Lecoeuche

► **To cite this version:**

Laurent Bako, Stéphane Lecoeuche. Identification récursive de systèmes à commutations. Journées Doctorales / Journées Nationales MACS, 2007, Reims, France. pp.Actes JnJdMacs. hal-00271687

**HAL Id: hal-00271687**

**<https://hal.science/hal-00271687v1>**

Submitted on 9 Apr 2008

**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.

# Identification récursive de systèmes à commutations

Laurent BAKO<sup>1,2</sup>, Stéphane LECOEUCE<sup>1,2</sup>

<sup>1</sup> Ecole des Mines de Douai  
Département Informatique et Automatique, 59508, Douai, France

<sup>2</sup>Laboratoire d'Automatique, Génie Informatique et Signal, UMR CNRS 8146  
Université des Sciences et Technologies de Lille, 59655, Villeneuve d'Ascq, France

bako@ensm-douai.fr, lecoeuce@ensm-douai.fr  
<http://www.ensm-douai.fr>

*Résumé*— Cette communication concerne l'identification de systèmes linéaires à commutations. Ce problème, particulièrement délicat à cause du couplage entre l'état discret et l'état continu est traité de façon récursive via une alternance entre la classification des données de régression et l'estimation des paramètres des sous-modèles. La méthode proposée, sans être moins performante, présente un coût calculatoire réduit par rapport aux méthodes existantes. De plus, son caractère récursif lui permet de surmonter, moyennant une bonne initialisation, de légères variations de paramètres, inévitables dans un processus réel. Une comparaison avec une approche bayésienne récemment proposée dans la littérature démontre l'avantage de notre méthode.

*Mots-clés*— systèmes hybrides, systèmes à commutations, identification récursive, modèle ARX, inférence bayésienne.

## I. INTRODUCTION

L'identification de systèmes hybrides à partir des données entrées-sorties est une tâche doublement complexe. En plus d'être un problème d'identification standard, c'est également un problème de partitionnement de données de régression. Ces deux étapes sont si couplées qu'elles sont difficilement dissociables. La connaissance des paramètres caractéristiques de chaque sous-modèle du système permettrait un partitionnement immédiat (estimation de l'état discret) de l'ensemble des données ; de même, si les séquences de données disponibles étaient séparées par sous-modèles générateurs, il ne resterait plus qu'à estimer les paramètres de chacun d'eux en recourant aux méthodes classiques de régression. Le problème d'identification de systèmes à commutation peut s'énoncer génériquement de la façon suivante : à partir de données entrées-sorties générées par l'ensemble des modèles en interaction et collectées de façon aléatoire, estimer aussi bien les ordres des sous-modèles que leur nombre et leurs paramètres sachant que les ordres peuvent différer d'un sous-modèle à un autre et que les commutations peuvent survenir à tout instant. Il est même possible que les sous-modèles soient de structures différentes (des modèles linéaires mélangés à des modèles non-linéaires par exemple). Dans le cadre de cet article nous nous limiterons à des systèmes linéaires commutants ou linéaires par morceaux représentés par des modèles ARX.

Parmi les approches récemment proposées pour l'estimation de ARX hybrides, on distingue celles qui estiment simultanément l'état discret et les paramètres des sous-modèles [8], [3], [11] et celles qui procèdent d'abord à une

estimation complète de l'un ([13] détermine d'abord les paramètres tandis que [6], [9] recherchent premièrement l'état discret) avant de calculer l'autre. Une simplification assez courante consiste à supposer que les différents sous-modèles sont d'un même ordre et que cet ordre ainsi que le nombre de modes sont connus a priori. Nous ferons de même dans ce papier bien qu'en pratique, la méthode proposée ne requiert pas la connaissance du nombre de sous-modèles mais seulement d'un majorant de celui-ci.

La plupart des méthodes publiées n'ont recours aux moindres carrés qu'après partitionnement des données pour calculer les paramètres du système. Nous proposons ici une méthode simple basée sur les moindres carrés récursifs pour l'estimation simultanée des paramètres et des états discrets.

Une autre famille de méthodes basées sur une estimation directe de modèles d'état traitent de l'identification de systèmes linéaires commutants [10], [2], [4]. L'application de ces méthodes requièrent un temps de séjour plus ou moins important dans les différents régimes locaux. Dans le présent travail, l'avantage des modèles ARX est qu'ils permettent de traiter des commutations complètement arbitraires à la condition que le système global reste stable. Le caractère récursif de la méthode proposée la prédispose à une application en ligne et lui permet de surmonter des variations éventuelles de paramètres.

Cet article est structuré de la façon suivante. La partie II présente la problématique. La section III présente le principe de l'approche bayésienne de Juloski [8] en mettant en avant le parallèle fort avec la contribution de la présente communication. Dans la section IV, nous proposons une procédure séquentielle, fondée sur une initialisation maîtrisée, pour l'estimation de systèmes hybrides. L'algorithme d'initialisation utilisé est celui proposé par Vidal dans [13]. Des résultats obtenus par simulation montrent dans la dernière partie qu'en comparaison avec l'approche bayésienne, notre méthode possède l'avantage d'être peu gourmande en temps de calcul pour des performances équivalentes voire, pour l'exemple donné, supérieures.

## II. FORMULATION DU PROBLÈME

Considérons un système MISO, commutant entre  $s$  modes opératoires représentés chacun par le modèle de pré-

diction suivant :

$$\mathcal{M}_i : y_t = \theta_i^\top x_t + e_t, \quad i = 1, \dots, s \quad (1)$$

où  $x_t \in \mathbb{R}^n$  est le vecteur de régression,  $\theta_i \in \mathbb{R}^n$  le vecteur de paramètres,  $n = n_a + n_b n_u$ .

$$x_t = [ y_{t-1} \quad \dots \quad y_{t-n_a} \quad u_{t-1}^\top \quad \dots \quad u_{t-n_b}^\top ]^\top$$

$u_t \in \mathbb{R}^{n_u}$  désigne le vecteur des entrées du système,  $y_t \in \mathbb{R}$  la sortie,  $e_t \in \mathbb{R}$  compte pour les perturbations, le bruit de mesure éventuel et les incertitudes. Les entiers  $n_a$  et  $n_b$  font référence aux ordres du système. Dans le cadre de cet article nous les supposons connus et invariants pour tous les sous-modèles. Dans le texte, les expressions régime opératoire, sous-modèle ou encore modèle local désignent le comportement local (rapporté à une certaine région de l'espace de régression par exemple) du système. Les termes état discret, classe ou mode renvoient à l'indice entier du mode opératoire.

Disposant des mesures entrées-sorties générées par le système caractérisé par (1) sur un certain horizon de temps  $[0, t]$ , on désire à la fois assigner chaque paire de données  $(x_t, y_t)$  au sous-modèle qui en est le générateur le plus vraisemblable et estimer le vecteur de paramètres de ce sous-modèle. Il est évident que la résolution de ce problème passe par un partitionnement des données selon les différents modes opératoires [6], [9]. Il faut bien noter qu'une classification basée sur la proximité spatiale des vecteurs de régression  $x_t$  ne permet pas un bon partitionnement des données. Le nombre de modes étant supposé connu a priori, l'idée la plus intuitive consiste à alterner l'assignation des données à une classe et l'estimation des paramètres de cette classe [8], [11]. En notant  $\bar{\theta}_i = [ 1 \quad \theta_i^\top ]^\top$ ,  $\bar{x}_t = [ y_t \quad -x_t^\top ]^\top$ , le problème revient à trouver les hyperplans  $\mathcal{X}_i : \{ \bar{x}_t : \bar{\theta}_i^\top \bar{x}_t = 0 \}$  admettant pour vecteurs normaux les  $\bar{\theta}_i$ . Autrement dit, il s'agit de déterminer un ensemble de sous-espaces disjoints ou non en s'appuyant sur une collection finie de points aléatoirement distribués sur ces derniers. Pour obtenir chaque hyperplan  $\mathcal{X}_i$ , on peut soit en estimer directement une base à partir des vecteurs de régression, soit déterminer une base de son complément orthogonal (ce qui revient à déterminer juste la direction de  $\bar{\theta}_i$ ). Bien que le problème d'identification se pose de ces deux manières, en réalité, la quasi-totalité des méthodes existantes sont basées sur une estimation du complément orthogonal. Parmi les plus récentes, la méthode itérative de Juloski [8] basée sur une approche bayésienne repose sur une idée simple consistant à alterner assignation et estimation. Cette approche spécifique à l'identification de systèmes hybrides partage la même idée de fond que la méthode qui sera développée ici.

Ainsi, pour bien mettre en évidence le parallèle entre les deux procédures, nous commencerons par présenter la méthode itérative de Juloski avant de développer la nôtre.

### III. PRÉSENTATION DE LA MÉTHODE DE JULOSKI

Le principe de cette méthode est le suivant. Chaque vecteur de paramètres  $\theta_i$  est considéré comme une variable aléatoire à laquelle on substitue, dans la procédure d'estimation, la fonction de densité de probabilité (f.d.p)  $p_{\theta_i}(\cdot)$

qui la caractérise. L'objectif de cette démarche est de se placer dans un espace qui permette l'application des règles de mise à jour bayésienne. Notons  $p_{\theta_i}(\theta; t)$  la f.d.p continue de  $\theta_i$  à l'instant (ou l'étape d'itération)  $t$ , définie sur un espace dense  $\Theta_i \subset \mathbb{R}^n$ . La théorie du filtrage particulière [1] permet d'approcher cette f.d.p de la façon suivante.

$$\hat{p}_{\theta_i}(\theta; t) = \sum_{l=1}^M w_{i,l}(t) \delta(\theta - \theta_{i,l}(t)) \quad (2)$$

où  $\delta$  correspond à la distribution de Dirac,  $w_{i,l}(t) > 0$  représentent des poids associés aux  $\theta_{i,l}(t)$  et vérifiant  $\sum_{l=1}^M w_{i,l}(t) = 1$ . L'échantillon fini  $\{\theta_{i,l}(t)\}_{l=1}^M$  est supposé être généré sur l'espace  $\Theta_i$  selon la distribution donnée par  $p_{\theta_i}(\theta; t)$ . De toute évidence, la fiabilité d'une telle approximation est étroitement liée au nombre  $M$  de particules. Soit  $p((x_t, y_t) | \hat{\mu}_t = i)$  la probabilité que le mode  $i$  ait généré la donnée  $(x_t, y_t)$ . L'état discret  $\hat{\mu}_t$  correspond alors au mode qui maximise cette probabilité.

$$\hat{\mu}_t = \arg \max_i p((x_t, y_t) | \hat{\mu}_t = i) \quad (3)$$

avec

$$p((x_t, y_t) | \hat{\mu}_t = i) = \int_{\Theta_i} p((x_t, y_t) | \theta) p_{\theta_i}(\theta; t-1) d\theta \quad (4)$$

et

$$p((x_t, y_t) | \theta) = p_e(y_t - \theta^\top x_t) \quad (5)$$

où  $p_e$  est la f.d.p a priori du bruit additif  $e$  du (1) supposé être une séquence indépendante et identiquement distribuée. En s'aidant de la formule (2), on peut facilement obtenir l'approximation suivante

$$p((x_t, y_t) | \hat{\mu}_t = i) \approx \sum_{l=1}^M w_{i,l}(t-1) p_e(y_t - x_t^\top \theta_{i,l}(t-1)) \quad (6)$$

L'algorithme « sous-optimal » suivant affecte les données aux modes tout en estimant les paramètres de chaque sous-modèle.<sup>1</sup>

1. Initialisation : fournir les f.d.p a priori  $p_{\theta_i}(\cdot; 0)$  pour  $i = 1, \dots, s$ , fournir  $p_e(\cdot)$
2. POUR  $t = 1, \dots, \infty$ 
  - Affecter  $(x_t, y_t)$  au mode  $\hat{\mu}_t$  défini par (3)
  - Evaluer la f.d.p a posteriori  $p_{\theta_{\hat{\mu}_t}}(\cdot; t)$  du paramètre  $\theta_{\hat{\mu}_t}$  comme suit :
    - Diversification des particules :

$$\theta_{\hat{\mu}_t, l}(t) = \theta_{\hat{\mu}_t, l}(t-1) + \varepsilon_l$$

avec  $\varepsilon_l \sim N(0, \Sigma_\varepsilon)$ ,  $l = 1, \dots, M$

- Calcul des poids :

$$w_{\hat{\mu}_t, l}(t) = p((x_t, y_t) | \theta_{\hat{\mu}_t, l}(t)) = p_e(y_t - x_t^\top \theta_{\hat{\mu}_t, l}(t))$$

$$l = 1, \dots, M$$

<sup>1</sup>A l'origine, cet algorithme fonctionne en hors ligne : il y a au moins une revisite des données

– Normalisation des poids :

$$w_{\hat{\mu}_t,l}(t) := \frac{w_{\hat{\mu}_t,l}(t)}{\sum_{l=1}^M w_{\hat{\mu}_t,l}(t)}, l = 1, \dots, M$$

– Rééchantillonnage : De la distribution (2) calculée avec  $\{\theta_{\hat{\mu}_t,l}(t), w_{\hat{\mu}_t,l}(t)\}$ , on génère ensuite indépendamment  $M$  points pour former de nouvelles particules uniformément pondérées  $(\theta_{\hat{\mu}_t,l}(t), \frac{1}{M})$

- Pour tout  $j \neq \hat{\mu}_t$ ,  $p_{\theta_j}(\cdot; t) = p_{\theta_j}(\cdot; t-1)$

Une caractéristique assez évidente de la démarche bayésienne présentée est la complexification qui accompagne l'idée de substituer à chaque variable  $\theta_i$ , une fonction  $p_{\theta_i}(\cdot)$  que l'on sait continue et non-linéaire. L'approximation de celle-ci par une fonction à support fini nécessite pour être efficace que  $M$  soit très grand et qu'un domaine probable de variation de  $\theta_i$  soit connu. Aussi, on estime les f.d.p au prix d'un effort de calcul accru mais sans atténuer sensiblement le caractère « sous-optimal » de l'algorithme. Par ailleurs, l'initialisation d'un nombre aussi élevé de variables  $\{\theta_{i,l}\}$ ,  $i = 1, \dots, s$ ,  $l = 1, \dots, M$  n'est pas sans incidence sur les performances qu'on pourrait obtenir.

Nous proposons dans la partie suivante une approche somme toute assez semblable à l'approche bayésienne. Elle calcule directement et de façon algébrique les paramètres des sous-modèles sans espace de représentation intermédiaire, ce qui lui confère d'être moins exigeante en termes de charge de calcul.

#### IV. ALTERNATIVE PROPOSÉE

L'identification *réursive* (en ligne) de systèmes hybrides a besoin, pour être efficace d'être basée sur un algorithme optimal. Malheureusement, la majorité des méthodes dédiées à l'identification de systèmes hybrides, linéaires par morceaux présentent la faiblesse commune d'être « sous-optimales ». Elles souffrent en effet d'une possible attraction de minima locaux du critère mixte de classification-identification. A notre connaissance, les seuls algorithmes optimaux existants à ce jour sont ceux proposés dans [12] et [13]. Le premier souffre d'une complexité calculatoire qui limite sévèrement son application à des problèmes de petites dimensions; le deuxième est analytique et idéal seulement dans un contexte déterministe. Cette dernière méthode a été étendue à l'estimation réursive de systèmes ARX commutants [7], [14]. C'est, à notre connaissance, la première approche réursive proposée pour l'identification de systèmes hybrides. Cependant, elle reste limité à un contexte déterministe.

La méthode que nous proposons ici bien que tombant aussi dans la catégorie des méthodes « sous-optimales », présente l'avantage d'être simple aussi bien en termes d'implémentation que de charge de calculs, et tout aussi performante que les méthodes existantes.

Le traitement opéré sur les données sera séquentiel c'est-à-dire que l'ensemble des données à traiter n'est pas supposé disponible en une seule fois; à un instant  $t$ , seul l'historique  $(x_k, y_k)$ ,  $k = 1, \dots, t$  constitue la base d'apprentissage. Au lieu de recourir à des techniques de filtrage particulaire comme dans [8] pour assurer l'estimation des paramètres des différents modes opératoires, l'approche adoptée ici est basée sur les moindres carrés rékursifs.

Chaque mode  $i$  sera représenté par le couple  $\{\hat{\theta}_i(t), P_i(t)\}$  avec

$$P_i(t) = \left( \sum_{k=1, \hat{\mu}_k=i}^t \omega_{t,k} x_k x_k^\top + \lambda^{n_i(t)} h I \right)^{-1} \quad (7)$$

$$\hat{\theta}_i(t) = P_i(t) \left( \sum_{k=1, \hat{\mu}_k=i}^t \omega_{t,k} x_k y_k \right)$$

où  $\omega_{t,k} = \lambda^{n_i(t)-n_i(k)}$ ,  $\hat{\mu}_k$  est le mode actif à l'instant  $k$ ,  $n_i(k)$  est le cardinal du mode  $i$  à l'instant  $k$ ,  $\lambda \in [0, 1]$  est un facteur d'oubli,  $h \in \mathbb{R}$ .  $\hat{\theta}_i(t)$  est le vecteur des paramètres du modèle  $i$  à l'instant  $t$ . L'adéquation entre une nouvelle donnée  $(x_t, y_t)$  et chacun des  $s$  sous-modèles est mesurée par un certain critère de décision. Seule la classe gagnante (au sens où ses paramètres minimisent le critère de décision) pourra se mettre à jour avec la donnée selon l'algorithme<sup>2</sup> :

---

#### Algorithme 1 Moindres carrés rékursifs

---

$$\{\hat{\theta}_i(t), P_i(t)\} = \mathbf{MCR} \left\{ \hat{\theta}_i(t-1), P_i(t-1) \mid \hat{\mu}_t = i \right\}$$

$$\varepsilon_i(t) = y_t - \hat{\theta}_i(t-1)^\top x_t$$

$$G_i(t) = \frac{P_i(t-1) x_t}{\lambda + x_t^\top P_i(t-1) x_t}$$

$$\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + G_i(t) \varepsilon_i(t)$$

$$P_i(t) = \frac{1}{\lambda} (I_n - G_i(t) x_t^\top) P_i(t-1)$$


---

Dans tous les cas, l'obtention de bonnes estimées est subordonnée au choix d'un critère de décision adéquat et à une initialisation appropriée des paramètres.

##### A. Choix d'un critère de décision

Le critère de décision pourrait être formellement choisi sous la forme  $J(\theta_1(t-1), \dots, \theta_s(t-1), (x_t, y_t))$ . Etant donnée la rapidité possible des commutations, ce critère a besoin de permettre une décision instantanée c'est-à-dire faisant intervenir le moins possible les informations passées. Ainsi,  $J$  peut être sélectionné comme une fonction directe

– de l'erreur a priori

$$\varepsilon_i(t) = y_t - \hat{\theta}_i(t-1)^\top x_t = \bar{\theta}_i(t-1)^\top \bar{x}_t \quad (8)$$

– ou de l'erreur a posteriori

$$\begin{aligned} \eta_i(t) &= y_t - \hat{\theta}_i(t)^\top x_t = \bar{\theta}_i(t)^\top \bar{x}_t \quad (9) \\ &= \frac{\varepsilon_i(t)}{1 + \lambda^{-1} x_t^\top P_i(t-1) x_t} \end{aligned}$$

Les erreurs (8) et (9) peuvent être interprétées comme des projections orthogonales du vecteur  $\bar{x}_t$  respectivement sur  $\hat{\theta}_i(t-1)$  et  $\bar{\theta}_i(t)$ . Puisque le but de l'exercice est de déterminer les directions  $\bar{\theta}_i$ , il convient de normaliser les vecteurs  $\bar{\theta}_i(t-1)$  et  $\bar{\theta}_i(t)$  dans les équations précédentes. Comme

<sup>2</sup>MCR fait référence dans la suite à l'algorithme des moindres carrés rékursifs

$\bar{\theta}_i \neq 0$ , on obtient :  $\varepsilon_i(t) = \frac{\bar{\theta}_i(t-1)^\top}{\|\bar{\theta}_i(t-1)\|} \bar{x}_t$  et  $\eta_i(t) = \frac{\bar{\theta}_i(t)^\top}{\|\bar{\theta}_i(t)\|} \bar{x}_t$ . La décision consistera à comparer les normes des projections orthogonales de  $\bar{x}_t$  sur les directions  $\bar{\theta}_1(t), \dots, \bar{\theta}_s(t)$ . La normalisation introduite réduit cette comparaison à celle des corrélations (cosinus de l'angle  $(\bar{x}_t, \bar{\theta}_i(t))$ ) de  $\bar{x}_t$  avec chacun des paramètres  $\bar{\theta}_i(t)$ ; elle permet par ailleurs d'interpréter les quantités  $\varepsilon_i(t)$  et  $\eta_i(t)$  comme les distances de  $\bar{x}_t$  aux hyperplans correspondants. Les paramètres du mode le plus vraisemblablement assorti à la paire de mesures  $(x_t, y_t)$  (la direction la moins corrélée avec  $\bar{x}_t$ ) seront ajustés en utilisant cette paire. Il paraît alors assez évident que, suivant cette procédure, la qualité de la décision sera assez tributaire de la nature des modes en compétition, de leur nombre ainsi que de leur degré de séparabilité. La forme du critère basé sur l'erreur a posteriori est particulièrement intéressante pour être remarquée. La minimisation du coût (9) comprend la minimisation du numérateur  $\varepsilon_i(t)$  (erreur a priori) mais aussi la maximisation de la quantité  $x_t^\top P_i(t-1)x_t$  qui influe sur l'évolution de  $\theta_i(t)$  s'il se mettait à jour avec  $(x_t, y_t)$ . Faisant intervenir à la fois le couple  $\{\theta_i(t-1), P_i(t-1)\}$  et la paire  $(x_t, y_t)$ , (9) paraît plus approprié pour décider s'il faut attribuer ou non  $(x_t, y_t)$  au mode  $i$ .

Puisque l'estimation de l'état discret ne tient qu'à une comparaison entre des grandeurs, il n'est pas nécessaire que  $\bar{\theta}_i(t)$ ,  $i = 1, \dots, s$  soit exactement égal à la vraie valeur de  $\theta_i$  pour avoir  $\hat{\mu}_t = \mu_t$ . En effet, la détermination de l'état sera correcte, si pour tout  $i$ ,  $\exists K$  fini,  $\forall t \geq K$ ,  $\bar{x}_t \in \mathcal{X}_i$  entraîne  $|\bar{\theta}_j^\top \bar{x}_t| > |\bar{\theta}_i^\top \bar{x}_t|, \forall j \neq i$ . Par ailleurs, tant que l'état discret  $i$  reste correctement estimé, la convergence de  $\bar{\theta}_i$  sera obtenue. Pour garantir cela, une bonne initialisation est déterminante voire indispensable pour l'efficacité de cette méthode.

---

### Algorithme 2 Classification basée sur l'erreur a posteriori

---

- Initialisation :  $P_i(0) := \frac{1}{h}I$ , avec  $h \ll 1$ , les  $\hat{\theta}_i(0)$ ,  $i = 1, \dots, s$  sont tirés de façon aléatoire  
POUR  $t = 1, \dots, \infty$
  - Calculer et normaliser  $J_i(t) = \frac{\eta_i(t)^2}{1 + \|\bar{\theta}_i(t)\|_2^2}$ ,  $i = 1, \dots, s$
  - $\hat{\mu}_t = \arg \min_i J_i(t)$
  - $\{\hat{\theta}_{\hat{\mu}_t}(t), P_{\hat{\mu}_t}(t)\} = \mathbf{MCR} \left\{ \hat{\theta}_{\hat{\mu}_t}(t-1), P_{\hat{\mu}_t}(t-1) \right\}$
  - $\forall j \neq \hat{\mu}_t, \{\hat{\theta}_j(t), P_j(t)\} = \{\hat{\theta}_j(t-1), P_j(t-1)\}$
- 

### B. Discussion autour de l'initialisation

L'initialisation est certainement la phase la plus critique dans l'approche récursive à l'identification de systèmes hybrides proposée. Cela est dû au fait que les tâches de classification et d'identification sont accomplies alternativement l'une après l'autre. La propagation d'erreurs se trouve donc être particulièrement amplifiée. Pour pallier cet inconvénient, il paraît assez intéressant de recourir à un algorithme d'initialisation.

Supposons que chaque mode a été visité au moins  $K$  fois ( $K > n$ ) sur la séquence de données  $\{1, \dots, T\}$  avec  $T > \binom{n+s}{n}$ . Nous noterons dans la suite  $M_s := \binom{n+s}{n}$ . Si on suppose de plus que la quantité de bruit susceptible d'affecter les données est modérée, alors, on peut appliquer la méthode algebro-géométrique développée par Vidal [13].

L'idée de base de cette méthode consiste à se débarrasser du couplage entre l'étape de filtrage de l'état discret et celle d'estimation des modes en trouvant une condition vérifiée par toutes les données sans exception. En effet, pour tout vecteur  $\bar{x}_t, t > n_a$  il existe (dans le cas purement déterministe)  $\bar{\theta}_i$  tel que  $\bar{\theta}_i^\top \bar{x}_t = 0$ . D'où la contrainte de découplage hybride :

$$\prod_{i=1}^s (\bar{\theta}_i^\top \bar{x}_t) = 0 \quad \forall t \geq n_a \quad (10)$$

Cette contrainte n'est rien d'autre que la caractérisation des racines d'un polynôme homogène  $P_s$  de degré  $s$  (égal au nombre de modes) et de  $n+1$  variables

$$P_s(z) = \prod_{i=1}^s (\bar{\theta}_i^\top z) = \sum_{s_1 + \dots + s_{n+1} = s} h_{s_1, \dots, s_{n+1}} z_1^{s_1} \dots z_{n+1}^{s_{n+1}}$$

La relation (10) peut encore se lire

$$P_s(z) = h_s^\top \nu_s(z) = 0 \quad (11)$$

où  $\nu_s : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{M_s}$  est l'application dite de Veronese de degré  $s$  qui, au vecteur  $z = [z_1 \dots z_{n+1}]^\top$  associe un vecteur de  $M_s$  monômes  $z_1^{s_1} \dots z_{n+1}^{s_{n+1}}$ ,  $0 \leq s_j \leq s$ ,  $s_1 + \dots + s_{n+1} = s$ , disposés dans un vecteur selon l'ordre lexicographique.

On estime d'abord les coefficients du polynôme  $P_s$  dont les  $\bar{\theta}_i$  sont ensuite déduits en une seule fois, sans itération. Il suffit alors de connaître un point  $z_j$  sur chaque hyperplan  $\mathcal{X}_j$  pour en obtenir une normale  $\bar{\theta}_j$  comme suit :

$$\bar{\theta}_j = \frac{DP_s(z_j)}{e_1^\top DP_s(z_j)}, j = 1, \dots, s \quad (12)$$

où  $DP_s$  désigne la dérivée de  $P_s$  par rapport à  $z$  et  $e_1$ , le vecteur de même dimension que  $\theta_j$  possédant 1 comme premier élément et des zéros partout ailleurs. Pour plus de lisibilité sur le principe de cette méthode, des précisions sont données en annexe, mais le lecteur intéressé pourra toujours se référer aux travaux de Vidal [13].

Sous l'hypothèse d'une assez bonne initialisation, il est possible d'effectuer, dans la tâche de classification, une décision plus précise en intégrant par exemple des heuristiques comme l'indécidabilité ou le rejet en ambiguïté. Par exemple, si  $\forall i, J_i(t) > \varepsilon_o$ ,  $\varepsilon_o$  étant un certain seuil de performance fixé, la paire  $(x_t, y_t)$  peut être mise en attente de recyclage, soit systématiquement, soit périodiquement. Une accumulation importante de points inclassables pourrait alors donner lieu par exemple à la création d'un nouveau mode de fonctionnement.

**Remarque 1.** *L'approximation particulière par les méthodes de Monte-Carlo [1],[5], charpente de l'approche bayésienne présentée à la section III se traduit dans notre cas par une prospection de l'espace  $\Theta_i \subset \mathbb{R}^{n+1}$ , pouvant consister dans l'initialisation et l'estimation en parallèle de  $M$  valeurs  $\theta_{i,l}$  pour chaque mode  $i$ . Ce procédé est à rapprocher des algorithmes génétiques ou plus généralement des algorithmes dits non-déterministes qui sont connus pour être d'une grande utilité dans la recherche d'optima de fonctions non convexes.*

## V. SIMULATIONS

Pour illustrer la méthode présentée ci-dessus, on considère un exemple de système commutant entre trois sous-modèles linéaires d'ordres invariants  $n_a = 2$  et  $n_b = 1$  définis tels que :  $y_t = \theta_i(1)y_{t-1} + \theta_i(2)y_{t-2} + \theta_i(3)u_{t-1}$ ,  $i = 1, 2, 3$ .

$$\theta_1 = \begin{bmatrix} -0.6159 \\ -0.5476 \\ 0.60 \end{bmatrix} \quad \theta_2 = \begin{bmatrix} 1.60 \\ -0.6175 \\ -1.20 \end{bmatrix} \quad \theta_3 = \begin{bmatrix} 0.5106 \\ -0.810 \\ 3.50 \end{bmatrix}$$

L'entrée du système est choisie comme un bruit blanc centré de variance unité ; un bruit additif de sortie  $e$  est choisi également blanc tel que le RSB soit de 35 dB. Il n'est fait aucune hypothèse sur la nature des commutations ; elles peuvent se produire à tout instant, c'est-à-dire qu'il n'y a pas de temps de séjour minimum requis entre deux changements successifs de modes. Pour comparer notre méthode à la méthode proposée par Juloski et al., une simulation de Monte-Carlo de 1000 jeux a été réalisée en utilisant les deux méthodes sur un horizon de 500 points. Les histogrammes Fig. 4 montrent que la méthode que nous proposons ici est préférable (même sans faire appel à un algorithme spécifique d'initialisation) en terme de performances à celle basée sur une approximation particulière. Sur les figures Fig. 1 et Fig. 2 sont représentés les paramètres des sous-modèles estimés en ligne par simulation. Sont donnés ci-après les résultats obtenus :

– par notre méthode (sans initialisation spéciale) :

$$\hat{\theta}_1 = \begin{bmatrix} -0.6160 \\ -0.5477 \\ 0.6253 \end{bmatrix} \quad \hat{\theta}_2 = \begin{bmatrix} 1.5998 \\ -0.6166 \\ -1.2187 \end{bmatrix} \quad \hat{\theta}_3 = \begin{bmatrix} 0.5105 \\ -0.8102 \\ 3.5160 \end{bmatrix}$$

– par la méthode de Juloski :

$$\hat{\theta}_1 = \begin{bmatrix} -0.6160 \\ -0.5475 \\ 0.5951 \end{bmatrix} \quad \hat{\theta}_2 = \begin{bmatrix} 1.6000 \\ -0.6175 \\ -1.1604 \end{bmatrix} \quad \hat{\theta}_3 = \begin{bmatrix} 0.5106 \\ -0.8099 \\ 3.5170 \end{bmatrix}$$

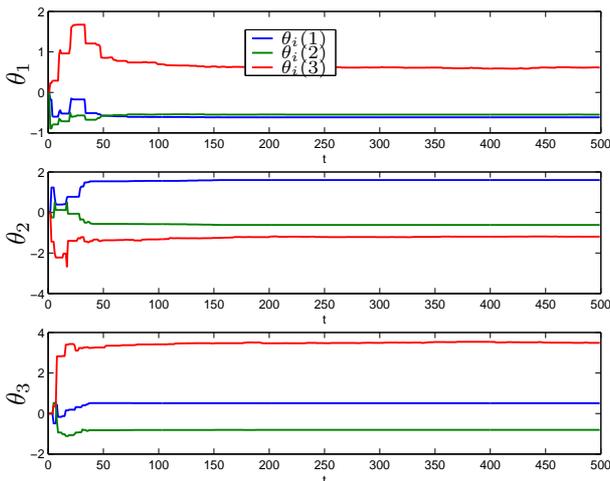


Fig. 1. Estimation récursive des paramètres  $\theta_i$  des trois sous-modèles du système par notre méthode sur un horizon de 500 points. 96% de points ont été bien classés

D'autres simulations montrent qu'en deçà de 30 dB, en faisant usage de l'algorithme algèbro-géométrique, l'efficacité de l'initialisation n'est plus très significative. Mais l'algorithme 2 fournit toujours même sans initialisation rigoureuse des résultats satisfaisants. En fait, la connaissance

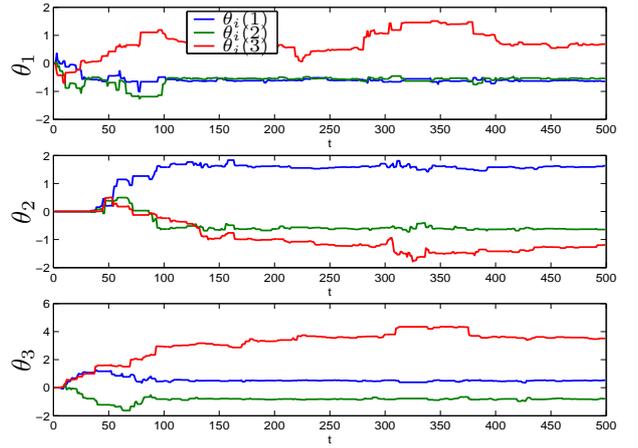


Fig. 2. Estimation récursive des paramètres  $\theta_i$  des trois sous-modèles du système par la méthode de Juloski :  $M = 200$ ,  $\Sigma_e = 1e^{-2}I$ , la f.d.p  $p_e$  est choisie normale de moyenne nulle et de variance 0.1, les  $p_{\theta_i}(\cdot; 0)$  sont des f.d.p uniformes. Les trajectoires des paramètres paraissent moins lisses que sur la Fig. 1

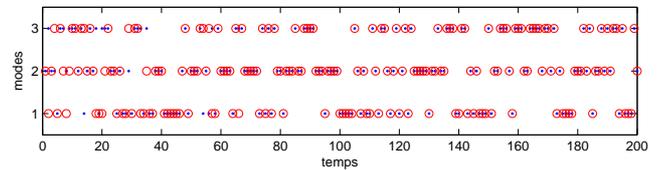


Fig. 3. Représentation de l'état discret : les points bleus représentent l'état discret réel, les cercles rouges indiquent son estimée

d'un majorant du nombre de sous-modèles suffit en pratique, dans le cadre de notre méthode, à la détermination du nombre réel de sous-modèles ainsi que de leurs paramètres. Par contre, une connaissance relativement précise de l'ordre est requise (mais  $n_a$  peut être pris égal à  $n_b$  sans inconvénient).

## VI. CONCLUSION

Nous avons suggéré dans ce papier une méthode pour l'identification récursive de systèmes pouvant commuter instantanément entre plusieurs modes de fonctionnement de même structure. L'approche proposée se construit autour d'une idée qui consiste à coupler les tâches de classification et d'estimation des paramètres. En comparaison avec la méthode bayésienne de Juloski, il a été montré que pour des performances équivalentes, notre méthode avait l'avantage d'être peu coûteuse en terme de charges de calcul. Il est difficile d'établir, dans le contexte récursif envisagé, des preuves de convergence sous quelques hypothèses tant le couplage classification-identification est inextricable. Nos prochains travaux s'étendront à l'identification récursive des systèmes hybrides MIMO sujets à des commutations pouvant être instantanées.

### ANNEXE : MÉTHODE ALGÈBRE-GÉOMÉTRIQUE

Dans la démarche algèbro-géométrique rappelée en IV, l'idée est de se départir du couplage entre l'étape de filtrage de l'état discret et celle d'estimation des modes en trouvant une équation satisfaite par toute donnée collectée quelqu'en soit le modèle générateur. Cette méthode est applicable dans le cas où les données sont presque parfaites (c'est à dire sans bruit).

En appliquant la relation (11) à l'ensemble des points  $\{\bar{x}_t\}_{t=n_a+1}^T$ , on obtient le système d'équations linéaires suivant :

$$L_s h_s = \begin{bmatrix} \nu_s(\bar{x}_{n_a})^\top \\ \nu_s(\bar{x}_{n_a+1})^\top \\ \vdots \\ \nu_s(\bar{x}_T)^\top \end{bmatrix} h_s = 0 \in \mathbb{R}^{T-n_a} \quad (13)$$

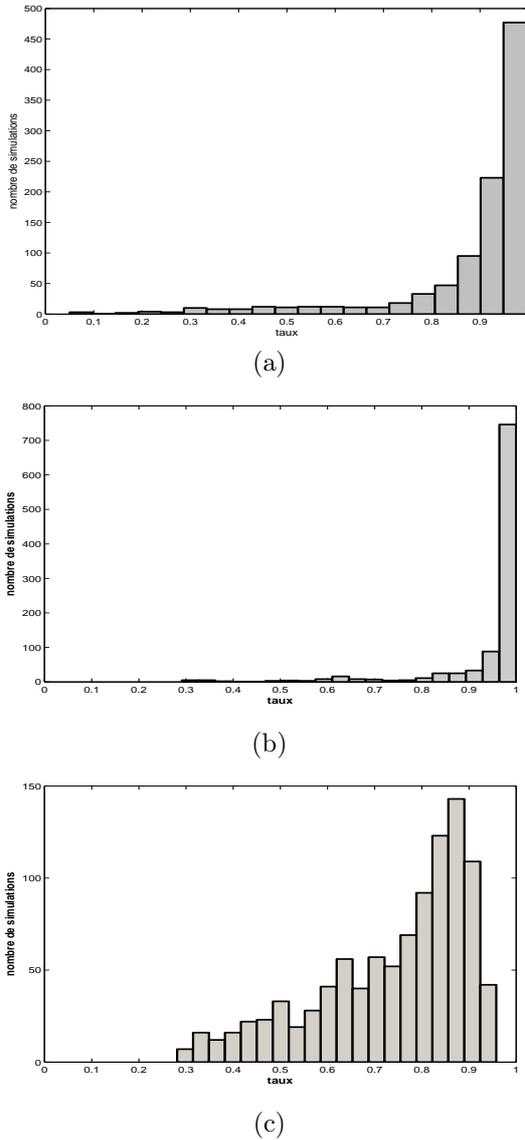


Fig. 4. Histogramme d'une simulation de Monte-Carlo (1000 simulations sur 500 points). L'axe des abscisses indique le taux de points bien classés et l'axe des ordonnées donne le nombre de simulations correspondant. Utilisant notre méthode sans initialisation spéciale (haut : (a)), 85% de simulations présentent un taux supérieur à 0.8 tandis que l'approche bayésienne (bas : (c)) n'en fournit que 50%. Avec l'algorithme d'initialisation (milieu : (b)) on obtient 95% de simulations présentant un taux supérieur à 0.8

$L_s \in \mathbb{R}^{(T-n_a) \times M_s}$  est constituée à partir des données,  $h_s \in \mathbb{R}^{M_s}$  est constitué d'une combinaison des paramètres de tous les sous-modèles.

Il est possible d'estimer  $\bar{\theta}_j$  à partir de la dérivée du polynôme multivariable  $P_s$  mais à condition de connaître au moins un point  $z_j$  dans chaque hyperplan  $\mathcal{X}_j, j = 1, \dots, s$ . Si  $z_0$  est un élément non nul quelconque de  $\mathbb{R}^{n+1}$  et  $v$  un élément de  $\mathbb{R}^{n+1}$  non colinéaire à  $z_0$  et vérifiant  $\bar{\theta}_j^T v \neq 0 \forall j, \mathcal{L} = \{z_0 + \alpha v, \alpha \in \mathbb{R}\}$  définit une droite affine dans l'espace  $\mathbb{R}^{n+1}$ . On peut facilement, compte tenues des contraintes imposées, montrer que la droite  $\mathcal{L}$  non parallèle à aucun des hyperplans  $\mathcal{X}_j$  admet un unique point d'intersection  $z_j = z_0 + \alpha_j v$  avec chacun d'eux,  $\alpha_j$  étant une racine du polynôme monovarié  $Q_s(\alpha) = P_s(z_0 + \alpha v)$ . Ainsi, on peut,

- en fixant  $z_0 \in \mathbb{R}^{n+1} \neq 0, v \in \mathbb{R}^{n+1}$  tel que,  $v \neq \lambda z_0 \forall \lambda$  et  $P_s(v) \neq 0$ ,
- en trouvant les racines  $\alpha_i$  de  $Q_s(\alpha)$  et en posant  $z_j = z_0 + \alpha_j v$ , calculer tous les vecteurs de paramètres

$$\bar{\theta}_j = \frac{DP_s(z_j)}{e_1^T DP_s(z_j)}, j = 1, \dots, s \quad (14)$$

où  $DP_s$  désigne la dérivée de  $P_s$  par rapport à  $z$ . Il faut penser à l'équation (10) comme à celle d'une surface de niveau et à  $\frac{\partial P_s}{\partial z} \Big|_{z_i}$  (=

$\nabla P_s|_{z_i}$ ), comme au vecteur gradient normal à cette surface autour de  $z_i$ . L'hyperplan  $\theta_i^T$  est tangent à cette surface dans le voisinage de  $z_i$ . Notons que la recherche des racines du polynôme  $Q_s(\alpha)$  peut vite devenir fastidieuse si  $s$  devient grand. Cette méthode permet par ailleurs une obtention du nombre de modes  $s$  comme une condition de rang vérifiée par la matrice  $L_s$ . Remarquons que si les données sont sujettes à des perturbations même légères, il n'y a plus aucune raison que la relation (10) soit égale à zéro ni même proche de zéro pour tout système linéaire commutant ARX. Une condition supplémentaire du type

$$\sigma_e \left| \prod_{i \neq \mu_t}^s (\bar{\theta}_i^T \bar{x}_t) \right| \leq \delta \forall t > n_a, \\ \delta \ll 1, \sigma_e = \sup |e(t)|,$$

a besoin d'être imposée pour garantir que le produit (10), à défaut d'être nul, sera au moins dans un voisinage de zéro. Cette remarque est la principale faiblesse de la méthode algèbro-géométrique. Néanmoins, elle représente une option (comme nous le suggérons ici) d'initialisation assez intéressante en présence de bruit.

## RÉFÉRENCES

- [1] M. Sanjeev Arulampalam, Simon Maskell, Neil Gordon, and Tim Clapp. A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50, no. 2 :174–188, 2002.
- [2] Laurent Bako, Guillaume Mercère, and Stéphane Lecoche. Identification en ligne de systèmes commutants à structure variable. In *Journées Identification, Modélisation Expérimentale, Poitiers*, 2006.
- [3] Alberto Bemporad, Andrea Garulli, Simone Paoletti, and Antonio Vicino. A bounded-error approach to piecewise affine system identification. *IEEE Transactions on Automatic and Control*, 50 :1567–1580, 2005.
- [4] José Borges, Vincent Verdult, Michel Verhaegen, and Miguel Ayala Botto. A switching detection method based on projected subspace classification. In *CDC-ECC05*, 2005.
- [5] Dan Crisan and Arnaud Doucet. A survey of convergence results on particle filtering methods for practitioners. *IEEE Transactions on Signal Processing*, 50, no. 3 :736–746, 2002.
- [6] Giancarlo Ferrari-Trecate, Marco Muselli, Diego Liberati, and Manfred Morari. A clustering technique for the identification of piecewise affine systems. *Automatica*, 39 :205–217, 2003.
- [7] Yasmin Hashambhoy and René Vidal. Recursive identification of switched arx models with unknown number of models and unknown orders. In *IEEE Conference on Decision and Control*, 2005.
- [8] A. Lj. Juloski, S. Weiland, and W.P.M.H. Heemels. A bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control*, 50, No 10 :1520–1533, 2005.
- [9] Hayato Nakada, Kiyotsugu Takaba, and Tooru Katayama. Identification of piecewise affine systems based on statistical clustering technique. *Automatica*, 41, No 5 :905–913, 2005.
- [10] Komi Midzodzi Pekpe and Stéphane Lecoche. Online classification of switching models based on subspace framework. In *IFAC Conference on Analysis and Design of Hybrid Systems, Alghero, Sardinia*, 2006.
- [11] José Ragot, Gilles Mourot, and Didier Maquin. Parameter estimation of switching piecewise linear systems. In *Proceedings of the 42nd IEEE Conference on Decision and Control, Maui, Hawaii, USA*, 2003.
- [12] Jacob Roll, Alberto Bemporad, and Lennart Ljung. Identification of piecewise affine systems via mixed-integer programming. *Automatica*, 40 :37–50, 2004.
- [13] R. Vidal, S. Soatto, Y. Ma, and S. Sastry. An algebraic geometric approach to the identification of a class of linear hybrid systems. In *Proceedings. 42nd IEEE Conference on Decision and Control*, 2003.
- [14] René Vidal and Brian D. O. Anderson. Recursive identification of switched arx hybrid models : exponential convergence and persistence of excitation. In *IEEE Conference on Decision and Control*, 2004.