Identifiability of a Switching Markov State-Space Model

Jana Kalawoun[1], Patrick Pamphile[2], Gilles Celeux[2]

[1] CEA, LIST, Laboratoire d'Analyse de Données et Intelligence des Systèmes
Centre CEA Saclay, 91191 Gif-sur-Yvette, France

[2]INRIA Université Paris Sud CNRS UMR8628
15 Rue Georges Clemenceau, 91405 Orsay, France
jana.kalawoun@cea.fr, patrick.pamphile@u-psud.fr
gilles.celeux@inria.fr

Résumé – Les modèles à espaces d'états gouvernés par une chaîne de Markov cachée sont utilisés dans de nombreux domaines appliqués comme le traitement de signal, la bioinformatique, etc. Cependant, il est souvent difficile d'établir leur identifiabilité, propriété essentielle pour l'estimation de leurs paramètres. Dans cet article, nous traitons un cas simple pour lequel l'état continu inconnu et les observations sont des scalaires. Nous démontrons que lorsque la chaîne de Markov est irréductible et apériodique, une information a priori reliant les observations et l'état continu inconnu à un instant t_0 suffit pour assurer "l'identifiabilité générale" de l'ensemble des paramètres du modèle. Nous montrons aussi qu'en intégrant ces contraintes dans un algorithme EM, les paramètres du modèle sont estimés efficacement.

Abstract – While switching Markov state-space models arise in many applied science applications like signal processing, bioinformatics, etc., it is often difficult to establish their identifiability which is essential for parameters estimation. This paper discusses the simple case in which the unknown continuous state and the observations are scalars. We demonstrate that if a prior information relating the observations to the unknown continuous state at a time t_0 is available, and if the Markov chain is irreducible and aperiodic, the set of the model parameters will be "globally structurally identifiable". In addition, we show that under these constraints, the model parameters can be efficiently estimated by an EM algorithm.

1 Introduction

One way of modeling changes in a time series consists of considering that the underlying dynamics of the system changes discontinuously at unknown points in time, indexed by a hidden discrete random variable. In this paper, we are interested in Switching Markov State-Space Models (SMSSM) which are widely used in several fields of applied science such as signal processing [1], econometrics [2] and bioinformatics [3]. A SMSSM can be viewed as a Linear Gaussian State-Space Model (LGSSM) with parameters indexed by a latent Markov chain, or as a Hidden Markov Model (HMM) with two latent states: a continuous system state, and a discrete Markov state.

For latent structure models, the parameters inference can fail due to identifiability issue. Leroux [4] establishes a sufficient condition for the HMM identifiability. For the LGSSM, the identifiability issue can be addressed by imposing some structural constraints on its parameters [5], or by taking it into account in the parameters inference algorithm [6].

We consider here a SMSSM where the unknown continuous state and the observations are scalars. We show that, if the hidden Markov chain is irreducible and aperiodic, a prior information relating the observations to the unknown continuous state at a time t_0, for instance $t_0 = 0$, is sufficient to ensure the identifiability of the model. Moreover, we check the relevance of these constraints by estimating the parameters of a SMSSM using a penalized EM algorithm. In this purpose, a SMSSM modeling the state of charge of an electric battery is considered. The results, using real electric vehicle data, show that the parameters are efficiently estimated.

The paper is organized as follows. Section 2 presents the model and the identifiability issue. In Section 3, the identifiability of a LGSSM is addressed. In Section 4, the previous identifiability result is extended for SMSSM. Section 5 shows the relevance of these constraints using real electric vehicle data. Section 6 concludes the paper.

2 Problem formulation

2.1 Model

Let s_t denote a discrete irreducible and aperiodic Markov chain on $\{1, \ldots, \kappa\}$, with initial distribution Π and transition matrix P. We consider a stochastic process (y_t, x_t, s_t) where y_t is observable, and s_t and x_t are unobservable. This process is defined by a switching Markov state-space model

$$
\begin{aligned}
x_t &= A(s_t)x_{t-1} + B(s_t)u_t + \omega_t, \quad (1)\\
y_t &= C(s_t)x_t + D(s_t)u_t + \varepsilon_t, \quad (2)
\end{aligned}
$$

where u_t denotes a known exogenous input, and ω_t and ε_t are independent Gaussian white noises with variances $\sigma_X^2(s_t)$ and $\sigma_Y^2(s_t)$ respectively. The observations y_t are assumed to be independent given (s_t, x_t). It is also assumed that x_0 is fixed and that the initial distribution Π is known. For simplicity, we consider that x_t and $y_t \in \mathbb{R}$. Let Θ denote the vector of parameters

$$
\Theta = \{P, \Gamma = (A(s), B(s), C(s), D(s), \sigma_X(s), \sigma_Y(s))\} \quad (3)
$$

for $s = 1, \ldots, \kappa$. The distributions $p(x_t \mid x_{t-1}, s_t, \Theta)$ and $p(y_t \mid x_t, s_t, \Theta)$ are assumed to be Gaussian, with parameters deduced from (1) and (2). It is noteworthy that given a specific sequence of Markov states $s_{0:T} = \{s_0, \ldots, s_T\}$, the likelihood $p(y_{0:T} \mid s_{0:T}, \Gamma)$ is

$$
p(y_{0:T}|s_{0:T}, \Gamma) = p(y_0|s_0, \Gamma) \prod_{t=1}^{T} p(y_t \mid y_{0:t-1}, s_{0:t}, \Gamma), \quad (4)
$$

where $y_t \mid y_{0:t-1}, s_{0:t}, \Gamma \sim \mathcal{N}(y_t; y_{t|t-1}, \Omega_{t|t-1})$, with $y_{t|t-1}$ and $\Omega_{t|t-1}$ deduced from (2)

$$
y_{t|t-1} = C(s_t)x_{t|t-1} + D(s_t)u_t, \quad (5)
$$

$$
\Omega_{t|t-1} = C(s_t)\Sigma_{t|t-1}C(s_t)' + \sigma_Y^2(s_t). \quad (6)
$$

The variables $x_{t|t-1}$ and $\Sigma_{t|t-1}$ are deduced from (1)

$$
x_{t|t-1} = A(s_t)x_{t-1|t-1} + B(s_t)u_t, \quad (7)
$$

$$
\Sigma_{t|t-1} = A(s_t)\Sigma_{t-1|t-1}A'(s_t) + \sigma_X^2(s_t), \quad (8)
$$

with $x_{t-1|t-1}$ and $\Sigma_{t-1|t-1}$ given by the correction step of a Kalman filter

$$
\begin{aligned}
x_{t-1|t-1} =& x_{t-1|t-2} + K_{t-1}(y_{t-1} - y_{t-1|t-2}), \quad (9)\\
\Sigma_{t-1|t-1} =& (I - K_{t-1}C(s_{t-1}))\Sigma_{t-1|t-2}(I - K_{t-1}C(s_{t-1}))'\\
&+ K_{t-1}\sigma_Y^2(s_{t-1})K'_{t-1}, \quad (10)
\end{aligned}
$$

where I is the identity matrix with an appropriate dimension and K_{t-1} is the Kalman gain [7] given by

$$
K_{t-1} = \Sigma_{t-1|t-2}C(s_{t-1})\Omega_{t-1|t-2}. \quad (11)
$$

2.2 Identifiability issue

The SMSSM, among many other mathematical models, is used to describe the dynamics of a given system using experimental data and understand observed phenomena. Indeed, it is desirable that every model parameter has a physical interpretation in order to easily integrate any prior knowledge and interpret the results of this modeling. Among the major issues arising from this modeling, this paper focuses on the identifiability one, which is essential for parameters estimation. According to the classical definition, a subset of parameters $F \subset \Theta$ is said "globally structurally (g.s.) identifiable" when $p(y_{0:T}|\Theta^*) = p(y_{0:T}|\Theta)$ implies $F^* = F$. However, it is well-known that a SMSSM is unidentifiable and constraints must be imposed to guarantee its identifiability, i.e. its interpretability. As a first step, we consider the identifiability of a LGSSM.

3 Identifiability of a LGSSM

Let us consider the LGSSM described by (1)-(2) with $\kappa = 1$. Each minimal representation of this model can be deduced from Γ by a state-space linear transformation H, which leads to the following system of equations

$$
\left\{
\begin{array}{lllll}
A^* & = & H & \cdot & A \cdot H^{-1}\\
B^* & = & H & \cdot & B\\
C^* & = & C & \cdot & H^{-1}\\
{\sigma_X^2}^* & = & H & \cdot & \sigma_X^2 \cdot H,
\end{array}
\right. \quad (12)
$$

where $H \in \mathbb{R}$. It has to be noted that D and σ_Y are always g.s. identifiable, i.e. they are invariant under any linear transformation H. And, A is also g.s. identifiable since $H \in \mathbb{R}$. The following prior information is considered at $t_0 = 0$

$$
y_0 = Cx_0 + Du_0, \quad (13)
$$

with $x_0 \cdot u_0 \neq 0$. Under (13), it is easily proved that the only solution of (12) is $H = 1$. Thus, the parameters of a LGSSM are g.s. identifiable under the constraint (13). In the following, this identifiability result is extended for SMSSM.

4 Identifiability of a SMSSM

First of all, it is noteworthy that Markov states s_t can be relabeled without changing the distribution of the observations. Thus, the identifiability of the SMSSM, up to state switching, is discussed below.

4.1 Specific sequence of Markov states

As a first step, we consider that the sequence of Markov states $s_{0:T}$ is known. It is assumed that at $t_0 = 0$

$$
y_0 = C(s_0)x_0 + D(s_0)u_0, \quad (14)
$$

for any hidden state $s_0 = 1, \ldots, \kappa$. Similarly to a LGSSM, each minimal representation of the model can be deduced from Γ by a state-space linear transformation $\mathcal{H}$ where

$$
\mathcal{H} = [H_1 \quad H_2 \quad \ldots \quad H_\kappa], \quad (15)
$$

with $\mathcal{H} \in \mathbb{R}^\kappa$. Indeed, given a specific Markov sequence $s_{0:T}$, the model can be transformed into κ LGSSMs with appropriate sampling times. Hence, $D(s)$, $\sigma_Y(s)$ and $A(s)$ are g.s. identifiable for $s = 1, \ldots, \kappa$, since they are invariant under any linear transformation H_s. In addition, the remaining parameters of these κ LGSSMs are g.s. identifiable under the constraint (14), as shown in Section 3. Accordingly given a specific Markov sequence $s_{0:T}$, the parameter Γ of a SMSSM with $\kappa \geq 1$ is g.s. identifiable under constraints (14), and

$$p(y_{0:T}|s_{0:T},\Gamma) = p(y_{0:T}|s_{0:T},\Gamma^*) \quad \Rightarrow \quad \Gamma = \Gamma^*. \tag{16}$$

The next section discusses the identifiability of a SMSSM when the sequence of Markov states is unknown.

4.2 Unknown sequence of Markov states

The marginal likelihood of the SMSSM (1)-(2) is given by

$$p(y_{0:T} \mid \Theta) = \sum_{s_{0:T} \in \mathcal{S}} p(s_{0:T}|P) \cdot p(y_{0:T}|s_{0:T},\Gamma), \tag{17}$$

where $\mathcal{S} = \{1,\dots,\kappa\}^{T+1}$ and $p(y_{0:T}|s_{0:T},\Gamma)$ is a Gaussian distribution whose parameters are recursively calculated (Section 2). Thus $p(y_{0:T} \mid \Theta)$ is a finite convex combinaison of Gaussian distributions. Following the line of proof in [8], we have

$$\begin{aligned} p(y_{0:T}|\Theta) = p(y_{0:T}|\Theta^*) \Rightarrow & \\ & 1) p(y_{0:T}|s_{0:T},\Gamma) = p(y_{0:T}|s_{0:T},\Gamma^*) \\ & 2) p(s_{0:T}|P) = p(s_{0:T}|P^*). \end{aligned} \tag{18}$$

Under the constraints (14), the first equation implies that $\Gamma = \Gamma^*$. Since s_t is an irreducible aperiodic Markov chain, the second equation implies that $P = P^*$ cf. Lemma 2 in [4]. As a result, we have the following proposition.

Proposition 1
The parameters of the SMSSM (1)-(2) *are g.s. identifiable if the following constraints are verified*

1. *A prior information at $t_0 = 0$ is available:*

   $$y_0 = C(s)x_0 + D(s)u_0 \text{ with } s = 1,\dots,\kappa \text{ and } x_0 \cdot u_0 \neq 0,$$

2. $\forall (i,j),\ P(i,j) \neq 0.$

Condition 2 implies that the hidden Markov chain is irreducible and aperiodic. In the next section the parameters of a SMSSM are estimated, under the proposed constraints, by the Maximum Likelihood (ML) methodology through the EM algorithm.

5 ML parameters inference

When faced with an identifiability problem, ML methods could fail to efficiently estimate the model parameters. In the following, we show that, since the identifiability issue is solved under the constraints (14), the ML parameter estimation works properly. The EM formulas are not detailed here. They are presented for instance in [9, 10].

The EM algorithm consists of iteratively estimating the parameter Θ by maximizing the expected log-likelihood of the complete data. Here this maximization must be conducted under the constraints (14). This is summarized in Algorithm 1. The Lagrangian associated to the

Algorithm 1 EM algorithm under constraints

Input $\leftarrow y_{0:T}, u_{0:T}, x_0, \Pi$
Init $\Theta^{(0)}, k = 0$
for $k < k_{max}$ **do**
  1- **E-Step** Compute
  $\mathcal{Q}(\Theta,\Theta^{(k)}) = \mathbb{E}_{y_{0:T},\Theta^{(k)}}\left[\log p_\Theta\left(x_{0:T}, s_{0:T}, y_{0:T}\right)\right]$
  2- **M-Step**
  $\Theta^{(k+1)} = \underset{\Theta}{\operatorname{argmax}}\, \mathcal{Q}(\Theta,\Theta^{(k)})$ while $\forall s = 1,\dots,\kappa$
  $y_0 - C(s)x_0 - D(s)u_0 = 0$ and $\sum_j P(s,j) = 1$
end for
Output $\leftarrow \Theta^{(k)}$

constraints is

$$\begin{aligned} \mathcal{L}(\Theta,\,\lambda,\,\mu) =& \mathcal{Q}(\Theta,\Theta') + \sum_{i=1}^{\kappa} \lambda_i [1 - \sum_j P(i,j)] \\ & + \sum_{i=1}^{\kappa} \mu_i [y_0 - C(i)x_0 - D(i)u_0], \end{aligned} \tag{19}$$

where λ_i and μ_i are the Lagrangian multipliers. Canceling the derivative equations of $\mathcal{L}(\Theta,\lambda,\mu)$ w.r.t. Θ requires performing summations over up to κ^{T+1} values of $s_{0:T}$. Actually, the optimal estimation of the unknown states of a SMSSM is a well-known NP problem [10]. To overcome this, a Monte Carlo (MC) approximation of the EM algorithm is generally developed [10]. It consists of using a set of N "particles" $\{s^i_{0:T}\}_{i=1}^N$ and importance weights $\{w^i_T\}_{i=1}^N$, such that $\partial\mathcal{Q}(\Theta,\Theta^{(k)})/\partial\Theta$ is estimated by

$$\sum_{i=1}^{N} w^i_T \frac{\partial}{\partial\Theta} \mathbb{E}_{s^i_{0:T},y_{1:T},\Theta^{(k)}}[\log p_\Theta(x_{0:T}, s^i_{0:T}, y_{1:T})]. \tag{20}$$

We illustrate the validity of these constraints using real electric vehicle data. A SMSSM of the State of Charge (SoC) of an electric battery, using voltage and current measurements, is considered. Indeed, the battery dynamics randomly changes according to uncontrolled usage conditions such as ambient temperature and driving behavior. The problem consists of estimating the parameters of this model using the proposed MC-EM algorithm. The observation and the transition equations are based on physical models [11], and the corresponding SMSSM is given by

$$\begin{aligned} x_t &= x_{t-1} + B(s_t)u_t + \omega_t, \\ y_t &= C(s_t)x_t + D_1(s_t)u_t + D_2(s_t) + \varepsilon_t, \end{aligned} \tag{21}$$

where y_t is the observed voltage, u_t the input current and x_t the SoC to be estimated. Here, $A(s)$ is physically identified: $\forall s$, $A(s) = 1$. In addition, we have a physical prior information at $t_0 = 0$

$$y_0 = C(s_0)x_0 + D_2(s_0), \tag{22}$$

where y_0 is the Open Circuit Voltage (OCV) measurement and x_0 its corresponding SoC. Indeed, in practice at

	B	C	D_1	D_2	σ_X	σ_Y
Ratio for $s = 1$	7.78	0.14	1.05	1	9.5	0.93
Ratio for $s = 2$	5.05	0.12	0.96	1	9.5	0.93
Ratio for $s = 3$	9.12	0.14	1.06	1	9.5	0.93

TAB. 1: Ratio of the parameters estimated without constraints to those with constraints, SMSSM (21) with $\kappa = 3$

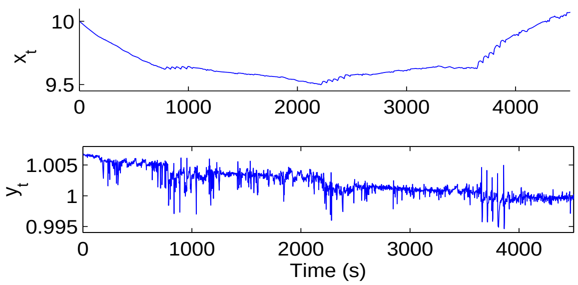

FIG. 1: Ratio of estimated x_t (top) and y_t (bottom) without constraints to the one with constraints

$t_0 = 0$, the battery is often in a resting state, and the SoC can be efficiently calculated using an OCV/SoC relationship. To verify the relevance of the proposed constraints, the unknown parameters are estimated based on the presented MC-EM algorithm, using 200 particles, with as well as without constraints (14). When the estimation is performed without constraints, we impose an upper bound to x_t to avoid numerical problems. Experimental results show that both marginal likelihoods $p(y_{0:T} \mid \Theta)$ are equal. Table 1 shows the ratio of the parameters estimated without constraints to those estimated with constraints. It can be seen that, as shown in (12), D and σ_Y are invariant in both cases, whereas C is divided and σ_X is multiplied by $H = 10$. For B, the ratio must be theoretically equal to $H = 10$. However, the experimental ratio is not fixed which is not surprising since B describes the evolution of the unknown x_t. Figure 1 represents the ratio of estimated x_t and y_t without constraints to the one with constraints. The results show that the estimated y_t is invariant whereas x_t is multiplied by $H = 10$. Thus, these results show that under the proposed constraints, the penalized EM algorithm efficiently estimates the parameters and that the identifiability issue is resolved. It is noteworthy that, under constraints, the order of magnitude of the estimated parameters and x_t is physically accurate; for instance the estimated state of charge $x_t \in [0, 100]$.

6 Conclusion

This paper has addressed the identifiability of switching Markov state-space models where the unknown continuous state and the observations are scalars. We prove that, in case of an irreducible and aperiodic Markov chain, if a constraint relating the observations to the continuous state at a time t_0 is available, the model parameters become generally structurally identifiable. In order to verify the relevance of this proposition, a comparison between ML parameters inference of a SMSSM, under and without the proposed constraints, is performed. The results show that, by considering the above constraints, the identifiability issue is resolved and the EM algorithm works efficiently.

References

[1] S.M. Oh, J.M. Rehg, T. Balch, and F. Dellaert, "Learning and inferring motion patterns using parametric segmental switching linear dynamic systems," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 103-124, 2008.

[2] M. Lammerding, P. Stephan, M. Trede, and B. Wilfling, "Speculative bubbles in recent oil price dynamics: evidence from a Bayesian Markov-switching state-space approach," *Energy Economics*, vol. 36, pp. 491-502, 2013.

[3] R. Yoshida, S. Imoto, and T. Higuchi, "Estimating time-dependent gene networks from time series microarray data by dynamic linear models with Markov switching," *Proceedings of the 2005 IEEE Computational Systems Bioinformatics Conference*, pp. 289-298, 2005.

[4] B.G. Leroux, "Maximum-likelihood estimation for hidden Markov models," *Stochastic Processes and their Applications*, vol. 40, no. 1, pp. 127-143, 1992.

[5] E. Walter, and Y. Lecourtier, "Unidentifiable compartmental models: what to do?," *Mathematical Biosciences*, vol. 56, no. 1-2, pp. 1-25, 1981.

[6] P.V. Overschee, and B.D. Moor, "Subspace algorithms for the stochastic identification problem," *Proceedings of the 30th IEEE Conference on Decision and Control*, vol. 2, pp. 1321-1326, 1991.

[7] R. Kalman, "Mathematical description of linear dynamical systems," *Journal of the Society for Industrial and Applied Mathematics Series A Control*, vol. 1, no. 2, pp. 152-192, 1963.

[8] S. J. Yakowitz and J. D. Spragins, "On the identifiability of finite mixtures," *The Annals of Mathematical Statistics*, vol. 39, pp. 209-214, Feb 1968.

[9] M. A. Tanner, *Tools for statistical inference - Observed data and data augmentation methods*, vol. 67, Springer-Verlag New York, 1991.

[10] A. Doucet and A. M. Johansen, "A tutorial on particle filtering and smoothing: Fifteen years later," *Handbook of Nonlinear Filtering*, vol. 12, pp. 656-704, 2009.

[11] G. Plett, "Extended Kalman filtering for battery management systems of LiPB-based HEV battery packs," *Journal of Power Sources*, vol. 134, pp. 262-276, 2004.