Identification of switched ARX models via convex optimization and expectation maximization

András Hartmann [a,c,*], João M. Lemos [a], Rafael S. Costa [a,c], João Xavier [b], Susana Vinga [c]

[a] INESC-ID, Instituto Superior Técnico, Universidade de Lisboa, R Alves Redol 9, 1000-029 Lisboa, Portugal
[b] ISR, Instituto Superior Técnico, Universidade de Lisboa, Av Rovisco Pais 1, 1049-001 Lisboa, Portugal
[c] IDMEC, Instituto Superior Técnico, Universidade de Lisboa, Av. Rovisco Pais 1, 1049-001 Lisboa, Portugal

ABSTRACT

This article addresses the problem of parameter identification for Switched affine AutoRegressive models with eXogenous inputs (SARX). The system includes continuous domain states that depend on discrete time-varying parameters. The identification of such systems typically results in non-convex problems that could be tackled as a mixed integer program. However, in this case, the computational complexity would be intractable in many practical applications. Another approach involves heuristics in order to deliver approximate solutions. This article proposes a three-step method based on solving a regularized convex optimization problem, followed by a clustering step, yielding a partial solution to the problem. When substituted back into the original problem, the partial solution renders it convex. Finally, this convex problem is solved in the third step, yielding an approximate solution. It is found that each step significantly improves the parameter estimation results on the systems considered. A beneficial property of the method is that it relies upon only one scalar tuning parameter, to which the final results are not highly sensitive. The performance of the algorithm is compared with other methods on a simulated system, and illustrated in an experimental biological dataset of diauxic bacterial growth.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

One limitation of many mathematical models describing dynamic systems is that the parameters are assumed to be invariant during the observation period. This premise does however not necessary hold in many cases, such as certain applications in economics [7], computer vision [29], systems biology [9] or longitudinal clinical datasets [15]. In the last decade, much attention has been given to dynamical modeling using Linear Parameter Varying (LPV) and closely related linear hybrid systems. The similarity between LPV and hybrid systems is that the system parameters may be subject to abrupt changes that are represented by different linear submodels. The main difference is that LPV systems do not limit the number of submodels like hybrid systems. In this sense hybrid systems are considered to be a special case of LPVs [26]. In the rest of this article we focus on the identification of a type of hybrid systems, for more information about the definition and identification of LPV systems, the interested reader is referred to the comprehensive book of [27] and references therein.

Hybrid models comprise both continuous and discrete states (parameters) [16]. Typically the dynamics can be modeled with continuous state evolution and the transitions between submodels being represented by the changes of discrete states. Various classes of hybrid systems were proposed [22], with different definitions of discrete and continuous states and their corresponding interactions. The equivalences between some of the classes (under mild conditions) were established by Heemels et al. [11,12], Weiland et al. [32].

In the literature, a great emphasis has been placed on hybrid system identification; for a comprehensive overview see Paoletti et al. [22], Lygeros [16] and references therein. Many contributions focus on PieceWise affine AutoRegressive models with eXogenous

input (PWARX) and other more restrictive model classes with identification techniques based on mixed integer programming [24], clustering [5], Bayesian [13] and bounded error [3] approaches. Presumably the first closed-form linear-algebraic solution to the identification of Switched affine AutoRegressive model with eXogenous inputs (SARX) systems was introduced by Ma and Vidal [17]. The proposed solution is based on decoupling the system into polynomial form, although this approach is known to be highly sensitive to noise [30,31]. Another identification scheme to linear/affine noise-free switching systems was recently reported by Bako et al. [2] which, being a recursive method, has the advantage of performing online identification. The method of Hartmann et al. [10] considers a noisy nonlinear switching system and proposes an iterative identification method that combines Particle Filter with Expectation Maximization (EM) algorithm.

Recently, the concept of sparsity emerged as a possible basis for approximate algorithms. One pioneer on this road was the work on segmentation by Ohlsson et al. [20] that identifies parameters of affine AutoRegressive model with eXogenous inputs (ARX) systems in order to construct an (LPV) model. The algorithm solves a Sum Of Norm (SON) regularized convex optimization problem and tunes the weights of the regularization in an iterative process. A similar idea was applied to identify the parameters of SARX [1] and PWARX systems in [19], and then estimate the partitions of parameters using clustering.

Inspired by this research line, a three-step algorithm to the identification of hybrid SARX models, called SON-EM is proposed here. In the first step, a SON regularized convex optimization is solved. The second step consists of applying Expectation Maximization (EM) to cluster the preliminary estimates. The first two steps can be interpreted as a change detection procedure and make the parameter identification problem fall back to a Quadratic Program (QP), which is solved in the third step. It is found that each step significantly improves the parameter estimation results. A beneficial property of the method is that it depends only on one scalar tuning parameter, to which the final results are not highly sensitive. An algorithm based on SON-EM is also introduced to determine the optimal cardinality of the submodel set. The performance of the methods is illustrated in a simulated system and in experimental diauxic bacterial growth data of [4].

The rest of this article is organized as follows: in Section 2 we introduce the problem and the method proposed to perform parameter identification. Results on a generated dataset and a real-world example are detailed in Section 3. The method and the results are further discussed in Section 4. Finally, conclusions are drawn in Section 5.

2. Methods

2.1. Problem formulation

SARX models are switched extensions to ARX models, defined as the concatenation of several different submodels. Each submodel corresponds to an ARX of fixed dimension. Consider the following system in input-output form

$$y(t) = \theta_{\eta(t)}^{\mathrm{T}} \Phi(t) + \varepsilon(t), \tag{1}$$

where the regression vector $\Phi(t)$ of dimension $n = n_a + n_b$ consists of past measurements and inputs

$$\Phi(t) = [y(t-1) \ \ldots \ y(t-n_a) \ u(t) \ \ldots \ u(t-n_b)]^{\mathrm{T}}. \tag{2}$$

At each time-instance, the time-dependent parameter vector $\theta(t) = \theta_{\eta(t)} \in \mathbb{R}^n$ belongs to a set $\Theta = \{\theta_1, \ldots, \theta_K\}$ of cardinality K, representing the submodel set. The discrete finite range function, $\eta(t) : \mathbb{R} \to \{1, \ldots, K\}$ indicates which submodel generates the output

at the time-instance t, and is referred as the switching sequence or discrete state. The input $u(t)$ is observed and so is the output $y(t)$ which is corrupted by an additional noise term, $\varepsilon(t)$.

SARX models represent a rather general class of switched systems, where the switching sequence is a function of the time only. It has been proven that every observable switched affine system has a switched autoregressive representation [32]. Another popular, but more restrictive model class is PWARX, where the switching sequence is defined by a polyhedral partition of the regression domain, see e.g. [22].

The parameter identification problem for SARX models can be stated as

Problem 1. Given a sequence of input-output pairs $\left\{u(t), y(t)\right\}_{t=1}^{T}$, the model orders n_a, n_b and the cardinality of the submodel set K, estimate the parameter vectors θ_k; $k = 1, \ldots, K$ and the switching sequence $\eta(t)$; $t = 1, \ldots, T$. An even more general problem is when the cardinality of submodel set is also unknown.

Problem 2. Given a sequence of input-output pairs $\left\{u(t), y(t)\right\}_{t=1}^{T}$ and the model orders n_a and n_b, estimate K, the cardinality of the submodel set, the parameter vectors θ_k; $k = 1, \ldots, K$ and the switching sequence $\eta(t)$; $t = 1, \ldots, T$.

From a practical point of view, a prerequisite to the identification is that all the submodels have to be excited within the input/output sample dataset. The states being sufficiently excited is a standard requirement for sufficient parameter estimation, even in the constant parameter case. If however there exists a submodel $s \in [1 \ldots K]$ such that it is not represented in the sample, meaning that $\nexists t \in \mathbb{R}, \eta(t) = s$, it follows that the submodel s remains hidden in the finite sample, and can not be identified. On the other hand, overestimating the number of submodels leads to overfitting, where two or more identified parameters may represent the same generating parameter. Another identifiability issue was pointed out by Bako [1] and Vidal et al. [28] when the model dimensions n_a and n_b in Eq. (2) are not *a priori* fixed. Therefore in Problem 1 and Problem 2, n_a and n_b are assumed to be known and equal for all submodels. Further theoretical aspects of identifiability of SARX systems can be found in Petreczky et al. [23].

For now, assume that K is known and let us focus on Problem 1, that can be recast in the following optimization problem

$$
\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \sum_{k=1}^{K} \ell \left(y(t) - \hat{\theta}_k^{\mathrm{T}} \Phi(t) \right) \mathcal{X}_{t,k} \\
s.t. \quad & \sum_{k=1}^{K} \mathcal{X}_{t,k} = 1 \quad \forall t \in [1 \ldots T] \\
& \mathcal{X}_{t,k} \in \left\{ 0, 1 \right\},
\end{aligned}
\tag{3}
$$

where $\ell(\cdot)$ is an arbitrary p-norm. The minimization is with respect to the parameter estimates $\hat{\theta}_k$, $k = 1 \ldots K$ and the matrix $\mathcal{X}$, consisting of $T \times K$ binary variables. The significance of the binary variable $\mathcal{X}_{t,k}$ is to indicate whether or not a data point belongs to submodel k at time instance t. The discrete state can be recovered as

$$\eta(t) = k \Leftrightarrow \mathcal{X}_{t,k} = 1. \tag{4}$$

Let us assume that $K < T$, as in most practical relevant situations. Then, Eq. (3) is known to be a mixed integer problem, being thus NP-hard[1] [22]. Solving this kind of problem needs exhaustive search

---

[1] In the case $K \geq T$ the solution is trivial, because to each time instance a different submodel can be assigned.

algorithms with worst-case complexity that has data length in the exponent ($\mathcal{O}(K^T)$), and therefore the mixed integer programming approach is only recommended for short data segments [24,22].

An equivalent problem formulation to Eq. (3) is described in [1]

$$\min_{\hat{\theta}_1, \ldots, \hat{\theta}_K} \sum_{t=1}^{T} \min_{i=1, \ldots, K} \ell\left(y(t) - \hat{\theta}_i^{\mathrm{T}} \Phi(t)\right). \tag{5}$$

This bilevel optimization problem clearly points out the non-convex nature of the problem, because the minimum of minima is, in general, not convex.

2.2. Identification method

Hereafter the main contributions of this paper, a three-steps procedure for an approximate solution to Problem 1 is described.

As a preliminary step, a Sum Of Norm (SON) regularized least squares optimization problem is solved

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \|y(t) - \tilde{\theta}^{\mathrm{T}}(t)\Phi(t)\|^2 \\
& + \lambda \sum_{t=2}^{T} \|\tilde{\theta}(t) - \tilde{\theta}(t-1)\|_p.
\end{aligned} \tag{6}$$

Here the optimization is with respect to $\tilde{\theta}(t)$, yielding the parameter estimates after the first step. The optimization problem in Eq. (6) implies a positive scalar regularization parameter, λ, which is the only one tuning parameter of the method. The expression $\|\cdot\|_p$ denotes for the p-norm, where the design variable p plays a crucial role by setting the norm of the regularization. Imposing a smoothing condition on the switches of the parameters has computationally convenient properties since, for $p \geq 1$, one obtains a continuous convex problem. If sparsity of the solution is favored, in the sense that the change of one dimension in the parameter vector is preferred against change in all the dimensions, p should be set to one, leading to a Least Absolute Shrinkage and Selection Operator (LASSO) flavor ℓ_1-norm regularization, similar to group LASSO [33].

Although the solution of Eq. (6) is an estimate of the parameters, since here the cardinality of the submodel set is not limited to K, it is not guaranteed to be a feasible point of the original problem. The rationale to apply EM clustering for mixture of Gaussians as described in details in Appendix A, in the second step is to constrain the cardinality of the submodel set and to estimate the switching sequence. Note that theoretically any other unsupervised learning algorithm could be applied here, that casts the preliminary estimates into K distinct sets, such as clustering or labeling methods.

The solution after the second step is feasible, but may not be optimal in the mean squared error sense because the clustering only operates on $\tilde{\theta}(t)$, independently from the observations. This problem is tackled in the third step by replacing the estimated switching sequence into Eq. (3) such as to obtain the following convex quadratic problem

$$\begin{aligned}
\text{minimize} \quad & \sum_{t=1}^{T} \|y(t) - \hat{y}(t)\|^2 \\
\text{s.t.} \quad & \hat{y}(t) = \hat{\theta}_{\hat{\eta}(t)}^{\mathrm{T}} \Phi(t).
\end{aligned} \tag{7}$$

Here the optimization is only with respect to $\hat{\theta}_k$, $k = 1 \ldots K$, because $\hat{\eta}(t)$ is given from step 2. Solving this optimization problem in the third step yields parameter estimates that are optimal according to the estimated switching sequence. This means that even if the inner optimization problem in Eq. (5) is an

approximation, the outer one is optimal for that specific approximation. The SON-EM algorithm is summarized in Algorithm 1.

Algorithm 1. SON-EM algorithm for parameter estimation of hybrid time-varying parameter systems

---
1. Solve the convex optimization problem of Eq. (6) to deliver preliminary estimates $\tilde{\theta}$
2. Proceed with EM clustering on $\tilde{\theta}$ to obtain K classes as described in Appendix A
3. Solve the convex optimization problem of Eq. (7)

2.3. Determining the cardinality of submodels

To solve Problem 2 and determine the value of K, an iterative procedure is proposed for model selection hereafter in Algorithm 2. The estimation is performed with increasing $K \in \mathbb{N}^+$, and computing the Mean Squared Error (MSE) of each reconstruction. As a rule of thumb, the MSE will decrease with increasing K. The algorithm terminates, when the change in MSE is smaller than a threshold ϵ, because no significant changes in MSE yields that introducing new submodels would not further reduce the error of the reconstruction.

Algorithm 2. Determining the cardinality of the submodel set

---
1. $K = 1$; $MSE_0 = \infty$
2. Proceed with Algorithm 1 using K submodels and calculate the error MSE_K
3. if ($|MSE_K - MSE_{K-1}| < \epsilon$), then return K
4. else $K = K + 1$; repeat from (2)

2.4. Software

There are many readily available solvers for convex problems, with Sum Of Norm (SON) regularization. Here the CVX MATLAB package was applied for specifying and solving convex programs [8]. The convergence of EM might be rather slow depending on the initial estimates values, which may hamper the identification of a good solution. Since in practice, good initial estimates can significantly accelerate the iterative procedure and reduce the number of runs needed to find a good local optimum, it is therefore advisable to choose a good seed, for example by pre-clustering the dataset with a fast algorithm. In our case, the MATLAB *kmeans* algorithm from the Statistics Toolbox was used for pre-clustering. An example implementation of the proposed algorithms is freely available under the terms of GNU Public License (GPLv3) from the authors webpage: http://andrashartmann.info/.

3. Results

3.1. Simulated SARX system

The following results were obtained using the linear model originally introduced in Lemos et al. [14]

$$\begin{aligned}
y(t+1) - \theta_{\eta(t)} y(t) + 0.7 y(t-1) = \\
u(t) - 0.5 u(t-1) + \varepsilon(t+1)
\end{aligned} \tag{8}$$

The system consists of one parameter and $K = 4$ submodels. Note that although $\eta(t)$ is driven by a Markov chain, as the identification algorithm does not rely on the Markov property, the system can be considered to be SARX. The states $[\theta_1 \ldots \theta_4]$ and the transition probabilities A_m are as follows:

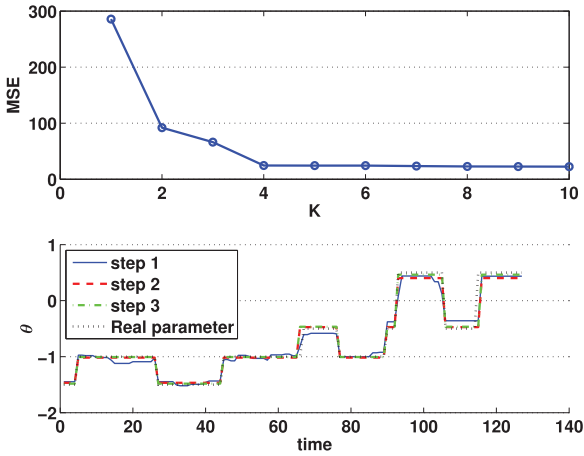$$\theta_1 = -1.5, \quad \theta_2 = -1, \quad \theta_3 = -0.5, \quad \theta_4 = 0.5 \tag{9}$$

Fig. 1. Parameter identification on generated data. MSE plotted against the values of K, in order to determine the cardinality of the submodel set (top) and parameter estimates after each step respectively (bottom) with $\lambda = 1$.

$$A_m = \begin{bmatrix} 0.98 & 0.02 & 0.00 & 0.00 \\ 0.01 & 0.98 & 0.01 & 0.00 \\ 0.00 & 0.01 & 0.94 & 0.05 \\ 0.00 & 0.00 & 0.02 & 0.98 \end{bmatrix}. \tag{10}$$

The input signal $u(t)$ is a ± 1 PseudoRandom Binary Sequence (PRBS). The noise is independent and identically distributed Gaussian with $\varepsilon(t) \sim \mathcal{N}(0, \sigma^2)$, $\sigma = 0.5$.

3.1.1. Some properties of SON-EM

The resulting estimates after each step of Algorithm 1 are shown in Fig. 1. The cardinality of the submodel set was estimated using Algorithm 2, and was clearly found to be $K = 4$, which coincides with the true value. It can be seen that the SON regularization estimates (first step) follow the changes of the parameter well, although there are some fluctuations around the real values. In fact, at some time instances the deviance to the real parameter values are significant. Applying the clustering (second step) clearly improves the estimates by constraining the cardinality of the submodel set, and the final estimates (third step) are even closer to the real parameters.

The sensitivity of the results to the regularization (tuning) parameter on the range $\lambda = [0.5, \ldots, 5]$ is illustrated in Fig. 2. From
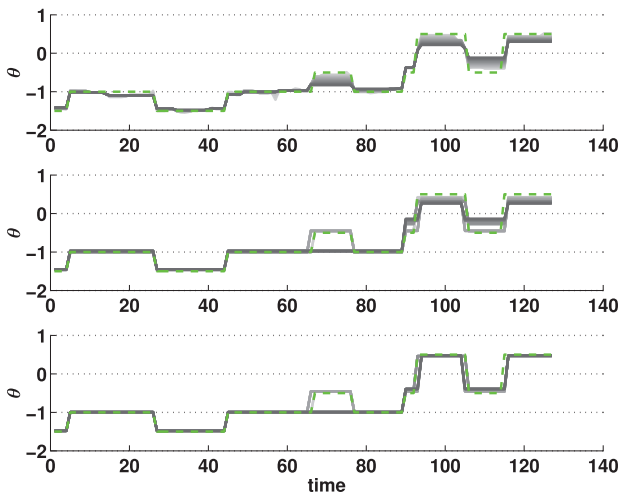

Fig. 2. Sensitivity of the estimates to the tuning parameter. Identification with $\lambda = [0.5 \ldots 5]$. Darker lines represent higher values of λ, the green dashed line is the real parameter. Estimates are shown after each step of SON-EM: SON regularization (top), EM (middle), final results (bottom).
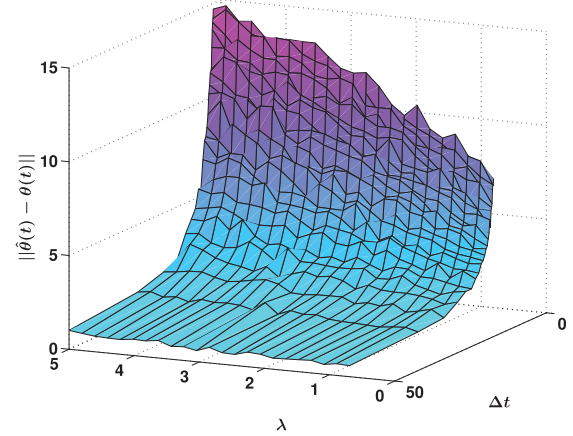

Fig. 3. Sensitivity to the dwell time and the compensation effect of the tuning parameter. The results are averaged over hundred runs on each $\{\Delta t, \lambda\}$ pairs. The relative standard deviation (σ/μ) was low (< 2.4).

this it can be seen that even if in the first two steps the preliminary estimates are sensitive to the choice of λ, the final result is independent of the tuning parameter on a large domain. In fact, in the example considered, only one significant change was observed in the estimated switching sequence between the time instances 65 and 76, with $\lambda \approx 2.5$. This is hypothesized to be the consequence of the regularization term in Eq. (6), that dominates the regression, and causes an error in the estimated switching sequence. It is important to note however that $\hat{\Theta}$, the parameter set, was estimated correctly independent from the choice of the value of λ.

The choice $\lambda = 0$ corresponds to an objective function containing MSE only, without regularization in the first step. In this case a significant degradation of the final results is observed. Reasonable λ values are typically of magnitude between 0.1 and 10. Note however that factors like the signal to noise ratio, and the values of the parameters may influence the ideal choice of λ. In the rest of this study λ was chosen to be equal to one if not stated otherwise.

The result is influenced by the time between the changes of the parameter (dwell time), see Fig. 3. Here, the transition matrix of the Markov chain, A_m is constructed to set the expected value of the dwell time to a predefined value Δt. As can be seen, shorter transition times are accompanied by exponential-like growing errors, but reducing the value of λ can compensate this effect to some extent.

3.1.2. Comparison with other methods

It is important to compare the parameter estimates of the SON-EM algorithm to results obtained with other methods. Comparisons were made with some of the latest developments in parameter identification of switched systems, such as the segmentation with regularization (segreg) method of Ohlsson et al. [20], the EM-PF method of Hartmann et al. [10], and the PWASON algorithm as described in [21]. Finally, because in some cases, simple approaches may outperform more complicated ones, the RANdom SAmple Consensus (RANSAC) algorithm of Fischler and Bolles [6] was also included. The method was adopted for hybrid systems as described in details in Appendix B.

Fig. 4 compares the estimates on 100 different generated time-series. These results show that SON-EM clearly outperforms EM-PF and RANSAC on the simulated linear system in terms of covariances of the estimates. It is however noteworthy that the median estimate of each method has similar good accuracy. Note that segreg and PWASON methods were disregarded from this comparison, because they do not explicitly restrict the cardinality of the parameter set, and hence the estimates may not be feasible points of the original problem Eq. (5).
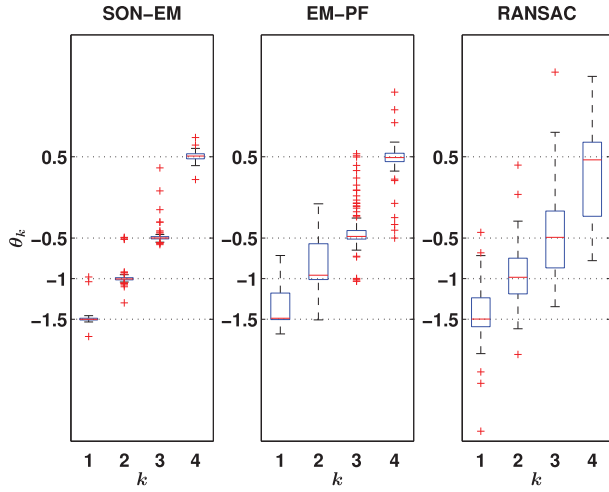
Fig. 4. Boxplots demonstrating the parameter estimates on 100 different simulated time-series. The EM-PF method was run for 20 iterations with 10.000 particles. RANSAC was parametrized with $\varepsilon = 0.1$ and 10.000 iterations. In SON-EM, the tuning parameter λ was set to one. Real parameter values are $\theta_1 = -1.5$, $\theta_2 = -1$, $\theta_3 = -0.5$, $\theta_4 = 0.5$.

Fig. 5. Comparison of the SON-EM, segreg, EM-PF and RANSAC methods. Segreg and PWASON methods were parametrized with $\lambda_{max}/\lambda = 0.001$, and run for 5 iterations with ℓ_1 regularization. The EM-PF method was run for 20 iterations with 10.000 particles. RANSAC was parametrized with $\varepsilon = 0.1$ and 10.000 iterations. In SON-EM, the tuning parameter λ was set to one.

In order to quantitatively compare the performance of the different methods, the following statistical measures were introduced.

- (measured) parameter accuracy

$$\Delta\theta(t) = \frac{1}{T}\sum_{t=1}^{T}\|\hat{\theta}(t) - \theta(t)\|_2 \tag{11}$$

- one-step ahead simulation error, Root Mean Squared Error (RMSE)

$$\text{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|y(t) - \hat{\theta}_{\hat{\eta}(t)}^{\mathsf{T}}\Phi(t)\|_2^2} \tag{12}$$

- sum squared residuals of simulation

$$\text{RMSE}_{\text{sim}} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}\|y(t) - \hat{y}(t)\|_2^2} \tag{13}$$

where $\hat{y}(t)$ is the simulated time-series. These measures characterize the accuracy of the estimated parameters and the quality of the reconstruction.

Table 1 contains statistics of 100 different simulations. The statistics also confirm that the best parameter accuracy with the smallest simulation error is achieved by the SON-EM method. Interestingly, the best one-step ahead predictions was obtained with PWASON, but this was not reflected on the parameter estimates nor on the simulation results. It is hypothesized that this was due to overfitting.

Fig. 5 shows one example run of the identification, explaining the above results. The RANSAC, the EM-PF and the PWASON algorithms show similar behavior at certain time instances, oscillating around the true values. This may be an effect of the noise, leading to the misclassification of particular data points. On the other hand, segreg and SON-EM captured well most of the switches, but the estimates of SON-EM are closer to the real parameters.

3.2. Case study on diauxic bacterial growth

Bacterial growth is the process in which an organism is growing in a substrate medium. Usually the biomass is monitored as
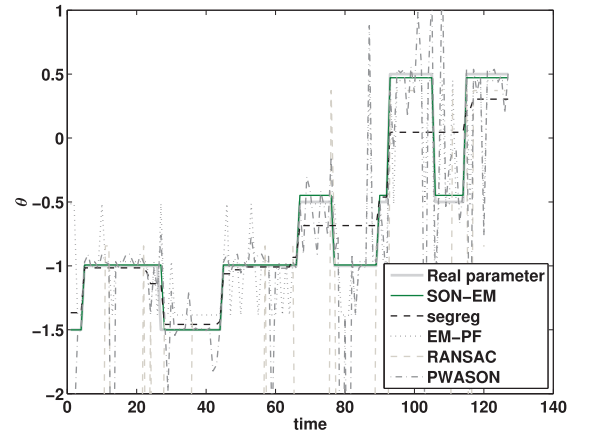
dry weight per cell volume. Such systems can be modeled as an exponential growth with saturation [18]. The term diauxic refers to a medium consisting of two different types of carbon sources. Diauxic growth was first studied in the 40's, see the comprehensive review [25] and references therein. When growing on a mixed medium, the organism first consumes the substrate that supports faster growth (preferred substrate), and then switches to the other one (secondary carbon source). Transition between carbon sources may take time, because the bacteria has to adopt its metabolism to the environmental change. The process has three distinct phases. After processing the preferred substrate in an initial exponential growth phase, a *diauxic lag* is followed, when the bacteria does not grow significantly, but synthesizes enzymes to process the secondary medium. Finally, the secondary carbon source is processed in another exponential growth phase. It is plausible to interpret the diauxic growth phenomenon as a hybrid SARX system, because a different submodel can be assigned to each phase.

Hereafter, data of *Streptococcus pneumoniae* growth experiments were used, where bacteria is grown on mixed medium of glucose as preferred carbon source and cellobiose as secondary medium [4]. The model fitted was a discrete version of the logistic model that is one of the earliest sigmoid type functions applied to bacterial growth. As shown in Appendix C, the model has a linear representation, therefore a switching extension of the model can be corresponded to a SARX model. The parameters were identified using SON-EM, the tuning parameter was set to $\lambda = 0.5$. Fig. 6 shows the estimates on experimental data with two different cellobiose concentration.

The EM clustering step is illustrated in the scatter plot of Fig. 7. Each point on the graph corresponds to a parameter estimate $\tilde{\theta}(t)$ after the first step of Algorithm refalgo:threestep. It can be seen how the smoothed parameter estimates were separated into non-overlapping clusters. Note the minor differences between $\tilde{\theta}_2(t)$ values with magnitude $\sim 10^{-9}$.

The method yielded a good fit in both experiments. In the case of cellobiose concentration of 0.3%, three different submodels were detected, corresponding to the first growth phase, the lag and the second growth phase. For lower cellobiose concentration (0.1%) the cardinality of the estimated parameters was only two, with no lag detected. The reconstructions after the second and the third steps are almost identical, however the parameters are different. This reflects that the system is not particularly sensitive to the first parameter.

Table 1
Mean ± standard deviation values of the statistical measures on 100 different simulated time-series. The measures RMSE and RMSE$_{sim}$ could not be interpreted for RANSAC because the presence of outliers.

	SON-EM	segreg	EM-PF	RANSAC	PWASON
$\Delta\theta(t)$	**0.023 ± 0.011**	0.113 ± 0.058	0.173 ± 0.054	0.592 ± 2.011	0.285 ± 0.087
RMSE	0.243 ± 0.018	0.389 ± 0.119	0.380 ± 0.211	na	**0.168 ± 0.017**
RMSE$_{sim}$	**0.441 ± 0.092**	0.744 ± 0.229	0.855 ± 0.501	na	0.909 ± 1.740

4. Discussion

Hybrid time-varying parameter models are of high importance and offer a simpler alternative to model nonlinear systems [7]. This work introduces a three-step method to estimate the parameters and the switching sequence of SARX models. The rationale is that given an (approximate) switching sequence, Problem 1 reduces to a convex optimization problem. Thus the first two steps approximate the switching sequence by first solving a regularized optimization problem, and then clustering the estimates that are close in the parameter space into K clusters. Similar approximation schemes were proposed by Roll et al. [24] to accelerate the identification of PWARX models, and by Ohlsson et al. [21] to identify PWARX and LPV systems.

In many real-world applications, the number of submodels is not known, and has to be estimated. An algorithm based on SON-EM is also provided to estimate the optimal number of submodels. The algorithm relies on the principle that if the number of submodels is underestimated, the estimation error is higher. There is a threshold, when adding new submodels to the system does not significantly decrease the estimation error. On the other hand, increasing the number of submodels above this threshold would lead to overfitting, where two or more identified parameters may represent the same generating parameter.

Noisy data is always challenging in context of parameter estimation. This is specially true if parameters are considered time-varying. A reasonable trade-off has to be found between goodness of fit to noisy data and the model complexity in order to avoid overfitting. This compromise can usually be controlled by choosing the tuning parameter(s) of the algorithms accordingly [20]. Unlike the algebraic [17] and the online recursive methods [2], SON-EM can handle noisy data by design, because the first step of the algorithm can be regarded as a smoothing applied to the data. On one hand, this filters out the majority of the noise depending on the tuning parameter. On the other hand, the smoothed estimates can be interpreted as a sophisticated initial guess for the clustering step. Another possibility to compensate the effect of noise would be to consider different kernels for regularization, as proposed by Ohlsson et al. [21].

Besides the tuning parameter, the estimates are influenced by the scale of the time between the changes of the parameters (dwell time). Approximate identification methods usually limit parameters to be slowly varying. Although this premise was not directly
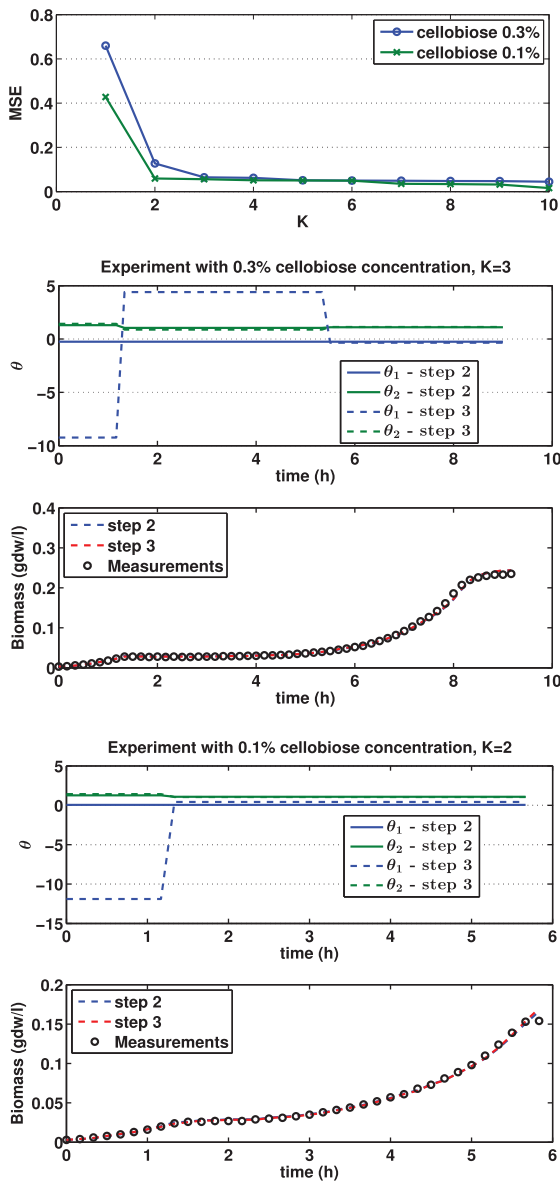

Fig. 6. Identification of growth models from experimental data of diauxic growth on a mixed medium of glucose and cellobiose. Mean Squared Error (MSE) values are plotted against K, in order to determine the optimal number of submodels to use (top). Parameter estimates and reconstruction after the second and the third step of SON-EM, where concentration of cellobiose is 0.3% (middle) and 0.1% (bottom). The tuning parameter was set to $\lambda = 0.5$.
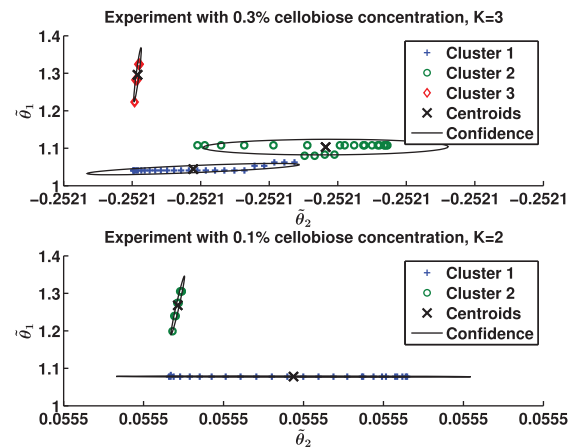

Fig. 7. Clustering of the parameter estimates in the second step of SON-EM, where concentration of cellobiose is 0.3% (top) and 0.1% (bottom). Confidence intervals were plotted around the centroids of the clusters as two times the standard deviation along the major and minor axes of the covariances.

imposed for SON-EM, it was observed that shorter dwell time causes growing estimation errors. This phenomenon is a consequence of the regularization, that acts just like group LASSO [33] on the differences of the time-varying parameter $\theta(t)$, grouping parameters of two consecutive time instances. It is hence natural, that degradation of the estimates due to decreasing dwell time can be considerably compensated by choosing lower λ values, as it was shown on an example.

In parameter estimation, naïve approaches like RANSAC [6] may also lead to good results on certain datasets, but a hard threshold has to be set in order to determine the submodel set. On the other hand, SON-EM includes a soft threshold that comes from the heuristic clustering step, and decreases the sensitivity of the results to the tuning parameter significantly. For instance on a slowly varying system, the estimates did not show changes on a rather large domain of the tuning parameter. This is a major advantage of the method proposed, because the tuning parameter does not have to be exactly determined in order to yield good final results. It is worth noting that convergence of EM to a good local optimum can be, in practice, difficult to achieve in a small number of runs. This problem can be addressed by choosing good initial partition, which can be estimated for example through k-means algorithm. On the simulated scenario, it has been shown how each step improves the estimates of the parameters. In the example considered, SON-EM was found superior to other methods both in terms of accuracy of parameters and simulation errors.

Reconstructions on *S. pneumoniae* diauxic growth show good fit to the experimental data. When the concentration of cellobiose was higher (0.3%), three phases could be distinguished by fitting different parameters. While on lower concentration (0.1%), only two phases were identified. It is presumed that in this case the lag phase is too short and hence not significant enough to be detected. The detailed analysis of the biological significance of the results is however out of the scope of this paper. It is noteworthy that the presented hybrid framework outperforms previous models for diauxic growth, specially in terms of model complexity.

5. Conclusion

A novel method for time-varying parameter identification on hybrid systems Switched affine AutoRegressive model with eXogenous inputs (SARX) models has been presented. The three-step algorithm, consists of a SON regularized convex optimization problem followed by a heuristic clustering step, and finally another optimization problem is solved. The first two steps approximate a switching sequence, while the third step adjusts the parameters according to the segmentation. The proposed method is not claimed to be superior to all existing approaches, but it was demonstrated that on a simulated scenario it outperforms some recently developed methods, such as segreg [20], EM-PF [10], PWASON [21], and the naïve RANSAC algorithm of Fischler and Bolles [6]. As the algorithm delivers a reasonable approximation to an NP-hard problem in tractable time, we hypothesize its usefulness in many areas of science and engineering. Besides having good results on simulated and *in vivo* experimental data, an important benefit of the introduced method is that it has only one scalar tuning parameter, and the final results show limited sensitivity to the tuning parameter.

Acknowledgements

Appendix A. Expectation Maximization (EM) algorithm for mixture of Gaussians

The EM algorithm maximizes the *a posteriori* likelihood of the estimates belonging in one of the of the K submodels.

$$\mathcal{L} = \prod_{t=1}^{T} P(\tilde{\theta}(t)) \tag{A.1}$$

For simplicity, a mixture of K Gaussians is used as the probability model, where means and covariances of the Gaussians are denoted by μ_k and σ_k, respectively.

$$P(\tilde{\theta}(t)) = \sum_{k=1}^{K} \mathcal{N}(\tilde{\theta}(t)|\mu_k, \sigma_k)P(k), \tag{A.2}$$

The EM algorithm consists of two steps, expectation (E-step) and maximization (M-step) described below.

Expectation: suppose we know the model (μ_k, σ_k and $P(k)$ for all $k \in [1 \ldots K]$), then the probability that the data at time instance t is generated by the submodel k is

$$P_{tk} \equiv P(k|t) = \frac{\mathcal{N}(\tilde{\theta}(t)|\mu_k, \sigma_k)P(k)}{P(\tilde{\theta}(t))}, \tag{A.3}$$

where $P(k)$ represents the population fraction, that is the probability that a randomly chosen data point is generated by the submodel k.

Maximization: suppose we know P_{tk} for all $k \in [1 \ldots K]$, $t \in [1 \ldots T]$, then the model can be updated as

$$
\begin{aligned}
\mu_k' &= \frac{\sum_{t=1}^{T} P_{tk}\tilde{\theta}(t)}{\sum_{t=1}^{T} P_{tk}} \\
\sigma_k' &= \frac{\sum_{t=1}^{T} P_{tk}(\tilde{\theta}(t) - \mu_k)^{\mathrm{T}}(\tilde{\theta}(t) - \mu_k)}{\sum_{t=1}^{T} P_{tk}} \\
P(k)' &= \frac{1}{N} \sum_{t=1}^{T} P_{tk}.
\end{aligned}
\tag{A.4}
$$

The algorithm is then described by the iterative process in Algorithm 3.

Algorithm 3. EM algorithm for mixture of Gaussians

---
1. Initialize a random model, μ_k, σ_k and $P(k)$ for all $k \in [1 \ldots K]$.
2. E-step as described in (A.3)
3. M-step as described in (A.4)
4. If the value of $\mathcal{L}$ is not increasing any more $\rightarrow$ exit
5. Else goto (2)

Finally, the switching sequence is taken as the maximum *a posteriori* estimate, being the cluster to which the estimate belongs with the highest probability.

$$\hat{\eta}(t) = \arg \max_{k}(P_{tk}) \tag{A.5}$$

Appendix B. RANSAC method for hybrid identification

The RANSAC algorithm is a naïve method for parameter estimation, for datasets that may contain outliers [6]. The algorithm is based on the concept that if the parameters can be identified on a subset of the measurements, then the set of measurements that fit poorly to the estimate are considered to be outliers.

Here the original algorithm is extended in order to identify hybrid models. The extension is based on the simple idea that from the point of the kth submodel, the output generated by the other submodels are outliers. In this way it is possible to identify the submodels one by one. The algorithm iterates until the cardinality of the outliers does not allow the estimation of a new submodel. The extended algorithm implicitly determines the number of the submodels, solving Problem 2. The extended algorithm contains the original RANSAC method as a special case, when the algorithm terminates after the first iteration, a situation for which the number of submodels equals one ($k = 1$).

Algorithm 4. RANSAC algorithm for hybrid models

---
1. Initialize the number of iterations $i = 1$, and the submodel $k = 1$
2. Select randomly a set of points $\mathcal{S}_k$ with cardinality $|\mathcal{S}_k| = s$ that is required to determine the model
3. Fit the model parameters $\hat{\theta}_k$ to the set $\mathcal{S}_k$
4. $i = i + 1$
5. Repeat from (2) until maximum iterations is reached ($i \leq i_{\max}$); In each iteration, keep only the best fitting parameter set
6. Simulate all the measurements with $\hat{\theta}_k$, and remove the set of points, $\mathcal{B}$ fitting better than a threshold ε
7. Optionally, adjust $\hat{\theta}_k$ to fit best to $\mathcal{B}$
8. If number of remaining points $< s$ then return
9. $k = k + 1$; goto (2)

Appendix C. Discrete logistic model

The discretized logistic model was derived by applying Euler's solver on the continuous logistic model [34] in equidistant time intervals.

$$x(t + 1) = x(t) + rx(t)\left(1 - \frac{x(t)}{C}\right) \quad (C.1)$$

Here, the state variable x represents the biomass, the parameter r refers to the maximum rate of growth, and C is the saturation level. The difference equation in the form shown in Eq. (C.1) is nonlinear. Without loss of generality we can suppose that the biomass is strictly positive, and divide both sides of the model by $x(t)$.

$$\frac{x(t + 1)}{x(t)} = 1 + r\left(1 - \frac{x(t)}{C}\right) \quad (C.2)$$

With the substitutions $y(t) = \frac{x(t+1)}{x(t)}$; $\theta_1 = 1 + r$; $\theta_2 = -\frac{r}{C}$, the model reduces to the following linear form

$$y(t) = \theta_1 + \theta_2 x(t). \quad (C.3)$$

References

[1] L. Bako, Identification of switched linear systems via sparse optimization, Automatica 47 (4) (2011) 668–677.
[2] L. Bako, K. Boukharouba, E. Duviella, S. Lecoeuche, A recursive identification algorithm for switched linear/affine models, Nonlinear Anal. Hybrid Syst. 5 (2) (2011) 242–253.
[3] A. Bemporad, A. Garulli, A bounded-error approach to piecewise affine system identification, IEEE Trans. Autom. Control 50 (10) (2005) 1567–1580.
[4] A. Boianelli, A. Bidossi, L. Gualdi, L. Mulas, C. Mocenni, G. Pozzi, A. Vicino, M.R. Oggioni, A non-linear deterministic model for regulation of diauxic lag on cellobiose by the pneumococcal multidomain transcriptional regulator CelR, PLoS ONE 7 (10) (2012).
[5] G. Ferrari-Trecate, M. Muselli, D. Liberati, M. Morari, A clustering technique for the identification of piecewise affine systems, Automatica 39 (2003) 205L 217.
[6] M.A. Fischler, R.C. Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography, Graph. Image Process. 24 (6) (1981) 381–395.
[7] C.W.J. Granger, Where do we go next: time varying parameter models? Stud. Nonlinear Dyn. Econ. 12 (3) (2008) 1–9.
[8] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.0 beta, 2013 http://cvxr.com/cvx
[9] A. Hartmann, J. Lemos, S. Vinga, Modeling multiple experiments using regularized optimization: a case study on bacterial glucose utilization dynamics, Comput. Biol. Med. (2014), http://dx.doi.org/10.1016/j.compbiomed.2014.08.027.
[10] A. Hartmann, S. Vinga, J.M. Lemos, Hybrid identification of time-varying parameter with particle filtering and expectation maximization, in: Mediterranean Conference on Control and Automation, Chania, Greece, 2013, pp. 884–889.
[11] W. Heemels, B.D. Schutter, A. Bemporad, Equivalence of hybrid dynamical models, Automatica 37 (2001) 1085–1091.
[12] W.P.M.H. Heemels, B.D. Schutter, A. Bemporad, On the Equivalence of Classes of Hybrid Dynamical Models, in: IEEE Conference on Decision and Control, Orlando, FL, USA, 2001, pp. 364–369.
[13] A.L. Juloski, S. Weiland, W.P.M.H. Heemels, A Bayesian approach to identification of hybrid systems, IEEE Trans. Autom. Control 50 (2005) 1520–1533.
[14] J.M. Lemos, L.M. Rato, J.S. Marques, Switching reconfigurable control based on hidden markov models, in: European Control Conference, Karlsruhe, Germany, 1999, pp. 3593–3598.
[15] H. Liang, H. Miao, H. Wu, Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model, Ann. Appl. Stat. 4 (1) (2010) 460–483.
[16] J. Lygeros, Lecture notes on hybrid systems, in: Notes for an ENSIETA workshop, 2004.
[17] Y. Ma, R. Vidal, Identification of deterministic hybrid ARX systems via the identification of Algebraic Varieties, in: Hybrid Systems: Computation and Control Lecture Notes in Computer Science, vol. 341(1), 2005, pp. 449–465.
[18] J. Monod, The growth of bacterial cultures, Annu. Rev. Microbiol. 3 (1949) 371L 394.
[19] H. Ohlsson, L. Ljung, Identification of switched linear regression models using sum-of-norms regularization, Automatica (49) (2013) 1045–1050.
[20] H. Ohlsson, L. Ljung, S. Boyd, Segmentation of ARX-models using sum-of-norms regularization, Automatica 46 (6) (2010) 1107–1111.
[21] H. Ohlsson, L. Ratliff, R. Dong, S. Sastry, Blind Identification of ARX Models with Piecewise Constant Inputs, 2013, arXiv preprint arXiv:1303.6719.
[22] S. Paoletti, A.L. Juloski, G. Ferrari-Trecate, R. Vidal, Identification of hybrid systems a tutorial, Eur. J. Control 13 (2-3) (2007) 242–260.
[23] M. Petreczky, L. Bako, S. Lecoeuche, Minimality and identifiability of SARX systems, in: IFAC Symposium on System Identification, Brussels, Belgium, 2012, pp. 541–546.
[24] J. Roll, A. Bemporad, L. Ljung, Identification of piecewise affine systems via mixed-integer programming, Automatica 40 (2004) 37L 50.
[25] M. Santillán, M.C. Mackey, Quantitative approaches to the study of bistability in the lac operon of *Escherichia coli*, J. R. Soc. Interface 5 (Suppl 1) (2008) 29–39.
[26] J.S. Shamma, An overview of LPV systems, in: Control of Linear Parameter Varying Systems with Applications, Springer, 2012, pp. 3–26.
[27] R. Tóth, Modeling and Identification of Linear Parameter-Varying Systems, Springer, 2010.
[28] R. Vidal, A. Chiuso, S. Soatto, Observability and identifiability of jump linear systems, in: IEEE Conference on Decision and Control, Las Vegas, NV, USA, 2002, pp. 3614–3619.
[29] R. Vidal, S. Soatto, A. Chiuso, Applications of hybrid system identification in computer vision, in: The European Control Conference, Kos, Greece, 2007, pp. 4853–4860.
[30] M. Vilela, C.C.H. Borges, S. Vinga, A.T.R. Vasconcelos, H. Santos, E.O. Voit, J.S. Almeida, Automated smoother for the numerical decoupling of dynamics models, BMC Bioinform. 8 (2007) 305.
[31] E.O. Voit, J. Almeida, Decoupling dynamical systems for pathway identification from metabolic profiles, Bioinformatics (Oxford, England) 20 (11) (2004) 1670–1681.
[32] S. Weiland, a.L. Juloski, B. Vet, On the equivalence of switched affine models and switched ARX models, in: IEEE Conference on Decision and Control, San Diego, CA, USA, 2006, pp. 2614–2618.
[33] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc. Ser. B 68 (2006) 49L 67.
[34] M.H. Zwietering, I. Jongenburger, F.M. Rombouts, v.K. Riet, Modeling of the bacterial growth curve, Appl. Environ. Microbiol. 56 (6) (1990) 1875–1881.