Estimating the probability of success of a simple algorithm for switched linear regression

Fabien Lauer *

Université de Lorraine, LORIA, UMR 7503, F-54506 Vandœuvre-lès-Nancy, France
CNRS, France
Inria, France

A R T I C L E I N F O

A B S T R A C T

This paper deals with the switched linear regression problem inherent in hybrid system identification. In particular, we discuss k-LinReg, a straightforward and easy to implement algorithm in the spirit of k-means for the nonconvex optimization problem at the core of switched linear regression, and focus on the question of its accuracy on large data sets and its ability to reach global optimality. To this end, we emphasize the relationship between the sample size and the probability of obtaining a local minimum close to the global one with a random initialization. This is achieved through the estimation of a model of the behavior of this probability with respect to the problem dimensions. This model can then be used to tune the number of restarts required to obtain a global solution with high probability. Experiments show that the model can accurately predict the probability of success and that, despite its simplicity, the resulting algorithm can outperform more complicated approaches in both speed and accuracy.

1. Introduction

This paper deals with hybrid dynamical system identification and more precisely with the switched linear regression problem. In this framework, a set of linear models are estimated in order to approximate a noisy data set with minimal error under the assumption that the data are generated by a switched system. We consider that the number of linear models is fixed a priori. Note that, even with this assumption, without knowledge of the classification of the data into groups associated to each one of the linear models, this problem is NP-hard [1] and is thus naturally intractable for data living in a high dimensional space. However, even for data in low dimension, algorithmic efficiency on large data sets remains an open issue.

Related work. As witnessed by the seminal works [2–6] reviewed in [1], the switched regression problem has been extensively studied over the last decade for the identification of hybrid dynamical systems. While methods in [3,4] focus on piecewise affine systems, the ones in [2,5,6] equivalently apply to arbitrarily switched systems. Due to the NP-hard nature of the problem, all methods, except the global optimization approach of [4], yield suboptimal solutions in the sense that they are based on local optimization algorithms or only solve an approximation of the original nonconvex problem. However, each approach increased our understanding of hybrid system identification with a different point of view. By studying the noiseless setting, the algebraic approach [2,7] showed how to cast the problem as a linear system of equations. The clustering-based method [3] proposed a mapping of the data into a feature space where the submodels become separable. The Bayesian approach [5] analyzed the problem in a probabilistic framework, while the bounded-error approach [6]

---

* Correspondence to: Université de Lorraine, LORIA, UMR 7503, F-54506 Vandœuvre-lès-Nancy, France.
   E-mail address: fabien.lauer@loria.fr.
   URL: http://www.loria.fr/~lauer.

switched the focus by investigating the estimation of the number of modes for a given error tolerance. Each one of these offers a practical solution to deal with a specific case: noiseless for [2], with few data for [4], with prior knowledge on parameters for [5] or on the noise level for [6].

But despite this activity, most of the proposed approaches have strong limitations on the dimension of the data they can handle and are mostly applicable to small data sets with less than a thousand points and ten regressors. The algebraic approach [2,7] provides a closed form solution, which can be very efficiently computed from large data sets, but which is only valid in the noise-free setting and rather sensitive to noise otherwise. Robust formulations of this approach exist [8,9], but these still suffer from a major limitation inherent in the algebraic approach: the complexity grows exponentially with respect to the dimension of the data and the number of modes. This issue is also critical for the approach of [10] which, for small data dimensions, efficiently deals with noise in large data sets through a global optimization strategy applied to a continuous cost function. The recent method of [11], based on sparse optimization, circumvents the issue of the number of modes by iteratively estimating each parameter vector independently, in the spirit of the bounded-error approach [6]. However, the method relies on an ℓ_1-norm relaxation of a sparse optimization problem, which requires restrictive conditions on the fraction of data generated by each mode to apply. In particular, when the number of modes increases, the assumption on the fraction of data generated by a single mode becomes less obvious. Other works on convex relaxation of sparse optimization formulations include [12,13], but the number of variables and of constraints in these convex optimization problems quickly grow with the number of data.

In a realistic scenario, say with noise, more than a thousand data points and more than two modes, globally optimal solutions (such as those obtained by Roll et al. [4]) cannot be computed in reasonable time and little can be stated on the quality of the models trained by approximate or local methods. Even for convex optimization based methods [11,9,13], the conditions under which the global minimizer of the convex relaxation coincides with the one of the original problem can be unclear or violated in practice. In this context, experimentally asserted efficiency and performance of an algorithm are of primary interest. In this respect, the paper shows empirical evidence that a rather straightforward approach to the problem can outperform other approaches from the recent literature in low dimensional problems while increasing the range of tractable problems towards larger dimensions.

Methods and contribution. This paper considers one of the most straightforward and easy to implement switched regression method and analyzes the conditions under which it offers an accurate solution to the problem with high probability. The algorithm discussed here is inspired by classical approaches to clustering and the k-means algorithm [14], hence its name: k-LinReg. As such, it is a local optimization method based on alternatively solving the problem (to be specified later) with respect to integer and real variables. The key issue in such an approach is therefore how to obtain a solution sufficiently close to the global one, which is typically tackled by multiple initializations and runs of the algorithm, but without guarantees on the quality of the result. In this paper, we focus on random initializations of the k-LinReg algorithm and the estimation of the probability of drawing an initialization leading to a satisfactory solution. In particular, we show how this probability is related to the number of data and how the number of random initializations can be chosen *a priori* to yield good results. This analysis is based on a random sampling of both the problem and initialization spaces to compute the estimates of the probability of success. Inspired by works on probabilistic methods [15], probabilistic bounds on the accuracy of these estimates are derived. These bounds provide the ground to consider the estimates as data for the subsequent estimation of a predictive model of the probability of success, from which the number of restarts from different initializations for a particular task can be inferred.

This study also reveals a major difference with the classical k-means problem, namely, that high dimensional problems can be solved efficiently and globally if the number of data is large enough. Compared with other approaches to switched linear regression, the computing time of the proposed method can even decrease when the number of data increases for high dimensional problems due to a reduced number of required restarts. Note that the approach developed in [16] has some similarities with the k-LinReg algorithm discussed here, but also some differences, for instance with respect to working in a recursive or batch manner. In addition, the issue of the convergence towards a global solution was not clearly addressed in [16], whereas it is the central subject of the proposed analysis.

Beside these aspects, the paper also provides new insights into the inherent difficulty of hybrid system identification problems measured through the probability of success for the proposed baseline method. In particular, numerical examples show that test problems typically considered in the literature can be solved with few randomly initialized runs of the k-LinReg algorithm.

Paper organization. The paper starts by formally stating the problem in Section 2. The k-LinReg algorithm is presented in Sections 3 and 4 is devoted to the study of its ability to find a solution close enough to the global one. Then, these results are used to build a model of its probability of success in Section 5, on the basis of which the number of restarts is computed in Section 5.2. Finally, Section 6 provides numerical experiments to test the proposed model and compares the final algorithm with some of the most recent approaches for hybrid system identification.

2. Problem formulation

Consider switched linear regression as the problem of learning a collection of n linear models $f_j(\boldsymbol{x}) = \boldsymbol{w}_j^T \boldsymbol{x}, j = 1, \ldots, n,$ with parameter vectors $\boldsymbol{w}_j \in \mathbb{R}^p$ from a training set of N pairs $(\boldsymbol{x}_i, y_i) \in \mathbb{R}^p \times \mathbb{R}$, where $\boldsymbol{x}_i$ is the regression vector and y_i the