Proceedings of Machine Learning Research vol 120:1–11, 2020 2nd Annual Conference on Learning for Dynamics and Control

Hierarchical Decomposition of Nonlinear Dynamics and Control for System Identification and Policy Distillation

Hany Abdulsamad † ABDULSAMAD@IAS.TU-DARMSTADT.DE
Jan Peters †* PETERS@IAS.TU-DARMSTADT.DE
† *Intelligent Autonomous Systems, Technische Universität Darmstadt*
* *Robot Learning Group, Max Planck Institute for Intelligent Systems*

Editors: A. Bayen, A. Jadbabaie, G. J. Pappas, P. Parrilo, B. Recht, C. Tomlin, M.Zeilinger

Abstract

The control of nonlinear dynamical systems remains a major challenge for autonomous agents. Current trends in reinforcement learning (RL) focus on complex representations of dynamics and policies, which have yielded impressive results in solving a variety of hard control tasks. However, this new sophistication and extremely over-parameterized models have come with the cost of an overall reduction in our ability to interpret the resulting policies. In this paper, we take inspiration from the control community and apply the principles of hybrid switching systems in order to break down complex dynamics into simpler components. We exploit the rich representational power of probabilistic graphical models and derive an expectation-maximization (EM) algorithm for learning a sequence model to capture the temporal structure of the data and automatically decompose nonlinear dynamics into stochastic switching linear dynamical systems. Moreover, we show how this framework of switching models enables extracting hierarchies of Markovian and auto-regressive locally linear controllers from nonlinear experts in an imitation learning scenario.

Keywords: Hybrid Systems, Hidden Markov Models, System Identification, Message Passing, Imitation Learning, Local Linear Controllers, Auto-Regressive Policies.

1. Introduction

The class of nonlinear dynamical systems governs a very wide range of real-world applications, and consequently underpins the most challenging problems of classical control and reinforcement learning (Fantoni and Lozano, 2002; Kober et al., 2013). Recent developments in learning-for-control have pushed towards deploying more complex and highly sophisticated representations, e.g. (deep) neural networks and Gaussian processes, to capture the structure of both dynamics and controllers, leading to overwhelming and unprecedented successes in the domain of RL (Mnih et al., 2015; Arulkumaran et al., 2017). This trend can be observed in both approximate optimal control approaches (Deisenroth and Rasmussen, 2011; Levine et al., 2016), and approximate value and policy iteration schemes (Schulman et al., 2015; Lillicrap et al., 2015; Haarnoja et al., 2018).

However, before the latest successful revival of neural networks in applications of control and robotics, researchers devised different paradigms for solving difficult control tasks. One interesting concept relied on decomposing nonlinear structures of dynamics and control into simpler local (linear) components, each responsible for an area of the state-action space. This decomposition is done with the aim of preserving interpretability and favorable mathematical properties studied over decades in classical control theory, such as local linear-quadratic assumptions (Liberzon, 2011).

© 2020 H. Abdulsamad & J. Peters.

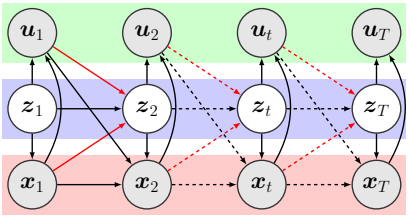

Figure 1: Graphical model of recurrent auto-regressive hidden Markov models (rAR-HMM) as presented in (Linderman et al., 2017), extended to support hybrid controls. rAR-HMMs are hybrid dynamic Bayesian networks (HDBN) that explicitly allow the discrete state z to depend on the continuous variables $\boldsymbol{x}$ and $\boldsymbol{u}$, as highlighted in red.

Parallels to this paradigm can be found in the control literature under the labels of *Hybrid Systems* or *Switched Models* (Liberzon, 2003; Haddad et al., 2006; Goebel et al., 2012; Borrelli et al., 2017), while in the machine and reinforcement learning communities the terminology of *Switching Dynamical Systems* (SDS) and *Switching State-Space Models* (SSM) is more widely adopted (Ghahramani and Hinton, 2000; Beal, 2003; Fox, 2009; Linderman et al., 2017).

Starting from this paradigm, we present in this work a view of data-driven automatic system identification and learning of composite control from the perspective of hybrid systems and switching linear dynamics. We are motivated by recent in-depth analysis and implications of using piecewise linear (PWL) activation functions such as rectified linear units (ReLU) (Montufar et al., 2014; Arora et al., 2016; Pan and Srikumar, 2016; Serra et al., 2017; Petersen and Voigtlaender, 2018), which show, that such representations effectively divide the input space into linear sub-regions.

Furthermore, our interest in hybrid systems is motivated by favorable properties inherent in such models. Besides the advantages of tractably acquiring generative models through powerful unsupervised Bayesian inference techniques, hybrid systems allow modeling of discrete events and hard nonlinearities. Moreover, they include built-in time recurrency, that enables them to capture correlations over extended time horizons, an important property when modeling physical systems governed by discrete-time ordinary differential equations. Finally, given the local scope of these systems, they are suitable for modeling phenomena with state-dependent noise.

In the following, we introduce the notation of fully observable stochastic switching linear systems and derive an expectation-maximization algorithm for inferring the parameters of a probabilistic graphical model of hybrid closed-loop dynamics. We use this inference technique to simultaneously learn the system dynamics and decompose nonlinear controllers from state-of-the-art algorithms into simpler, local Markovian and auto-regressive policies, thus dramatically reducing the complexity of dynamics and policy models. We benchmark the predictive power of the learned dynamics against a set of other commonly used representations on simulated classical control tasks.

2. Hybrid Systems as Dynamic Bayesian Networks

We focus on *Recurrent Auto-Regressive Hidden Markov Models* (rAR-HMM), a special case of Switching Linear Dynamical Systems (SLDS) as presented in (Linderman et al., 2017) and implicitly discussed in (Barber, 2006). We extend them to support exogenous and endogenous inputs in order to simulate the open- and closed-loop behaviors of driven dynamics. Figure 1 depicts the extended rAR-HMM as a hybrid dynamic Bayesian network (HDBN). The resulting graphical model is, in our opinion, most suitable for modeling dynamical systems and closely resembles *Piecewise Auto-Regressive Models* (PWARX) (Paoletti et al., 2007), widely used in the control literature.

An rAR-HMM with K regions models the trajectory of a hybrid system as follows. The initial continuous state $\boldsymbol{x}_1$ and continuous action $\boldsymbol{u}_1$ are are drawn from a conditional Gaussian distribu-

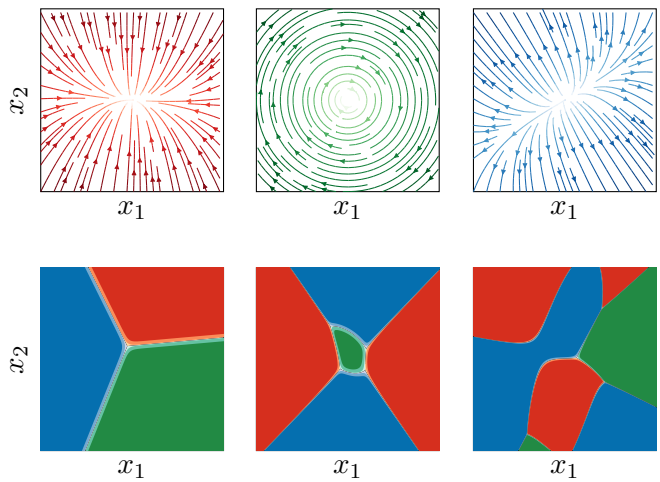

Figure 2: A hybrid system with $K = 3$ local linear regimes. The top figures depict the mean unforced continuous transition dynamics in the phase space. The lower figures show the probability of switching, with corresponding color, as a function of the state. We show different decision boundary models: linear (left), quadratic (middle) and third order polynomial (right) functions.

tions, while the initial region indicator variable z_1 is modeled by a categorical distribution

$$z_1 \sim \text{Cat}(K, \boldsymbol{\pi}), \quad \boldsymbol{x}_1 | z_1 \sim \mathcal{N}(\boldsymbol{\mu}_{z_1}, \boldsymbol{\Omega}_{z_1}), \quad \boldsymbol{u}_1 | z_1, \boldsymbol{x}_1 \sim \mathcal{N}(\mathbf{K}_{z_1}\boldsymbol{x}_1, \boldsymbol{\Sigma}_{z_1}). \tag{1}$$

The transition of the continuous state $\boldsymbol{x}_{t+1}$ and actions $\boldsymbol{u}_t$ are modeled by linear-Gaussian dynamics

$$\boldsymbol{x}_{t+1} = \mathbf{A}_{z_{t+1}}\boldsymbol{x}_t + \mathbf{B}_{z_{t+1}}\boldsymbol{u}_t + \boldsymbol{c}_{z_{t+1}} + \boldsymbol{\lambda}_{z_{t+1}}, \qquad \boldsymbol{\lambda}_z \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_z), \tag{2}$$

$$\boldsymbol{u}_t = \mathbf{K}_{z_t}\boldsymbol{x}_t + \boldsymbol{\sigma}_{z_t}, \qquad \boldsymbol{\sigma}_z \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_z), \tag{3}$$

where $\{\mathbf{A}, \mathbf{B}, \boldsymbol{c}, \mathbf{K}, \boldsymbol{\Sigma}, \boldsymbol{\Lambda}\}$ are matrices and vectors of appropriate dimensions with respect to $\boldsymbol{x}$ and $\boldsymbol{u}$, $\forall k \in [1, K]$. The transition probability $p(z_{t+1}|z_t, \boldsymbol{x}_t, \boldsymbol{u}_t)$, marking the next active dynamical regime, is governed by K categorical distributions, whose probabilities are determined by K state-action dependent generalized linear models (GLM) (Gelman et al., 2013) with logistic link functions

$$\psi_{ij} = p(z_{t+1} = i | z_t = j, \boldsymbol{x}_t, \boldsymbol{u}_t) = \frac{\exp\left(f(\boldsymbol{x}_t, \boldsymbol{u}_t; \boldsymbol{\omega}_{ij})\right)}{\sum_k \exp\left(f(\boldsymbol{x}_t, \boldsymbol{u}_t; \boldsymbol{\omega}_{kj})\right)}, \tag{4}$$

where f may contain any type of features of $\boldsymbol{x}$ and $\boldsymbol{u}$ and is parameterized by $\boldsymbol{\omega}_{ij}$, a specific vector for each transition $j \to i \; \forall i, j \in [1, K]$. Figure 2 depicts random realizations of different transition functions that lead to a variety of decompositions of the state space.

This representation of switching dynamics has a major advantage in comparison to other suggested switching models (Fox et al., 2009), as it allows for modeling the discrete transition dynamics as a function of both the continuous state $\boldsymbol{x}$ and action $\boldsymbol{u}$, thus coupling continuous and discrete dynamics of an HMM in both directions. This aspect has significant implications on the expressiveness of the model and its ability to capture the underlying dynamics of many interesting physical applications and avoiding redundancies in the number of regions required to explain the data.

The remainder of this paper will focus on using the proposed hybrid models in two manners: An **open-loop** setting that treats the control $\boldsymbol{u}$ as an exogenous input and is used for automatic system identification of general nonlinear dynamics via decomposition into continuous and discrete switching dynamics. Here we assume all regions to be first-order Markovian linear systems in the state $\boldsymbol{x}$. A **closed-loop** setting, that assumes the controls $\boldsymbol{u}$ to originate from a general nonlinear controller that is to be decomposed and inferred simultaneously with the dynamics. In this setting, the controllers can be easily extended to be polynomial in the state $\boldsymbol{x}$ or multi-step auto-regressive to incorporate past observations. Such controllers can approximate time-variant policies and lead to overall smoother control, possibly more suitable for the application on real platforms.

3. Inference of Switching Dynamics and Control

Different schemes of Bayesian inference approaches have been proposed for learning (r)AR-HMMs (Fox et al., 2009; Linderman et al., 2017), which have relied on computationally costly Gibbs sampling techniques. In this section, we propose an efficient expectation-maximization/Baum-Welch (EM/BW) algorithm (Baum et al., 1970; Dempster et al., 1977) to infer the set of parameters $\boldsymbol{\theta}$ of an rAR-HMM given multiple time series of observations $\mathbf{Y}$. Our approach is closely related to the Baum-Welch algorithms proposed in (Bengio and Frasconi, 1995) and (Daniel et al., 2016).

Baum-Welch Lower Bound: Consider again the model of an rAR-HMM in Figure 1, in which the continuous state $\boldsymbol{x}$ and action $\boldsymbol{u}$ are observed quantities, while the K-region indicators z are hidden. For inferring the model parameters we generate a dataset $\boldsymbol{\mathcal{D}}$ consisting of N trajectories of length T with $\mathbf{Y}^n_{1:T} = \{(\boldsymbol{x}^n_1, \boldsymbol{u}^n_1), ..., (\boldsymbol{x}^n_T, \boldsymbol{u}^n_T)\}$ and $\boldsymbol{z}^n_{1:T} = \{z^n_1, ..., z^n_T\}$. We drop the time indexing to avoid a cluttered notation. The learning objective is to maximize the log-likelihood of the observed data $\boldsymbol{\mathcal{D}}$ given the parameters $\boldsymbol{\theta} = \{\boldsymbol{\pi}, \boldsymbol{\mu}_{1,k}, \boldsymbol{\Omega}_{1,k}, \boldsymbol{\psi}(\boldsymbol{\omega}_k), \mathbf{A}_k, \mathbf{B}_k, \boldsymbol{c}_k, \boldsymbol{\Lambda}_k, \mathbf{K}_k, \boldsymbol{\Sigma}_k\}\ \forall k \in [1, K]$

$$\boldsymbol{\theta}_{\text{ML}} = \arg\max_{\boldsymbol{\theta}}\ \ln p(\boldsymbol{\mathcal{D}}|\boldsymbol{\theta}) = \arg\max_{\boldsymbol{\theta}}\ \ln \prod_{n=1}^{N} \sum_{\boldsymbol{z}^n} p(\mathbf{Y}^n, \boldsymbol{z}^n|\boldsymbol{\theta}), \tag{5}$$

where $p(\mathbf{Y}^n, \boldsymbol{z}^n|\boldsymbol{\theta})$ is the complete-data likelihood of a single trajectory and factorizes to the following form according to our graphical model

$$p(\mathbf{Y}^n, \boldsymbol{z}^n|\boldsymbol{\theta}) = p(z^n_1)p(\boldsymbol{x}^n_1|z^n_1) \prod_{t=1}^{T} p(z^n_{t+1}|z^n_t, \boldsymbol{x}^n_t, \boldsymbol{u}^n_t)p(\boldsymbol{u}^n_t|\boldsymbol{x}^n_t, z^n_t)p(\boldsymbol{x}^n_{t+1}|\boldsymbol{x}^n_t, \boldsymbol{u}^n_t, z^n_{t+1}). \tag{6}$$

The optimization of Equation (5) poses a technical difficulty represented by the summation over all possible trajectories of the hidden variables $\boldsymbol{z}^n$, which is of computational complexity $\mathcal{O}(NK^T)$ and is intractable in most cases. This issue is overcome by introducing the variational posterior distributions over the hidden variables $q_{\boldsymbol{z}^n}(\boldsymbol{z}^n)$, and deriving a lower bound on the log-likelihood

$$\ln \prod_{n=1}^{N} p(\mathbf{Y}^n|\boldsymbol{\theta}) = \ln \prod_{n=1}^{N} \sum_{\boldsymbol{z}^n} p(\mathbf{Y}^n, \boldsymbol{z}^n|\boldsymbol{\theta}) \geq \sum_{n=1}^{N} \sum_{\boldsymbol{z}^n} q_{\boldsymbol{z}^n}(\boldsymbol{z}^n) \ln \frac{p(\mathbf{Y}^n, \boldsymbol{z}^n|\boldsymbol{\theta})}{q_{\boldsymbol{z}^n}(\boldsymbol{z}^n)}. \tag{7}$$

We can find a point estimate of the parameters $\boldsymbol{\theta}_{\text{ML}}$ by following the standard scheme of EM of alternating between an expectation step (E-step), in which the lower bound in Equation (7) is maximized with respect to the variational distributions $q_{\boldsymbol{z}^n}(\boldsymbol{z}^n)$ given an estimate $\hat{\boldsymbol{\theta}}$, and a maximization step (M-step), that updates $\boldsymbol{\theta}$ given $q_{\boldsymbol{z}^n}(\boldsymbol{z}^n)$.

Expectation Step: By further decomposing Equation (7) and isolating $q_{\boldsymbol{z}^n}(\boldsymbol{z}^n)$ we arrive at

$$\ln \prod_{n=1}^{N} p(\mathbf{Y}^n|\boldsymbol{\theta}) \geq \sum_{n=1}^{N} \ln p(\mathbf{Y}^n|\boldsymbol{\theta}) - \sum_{n=1}^{N} \text{KL}\left(q_{\boldsymbol{z}^n}(\boldsymbol{z}^n)||p(\boldsymbol{z}^n|\mathbf{Y}^n, \boldsymbol{\theta})\right), \tag{8}$$

resulting in the optimal posterior distributions $q_{\boldsymbol{z}^n}(\boldsymbol{z}^n) = p(\boldsymbol{z}^n|\boldsymbol{x}^n_{1:T}, \boldsymbol{u}^n_{1:T}, \boldsymbol{\theta})$ after which the bound is tight, if the modeled posteriors $q_{\boldsymbol{z}^n}(\boldsymbol{z}^n)$ lie in the same family of the true distributions. These quantities are the smoothed posterior marginals $p(z^n_t|\boldsymbol{x}^n_{1:T}, \boldsymbol{u}^n_{1:T})$ and can be computed by a forward-backward algorithm (Murphy, 2012)

$$p(z^n_t = k|\boldsymbol{x}^n_{1:T}, \boldsymbol{u}^n_{1:T}) \propto p(z^n_t = k|\boldsymbol{x}^n_{1:t}, \boldsymbol{u}^n_{1:t})p(\boldsymbol{x}^n_{t+1:T}, \boldsymbol{u}^n_{t+1:T}|z^n_t = k, \boldsymbol{x}^n_t, \boldsymbol{u}^n_t). \tag{9}$$

Maximization Step: By taking the result of the E-step and reformulating Equation (7) we can write the lower bound in terms of $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$, where $\hat{\boldsymbol{\theta}}$ is the parameter vector from the last M-step

$$\mathcal{L} = Q(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) = \sum_{n=1}^{N} \sum_{\boldsymbol{z}^n} p(\boldsymbol{z}^n | \mathbf{Y}^n, \hat{\boldsymbol{\theta}}) \ln p(\mathbf{Y}^n, \boldsymbol{z}^n | \boldsymbol{\theta}). \tag{10}$$

This function is first treated under distribution-normalizing constraints via the method of Lagrangian multipliers before arriving at the complete-log-likelihood that can be maximized w.r.t. $\boldsymbol{\theta}$.

4. Empirical Evaluation

To analyze the power of auto-regressive models we construct two evaluation scenarios [1]. On the one hand, we want to quantify the quality of open-loop learned models and their ability to capture the underlying dynamics and compare them to state-of-the-art models in *long-horizon* and *small-data* regimes. On the other hand, we use the closed-loop models to learn local switching controllers in an imitation learning setting and replace complicated policy structures.

4.1. Model Learning and System Identification

In this experiment, we focus on rAR-HMMs with exogenous inputs. Our aim is to learn the dynamics of three simulated systems; a bouncing ball, an actuation-constrained pendulum, and a cart-pole system. We compare the predictive power of rAR-HMMs to classical AR-HMMs, Feedforward Neural Nets (FNN), Gaussian Processes (GP), Long-Short-Term Memory Networks (LSTM) (Hochreiter and Schmidhuber, 1997) and Recurrent Neural Networks (RNN). During evaluation, we collect a training and a test dataset. The training dataset is randomly split into 24 groups each containing a subset of trajectories and used to train different instances of all models, which are then all tested on the test dataset. All neural models have 2 hidden layers, which we test for a variety of different layer sizes, $[16, 32, 64, 128, 256, 512]$ for FNNs, $[16, 32, 64, 128, 256]$ for RNNs, and $[16, 32, 64, 128]$ for LSTMs. In the case of the auto-regressive models (r)AR-HMMs, we test a number of components K, dependent on the task. As a comparison metric, we evaluate the Normalized Mean Square Error (NMSE), averaged over the 24 splits, for a range of horizons by combing through the test trajectories step by step and predicting the given horizon. We also briefly compare the model complexity in terms of the total number of parameters of each representation.

Bouncing Ball A ball is simulated falling under the influence of gravity until it hits the ground and bounces upwards again. This system is a canonical example of a two-regime two-dimensional hybrid system, because of the hard velocity switch at the moment of impact. We simulate the dynamics with a frequency of 20 Hz and collect 25 trajectories for training with different initial heights and velocities, each 30 seconds long. This dataset is split 24 folds with 10 trajectories, 10×150 data points in each subset. The test dataset consists of 5 trajectories, each 30 seconds long. We evaluate the NMSE for horizons $h = \{1, 20, 40, 60, 80\}$ time steps. We did not evaluate a GP model in this setting, due to the long prediction horizons that result in a very high computation effort during evaluation time. The (r)AR-HMM models are tested for $K = 2$. The logistic link function of the rAR-HMM is a neural net with one hidden layer containing 16 neurons. The results depicted in Figure 3 show that the rAR-HMM model is able to approximate the true dynamics very well and outperforms the standard AR-HMM and all other neural models.

1. Source code can be found on https://github.com/hanyas/sds

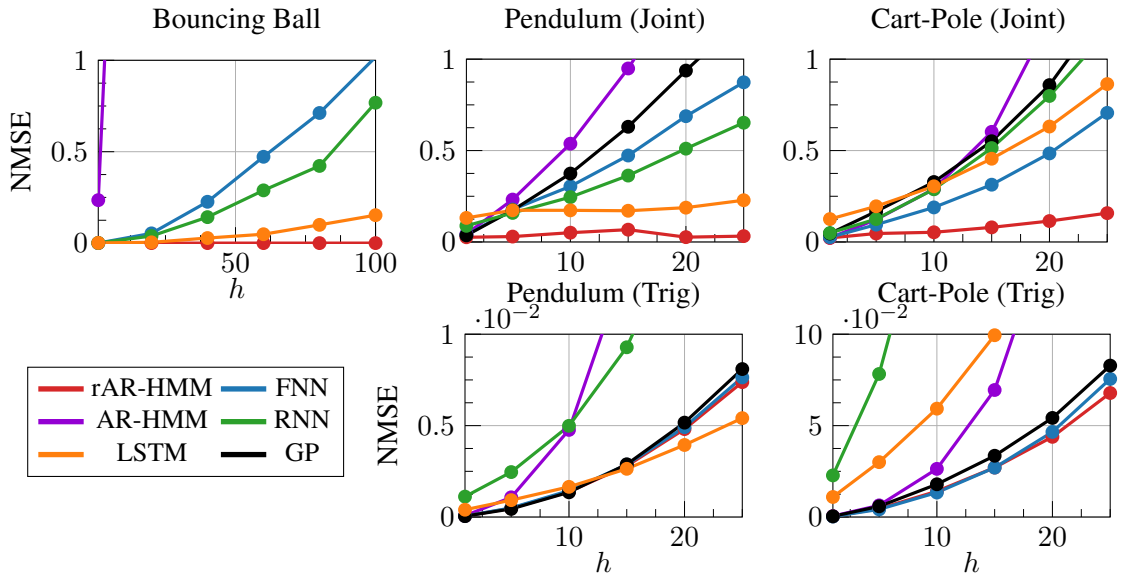

Figure 3: Comparing the h-step NMSE of rAR-HMMs to other expressive models, averaged over 24 data splits. Evaluation on three dynamical systems, a bouncing ball, a pendulum and a cart-pole. In this small-data regime scenario, rAR-HMMs exhibit the most consistent approximation capabilities.

	Bouncing	Pend. (Joint)	Pend. (Trig)	Cart. (Joint)	Cart. (Trig)
AR-HMM	22 (2)	180 (9)	130 (5)	287 (7)	275 (5)
rAR-HMM	86 (2)	468 (9)	582 (9)	575 (7)	711 (7)
FNN	1250 (32)	546 (64)	1315 (32)	1380 (32)	1445 (32)
RNN	12866 (64)	50306 (128)	3427 (32)	50820 (128)	51077 (128)
LSTM	200450 (128)	51074 (64)	51395 (64)	201732 (128)	202373 (128)

Table 1: Qualitative comparison of model complexity of the best performing representations in Figure 3. The values reflect the total number of parameters of each model. The values in parentheses reflect the size of the hidden layers of the neural models and the number of discrete components of the auto-regressive models, respectively.

Pendulum and Cart-Pole These two systems are classical benchmarks from the control literature. Here we consider two different observation models, one in the wrapped joint space, which includes a sharp discontinuity in the pendulum/pole angles, and a second model in a smooth trigonometric space of the angles. Both dynamics are simulated with a frequency of 100 Hz. We collect 25 training trajectories starting from different initial conditions and applying random explorative actions. Each trajectory is 2.5 seconds long. The 24-splits each consist of 10 trajectories, 10×250 data points. The test dataset consists of 5 trajectories, each 2.5 seconds long. Forecasting accuracy is evaluated for horizons $h = \{1, 5, 10, 15, 20, 25\}$. The (r)AR-HMM models are tested for $K = \{3, 5, 7, 9\}$ on both tasks. The logistic link function of the rAR-HMM, in both tasks, is a neural net with one hidden layer containing 24 neurons. The forecast evaluation of both tasks as shown in Figure 3 provides empirical evidence for the representation power of rAR-HMMs in both smooth and discontinuous state spaces. FNNs and GPs perform equally well in the smooth trigonometric space of observations, struggle however in the discontinuous space, similar to RNNs and LSTMs.

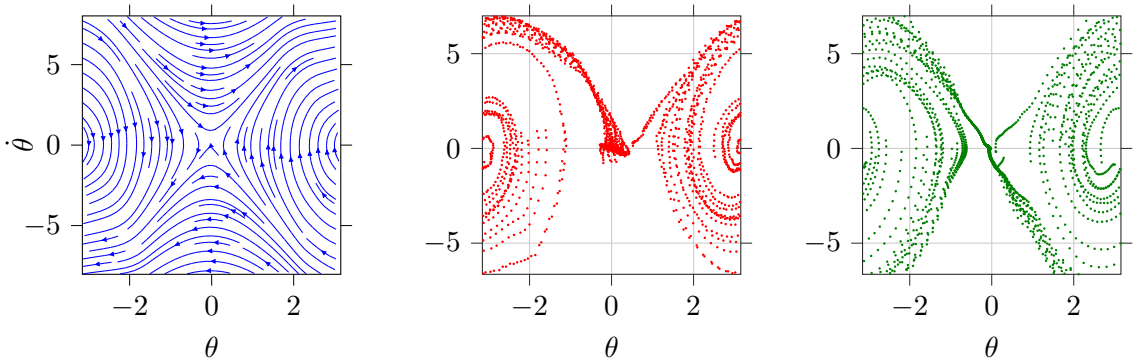

Figure 4: Phase portraits of the pendulum experiments. The identified unforced pendulum dynamics is in blue (left), while the learned stationary hybrid policy by imitation of time-variant Optimal Control (OC) trajectories is depicted in red (middle) and the stationary hybrid policy learned by imitation of a global Soft Actor-Critic (SAC) policy is seen in green (right).

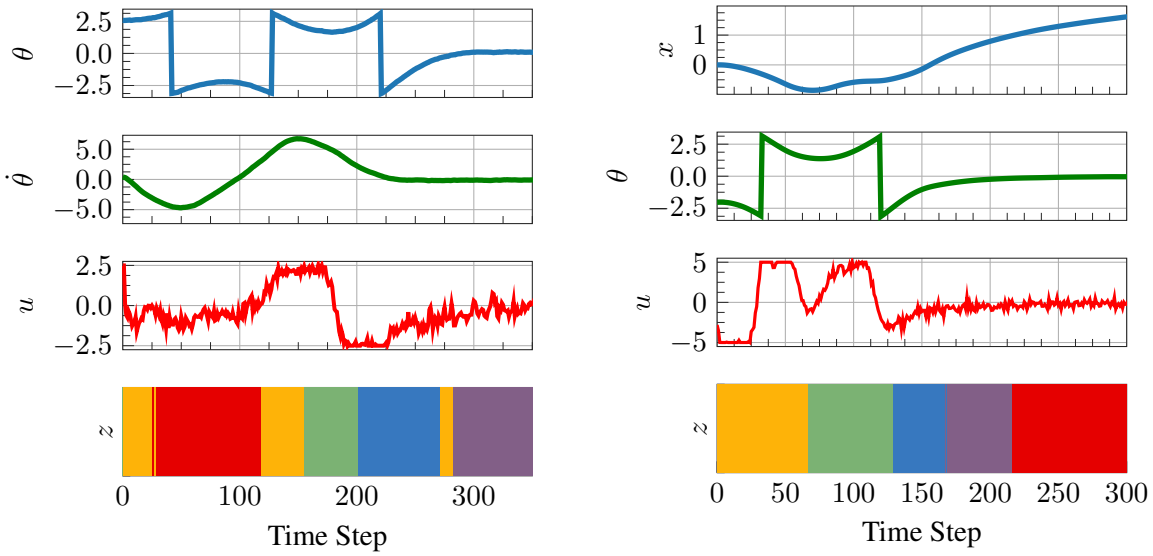

Figure 5: Sample trajectories from the learned stationary hybrid policies on the pendulum (left) and cart-pole (right) environments. Both policies are able to consistently solve both tasks, while relying on simple local representations of the feedback controllers.

Policy Distillation: For this evaluation, we consider the full rAR-HMM with endogenous inputs, that stem from an internal policy whose parameters are to be inferred. For this purpose, we supply policies from two different experts for the pendulum and cart-pole environments. One policy is a time-variant optimal controller computed via iterative Differential Dynamic Programming (Jacobson and Mayne, 1970) and the other is a reinforcement learning agent trained by Soft Actor-Critic (SAC) (Haarnoja et al., 2018), with 4545 parameters for pendulum and 17537 parameters for cart-pole. For imitation, we chose 1-step auto-regressive hybrid policies with 5 regimes each. The resulting controllers were able to complete the task of swinging up and stabilization of both systems with over 95% success rate. Figure 4 shows the phase portrait of the open-loop identified dynamics and closed-loop control during imitation. Figure 5 depicts sampled trajectories of the hybrid policies applied to the environments highlighting the switching behavior.

5. Related Work

Hybrid systems have been extensively studied in the control community and are widely used in real-world/real-time applications (Borrelli et al., 2006; Menchinelli and Bemporad, 2008). For specific research on the topic of system identification of hybrid systems, we refer the reader to the survey work in (Paoletti et al., 2007), in which the authors introduce a taxonomy of different sub-classes of hybrid systems with linear separation boundaries, and review different procedures for clustering and identifying sub-regimes of dynamics, varying from mixed-integer optimization (Bemporad et al., 2001) to Bayesian methods (Juloski et al., 2005).

Hybrid representations also play a central role in the domain of data-driven, general-purpose process modeling and state estimation (Ackerson and Fu, 1970; Hamilton, 1990). Moreover, different classes of stochastic hybrid systems serve as powerful generative models for complex dynamical behaviors (Pavlovic et al., 2001; Mesot and Barber, 2007; Oh et al., 2005). The analysis of these models has been traditionally considered under the paradigms of message passing and Markov chain Monte Carlo (Fox et al., 2009; Linderman et al., 2017), however, the recent rise of variational auto-encoders (Kingma and Welling, 2013) has enabled a new and powerful view on possible inference techniques (Becker-Ehmck et al., 2019). A distinct property of approaches that rely on auto-encoders is the need to relax the discrete variables in order to do inference.

Switching systems have also served as a powerful tool in various imitation learning approaches. Calinon et al. (2010) combine traditional HMMs with Gaussian mixture regression to represent trajectory distributions, while Daniel et al. (2016) use a hidden semi-Markov model to learn hierarchical policies. Finally, excellent work on hierarchical decomposition of policies in a fully Bayesian framework is introduced by Šošić et al. (2017), albeit under known transition dynamics.

6. Conclusion

We have presented a data-driven view for the hierarchical decomposition of nonlinear dynamics and control into simpler units based on the paradigms of hybrid systems and probabilistic graphical models. We have discussed the advantages of such models and demonstrated on a range of classical benchmarks that auto-regressive locally linear models are powerful representations that can capture long-horizon nonlinear dynamics in scenarios with limited data. Furthermore, we have shown that the extension to support endogenous inputs is useful in an imitation learning setup, and can drastically compress over-parameterized neural polices. Moreover, we have used a natural extension of this framework to represent and learn non-Markovian auto-regressive policies.

We view this work as a first step towards a more structured application of machine learning principles in learning-for-control, hoping that by applying regularization through simplicity (Occam's razor), such structures may have a positive effect on the learning procedures that have recently drifted towards ever larger models. Future considerations will focus on more efficient learning algorithms of these graphical models and pursue to highlight their main advantages for the domain of learning-for-control, by extending them to become Bayesian and applying them to environments with discrete events and state-dependent noise.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement #640554 (SKILLS4ROBOTS).

References

G Ackerson and K Fu. On state estimation in switching environments. *IEEE Transactions on Automatic Control*, 1970.

Raman Arora, Amitabh Basu, Poorya Mianjy, and Anirbit Mukherjee. Understanding deep neural networks with rectified linear units. *arXiv preprint arXiv:1611.01491*, 2016.

Kai Arulkumaran, Marc Deisenroth, Miles Brundage, and Anil Anthony Bharath. A brief survey of deep reinforcement learning. *arXiv preprint arXiv:1708.05866*, 2017.

David Barber. Expectation correction for smoothed inference in switching linear dynamical systems. *Journal of Machine Learning Research*, 2006.

Leonard E Baum, Ted Petrie, George Soules, and Norman Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 1970.

Matthew J Beal. *Variational Algorithms for Approximate Bayesian Inference*. PhD thesis, Gatsby Computational Neuroscience Unit, University College London, 2003.

Philip Becker-Ehmck, Jan Peters, and Patrick Van Der Smagt. Switching linear dynamics for variational bayes filtering. *arXiv preprint arXiv:1905.12434*, 2019.

Alberto Bemporad, Jacob Roll, and Lennart Ljung. Identification of hybrid systems via mixed-integer programming. In *Proceedings of the 40th IEEE Conference on Decision and Control*. IEEE, 2001.

Yoshua Bengio and Paolo Frasconi. An input output HMM architecture. In *Advances in Neural Information Processing Systems*, 1995.

Francesco Borrelli, Alberto Bemporad, Michael Fodor, and Davor Hrovat. An MPC/hybrid system approach to traction control. *IEEE Transactions on Control Systems Technology*, 2006.

Francesco Borrelli, Alberto Bemporad, and Manfred Morari. *Predictive control for linear and hybrid systems*. Cambridge University Press, 2017.

Sylvain Calinon, Florent D'halluin, Eric L Sauser, Darwin G Caldwell, and Aude G Billard. Learning and reproduction of gestures by imitation. *IEEE Robotics & Automation Magazine*, 2010.

Christian Daniel, Herke Van Hoof, Jan Peters, and Gerhard Neumann. Probabilistic inference for determining options in reinforcement learning. *Machine Learning*, 2016.

Marc Deisenroth and Carl E Rasmussen. PILCO: A model-based and data-efficient approach to policy search. In *Proceedings of the 28th International Conference on Machine Learning*, 2011.

Arthur P Dempster, Nan M Laird, and Donald B Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society*, 1977.

Isabelle Fantoni and Rogelio Lozano. *Non-linear control for underactuated mechanical systems*. Springer Science & Business Media, 2002.

Emily Fox. *Bayesian nonparametric learning of complex dynamical phenomena*. PhD thesis, Massachusetts Institute of Technology, 2009.

Emily Fox, Erik B Sudderth, Michael I Jordan, and Alan S Willsky. Nonparametric Bayesian learning of switching linear dynamical systems. In *Advances in Neural Information Processing Systems*, 2009.

Andrew Gelman, Hal S Stern, John B Carlin, David B Dunson, Aki Vehtari, and Donald B Rubin. *Bayesian data analysis*. Chapman and Hall/CRC, 2013.

Zoubin Ghahramani and Geoffrey E Hinton. Variational learning for switching state-space models. *Neural Computation*, 2000.

Rafal Goebel, Ricardo G Sanfelice, and Andrew R Teel. *Hybrid Dynamical Systems: modeling, stability, and robustness*. Princeton University Press, 2012.

Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. *arXiv preprint arXiv:1801.01290*, 2018.

Wassim M Haddad, VijaySekhar Chellaboina, and Sergey G Nersesov. Impulsive and hybrid dynamical systems. *Princeton Series in Applied Mathematics*, 2006.

James D Hamilton. Analysis of time series subject to changes in regime. *Journal of Econometrics*, 1990.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 1997.

David H Jacobson and David Q Mayne. Differential dynamic programming. 1970.

Aleksandar Lj Juloski, Siep Weiland, and WPMH Heemels. A Bayesian approach to identification of hybrid systems. *IEEE Transactions on Automatic Control*, 2005.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Jens Kober, J Andrew Bagnell, and Jan Peters. Reinforcement learning in robotics: A survey. *The International Journal of Robotics Research*, 2013.

Sergey Levine, Chelsea Finn, Trevor Darrell, and Pieter Abbeel. End-to-end training of deep visuomotor policies. *The Journal of Machine Learning Research*, 2016.

Daniel Liberzon. *Switching in systems and control*. Springer Science & Business Media, 2003.

Daniel Liberzon. *Calculus of variations and optimal control theory: A concise introduction*. Princeton University Press, 2011.

Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *arXiv preprint arXiv:1509.02971*, 2015.

Scott W Linderman, Matthew J Johnson, Andrew C Miller, Ryan P Adams, David M Blei, and Liam Paninski. Bayesian learning and inference in recurrent switching linear dynamical systems. In *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics*, 2017.

P Menchinelli and Alberto Bemporad. Hybrid model predictive control of a solar air conditioning plant. *European Journal of Control*, 2008.

Bertrand Mesot and David Barber. Switching linear dynamical systems for noise robust speech recognition. *IEEE Transactions on Audio, Speech, and Language Processing*, 2007.

Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 2015.

Guido F Montufar, Razvan Pascanu, Kyunghyun Cho, and Yoshua Bengio. On the number of linear regions of deep neural networks. In *Advances in Neural Information Processing Systems*, 2014.

Kevin P Murphy. *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.

Sang Min Oh, James M Rehg, Tucker Balch, and Frank Dellaert. Data-driven MCMC for learning and inference in switching linear dynamic systems. In *Association for the Advancement of Artificial Intelligence*. Georgia Institute of Technology, 2005.

Xingyuan Pan and Vivek Srikumar. Expressiveness of rectifier networks. In *International Conference on Machine Learning*, 2016.

Simone Paoletti, Aleksandar Lj. Juloski, Giancarlo Ferrari-Trecate, and René Vidal. Identification of hybrid systems: A tutorial. *European Journal of Control*, 2007.

Vladimir Pavlovic, James M Rehg, and John MacCormick. Learning switching linear models of human motion. In *Advances in Neural Information Processing Systems*, 2001.

Philipp Petersen and Felix Voigtlaender. Optimal approximation of piecewise smooth functions using deep ReLU neural networks. *Neural Networks*, 2018.

John Schulman, Sergey Levine, Pieter Abbeel, Michael Jordan, and Philipp Moritz. Trust region policy optimization. In *International Conference on Machine Learning*, 2015.

Thiago Serra, Christian Tjandraatmadja, and Srikumar Ramalingam. Bounding and counting linear regions of deep neural networks. *arXiv preprint arXiv:1711.02114*, 2017.

Adrian Šošić, Abdelhak M Zoubir, and Heinz Koeppl. A bayesian approach to policy recognition and state representation learning. *IEEE transactions on pattern analysis and machine intelligence*, 2017.