2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)
June 28-30, 2023. Seattle, Washington, USA

Data-driven Identification of Stochastic System Dynamics under Partial Observability Using Physics-Based Model Priors with Application to Acrobot

Victor Vantilborgh*, Tom Lefebvre and Guillaume Crevecoeur

***Abstract*—Accurate dynamical models form a main driver for high performance mechatronic applications. Conventional modeling of mechatronic systems is often limited in its ability to handle poorly understood phenomena and may not be adequate in instances where the underlying dynamics are not fully known nor fully captured by sensory data. To overcome these limitations, we propose a physics-based data-driven state-space modeling approach. We phrase the problem as a probabilistic representation learning problem. The hybrid model combines known physical relations with parametrized functions, represented as neural networks, to serve as substitutes for the previously unidentified substructures. The identification problem is solved using the Expectation-Maximization (EM) algorithm. In the Expectation step, Bayesian smoothers are utilized to provide complete state estimates from partial observations. In the M-step, the hybrid model is fitted onto the smoothed data. Although the physics based prior model comes at the loss of expressiveness, it serves as a strong model prior. The use of a physical model prior is beneficial both to improve the accuracy of the inference during the E-step as well as to reduce the complexity of the M-step. The proposed methodology is applied and validated for the identification of friction in both joints of an acrobat, with only measurements available in one joint. Numerical experiments demonstrate the methods capability of identifying comprehensive representations of the friction characteristics in both joints and possessing accurate predictive abilities.**

I. INTRODUCTION

Accurate modeling of mechatronic systems is crucial for realizing their full potential and enabling advanced capabilities such as precise motion control, predictive maintenance, and informed decision-making. Conventional modeling of mechatronic systems is based on an expert's understanding of the system dynamics, represented through physics-based models [1]. These models provide a high degree of interpretability and robustness, but they are limited in their ability to handle poorly understood phenomena. On the other hand, when large amounts of observational data are available, machine learning techniques can be employed to infer the underlying system dynamics [2], [3]. A wide variety of data-driven system identification techniques have been developped, such as Hammerstein-Wiener structures [4], nonlinear ARMAX models [5] and Neural Network (NN) models [6]. These data-driven methods can achieve high accuracy, but they often lack the interpretability of physics-based models due to their lack of physical intuition.

At the intersection of these two fields, a persistent research effort is aimed at integrating the advantages of both physics-based and data-driven modeling methods, while mitigating their shortcomings. A spectrum of modeling strategies exists, which endeavors to find a balance between interpretability and precision by leveraging both expert knowledge and observational data to generate a comprehensive representation of the system dynamics. These modeling formalisms generally focus on specific system classes, such as those described by Physics-Informed Neural Networks (PINN) [7], Langrangian Neural Networks (LNN) [8] and Neural Network Augmented Physics (NNAP) [9], to name a few examples.

However, in numerous instances, the underlying dynamics of a dynamical system are not fully known and can only be partially observed. In these scenarios, traditional machine learning techniques that merely construct an input-output mapping are inadequate, as the governing dynamics are not fully captured in the observations. Generative models such as Hidden Markov Models (HMM) [10] provide a more appropriate framework. The State-Space models (SSM) describing such an HMM decompose the model into two components: a *transition* model that captures the underlying latent dynamics of the observed data and an *emission* model that maps the latent variables to the observation domain, accounting for system and measurement noise respectively. A key aspect of HMMs is their computational flexibility, enabling the inversion of the emission model into an *inference* model which allows to render estimates of the latent variables given the observations [11].

In this study, we combine the benefits of the NNAP modeling formalism and HMMs for nonlinear probabilistic state-space model identification in mechatronic systems with incomplete state observability. This problem is considered as a special case of the generic variational representation learning problem [11], [12]. We explore the potential of the NNAP architecture as a physics-informed model prior for the latent state transition model in conjunction with approximate Bayesian inference algorithms. For simplicity, in this work we impose the mild assumption that there is a known relation between the observed variables and part of the state variables occurring in the state-space model.

Overall, the contributions of this paper are: (i) the integration of physics-based neural network models in the theoretical Hidden Markov Model framework within the larger scope of

*Corresponding author.
V. Vantilborgh, T. Lefebvre and G. Crevecoeur are with the Department of Electromechanical, Systems and Metal Engineering, Ghent University, B-9052 Ghent, Belgium e-mail: {victor.vantilborgh, tom.lefebvre, guillaume.crevecoeur}@ugent.be.
V. Vantilborgh, T. Lefebvre and G. Crevecoeur are member of Core Lab MIRO, Flanders Make, Belgium.

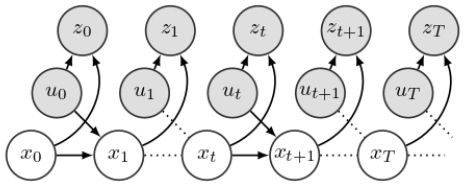


Fig. 1: Graphical representation of a Hidden Markov Model.

probabilistic representation learning. (ii) Assessment of the performance of different approximate Bayesian inference algorithms tailored to nonlinear models, specifically the Extended Kalman Smoother (EKS) and the Particle Smoother (PS). (iii) Analysis of the performance of our approach in terms of interpretability and predictive capabilities w.r.t. to pure data-driven approaches, with application on the acrobot.

II. Problem formulation

We consider the problem of identifying a substructure of a probabilistic nonlinear state-space model under incomplete observability. This problem has enjoyed considerable attention in the system identification community, however mainly in a deterministic setting or assuming measurement noise but no system noise. We generalize the scope of the problem to general nonlinear and stochastic systems in discrete time.

Given N data sequences $\mathcal{D} = \{z^n_{0:T}, u^n_{0:T}\}_{n=1}^N$, each consisting of $T+1$ samples, with input data $u_t \in \mathbb{R}^{n_u}$ and output data $z_t \in \mathbb{R}^{n_z}$, we are interested in learning a representation of the following measurement likelihood [11]. Under the assumption of independently and identically distributed (i.i.d) experiments the probability decomposes into the product of the probability of the N individual experiments.

$$p(\{z^n_{0:T}\}_{n=1}^N | \{u^n_{0:T}\}_{n=1}^N) = \prod_{n=1}^N p(z^n_{0:T} | u^n_{0:T}) \tag{1}$$

In the setting where the observed variables do not allow to reconstruct the full state, it is common to assume a sequential latent space $x_t \in \mathbb{R}^{n_x}$. Representing the time-series data as Hidden Markov Models allows to elegantly combine the observed variables z_t and u_t with the latent variables x_t. We refer to Fig. 1 for a graphical representation.

A HMM is completely characterised by the conditional probabilities associated to the graph vertices and the root node. For HMMs the conditional dependencies are also known as the Markov properties. Clearly this formulation coincides with a probabilistic version of a standard discrete time state-space model or a PSSM.

$$\begin{aligned} x_0 &\sim p(x_0) \\ x_t &\sim p(x_t | x_{t-1}, u_{t-1}) \\ z_t &\sim p(z_t | x_t, u_t) \end{aligned} \tag{2}$$

The conditional structure of the HMM allows to decompose (1) as follows

$$p_\theta(z_{0:T} | u_{0:T}) = \int p_\theta(z_{1:T}, x_{0:T} | u_{0:T}) dx_{0:T} \tag{3}$$

where

$$\begin{aligned} &p(z_{1:T}, x_{0:T} | u_{0:T}) \\ &\quad = p(x_0) \prod_{t=0}^T p(z_t | x_t, u_t) \prod_{t=1}^T p(x_t | x_{t-1}, u_{t-1}) \end{aligned} \tag{4}$$

The goal in this work is thus to determine representations for the probabilities $p(x_t|x_{t-1}, u_{t-1})$ and $p(z_t|x_t, u_t)$ presented with the data $\mathcal{D}$. In practice this can be achieved by parameterizing the probabilities and determining a Maximum A Posteriori estimate optimizing the likelihood of the data

$$\hat{\theta}_{ML} = \max_\theta \mathcal{L}_\theta(\{z^n_{0:T}\}_{n=1}^N) = \max_\theta \sum_{n=1}^N \log p_\theta(z^n_{0:T} | u^n_{0:T}) \tag{5}$$

We propose to use the following parameterization where the random variables w_t and v_t serve as the sources of process and observation noise, respectively.

$$\begin{aligned} p(x_t | x_{t-1}, u_{t-1}) &= f_\theta(x_{t-1}, u_{t-1}) + w_t, \;\; w_t \sim \mathcal{N}(0, \mathrm{Q}) \\ p(z_t | x_t, u_t) &= g_\theta(x_t, u_t) + v_t, \;\; v_t \sim \mathcal{N}(0, \mathrm{R}) \end{aligned} \tag{6}$$

Ideally, the latent space coincides with the physical state of the system. In general however the latent space can be identified up to a similarity transform from observations alone, i.e. any invertible transformation $\mathcal{T}$ such that $\zeta_t = \mathcal{T}^{-1} f(\mathcal{T}(\zeta_{t-1})) = f_\mathcal{T}(\zeta_{t-1})$ and $z_t = g(\mathcal{T}(\zeta_t)) = g_\mathcal{T}(\zeta)$ results in an equivalent solution. Therefore, without including prior knowledge, we cannot identify the latent space uniquely nor make it coincide with any physically interpretable space.

In this work specifically, we examine the situation where a SSM has been partially identified but is still affected by multiple unidentified nonlinear phenomena. Furthermore, we consider scenarios where it is impossible to observe the complete system's state x, resulting in an incomplete observation represented by z, i.e. $\dim[z] < \dim[x]$. However, we assume that the function g is known and invertible such that from z we can construct $\bar{x}$ with $\bar{x} \subset x$ the *observable* state and that $\bar{x}$ together with the physics based model prior contains sufficient information to identify the unknown substructures and as such to reconstruct the complete state $x = \{\bar{x}, \underline{x}\}$, with $\underline{x}$ the *unobservable* states.

To address this, we propose a data-driven system identification method for identifying the unknown phenomena complementing the physical model. The transition model f_θ will be parametrized by substituting neural networks for the unidentified terms in the prior physics model. Our motivation is to solve (1) more efficiently and gain additional insights in the dynamical behaviour of the system.

III. Methodology

A. NNAP Architecture

The Neural Network Augmented Physics architecture is a modeling formalism for constructing dynamical models capable of predicting the next state, x_{t+1}, given the current state, x_t, and input, u_t, of a systems with nonlinear dynamics.

The model is particularly useful for systems that can be governed by Ordinary Differential Equations (ODE) of the form $f(x_t, u_t, y_t)$, where the unknown static relation $y_t = h(x_t, u_t)$ is not analytically expressible nor directly measurable. Within the NNAP framework these unknown phenomena are approximated by a neural network h_θ that is parametrized by θ denoting its weights and biases.

By utilizing this NNAP architecture as a physics-informed model prior, the unknown ODE substructure can be identified while the structure of the model remains fixed. The resulting model is interpretable and allows for the unique identification of unknown phenomena. The substitution of the unidentified terms represented by h_θ in the prior ODE model results in the following general expression

$$\dot{x}_t = f'(x_t, u_t, h_\theta(q_t)) = f'_\theta(x_t, u_t) \tag{7}$$

where $q_t = \phi(x_t, u_t)$ denotes the input feature vector, with ϕ some arbitrary preprocessing function.

Numerical integration techniques can be employed to solve this derivative function over the sampling time of the measurement sequence, resulting in a discrete-time state-space model

$$x_{t+1} = f_\theta(x_t, u_t) \tag{8}$$

B. Expectation-Maximization algorithm

It is widely acknowledged that the optimization of the objective in equation (5) is a challenging task, provided that x is unobserved and the distribution $p_\theta(x_t|x_{t-1}, u_{t-1})$ is unknown prior to the estimation of the parameter θ. To circumvent this, we approximate $\mathcal{L}_\theta$ by its minimum variance estimate $\mathcal{Q}_{\theta,\theta^*}$ given the observed data and an assumption θ^* on the true value of the parameter θ.

$$\begin{aligned}\mathcal{Q}_{\theta,\theta^*} &= \mathbb{E}_{\theta^*}\left[\log p_\theta(z_{0:T}, x_{0:T}|u_{0:T})|z_{0:T}\right] \\ &= \int \log p_\theta(z_{0:T}, x_{0:T}|u_{0:T}) p_{\theta^*}(x_{0:T}|u_{0:T}, z_{0:T}) dx_{0:T}\end{aligned} \tag{9}$$

This gives rise to an iterative procedure where Q_{θ,θ_k} is evaluated for θ_k and optimized for θ yielding the next iterate θ_{k+1}. This procedure is known as the Expectation-Maximization algorithm, see Algorithm 1.

The function Q_{θ,θ^*} can be further decomposed by substitution of (4) into three separate terms each targeting one of the components of the HMM (2).

$$Q_{\theta,\theta^*} = Q^0_{\theta,\theta^*} + Q^f_{\theta,\theta^*} + Q^g_{\theta,\theta^*} \tag{10}$$

where

$$\begin{aligned}Q^0_{\theta,\theta^*} &= \int \log p_\theta(x_0) p_{\theta^*}(x_0|u_{0:T}, z_{0:T}) dx_0 \\ Q^f_{\theta,\theta^*} &= \sum_{t=0}^{T-1} \int\int \log p_\theta(x_{t+1}|x_t, u_t) p_{\theta^*}(x_{t+1}, x_t|u_{0:T}, z_{0:T}) dx_t dx_{t+1} \\ Q^g_{\theta,\theta^*} &= \sum_{t=0}^{T} \int \log p_\theta(z_t|x_t, u_t) p_{\theta^*}(x_t|u_{0:T}, z_{0:T}) dx_t\end{aligned} \tag{11}$$

Evaluation of the different terms requires evaluation of the posterior expectation. Thus if we want to use the exact solution we should be able to evaluate the posterior distribution $p(x_{0:T}|u_{0:T}, z_{0:T})$, which is intractable for nonlinear probabilistic state-space models.

Algorithm 1: Expectation-Maximization Algorithm

1: Set $k = 0$, initialize θ_0
2: **repeat**
3: E-step: calculate $Q(\theta, \theta_k)$
4: M-step: $\theta_{k+1} = \arg\max_\theta Q(\theta, \theta_k)$
5: **until** convergence: $Q(\theta_k, \theta_{k-1}) - Q(\theta_{k-1}, \theta_{k-2}) \to 0$

C. Approximating the Posterior distribution

The computation of $\mathcal{Q}_{\theta,\theta^*}$ primarily depends on the evaluation of the posterior probability $p(x_{0:T}|u_{0:T}, z_{0:T})$. Here we demonstrate how the computational flexibility of the HMM architecture can be exploited to compute this distribution, which is known as the smoothing distribution.

In the theory of HMMs, the probability of state x_t given the sequence of observations $z_{0:t}$ is referred to as the filtering distribution. This distribution can be calculated through a recursive process.

$$\begin{aligned}&p(x_{t+1}|z_{0:t+1}, u_{0:t+1}) \\ &\quad \propto p(z_{t+1}|x_{t+1}) \int p(x_{t+1}|x_t, u_t) p(x_t|z_{0:t}, u_{0:t}) \mathrm{d}x_t\end{aligned} \tag{12}$$

The smoothing distribution can be evaluated in a backward recursive manner, once the filtering distribution is available

$$\begin{aligned}&p(x_t|z_{0:T}, u_{0:T}) \\ &\quad \propto p(z_t|x_t) \int p(x_{t+1}|x_t, u_t) p(x_t|z_{0:T}, u_{0:T}) \mathrm{d}x_t\end{aligned} \tag{13}$$

These are typically approximated as it quickly becomes intractable to evaluate exactly for nonlinear systems. In the present work we consider two popular choices. We put forth the Extended Rauch-Tung-Striebel (RTS) Smoother and the Particle Smoother (PS) as respectively a gradient and population based approximation of the exact Bayesian Smoother in order to calculate this posterior distribution.

1) Extended Rauch-Tung-Striebel Smoother: The Extended RTS smoother assumes the transition (f) and emission (g) functions from (6) to be differentiable and the random variables w_t and v_t to be sources of Gaussian white noise with covariance matrices Q_t and R_t respectively. The RTS filter and smoother perform approximate inference through linearization of the nonlinear equations of the systems model

$$\begin{aligned}\mathrm{A}_t &= \frac{\partial f(\mu_t, u_t)}{\partial \mu_t} \\ \mathrm{C}_t &= \frac{\partial g(\mu_t, u_t)}{\partial \mu_t}\end{aligned} \tag{14}$$

where μ_t is the expected state conditioned onto measurements $z_{0:t-1}$ and the Jacobian matrices A_t and C_t correspond to the transition and observation matrix of linear state-space models, respectively. The recursive schemes for calculating the filtered

and smoothed distributions given these linearized functions are then identical to those of the standard linear Kalman filter and smoother. A detailed discussion is presented in [11].

2) Particle Smoother: The fundamental idea underlying particle smoothers, more formally known as sequential importance sampling (SIR) methods, is to approximate the integrals of (12) and (15) by a sum of *sufficiently many* (N_p) uncorrelated samples $\hat{x}^i$, i.e. the particles, such that

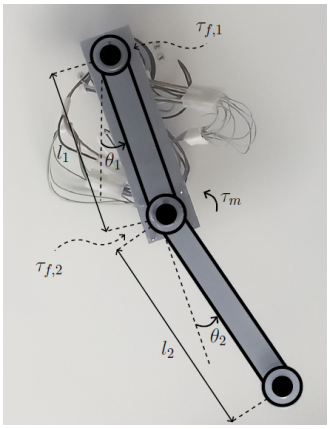
$$p(x_t|z_{0:T}, u_{0:T}) \approx \sum_{i=1}^{N_p} w^i_{t|0:T} \delta(x_t - \hat{x}^i_t) \tag{15}$$

with δ the Dirac delta operator and $w^i_{t|0:T}$ the weights denoting the probability of the occurrence of particle $\hat{x}^i_t$ at moment t given the sequence of observations $z_{0:T}$. An in depth description can be found in [11], [14].

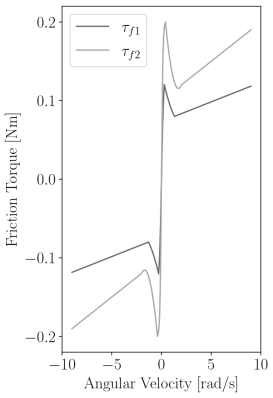
IV. Numerical validation on acrobot

We verify the combined architecture described in the previous section on the Acrobot system.

1) System description: The Acrobot system is a double pendulum composed of two links connected by joints. It is an underactuated and highly nonlinear system, with only the joint connecting the two links being actuated by a servomotor.



(a) Schematic representation of the acrobot.



(b) True friction characteristics for the joint friction torques.

Fig. 2: The acrobot system.

The simulation environment used in this study is based on OpenAI's Gym [15] implementation of the acrobot, as described by Sutton [16], with modifications to be consistent with an experimental lab setup of the acrobot. A schematic representation is given in Fig. 2a. The modifications include the integration of friction terms in the form of joint friction torques $\tau_f(q, \dot{q}) = \{\tau_{f_1}, \tau_{f_2}\}$, which were derived from friction characteristics identified on the experimental lab setup. These static friction characteristics are depicted in Fig. 2b. The friction characteristics are represented as an angular velocity to friction torque mapping and encompass friction effects, such as Coulomb friction, viscous friction, and the Stribeck effect. These friction terms are incorporated into the governing equations, which can be represented as

$$\mathrm{M}(q)\ddot{q} + c(q, \dot{q}) = \mathrm{B}\tau + \tau_f(q, \dot{q}) \tag{16}$$

where q refers to the generalised coordinate vector, $\mathrm{M}(q)$ and $c(q, \dot{q})$ denote the inertia matrix and the vector of Coriolis, centrifugal and gravitational forces, respectively, τ represents the control input of the second joint and finally B denotes a selection matrix. The parameter values used in the simulation are summarized in Table I. Additionally, the action space was altered to a continuous input space. The state vector $x = \{q, \dot{q}\} = \{\theta_1, \theta_2, \dot{\theta}_1, \dot{\theta}_2\}$ defines the system.

TABLE I: The acrobot's parameters.

Parameter	m_1	m_2	l_1	l_2
Value	0.235	0.250	0.275	0.285
Unit	[kg]	[kg]	[m]	[m]

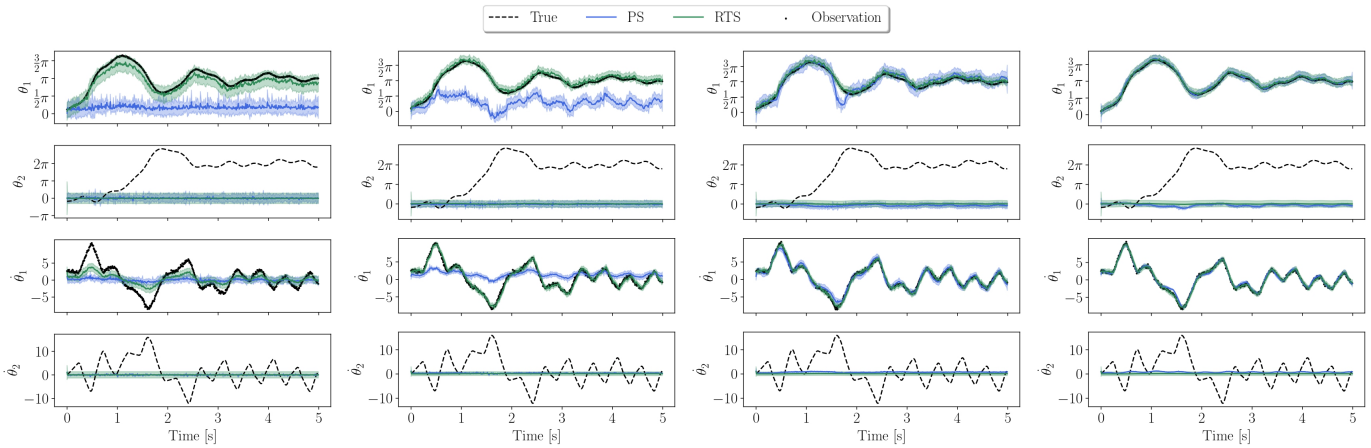
We aim to validate our method by identifying the friction characteristics τ_f in (16) of the acrobot under partial observability. In this scenario, partial observability refers to the fact that only the angle and angular velocity of the first joint are measured, i.e $\bar{x} = \{\theta_1, \dot{\theta}_1\}$, while the second link's angle and angular velocity are unobservable, i.e. $\underline{x} = \{\theta_2, \dot{\theta}_2\}$.

A. Results and discussion

The effectiveness of a hybrid model learned via the proposed probabilistic modeling framework is validated and compared to that of a purely data-driven model in terms of its learning behavior, filtering and prediction performance. Additionally, the physical interpretability of the learned hybrid model is assessed. The data-driven model has no physics-informed prior, i.e. a neural network represents the complete transition model f_θ. Furthermore, the performance of the proposed framework is compared using both the extended RTS smoother and the particle smoother for the implementation of the models and the EM training algorithm. We refer to the *RTS implementation* and the *PS implementation* respectively. For this numerical case study, the models are learned from a relatively limited training data set consisting of $N = 20$ data sequences, each with $T = 500$ time steps, sampled at $100Hz$. We have 5 additional test sequences of identical length for the purpose of validation. The starting position of the acrobot in each sequence is randomly initiated from the upper half of the plane and the system is subject to a sinusoidal control input u_t with different amplitude and period for each sequence.

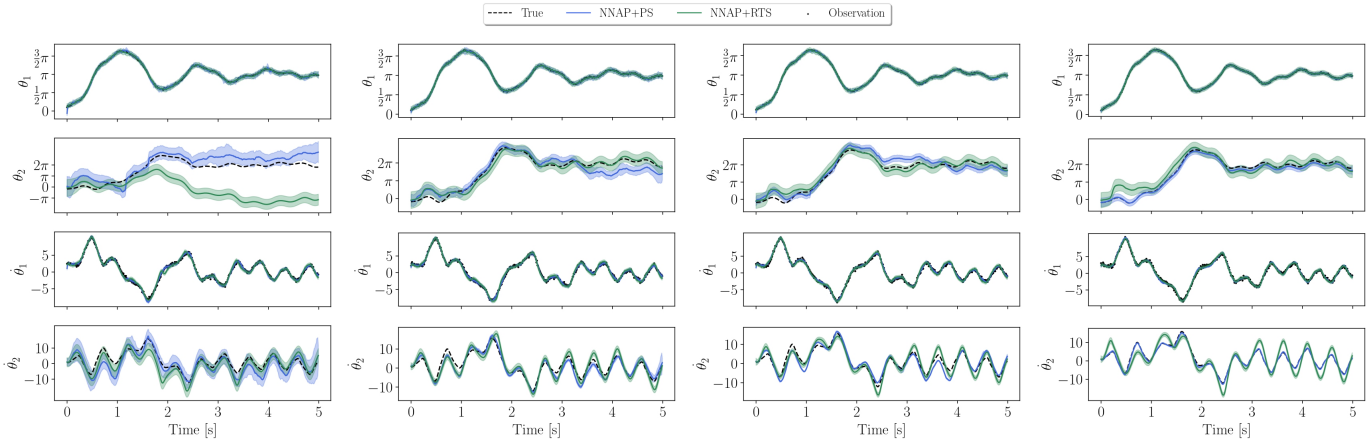
1) Learning behavior and filtering performance: First, we evaluate the application of the models in state estimation algorithms, such as the extended Kalman filter (EKF) or particle filter (PF). The EKF and PF distributions are computed during the E-step of the EM algorithm and thus offer valuable information on the training process of the RTS and PS implementations of the modeling framework, as described in Section III-C.

The evolution of the filtered distribution for one of the the test sequences at various stages in the training process for the data-driven models are depicted in Fig. 3. The data-driven models initiate with a random prior model. The filtered



(a) Epoch 1 (b) Epoch 10 (c) Epoch 25 (d) Epoch 50

Fig. 3: Filtered distributions during training of the data-driven models.



(a) Epoch 1 (b) Epoch 5 (c) Epoch 10 (d) Epoch 20

Fig. 4: Filtered distributions during training of the physics-informed models.

distribution of the RTS model quickly reflects the dynamics of the observed variables $\bar{x} = \{\theta_1, \dot{\theta}_1\}$, but fails to learn any dynamic relation for the unobserved variables $\underline{x} = \{\theta_2, \dot{\theta}_2\}$ throughout the learning process. After 50 epochs of training the RTS model does not longer improve, but is still unable to provide more accurate estimates of the observed state variables than the measurements themselves, indicating its failure to learn a good dynamics model. The filtered distribution of the PS model takes longer to converge for the observed variables, but eventually produce more accurate estimates of measured variables. The PS model also displays more dynamics for the unobserved variables, but these can not be directly related to the actual dynamics, due to non-uniqueness of the latent variable space, as discussed in Section II.

The physics-informed models benefit from the physical prior and already represent accurate complete state estimates at the start of the training process, as can be seen in Fig. 4. The neural network component should only learn the friction characteristics instead of the entire dynamics, resulting in a much more efficient training process. Both implementations have converged after 20 epochs. The PS model is capable of producing highly accurate complete state estimates x. The RTS model produces accurate estimates of the observed variables, but is not capable of reflecting the unobserved variables perfectly.

2) Prediction performance: We briefly demonstrate the multistep prediction performance of the converged models. Fig. 5 illustrates the prediction accuracy of the data-driven models. The RTS model displays poor prediction capabilities. The PS model displays acceptable predictive performance of the observed variables $\bar{x}$ for approximately $1s$ (100 samples), but deviates from the true trajectory beyond that.

The predictions of the physics-inspired models are depicted in Fig. 6. The physical backbone of these hybrid models result in significantly more accurate predictions for both observed and unobserved state variables. Again, the PS model outperforms the RTS model.

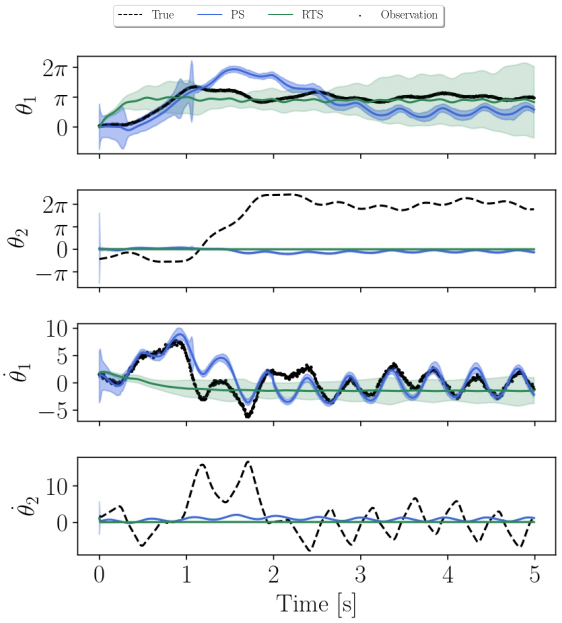


Fig. 5: Predicted distributions of the data-driven models.

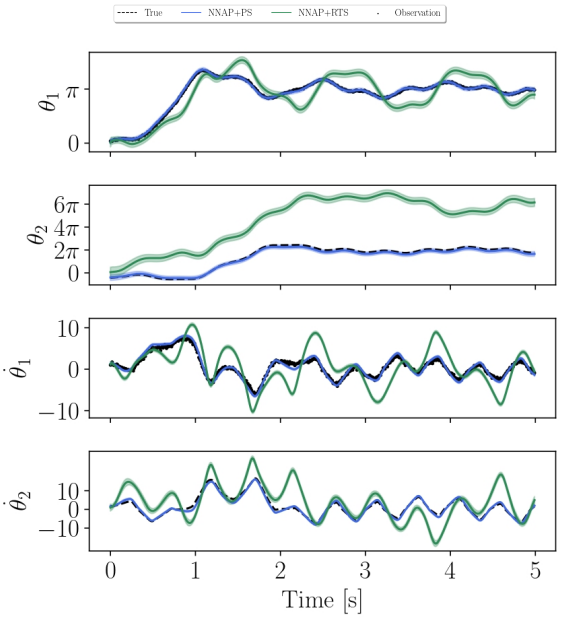


Fig. 6: Predicted distributions of the physics-informed models.

3) Physical insights: In addition to providing accurate predictions, the objective of the hybrid modeling approach is to gain deeper insights into the poorly understood components of the system dynamics. The modular design of the hybrid model allows to isolate the data-driven component h from the rest of the model, enabling a systematic analysis of the learned dynamics. The learned friction characteristics during different phases of the training process are shown in Fig. 7. These results can be aligned with the intermediate filtered distributions as depicted in Fig. 4. At the beginning of training the models are only able to reconstruct a vague approximation of the true friction characteristics, hence explaining the poor filtered distributions. At the end of training, the PS model has identified an accurate friction model, resulting in good filtering and prediction performance, as discussed above. The incapability of the RTS model in accurately capturing the highly nonlinear dynamics of the acrobot system can be attributed to the linearizations of the model utilized to obtain the required filtered and smoothed distributions for the EM training algorithm.

V. Conclusion

This paper presents a probabilistic system identification framework aimed at characterizing the behavior of mechatronic systems subject to unknown dynamics under partial observability. The proposed approach employs a hybrid model consisting of a partially identified state-space model and neural network layers, where the neural network provides an estimation of the unknown SSM substructures. The hybrid model is placed is a physics-informed prior over the latent space of a Hidden Markov Model. The variational inference framework allows to decompose the problem of learning the dynamical model of the underlying system dynamics from the mapping of this latent space to the observed variables. The methodology is validated on an acrobot system with unknown joint friction characteristics, where only observations from a single joint are available. The results demonstrate that the proposed approach is capable of accurately identifying the behavior of the entire system, despite incomplete state observations, as well as capturing the joint friction characteristics and the obtained hybrid models outperform pure data-driven models in terms of prediction accuracy.

Acknowledgment

This work was supported by the Flanders Make projects QUASIMO and HAIEM and the "Onderzoeksprogramma Artificiële Intelligentie (AI) Vlaanderen" programme.

References

[1] G. Kerschen, K. Worden, A.F. Vakakis and J.C. Golinval, "Past, present and future of nonlinear system identification in structural dynamics," *Mechanical systems and signal processing*, vol. 20, no. 3, pp. 505-592, 2006.

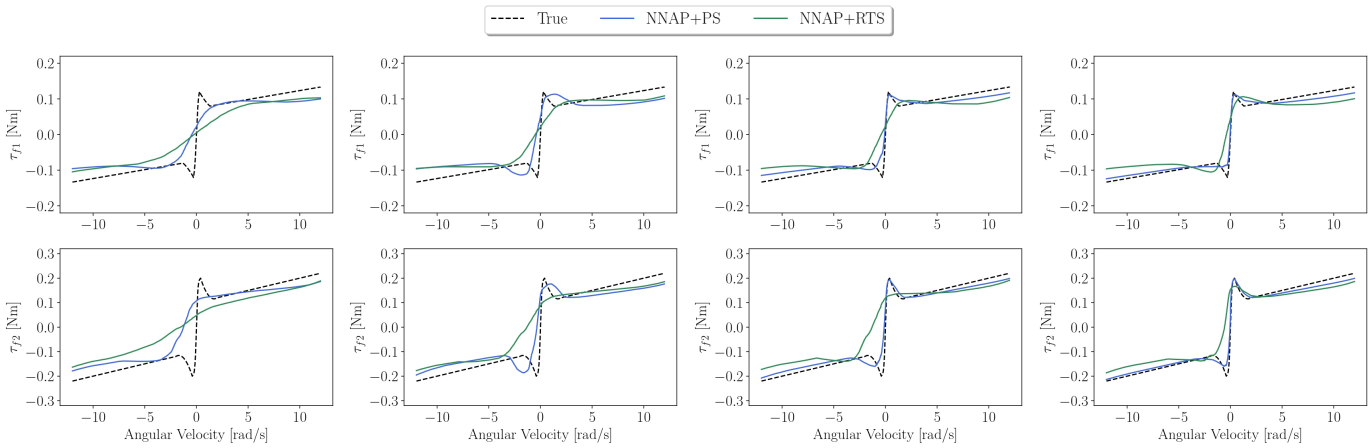
[2] S.L. Brunton and J.N. Kutz, "Data-driven science and engineering: Machine learning, dynamical systems, and control," *Cambridge University Press*, 2019.

[3] L. Ljung, C. Andersson, K. Tiels and T.L. Schön, "Deep learning and system identification", *IFAC-PapersOnLine 2020*, vol. 53, no. 2, pp.1175-1181, 2020.

[4] S. Rangan, G. Wolodkin and K. Poolla, "New results for Hammerstein system identification," *In Proceedings of the 34th IEEE conference on decision and control*, pp. 697-702, 1995.

[5] S. Chen, S. A. Billings, C. F. Cowan and P. M. Grant, "Practical identification of narmax models using radial basis functions," *International Journal of Control*, vol. 52, no. 6, pp. 1327-1350, 1990.

[6] Y. Wang, "A new concept using lstm neural networks for dynamic system identification," *IEEE 2017 American Control Conference*, pp. 5324-5329, 2017.



(a) Epoch 1 (b) Epoch 5 (c) Epoch 10 (d) Epoch 20

Fig. 7: Evolution of friction characteristics during training of the physics-informed models.

[7] M. Raissi, P. Perdikaris and G.E. Karniadakis, "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations," *Journal of Computational physics*, vol.378, pp. 686-707, 2019.

[8] M. Cranmer, S. Greydanus, S. Hoyer, P. Battaglia, D. Spergel and S. Ho, "Lagrangian neural networks," *arXiv preprint:2003.04630*, 2020.

[9] W. De Groote, E. Kikken, E. Hostens, S. Van Hoecke, and G. Crevecoeur, "Neural network augmented physics models for systems with partially unknown dynamics: Application to slider-crank mechanism," *IEEE/ASME Transactions on Mechatronics*, vol. 27, no. 1, 2021.

[10] F. Pernkopf, R. Peharz, and S. Tschiatschek, "Introduction to probabilistic graphical models," *Academic Press Library in Signal Processing, Elsevier*, vol. 1, pp. 989-1064, 2014.

[11] Särkkä, "Bayesian filtering and smoothing," *Cambridge University Press*, 2013.

[12] C. M. Bishop and N. M. Nasrabadi, "Pattern recognition and machine learning," *Springer*, vol. 4, 2006.

[13] K.P. Murphy, "Machine learning: a probabilistic perspective," *MIT press*, 2012.

[14] T.B. Schon, A. Wills, B. Ninnes, "System identification of nonlinear state-space models," *Automatica*, vol. 47, pp. 39-49, 2011.

[15] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, W. Zaremba, "Openai gym," *arXiv preprint arXiv:1606.01540*, 2016.

[16] R.S. Sutton, "Generalization in Reinforcement Learning: Successful Examples Using Sparse Coarse Coding," *Advances in Neural Information Processing Systems*, vol. 8, 1995.