

Kalman-Bucy-Informed Neural Network for System Identification

Tobias Nagel and Marco F. Huber

Abstract—Identifying parameters in a system of nonlinear, ordinary differential equations is vital for designing a robust controller. However, if the system is stochastic in its nature or if only noisy measurements are available, standard optimization algorithms for system identification usually fail. We present a new approach that combines the recent advances in physics-informed neural networks and the well-known achievements of Kalman filters in order to find parameters in a continuous-time system with noisy measurements. In doing so, our approach allows estimating the parameters together with the mean value and covariance matrix of the system’s state vector. We show that the method works for complex systems by identifying the parameters of a double pendulum.

I. INTRODUCTION

Controlling a dynamical system in a safe manner requires a model that describes the system properties precisely. Ordinary differential equations (ODEs) are often used to satisfy this requirement. Besides setting up the corresponding equation operators, it is also inevitable to identify the real-valued coefficients that define the characteristics of the system. Estimating these parameters by using measurements is termed as “inverse problem” and can be a difficult task, depending on the system’s complexity. This work presents a new method that is capable of identifying unknown parameters in a nonlinear ODE system, based on noisy measurements by using an extended Kalman-Bucy filter (EKBf) in a machine learning framework.

For linear systems, the subspace-based state space identification methods are well established. They aim at finding a linear state space model by using a regularized least-squares algorithm [7]. If the system comprises nonlinear behavior, the most straightforward solution approaches for parameter identification are standard minimization techniques like gradient-based [9] or gradient-free [1] methods. For a system with noisy measurements, the problem becomes even more difficult and requires incorporating stochastic moments in the optimization. Raue et. al. summarize their experiences of fitting measurements of biological systems to their corresponding ODE system by maximizing a log-likelihood function that comprises a normally distributed measurement noise [13]. However, these methods require a numerical solution of the ODE, repeatedly for each optimization iteration. Besides being very time consuming, this approach often fails

because of the system’s nonlinearity, the noise influence or an unstable behavior in the numerical solution [6].

A possibility to circumvent the problem with a machine learning approach is described in [16] by using a neural network to improve a Kalman filter system in order to obtain a better state estimate. Though, this does not give us the actual system parameter values but compensates for model errors. In 2017, Raissi et. al. presented how physics-informed neural networks (PINNs) can be trained by using modern automatic differentiation frameworks [12]. The approach utilizes deep neural networks to discover and solve nonlinear differential equation systems. This is achieved by training a neural network to represent an approximate solution to the differential equation. The method also enables a parameter search, by including the unknown parameters as additional network weights. The concept has been applied in numerous research fields, e.g., mechanics [8], thermodynamics [10] or in chemical reaction equations [4]. PINNs also enable the possibility to include stochastic behavior in the modeling process. Recently, this has been addressed by O’Leary et. al. who incorporate a mean value of the state and its covariance matrix in the framework, leveraging it to a stochastic physics-informed neural network (SPINN) [11]. The authors do so by propagating the first two central moments of a state variable through the known differential equation systems. Afterwards a neural network is trained in order to match these estimated central moments to measured ones. However, the authors do not address the problem of identifying parameters in the system. Another option is to use a Bayesian neural network (BNN) in a PINN environment which allows an embedding of uncertainty and, hence, the usage of stochastic differential equations. Yang et. al. use a BNN [18] to include noisy data into a partial differential equation problem in order to solve as well as identify the system [18]. However, BNNs are often not capable of achieving the same approximation accuracy as standard neural networks and are significantly more difficult to train.

In this paper, we present a new physics-informed machine learning approach that we call *Kalman-Bucy-informed neural network* (KBINN). A Kalman-Bucy filter incorporates two ODEs that describe the temporal evolution of the mean value and the covariance matrix of the system’s state. In our method, we include two neural networks that are implemented in a PINN framework in order to approximate a solution to the Kalman-Bucy equations. This allows an implicit identification of unknown system parameters by incorporating them into the network training. The rest of the paper is organized as follows: In Section II, we give a short mathematical formulation of the problem. Section III intro-

Tobias Nagel and Marco F. Huber are with the Fraunhofer Institute for Manufacturing Engineering and Automation IPA, Center for Cyber Cognitive Intelligence (CCI), 70569 Stuttgart, Germany {tobias.nagel, marco.huber}@ipa.fraunhofer.de

Marco F. Huber is with the Institute of Industrial Manufacturing and Management IFF, University of Stuttgart, 70569 Stuttgart, Germany marco.huber@ieee.org

duces the *extended Kalman-Bucy filter* (EKBF) and gives a short summary of neural networks. Section IV describes the KBINN method, followed by validation experiments in Section V. We discuss the strengths and limitations of our method in Section VI and close the paper with a conclusion in Section VII.

II. PROBLEM FORMULATION

The state space representation of a continuous-time, non-linear, dynamic and time-variant system of rank $n \in \mathbb{N}$ is defined by means of

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{w}(t), t, \boldsymbol{\theta}) \\ \mathbf{y}(t) &= \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), \mathbf{v}(t), t),\end{aligned}\quad (1)$$

where $f(\cdot)$ is the nonlinear ODE and $g(\cdot)$ is the measurement function. Both are assumed to be known, except for a set of unknown parameters. Furthermore, $\mathbf{x}(t) \in \mathbb{R}^n$ with $t \geq 0$ denotes the state vector, $\mathbf{u}(t) \in \mathbb{R}^p$ and $\mathbf{y}(t) \in \mathbb{R}^q$ denote the input and output signal with dimensions $p, q \in \mathbb{N}$, respectively. The vectors $\mathbf{w}(t) \in \mathbb{R}^n$ and $\mathbf{v}(t) \in \mathbb{R}^q$ denote white process noise and white measurement noise, respectively, which are both assumed to be zero-mean Gaussian with covariance matrices $\mathbf{Q}(t) \in \mathbb{R}^{n \times n}$ and $\mathbf{R}(t) \in \mathbb{R}^{q \times q}$, respectively. This induces the state $\mathbf{x}(t)$ to be a random variable as well. $\boldsymbol{\theta} \in \mathbb{R}^d$ denotes a vector of $d \in \mathbb{N}$ unknown parameters of the ODE system. If we acquire noisy measurements $\bar{\mathbf{y}}(t_i)$ of the system at N discrete time steps t_i with $i = 1, \dots, N$, the aim of our method is to find a parameter vector $\boldsymbol{\theta}^*$ that minimizes the objective function

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta}} (\mathbf{y}(t_i) - \bar{\mathbf{y}}(t_i))^2. \quad (2)$$

Standard minimization algorithms require to solve the ODE in Eq. (1) at each iteration numerically in order to minimize Eq. (2). As has been mentioned in Section I, these algorithms often fail due to the necessary small step width which is imposed by the system's nonlinearity or the noise.

III. PRELIMINARIES

In this section, we briefly introduce the EKBF, which is based on the well-known Kalman filter, applied to the nonlinear and continuous-time case. Afterwards, we give a short introduction to artificial neural networks.

A. Extended Kalman-Bucy Filter

It is obvious from Eq. (1) that the state \mathbf{x} is not directly measurable. To compensate for this problem, Rudolf E. Kalman introduced in 1960 the Kalman filter first for discrete-time, linear systems and a year later for the continuous-time case, together with Richard S. Bucy [5]. This concept allows estimating the mean value and the covariance matrix of the system's state. If the noise is assumed to be Gaussian, the first two central moments are sufficient to describe the state's probability distribution exactly. If a nonlinear system is considered, it is necessary to perform linearizations at each time step which leads to the EKBF. It

is composed of two initial value problems that comprises the state's estimated mean value

$$\begin{aligned}\dot{\hat{\mathbf{x}}}(t) &= \mathbf{f}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \mathbf{0}, t) \\ &+ \mathbf{K}(t) \cdot (\bar{\mathbf{y}}(t) - \mathbf{g}(\hat{\mathbf{x}}(t), \mathbf{u}(t), \mathbf{0}, t))\end{aligned}\quad (3)$$

with a known initial value $\hat{\mathbf{x}}(0) = \hat{\mathbf{x}}_0$ and a Kalman Gain

$$\mathbf{K}(t) = \hat{\mathbf{P}}(t) \cdot \hat{\mathbf{C}}^T(t) \cdot \hat{\mathbf{R}}^{-1}(t) \quad (4)$$

as well as its covariance matrix

$$\begin{aligned}\dot{\hat{\mathbf{P}}}(t) &= \hat{\mathbf{A}}(t)\hat{\mathbf{P}}(t) + \hat{\mathbf{P}}(t)\hat{\mathbf{A}}(t)^T \\ &- \hat{\mathbf{P}}(t)\hat{\mathbf{C}}^T(t)\hat{\mathbf{R}}^{-1}(t)\hat{\mathbf{C}}(t)\hat{\mathbf{P}}(t) + \hat{\mathbf{Q}}(t)\end{aligned}\quad (5)$$

with a known initial value $\hat{\mathbf{P}}(0) = \hat{\mathbf{P}}_0$. In Eq. (4) and (5), the involved matrices are obtained by linearization of Eq. (1) according to

$$\begin{aligned}\hat{\mathbf{A}}(t) &= \left. \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}, t)}{\partial \mathbf{x}(t)} \right|_{\wedge}, \quad \hat{\mathbf{C}}(t) = \left. \frac{\partial \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{v}, t)}{\partial \mathbf{x}(t)} \right|_{\wedge}, \\ \hat{\mathbf{G}}(t) &= \left. \frac{\partial \mathbf{f}(\mathbf{x}, \mathbf{u}, \mathbf{w}, t)}{\partial \mathbf{w}(t)} \right|_{\wedge}, \quad \hat{\mathbf{V}}(t) = \left. \frac{\partial \mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{v}, t)}{\partial \mathbf{v}(t)} \right|_{\wedge}.\end{aligned}\quad (6)$$

The \wedge -symbol denotes that the linearization is performed repeatedly for each new mean value $\hat{\mathbf{x}}(t)$. This also allows obtaining the noise covariance matrices

$$\begin{aligned}\hat{\mathbf{Q}}(t) &= \hat{\mathbf{G}}(t) \cdot \mathbf{Q}(t) \cdot \hat{\mathbf{G}}^T(t), \\ \hat{\mathbf{R}}(t) &= \hat{\mathbf{V}}(t) \cdot \mathbf{R}(t) \cdot \hat{\mathbf{V}}^T(t)\end{aligned}\quad (7)$$

in Eq. (4) and (5), respectively. Note that we omitted $\boldsymbol{\theta}$ from Eq. (1) for introducing the EKBF, since the filtering problem does not aim at identifying parameters in a system, but only enables us to calculate the state's mean value and its covariance matrix in a stochastic environment. The necessity of performing a linearization for every new state usually leads to a considerable computing effort which lowers the attractiveness of the EKBF for many applications. But the recent advances in automatic differentiation and its usage in neural networks make this method perfect for our purpose.

B. Neural Networks

A neural network is a type of machine learning algorithm that maps an input signal of rank ι to an output signal of rank κ by approximating a desired function. In its simplest form, it consists of at least three layers: the input layer comprises ι neurons and does not perform any transformations but only distributes the input signal to the successive layer. The hidden layer comprises a variable count of neurons. Each performs a weighted, nonlinear transformation by means of

$$o = \sigma \left(\sum_{k=1}^l i_k w_k + w_0 \right) = \sigma(\mathbf{i}^T \cdot \mathbf{w}). \quad (8)$$

Here, $\mathbf{i} = [1, i_1, \dots, i_l]^T \in \mathbb{R}^{l+1}$ denotes the output of the previous layer with l neurons and $\mathbf{w} = [w_0, w_1, \dots, w_l]^T \in \mathbb{R}^{l+1}$ denotes a weighting vector. If the neural network is used to approximate a nonlinear behavior, the activation function $\sigma(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ needs to be of nonlinear nature as

well. A neural network can comprise several successive hidden layers. The last layer, called output layer, is composed of κ neurons and also performs the calculation of Eq. (8). It can be shown that a neural network with at least one hidden layer but an arbitrary count of neurons and bounded, continuous, non-constant activation functions is able to approximate any given function [2].

IV. PHYSICS-INFORMED NEURAL NETWORKS FOR SYSTEM IDENTIFICATION

In this section, we first introduce the concept of PINNs in order to perform an identification of an ODE system. The idea of using PINNs as a tool for system identification has first been mentioned in [12]. We extend this approach by incorporating noise signals and take advantage of an EKBF in order to estimate a system's mean value and covariance matrix. This allows estimating parameters even if there are only noisy measurements available or if the model is imprecise.

A. Physics-Informed Neural Network

PINNs consider an initial value problem of the form

$$\begin{aligned}\dot{\mathbf{x}}(t) &= \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t), t, \boldsymbol{\theta}) \\ \mathbf{y}(t) &= \mathbf{g}(\mathbf{x}(t), \mathbf{u}(t), t)\end{aligned}\quad (9)$$

with equivalent variables to Eq. (1) and an initial value $\mathbf{x}(0) = \mathbf{x}_0$. Note that we do not consider any noise in this case. Let $\mathbf{x}(t)$ be an (approximate) solution to Eq. (9), depending on the existence. The idea is to train a neural network $\pi(t, \mathbf{W}): \mathbb{R}^{p+1} \rightarrow \mathbb{R}^n$ that approximates the solution $\mathbf{x}(t)$ by adjusting the network weights which are summarized in a matrix \mathbf{W} . Parallel to the training, a vector $\hat{\boldsymbol{\theta}} \in \mathbb{R}^d$, that comprises an estimate of the original parameters $\boldsymbol{\theta}$, becomes optimized. We do so by setting up a loss function according to

$$J_{\text{PINN}} = \sum_{i=1}^N \left(\frac{d}{dt} \pi(t_i) - \mathbf{f}(\pi(t_i), \mathbf{u}, t_i, \hat{\boldsymbol{\theta}}) \right)^2 + (g(\pi(t_i), \mathbf{u}, t_i) - \bar{\mathbf{y}}_i)^2, \quad (10)$$

with $\bar{\mathbf{y}}_i$ denoting a measurement of the system at time t_i . For the sake of readability, we abbreviated the full argument list of $\pi(t, \mathbf{W})$. Subsequently, we execute a minimization algorithm in order to find the optimal values for the weight matrix \mathbf{W} and the estimated parameter vector $\hat{\boldsymbol{\theta}}$. If successful, minimizing the loss function in Eq. (10) yields the true parameters $\boldsymbol{\theta}$ and a neural network that features the same derivative and initial value as $\mathbf{x}(t)$.

B. Kalman-Bucy-Informed Neural Network

In our approach, we consider a system with noise as in Eq. (1). The key idea is to implement a PINN that approximates the solution of the initial value problems from Eq. (3) and Eq. (5). This results in a method that allows estimating the parameters of a stochastic ODE system given noisy measurements and a possibly erroneous model. We do so by constructing two neural networks that receive a

time signal as input and predict the state's mean value $\hat{\mathbf{x}}(t)$ as well as its covariance matrix $\hat{\mathbf{P}}(t)$, respectively. Fig. 1 depicts a sketch of our approach. The *mean network* $\xi(t, \mathbf{W}_x)$ with weight matrix \mathbf{W}_x approximates the state's mean value $\hat{\mathbf{x}}(t)$. Its counterpart for the covariance matrix is the *covariance network* $\psi(t, \mathbf{W}_P)$ with a weighting matrix \mathbf{W}_P . The vector $\hat{\boldsymbol{\theta}}$ contains the estimated parameters. We optimize both weighting matrices \mathbf{W}_x and \mathbf{W}_P , together with $\hat{\boldsymbol{\theta}}$ simultaneously by minimizing a loss function J that is composed of three weighted terms according to

$$J = \sum_{i=1}^N (\alpha_1 L_{1,i} + \alpha_2 L_{2,i} + \alpha_3 L_{3,i}) \quad (11)$$

with weights $\alpha_j > 0$, $j \in \{1, 2, 3\}$ that can be chosen arbitrarily. To improve readability, we again drop the weighting matrices from both neural network functions. The first term $L_{1,i}$ in (11) is given by

$$L_{1,i} = \|\xi(0) - \mathbf{x}_0\|_2 + \|\dot{\xi}(t_i) - \Xi(\xi, t_i)\|_2 \quad (12)$$

with

$$\begin{aligned}\Xi(\xi, t_i) &= \mathbf{f}(\xi(t_i), \mathbf{u}, \mathbf{0}, t_i, \hat{\boldsymbol{\theta}}) \\ &\quad + \mathbf{K}(t_i) \cdot (\bar{\mathbf{y}}_i - \hat{\mathbf{y}}(t_i)).\end{aligned}\quad (13)$$

and $\mathbf{K}(t_i) = \psi(t_i) \cdot \hat{\mathbf{C}}^T(t_i) \cdot \hat{\mathbf{R}}^{-1}(t_i)$ as a Kalman gain. $\hat{\mathbf{y}}(t_i) = \mathbf{g}(\xi(t_i), \mathbf{u}, \mathbf{0}, t_i)$ describes the estimated output and $\|\cdot\|_2$ denotes the Euclidean norm. Hence, $L_{1,i}$ utilizes Eq. (3) and (4) by replacing $\hat{\mathbf{x}}(t)$ and $\hat{\mathbf{P}}(t)$ with the corresponding neural networks. It converges to 0 if $\dot{\xi}(t)$ and $\xi(0)$ approach $\dot{\hat{\mathbf{x}}}(t)$ and \mathbf{x}_0 . The second term $L_{2,i}$ is an equivalent to $L_{1,i}$, applied to the covariance matrix. It is defined by means of

$$L_{2,i} = \|\psi(0) - \hat{\mathbf{P}}_0\|_F + \|\dot{\psi}(t_i) - \Psi(\psi, t_i)\|_F \quad (14)$$

with the Frobenius norm $\|\cdot\|_F$ and

$$\begin{aligned}\Psi(\psi, t_i) &= \hat{\mathbf{A}}(t_i)\psi(t_i) + \psi(t_i)\hat{\mathbf{A}}(t_i)^T \\ &\quad - \psi(t_i)\hat{\mathbf{C}}^T(t_i)\hat{\mathbf{R}}^{-1}(t_i)\hat{\mathbf{C}}(t_i)\psi(t_i) + \hat{\mathbf{Q}}(t_i),\end{aligned}\quad (15)$$

which utilizes Eq. (5) where we substitute the covariance matrix $\hat{\mathbf{P}}$ by the corresponding neural network $\psi(t_i)$. The third term $L_{3,i}$ aims at keeping the original measurement values close to the uncertainty interval that is spanned by the mean and its covariance by incorporating a maximum likelihood estimator. Since we assume the measurement noise to be additive and Gaussian, we use a normal distribution by means of

$$p_j(\bar{\mathbf{y}}_j(t_i); \xi(t_i), \psi(t_i)) = \mathcal{N}(\bar{\mathbf{y}}_j(t_i); \boldsymbol{\mu}_j(t_i), \boldsymbol{\sigma}_j^2(t_i)) \quad (16)$$

with $j = 1, \dots, q$ for each output dimension. We calculate the mean value by means of $\boldsymbol{\mu}(t_i) = \text{mean}(\mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{v}, t_i))$ and the variance according to $\boldsymbol{\sigma}^2(t_i) = \text{var}(\mathbf{g}(\mathbf{x}, \mathbf{u}, \mathbf{v}, t_i))$ with $\mathbf{x}(t_i) \sim \mathcal{N}(\mathbf{x}(t_i); \xi(t_i), \psi(t_i))$. Both, the mean and the variance can be calculated in closed form if the measurement function exhibits a special type, e.g., linear, polynomial or trigonometric [3]. In other cases, approximations are

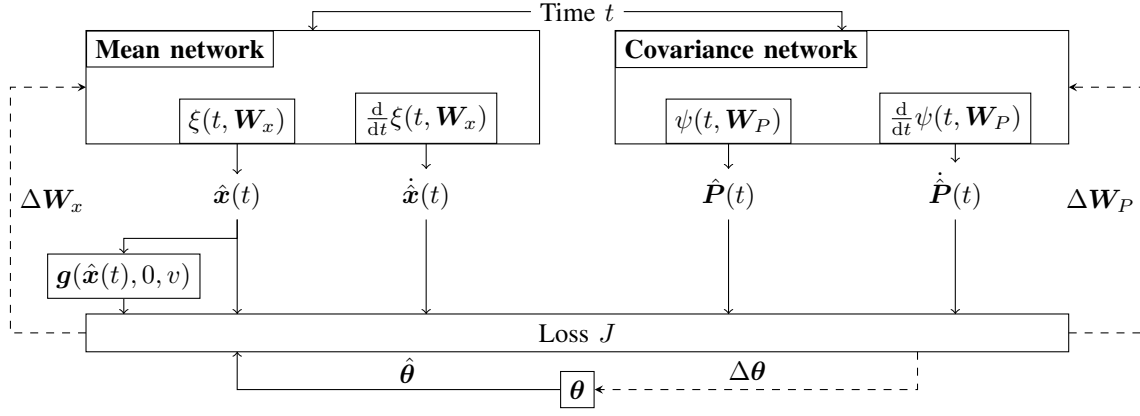


Fig. 1: Sketch of the KBINN. Two neural networks are trained in order to approximate a PINN-based solution for the EKBF. The output of both networks as well as the estimated parameters $\hat{\boldsymbol{\theta}}$ are fed into the loss function J . Changes of the training parameters are indicated by dashed lines and a Δ in front of the variable.

necessary, e.g. sampling or numerical integration. We then obtain the loss $L_{3,i}$ by using the negative natural logarithm

$$L_{3,i} = - \sum_{j=1}^q \log(p_j(\bar{\mathbf{y}}_j(t_i); \xi(t_i), \psi(t_i))) . \quad (17)$$

C. Implementation

In this section, we explain the implementation details of our approach. For both neural networks, all activation functions need to be at least twice continuous differentiable, because we use the network's temporal derivative and a gradient-based weight optimization. The neural network $\xi(t, \mathbf{W}_x)$ possesses one input neuron and n output neurons, with n denoting the system's rank. The neuron and layer count in the hidden layers depends on the particular use case. We choose the hyperbolic tangent as activation function for all hidden neurons. The output layer comprises a linear activation function. The covariance network $\psi(t, \mathbf{W}_P)$ is slightly more complex. While the architecture of input and hidden layers are identical to $\xi(t, \mathbf{W}_x)$, we must ensure that the network's output comprises a symmetric positive semi-definite matrix of rank n . We do so by constructing $(n+1) \cdot n/2$ output neurons and build an upper triangular matrix. Afterwards, we multiply the upper triangular matrix with its transposed version.

After setting up the neural networks as well as $\hat{\boldsymbol{\theta}}$, the weights \mathbf{W}_x and \mathbf{W}_P as well as the parameters are optimized simultaneously by minimizing the loss function of Eq. (11). Therefore we use the backpropagation algorithm [14], which utilizes a gradient-based optimization to efficiently fit the weights of a neural network.

V. EXPERIMENT: IDENTIFICATION OF A DOUBLE PENDULUM SYSTEM

We validate the proposed KBINN by solving the inverse problem of a double pendulum system. For comparison, we employ MATLAB's system identification toolbox, which is dedicated to identify parameters by using system measurements. The function `nlgreyest` allows estimating model

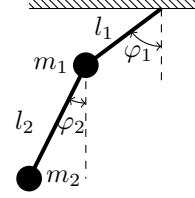


Fig. 2: Sketch of a double pendulum.

parameters of a nonlinear ODE by iteratively solving it numerically and using a gradient-based optimization algorithm.

A double pendulum is composed of a conventional pendulum with mass m_1 and rod length l_1 , extended by another one with values m_2 and l_2 . Generally, the system's parameters are hard to identify because of the chaotic motion behavior that lets small deviations in the initial position cause substantial trajectory differences. Fig. 2 depicts a sketch of the system. The state is defined by the pendulum angles and their angular velocities $\mathbf{x} = [\varphi_1 \ \dot{\varphi}_1 \ \varphi_2 \ \dot{\varphi}_2]^T$. We define $c_{12} = \cos(\varphi_1 - \varphi_2)$, $s_{12} = \sin(\varphi_1 - \varphi_2)$ and $M = \frac{m_2}{m_1 + m_2}$. The corresponding ODE is given by [15]

$$\begin{aligned} \ddot{\varphi}_1 &= -M \frac{l_2}{l_1} (\ddot{\varphi}_2 c_{12} + \dot{\varphi}_2^2 s_{12}) - \frac{g}{l_1} \sin(\varphi_1) \\ \ddot{\varphi}_2 &= -\frac{l_1}{l_2} (\ddot{\varphi}_1 c_{12} - \dot{\varphi}_1^2 s_{12}) - \frac{g}{l_2} \sin(\varphi_2) \end{aligned} \quad (18)$$

with the gravitational constant $g = 9.81 \text{ m/s}^2$. All state variables are time dependent. Our objective is to find the parameter vector $\boldsymbol{\theta} = [l_1 \ l_2 \ M]^T$ that comprises the pendulum's lengths and masses. Since the KBINN is capable of dealing with small modeling errors, we exacerbate the problem by adding a damping term to Eq. 18 which yields

$$\begin{aligned} \ddot{\varphi}_1 &= -M \frac{l_2}{l_1} (\ddot{\varphi}_2 c_{12} + \dot{\varphi}_2^2 s_{12}) - \frac{g}{l_1} \sin(\varphi_1) - 0.05 \dot{\varphi}_1 \\ \ddot{\varphi}_2 &= -\frac{l_1}{l_2} (\ddot{\varphi}_1 c_{12} - \dot{\varphi}_1^2 s_{12}) - \frac{g}{l_2} \sin(\varphi_2) - 0.05 \dot{\varphi}_2 . \end{aligned} \quad (19)$$

We assume all state variables in \mathbf{x} to be measurable after adding a noise according to $\mathbf{y}(t) = \mathbf{x}(t) + \mathbf{v}(t)$ with

TABLE I: Mean absolute errors and standard deviations of the identified parameters over ten identification runs.

| Method | Δl_1 | Δl_2 | ΔM |
|-----------|-----------------|-----------------|-----------------|
| KBINN | 0.03 ± 0.11 | 0.02 ± 0.05 | 0.08 ± 0.24 |
| nlgreyest | 0.49 ± 0.28 | 0.30 ± 0.25 | 0.30 ± 0.17 |

$v(t)$ being zero-mean Gaussian that comprises a covariance matrix of $R(t) = I \cdot 0.25$ with the identity matrix I of rank $n = 4$. Without assuming all states to be measurable, the parameter identification turns out to be too error-prone. Nevertheless, we successfully tested the algorithm on less complex systems (e.g., an elastic pendulum) without a fully observable state vector. We use Eq. (19) to simulate system trajectories by solving the ODE numerically and add noise as mentioned above to acquire $\bar{y}(t_i)$ with $N = 3,000$ in a time interval of 3 s which yields a sampling frequency of 1 kHz.

To prove the functionality of our algorithm statistically, we perform ten system identification runs by sampling a new parameter vector θ from a uniform distribution $\mathcal{U}_{[0,1]}$ for each trial. Additionally, we generate new initial state vectors $x_0 = [\varphi_{10} \ \dot{\varphi}_{10} \ \varphi_{20} \ \dot{\varphi}_{20}]$ by sampling $\varphi_{10}, \varphi_{20}$ from another uniform distribution $\mathcal{U}_{[-\pi/4, \pi/4]}$. The initial angular velocities are kept to 0. We also use new noise samples and new initial weight matrices for each cycle. We measure the success of the method by calculating the absolute error between the identified parameters. Table I shows the mean absolute errors of all identification runs for each parameter Δl_1 , Δl_2 and ΔM . We can see that on average, the KBINN leads to a more precise estimate, compared to the nlgreyest-method. Additionally, Fig. 3(a) shows a box plot of the absolute errors. The ten identification runs depict one outlier sample which is clearly visible in the boxplot. The remaining nine trials all lead to an estimation error that is close to 0. The nlgreyest-function on the other hand leads to estimates that differ heavily from the original values.

We observed that the KBINN-based identification method requires obtaining measurements at a high sampling frequency. This is caused by the linearizations of the EKBF in Eq. (6). A higher sampling frequency comes along with more data which requires more computing power and time. Thus, we want to find out the minimal sampling frequency that still allows a good parameter identification. We obtain five parameter and initial state scenarios by sampling from the same uniform distributions, as we used before. Afterwards, we identify the parameters based on different sampling frequencies which start at 3 Hz and increase up to 1 kHz. We then calculate the mean value of Δl_1 , Δl_2 and ΔM to acquire the box plot shown in Fig. 3(b). As expected, a higher sampling frequency leads to a smaller error, on average. For the double pendulum, at least 100 Hz are required to get a steady deviation of less than one percent.

In the following, we take one parameter scenario with $\theta = [0.6 \text{ m} \ 0.9 \text{ m} \ 0.57]$ as a showcase on how the KBINN networks behave. The parameter $M = 0.57$ has been acquired by the masses $m_1 = 0.3 \text{ kg}$ and $m_2 = 0.4 \text{ kg}$.

Fig. 4 shows the noisy trajectories of all states in black. We also plot the predictions of the KBINN’s mean and covariance networks $\xi(t)$ and $\psi(t)$. We can observe that the noisy trajectories stay mostly within the area that is spanned by the KBINN’s predicted mean and variance.

VI. DISCUSSION

In this section, we discuss the advantages and limitations of our approach. As we showed in Section V, we were always able to solve an inverse problem, although only noisy measurement data is given and the used ODE model contains modeling errors. In this paper, we only showed the identification results of three parameters, but further experiments indicate that our method scales really well with an increasing parameter count (ODEs with up to eight parameters were examined). We also want to stress that the identification is continuous-time as a whole and, thus, also works for systems that become unstable if discretized.

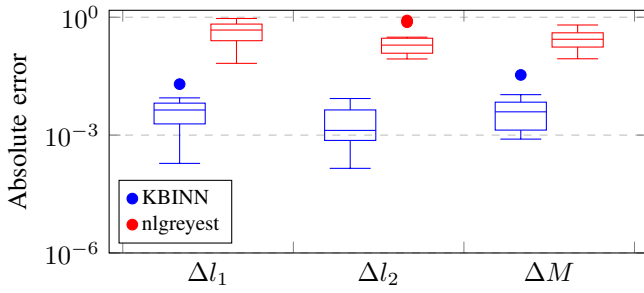
However, the KBINN method also has some limitations. The most severe one is the architecture choice of the neural networks. The only possibility to find a good architecture is by trial and error or by using methods of neural architecture search [17]. Both, too few and too many neurons will lead to an unsuccessful identification. Furthermore, we sometimes observe a bad fit for the KBINN networks in the last measurement points. The reason is a bad derivative estimate in the last time step that cannot influence the further trend anymore and, thus, does not affect the loss function heavily. We successfully experimented to solve this by extrapolating the measurements with a constant slope. Fortunately, the KBINN indicates this false prediction by an increased variance. As we mentioned in Section V, we also sometimes observe an outlier after several successful identification runs. We suspect that this occurs when the backpropagation algorithm is stuck in a local minimum during training. A possible work-around is to restart the training with a bigger step width. In our cases, we could correct the outliers by simply restarting the training with new initial weight matrices W_x and W_P . Note that we did not perform this correction in our analysis in Section V.

VII. CONCLUSIONS AND OUTLOOK

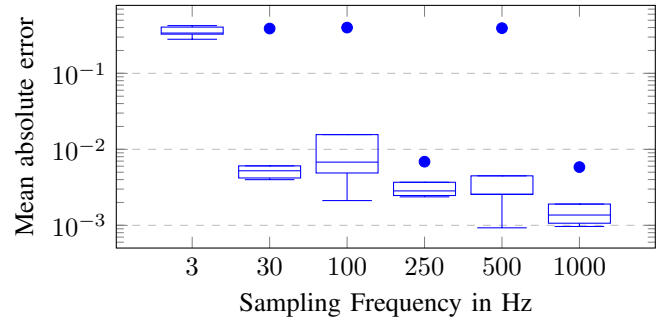
We introduced a new method to identify unknown parameters in an ODE system with noisy measurements and imprecise models. The method uses recent advances in physics-informed machine learning by incorporating an extended Kalman-Bucy filter. We prove the performance of our method on a double pendulum system and compare the results to the corresponding MATLAB-function. Future work is devoted to not only estimate the ODE’s parameters but the entire structure of the equations.

VIII. ACKNOWLEDGEMENTS

This work was partially supported by the Baden-Wuerttemberg Ministry for Economic Affairs, Labour and Tourism (Project KI-Fortschrittszentrum “Lernende Systeme und Kognitive Robotik”) and the German Federal Ministry of Education and Research (Project Degrad-EL³-Q).

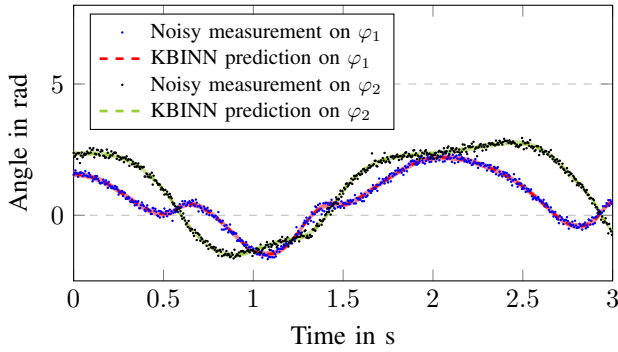


(a) Box plots of ten identification runs. The y-axis shows the absolute error between the original and identified parameter in accordance to its unit.

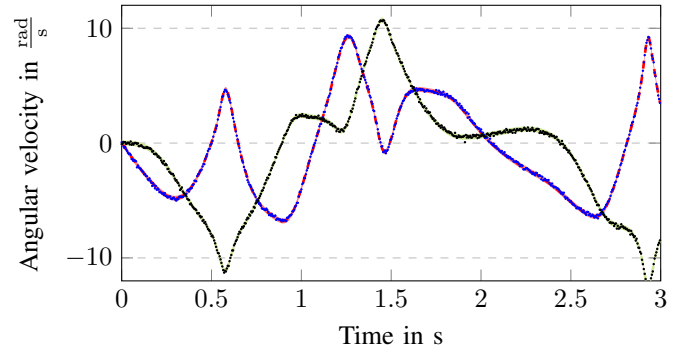


(b) Box plots of five identification runs per sampling frequency. Each data point describes the average error over all parameters Δl_1 , Δl_2 , ΔM .

Fig. 3: Box plots of several identification runs to highlight the performance regarding (a) the estimate error for each parameter and (b) the performance as a function of the sampling frequency.



(a) Trajectories and predictions of φ_1 and φ_2 .



(b) Trajectories and predictions of $\dot{\varphi}_1$ and $\dot{\varphi}_2$ with equivalent legend.

Fig. 4: Plots of the simulated, noisy trajectories and the predictions of both neural networks $\xi(t)$ and $\psi(t)$ for the double pendulum. The uncertainty is depicted as colored 2σ -interval, but it is barely visible as the values are close to zero.

REFERENCES

- [1] Devin Akman, Olcay Akman, and Elsa Schaefer. Parameter estimation in ordinary differential equations modeling via particle swarm optimization. *Journal of Applied Mathematics*, 2018:1–9, 2018.
- [2] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366, 1989.
- [3] Marco Huber. *Nonlinear Gaussian Filtering: Theory, Algorithms, and Applications*. KIT Scientific Publishing, Karlsruhe, Mar 2015.
- [4] Weiqi Ji, Weilun Qiu, Zhiyu Shi, Shaowu Pan, and Sili Deng. Stiff-pinn: Physics-informed neural network for stiff chemical kinetics. *The Journal of Physical Chemistry A*, 125(36):8098–8106, 2021.
- [5] R. E. Kalman and R. S. Bucy. New results in linear filtering and prediction theory. *Journal of Basic Engineering*, 83(1):95–108, 1961.
- [6] Hua Liang and Hulin Wu. Parameter estimation for differential equation models using a framework of measurement error in regression models. *Journal of the American Statistical Association*, 103(484):1570–1583, 2008.
- [7] Lennart Ljung. Perspectives on system identification. *Annual Reviews in Control*, 34(1):1–12, 2010.
- [8] George S. Misyris, Andreas Venzke, and Spyros Chatzivasileiadis. Physics-informed neural networks for power systems. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*. IEEE, 2020.
- [9] Eshan D. Mitra and William S. Hlavacek. Parameter estimation and uncertainty quantification for systems biology models. *Current opinion in systems biology*, 18:9–18, 2019.
- [10] Sina Amini Niaki, Ehsan Haghighat, Trevor Campbell, Anoush Pour-sartip, and Reza Vaziri. Physics-informed neural network for modelling the thermochemical curing process of composite-tool systems during manufacture. *Computer Methods in Applied Mechanics and Engineering*, 384(3–4):113959, 2021.
- [11] Jared O’Leary, Joel A. Paulson, and Ali Mesbah. Stochastic physics-informed neural networks (spinn): A moment-matching framework for learning hidden physics within stochastic differential equations, 2021.
- [12] Maziar Raissi, Paris Perdikaris, and George Em Karniadakis. Physics informed deep learning (part i): Data-driven solutions of nonlinear partial differential equations, 2017.
- [13] Andreas Raue, Marcel Schilling, Julie Bachmann, Andrew Matteson, Max Schelker, Max Schelke, Daniel Kaschek, Sabine Hug, Clemens Kreutz, Brian D. Harms, Fabian J. Theis, Ursula Klingmüller, and Jens Timmer. Lessons learned from quantitative dynamical modeling in systems biology. *PloS one*, 8(9):e74335, 2013.
- [14] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. Learning representations by back-propagating errors. *Nature*, 323(6088):533–536, 1986.
- [15] Tomasz Stachowiak and Toshio Okada. A numerical analysis of chaos in the double pendulum. *Chaos, Solitons & Fractals*, 2006.
- [16] S.C. Stubberud, R.N. Lobbia, and M. Owen. An adaptive extended kalman filter using artificial neural networks. In *Proceedings of 1995 34th IEEE Conference on Decision and Control*, volume 2, pages 1852–1856 vol.2, 1995.
- [17] Colin White, Willie Neiswanger, and Yash Savani. Bananas: Bayesian optimization with neural architectures for neural architecture search. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, pages 10293–10301, 2021.
- [18] Liu Yang, Xuhui Meng, and George Em Karniadakis. B-pinns: Bayesian physics-informed neural networks for forward and inverse pde problems with noisy data. *Journal of Computational Physics*, 425:109913, 2021.