

Learning Flow Functions from Data with Applications to Nonlinear Oscillators¹

Miguel Aguiar¹, Amritam Das² and Karl H. Johansson¹

¹*Division of Decision and Control Systems and Digital Futures,
KTH Royal Institute of Technology, Stockholm, Sweden and*

²*Control Systems Group, Dept. of Electrical Engineering, Eindhoven
University of Technology, Eindhoven, Netherlands.*

{aguiar,kallej}@kth.se and am.das@tue.nl

Abstract: We describe a recurrent neural network (RNN) based architecture to learn the flow function of a causal, time-invariant and continuous-time control system from trajectory data. By restricting the class of control inputs to piecewise constant functions, we show that learning the flow function is equivalent to learning the input-to-state map of a discrete-time dynamical system. This motivates the use of an RNN together with encoder and decoder networks which map the state of the system to the hidden state of the RNN and back. We show that the proposed architecture is able to approximate the flow function by exploiting the system's causality and time-invariance. The output of the learned flow function model can be queried at any time instant. We experimentally validate the proposed method using models of the Van der Pol and FitzHugh-Nagumo oscillators. In both cases, the results demonstrate that the architecture is able to closely reproduce the trajectories of these two systems. For the Van der Pol oscillator, we further show that the trained model generalises to the system's response with a prolonged prediction time horizon as well as control inputs outside the training distribution. For the FitzHugh-Nagumo oscillator, we show that the model accurately captures the input-dependent phenomena of excitability.

Copyright © 2023 The Authors. This is an open access article under the CC BY-NC-ND license (<https://creativecommons.org/licenses/by-nc-nd/4.0/>)

Keywords: Oscillator, Learning, Recurrent Neural Network, Excitability

1. INTRODUCTION

Models play a vital role in designing control systems. For instance, in receding horizon control (Maciejowski (2002)), the model is used to predict the future evolution of the state variables and acts as a constraint in the formulation of the optimal control problem. With increasing complexity, the curse of dimensionality limits the usefulness of standard first-principle models. This limitation has motivated research on data-driven approximation of such physical models. Besides fast predictions for arbitrary initial conditions, another advantage of many data-driven methods in designing control systems is the efficient computation of the gradients of the model with respect to initial conditions, parameters or input signals. To leverage these advantages, in (Li et al., 2021; Geneva and Zabaras, 2022; Lu et al., 2021; Kissas et al., 2022), the focus is on learning the map from initial conditions, parameters and inputs to the solution of a differential equation.

The problem of approximating physical models from data is also tackled in system identification, see Schoukens and Ljung (2019) for an overview. Forgone and Piga (2021) have proposed methods for identifying dynamics

of continuous-time control systems using neural Ordinary Differential Equations (ODEs). However, as the dynamics correspond to the time derivative of the flow, the neural ODEs must be integrated through an ODE solver to obtain the system trajectories, representing an extra computational burden both for prediction and for computing gradients. Furthermore, errors in the learned dynamics will accumulate over time when the dynamics are integrated, and the error in the simulated trajectory can become unbounded.

To directly learn the flow function of an autonomous dynamical system, Biloš et al. (2021) has proposed an alternative to neural ODEs that avoids the step of using an ODE solver and allows for faster prediction. This motivates the search for a corresponding learning scheme for controlled dynamical systems where inputs are present. However, this is a harder problem since the domain of the flow of a continuous-time control system is infinite-dimensional.

Hanson and Raginsky (2020) have shown that continuous-time recurrent neural networks are universal approximators for flow functions of stable continuous-time dynamical systems, where the approximation quality is uniform over time. But the question of whether learning such a model from data is feasible in practice is to the best of our knowledge open.

¹ © 2023 the authors. This work has been accepted to IFAC for publication under a Creative Commons Licence CC-BY-NC-ND. An early version of this work was accepted for (non-archival) poster presentation to the NeurIPS 2022 workshop *The Symbiosis of Deep Learning and Differential Equations*.

Our main contributions are a neural network-based architecture for learning flow functions of controlled dynamical systems in continuous-time and a detailed experimental demonstration of the performance and generalisation capabilities of the proposed architecture in predicting the responses of nonlinear oscillators. We provide a mathematical formulation of the problem of learning the flow function of a control system, showing that it can be reduced to a tractable optimisation problem. The inputs are restricted to the class of piecewise constant functions, which are practically relevant since digital controllers typically produce such control inputs. Leveraging the causality and time-invariance properties of the considered class of systems, we show that the continuous time flow function can be efficiently approximated by a discrete-time recurrent neural network-based architecture. We demonstrate the capabilities of the proposed architecture in predicting the input-dependent response of nonlinear oscillators such as the Van der Pol and FitzHugh-Nagumo oscillators.

The organisation of the paper is as follows. After a detailed description of the considered class of control systems and formulating the learning problem in Section 2, in Section 3 we present the proposed architecture to learn the flow function of a control system. In Section 4, the proposed methodology is experimentally evaluated to predict the periodic response of the non-autonomous Van der Pol oscillator. In Section 5, the proposed methodology is evaluated to predict the occurrence of excitable behaviour in the FitzHugh-Nagumo oscillator. Finally, in Section 6 we provide some concluding remarks and future directions for further research.

2. PROBLEM FORMULATION

2.1 Considered class of control systems and flow functions

A control system Σ consists of the following quadruple (see Sontag (1998), pp. 26)

$$\Sigma = (\mathcal{T}, \mathcal{X}, \mathbb{U}, \varphi), \quad (1)$$

describing the evolution of state-variables of the dynamical system over a time interval \mathcal{T} depending on its initial condition $x \in \mathcal{X}$ and input $u \in \mathbb{U}$, where \mathbb{U} is a set of functions $u : \mathcal{T} \rightarrow \mathcal{U}$. The *flow*, dictating this evolution, is defined as a mapping $\varphi : \mathcal{T} \times \mathcal{X} \times \mathbb{U} \rightarrow \mathcal{X}$.

We assume that Σ is time-invariant and finite dimensional. In particular, $\mathcal{T} \subseteq \mathbb{R}_{\geq 0}$, $\mathcal{X} \subset \mathbb{R}^n$ and $\mathcal{U} \subset \mathbb{R}^m$. We also assume that \mathbb{U} is the set of piecewise constant controls² of period $\Delta > 0$. In other words, given a sequence $\{u_k\}_{k=1}^{\infty}$ with $u_k \in \mathcal{U}$, the control input u is defined by

$$u(t) = u_k, \quad (k-1)\Delta \leq t < k\Delta, \quad k \in \mathbb{N}. \quad (2)$$

We will exploit two properties of the flow: *causality* which implies that the flow at time $T \geq 0$, $\varphi(T, x, u)$ depends only on the values of $u(t)$ for $0 \leq t < T$, and *continuity* in the sense that $t \mapsto \varphi(t, x, u)$ is continuous for each x, u .

As an example, we can consider the flow function generated by a system of ordinary differential equations $\dot{\xi}(t) = f(\xi(t), u(t))$, $t \geq 0$ with initial condition x where u is generated by a digital controller.

² The approach can be generalised to any class of input signals that admit a finite-dimensional causal parameterisation.

2.2 Mathematical formulation of the learning problem

We are interested in learning the flow from data on a time interval $[0, T]$ with $T > \Delta$. The input signals u and initial conditions x of interest are assumed to be drawn from probability distributions P_u on \mathbb{U} and P_x on \mathcal{X} , respectively.

Given an hypothesis class $\mathcal{H} \subset \{\hat{\varphi} : \mathbb{R}_{\geq 0} \times \mathcal{X} \times \mathbb{U} \rightarrow \mathcal{X}\}$, we define for $\hat{\varphi} \in \mathcal{H}$ the loss function

$$\ell_T(\hat{\varphi}) := \mathbf{E} \left[\frac{1}{T} \int_0^T \|\hat{\varphi}(t, X, U) - \varphi(t, X, U)\|^2 dt \right], \quad (3)$$

(the average squared prediction error over $[0, T]$), where $X \sim P_x$ and $U \sim P_u$ are independent random variables. The problem of finding the best approximation to the true flow φ amounts to minimising ℓ_T over \mathcal{H} .

In practice, the data consists of discrete-time samples from N different trajectories:

$$\xi_k^i = \varphi(t_k^i, x^i, u^i) + v_k^i, \quad k = 1, \dots, K, \quad i = 1, \dots, N, \quad (4)$$

where K is the number of samples of each trajectory, $t_k^i \in [0, T]$ is an increasing sequence of time samples, v_k^i is measurement noise, and x^i, u^i are sampled i.i.d. from P_x and P_u . Thus, in order to learn φ from the data (4), we define the empirical loss function

$$\hat{\ell}_T(\hat{\varphi}) := \frac{1}{N} \sum_{i=1}^N \frac{1}{K} \sum_{k=1}^K \|\xi_k^i - \hat{\varphi}(t_k^i, x^i, u^i)\|^2 \quad (5)$$

and search for a minimiser of $\hat{\ell}_T$ in \mathcal{H} .

The objective of this paper is to define an hypothesis space \mathcal{H} which renders the above problem tractable and provides an approximation $\hat{\varphi}$ of the true flow function φ while preserving causality and continuity. In the next section, we propose a neural network-based architecture to solve this problem.

3. PROPOSED ARCHITECTURE

3.1 Motivation

Due to causality and the considered class of inputs (2), the flow $\varphi(s, x, u)$ at a time instant $s \in \mathbb{R}_{\geq 0}$ depends only on a finite number of the input values $\{u_k\}$. Thus, at any time during the first control period $[0, \Delta]$, only the value of $u_1 \in \mathcal{U}$ and the initial condition $x \in \mathcal{X}$ are required to define φ . Therefore, we define $\Phi : [0, 1] \times \mathcal{X} \times \mathcal{U} \rightarrow \mathcal{X}$ as

$$\Phi(\tau, x, u_1) := \varphi(\tau\Delta, x, u),$$

such that a finite-dimensional vector of parameters (as opposed to functions) directly maps to the flow. For an arbitrary time instant $s \in \mathbb{R}_{\geq 0}$, the flow $\varphi(s, x, u)$ can be computed as follows:

- (1) Construct a map $d_\Delta : (s, u) \mapsto \{\tau_k, u_k\}_{k=1}^{k_s+1}$ such that u_k is given according to (2) and

$$k_s := \left\lfloor \frac{s}{\Delta} \right\rfloor, \quad \tau_k := \begin{cases} 1, & k \leq k_s \\ \frac{s - k_s\Delta}{\Delta}, & k = k_s + 1. \end{cases}$$

- (2) Define the sequence $x_k \in \mathcal{X}$ for all $k = 1, \dots, k_s + 1$ as

$$\begin{aligned} x_0 &= x, \\ x_k &= \Phi(\tau_k, x_{k-1}, u_k). \end{aligned} \quad (6)$$

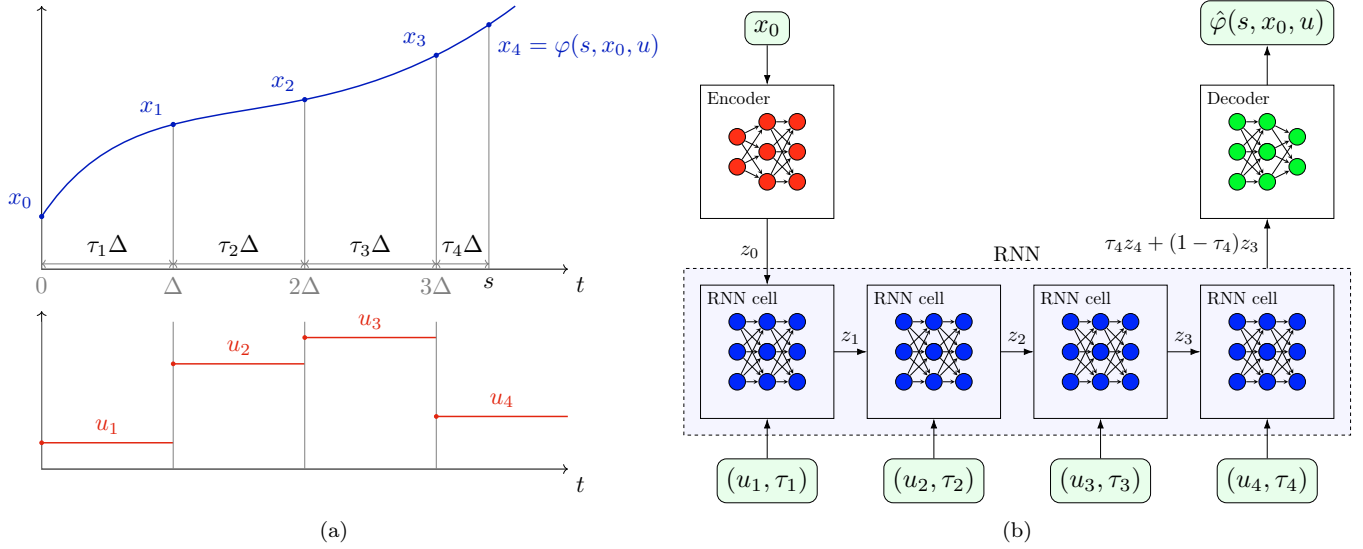


Fig. 1. (a) Schematic illustration of true flow function φ for parameters $\{u_k, \tau_k\}_{k=1}^4$. (b) Corresponding model for the approximated flow $\hat{\varphi}$. In the approximated model, we first map the initial condition to a higher dimensional space through a feedforward *encoder* network. Then, the encoded state is propagated in time through a Recurrent Neural Network (RNN). Each cell of the RNN sequentially takes (u_k, τ_k) as inputs. The two last hidden states are interpolated and mapped back to \mathcal{X} through another feedforward *decoder* network.

Then we have that $x_{k_s+1} = \varphi(s, x, u)$. Thus, trajectories of φ can be equivalently represented by the trajectories of a discrete-time dynamical system with inputs (τ_k, u_k) , as represented in (6) and illustrated in Figure 1a.

3.2 Definition of the architecture

Based on the previous discussion, we focus on approximating the set of difference equations (6), and Recurrent Neural Network (RNN) models are suitable for this task since they are universal approximators of such mappings (Schäfer and Zimmermann, 2006).

To increase the flexibility of the model, we first map the initial state to a feature space \mathcal{Z} using an *encoder* network, a deep neural network (DNN) whose map we denote by h_{enc} . Denoting the mapping defined by the RNN as h_{RNN} , we then have

$$z_0 = h_{\text{enc}}(x),$$

$$(z_1, \dots, z_{k_s+1}) = h_{\text{RNN}}\left(z_0, \{(u_k, \tau_k)\}_{k=1}^{k_s+1}\right),$$

where $z_i \in \mathcal{Z}$ are the hidden states of the RNN. To ensure that the learned flow function is continuous in time, we combine the two last RNN states using the following map:

$$g\left(\{z_k\}_{k=1}^{k_s+1}, \{\tau_k\}_{k=1}^{k_s+1}\right) := (1 - \tau_{k_s+1})z_{k_s} + \tau_{k_s+1}z_{k_s+1}.$$

Note that this does not amount to linear interpolation since z_{k_s+1} depends on τ_{k_s+1} . To map the output of g back to a state vector in \mathcal{X} we use a *decoder* DNN which we denote by h_{dec} , whose output yields the approximated flow $\hat{\varphi}$ at the time instant s , i.e.

$$\hat{\varphi}(s, x, u) = h_{\text{dec}}(z), \quad (7)$$

where z is defined as

$$z := g\left(h_{\text{RNN}}\left(h_{\text{enc}}(x), \{u_k, \tau_k\}_{k=1}^{k_s+1}\right), \{\tau_k\}_{k=1}^{k_s+1}\right). \quad (8)$$

As a result, the hypothesis space \mathcal{H} is defined by the set of functions $\hat{\varphi}$ of the form (7)-(8) that are parameterised by the parameters of the networks h_{RNN} , h_{enc} and h_{dec} .

A block diagram of the architecture is shown in Figure 1b.

4. EXPERIMENTAL EVALUATION: VAN DER POL OSCILLATOR

We illustrate the proposed method by evaluating its performance in learning the flow function of the Van der Pol oscillator. Additionally, we study the generalisation capabilities of the trained model with respect to the simulation time horizon and the input distribution.

4.1 Learning the flow function

Data generation: The Van der Pol oscillator is described by the system of ordinary differential equations

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) \\ \dot{x}_2(t) &= -x_1(t) + (1 - x_1(t)^2)\mu x_2(t) + u(t). \end{aligned} \quad (9)$$

We take $\mu = 1$ and $x(0) \sim N(0, I)$, i.e. a standard normal distribution. The control input sampling time is $\Delta = 0.2$ and the inputs considered are square wave inputs with period 5Δ and amplitudes sampled i.i.d. from $N(0, \sigma = 5)$, i.e.

$$\begin{aligned} u_{1+5k} &\sim N(0, 5), \\ u_{j+5k} &= u_{1+5k}, \quad j = 2, 3, 4, 5 \end{aligned}$$

holds for all $k \geq 0$.

To generate the data used to train the model, we integrate (9) with an RK45 solver. A total of $N = 30$ trajectories are generated, and these are divided into train, validation and test sets following a 60% – 20% – 20% random split. For each trajectory, $K = 200$ time points t_k^i are sampled using Latin hypercube sampling. The measurement noise in (4) is zero-mean Gaussian noise with standard deviation of 0.1.

Training: We train the model using the stochastic gradient descent algorithm **Adam** with a batch size of 1024.

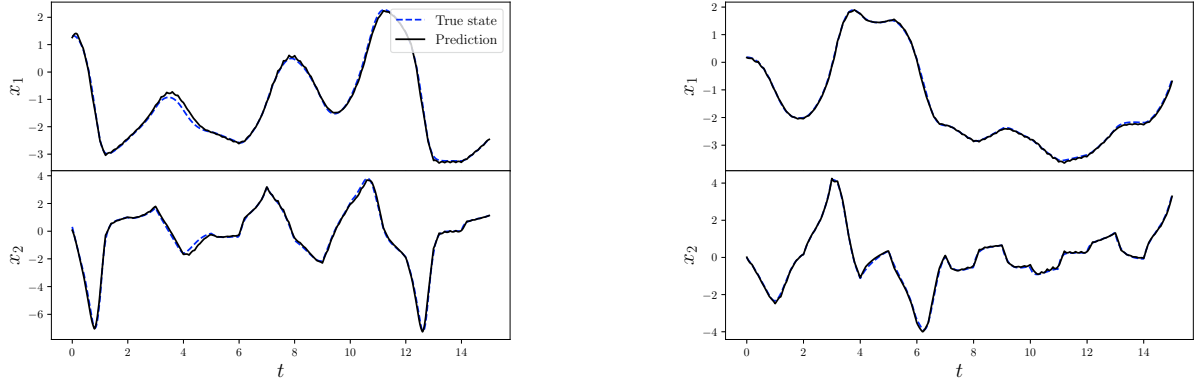


Fig. 2. Actual (blue, dashed) and predicted (black) trajectories for the Van der Pol model with initial condition and input drawn from the corresponding distributions.

The learning rate is reduced 5-fold for every 5 epochs in which the validation loss is not reduced, and the training is stopped when the validation loss does not decrease more than 5×10^{-4} for 30 consecutive epochs.

Results: We used random search to determine the best values for the size of the encoder, decoder and recurrent networks, and the initial learning rate. The optimal initial learning rate was found to be 1×10^{-2} and the training took 186 seconds on a cluster node with an NVIDIA T4 GPU and an Intel(R) Xeon(R) Gold 6226R CPU @ 2.90GHz. The RNN is a single-layer LSTM network with 8 hidden states. The encoder network maps the initial state x to the initial LSTM hidden and cell states, and is a 3-layer feedforward net with 96 nodes in the hidden layers. The decoder network maps the hidden LSTM state to the flow value and is a 3-layer feedforward network with 48 nodes in the hidden layers. All networks use tanh activations.

Figure 2 shows two predicted trajectories on $[0, 15]$ for two new pairs of initial conditions and inputs drawn from P_x and P_u (i.e., unseen during training). Note that the actual trajectory and the predicted trajectory are nearly indistinguishable.

4.2 Prediction over large time horizons

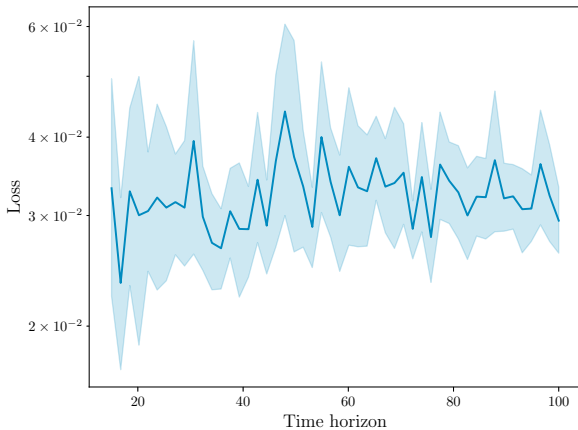


Fig. 3. Estimate of $\ell_t(\hat{\phi})$ as a function of t for the Van der Pol model trained with $T = 15$.

The trajectories used for training the model of the previous section have a time horizon of 15 seconds. Due to the recurrent structure of the model and the stability of the system under consideration, we expect that the performance is maintained for longer prediction horizons. To this end, we compute the test loss $\ell_t(\hat{\phi})$ on a different set of 100 trajectories on $[0, t]$ for a set of gridded values $15 \leq t \leq 100$. The result is shown in Figure 3, where the coloured area represents the 95% confidence interval approximated using the empirical variance estimate. We observe that ℓ_t remains approximately constant as t increases, indicating that the model gives reliable predictions for t much larger than the value of T used for the training trajectories.

4.3 Generalisation to different input distributions

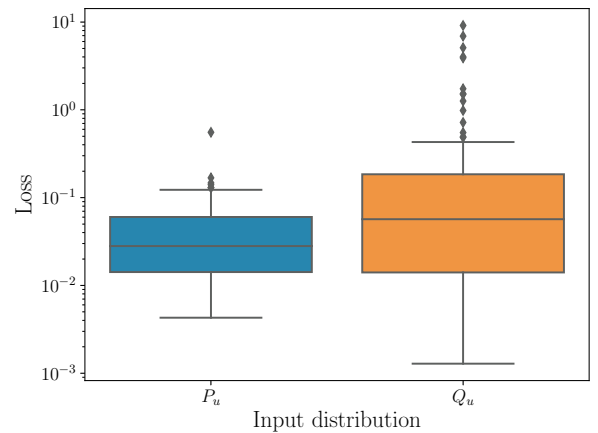


Fig. 4. Distribution of the loss $\ell_T(\hat{\phi})$ with input distributions P_u (blue) and Q_u (orange).

We additionally investigate the performance of the model trained in section 4 with an input distribution different from the training distribution P_u . In particular, we consider a new distribution Q_u on \mathbb{U} consisting of sinusoidal sequences with random amplitude and frequency, that is,

$$u_k = A \sin(\Omega k \Delta),$$

where $A \sim \text{LogNormal}(0, 1)$ and $\Omega \sim \text{Uniform}(0, 2\pi)$. This corresponds to sinusoidal signals with a maximum frequency of 1 Hz.

To verify the performance on this class of inputs, we compute an estimate of $\ell_T(\hat{\varphi}; Q_u)$, defined as in equation (3) with the expectation taken with $U \sim Q_u$. Figure 4 shows a box plot of the distribution of the estimate of ℓ_T computed on 100 trajectories for each of the two input distributions P_u and Q_u . As expected, the mean and variance of the prediction loss for the distribution Q_u are slightly higher than for P_u , but remain reasonably close to that of P_u .

5. PREDICTING EXCITABILITY IN THE FITZHUGH-NAGUMO OSCILLATOR

Excitability is the system property of biological oscillators that constitute neurons, muscle cells, and endocrine cells. Here, based on the input energy, the output response either exhibits resting behaviour or stereotypical spike trains (also known as an all-or-none response, see Sepulchre et al. (2017)). Predicting the excitable behaviour has been a key problem in neurophysiology. In this paper, we provide a data-driven method where we apply the proposed model architecture for learning the flow of FitzHugh-Nagumo oscillator and predict its excitable behaviour.

5.1 Learning the flow function

Data generation: The FitzHugh-Nagumo oscillator is described by the following set of nonlinear differential equations:

$$\begin{aligned}\eta \dot{x}_1(t) &= x_1(t) - x_1(t)^3 - x_2(t) + u(t), \\ \eta \gamma \dot{x}_2(t) &= x_1(t) + a - b x_2(t),\end{aligned}\quad (10)$$

where η, γ, a and b are positive constants, which we choose as $\eta = 1/50$, $\gamma = 40$, $a = 0.3$, $b = 1.4$.

As before, we take $x(0) \sim N(0, I)$. The control period is $\Delta = 0.1$ and the input distribution P_u is given by

$$\begin{aligned}u_{1+40k} &\stackrel{\text{i.i.d.}}{\sim} \text{LogNormal}(\mu = \log(0.2), \sigma = 0.5), \\ u_{j+40k} &= u_{1+40k}, \quad j = 2, \dots, 40\end{aligned}$$

for all $k \geq 0$. This is chosen so that the multi-stable behaviour of the oscillator is observed.

We generated $N = 300$ trajectories on $[0, 20]$ using an RK45 solver, sampling $K = 300$ time points from each trajectory using Latin hypercube sampling. The measurement noise in (4) is zero-mean Gaussian noise with standard deviation equal to 0.05.

Results: We trained a model where the RNN is an LSTM with 16 hidden states, the encoder has 2 hidden layers with 64 nodes each, and the decoder has 2 hidden layers with 32 nodes each. The training took 1196 seconds with an initial learning rate of 2×10^{-2} , with the same algorithm and hardware as in Section 4.

Figure 5 shows two predicted trajectories on $[0, 40]$ for two new pairs of initial conditions and inputs drawn from P_x and P_u (i.e., unseen during training). It can be observed (in Figure 6) that even though the trained model occasionally fails to predict the peak value of the oscillations it subsequently recovers from the error.

5.2 Prediction of excitability

We further predict the occurrences of excitable behaviour in FitzHugh-Nagumo oscillator by carefully changing the

amplitude of the input signal. Note that, for the FitzHugh-Nagumo model (10), excitability is well-studied, allowing us to select the amplitude of the input suitable for exhibiting excitability. In Figure 6a, we show that the learned model is able to predict that, as the amplitude of the input is gradually decreased over time, the oscillator's response traverses from a higher resting potential (constant response), passes through the excitable region (periodic spike train), and returns to a lower resting potential (constant response). In Figure 6b, the model predicts the occurrence of two distinct excitable regions with two sets of spike trains.

6. CONCLUDING REMARKS

We presented a recurrent neural network architecture to learn the flow of a causal and time-invariant control system in continuous time from trajectory data. Exploiting causality and time-invariance, we show that the problem of learning the flow function can be cast as the problem of learning a discrete-time dynamical system, motivating the use of an RNN-based architecture. Our experimental results on the Van der Pol and FitzHugh-Nagumo oscillators show that the learned model has good prediction performance, and demonstrate that the model is able to generalise to longer prediction time horizons and new classes of input signals. We expect that our approach can provide an alternative to traditional modelling approaches in control problems, bypassing the need of solving complex dynamics equations by directly predicting trajectories using the flow function.

There are many possible avenues of research to improve upon and develop the method we have proposed here. An immediate extension would be the removal of the restriction to piecewise constant inputs, in order to obtain a more general class of continuous-time models. Regarding further experimental validation, we intend to evaluate the method on other classes of systems. A scalability study to systems with a large number of states would be particularly interesting. It is also important to evaluate the model in the context of control applications where specific forms of feedback are typically involved. Theoretical properties of the problem of learning the flow function, as formulated in Section 2, have not been studied, to the best of our knowledge. It would be of interest to investigate whether it is possible to obtain error bounds or scaling laws for the size of the optimal model.

Additionally, this work has significant implications for neuro-biological study as well as negative resistance circuits. Learning the flow function can be viewed as a new method for in silico studies that is data-driven, independent of numerical integration and does not suffer from numerical instability. For instance, learning excitable spiking patterns from an up-scaled network of neurons (e.g., a network of Hodgkin-Huxley oscillators with multiple ion-channels) will directly follow from the methods developed in this paper. From the point of view of electrical circuits, learning the flow function is equivalent to learning the conductance of a negative resistance circuit if we consider the input to be a current signal and the state trajectories to be branch voltages.

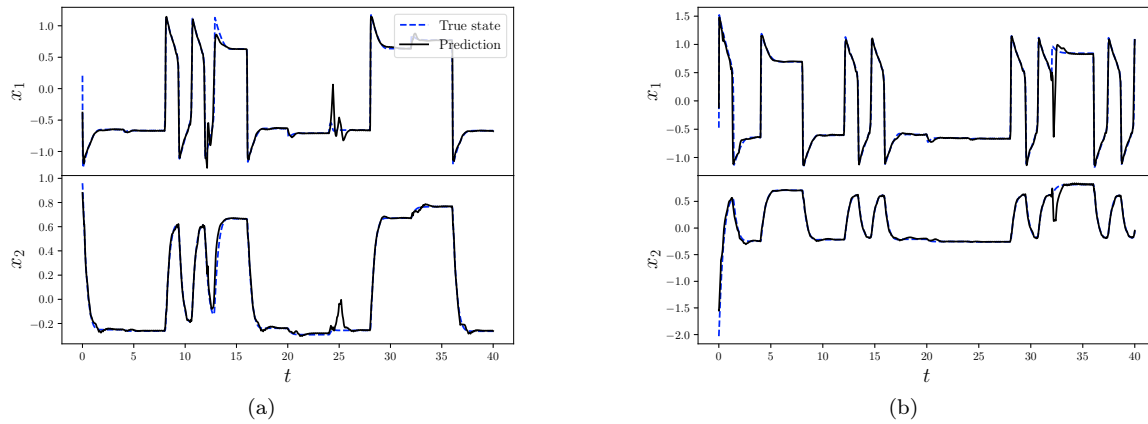


Fig. 5. Actual (blue, dashed) and predicted (black) trajectories for the FitzHugh-Nagumo oscillator with initial condition and input drawn from the corresponding distributions.

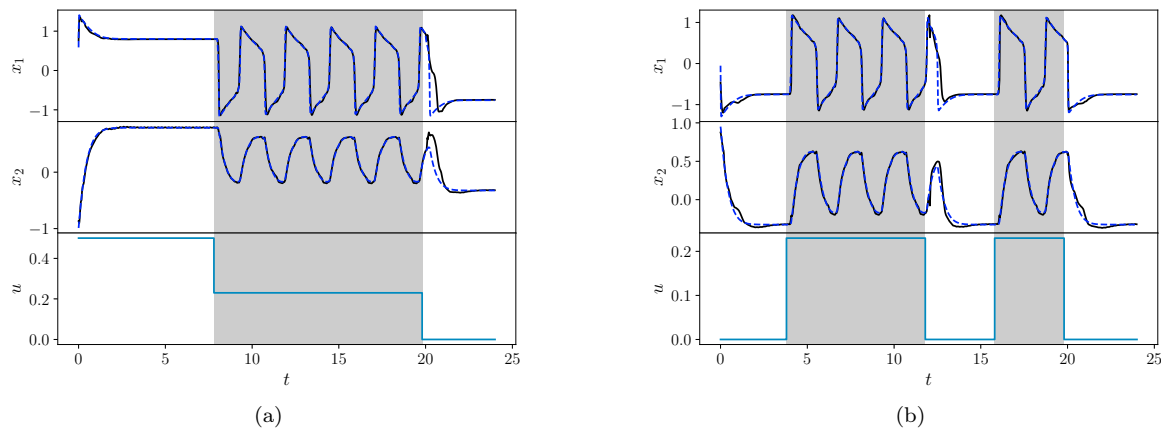


Fig. 6. In (a) and (b), actual (blue, dashed) and predicted (black) trajectories demonstrate two distinct cases of excitability for the FitzHugh-Nagumo oscillator. Excitable regions are shaded in grey.

7. ACKNOWLEDGEMENTS

The computations were enabled by resources provided by the Swedish National Infrastructure for Computing (SNIC) at Chalmers Centre for Computational Science and Engineering (C3SE) partially funded by the Swedish Research Council through grant agreement no. 2018-05973.

REFERENCES

- Biloš, M., Sommer, J., Rangapuram, S.S., Januschowski, T., and Günnemann, S. (2021). Neural flows: Efficient alternative to neural ODEs. In *Advances in Neural Information Processing Systems*.
- Forgione, M. and Piga, D. (2021). Continuous-time system identification with neural networks: Model structures and fitting criteria. *European Journal of Control*, 59, 69–81.
- Geneva, N. and Zabaraz, N. (2022). Transformers for modeling physical systems. *Neural Networks*, 146, 272–289.
- Hanson, J. and Raginsky, M. (2020). Universal simulation of stable dynamical systems by recurrent neural nets. In *Proceedings of the 2nd Conference on Learning for Dynamics and Control*, volume 120 of *Proceedings of Machine Learning Research*, 384–392. PMLR.
- Kissas, G., Seidman, J.H., Guilhoto, L.F., Preciado, V.M., Pappas, G.J., and Perdikaris, P. (2022). Learning operators with coupled attention. URL <https://arxiv.org/abs/2201.01032>.
- Li, Z., Kovachki, N.B., Azizzadenesheli, K., Liu, B., Bhattacharya, K., Stuart, A., and Anandkumar, A. (2021). Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*.
- Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G.E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3), 218–229.
- Maciejowski, J.M. (2002). *Predictive Control with Constraints*. Prentice Hall, England.
- Schäfer, A.M. and Zimmermann, H.G. (2006). Recurrent neural networks are universal approximators. In *Artificial Neural Networks – ICANN 2006*, 632–640. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Schoukens, J. and Ljung, L. (2019). Nonlinear system identification: A user-oriented road map. *IEEE Control Systems*, 39(6), 28–99.
- Sepulchre, R., Drion, G., and Franci, A. (2017). Excitable behaviors. *arXiv e-prints*, arXiv:1704.04989.
- Sontag, E.D. (1998). *Mathematical Control Theory: Deterministic Finite Dimensional Systems*, 25–80. Springer New York, New York, NY.