Friedrich-Alexander-Universität
Faculty of Sciences

Friedrich-Alexander-Universität
DYNAMICS, CONTROL, MACHINE LEARNING AND NUMERICS

# APPROXIMATION THEORY AND APPLICATIONS OF SEMI-AUTONOMOUS NEURAL ODES

**Ziqian Li   Lorenzo Liverani   Enrique Zuazua**
Chair in Dynamics, Control, and Numerics – AvH Professorship
FAU Erlangen-Nürnberg

**Kang Liu**
Institut de Mathématiques de Bourgogne
Université Bourgogne Europe

## Introduction

Neural ordinary differential equations (NODEs) represent a groundbreaking fusion of deep learning and differential equations [1]. Mathematically, NODEs rule the evolution of an absolutely continuous state trajectory $\boldsymbol{x} = \boldsymbol{x}(t) : [0, T] \to \mathbb{R}^d$ via an ordinary differential equation parameterized by a neural network,

$$\begin{cases} \dot{\boldsymbol{x}} = \displaystyle\sum_{i=1}^{P} W_i(t) \circ \boldsymbol{\sigma}(A_i(t)\boldsymbol{x} + B_i(t)), \\ \boldsymbol{x}(0) = x_0, \end{cases}$$

where $P$ is the number of neurons per layer, and $\{A_i(t), W_i(t), B_i(t)\}_{i=1}^{P}$ are time-dependent network parameters.

NODEs excel at interpolating irregular, time-stamped data, yet a rigorous approximation theory for general ODEs remains underdeveloped. This work addresses that gap by establishing universal approximation results and convergence rates for NODE-based models.

## SA-NODEs

In this work [2], we focus on a particular instance of NODEs, namely,

$$\begin{cases} \dot{\boldsymbol{x}} = \displaystyle\sum_{i=1}^{P} W_i \circ \boldsymbol{\sigma}(A_i^1\boldsymbol{x} + A_i^2 t + B_i), \\ \boldsymbol{x}(0) = x_0. \end{cases} \tag{1}$$

Because the weights $W_i$, matrices $A_i^1$, $A_i^2$, and biases $B_i$ do not depend on $t$, we call this equation *semi-autonomous NODEs* (SA-NODEs).

The approximated vector field $f_\Theta = \sum_{i=1}^{P} W_i \circ \boldsymbol{\sigma}(A_i^1\boldsymbol{x} + A_i^2 t + B_i)$ is uniformly Lipschitz continuous in $\boldsymbol{x}$ with the estimate:

$$\|f_\Theta(\boldsymbol{x}, t) - f_\Theta(\boldsymbol{y}, t)\| \leq \left\| \sum_{i=1}^{P} |W_i| \circ \|A_i^1\|_{\ell^2} \right\| \|\boldsymbol{x} - \boldsymbol{y}\|. \tag{2}$$

## Approximation theory

- **Approximation for ODE systems.** Consider the ODE system

$$\begin{cases} \dot{\boldsymbol{z}} = f(\boldsymbol{z}, t), \; t \in (0, T), \\ \boldsymbol{z}(0) = z_0. \end{cases} \tag{3}$$

  Under the sole assumption of $f$ being continuous in time and uniformly Lipschitz in space, the associated SA-NODE approximation $x_{z_0}(t)$ satisfies the following universal approximation property:

$$\|\boldsymbol{z}_{z_0}(\cdot) - \boldsymbol{x}_{z_0}(\cdot)\|_{\mathbb{L}^\infty([0,T];\mathbb{R}^d)} \leq \varepsilon.$$

  By further assuming $f \in \mathcal{H}^k_{\mathrm{loc}}$ with $k > (d+1)/2 + 2$, one obtains an upper bound on the approximation rate

$$\|\boldsymbol{z}_{z_0}(\cdot) - \boldsymbol{x}_{z_0}(\cdot)\|_{\mathbb{L}^\infty([0,T];\mathbb{R}^d)} \leq \frac{C_{T,f}}{\sqrt{P}}, \quad \forall z_0 \in [-1, 1]^d.$$

- **Approximation for transport equations.** For the transport equation

$$\begin{cases} \partial_t \rho + \mathrm{div}_x(f(x, t)\, \rho) = 0, \quad (x, t) \in \mathbb{R}^d \times [0, T], \\ \rho(\cdot, 0) = \rho_0, \end{cases} \tag{4}$$

  the SA-NODE–based neural transport equation [3]

$$\begin{cases} \partial_t \rho_\Theta + \mathrm{div}_x((\displaystyle\sum_{i=1}^{P} W_i \circ \boldsymbol{\sigma}(A_i^1 x + A_i^2 t + B_i))\rho_\Theta) = 0, \\ \rho_\Theta(\cdot, 0) = \rho_0. \end{cases} \tag{5}$$

  achieves the approximation bound

$$\sup_{t\in[0,T]} \mathbb{W}_1(\rho(\cdot, t), \rho_\Theta(\cdot, t)) \leq \frac{C_{T,f,\rho_0}}{\sqrt{P}},$$

  where $\mathbb{W}_1(\cdot, \cdot)$ is the Wasserstein-1 distance.

## Training strategy

To approximate the ODE system (3) by the SA-NODE (1), we collect a training set

$$\mathcal{D} = \{\boldsymbol{z}_k(t_l)\}_{k,l} \subset \mathbb{R}^d, \quad k = 1, \ldots, N, l = 1, \ldots, M,$$

where $N$ is the number of trajectories and $M$ is the number of time steps. We write the SA-NODE prediction as $\boldsymbol{x}_k(t_l, \Theta)$. Since the network's Lipschitz constant is controlled by its parameters (cf. (2)), we train by minimizing

$$L(\Theta) = \frac{1}{NM} \sum_{k=1}^{N} \sum_{l=1}^{M} (\boldsymbol{z}_k(t_l) - \boldsymbol{x}_k(t_l, \Theta))^2 + \lambda \left\| \sum_{i=1}^{P} |W_i| \circ \|A_i^1\|_{\ell^2} \right\|.$$

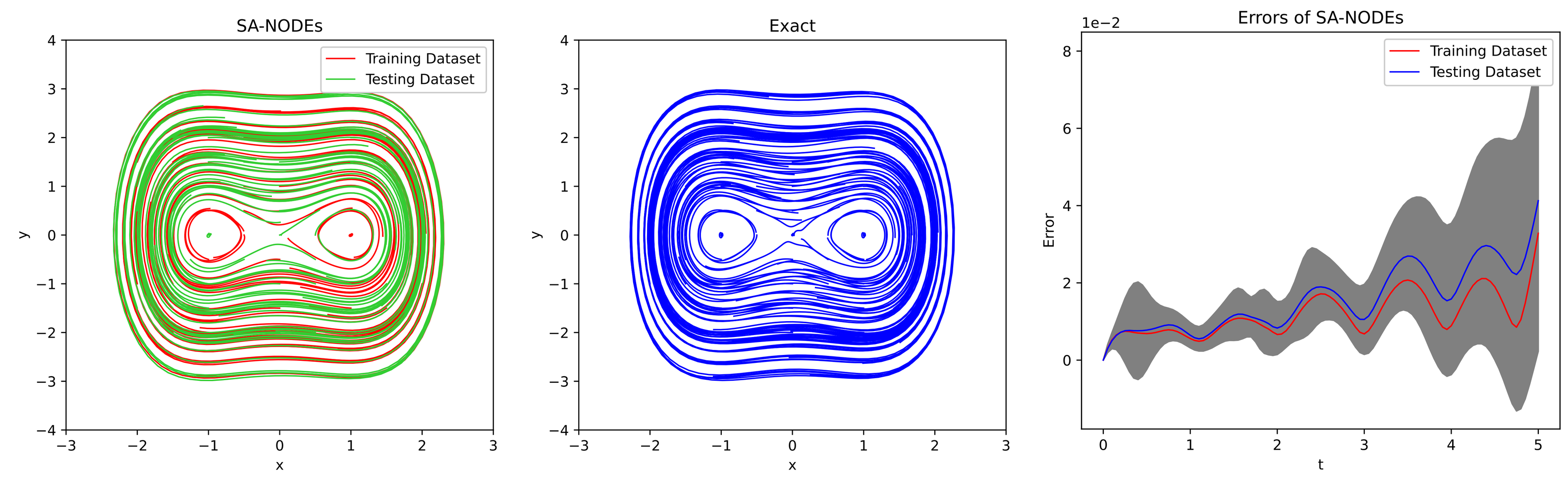
To extend training to the transport equation (4) by the neural transport equation (5), we add $L(\Theta)$ with an additional term measuring the discrepancy between $\rho_\Theta$ and $\rho$ along each trajectory, then obtain the loss function for training transport equations.

## Numerical experiments

- **Simulations of ODEs.** We simulate the Duffing oscillator

$$\begin{cases} \dot{z}_1 = z_2, \\ \dot{z}_2 = z_1 - z_1^3 + \delta \cos(\omega t), \end{cases}$$

  using SA-NODE where $\delta = 0.1$ and $\omega = \pi$. Half of the trajectories ($N/2$) form the training set; the remainder are used for testing. The following result shows SA-NODEs simulates well with the system.




(a) SA-NODEs and exact solution. (b) Errors.

Figure: SA-NODEs solution, exact solution and errors for ODE systems.

- **Simulations of transport equations.** We model two-dimensional frontogenesis via the transport equation

$$\begin{cases} \partial_t \rho(x, y, t) + \mathrm{div}\left((-yg(r(x, y)), xg(r(x, y)))\, \rho(x, y, t)\right) = 0, \\ \rho(\cdot, 0) = \rho_0, \end{cases}$$

  where

$$g(r(x, y)) = \frac{1}{r(x, y)}\, \overline{v}\, \mathrm{sech}^2(r(x, y)) \tanh(r(x, y)),$$

  with $r(x, y) = \sqrt{x^2 + y^2}$ and $\overline{v} = 2.59807$. Over $t \in [0, 4]$, SA-NODE achieves near-perfect alignment with the analytic solution.

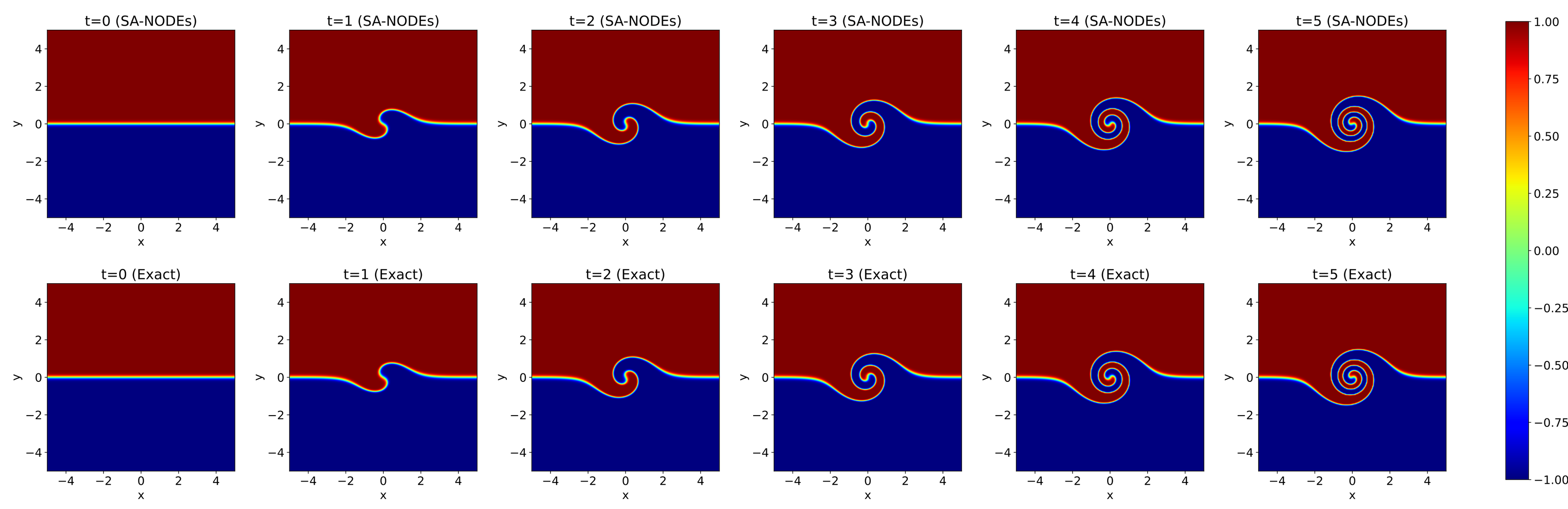



Figure: SA-NODEs and exact solution for Doswell frontogenesis.

## Conclusions and Perspectives

- **Conclusions**
  - We introduced SA-NODEs, a unified framework for modeling and approximating both ODE and transport-PDE dynamics.
  - We proved their universal approximation property and established explicit convergence rates.
  - Numerical experiments validate SA-NODEs' accuracy and robustness across multiple test cases.
- **Perspectives**
  - Extend SA-NODEs to inverse problems for system identification and parameter recovery.
  - Explore their predictive performance for long-term forecasting in complex dynamical systems.
  - Incorporate SA-NODEs into model predictive control (MPC) pipelines for real-time decision making.

## References

[1] R. T. Q. Chen, Y. Rubanova, J. Bettencourt, D. Duvenaud. Neural ordinary differential equations. NeurIPS 2018.
[2] Z. Li, K. Liu, L. Liverani, E. Zuazua (2024). Universal Approximation of Dynamical Systems by Semi-Autonomous Neural ODEs and Applications. submitted, under review.
[3] D. Ruiz-Balet, E. Zuazua (2024). Control of neural transport for normalising flows. J. Math. Pures Appl, 181: 58-90.

Department MATHEMATIK

Unterstützt von / Supported by
Alexander von Humboldt Stiftung / Foundation