a Neural Network Approach to Learn Delay Differential Equations via Pseudospectral Collocation

Dimitri Breda, Muhammad Tanveer*

CDLab – Computational Dynamics Laboratory
Department of Mathematics, Computer Science and Physics – University of Udine

*PhD in Mathematical and Physical Sciences (PNRR)

Funded by the European Union
NextGenerationEU

"MOdellistica Numerica e Data-driven per l'Innovazione sostenibile – MONDI"
(CUP: G25F21003390007)

PINN-PAD
February 22, 2024 @ Padova

introducing DDEs [3, 5, 6]

- let $\tau > 0$, $X \coloneqq C([-\tau, 0]; \mathbb{R})$ and consider

$$x'(t) = f(x_t)$$

  for smooth $f : X \to \mathbb{R}$ and $x_t(\theta) \coloneqq x(t + \theta)$, $\theta \in [-\tau, 0]$

- e.g., Mackey-Glass [9]:

$$f(\varphi) \coloneqq \frac{\beta\varphi(-\tau)}{1 + \varphi(-\tau)^n} - \gamma\varphi(0) \quad \Rightarrow \quad x'(t) = \frac{\beta x(t - \tau)}{1 + x(t - \tau)^n} - \gamma x(t)$$

(in all the experiments: $\beta = 4, \gamma = 2, n = 9.6, \tau = 1$)

[5] Hale – Springer 1977
[3] Diekmann, van Gils, Verduyn Lunel, Walther – Springer 1995
[6] Erneux – Springer 2009
[9] Mackey, Glass – Science 1977
(full references at the end)

the abstract Cauchy problem [4]

- the IVP in $\mathbb{R}$

$$\begin{cases} x'(t) = f(x_t), & t \geqslant 0 \\ x(\theta) = \varphi(\theta), & \theta \in [-\tau, 0], \end{cases}$$

is equivalent to the IVP in X

$$\begin{cases} u'(t) = \mathcal{A}(u(t)), & t \geqslant 0, \\ u(0) = \varphi \end{cases} \tag{1}$$

through $u(t) = x_t$ for $\varphi \in \mathcal{D}(\mathcal{A})$, where $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subseteq X \to X$ given by

$$\mathcal{A}(\psi) = \psi', \quad \mathcal{D}(\mathcal{A}) = \{\psi \in X \ : \ \psi' \in X \text{ and } \psi'(0) = f(\psi)\}$$

generates the associated semigroup $\{T(t)\}_{t \geqslant 0}$ of solution operators

$$T(t) : X \to X, \qquad T(t)\varphi \coloneqq x_t$$

- DDEs give rise to dynamical systems on X, hence ∞-dimensional:
  - a general approach: reduce (1) to a system of ODEs + tools for ODEs
  - [7] uses Euler in view of applying NODEs and TDNNs to learn DDEs

[4] Engel, Nagel – Springer 1999
[7] Ji, Orosz – IFAC PoL 2022

PseudoSpectral Collocation (PSC)

- let $-\tau = \theta_M < \cdots < \theta_1 < \theta_0 = 0$ be Chebyshev extrema
- for $X_M \coloneqq \mathbb{R}^{M+1}$ set
  - $R_M : X \to X_M$, $\quad R_M\psi \coloneqq (\psi(\theta_0), \psi(\theta_1), \ldots, \psi(\theta_M))$
  - $P_M : X_M \to X$, $\quad (P_M\Psi)(\theta) \coloneqq \sum_{j=0}^{M} \ell_j(\theta)\Psi_j,\ \theta \in [-\tau, 0]$
- discretize $\mathcal{A} : \mathcal{D}(\mathcal{A}) \subseteq X \to X$

$$\mathcal{A}\psi = \psi', \quad \mathcal{D}(\mathcal{A}) = \{\psi \in X \ :\ \psi' \in X \ \text{ and } \ \psi'(0) = f(\psi)\}$$

with $\mathcal{A}_M : X_M \to X_M$

$$[\mathcal{A}_M(\Psi)]_{i=1,\ldots,M} = [R_M(P_M\Psi)']_{i=1,\ldots,M}, \quad [\mathcal{A}_M(\Psi)]_0 = f(P_M\Psi)$$

[2] B., Maset, Vermiglio – SISC 2005 $\rightarrow$ f linear

[1] B., Diekmann, Gyllenberg, Scarabel, Vermiglio – SIADS 2016 $\rightarrow$ f nonlinear

from DDEs to ODEs

- let
$$U(t) \coloneqq (U_0(t), U_1(t), \ldots, U_M(t))^T \in X_M, \qquad U_i(t) \approx u(t)(\theta_i)$$
- the abstract Cauchy problem
$$\begin{cases} u'(t) = \mathcal{A}(u(t)), & t \geqslant 0, \\ u(0) = \varphi \end{cases}$$
with
$$\mathcal{A}\psi = \psi', \quad \mathcal{D}(\mathcal{A}) = \{\psi \in X \ : \ \psi' \in X \ \text{ and } \ \psi'(0) = f(\psi)\}$$
is reduced to the $M+1$ ODEs
$$\begin{cases} U_0'(t) = f(P_M U(t)) \\ U_i'(t) = [D_M U(t)]_i, & i = 1, \ldots, M, \\ U(0) = (\varphi(\theta_0), \varphi(\theta_1), \ldots, \varphi(\theta_M))^T \end{cases}$$
for $[D_M]_{i,j} \coloneqq \ell_j'(\theta_i)$, $i = 1, \ldots, M$, $j = 0, 1, \ldots, M$
- the RHS f affects only the first ODE: easy to code [8]

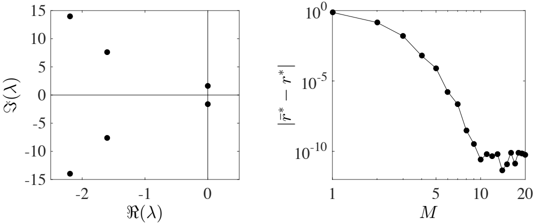
[8] Liessi, Santi, Vermiglio – IFAC TDS 2022

a simple test

- the logistic DDE

$$x'(t) = rx(t)[1 - x(t-1)]$$

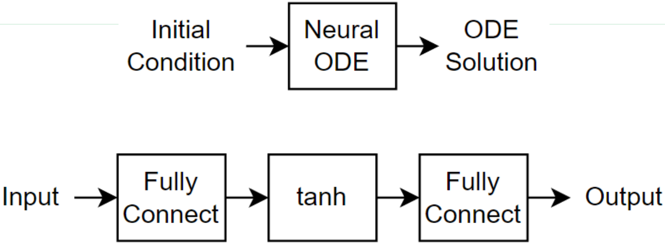
  has equilibrium $\bar{x} = 1$ for all r:
  - asymptotically stable for $r \in (0, \pi/2)$
  - unstable for $r > \pi/2$
  - Hopf bifurcation at $r^* := \pi/2$
- error in approximating r^* via `MatCont` on approximating ODE:



Neural ODEs [10]

NODEs are deep learning operations whose output is defined by the (numerical) solution of an associated ODE:

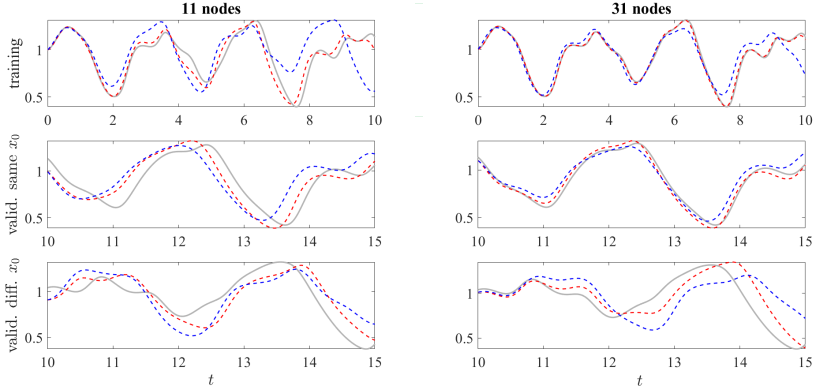
$$y' = g(t, y, \theta)$$



Implementation: We use MATLAB tools for DL (e.g., `dlarray`, `dlode45`)

[10] Chen, Bettencourt, Rubanova, Duvenaud – NeurIPS 2018

ODEs from DDEs via Euler

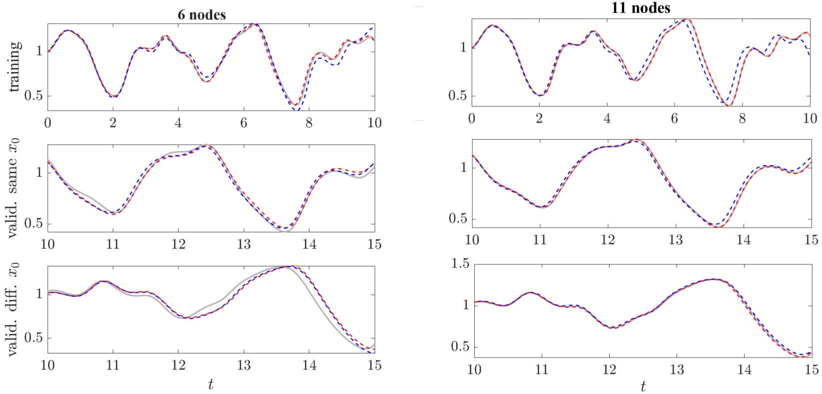


grey: ground truth DDE, red: ground truth ODE, blue: NODE

CPU times:

Euler: 9 minutes for 11 nodes, 12 m for 31 nodes

ODEs from DDEs via PSC



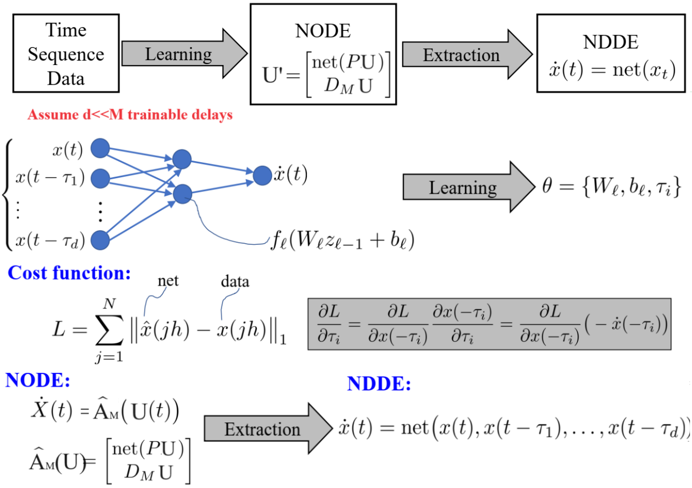
grey: ground truth DDE, red: ground truth ODE, blue: NODE

CPU times:

PSC: 9 minutes for 6 nodes, 9m for 11 nodes

Time delay neural networks [11]

TDNNs are capable of handling delayed input sequences



[11] Waibel, Hanazawa, Hinton, Shikano, Lang – IEEE TASSP (1989)

Resume

ODE:

$$U'(t) = \begin{bmatrix} \frac{\beta U_M(t)}{1+(U_M(t))^n} - \gamma U_0(t) \\ D_M U(t) \end{bmatrix}$$

NODE:

$$U'(t) = \begin{bmatrix} W_3 \tanh(W_2 \tanh(W_1 P U(t) + b_1) + b_2) \\ D_M U(t) \end{bmatrix}$$

NDDE:

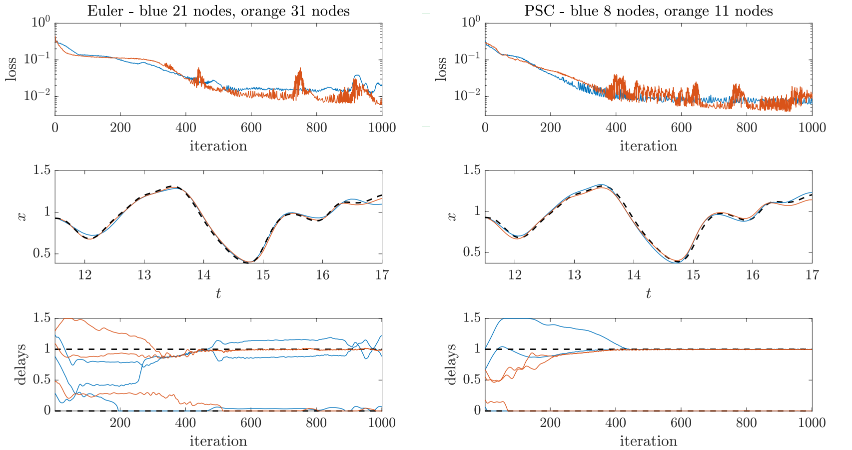
$$x'(t) = W_3 \tanh(W_2 \tanh(W_1 \begin{bmatrix} x(t) \\ x(t-\tau_1) \\ \dots \\ x(t-\tau_d) \end{bmatrix} + b_1) + b_2),$$

Here $P : \mathbb{R}^{M+1} \to \mathbb{R}^{d+1}$ transforms state collocated samples $(x(t+\theta_0), \dots, x(t+\theta_M))$ into delayed states $(x(t), x(t-\tau_1), \dots, x(t-\tau_d))$ and is computed through barycentric interpolation [12] on Chebyshev extrema

Info: simulation horizon = 0.5, batch size = 1000, and update parameters using adaptive moment estimation method

[12] J. Berrut, L. Trefethen – SIAM review 2004

TDNN Results



CPU times:

- Euler: 16 minutes for 21 nodes, 17m for 31 nodes
- PSC: 8m for 8 nodes, 10m for 11 nodes

SINDy [13] for DDEs

- Collect measurement data $\mathbf{x}(t)$ and (approximate) $\mathbf{x}'(t)$
- Construct a library $\Theta(\mathbf{x})$ of candidate functions
- Solve $\mathbf{x}' = \Theta(\mathbf{x})\boldsymbol{\Xi}$ by sparse regression promoting the sparsity of $\boldsymbol{\Xi}$

"Expert" SINDy for DDEs [14, 15]

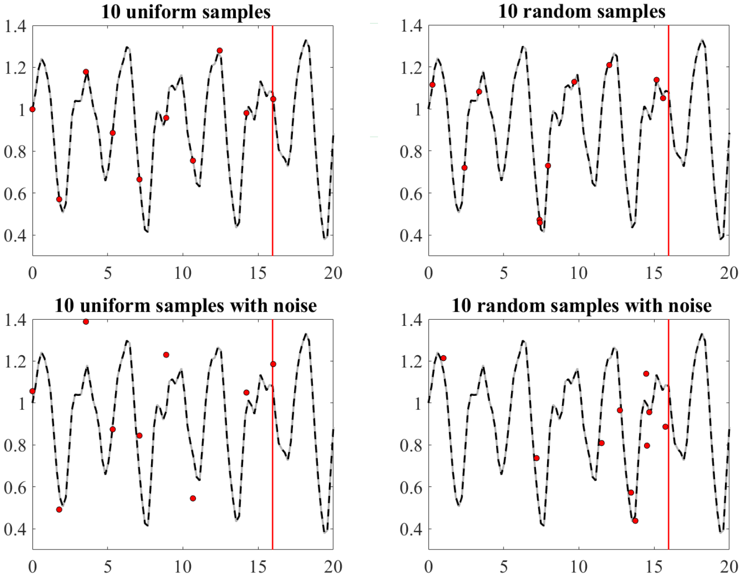
- direct on the DDE
- via library extension to the delayed data-set $x(t - \tau_i)$
- delays assumed to be known

"Pragmatic" SINDy for DDEs

- on ODEs via PSC of the DDE
- via library extension to the collocated data-set $x(t + \theta_i)$
- delays assumed not to be known

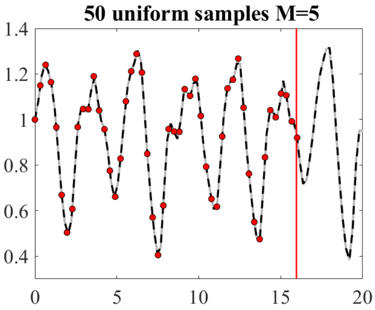
[13] Brunton, Proctor, Kutz – PNAS (2016)
[14] Breda, Demo, Pecile, Rozza – IFAC TDS 2022
[15] Sandoz, Ducret, Gottwald, Vilmart, Perron – PMPES 2023

Expert SINDy



grey: ground truth, black: SINDy, red: samples

Pragmatic SINDy



Concluding remarks

- data-driven methods for DDEs
- Need to explore and understand more
- Challenge: handling of delay terms
- ongoing and future:
  - SINDy for stochastic DDEs, with R. D'Ambrosio (L'Aquila), D. Conte and I. Santaniello (Salerno)
  - extend to structured population models (PNRR)

IFAC

INTERNATIONAL FEDERATION OF AUTOMATIC CONTROL

τDS 2024

18th IFAC Workshop on Time Delay Systems

September 24–27, 2024, Udine, Italy

IFAC
INTERNATIONAL FEDERATION
OF AUTOMATIC CONTROL

τDS
2024

18th IFAC Workshop on Time Delay Systems

September 24–27, 2024, Udine, Italy

bibliografia I

[1] D. Breda, O. Diekmann, M. Gyllenberg, F. Scarabel, and R. Vermiglio, Pseudospectral discretization of nonlinear delay equations: new prospects for numerical bifurcation analysis, *SIAM J. Appl. Dyn. Sys.*, 15(1):1–23, 2016.

[2] D. Breda, S. Maset, and R. Vermiglio, Pseudospectral differencing methods for characteristic roots of delay differential equations, *SIAM J. Sci. Comput.*, 27(2):482–495, 2005.

[3] O. Diekmann, S. A. van Gils, S. M. Verduyn Lunel, and H.-O. Walther, *Delay Equations – Functional, Complex and Nonlinear Analysis*, Number 110 in Applied Mathematical Sciences. Springer Verlag, New York, 1995.

[4] K. Engel and R. Nagel, *One-Parameter Semigroups for Linear Evolution Equations*, Number 194 in Graduate Texts in Mathematics. Springer-Verlag, New York, 1999.

[5] J. K. Hale, *Theory of functional differential equations*, Number 99 in Applied Mathematical Sciences. Springer Verlag, New York, first edition, 1977.

[6] T. Erneux, Applied delay differential equations. Springer, 2009.

bibliografia II

[7] X. A. Ji and G. Orosz, Learning time delay systems with neural ordinary differential equations, *IFAC PapersOnLine*, 55(36):79–84, 2022.

[8] D. Liessi, E. Santi, and R. Vermiglio, Numerical bifurcation analysis of delay equations: a user-friendly matcont interface, In *IFAC TDS*. IFAC, 2022.

[9] M. C. Mackey and L. Glass, Oscillations and chaos in physiological control systems, *Science*, 197:287–289, 1977.

[10] R. Chen, R. Yulia, B. Jesse, and K. David "Neural ordinary differential equations." Advances in neural information processing systems 31, 2018.

[11] A. Waibel, T. Hanazawa, G. Hinton, K. Shikano, K.J. Lang Phoneme recognition using time-delay neural networks IEEE Transactions on Acoustics, Speech, and Signal Processing, 37 (3): 328-339 1989,

[12] J. Berrut, L. Trefethen, Barycentric lagrange interpolation. SIAM review ;46(3):501-17 2004

bibliografia III

[13] S. Brunton, J. Proctor, & J. Kutz, Discovering governing equations from data by sparse identification of nonlinear dynamical systems. Proceedings of the national academy of sciences, 2016

[14] D. Breda, N. Demo, A. Pecile, and G. Rozza, Data-driven methods for delay differential equations, In Proceedings of the 17th IFAC Workshop on Time Delay Systems. IFAC, 2022

[15] A. Sandoz, V. Ducret, G. A. Gottwald, G. Vilmart, and K. Perron, SINDy for delay-differential equations: application to model bacterial zinc response, Proc. Roy. Soc. Lond. A, 479:20220556, 2023