Orthogonal Laguerre Recurrent Neural Networks

Sergio A. Dorado-Rojas
sergio.dorado.rojas@gmail.com
Rensselaer Polytechnic Institute

Bhanukiran Vinzamuri
bhanu.vinzamuri@ibm.com
IBM

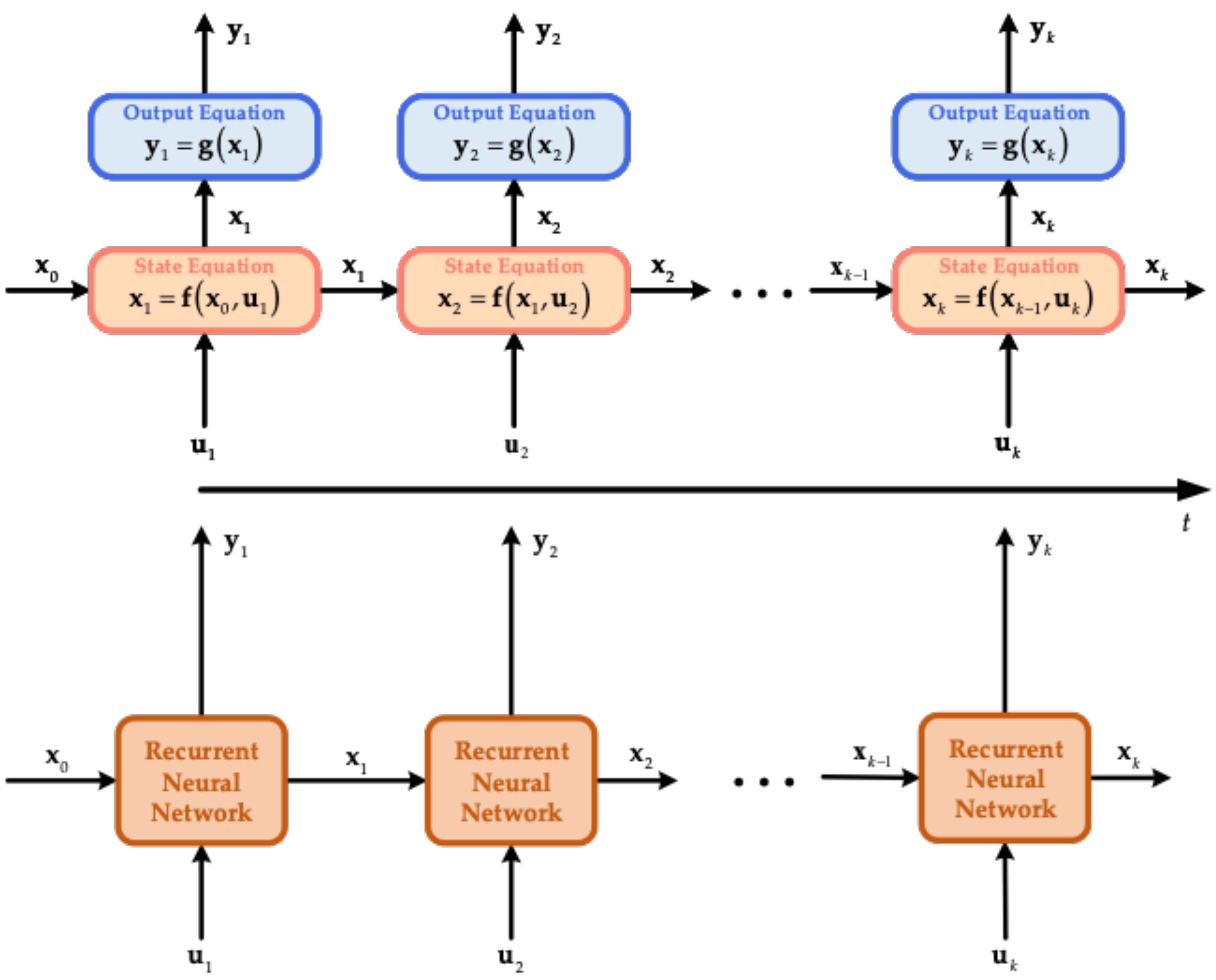
Luigi Vanfretti
luigi.vanfretti@gmail.com
Rensselaer Polytechnic Institute

Overview

Recurrent Neural Networks (RNNs) are a kind of Discrete-Time (DT) systems. Despite remarkable efforts to *analyze* their behavior from the dynamical systems perspective, just recently a dynamical systems approach was used to *design* RNN architectures. In Legendre Memory Units (LMUs), the state equation of a DT Linear Time-Invariant (DTLTI) system serves as a memory update policy by orthogonal time projections. Likewise, in Bistable Recurrent Cells (BRCs), a nonlinearity is used to force bistable behavior in the cell, increasing their memory capabilities mimicking biological neurons. So, the close relationship between RNNs and DT systems can be exploited not only for analysis but also for design.

$$\mathbf{x}_k = \bar{\mathbf{f}}\left(k, \mathbf{x}_{k-1}, \mathbf{u}_{k-1}\right) \Leftrightarrow \mathbf{x}_k = \sigma\left(\mathbf{u}_k, \mathbf{y}_k, \mathbf{x}_{k-1}\right)$$

$$\mathbf{y}_k = \bar{\mathbf{g}}\left(k, \mathbf{x}_k, \mathbf{u}_k\right) \Leftrightarrow \mathbf{y}_k = \gamma\left(\mathbf{h}_k\right)$$



Relevant Papers

[1] Sussillo, D., & Barak, O. (2013). Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. Neural computation, 25(3), 626-649.

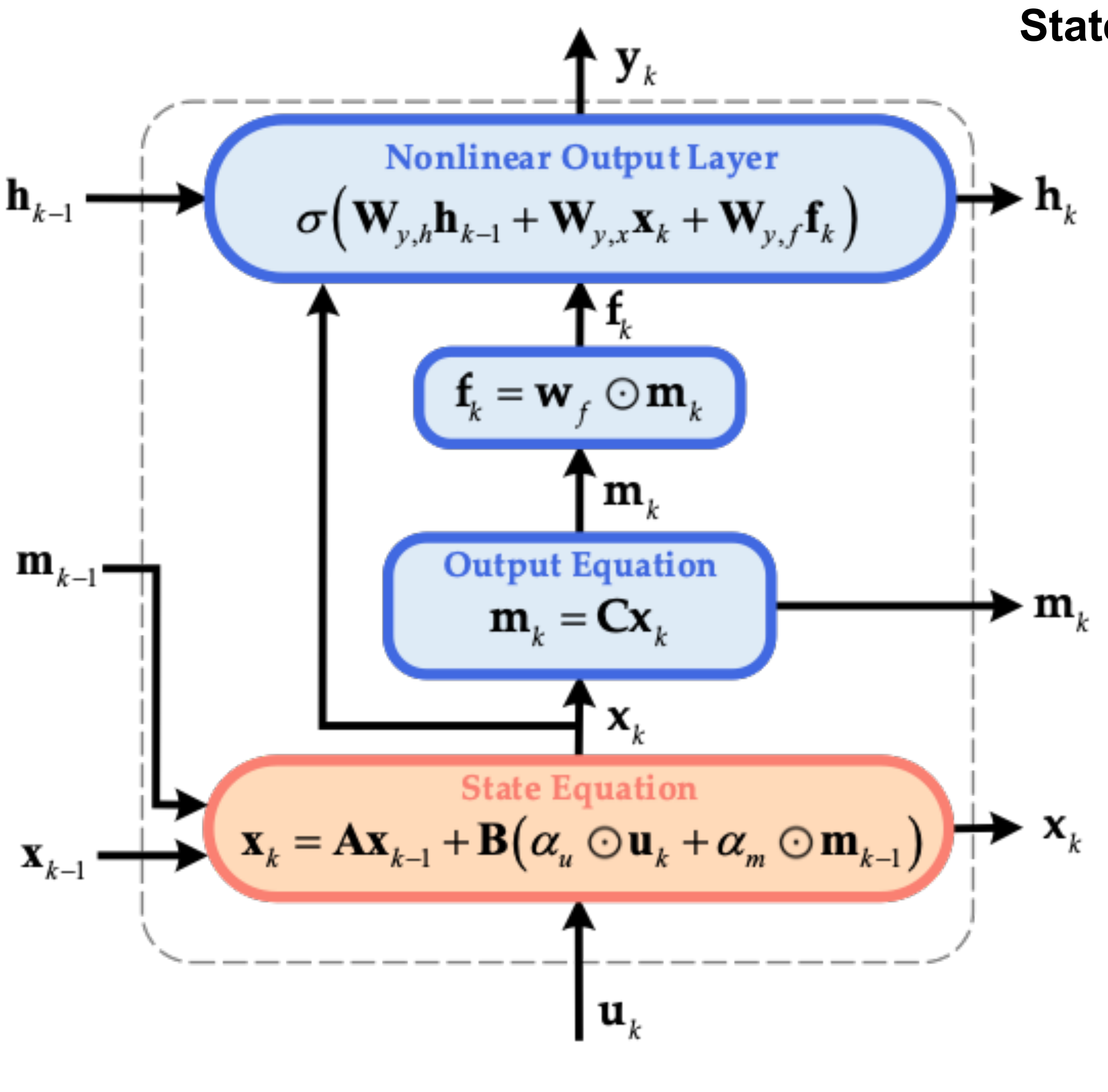
[2] Haviv, D., Rivkind, A., & Barak, O. (2019). Understanding and controlling memory in recurrent neural networks. arXiv preprint arXiv:1902.07275.

[3] Voelker, A., Kajić, I., & Eliasmith, C. (2019). Legendre Memory Units: Continuous-Time Representation in Recurrent Neural Networks. Advances in Neural Information Processing Systems, 15570-15579.

[4] Vecoven, N., Ernst, D., & Drion, G. (2020). A bio-inspired bistable recurrent cell allows for long-lasting memory. arXiv preprint arXiv:2006.05252.

Proposal

Our work exploits a complete state-space representation of a DTLTI system inside an RNN. We employ Laguerre functions, a DTLTI system orthogonal basis, to construct a state update policy and the output equation to perform memory computation thus enhancing RNN expressivity and generalizability. We also introduce a *Ladder network,* a special case of Laguerre networks that introduces delay dynamics inside the RNN.



State and Memory Update

$$\mathbf{x}_k = \mathbf{A}\mathbf{x}_{k-1} + \mathbf{B}\left(\mathbf{u}_k + \mathbf{m}_{k-1}\right)$$
$$\mathbf{m}_k = \mathbf{C}\mathbf{x}_k$$

Full DTLTI system representation

Laguerre Network

The state-space representation is constructed from Laguerre polynomials

$$\ell_i(t) = \left(\sqrt{2p}\right)\frac{e^{pt}}{(i-1)!}\frac{d^{i-1}}{dt^{i-1}}\left[t^{i-1}e^{-2pt}\right]$$

Laguerre polynomials can be computed from $\mathbf{1}(t) = e^{\mathbf{A}_\ell t}\mathbf{1}(0)$ where

$$\mathbf{1}(t) = \left[\ell_1(t) \quad \ell_2(t) \quad \dots \quad \ell_N(t)\right]^T$$

$$\mathbf{A}_\ell = \begin{bmatrix} -p & 0 & \dots & 0 \\ -2p & -p & \dots & 0 \\ \vdots & \dots & \ddots & \vdots \\ -2p & -2p & \dots & -p \end{bmatrix}$$

The $(\mathbf{A}, \mathbf{B}, \mathbf{C})$ matrices are $(\mathbf{A}_\ell, \mathbf{B}_\ell, \mathbf{C}_\ell)$ where $\mathbf{B}_\ell = \mathbf{1}(0)$ and $\mathbf{C}_\ell$ is learned during training

Ladder Network

Let be a delay m_i (number of samples) set by the designer. The ladder network is

$$\mathbf{G}(z) = \begin{bmatrix} \frac{1}{z^{m_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{z^{m_2}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{z^{m_d}} \end{bmatrix}$$

$(\mathbf{A}, \mathbf{B}, \mathbf{C})$ obtained through a state-space representation of $\mathbf{G}(z)$

Results

We validated the performance of our RNN architectures in two dynamical systems-based experiments. In the time-series prediction benchmark, the goal is to predict a nonlinear system's behavior from a set of available measurements. For the system identification experiment, the learning problem aims at determining the state-space representation of a system from data as a regression task. Both experiments are performed for a pendulum with no friction and a fluid flow system.

Model	Pendulum (no noise; linear)	Pendulum (noisy; linear)	Pendulum (no noise; linear)	Pendulum (noisy; linear)	Fluid Flow (noisy)
Ladder	0.0150 ± 0.0050	**0.0116 ± 0.0018**	0.1819 ± 0.1041	0.1929 ± 0.1261	**0.5135 ± 0.0626**
Laguerre	**0.0146 ± 0.0029**	0.0168 ± 0.0057	**0.1249 ± 0.0060**	**0.1257 ± 0.0059**	0.7973 ± 0.5440
LMU	0.2261 ± 0.0986	0.1914 ± 0.0817	0.5132 ± 0.2465	0.5012 ± 0.1133	1.3942 ± 0.0025
BRC	21.8499 ± 1.8473	23.1749 ± 1.8363	14.0479 ± 4.1699	18.0374 ± 1.7233	0.6973 ± 0.1684
nBRC	2.1914 ± 0.9910	1.7015 ± 0.4043	2.6853 ± 1.5883	3.0102 ± 1.0030	0.5656 ± 0.0481