Recurrent Neural Networks with intrinsically critical dynamics

Andrea Ceni, Claudio Gallicchio
andrea.ceni@di.unipi.it, claudio.gallicchio@unipi.it
Department of Computer Science, University of Pisa, Italy

Summary. We introduce two novel Reservoir Computing (RC)-based approaches to design recurrent neural dynamics that operate intrinsically at the edge of criticality, namely the *Edge-of-Critical* reservoir network, and the *Euler State Network*. Both are based on randomized weights and are amenable to neuromorphic hardware implementations. At the same time they show great application potential in comparison with both conventional RC networks and fully trained recurrent networks.

1 Introduction

Reservoir Computing (RC) is a powerful technique for efficiently training Recurrent Neural Networks (RNNs) [6]. RC exploits the inherent stability of neural dynamics by using a fixed, untrained recurrent reservoir layer, with a trainable readout for output computation. This approach has been particularly effective in embedded systems for distributed learning functionalities, and as a reference for neuromorphic hardware implementations [7]. A critical aspect of RC is the development of stable dynamics in the untrained reservoir, based on a global asymptotic stability property called the Echo State Property in the widely known Echo State Network (ESN) model [5]. This property ensures that the reservoir develops stable dynamics while also having a fading memory and limited state space structure. At the same time, though, it crucially limits the ability to transmit input information through several time-steps.

Under the umbrella of RC, we introduce two approaches to enable RNN systems to develop long-term memorization abilities, and effectively latch temporal information across long time-series. The fundamental aspect of our proposal is to modify the recurrent equations of the reservoir system in such a way to determine dynamics that operate *by design* at the edge of criticality.

2 Reservoir Computing with critical dynamics by design

Our first proposal, based on [1, 2] and called *Edge-of-Critical* reservoir network (ECN), consists in modifying the mathematical formulation of the state dynamics in a leaky integrator reservoir. Specifically, we design a reservoir layer with the following state transition function:

$$\mathbf{h}(t) = (1-\beta)\,\mathbf{O}\,\mathbf{h}(t-1) + \beta\,\phi(\mathbf{W}_h\mathbf{h}(t-1) + \mathbf{W}_x\mathbf{x}(t) + \mathbf{b}). \tag{1}$$

Here, $\mathbf{h}(t)$ and $\mathbf{x}(t)$ respectively denote the reservoir state and the externally applied input at time t, $\mathbf{W}_h$ is the recurrent weight matrix, $\mathbf{W}_x$ is the input weight matrix, $\mathbf{b}$ is the bias vector, $\mathbf{O}$ is an orthogonal matrix. Eq. 1 includes a model hyper-parameter $\beta \in (0,1]$ that modulates the influence of the additive term (i.e., $\mathbf{O}\,\mathbf{h}(t-1)$) versus the non-linear update (i.e., $\phi(\mathbf{W}_h\mathbf{h}(t-1) + \mathbf{W}_x\mathbf{x}(t) + \mathbf{b})$) in the new reservoir state. As in conventional RC, all the parameters in the above eq. 1 are randomly generated and left untrained. However, differently from standard ESNs, the value of β can be used to tune the proximity of the reservoir dynamics to the edge of criticality. Such a property follows from the peculiar structure implied in the Jacobian of eq. 1, given as $\mathbf{J}(t) = (1-\beta)\,\mathbf{O} + \beta\,\mathbf{D}(t)\,\mathbf{W}_h$, where $\mathbf{D}(t) = diag(\phi'(\mathbf{W}_h\mathbf{h}(t-1) + \mathbf{W}_x\mathbf{x}(t) + \mathbf{b}))$. A closer analysis reveals that all the eigenvalues of $\mathbf{J}(t)$ lie in an annular neighborhood of radius $\beta\gamma\|\mathbf{W}_r\|$ of the circle centered in the origin of radius $1-\beta$, where $\gamma = \|\mathbf{D}(t)\|$. Moreover, it is possible to bound the maximum local Lyapunov exponent (MLLE) of eq. 1, denoted by Λ, as $\ln\big(1-\beta(\|\mathbf{W}_h\|+1)\big) \leq \Lambda \leq \ln\big(1+\beta(\|\mathbf{W}_h\|-1)\big)$. Leveraging its arbitrary proximity to critical dynamics, ECN sensibly improves over ESN in terms of short-term memory (see [2]) and auto-regressive modeling of complex dynamics (see Figure 1).

Our second proposal stems from the analysis of neural networks architectures under the prism of dynamical systems. In our case, we focus our analysis on the operation of a continuous-time recurrent neural system modeled by the following ODE: $\mathbf{h}'(t) = \phi(\mathbf{W}_h\,\mathbf{h}(t) + \mathbf{W}_x\,\mathbf{x}(t) + \mathbf{b})$. We impose two types of constraints on this ODE: stability, which is essential to preventing input perturbations from causing poor generalization, and non-dissipativity, which is critical to avoid the development of lossy dynamics that can lead to catastrophic forgetting of past inputs during state evolution. We thereby seek for a critical condition under which the eigenvalues of the Jacobian of our ODE are featured by ≈ 0 real parts. Interestingly, a simple *architectural* way of achieving this condition, rather than learning it, is to impose

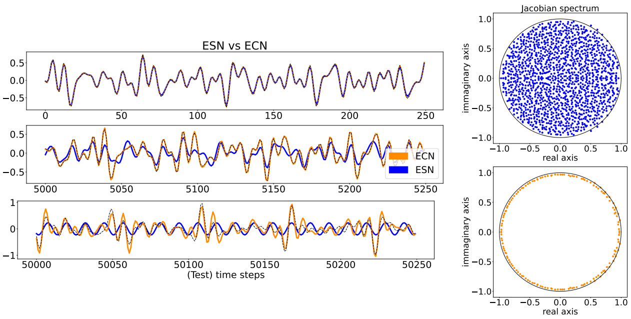

Figure 1: Autoregressive modeling of the multiple superimposed oscillator (MSO) task with 8 frequencies. **Left**: A large fine-tuned ESN of 1500 neurons and a small fine-tuned ECN with 150 neurons have been trained to reproduce with the feedback of the output the MSO8 signal (dashed line). The output signals generated by ESN (blue) and ECN (orange) after various time intervals Δt of running in auto-generation mode are plotted, from top to bottom, $\Delta t = 0, 5000, 50000$. **Right**: The eigenvalues of the Jacobians of the corresponding trained ESN model and ECN model.

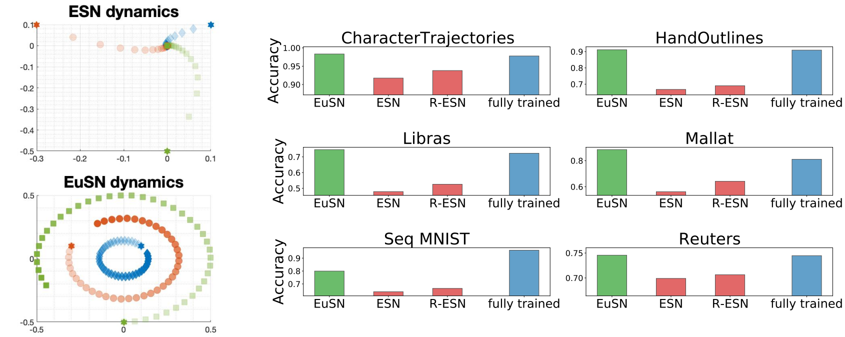

Figure 2: **Left**: Autonomous reservoir dynamics around the origin for a stable ESN (top), and an EuSN (bottom). **Right**: Test set accuracy on a selection of time-series classification benchmarks, achieved by EuSN, ESN, ESN with ring-reservoir (R-ESN), and fully trainable RNNs (best results among vanilla RNN, Antisymmetric-RNN and GRU). See further details in [3].

an antisymmetric structure to the recurrent weight matrix, i.e., $\mathbf{W}_h = \mathbf{W} - \mathbf{W}^T$. In this case, the eigenvalues of the ODE are by construction on the imaginary axis, and the continuous system dynamics operate intrinsically at the edge of criticality. The weight values involved in the ODE defined above are randomly chosen in line with the RC practice. To develop our discrete-time reservoir dynamics, we solve the ODE numerically by using the forward Euler method, resulting in the following reservoir equation:

$$\mathbf{h}(t) = \mathbf{h}(t-1) + \varepsilon\phi((\mathbf{W} - \mathbf{W}^T - \gamma\mathbf{I})\,\mathbf{h}(t-1) + \mathbf{W}_x\,\mathbf{x}(t) + \mathbf{b}), \tag{2}$$

where ε and γ are two hyper-parameters denoting, respectively, the step size of integration and the diffusion coefficient (for stabilizing the discretization). The resulting method is called *Euler State Network* (EuSN) [3]. As for the case of ECNs, the values of the hyper-parameters can be used to control the proximity of the EuSN dynamics to the critical behavior. In this case, with $\mathbf{D}(t) = diag(\phi'((\mathbf{W} - \mathbf{W}^T - \gamma\mathbf{I})\,\mathbf{h}(t-1) + \mathbf{W}_x\,\mathbf{x}(t) + \mathbf{b}))$, the Jacobian of the system in eq. 2 takes the form $\mathbf{J}(t) = \mathbf{I} + \varepsilon\mathbf{D}(t)(\mathbf{W} - \mathbf{W}^T) - \varepsilon\gamma\mathbf{D}(t)$. The eigenvalues of $\mathbf{J}(t)$ are bounded within a ball of radius $\varepsilon(\|\mathbf{W} - \mathbf{W}^T\|) + \gamma)$ around unity, causing the resulting MLLE to be constrained as follows: $\ln(1 - \varepsilon(\|\mathbf{W} - \mathbf{W}^T\| + \gamma)) \leq \Lambda \leq \ln(1 + \varepsilon(\|\mathbf{W} - \mathbf{W}^T\| + \gamma))$. While based on randomized weights as the conventional ESN, EuSN does not suffer from lossy dynamics, and it is able to effectively preserve input information through several time-steps (see Figure 2, left). Relevantly, this new type of architectural bias for recurrent systems allows EuSNs to sensibly reduce the accuracy gap with fully trainable RNNs in tasks involving classification of time-series (see Figure 2, right). The EuSN approach has also shown its effectiveness in broader contexts related to learning on graphs (see [4]), allowing to clearly outperform state-of-the-art models in deep graph networks.

References

[1] Andrea Ceni. Random orthogonal additive filters: a solution to the vanishing/exploding gradient of deep neural networks. *arXiv preprint arXiv:2210.01245*, 2022.

[2] Andrea Ceni and Claudio Gallicchio. Orthogonality in additive reservoir computing. In *ESANN*, 2022.

[3] Claudio Gallicchio. Euler state networks: Non-dissipative reservoir computing. *arXiv preprint arXiv:2203.09382*, 2023.

[4] Alessio Gravina, Davide Bacciu, and Claudio Gallicchio. Anti-Symmetric DGN: a stable architecture for Deep Graph Networks. In *Proceedings of ICLR*, 2023. arXiv preprint arXiv:2210.09789.

[5] Herbert Jaeger and Harald Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *science*, 304(5667):78–80, 2004.

[6] Mantas Lukoševičius and Herbert Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127–149, 2009.

[7] Gouhei Tanaka, Toshiyuki Yamane, Jean Benoit Héroux, Ryosho Nakane, Naoki Kanazawa, Seiji Takeda, Hidetoshi Numata, Daiju Nakano, and Akira Hirose. Recent advances in physical reservoir computing: A review. *Neural Networks*, 115:100–123, 2019.