Weakly Nonlinear Theory of Echo-State Networks

Vincent Hakim[1], Alain Karma[2]

[1] Laboratoire de Physique de l'Ecole Normale Supérieure, CNRS, Ecole Normale Supérieure, PSL University, Sorbonne Université, Université Paris-Diderot, Paris, France
[2] Physics Department and Center for Interdisciplinary Research in Complex Systems, Northeastern University, Boston, MA, USA
vincent.hakim@ens.fr, a.karma@northeastern.edu

Echo-state-networks (ESN) [1,2] are known for their remarkable property of producing prescribed autonomous dynamics by learning a simple feedback to a large recurrent network. They have served as conceptual models for how the brain produces movement [2] and continue to inspire the design of various artificial dynamical devices [3]. However, the principles that underly their seemingly miraculous success remain incompletely understood. Here, we develop a weakly nonlinear theory of ESN that explains this success in the regime where the recurrent network evolution is stable and the feedback is weak.

We analyze the prototypical network of N recurrent units (see Figure. 1) described by

$$\frac{d\mathbf{x_i}}{dt} = -\mathbf{x_i} + g\sum_j M_{ij} r(x_j) + b_i \sum_j w_j r(x_j) \tag{1}$$

where $\mathbf{M}$ is a gaussian random matrix , b is a random "feedback" vector and $r(x)$ is a nonlinear "activation" function, taken as usual to be $\tanh(x)$ for simplicity. The aim of the learning procedure is to choose the vector of output synaptic weights $\mathbf{w}$ such that $z(t) = \sum_j w_j r[x_j(t)])$ reproduces, as best as possible, a desired function $f(t)$.

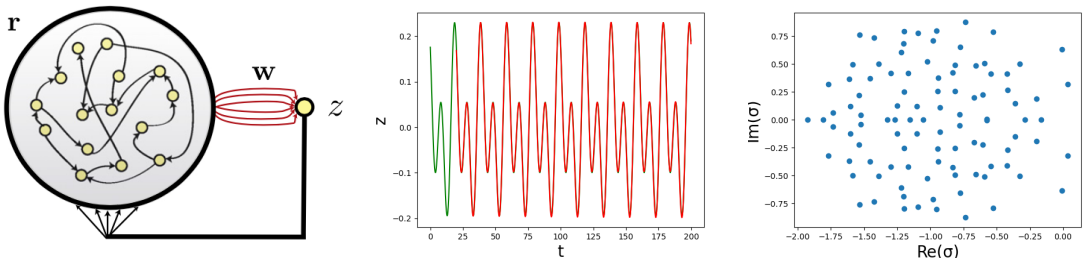

Figure 1. Left: Network schematic (adapted from [2]); (Center) One period of function f to be learnt (green) and network approximation z (red); (Right) Spectrum of the linear dynamics (N=100).

When the amplitude of $f(t)$ is small, the ESN operates in a weakly nonlinear regime. We show that the learning of the output weights $\mathbf{w}$ amounts to i) positioning the eigenvalues of the linear dynamics at locations close to those of the Fourier frequencies of $f(t)$ and ii) constraining the weakly nonlinear dynamics to converge to the correct amplitudes for these Fourier modes. We further provide analytical predictions for when the nonlinear dynamical attractors are stable corresponding to successful learning that is strongly N-dependent. We expect our theory to be applicable to various generalizations of Eq. (1), for instance based on other dynamical units [3], or for more structured networks with different unit classes.

References

1. H. Jaeger and H. Haas, Science **304**, 78 (2004)
2. D. Sussillo and L. F. Abbott, Neuron **63**, 544 (2009); D. Sussillo, et al, Nat. Neuro. **18**, 1025 (2015)
3. K. Nakajima and I. Fischer, *Reservoir computing* (Springer, 2021)