2016 IEEE International Conference on Computational Science and Engineering, IEEE International Conference on Embedded and Ubiquitous Computing, and International Symposium on Distributed Computing and Applications to Business, Engineering and Science

Time-delay reservoir computers and high-speed information processing capacity

Lyudmila Grigoryeva*, Julie Henriques†, Laurent Larger‡ and Juan-Pablo Ortega§¶

*Department of Mathematics and Statistics, Universität Konstanz, Box 146. D-78457 Konstanz. Germany.

†Methodology and Quality of Life in Oncology Unit, CHRU Besançon. F-25000 Besançon. France.

‡FEMTO-ST, University Bourgogne Franche-Comté. 15, Avenue des Montboucons. F-25030 Besançon cedex. France.

§Faculty of Mathematics and Statistics, Universität Sankt Gallen, Bodanstrasse 6, CH-9000 Sankt Gallen. Switzerland.

¶Centre National de la Recherche Scientifique (CNRS). France

***Abstract*—The aim of this presentation is to show how various ideas coming from the nonlinear stability theory of functional differential systems, stochastic modeling, and machine learning, can be put together in order to create an approximating model that explains the working mechanisms behind a certain type of reservoir computers. Reservoir computing is a recently introduced brain-inspired machine learning paradigm capable of excellent performances in the processing of empirical data. We focus on time-delay based reservoir computers that have been physically implemented using optical and electronic systems and have shown unprecedented data processing rates. Reservoir computing is well-known for the ease of the associated training scheme but also for the problematic sensitivity of its performance to architecture parameters. The reservoir design problem is addressed, which remains the biggest challenge in the applicability of this information processing scheme. Our results use the information available regarding the optimal reservoir working regimes in order to construct a functional link between the reservoir parameters and its performance. This function is used to explore various properties of the device and to choose the optimal reservoir architecture, thus replacing the tedious and time consuming parameter scannings used so far in the literature.**

***Keywords*-Reservoir computing; echo state networks; neural computing; time-delay reservoir; memory capacity; architecture optimization.**

I. Introduction

Reservoir computing (RC) [1], [2], [3], [4] is a new brain-inspired machine learning paradigm that has shown much potential in overcoming the physical limitations of the Turing or von Neumann machine methods implemented in most computational systems. In this talk we discuss a model introduced in [5] for a particular implementation of this paradigm suggested by the intrinsic information processing abilities of dynamical systems. More specifically, we will focus on a RC realization based on the sampling of the solution of a time-delay differential equation [6], [7]. We refer to this combination as time-delay reservoirs (TDRs). Physical implementations of this scheme carried out with dedicated hardware are already available and have shown excellent performances in the processing of empirical data: spoken digit recognition [8], [9], [10], [11], [12], the NARMA model identification task [13], [6], continuation of chaotic time series, and volatility forecasting [14]. A recent example that shows the potential of this combination are the results in [12] where an optoelectronic implementation of a TDR is capable of achieving the lowest documented error in the speech recognition task at unprecedented speed in an experiment design in which digit and speaker recognition are carried out in parallel.

A major advantage of RC is the linearity of its training scheme. This choice makes its implementation easy when compared to more traditional machine learning approaches like recursive neural networks, which usually require the solution of convoluted and sometimes ill-defined optimization problems. In exchange, as it can be seen in most of the references quoted above, the system performance is not robust with respect to the choice of the parameter values $\boldsymbol{\theta}$ of the nonlinear kernel used to construct the RC (see below). This observation makes the kernel parameter optimization a very important step in the RC design and has motivated the introduction of alternative parallel-based architectures [15], [14] to tackle this difficulty.

The main contribution that we will here discuss is an approximated model that, to our knowledge, provides the first rigorous analytical description of the delay-based RC performance. This powerful theoretical tool can be used to systematically study the delay-based RC properties and to replace the trial and error approach in the choice of architecture parameters by well structured optimization problems.

II. Time-delay reservoirs (TDR)

TDRs are based on the interaction of the time-dependent input signal $z(t) \in \mathbb{R}$ that we are interested in, with the solution space of a non-autonomous time-delay differential equation of the form

$$\dot{x}(t) = -x(t) + f(x(t-\tau), I(t), \boldsymbol{\theta}), \qquad (1)$$

where f is a nonlinear kernel that depends on the K parameters in the vector $\boldsymbol{\theta} \in \mathbb{R}^K$, $\tau > 0$ is the delay, $x(t) \in \mathbb{R}$, and $I(t) \in \mathbb{R}$ is obtained using a temporal multiplexing over the delay period of the input signal $z(t)$ that we explain later on. The choice of nonlinear kernel is determined by the intended physical implementation of the computing system; we focus on two parametric sets of kernels that have already been explored in the literature,

978-1-5090-3593-9/16 $31.00 © 2016 IEEE
DOI 10.1109/CSE-EUC-DCABES.2016.230

namely, the Mackey-Glass [16] and the Ikeda [17] families. These kernels were used for reservoir computing purposes in the RC electronic and optic realizations in [9] and [10], respectively.

In order to visualize the TDR construction using a neural networks approach it is convenient, as in [9], to consider the Euler time-discretization of (1) with integration step $d := \tau/N$, namely,

$$\frac{x(t)-x(t-d)}{d} = -x(t) + f(x(t-\tau), I(t), \boldsymbol{\theta}). \quad (2)$$

The design starts with the choice of a number $N \in \mathbb{N}$ of virtual neurons and of an adapted input mask $\mathbf{c} \in \mathbb{R}^N$. Next, the input signal $z(t)$ at a given time t is multiplexed over the delay period by setting $\mathbf{I}(t) := \mathbf{c}z(t) \in \mathbb{R}^N$. We then organize it, as well as the solutions of (2), in neuron layers $\mathbf{x}(t)$ parametrized by a discretized time $t \in \mathbb{Z}$ by setting

$$x_i(t) := x(t\tau - (N-i)d), \quad I_i(t) := I(t\tau - (N-i)d),$$

$i \in \{1, \ldots, N\}, \quad t \in \mathbb{Z}$, where $x_i(t)$ and $I_i(t)$ stand for the ith-components of the vectors $\mathbf{x}(t)$ and $\mathbf{I}(t)$, respectively, with $t \in \mathbb{Z}$. With this convention, the solutions of (2) are described by the following recursive relation:

$$x_i(t) := e^{-\xi} x_{i-1}(t) + (1-e^{-\xi}) f(x_i(t-1), I_i(t), \boldsymbol{\theta}), \quad (3)$$

with $x_0(t) := x_N(t-1)$, and $\xi := \log(1+d)$. These recursions uniquely determine a smooth map $F : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^K \to \mathbb{R}^N$ that specifies the neuron values as a recursion on the neuron layers via an expression of the form

$$\mathbf{x}(t) = F(\mathbf{x}(t-1), \mathbf{I}(t), \boldsymbol{\theta}), \quad (4)$$

where F is constructed out of the nonlinear kernel map f that depends on the K parameters in the vector $\boldsymbol{\theta}$; F is referred to as the reservoir map. The construction of the TDR computer is finalized by connecting the reservoir output to a linear readout $\mathbf{W}_{\mathrm{out}} \in \mathbb{R}^N$ that is calibrated using a training sample by minimizing the associated task mean square error via a (ridge regularized) linear regression.

III. Optimal performance: stability and unimodality

The performance of the TDR for a given task is much dependent on the value of the kernel parameters $\boldsymbol{\theta}$ and, in some cases, on the entries of the input mask $\mathbf{c}$. When comparing RC to standard neural networks and thinking of it as a machine learning paradigm, the RC training phase can be assimilated to the determination of both the linear readout $\mathbf{W}_{\mathrm{out}}$ (straightforward in this case using a linear regression) and the optimal parameters $\boldsymbol{\theta}$. Unlike the situation encountered in the neural networks context for which efficient training algorithms have been developed over the years, the optimal parameters $\boldsymbol{\theta}$ are usually determined in the RC context by trial and error or using computationally costly systematic scannings.

In [5] (see further extensions in [18], [19]) we constructed an approximate model that we used to to establish a functional link between the RC performance and the parameters $\boldsymbol{\theta}$ and the input mask values $\mathbf{c}$ that is based on the observation that the optimal RC performance is always obtained when the TDR is working in a stable unimodal regime, that is, the reservoir is initialized at a stable equilibrium of the autonomous system ($I(t) = 0$) associated to (1) and the mean and variance of the input signal $I(t)$ are designed using the input mask $\mathbf{c}$ so that the reservoir output remains around it and does not visit other stable equilibria or dynamical elements (see [5] for empirical and theoretical arguments in this direction). In the context of recent successful physical realizations of RC, experimental parameters are systematically chosen so that these conditions are satisfied (for example in [9], [10] these conditions are ensured via a proper tuning of the gain of the delayed feedback function. This approach differs from the one in [12], where the conditions are met by choosing a laser injection current strictly smaller but close to the lasing threshold, as well as by using a moderate feedback, which prevents eventual self sustained external cavity mode oscillations. An additional important observation suggested by all these experimental setups is the need for a nonlinearity at the level of the input injection. In [9], [10] this feature is obtained using a strong enough input signal amplitude and via the transformation associated to the nonlinear delayed feedback. In [12] the delayed feedback is linear but an external Mach-Zehnder modulator is used that implicitly provides a nonlinear transformation of the input signal as it is optically seeded through the nonlinear electro-optic modulation transfer function of the Mach-Zehnder.

Consider an equilibrium $x_0 \in \mathbb{R}$ of the continuous time model (1). It can be shown using a Lyapunov-Krasovskiy-type analysis (see [5]) that the asymptotic stability of x_0 is guaranteed as long as $|\partial_x f(x_0, 0, \boldsymbol{\theta})| < 1$.

IV. The approximating model and the nonlinear memory capacity of the TDR

As optimal operation is attained when the TDR functions in a unimodal fashion around an asymptotically stable steady state, we can approximate it by its partial linearization with respect to the delayed self feedback term at that point and keeping the nonlinearity for the input injection. For statistically independent input signals of the type used to compute nonlinear memory capacities of the type introduced in [20], this approximation allows us to visualize the TDR as a N-dimensional (N is the number of neurons) vector autoregressive stochastic process of order one (we denote it as VAR(1)) for which the value of the associated nonlinear memory capacities can be explicitly computed. The quality of this approximation at the time of evaluating the memory capacities of the original system is excellent and the resulting function can be hence used for RC optimization purposes regarding the nonlinear kernel parameter values $\boldsymbol{\theta}$ and the input mask $\mathbf{c}$.

Consider a stable equilibrium $x_0 \in \mathbb{R}$ of the autonomous system associated to (1) or, equivalently, a stable fixed point of (4) of the form $\mathbf{x}_0 := (x_0, \ldots, x_0)^\top \in \mathbb{R}^N$. If we approximate (4) by its partial linearization at $\mathbf{x}_0$ with respect to the delayed self feedback and by the R-order Taylor series expansion of the functional that describes the signal injection, we obtain an expression of the form:

$$\mathbf{x}(t) = F(\mathbf{x}_0, \mathbf{0}_N, \boldsymbol{\theta}) + A(\mathbf{x}_0, \boldsymbol{\theta})(\mathbf{x}(t-1) - \mathbf{x}_0) + \boldsymbol{\varepsilon}(t), \quad (5)$$

where $A(\mathbf{x}_0, \boldsymbol{\theta}) := D_{\mathbf{x}} F(\mathbf{x}_0, \mathbf{0}_N, \boldsymbol{\theta})$ is the linear connectivity matrix and $\boldsymbol{\varepsilon}(t)$ is given by:

$$\begin{aligned}\boldsymbol{\varepsilon}(t) = (1 - e^{-\xi})(&q_R\left(z(t), c_1\right), \\ &q_R\left(z(t), c_1, c_2\right), \ldots, q_R\left(z(t), c_1, \ldots, c_N\right))^\top, \quad (6)\end{aligned}$$

with

$$\begin{aligned}&q_R\left(z(t), c_1, \ldots, c_r\right) := \\ &\qquad \sum_{i=1}^{R} \frac{z(t)^i}{i!} (\partial_I^{(i)} f)(x_0, 0, \boldsymbol{\theta}) \sum_{j=1}^{r} e^{-(r-j)\xi} c_j^i, \quad (7)\end{aligned}$$

and $(\partial_I^{(i)} f)(x_0, 0, \boldsymbol{\theta})$ is the ith order partial derivative of the nonlinear kernel f with respect to the second argument $I(t)$, evaluated at the point $(x_0, 0, \boldsymbol{\theta})$.

If we now use as input signal $z(t)$ independent and identically distributed random variables with mean 0 and variance σ_z^2 (we denote it by $\{z(t)\}_{t\in\mathbb{Z}} \sim \mathrm{IID}(0, \sigma_z^2)$) then the recursion (5) makes the reservoir layer dynamics $\{\mathbf{x}(t)\}_{t\in\mathbb{Z}}$ into a discrete time random process that, as we show in what follows, is the solution of a N-dimensional vector autoregressive model of order 1 (VAR(1)). Indeed, it is easy to see that the assumption $\{z(t)\}_{t\in\mathbb{Z}} \sim \mathrm{IID}(0, \sigma_z^2)$ implies that $\{\mathbf{I}(t)\}_{t\in\mathbb{Z}} \sim \mathrm{IID}(\mathbf{0}_N, \Sigma_I)$, with $\Sigma_I := \sigma_z^2 \mathbf{c}^\top \mathbf{c}$, and that $\{\boldsymbol{\varepsilon}(t)\}_{t\in\mathbb{Z}}$ is a family of N-dimensional independent and identically distributed random variables with mean $\boldsymbol{\mu}_\varepsilon$ and covariance matrix Σ_ε given by the following expressions:

$$\begin{aligned}\boldsymbol{\mu}_\varepsilon = \mathrm{E}\left[\boldsymbol{\varepsilon}(t)\right] = (1 - e^{-\xi})(&q_R\left(\mu_z, c_1\right), \\ &q_R\left(\mu_z, c_1, c_2\right), \ldots, q_R\left(\mu_z, c_1, \ldots, c_N\right))^\top, \quad (8)\end{aligned}$$

where the polynomial q_R is the same as in (7) and where we use the convention that the powers $\mu_z^i := \mathrm{E}\left[z(t)^i\right]$, for any $i \in \{1, \ldots, R\}$ and with $\mathrm{E}[\cdot]$ denoting the mathematical expectation. Additionally, $\Sigma_\varepsilon := \mathrm{E}\left[(\boldsymbol{\varepsilon}(t) - \boldsymbol{\mu}_\varepsilon)(\boldsymbol{\varepsilon}(t) - \boldsymbol{\mu}_\varepsilon)^\top\right]$ has entries determined by the relation:

$$\begin{aligned}(\Sigma_\varepsilon)_{ij} = (1 - e^{-\xi})^2 (&(q_R(\cdot, c_1, \ldots, c_i) \cdot q_R(\cdot, c_1, \ldots, c_j))(\mu_z) \\ &- q_R(\mu_z, c_1, \ldots, c_i) q_R(\mu_z, c_1, \ldots, c_j)),\end{aligned}$$

where the first summand stands for the multiplication of the polynomials $q_R(\cdot, c_1, \ldots, c_i)$ and $q_R(\cdot, c_1, \ldots, c_j)$ and the subsequent evaluation of the resulting polynomial at μ_z, and the second one is made out of the multiplication of the evaluation of the two polynomials.

Using these observations, we can consider (5) as the prescription of a VAR(1) model driven by the independent noise $\{\boldsymbol{\varepsilon}(t)\}_{t\in\mathbb{Z}}$. If the nonlinear kernel f satisfies the generic condition that the polynomial in u given by $\det\left(\mathbb{I}_N - A(\mathbf{x}_0, \boldsymbol{\theta})u\right)$, does not have roots in and on the complex unit circle, then (5) has a second order stationary solution $\{\mathbf{x}(t)\}_{t\in\mathbb{Z}}$ with time-independent mean given by

$$\begin{aligned}\boldsymbol{\mu}_x &= \mathrm{E}\left[\mathbf{x}(t)\right] \\ &= (\mathbb{I}_N - A(\mathbf{x}_0, \boldsymbol{\theta}))^{-1} (F(\mathbf{x}_0, \mathbf{0}_N, \boldsymbol{\theta}) - A(\mathbf{x}_0, \boldsymbol{\theta})\mathbf{x}_0 + \boldsymbol{\mu}_\varepsilon)\end{aligned}$$

and an also time independent autocovariance function $\Gamma(k) := \mathrm{E}\left[\left(\mathbf{x}(t) - \boldsymbol{\mu}_x\right)\left(\mathbf{x}(t-k) - \boldsymbol{\mu}_x\right)^\top\right]$, $k \in \mathbb{Z}$, recursively determined the Yule-Walker equations:

$$\mathrm{vec}(\Gamma(0)) = \left(\mathbb{I}_{N^2} - A(\mathbf{x}_0, \boldsymbol{\theta}) \otimes A(\mathbf{x}_0, \boldsymbol{\theta})\right)^{-1} \mathrm{vec}(\Sigma_\varepsilon).$$

The moments that we just spelled out are all that is needed in order to characterize the memory capacities of the RC.

Figure 1 represents the normalized mean square error (NMSE) surfaces (which amounts to one minus the capacity) exhibited by a specific reservoir on a memory task (see caption) when evaluated using the theoretical formula based on the previously described approximation as well as with Monte Carlo estimations (using 50,000 occurrences each) of the NMSE exhibited by the discrete and continuous time TDRs, respectively. The time-evolution of the time-delay differential equation (continuous time model) was simulated using a Runge-Kutta fourth-order method with a discretization step equal to $d/5$. A quick inspection of Figure 1 reveals the ability of the model to accurately capture most of the details of the error surface and, most importantly, the location in parameter space where optimal performance is attained; it is very easy to visualize in this particular example how sensitive the magnitude of the error and the corresponding memory capacity are to the choice of parameters and how small in size the region in parameter space associated with acceptable operation performance may be.

Acknowledgment

We acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493), the ANR "BIPHOPROC" project (ANR-14-OHRI-0002-02), the European project PHOCUS (FP7 Grant No. 240763), the Labex ACTION program (Contract No. ANR-11-LABX-01-01), and Deployment S.L. LG acknowledges financial support from the Faculty for the Future Program of the Schlumberger Foundation.

References

[1] H. Jaeger, "The 'echo state' approach to analysing and training recurrent neural networks," *German National Research Center for Information Technology*, 2001.

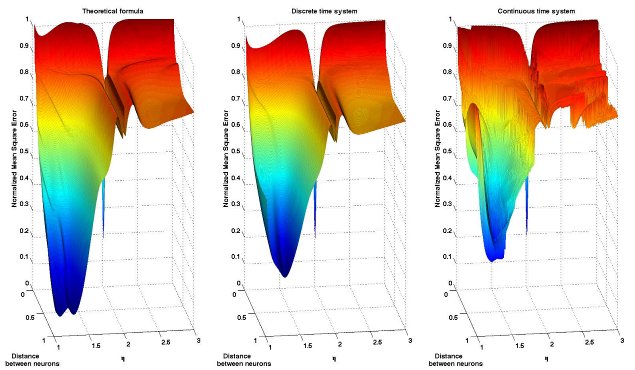

Figure 1. Error surfaces exhibited by a Mackey-Glass kernel based reservoir computer in a 6-lag quadratic memory task, as a function of the distance between neurons and the parameter η. The points in the surfaces of the middle and right panels are the result of Monte Carlo evaluations of the NMSE exhibited by the discrete and continuous time TDRs, respectively. The left panel was constructed using the capacity formula that is obtained as a result of modeling the reservoir with an approximating VAR(1) model. The computational convenience of such formula can be visualized by noticing that each point in the middle and right panels took 37 and 41 seconds, respectively, to be estimated using a computer code written down in a high level programming language running on a single 2.7 GHz Intel i7 core; the same computation using the approximating model in the left panel took only 1.1 seconds.

[2] H. Jaeger and H. Haas, "Harnessing Nonlinearity: Predicting Chaotic Systems and Saving Energy in Wireless Communication," *Science*, vol. 304, no. 5667, pp. 78–80, 2004.

[3] W. Maass, T. Natschläger, and H. Markram, "Real-time computing without stable states: a new framework for neural computation based on perturbations," *Neural Computation*, vol. 14, pp. 2531–2560, 2002.

[4] W. Maass, "Liquid state machines: motivation, theory, and applications," in *Computability In Context: Computation and Logic in the Real World*, S. S. Barry Cooper and A. Sorbi, Eds., 2011, ch. 8, pp. 275–296.

[5] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, "Optimal nonlinear information processing capacity in delay-based reservoir computers," *Scientific Reports*, vol. 5, no. 12858, pp. 1–11, 2015.

[6] A. Rodan and P. Tino, "Minimum complexity echo state network." *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 22, no. 1, pp. 131–44, jan 2011.

[7] J. M. Gutiérrez, D. San-Martín, S. Ortín, and L. Pesquera, "Simple reservoirs with chain topology based on a single time-delay nonlinear node," in *20th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*, 2012, pp. 13–18.

[8] H. Jaeger, M. Lukoševičius, D. Popovici, and U. Siewert, "Optimization and applications of echo state networks with leaky-integrator neurons," *Neural Networks*, vol. 20, no. 3, pp. 335–352, 2007.

[9] L. Appeltant, M. C. Soriano, G. Van der Sande, J. Danckaert, S. Massar, J. Dambre, B. Schrauwen, C. R. Mirasso, and I. Fischer, "Information processing using a single dynamical node as complex system," *Nature Communications*, vol. 2, p. 468, jan 2011.

[10] L. Larger, M. C. Soriano, D. Brunner, L. Appeltant, J. M. Gutierrez, L. Pesquera, C. R. Mirasso, and I. Fischer, "Photonic information processing beyond Turing: an optoelectronic implementation of reservoir computing," *Optics Express*, vol. 20, no. 3, p. 3241, jan 2012.

[11] Y. Paquot, F. Duport, A. Smerieri, J. Dambre, B. Schrauwen, M. Haelterman, and S. Massar, "Optoelectronic reservoir computing," *Scientific reports*, vol. 2, p. 287, jan 2012.

[12] D. Brunner, M. C. Soriano, C. R. Mirasso, and I. Fischer, "Parallel photonic information processing at gigabyte per second data rates using transient states," *Nature Communications*, vol. 4, no. 1364, 2013.

[13] A. F. Atiya and A. G. Parlos, "New results on recurrent network training: unifying the algorithms and accelerating convergence," *IEEE transactions on neural networks / a publication of the IEEE Neural Networks Council*, vol. 11, no. 3, pp. 697–709, jan 2000.

[14] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, "Stochastic time series forecasting using time-delay reservoir computers: performance and universality," *Neural Networks*, vol. 55, pp. 59–71, 2014.

[15] S. Ortin, L. Pesquera, and J. M. Gutiérrez, "Memory and nonlinear mapping in reservoir computing with two uncoupled nonlinear delay nodes," in *Proceedings of the European Conference on Complex Systems*, T. Gilbert, M. Kirkilionis, and G. Nicolis, Eds. Springer International Publishing Switzerland, 2012, pp. 895–899.

[16] M. C. Mackey and L. Glass, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, pp. 287–289, 1977.

[17] K. Ikeda, "Multiple-valued stationary state and its instability of the transmitted light by a ring cavity system," *Optics Communications*, vol. 30, no. 2, pp. 257–261, aug 1979.

[18] L. Grigoryeva, J. Henriques, L. Larger, and J.-P. Ortega, "Nonlinear memory capacity of parallel time-delay reservoir computers in the processing of multidimensional signals," *Neural Computation*, vol. 28, pp. 1411–1451, 2016. [Online]. Available: http://arxiv.org/abs/1510.03891

[19] L. Grigoryeva, J. Henriques, and J.-P. Ortega, "Quantitative evaluation of the performance of discrete-time reservoir computers in the forecasting, filtering, and reconstruction of stochastic stationary signals," *Preprint*, 2015.

[20] J. Dambre, D. Verstraeten, B. Schrauwen, and S. Massar, "Information processing capacity of dynamical systems," *Scientific reports*, vol. 2, no. 514, 2012.