Optimal nonlinear information processing capacity in delay-based reservoir computers

Lyudmila Grigoryeva[1], Julie Henriques[1,2], Laurent Larger[1], and Juan-Pablo Ortega[3]

[1]Université de Franche-Comté [2]Cegos Deployment [3]CNRS, Université de Franche-Comté

Context and objectives

Reservoir computing is a recently introduced brain-inspired machine learning paradigm. We focus on time-delay reservoir (TDR) computers that have been physically implemented using optical and electronic systems and show excellent computational performances at unprecedented information processing speeds. TDRs are known for their easy-to-implement training but also for their problematic sensitivity to architecture parameters. We address the reservoir design problem, which remains the biggest challenge at the time of applying the reservoir computing techniques to sophisticated machine learning tasks. More specifically, the information available regarding the optimal operating regimes of a reservoir is used to construct a functional link between its parameters and its performance. This function is then used to explore various properties of the device and to choose the optimal architecture, thus replacing tedious and time consuming parameter scannings by well-structured optimization problems.

Architecture of the time-delay reservoir (TDR) computer

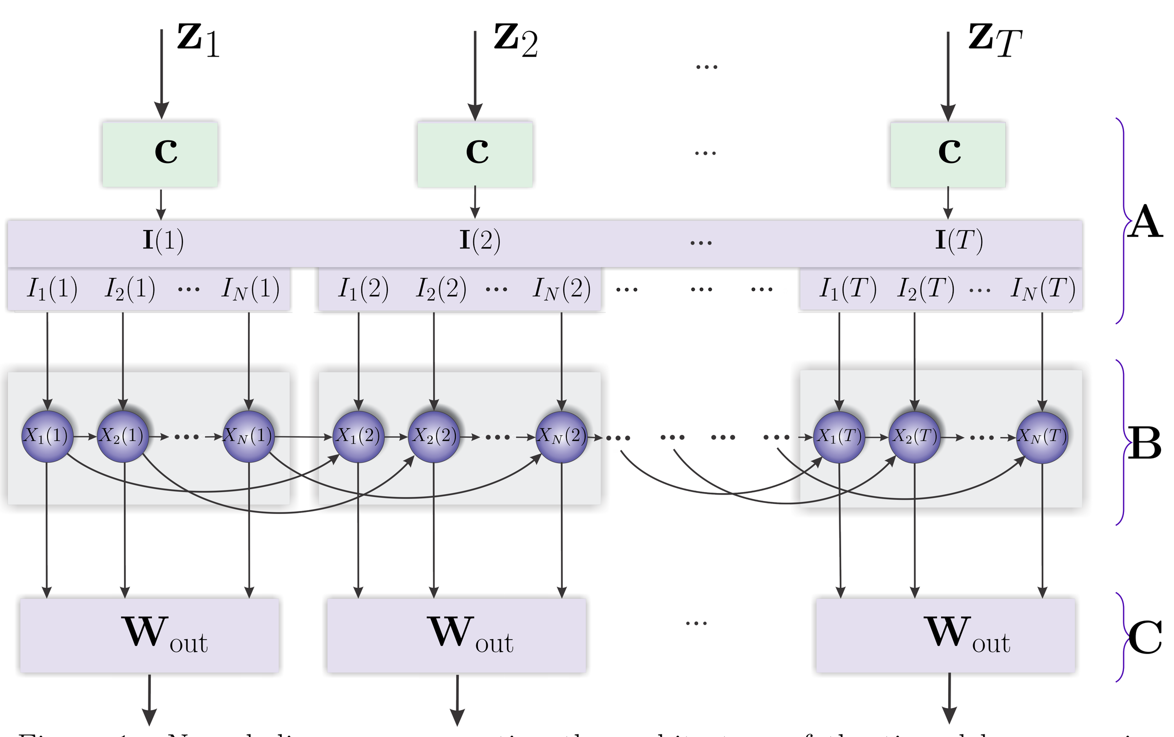

Figure 1. Neural diagram representing the architecture of the time-delay reservoir (TDR) and the three modules of the reservoir computer (RC): A is the input layer, B is the time-delay reservoir, and C is the readout layer.

Optimal performance: stability and unimodality

Stability plays an essential role in the performance of the time-delay reservoirs.

Stability of the TDR: the **continuous time model** of the TDR is

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

where f is the **nonlinear kernel** function, $\boldsymbol{\theta} \in \mathbb{R}^K$ is the reservoir parameters vector, $\tau > 0$ is the **delay**, $x(t) \in \mathbb{R}$, and $I(t) \in \mathbb{R}$ is obtained via temporal multiplexing over τ of the input signal $z(t)$. We show using a Lyapunov-Krasovskiy functional [2] that given an equilibrium $x_0 \in \mathbb{R}$ of the continuous time model (1) in autonomous regime ($I(t) = 0$), its asymptotic stability is guaranteed whenever there exist $\varepsilon > 0$ and $|k_\varepsilon| < 1$ such that

$$\left(f(x + x_0, 0, \boldsymbol{\theta}) - x_0\right)/x \leq k_\varepsilon \quad \text{for all } x \in (-\varepsilon, \varepsilon).$$

This relation implies that if f is differentiable at x_0 then this point is stable whenever

$$|\partial_x f(x_0, 0, \boldsymbol{\theta})| < 1$$

The **discrete time approximation** of the TDR is

$$x_i(t) := e^{-\xi} x_{i-1}(t) + (1 - e^{-\xi}) f(x_i(t-1), I_i(t), \boldsymbol{\theta}), \qquad (2)$$

with $x_0(t) := x_N(t-1)$, $\xi := \log(1+d)$, $i \in \{1, \ldots, N\}$, $x_i(t)$ is the **ith neuron value of the tth layer of the reservoir**, and d is referred to as the **separation between neurons**. The recursions (2) uniquely determine a **reservoir map** $F : \mathbb{R}^N \times \mathbb{R}^N \times \mathbb{R}^K \to \mathbb{R}^N$ such that

$$\mathbf{x}(t) = F(\mathbf{x}(t-1), \mathbf{I}(t), \boldsymbol{\theta}). \qquad (3)$$

The point $x_0 \in \mathbb{R}$ is an equilibrium of (1) if and only if $\mathbf{x}_0 := (x_0, \ldots, x_0)^\top \in \mathbb{R}^N$ is a fixed point of (3). The asymptotic stability of this fixed point is ensured whenever the **connectivity matrix** $D_{\mathbf{x}} F(\mathbf{x}_0, \mathbf{0}_N, \boldsymbol{\theta})$ has a spectral radius smaller than one, which is the case whenever

$$|\partial_x f(x_0, 0, \boldsymbol{\theta})| < 1. \qquad (4)$$

Conclusions: *Optimal TDR performance is attained when the TDR operates in a unimodal regime around an asymptotically stable state. We find common stability conditions for the continuous and discrete time systems.*

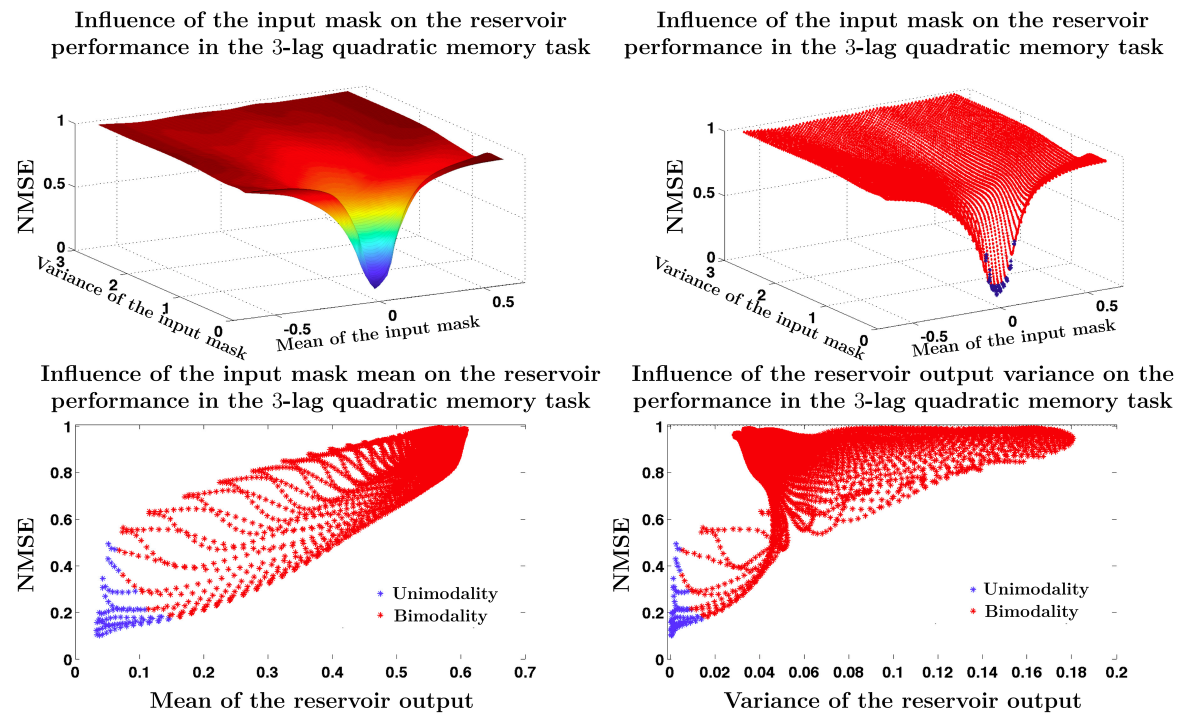

Figure 2. Behavior of the reservoir performance in a quadratic memory task as a function of the mean and the variance of the input mask. The top panels show how the performance degrades very quickly as soon as the mean and the variance of the input mask separate from zero. The bottom panels depict the reservoir performance as a function of the various output means and variances. We have indicated with red markers the cases in which the reservoir visits the stability basin of a contiguous stable equilibrium hence showing how unimodality is associated to optimal performance.

The approximating model and the nonlinear memory capacity

(1) We construct an approximation of the TDR via its partial linearization at the equilibrium point with respect to the delayed self feedback term and respecting the nonlinearity of the input injection.

Consider a stable equilibrium $x_0 \in \mathbb{R}$ of the autonomous system associated to (1) or, equivalently, a stable fixed point $\mathbf{x}_0 := (x_0, \ldots, x_0)^\top \in \mathbb{R}^N$ of (3). We construct the approximation of (3) by using its linearization at $\mathbf{x}_0$ with respect to the delayed self-feedback and its Rth-order Taylor expansion with respect to its dependence on the signal injection:

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

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

$$\boldsymbol{\varepsilon}(t) = (1 - e^{-\xi}) \left(q_R(z(t), c_1), \ldots, q_R(z(t), c_1, \ldots, c_N)\right)^\top,$$

with

$$q_R(z(t), c_1, \ldots, c_r) := \sum_{i=1}^{R} \frac{z(t)^i}{i!} (\partial_I^{(i)} f)(x_0, 0, \boldsymbol{\theta}) \sum_{j=1}^{r} e^{-(r-j)\xi} c_j^i,$$

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

(2) For statistically independent input signals the approximation (5) allows us to visualize the TDR as a N-dimensional vector autoregressive stochastic process of order one (VAR(1), [3]).

Let the input signal be $\{z(t)\}_{t \in \mathbb{Z}} \sim \text{IID}(0, \sigma_z^2)$, then $\{\mathbf{I}(t)\}_{t \in \mathbb{Z}} \sim \text{IID}(\mathbf{0}_N, \Sigma_I)$, with $\Sigma_I := \sigma_z^2 \mathbf{c}^\top \mathbf{c}$, and $\{\boldsymbol{\varepsilon}(t)\}_{t \in \mathbb{Z}} \sim \text{IID}(\boldsymbol{\mu}_\varepsilon, \Sigma_\varepsilon)$ with

$$\boldsymbol{\mu}_\varepsilon = (1 - e^{-\xi}) \left(q_R(\mu_z, c_1), \ldots, q_R(\mu_z, c_1, \ldots, c_N)\right)^\top,$$

where $\mu_z^i := \mathrm{E}\left[z(t)^i\right]$, for any $i \in \{1, \ldots, R\}$, and with $\Sigma_\varepsilon := \mathrm{E}\left[(\boldsymbol{\varepsilon}(t) - \boldsymbol{\mu}_\varepsilon)(\boldsymbol{\varepsilon}(t) - \boldsymbol{\mu}_\varepsilon)^\top\right]$ with entries given by:

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

The process (5) is a VAR(1) model driven by the independent noise $\{\boldsymbol{\varepsilon}(t)\}_{t \in \mathbb{Z}}$ with time-independent mean $\boldsymbol{\mu}_x$ and an autocovariance function $\Gamma(k)$ recursively determined by the Yule-Walker equations [3].

(3) The approximation (5) allows us to write the nonlinear capacities of the TDR as the function of the intrinsic architecture parameters $\boldsymbol{\theta}$ and the input mask $\mathbf{c}$.

A **h-lag memory task** is determined by a (in general nonlinear) function $H : \mathbb{R}^{h+1} \to \mathbb{R}$ that is used to generate a one-dimensional signal $y(t) := H(z(t), z(t-1), \ldots, z(t-h))$ out of the reservoir input. Given a TDR computer, the optimal linear readout $\mathbf{W}_{\text{out}}$ adapted to the memory task H is obtained as the solution of a ridge regression problem with regularization parameter $\lambda \in \mathbb{R}$ in which the covariates are the neuron values corresponding to the reservoir output and the explained variables are the values $\{y(t)\}$ of the memory task function. The H-memory capacity $C_H(\boldsymbol{\theta}, \mathbf{c}, \lambda)$ of the TDR computer characterized by a nonlinear kernel f with parameters $\boldsymbol{\theta}$, an input mask $\mathbf{c}$, and a ridge parameter λ, using a VAR(1) approximation, is given by

$$\begin{aligned}C_H(\boldsymbol{\theta}, \mathbf{c}, \lambda) = &(\mathrm{Cov}(y(t), \mathbf{x}(t)))^\top (\Gamma(0) + \lambda \mathbb{I}_N)^{-1} (\Gamma(0) + 2\lambda \mathbb{I}_N) \\ &\times (\Gamma(0) + \lambda \mathbb{I}_N)^{-1} \mathrm{Cov}(y(t), \mathbf{x}(t)))/\mathrm{var}\,(y(t)). \quad (6)\end{aligned}$$

Conclusions: *The VAR(1) model (5) and its associated moments provide an explicit approximation of the memory capacities of the RC.*

Optimal nonlinear capacity

The h-lag quadratic memory task. Take a quadratic task function of the form $H(\mathbf{z}^h(t)) := \mathbf{z}^h(t)^\top Q \mathbf{z}^h(t)$, for some symmetric $h+1$-dimensional matrix Q. In this case $\mathrm{var}(y(t)) = (\mu_z^4 - \sigma_z^4) \sum_{i=1}^{h+1} Q_{ii}^2 + 4\sigma_z^4 \sum_{i=1}^{h+1} \sum_{j>i}^{h+1} Q_{ij}^2$, and

$$\begin{aligned}\mathrm{Cov}(y(t), x_i(t)) = &(1 - e^{-\xi}) \sum_{j=1}^{h+1} \sum_{r=1}^{N} Q_{jj} (A^{j-1})_{ir} \\ &\times (s_R(\mu_z, c_1, \ldots, c_r) - \sigma_z^2 q_R(\mu_z, c_1, \ldots, c_r)),\end{aligned}$$

where the polynomial s_R on the variable x is defined as $s_R(x, c_1, \ldots, c_r) := x^2 \cdot q_R(x, c_1, \ldots, c_r)$.

Conclusions: *The quality of the approximation (5) 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 intrinsic TDR architecture parameters $\boldsymbol{\theta}$ and the input mask $\mathbf{c}$.*

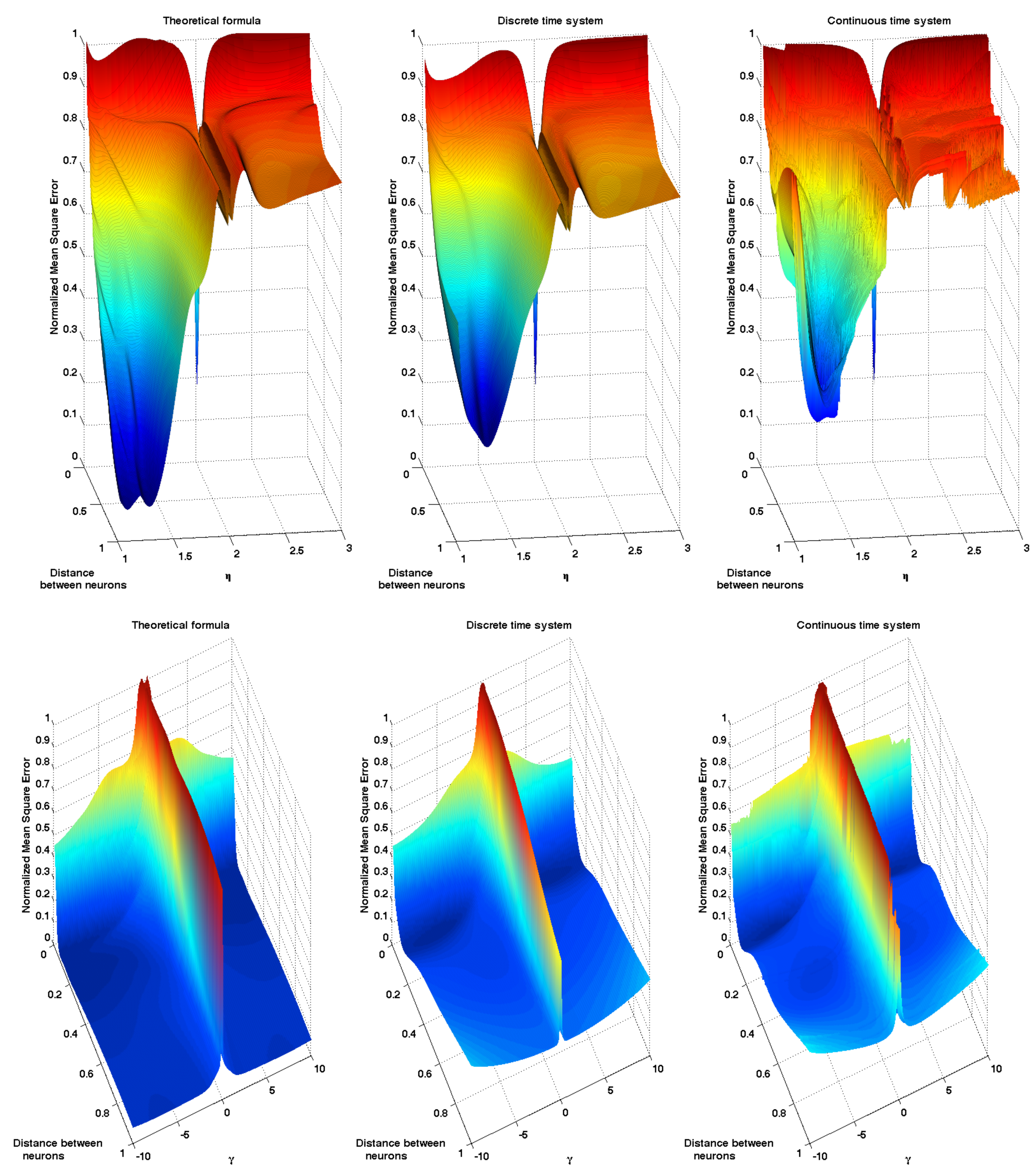

Figure 3. Error exhibited by a TDR computer with a Mackey-Glass kernel in a 6-lag quadratic memory task (the upper figure) and a 3-lag quadratic memory task (the lower figure) as a function of the separation between neurons d and the parameter η (the upper figure), parameter γ (the lower figure), respectively. 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 formula (6) that is obtained as a result of modeling the reservoir with an approximating VAR(1) model.

Perspectives

1. Modeling of the reservoir computing working principle and the design of optimal architectures
   - Extension to non-independent and multivariate signals
   - Theoretical treatment of classification problems
   - Modeling parallel reservoir computers [1] and their properties
   - Use of the reservoir model to establish the reservoir computing defining features
2. Technological implementation of optimal reservoir architectures
3. Applications to biomedical signal classification and forecasting

References

[1] Lyudmila Grigoryeva, Julie Henriques, Laurent Larger, and Juan-Pablo Ortega. Stochastic time series forecasting using time-delay reservoir computers: performance and universality. *Neural Networks*, 55:59–71, 2014.

[2] N. N. Krasovskiy. *Stability of Motion.* Stanford University Press, 1963.

[3] Helmut Lütkepohl. *New Introduction to Multiple Time Series Analysis.* Springer-Verlag, Berlin, 2005.

Acknowledgements

We acknowledge partial financial support of the Région de Franche-Comté (Convention 2013C-5493), 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.

Labex Action Integrated smart systems