Reservoir computing with solitons

Nuno Azevedo Silva, Tiago D Ferreira and Ariel Guerreiro

1 INESC TEC, Centre for Applied Photonics, Rua do Campo Alegre 687, 4169-007 Porto, Portugal
2 Faculdade de Ciencias da Universidade do Porto, Rua do Campo Alegre 687, 4169-007 Porto, Portugal

E-mail: nuno.a.silva@inesctec.pt

Keywords: reservoir computing, nonlinear physics, solitons

Abstract

Reservoir computing is a promising framework that facilitates the approach to physical neuromorphic hardware by enabling a given nonlinear physical system to act as a computing platform. In this work, we exploit this paradigm to propose a versatile and robust soliton-based computing system using a discrete soliton chain as a reservoir. By taking advantage of its tunable governing dynamics, we show that sufficiently strong nonlinear dynamics allows our soliton-based solution to perform accurate regression and classification tasks of non-linear separable datasets. At a conceptual level, the results presented pave a way for the physical realization of novel hardware solutions and have the potential to inspire future research on soliton-based computing using various physical platforms, leveraging its ubiquity across multiple fields of science, from nonlinear optical media to quantum systems.

1. Introduction

Motivated by the search for ultra-fast all-optical technologies, soliton computing is a research line that explores the particle-like behavior of these self-localized wave solutions to implement unconventional computational systems [1–4]. From particle machines [1, 3], logic gates [1] and algorithms [5, 6], past advances reported the potential of the approach in all-optical environments and its Turing completeness [3]. Still, like many efforts in optical computing, the theoretical promise of this ultra-fast and energy-efficient computing solution crashed into the lack of control at the experimental level to compete with the more robust electronic-based von Neumann-like architectures. A workaround to this problem is to pursue new, alternative, and versatile computing frameworks to bridge this technological gap, establishing opportunities to revisit soliton-based solutions in light of new paradigms.

In this context, artificial neural networks established in the last decades as the most successful alternative computing framework. They enjoy a solid reputation across many fields of science not only as algorithms but also as neuromorphic hardware. Concerning the former, reservoir computing (RC) stands as one of the most versatile approaches. Historically, RC appeared as an easier training approach to recurrent neural networks [7, 8], but it was its possible transference to physical hardware that caught the attention due to its simple inner-workings. In short, the framework uses an untrained nonlinear system—the reservoir—to expand an input signal into an output high-dimensional space, where good linear cuts for a given abstract task can be found [9, 10]. Put in this way, the RC strategy enables any nonlinear physical system to work as a hypothetical computing device, bypassing the complicated fine-tuning processes of other approaches (e.g. hardware implementations of feed-forward neural networks) [9, 11]. From buckets of water [12] to quantum systems [13], multiple works recently confirmed this versatility. Regarding all-optical computing, RC-inspired solutions are now enabling photonic systems as computing solutions, capitalizing the native properties of light to significantly enhance or surpass its electronic counterparts [11, 14–17].

Leveraging on these opportunities, this work revisits soliton-based computing taking the RC paradigm as a blueprint. First, we focus on the nonlinear Schrödinger equation (NLSE) to explore how the dynamics of interacting solitons arranged in a one-dimensional chain can act as a nonlinear reservoir. Deriving the Toda lattice equation as an approximate model for the dynamical system, we demonstrate that the ability of...
the reservoir to perform accurate regression and classification tasks is crucially tied to the dimensionality of the output space and to the nonlinear dynamics of the reservoir. Our findings suggest that nonlinear regimes allow for better accuracy in regression tasks and classification nonlinear separable datasets, which aligns with previous results for propagating waves and extend them to classification tasks [18]. With the nonlinear properties easily controlled through the scaling amplitude at the input, our system may foster future research and experimental testbeds, not only in all-optical environments [1, 19] but also in other physical systems where soliton solutions are known to exist [20–22].

2. Physical model

Before getting to the physical system itself, we first introduce the basic concepts of RC, aiming to help in the understanding of the construction and results of this work. Simply put, RC starts by taking the input state of the reservoir. The state then evolves under the nonlinear dynamics of a reservoir, reaching a final output state that is sampled over \( \mathcal{N}_I \) output channels, producing an output vector \( \mathbf{Y}^{(i)} \in \mathbb{R}^{\mathcal{N}_C} \). Finally, \( \mathbf{Y}^{(i)} \) is transformed into a forecast via a transformation \( \mathbf{W} : \mathbb{R}^{\mathcal{N}_C} \rightarrow \mathbb{R}^{\mathcal{N}_D} \), which shall be obtained from a training procedure.

In RC the training process consists in determining \( \mathbf{W} \) that minimizes a given loss function [7, 8]. For example, for regression tasks a loss function may be the root mean square error

\[
\varepsilon = \sqrt{\frac{1}{\mathcal{N}_D} \sum_{j=1}^{\mathcal{N}_D} \left\| \mathbf{W} \left[ \mathbf{Y}^{(j)} \right] - \mathbf{T}^{(j)} \right\|^2},
\]

with \( \mathbf{T}^{(j)} \in \mathbb{R}^{\mathcal{N}_T} \) (with dimension \( \mathcal{N}_T \)) being the target associated to each input vector \( \mathbf{X}^{(j)} \). \( \mathbf{W} \) is often chosen as a linear transformation (making \( \mathbf{W} \) in practice a weight matrix), making the minima of equation (1) possible to be calculated using a Moore–Penrose pseudo-inverse [7]. Alternatively, a Ridge regression with Tikhonov regularization can also be employed, a method that has the advantage of being faster to compute numerically, and the ability to penalize large matrix entries, allowing to prevent overfitting that typically occurs in the pseudo-inverse technique [23]. Concerning classification tasks, even though the previous methods are still applicable (by selecting the class with highest forecast value), the use of logistic regression on the output data was reported in recent works yielding promising results [24, 25].

Put in this way, it is clear that the versatility of RC paradigm resides in the freedom of choice of the reservoir, making that role accessible to any nonlinear physical system. To deploy an RC framework with interacting solitons as a reservoir, we first start by introducing the NLSE

\[
i \partial_t \psi + \frac{1}{2} \partial_x^2 \psi + |\psi|^2 \psi = 0,
\]

a general model that describes the propagation of a wave envelope along the optical axis \( z \) in optical media with cubic-type nonlinearity. Equation (2) admits single soliton solutions localized along the transverse direction \( x \)

\[
\psi_i(x, z) = a (x - \bar{x}_i) \text{sech} \left[ a (x - \bar{x}_i) \right] e^{2i \bar{v}_i (x - \bar{x}_i) + i \delta_i},
\]

with \( a, \bar{x}_i, v_i \) and \( \delta_i \) being the amplitude, position, velocity and phase of the soliton, respectively. The dynamics of single solitons are essentially linear and therefore would be insufficient to create an effective reservoir. On the contrary, the case of multiple interacting solitons and in specific, the N-soliton chain, are known to feature highly nonlinear dynamics, therefore constituting a good candidate for a soliton-based reservoir [26].

The N-soliton chain solution for equation (2) can be obtained analytically from the original inverse scattering methodology [27], but a complete theoretical approach results either in a cumbersome mathematical description or is strongly limited by approximations. As a workaround, a simplified model can still be obtained taking the assumption of a small overlap between consecutive solitons [26, 28, 29]. In this case, we can propose an N-soliton ansatz solution for equation (2) as a one-dimensional chain of \( N \) single soliton solutions, i.e.

\[
\psi = \sum_{i=1}^{N} \psi_i(x, z),
\]

with positions \( \bar{x}_i = (i - 1/2) \Delta + x_i \) with \( \Delta \) working as a lattice parameter and \( x_i \) as a displacement from the equilibrium position. For now on we will assume that consecutive solitons are initialized out-of-phase, i.e. \( \delta_{i+1} - \delta_i = \pi \), making the interaction between adjacent solitons of repulsive-type [28, 29].


Figure 1. Illustration of reservoir computing procedure using a soliton chain. (A) The input signal consists of a soliton chain where the data is encoded in the displacement $x_i$ of each soliton (here represented an example of the spiral dataset test, described in main text); (B) Following the dynamics imposed by equation (2), the values of $x_i$ evolve along $z$ until reaching the final propagation distance $z_f$; (C) The Information is retrieved by integrating $|\psi|^2$ across each of $N_c$ output channels, yielding the output vector $Y^{(j)}$. 

The perturbative approach (see appendix A for a complete analysis and accuracy considerations) results then in the Toda Lattice equation (TLE) [28, 29]

$$\partial_z^2 x_i = 4a^3 e^{-a\Delta} \left[ e^{-a(x_i+1-x_i)} - e^{-a(x_i-x_i-1)} \right].$$

(5)

as an approximate model to describe the dynamics of each soliton position. The TLE is a model well-known for its nonlinear dynamics, a fact that confirms the soliton-chain as a good candidate for a soliton-based reservoir. Additionally, we call attention to an interesting crossover between linear and nonlinear regimes that is achieved by varying the amplitude of the oscillations. In fact, considering small propagation distances, small amplitude oscillations (when compared with the lattice parameter, i.e. $x_i \ll \Delta$) shall follow a linear or weak nonlinear regime close to Hooke’s law. On the other hand, higher amplitude oscillations ($x_i \sim \Delta$) feature richer stronger nonlinear dynamics, extensively reported in the literature [30]. In practice, this means that in the context of RC framework and considering the soliton chain as the reservoir, the strength of the nonlinear dynamics of the system can be easily tuned between quasi-linear and nonlinear regimes just by varying the amplitude of the displacements at the input (figure 1).

To deploy the soliton chain as the reservoir on an RC framework, our strategy is to take the $N_I$ real-valued features of each element $j$ of a dataset and encode them in the initial displacements of $N_S \geq N_I$ solitons. For that, we normalize each real-valued feature $f_i$ to the range $[-1, 1]$ and multiply it by a scaling amplitude $A$ according to

$$x_i(z = 0) = A \left( 2 \frac{f_i - \min f}{\max f - \min f} - 1 \right).$$

(6)

The soliton chain can then be used as the reservoir, evolving the encoded input signal $\psi_j(x, 0)$ according to equation (2) and with the strength of the nonlinear behavior being controlled by the scaling amplitude $A$ as discussed before. The output signal is later retrieved at $z_f$ with the information of each $n \in [1, N_C]$ output channel corresponding to the amount of power in each channel division of the transverse dimension $L_x$ according to

$$Y^{(j)}_n = \int_{(n-1)\Delta x/L_x}^{n\Delta x/L_x} |\psi(x, z_f)|^2 dx.$$

(7)

This information is collected from each simulation of element $j$ of the dataset and stored to be used in subsequent training and test stages. Contextualizing in view of reservoir computing principles, the number
of output channels $N_c$ in our system works as an high dimensional space where data is projected, while the nonlinear dynamics of the soliton centroids allow to separate the data in the output space. Before advancing and to avoid confusion, we call the attention of the reader to the fact that the success of our approach is not critically tied to the accuracy of the TLE describing the dynamics of the centroids. In fact, we introduce it only used to get a better insight on how to control the nonlinearity strength of the dynamics of the reservoir. Therefore, even if for some cases the model is less accurate due to significant overlap between solitons (see appendix A), it shall not hurt the validity of the subsequent analysis as they are all based on simulations of the complete NLSE model rather than the TLE.

3. Results

3.1. Regression tasks

To first test the ability of the system to learn (rather than memorizing), we focus on a regression task of approximating a given function. In short, the task consists in predicting $f(x)$ for a total of $N_D = 200$ evenly spaced numbers $x \in [-1, 1]$ which are encoded as in equation (6) in the displacement of the 1st soliton of a chain of 8 solitons of amplitude $a = 2$ and $\Delta = 4$.

After encoding each element of the sample dataset into the input signal, we computed numerically its evolution according to equation (2) using a standard split-step Fourier method with periodic conditions [31]. The output signal was retrieved and stored at a final distance $z_f = 50$ for two scaling amplitudes, $A = 0.25$ and $A = 1$. To prevent model overfitting (see appendix B) and to approximate the simulation to real experimental conditions, we performed two simulation rounds for each element of the dataset adding a 5% white noise on top of the input signal. Subsequently, for a varying number of channels $N_c$, we computed the corresponding weight matrix $W$ using the Moore–Penrose pseudo-inverse technique on the data from the first round of simulations. Finally, we use the computed weight matrix to predict $f(x)$ from the output data of the second round. The results of the described process are depicted in figure 2 for the function $f(x) = \text{sinc}(\pi x)$, which confirm that the system is both able to learn and to provide a good approximation of the target function using the RC paradigm.

On one hand, regarding the dimension of the output space, subfigures A-B suggest that increasing the number of channels—which corresponds to increasing the dimensionality of the output space—can increase the ability of the system to learn for both the cases $A = 1$ and $A = 0.25$. Yet, computing the error as represented in subfigure C, it shall be noticed that increasing the number of channels does not always increase the accuracy. Indeed, a decrease is observed as $N_c$ gets closer to the total size of the dataset, which is explained by overfitting effects commonly associated with the pseudo-inverse technique [23] and that becomes evident when using noisy data. Comparing with previous implementations [18], our data suggests that our implementation can deal with these effects better at a higher signal-to-noise ratio, a key result to warrant the robustness of our framework on real-world experimental settings (see appendix B).

On the other hand, to discuss the role of nonlinearity, we first observe that the results suggest that the system is unable to learn directly from the initial input signal. This means that simply projecting the data with position encoding into the higher dimensionality space of $N_c$ channels is insufficient for the regression task, requiring the nonlinear expansion provided by the Toda Lattice dynamics. To confirm this argument we introduce a common metric called capacity [32–34], defined as

$$C[T] = 1 - \frac{\sum_{j=1}^{N_D} \| W [y(j)] - T(j) \|^2}{\sum_{j=1}^{N_D} \| T(j) \|^2}. \quad (8)$$

The capacity, which varies between 0 and 1, measures the success of a given system at approximating a given target $T$, with 1 corresponding to a perfect reconstruction and 0 to a system that is unable to do that. In figure 3 we present the results obtained for the function under consideration for varying input scaling values. It is clear to see that the nonlinear regime allows for better performances in the regression task, yet, while weak nonlinear dynamics are insufficient for the approximation of this function, a nonlinear regime that is too strong can also be detrimental as represented by the decrease registered around $A = 1.5$. The results are again consistent with previous literature, in particular with reference [18], which claims that learning an arbitrary dataset is crucially tied to a non-polynomial evolution of the information at the output channels. Finally, and before advancing, we would like to call attention to the fact that the capacity measured here refers only to the function under consideration. To better understand the performance of our system across a given space of functions and how it compares against other RC implementations, one should follow a methodology similar to that previously described in the literature [32, 33]. This research line is left outside the scope of the present manuscript and shall be addressed in future works.
Figure 2. Results of the regression task for the function $f(x) = \text{sinc} (\pi x)$ using a system of $N_s = 8$ solitons. Quantitative prediction trained with (A) $N_c = 25$ (B) and (C) $N_c = 50$ channels directly from the input signal before propagation (orange crosses) and from the output signal for scaling inputs with amplitude $A = 0.25$ (green squares) and $A = 1$ (blue circles). The dotted line represents ideal predictions. (D) Computed error as given by equation (1) in function of the number of channels $N_c$.

Figure 3. Capacity of the soliton-based RC measured for the regression task of the function $f(x) = \text{sinc} (\pi x)$ for varying values of input scaling parameter $A$.

3.2. Classification tasks
So far we have shown that nonlinear dynamics combined with a high-dimensional output space provides our system the ability to effectively learn or memorize a dataset. However, the typical machine learning problem is not to memorize an entire dataset but to infer the best model using just a training subset, and from it, accurately predict the result on the remaining untrained test data. To evaluate the ability of our
In this task, we consider a two-feature dataset of size $N_D = 250$ that shall be classified into the corresponding target class as illustrated in figure 4(A)). Following an approach similar to the one used before, we encoded the two features of each element in the positions of the 1st and 5th soliton of the same 8-soliton chain used for the regression task (see figure 1). For each input signal we performed numerical simulations of the NLSE, again up to the same propagation distance $z_f = 50$ and for two distinct scaling encoding amplitudes $A = 0.25$ and $A = 1$. In the training procedure, we divided the dataset into randomly chosen 80%–20% train-test subsets, and computed the model $W$ using logistic regression on the output signal data of the training subset data (function LogisticRegression of Python sklearn library [35]) for a varying number of channels $N_c$. Subsequently, the computed model $W$ was used to classify the remaining test dataset. The training procedure was repeated for 10 randomly chosen train-test subsets, obtaining the average results depicted in figure 5.

As observed for the regression task, the results suggest that both increasing the number of channels and the nonlinear strength play important roles in the overall accuracy. Indeed, looking at the accuracy of the model directly predicted from the input signal, we see that it barely beats the pure guessing probability (50%). This performance is easily surpassed by the models obtained using the output signal collected at the end of the reservoir, which again confirms the importance of the nonlinear dynamics. Furthermore, it is also clear that increasing the scaling amplitude increases the accuracy performance of the predicted model in both the training and test subsets, achieving a typical accuracy in excess of 90% in the test subset for amplitude $A = 1$.

Finally, in order to get a better insight of the model predicted in each case, we performed numerical simulations corresponding to each point of a $50 \times 50$ rectangular grid with $feature \ 1 \in [-3.4, 3.6]$ and $feature \ 2 \in [-2.8, 3.2]$, and classified each point using a previously trained model with $N_c = 25$ channels. Figures 4(B)–4(D)) show the results obtained: it becomes not only clear that the correct spiral-like separation of the data is only possible in the nonlinear regime of the reservoir, but also that the predicted
model for \( A = 1 \) generalizes well in the feature space. This proves that our soliton-based RC system is able to perform classification tasks involving nonlinear separable data, setting the field for future research involving more complicated datasets.

4. Conclusions

In this work, we explored the possibility of using an interacting soliton-chain to work as a reservoir on a computing system based on the RC paradigm. Taking advantage of its tunable nonlinear dynamics, governed by the Toda lattice equation and easily controlled using the initial displacement amplitude of each soliton, we were able to investigate the importance of the role played by nonlinear dynamics for the success of the system performing of regression and classification tasks. We demonstrated that providing the system with a sufficiently large output space and nonlinear dynamics, our soliton-based computing system is able to perform both regression and classification tasks. In spite of having used computational methods, the results presented were all obtained with noisy data which approximates the experimental conditions and therefore, are more plausible to be confirmed in future experiments.

To our best knowledge, this work presents the first time that a soliton-based system is considered to deploy a reservoir computer, a fact that in our opinion may inspire future research in two directions. On one hand, the versatility, tunability, and robustness of the system may inspire the development of all-optical computing technologies, not only restricted to propagation geometries and continuous-wave beams but also with temporal optical pulses \([24, 28]\), possibly bypassing the obstacle of the absence of periodic conditions in real-world setups by establishing a Newton’s cradle-like configuration \([26, 29, 36]\). On the other hand, considering the universality of the soliton solutions in nonlinear science, not only in optical systems \([19]\) but also in fluids \([37]\), plasmas \([20]\) and quantum gases \([21, 22]\), just to name a few, the present work establishes a general blueprint that can easily be extended towards computing solutions with a great variety of physical systems.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Acknowledgments

This work is financed by the European Regional Development Fund (ERDF) through the Operational Programme for Competitiveness and Internationalisation, COMPETE 2020 Programme, and by National Funds through the FCT – Fundação para a Ciência e a Tecnologia (Portuguese Foundation for Science and Technology), within Project No. POCI-01-0145-FEDER-032257 and project UIDB/50014/2020, as well as by the North Portugal Regional Operational Programme (NORTE 2020), under the PORTUGAL 2020 Partnership Agreement. TDF is supported by Fundação para a Ciência e a Tecnologia through Grant No. SFRH/BD/145119/2019. NAS and AG acknowledge the Texas Advanced Computing Center (TACC) at The
University of Texas at Austin for providing visualization resources that have contributed to the research results reported within this paper. NAS also acknowledge the financial support of the project “Quantum Fluids of Light in Hot Atomic Vapors”, supported by FCT in collaboration with the Ministry of Education, Science and Technological Development of the Republic of Serbia.

Appendix A. Perturbative approach and the Toda Lattice equation

The key idea of this work is based on using a soliton-chain and their particle-like interaction as a nonlinear reservoir in a reservoir computing framework. As starting point of our analysis we focused on the NLSE,

\[ i\partial_t \psi + \frac{1}{2} \partial_x^2 \psi + |\psi|^2 \psi = 0, \]  

(A.1)

a well-known model that describes in a universal manner the propagation of an envelope wave along the propagation direction \( z \) in a nonlinear media with cubic-type nonlinearity. Furthermore, in one-dimensional settings, equation (A.1) admits single soliton solutions localized along the transverse direction \( x \) of the form

\[ \psi_j(x,z) = a \sech[a(x-x_j)] \exp[i\psi_j(x-z)], \]  

(A.2)

with \( a, x_j, \mu_j \) and \( \delta \) being the amplitude, position, velocity and phase of the soliton respectively. To construct the soliton chain solution we use as reservoir, we assume small overlapping between consecutive solitons \([26, 28, 29]\), a case for which we can propose an \( N \)-soliton ansatz solution for equation (A.1) as a one-dimensional chain of \( N \) single soliton, i.e.

\[ \psi = \sum_{i=1}^{N} \psi_i(x,z), \]  

(A.3)

with centroid positions \( x_i = (i - 1/2) \Delta + x_i \), where \( \Delta \) works as a lattice parameter and \( x_i \) as a displacement from the equilibrium position. By replacing the ansatz into equation (1) we obtain a system of coupled NLSE-like equations reading for each soliton

\[ \frac{i}{\partial_t} \psi_i + \frac{1}{2} \partial_x^2 \psi_i + |\psi_i|^2 \psi_i = iR_i[\psi]. \]  

(A.4)

To treat the perturbation \( R_i \), which is associated with the inter-soliton interaction, we assumed it as a small perturbation that can be described using a quasi-particle approach \([28]\). First assuming that only first neighbors can interact we get

\[ R_i[\psi] = 2|\psi_i|^2 (\psi_{i+1} + \psi_{i-1}) + |\psi_i|^4 (\psi_{i+1}^* + \psi_{i-1}^*) \]  

(A.5)

that in the quasi-particle framework results in the following evolution equations for the parameters of each \( j \)th soliton \([28]\)

\[ \frac{d\mu_j}{dz} = 2a^3 \left[ \cos(\phi_{i+1,j}) \exp(-\Delta_{ij+1}) - \cos(\phi_{i-1,j}) \exp(-\Delta_{ij-1}) \right], \]  

(A.6)

\[ \frac{dx_j}{dz} = 2\mu_i - 2a \sin(\phi_{i+1,j}) \exp(-\Delta_{ij+1}) - \sin(\phi_{i-1,j}) \exp(-\Delta_{ij-1}) \]  

(A.7)

\[ \frac{d\delta_j}{dz} = 2 \left( a^2 + \mu_j^2 \right) + 2\mu \left( \frac{dx_j}{dz} - 2\mu \right) + 6a^2 \cos(\phi_{i+1,j}) \exp(-\Delta_{ij+1}) \]  

\[ + 6a^2 \cos(\phi_{i-1,j}) \exp(-\Delta_{ij-1}) \]  

(A.8)

In this system, \( \phi_{i,j} = 2\mu \left( x_j - x_i \right) + \theta_{ij} \) is the complex phase between solitons with \( \theta_{ij} = \delta_j - \delta_i \), \( \Delta_{ij} = 2a \left| x_i - x_j \right| \) is the spacing between solitons in the \( i \)th and \( j \)th position, and \( a \) and \( \mu \) are the mean values for the amplitude and velocity, respectively. These equations are a first order approximation that assume a small overlap, i.e. \( a \left| x_i - x_j \right| \gg 1 \) and that the solitons start and remain with similar amplitudes \( a \). To deploy a stable interacting chain, we consider that the system is initialized with consecutive solitons having a phase difference of \( \pi \), which results in a repulsive inter-soliton interaction. Note, that while equation (A.8) shows that the phase between consecutive solitons can actually vary \([29]\), we found numerically that it can be considered constant for the distance of interest we considered in our simulations. In this conditions, we end with the Toda lattice equation

\[ \partial_x^2 x_i = 4a^3 e^{-\alpha \Delta} \left[ e^{-a(x_{i+1} - x_i)} - e^{-a(x_i - x_{i-1})} \right], \]  

(A.9)
Figure A1. Results of numerical simulation of the point of \((f_1, f_2) = (1, 1)\) of the two-spiral dataset, encoded as described in the main text for \(A = 1\). In this figure we present a qualitative comparison between the numerical simulation of the soliton chain using the NLSE (solid black line, background intensity image) and that of the centroids motion using the TLE (dashed blue line).

Figure A2. Results of the regression task for the function \(f(x) = \text{sinc}(\pi x)\) using a system of \(N_s = 8\) solitons. Quantitative prediction using the same train and test output signals for (A) \(N_c = 25\) (B) \(N_c = 50\) channels for scaling encoding amplitudes \(A = 0.25\) (green squares) and \(A = 1\) (blue circles). Also represented are the results obtained directly from the input signal before propagation (orange crosses). The dotted line represent ideal predictions. (C) Computed error in function of the number of channels \(N_c\).

Figure A3. Same as in figure 2 but with distinct train-test output signals with noise.

that we presented in the main text. Our interest in this model is based on the fact that simply varying the amplitude of oscillation can allow us to control the nonlinearity strength of the dynamics. While it is known that in the regime of large displacements the Toda lattice equation is a nonlinear model, for small displacements it can be approximated at the first order by

\[
\frac{\partial^2 x_i}{\partial z^2} \approx C (x_{i+1} - 2x_i + x_{i-1}) + \mathcal{O}(x^2)
\]

with \(C = 4a^4 \exp(-a\Delta)\), which resembles the equation of motion of a system governed by linear Hooke’s law and that means that a crossover between linear and nonlinear behavior exists depending on the amplitude of the initial conditions.

Figure A.1 represents results for the centroid motion obtained for a simulation of the two-spiral dataset as described in the main text. In spite some noticeable differences for larger propagation distances that can
be attributed to the significant overlap between solitons of the extreme case we choose here, the results obtained still show a good qualitative agreement between those obtained numerically from the simulation of the full NLSE and those of Toda Lattice equation. This confirms the possibility to explore the reservoir’s nonlinearity control at the light of the Toda Lattice equation.

Appendix B. Pseudoinverse and overfitting in regression tasks

In the main text, the first test to our soliton-based computing system is a regression task consisting of approximating the function \( f(x) = \text{sinc}(\pi x) \). As described, we trained our system with a total of \( N_D = 200 \) evenly spaced numbers \( x \in [-1, 1] \), propagating the corresponding encoded signals up to a distance \( z_f = 50 \) according to equation (A.1) for scaling amplitudes \( A = 0.25 \) and \( A = 1 \). Subsequently, using the output signals and for a varying number of intensity channels \( N_c \) as explained in the main text, we computed the corresponding weight matrix \( \bar{W} \) using the Moore–Penrose pseudo-inverse technique. To prevent overfitting and approximate the results obtained from real experimental conditions, we choose to perform two rounds of simulations for each element of the dataset adding a 5\% white noise on top of the input signal, using the first data to compute \( \bar{W} \) and the second data to get a prediction and compute the error.

To understand the importance of the above technique we represent in figure A.2 the results obtained for the regression task computed with just the first simulation round, i.e. using the same output signal for training the model \( W \) and subsequently predict the result. As it can be seen if we used such strategy we would be lead to the belief that the best performance—corresponding actually to an error of the order of the computer precision used—is achieved for \( N_c = N_D \). Furthermore, accurate results can be obtained not only in the nonlinear regime but also in the linear one, and even using the input signal itself seems sufficient. However, such conclusion that would partly align with the results of reference [18] is inaccurate. Indeed, if we apply the system to a dataset with different noise like we did in the main text (see figure A.3)—and which is the case in real experimental conditions—we see exactly the opposite: the accuracy decreases approaching the \( N_c = N_D \) case. This is the typical manifestation of an overfitting problem, well known to occur using the Moore-Penrose pseudoinverse [23]. The alternative is either to use noise as we did or alternatively, to use Ridge regression techniques which can also achieve good quantitative results [23] (see figure A.4).

ORCID iDs

Nuno Azevedo Silva https://orcid.org/0000-0002-8467-4357

References

[1] Jakubowski M H, Steiglitz K and Squier R 2002 Computing with solitons: a review and prospectus Collision-based Computing (Berlin: Springer) pp 277–97
[2] Adamatzky A 2012 Collision-based Computing (Berlin: Springer)
[3] Jakubowski M H, Steiglitz K and Squier R 1996 When can solitons compute? Complex Syst. 10 1–22
[4] Jakubowski M H, Steiglitz K and Squier R 1997 Information transfer between solitary waves in the saturable Schrödinger equation Phys. Rev. E 56 7267
[5] Bakaoukas A G 2018 An all-optical soliton FFT computational arrangement in the 3NLSE-domain Nat. Comput. 17 231–48
[6] Bakaoukas A G 2019 The two separate optical fibres approach in computing with 3nlse–domain optical solitons Intelligent Computing-Proc. of the Computing Conf. (Berlin: Springer) pp 253–80
[7] Jaeger H 2001 The echo state approach to analysing and training recurrent neural networks—with an erratum note Technical Report German National Research Center for Information Technology GMD Bonn, Germany148
[8] Maass W, Natschläger T and Markram H 2002 Real-time computing without stable states: a new framework for neural computation based on perturbations Neural Comput. 14 2531–60
[9] Tanaka G, Yamane T, Héroux J B, Nakane R, Kanazawa N, Takeda S, Numata H, Nakano D and Hirose A 2019 Recent advances in physical reservoir computing: a review Neural Netw. 115 100–23
[10] Seoane L F 2019 Evolutionary aspects of reservoir computing Phil. Trans. R. Soc. B 374 20180377
[11] Van der Sande G, Brunner D and Soriano M C 2017 Advances in photonic reservoir computing Nanophotonics 6 561–76
[12] Fernando C and Sojakka S 2003 Pattern recognition in a bucket European Conf. on Artificial Life (Berlin: Springer) pp 588–97
[13] Ghosh S, Opala A, Matuszewski M I, Peterek T and Liew T C H 2019 Quantum reservoir processing npj Quantum Inf. 5 1–6
[14] Lin X, Rivenson Y, Yardimci N T, Veli M, Luo Y, Jarrahi M and Ozcan A 2018 All-optical machine learning using diffractive deep neural networks Science 361 1004–8
[15] Shen Y et al 2017 Deep learning with coherent nanophotonic circuits Nat. Photon. 11 441
[16] Sunada S and Uchida A 2019 Photonic reservoir computing based on nonlinear wave dynamics at microscale Sci. Rep. 9 1–10
[17] Ballarini D et al 2020 Polaritonic neuromorphic computing outperforms linear classifiers Nano Lett. 20 3506–12
[18] Marcucci G, Pierangeli D and Conti C 2019 Theory of neuromorphic computing by waves: machine learning by rogue waves, dispersive shocks, and solitons Phys. Rev. Lett. 125 093901
[19] Kivshar Y S and Agrawal G P 2003 Optical Solitons: From Fibers to Photonic Crystals (New York: Academic press)
[20] Kuznetsov E A, Rubenchik A M and Zakharov V E 1986 Soliton stability in plasmas and hydrodynamics Phys. Rep. 142 103–65
[21] Khawaja U A, Stooft H T C, Hulet R G, Strecer K E and Partridge G B 2002 Bright soliton trains of trapped Bose–Einstein condensates Phys. Rev. Lett. 89 200404
[22] Burger S, Bongs K, Dettmer S, Ertmer W, Sengstock K, Sampera A, Shlyapnikov G V and Lewenstein M 1999 Dark solitons in Bose–Einstein condensates Phys. Rev. Lett. 83 5198
[23] Schrauwen B, Verstraeten D and Van Campenhout J 2007 An overview of reservoir computing: theory, applications and implementations Proc. of the 15th European Symp. on Artificial Neural Networks pp 471–82
[24] Borlenghi S, Boman M and Delin A 2018 Modeling reservoir computing with the discrete nonlinear Schrödinger equation Phys. Rev. E 98 052201
[25] Opala A, Ghosh S, Liew T C H and Matuszewski M 2019 Neurromorphic computing in ginzburg-landau polariton-lattice systems Phys. Rev. Appl. 11 064029
[26] Novoa D, Malomed B A, Michinel H and Perez-Garcia V M 2008 Supersolitons: solitonic excitations in atomic soliton chains Phys. Rev. Lett. 101 144101
[27] Zakharov V and Shabat A 1972 Exact theory of two-dimensional self-focusing and one-dimensional self-modulation of waves in nonlinear media Sov. Phys. JETP 34 62
[28] Gerdjikov V S, Kaup D J, Uznov I M and Evtovichei E G 1996 Asymptotic behavior of soliton trains of the nonlinear schrödinger equation Phys. Rev. Lett. 77 3943
[29] Ma M, Navarro R and Carretero-González R 2016 Solitons riding on solitons and the quantum Newton’s cradle Phys. Rev. E 93 022202
[30] Krüger H and Teschl G 2009 Long-time asymptotics of the toda lattice for decaying initial data revisited Rev. Math. Phys. 21 61–109
[31] Silva N A, Mendonça J T and Guerreiro A 2017 Persistent currents of superfluidic light in a four-level coherent atomic medium J. Opt. Soc. Am. B 34 2220–6
[32] Dambre J, Verstraeten D, Schrauwen B and Massar S 2012 Information processing complexity of dynamical systems Sci. Rep. 2 1–7
[33] Inubushi M and Yoshimura K 2017 Reservoir computing beyond memory-linearity trade-off Sci. Rep. 7 1–10
[34] Nakajima K 2020 Physical reservoir computing—an introductory perspective Japan J. Appl. Phys. 59 060501
[35] Pedregosa F et al 2011 Scikit-learn: machine learning in python J. Mach. Learn. Res. 12 2825–30
[36] Driben R, Malomed B A, Yulin A V and Skryabin D V 2013 Newton’s cradles in optics: from n-soliton fission to soliton chains Phys. Rev. A 87 063808
[37] Johnson R S 1980 Water waves and Korteweg–de Vries equations J. Fluid Mech. 97 701–19