ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

A Markovian Characterization of Redundancy in Echo State Networks by PCA

Claudio Gallicchio and Alessio Micheli

Department of Computer Science - University of Pisa
Largo B. Pontecorvo, 3 56127 - Italy

Abstract. Richness of dynamics is a desirable feature of Echo State Networks (ESNs) limited by a known high redundancy of state units activations. We show how this feature is mainly influenced by the Markovian state space organization of ESNs through a Principal Component Analysis (PCA) of the reservoir space. Relevances of principal components are coherent with Markovianity, whose role is further enlightened by investigating the strong relation among the suffix elements of the input sequence and the most relevant directions of variability in the state space.

1 Introduction

Echo State Networks (ESNs) [1, 2] are emerging as an efficient approach to Recurrent Neural Networks (RNNs) modeling. An ESN (typically) consists in a large, random and sparsely connected *reservoir* hidden layer of recurrent units plus a linear *readout* layer of feed-forward units. The key feature of ESNs is that the reservoir can be left untrained after initialization, restricting training to the readout part, whose parameters can be adapted by efficient linear regression.

The reservoir should provide the readout with a richly varied pool of dynamics from which to linearly combine. Dimensionality of the reservoir is a relevant factor responsible for increasing the dynamics diversification and network performance (e.g. [3, 4]). However, high correlation among reservoir units activations is a well known fact in ESN modeling (e.g.[5, 6, 7]), making the approach to complex tasks more difficult. Several measures for reservoir goodness have been proposed in ESN literature. The pairwise correlation among reservoir units and the entropy of state distributions are two of the most popular metrics [5, 7], leading to alternative ESN architectures, such as the Decoupled ESN (DESN) [6], and to methods for optimizing reservoirs by using Intrinsic Plasticity (IP) (e.g. [8]) or maximization of a time-averaged entropy of echo states [9]. However, simple tools still lack to be exploited.

It is known that RNNs with contractive state dynamics have the inherent ability to discriminate among different (recent) input histories in a *Markovian* flavor even without training the recurrent connection weights (e.g. [10]). Input sequences sharing a common suffix drive the same network into states which are close to each other proportionally to the length of the common suffix. This also applies to ESNs [11]. Markovianity has also revealed a major role in determining the success and limitations of ESNs in predictive applications [4].

In this paper, the issue of redundancy (intended as high pairwise correlation) among reservoir units activations is put in relation to the Markovian organization of ESNs dynamics. Markovianity rules both the global behavior of the network and the local behaviors of each state unit. It is therefore likely to expect that this characterization leads to similar activations among reservoir units and thus to redundancy, with higher redundancy induced by stronger degrees of Markovianity. Our investigation exploits Principal Component Analysis (PCA) [12] as a simple and useful tool to analyze ESN dynamics, by isolating interesting orthogonal directions of variability in the reservoir state space. The relevance of principal components (PCs) of reservoir states and their relations with suffix elements of the input ground our analysis and allow us to trace redundancy back to the Markovian nature of ESN dynamics.

2 Echo State Networks

ESNs have been introduced in [1]. The equations describing the basic model are:

$$\mathbf{x}(n) = f(\mathbf{W}_{in}\mathbf{u}(n) + \hat{\mathbf{W}}\mathbf{x}(n-1)); \quad \mathbf{y}(n) = \mathbf{W}_{out}\mathbf{x}(n) \tag{1}$$

where $\mathbf{u}(n)$, $\mathbf{y}(n)$ and $\mathbf{x}(n)$ respectively represent the input, output and state of the network at pass n. $\mathbf{W}_{in}$ is the input-to-reservoir weight matrix, $\hat{\mathbf{W}}$ is the recurrent reservoir weight matrix and $\mathbf{W}_{out}$ is the reservoir-to-output weight matrix. $\hat{\mathbf{W}}$ is usually a sparse matrix, with a percentage of connectivity below 20%. Typically, the activation function is *tanh* for the reservoir units and is linear for the output units. Common variants to equation 1 can be found in [1].

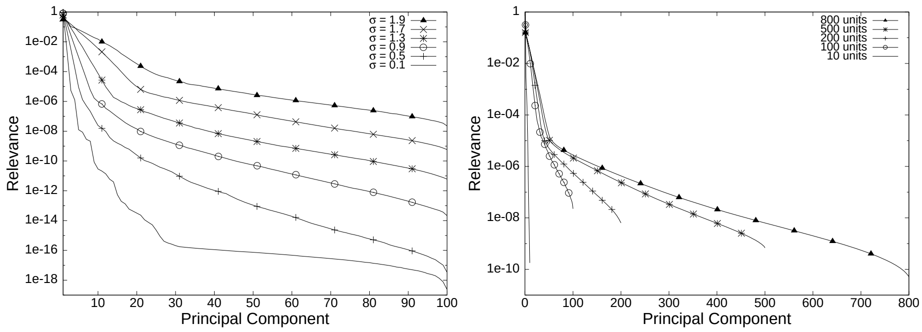
To get a valid ESN, the "echo state property" must hold. Two conditions for the echo state property, in case of *tanh* as reservoir activation function, are provided in [1]. They both involve a scaling of the matrix $\hat{\mathbf{W}}$ to have a spectral radius less than unity (necessary condition) or a maximum singular value less than unity (sufficient condition). The maximum singular value of matrix $\hat{\mathbf{W}}$ is its Euclidean norm and governs the contractivity and consequent Markovianity of the state transition function in equation 1 [4]. For this reason we call it the *contraction coefficient* of the network, indicated by the symbol σ, and use it instead of the spectral radius to scale the matrix $\hat{\mathbf{W}}$. A smaller value of σ corresponds to a stronger Markovian characterization of network state dynamics.

3 Principal Component Analysis of Echo State Networks

In the following, we will refer to simple symbolic input sequences without covariance among input elements to feed the networks in experiments. We considered an alphabet of 10 symbols, $\mathcal{A} = \{a, b, \ldots, j\}$. Each symbol in a sequence was randomly selected from a uniform distribution over $\mathcal{A}$. A uniform spacing in $(0, 1]$ was adopted to numerically represent each symbol, such that symbol a was coded as 0.1, symbol b was coded as 0.2, and so on up to symbol j, coded as 1.0. For each experiment, an input sequence of length 100 was used as initial transient, while a 1000-long sequence was used to collect reservoir states. For every time step $n = 1, \ldots, 1000$, the network state $\mathbf{x}(n)$ was computed and stored as a column in a matrix $\mathbf{X}$, with a number of rows equal to the reservoir dimension and a number of columns equal to the number of time steps. PCA was then applied to matrix $\mathbf{X}$.

ESNs with different reservoir dimensions and σ were considered, while the sparse connectivity among reservoir units was fixed to 10%. Weight values in $\mathbf{W}_{in}$ were randomly chosen according to a uniform distribution over $[-0.1, 0.1]$.

Redundancy of reservoir units activations can be clearly shown by the relevance of PCs of ESN states (i.e. the eigenvalues of the covariance matrix of $\mathbf{X}$). This study can be used as a simple tool to assess the richness of ESN dynamics. As PCs represent orthogonal directions of variability in the state space, a greater number of relevant PCs indicates a richer set of state dynamics and reservoir units able to better diversify their behavior. We collected the relative relevance of PCs of reservoir states with varying σ and the reservoir dimension. Results, averaged over 10 independent trials and scaled in $[0, 1]$, are presented in Fig. 1. A first general observation is that, since we used a semilog scale, the relevance of PCs shows an exponential decay with the associated PCs ranking, with values eventually falling under machine precision. A small number of PCs represent almost all the variability in the state space, confirming the expected high redundancy. For $\sigma = 0.9$ and reservoir dimension equal to 100, the first three PCs collected on average the 99.4% of the total relevance.



(a) (b)

Fig. 1: Relevance of PCs of reservoir states. For each PC, the ratio between its variance and the total variance in the PC space is reported. **(a):** σ varying in $[0.1, 1.9]$ and reservoir dimension fixed to 100. **(b):** Reservoir dimension varying between 10 and 800 ($\sigma = 1.9$). The relevance dimension is in log scale.

More in detail, Fig. 1a reports the relative relevance of PCs for ESNs with 100 reservoir units, varying σ from 0.1 to 1.9 (corresponding to mean spectral radius values from 0.05 to 0.98). It is evident that decreasing the value of σ (i.e. increasing the Markovianity of the state transition function) leads to lower relative relevances of PCs with smaller variance. More contractive ESN dynamics present a smaller number of orthogonal directions with non-negligible relevance and thus are characterized by higher redundancy. This observation is coherent with the Markovian nature of the ESN dynamics. Indeed, the range of behaviors

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

of each state unit shrinks towards the global Markovian dynamics as σ decreases.

Fig. 1b shows the relative relevance of PCs for ESNs with $\sigma = 1.9$ and with varying reservoir dimension from 10 to 800 units. The application of PCA graphically shows how much increasing the reservoir dimension can be effective in allowing a differentiation of the state dynamics. From Fig. 1b we may also observe that the enhancement of the richness of ESN dynamics is attenuated as the reservoir dimension gets larger. While a great difference can be noted between the 10 units case and the 100 units case, a much less evident difference emerges from a comparison between the 500 units case and the 800 units one, suggesting that for a given degree of Markovianity there exists a saturation effect in the number of units.

The following investigation on the meaning of the PCs of the reservoir states can enlighten the Markovian organization of ESN state space and eventually its relationships with the richness of the dynamics.

We considered a 100 units reservoir with $\sigma = 0.3$. In Fig. 2 we plotted the projections of the reservoir states into the first two PCs space (PC scores). Each point represents a state of the network and is labeled with the suffix of length 2, i.e. the couple (next-to-last, last), of the input sequence which drove the network in that state. Fig. 2 clearly reveals that the first two PCs contain sufficient information to reconstruct the input sequence suffix of length 2. In particular, the first PC groups states according to the last input symbol, while the second PC groups them according to the 2nd-last input symbol. The influence of the previous parts of the input sequence is not graphically appreciable.

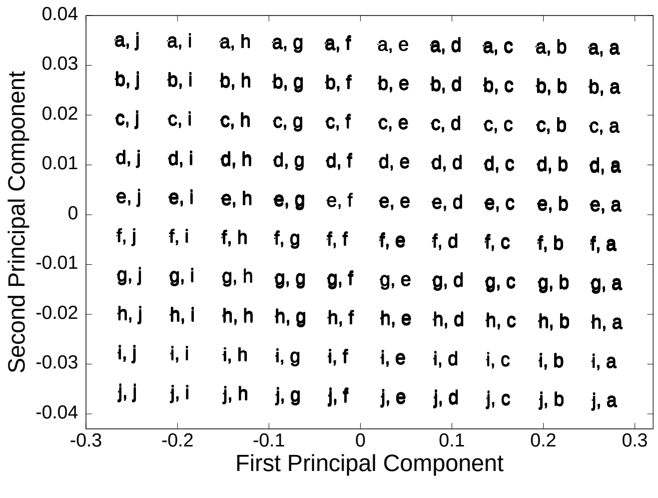


Fig. 2: Markovian organization (suffix based clusters) resulting on the first two PCs space of reservoir states. Note the different scales between the two axes.

The relation between symbols of the input suffix and the PCs of reservoir states is further and explicitly enlightened in Fig. 3, referring to the same ESN setting introduced for Fig. 2. A nearly linear relation between the last input

ESANN 2010 proceedings, European Symposium on Artificial Neural Networks - Computational Intelligence
and Machine Learning. Bruges (Belgium), 28-30 April 2010, d-side publi., ISBN 2-930307-10-2.

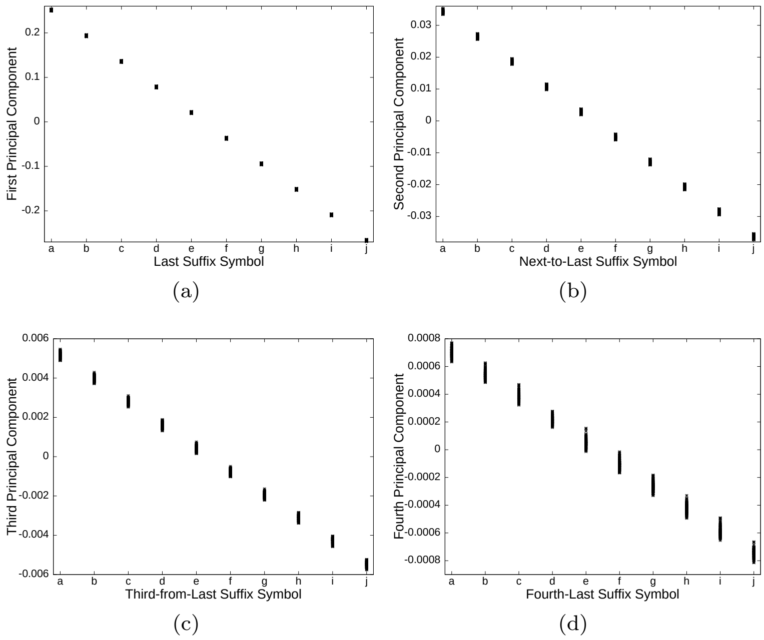


Fig. 3: Suffix input symbols vs most relevant PCs of reservoir states. **(a):** Last input symbol vs first PC. **(b):** 2nd-last input symbol vs second PC. **(c):** 3rd-last input symbol vs third PC. **(d):** 4th-last input symbol vs fourth PC.

symbol and the first PC can be seen from Fig. 3a. The first PC, i.e. the direction of greater variance in the reservoir space, may be almost identified by the variability on the last input symbol and the symbols are discriminable on intervals of the PC values. Even more noteworthy is the presence of analogous relations among other suffix elements of the input sequence and subsequent PCs with decreasing variance. While the dynamics of each unit is ruled by the same type of Markovian behavior, well represented by the first PC, the differences among the Markovian dynamics of the state units activate orthogonal directions of variability. In particular, being isolated from the first PC, the second PC reveals its nature mainly related to the second element in the Markovian ranking of relevance. Such relations were found up to the 4th-last input element and the fourth PC (Fig. 3b, 3c, 3d), whereas the strength of the relationship rapidly decreases in the past temporal direction.

These relations revealed the nature of ESN state space organization, enlightening its Markovian flavor: most recent input symbols directly influenced most relevant directions of variance in reservoir state spaces, with dramatically decreasing strength for older input entries. Moreover, although the dynamics of each unit is Markovian, and hence dominated by the last observed input element, the PCA of the reservoir states reveals a nice *spectral property* on older input elements (up to a certain order).

4 Conclusions

The study of relative relevances of PCs has been used as simple tool to assess the richness of network dynamics. In fact, the effect of diversification among the Markovian dynamics of each single unit has been revealed by showing the increasing relevance of PCs with increasing reservoir dimension. The main result concerns the strong relation between redundancy of reservoir units activations and Markovianity of ESN dynamics. PCA have revealed high redundancy of reservoir units, and stronger Markovian characterizations of state transition functions have resulted in higher redundant reservoirs. Markovian ordering on the suffix elements of the input sequence has shown a tight relation to the most relevant PCs, providing also an insight on the order of Markovianity involved. Therefore, we can conclude that main redundancy effects in ESNs follows from the inherent Markovian nature of reservoir state space organization.

We believe that this investigation, especially because of the simplicity of the exploited tool, can contribute to an easier understanding of features and limitations of the ESN class of neural network models.

References

[1] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. Technical report, GMD - German National Research Institute for Computer Science, 2001.

[2] H. Jaeger and H. Haas. Harnessing nonlinearity: Predicting chaotic systems and saving energy in wireless communication. *Science*, 304(5667):78–80, April 2004.

[3] D. Verstraeten, B. Schrauwen, M. D'Haene, and D. Stroobandt. An experimental unification of reservoir computing methods. *Neural Networks*, 20(3):391 – 403, 2007.

[4] C. Gallicchio and A. Micheli. Architectural and markovian factors of echo state networks. Submitted to journal. Technical Report TR-09-22 (November 2009) available at `http://compass2.di.unipi.it/TR/Files/TR-09-22.pdf.gz`.

[5] M. Lukosevicius and H. Jaeger. Reservoir computing approaches to recurrent neural network training. *Computer Science Review*, 3(3):127 – 149, 2009.

[6] X. Yanbo, Y. Le, and S. Haykin. Decoupled echo state networks with lateral inhibition. *Neural Networks*, 20(3):365–376, 2007.

[7] H. Jaeger. Reservoir riddles: suggestions for echo state network research. *Proceedings of IJCNN 2005*, 3:1460–1462 vol. 3, July-4 Aug. 2005.

[8] B. Schrauwen, M. Wardermann, D. Verstraeten, J.J. Steil, and D. Stroobandt. Improving reservoirs using intrinsic plasticity. *Neurocomputing*, 71(7-9):1159 – 1171, 2008.

[9] M.C. Ozturk, D. Xu, and J.C. Principe. Analysis and design of echo state networks. *Neural Computation*, 19(1):111–138, 2007.

[10] B. Hammer and P. Tiño. Recurrent neural networks with small weights implement definite memory machines. *Neural Computation*, 15(8):1897–1929, 2003.

[11] P. Tiño, B. Hammer, and M. Bodén. Markovian bias of neural-based architectures with feedback connections. In *Perspectives of Neural-Symbolic Integration*, pages 95–133. Springer-Verlag, 2007.

[12] I.T. Jolliffe. *Principal Component Analysis*. Springer, second edition, October 2002.