Supplementary materials for: Learning Reservoir Dynamics with Temporal Self-Modulation

Yusuke Sakemi[1,2], Sou Nobukawa[1,3,4], Toshitaka Matsuki[5], Takashi Morie[6], and Kazuyuki Aihara[1,2]

[1]Research Center for Mathematical Engineering, Chiba Institute of Technology, Narashino, Japan
[2]International Research Center for Neurointelligence (WPI-IRCN), The University of Tokyo, Tokyo, Japan
[3]Department of Computer Science, Chiba Institute of Technology, Narashino, Japan
[4]Department of Preventive Intervention for Psychiatric Disorders, National Institute of Mental Health, National Center of Neurology and Psychiatry, Tokyo, Japan
[5]Oita University, Oita, Japan
[6]Graduate School of Life Science and Systems Engineering, Kyushu Institute of Technology, Kitakyushu, Japan

March 31, 2023

Dependence of timesteps t_p in simple attention task

We performed the sensitivity analysis for the simple attention task using different timesteps t_p, at which the magnitudes of expansion of the perturbations were measured. Figure 1 shows the results for the RC and SM-RC models.

Learning stability

Recurrent neural network (RNN) learning is unstable because there are many regions where the loss function changes the value sharply [1, 2]. Reservoir computing (RC) avoids this problem by learning only the output layer [3]. In self-modulated RC (SM-RC), we found that the learning process was unstable, which is attributed to the fact that the RNN was indirectly trained through gates. Figure 2 shows the MSE for each learning epoch in 50 different trials. The learning task involved the Lorentz model, with $N^{\text{forward}} = 20$ and $\sigma^{\text{Lorentz}} = 0.03$. Figure 2 A shows the case where all the trainable weights, including the output layer, were trained via the gradient descent method. In this case, the learning process was unstable, resulting in poor performance. This implies that SM-RC training has the same problems as RNN training described above. Figure 2 B shows the case where the weights of the output layer were trained via the pseudo-inverse method and the weights related to the gates were learned via the gradient method. In this case, the learning process was stabilized, and the performance was significantly improved. By training the output layer in advance, the loss is reduced; consequently, the backpropagating error signal is suppressed, which is thought to lead to stabilization. However, because the resulting performance still varied among trials, we evaluated the learning performance of SM-RC by repeating the training procedure with different initial weights. Gradient clip [2], which is known as a method to stabilize learning, did not improve the performance

Time evolution of SM-RC for Lorentz model tasks

Figure 3 shows the time evolution of the SM-RC models trained on the Lorentz tasks when the task conditions σ^{in}, N^{forward} were changed. As shown, in all the cases, the gates were modulated to adapt to the input

signal. The behaviors of the gates were complex and depended on the task conditions, indicating that the self-modulation mechanism can adapt the reservoir dynamics to tasks in a flexible manner.

References

[1] K. Doya. Bifurcations in the learning of recurrent neural networks. In *[Proceedings] 1992 IEEE International Symposium on Circuits and Systems*, volume 6, pages 2777–2780 vol.6, 1992.

[2] Sanjoy Dasgupta and David McAllester, editors. *On the difficulty of training recurrent neural networks*, volume 28 of *Proceedings of Machine Learning Research*, Atlanta, Georgia, USA, 17–19 Jun 2013. PMLR.

[3] H. Jaeger. The "echo state" approach to analysing and training recurrent neural networks. *Technical Report GMD Report 148, German National Research Center for Information Technology*, 2001.

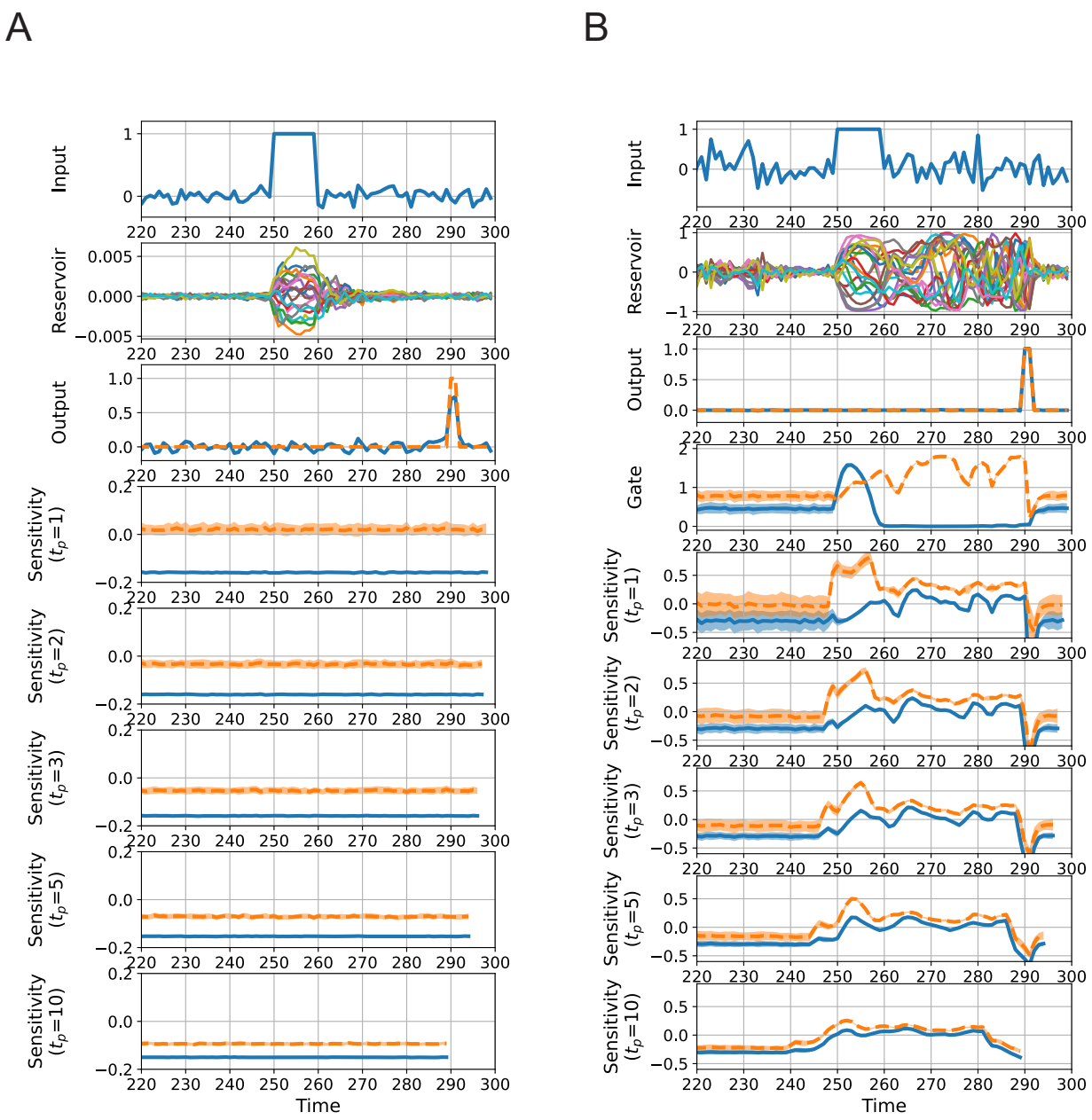

Figure 1: Results of the sensitivity analysis for the simple attention task. A. Results for the case of RC with $N^{\mathrm{res}} = 100$ and $\sigma^{\mathrm{in}} = 0.1$. From the top, the time evolution of the input signal, reservoir states, output, and sensitivity are shown. B. Results for the case of SM-RC with $N^{\mathrm{res}} = 100$ and $\sigma^{\mathrm{in}} = 0.3$. From the top, the time evolution of the input signal, reservoir states, outputs, gates, and sensitivity are shown. In A and B, the sensitivity is analyzed using different timesteps t_p, at which the magnitudes of expansion of the perturbations were measured. For the output layer, the solid line indicates the predicted output, and the dashed line indicates the teacher signal. In the panels displaying the reservoir states, only 20 reservoir neurons are shown.

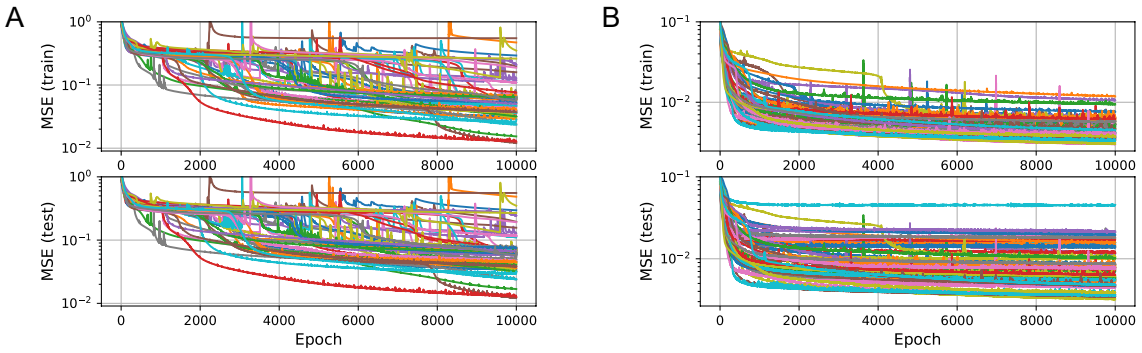

Figure 2: MSEs as a function of the learning epoch in Lorentz model tasks with $N_{\text{forward}} = 20$, $\sigma = 0.03$. 50 different trials are shown with different initial weights for cases where (A) the output layer was also trained via BPTT and (B) the output layer was trained via the pseudo-inverse method.

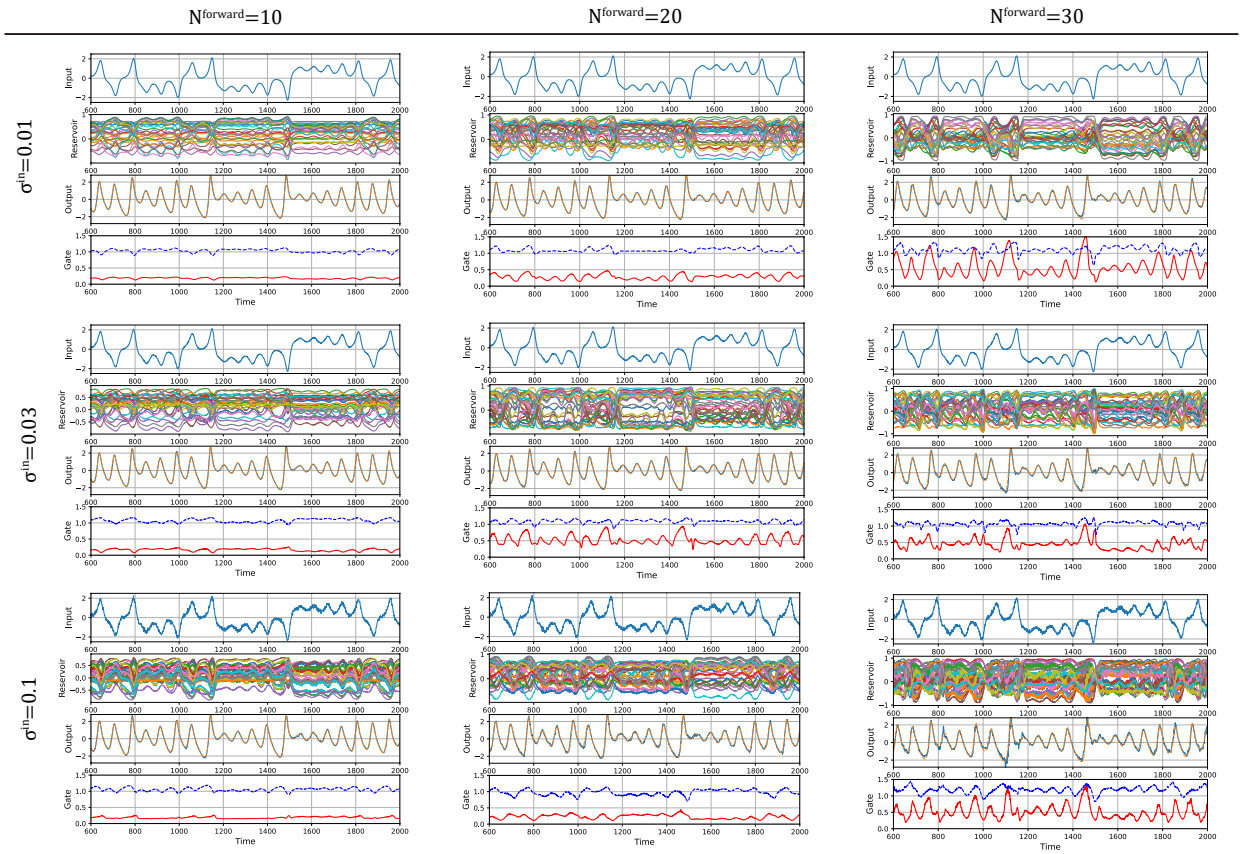

Figure 3: Time evolution of SM-RC for the Lorentz model tasks with various input noises σ^{in} and numbers of prediction steps N^{forward}. For the output layer, the solid lines indicate the predicted outputs, and the dashed lines indicate the teacher signals. For the gates, the solid red lines indicate the input gates, and the blue dashed lines indicate the spectral radius. In the panels displaying the reservoir states, only 20 reservoir neurons are shown.