Local variance optimization for the autonomous regulation of echo state networks

1 Summary

Echo state networks have proven to be a powerful tool in the field of time series prediction (Jaeger, 2001, Lukoševičius and Jaeger, 2009). Several approaches to the optimization of the dynamic reservoir have been investigated in the past, including global tuning for criticality (Livi et al., 2016), as well as local adaptation towards a given output distribution (Schrauwen et al., 2008, Boedecker et al., 2009). The spectral radius $|\Lambda_{\max}|$ of the synaptic weight matrix provides a measure to regulate the network in an appropriate working regime (Caluwaerts et al., 2013). We show that $|\Lambda_{\max}|$ can be regulated by local homeostasis of the variance σ_y^2 of neural activity. This variance control operates on the gain of the neural transfer function and its optimization target depends on the variance σ_{ext}^2 of external inputs: by deriving a relation between recurrent and external input variance, we transferred the explicit spectral radius condition into an implicit expression that defines a manifold in the space of external and recurrent input variances. While our analytically derived condition is only exact for statistically independent external driving across nodes, we have conducted further numerical experiments, suggesting that it also holds under less restrictive conditions. This optimization rule is biologically plausible since it only relies on locally available information. In contrast to previously proposed optimization rules via local intrinsic plasticity, our model relies on the assumption that external and recurrent input signals can be treated as two separate streams of information. The network can hence react autonomously to changes of the input statistics. We demonstrate the importance of this separation by means of network performance —quantified by a nonlinear memory recall task— under varying input statistics.

2 Additional Information

Our analytical analysis is based on a mean-field approach, where one assumes that all presynaptic neurons are statistical independent. Optimizing the variance of the neural activity then consists of solving

$$\sigma_{\mathrm{t}}^2 = \int_{-\infty}^{\infty} \mathrm{dx} \tanh^2(ax) N_{\mu,\sigma}(x), \qquad \sigma^2 = \sigma_{\mathrm{w}}^2 \sigma_{\mathrm{t}}^2 + \sigma_{\mathrm{ext}}^2 \,, \tag{1}$$

where the distribution $N_{\mu,\sigma}(x)$ of the membrane potential x is a Gaussian, with mean $\mu = 0$ and variance σ^2, which is a combination of external input variance σ_{ext}^2, recurrent presynaptic variance σ_{t}^2 and the variance of recurrent synaptic weights σ_{w}^2. An exact analytic expression for this integral does not exist. An approximation with the correct scaling to second order as well as the right convergence is

$$\tanh^2(x) \approx 1 - \exp\left(-x^2\right) \,. \tag{2}$$

Using this approximation in (1) and solving the integral yields

$$\sigma_{\mathrm{t}}^2 = 1 - 1/\sqrt{1 + 2a^2\sigma_{\mathrm{w}}^2 \left(\sigma_{\mathrm{t}}^2 + \sigma_{\mathrm{ext}}^2/\sigma_{\mathrm{w}}^2\right)} \,. \tag{3}$$

A unit spectral radius translates to the condition $a = \sigma_{\mathrm{w}}^{-1}$, allowing us to state a further simplified relation, which is what is required for tuning the network activity into an optimal state:

$$\sigma_{\mathrm{t}} \approx \left[\sqrt{\frac{3}{2}}\frac{\sigma_{\mathrm{w}}}{\sigma_{\mathrm{ext}}} + 1\right]^{-1/2} \,. \tag{4}$$

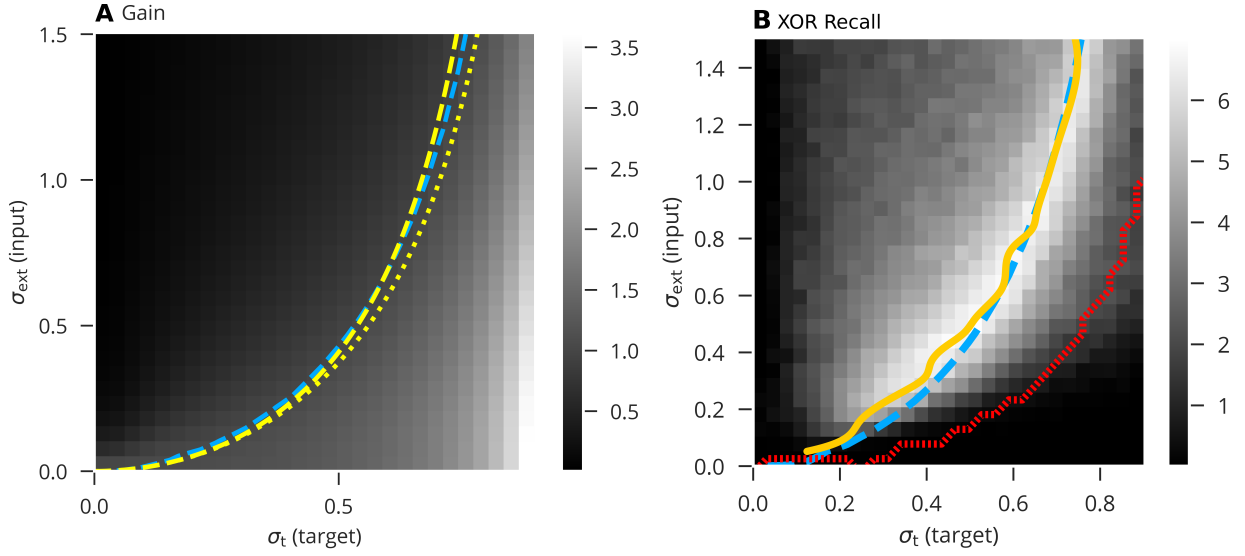A comparison between a numerical evaluation of this manifold and the analytical results is shown in Fig. 1A.

Figure 1: **Spectral Radius Transition Approximation A**: Colormap encodes the spectral radius of recurrent connectivity. Blue dashed line marks unit spectral radius. Yellow dashed line is given by (4), dotted line by the solution of (3) for $a = \sigma_{\mathrm{w}}^{-1}$. **B**: Colormap encodes $\mathrm{MC_{XOR}}$ as defined in (6). Blue dashed line as in **A**. The orange line plots the position of maximal $\mathrm{MC_{XOR}}$ for a given σ_{ext}. Red striped line marks the loss of the Echo-State property.

Furthermore, we investigated how the network performs under different σ_{t} and σ_{ext} if given a task that cannot be solved without nonlinear transformations on the input. For this purpose, we defined a XOR-recall task with delay τ by

$$f(t) = \mathrm{XOR}\left[u(t-\tau), u(t-\tau-1)\right], \qquad \mathrm{XOR}\left[x, y\right] \equiv \begin{cases} 0 & x = y \\ 1 & \text{else} \end{cases}. \qquad (5)$$

We defined the total XOR memory capacity by

$$\mathrm{MC_{XOR}} \equiv \sum_{k=1}^{\infty} \mathrm{MC}_{\mathrm{XOR},k}, \qquad \mathrm{MC}_{\mathrm{XOR},k} \equiv \max_{w_{\mathrm{out}}} \frac{\mathrm{Cov}^2\left[\mathrm{XOR}\left[u(t-k), u(t-k-1)\right], y_{\mathrm{out}}(t)\right]_t}{\mathrm{Var}\left[\mathrm{XOR}\left[u(t), u(t-1)\right]\right]_t \mathrm{Var}\left[y_{\mathrm{out}}(t)\right]_t}. \qquad (6)$$

The results are shown in Fig. 1B. The maximally achievable performance remains relatively constant over a wide range of σ_{ext}. This local maximum as a function of σ_{ext} is well described by the condition of unit spectral radius and could therefore be homeostatically achieved by means of our implicit relation (4).

References

Boedecker, J., Obst, O., Mayer, N. M., and Asada, M. (2009). Initialization and self-organized optimization of recurrent neural network connectivity. *HFSP Journal*, 3(5):340–349.

Caluwaerts, K., Wyffels, F., Dieleman, S., and Schrauwen, B. (2013). The spectral radius remains a valid indicator of the echo state property for large reservoirs. In *IEEE International Joint Conference on Neural Networks (IJCNN)*, page 6.

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks. GMD Report 148, GMD - German National Research Institute for Computer Science.

Livi, L., Bianchi, F. M., and Alippi, C. (2016). Determination of the edge of cricritical in echo state networks through Fisher information maximization. *arXiv:1603.03685v2*.

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

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