

Local Variance Optimization for the Autonomous Regulation of Echo State Networks

Fabian Schubert, Claudius Gros

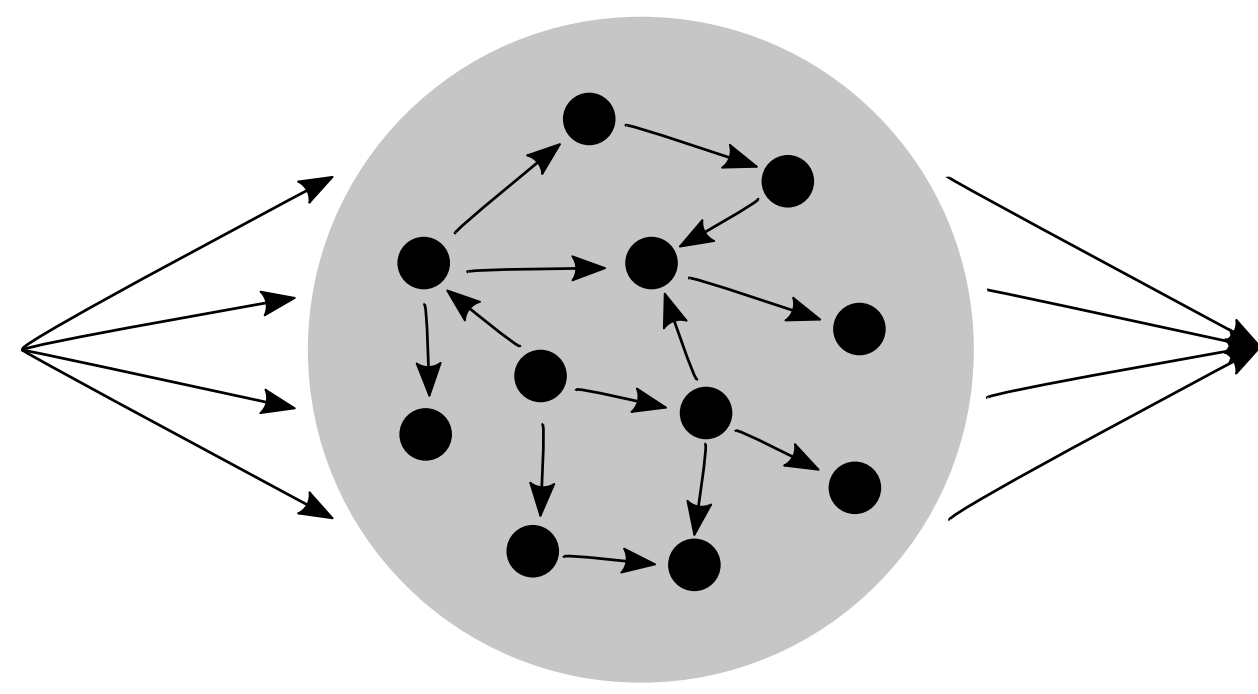
Institute for Theoretical Physics, Goethe University Frankfurt a.M.

Introduction

- Echo state networks have proven to be a powerful tool in the field of time series prediction [1].
- The spectral radius $|\Lambda_{\max}|$ of the synaptic weight matrix provides a measure to regulate the network in an appropriate working regime [4]. 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 input.
- 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.

Model Description

$$\begin{aligned} y_i^{t+1} &= \tanh(a_i^t X_i^{t+1}) & X_i^{t+1} &= \sum_{j=1}^N W_{ij} y_j^t + E_i^{t+1} \\ b_i^{t+1} &= b_i^t + \epsilon_b [\langle y_i \rangle_T - \mu_t] & a_i^{t+1} &= a_i^t + \epsilon_a [\sigma_t^2 - (y_i^t - \bar{y}(t)_i)^2] \\ \bar{y}_i^{t+1} &= \epsilon_{\text{trail}} [y_i^{t+1} - \bar{y}_i^t] \end{aligned}$$



- W_{ij} is a sparse random matrix with connection probability p . Nonzero entries were drawn from a Gaussian distribution $\mathcal{N}(\mu = 0, \sigma = 1/\sqrt{Np})$. Diagonal entries were always set to zero.
- E_i^t are random vectors of size N with independent entries drawn from a Gaussian distribution $\mathcal{N}(\mu = 0, \sigma = \sigma_{\text{ext}})$.
- By changing individual gain and bias values a_i and b_i , the homeostatic control tries to drive the activity standard deviation and mean of every cell to the value given by σ_t and μ_t .
- By controlling the output variance of the neural activity, we can tune the network into a regime that exhibits subcritical, but transiently active dynamics in the absence of external input.
- This led to the question how a homeostatic variance control could be used to tune network properties into a desired regime, even under changing input statistics.

Self-Consistency Equation

If we assume no correlations in time across the neural population, we can formulate a self-consistency equation, given by

$$\sigma_t^2 = \int_{-\infty}^{\infty} \tanh^2(ay) \mathcal{N}(y, \mu = 0, \sigma = \sqrt{\sigma_t^2 + \sigma_{\text{ext}}^2}) dy.$$

A useful approximation with the correct scaling to second order as well as the right convergence is $\tanh^2(x) \approx 1 - \exp(-x^2)$.

From this approximation we can derive a relation between a target variance σ_t and the external input variance σ_{ext} that marks the point where the spectral radius is unity:

$$\sigma_t \approx \left[\sqrt{\frac{3}{2}} \frac{\sigma_w}{\sigma_{\text{ext}}} + 1 \right]^{-1/2}.$$

This approximate solution is shown in Fig. 2.

Scaling of Spectral Radius for Inhomogeneous Gains

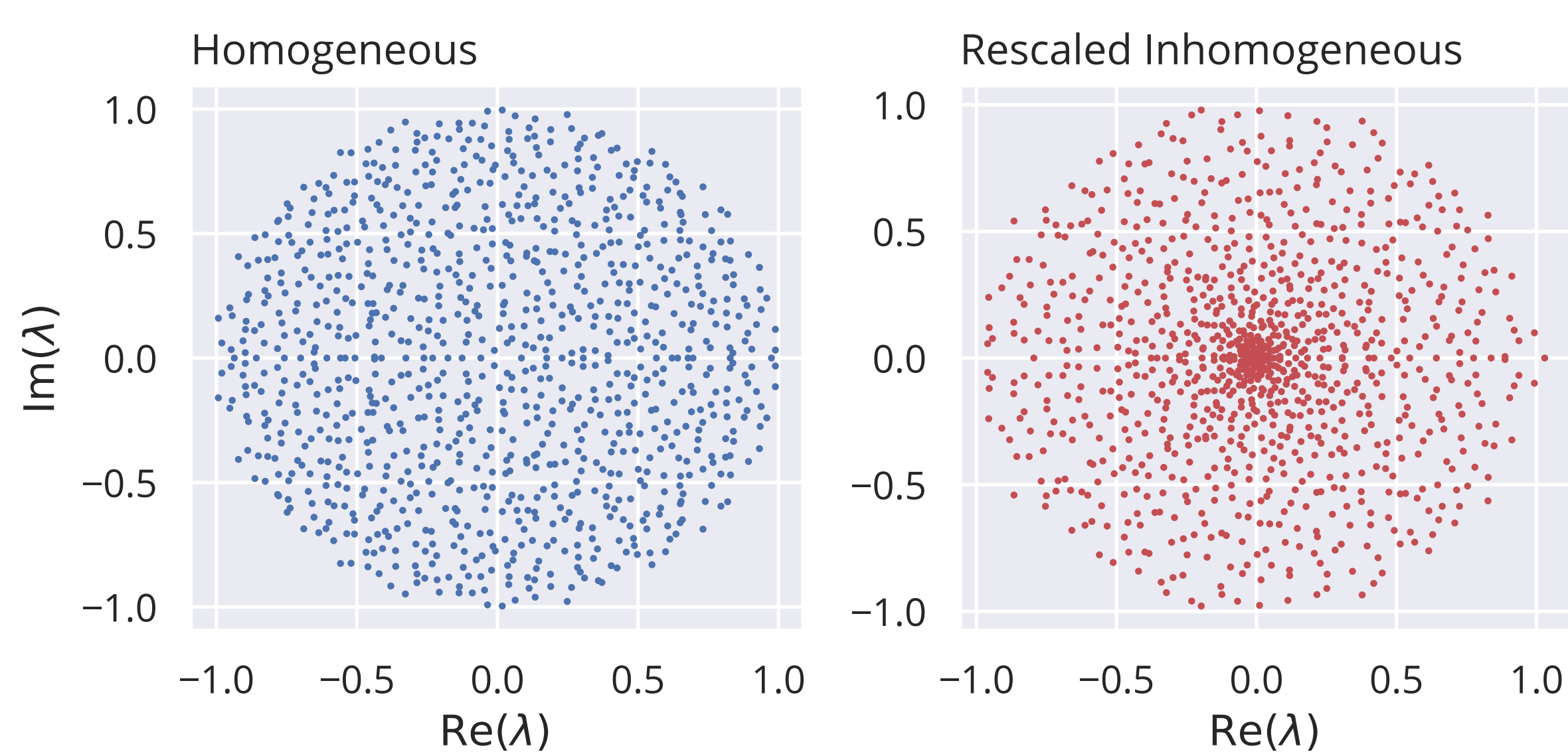
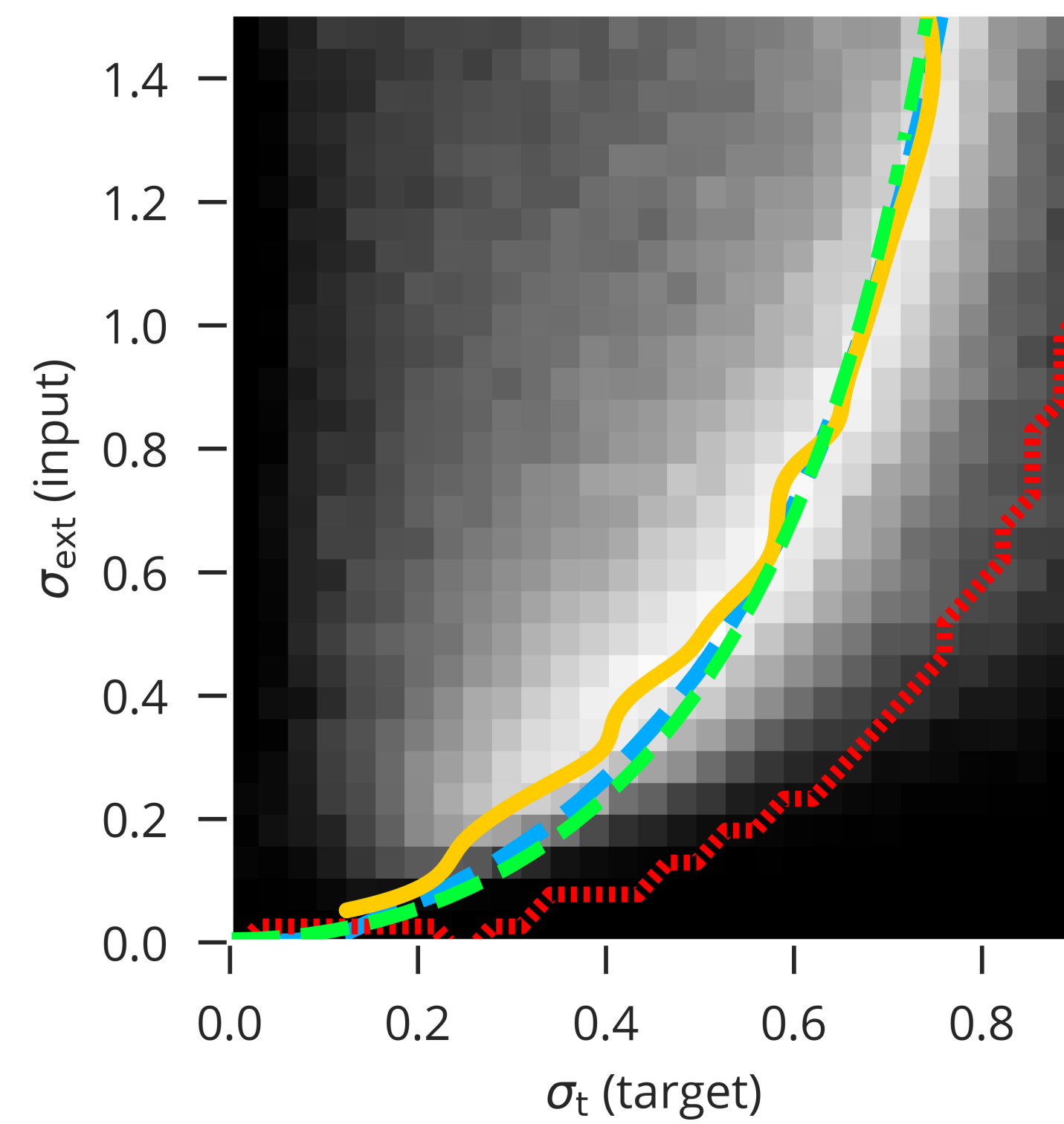


Figure 1: A: Distribution of eigenvalues for all gains set to 1. B: Gains drawn from a uniform distribution, such that $\sqrt{\sum_i a_i^2/N} = 1$.

In the self-consistency equation, we assumed a single gain parameter a controlling the spectral radius. For an *inhomogeneous* distribution of gains, we found that the spectral radius is preserved if $\sqrt{\sum_i a_i^2/N}$ remains constant. However, inhomogeneity caused a bunching of eigenvalues towards zero and reduced the task performance of the network.

Nontrivial Memory Task



- We defined a XOR-Memory recall task given by $f(t) = \text{XOR}[u(t - \tau), u(t - \tau - 1)]$.
- We tested the performance of the ESN after homeostasis has equilibrated for different pairs of input and target variances, see Fig. 2.
- The analytically derived expression provides a good fit to the optimal performance marked by the orange line.

Figure 2: XOR - Memory capacity, orange line marks max. MC for a given σ_{ext} , red line the loss of the echo state property. The blue line marks where the spectral radius is unity, the green line is our analytically derived prediction. The loss of the echo state property is marked by the red line.

Gain Optimization: Local vs Full Gradient

- We were interested in the gain adaptation dynamics for our particular task. In particular, we compared a simplified, local a error gradient based gain adaptation against its full counterpart.
- We calculated the full gradient using the Realtime-Recurrent-Learning (RTRL) scheme with a squared error loss function \mathcal{L} . The nonlocal term was dropped in the approximate rule.

$$\begin{aligned} \mathcal{L} &= \left\langle \frac{1}{2} \epsilon^2(t) \right\rangle_t = \left\langle \frac{1}{2} (\mathbf{w}_{\text{out}} \mathbf{y}(t) - f(t))^2 \right\rangle_t \\ \Delta a &= \partial_a \mathcal{L} = \langle \epsilon(t) \mathbf{w}_{\text{out}} \partial_a \mathbf{y}(t) \rangle_t \\ \partial_a \mathbf{y}(t) &= \mathbf{y}'(t) * [\mathbf{X}(t) + a \mathbf{W} \partial_a \mathbf{y}(t - 1)] \\ \Delta a_{\text{approx}} &= \langle \epsilon(t) \mathbf{w}_{\text{out}} \mathbf{y}'(t) * \mathbf{X}(t) \rangle_t \\ y_i'(t) &= 1 - y_i^2(t) \end{aligned}$$

- This approximation is similar to recently proposed local synaptic learning rules for recurrent networks [5].

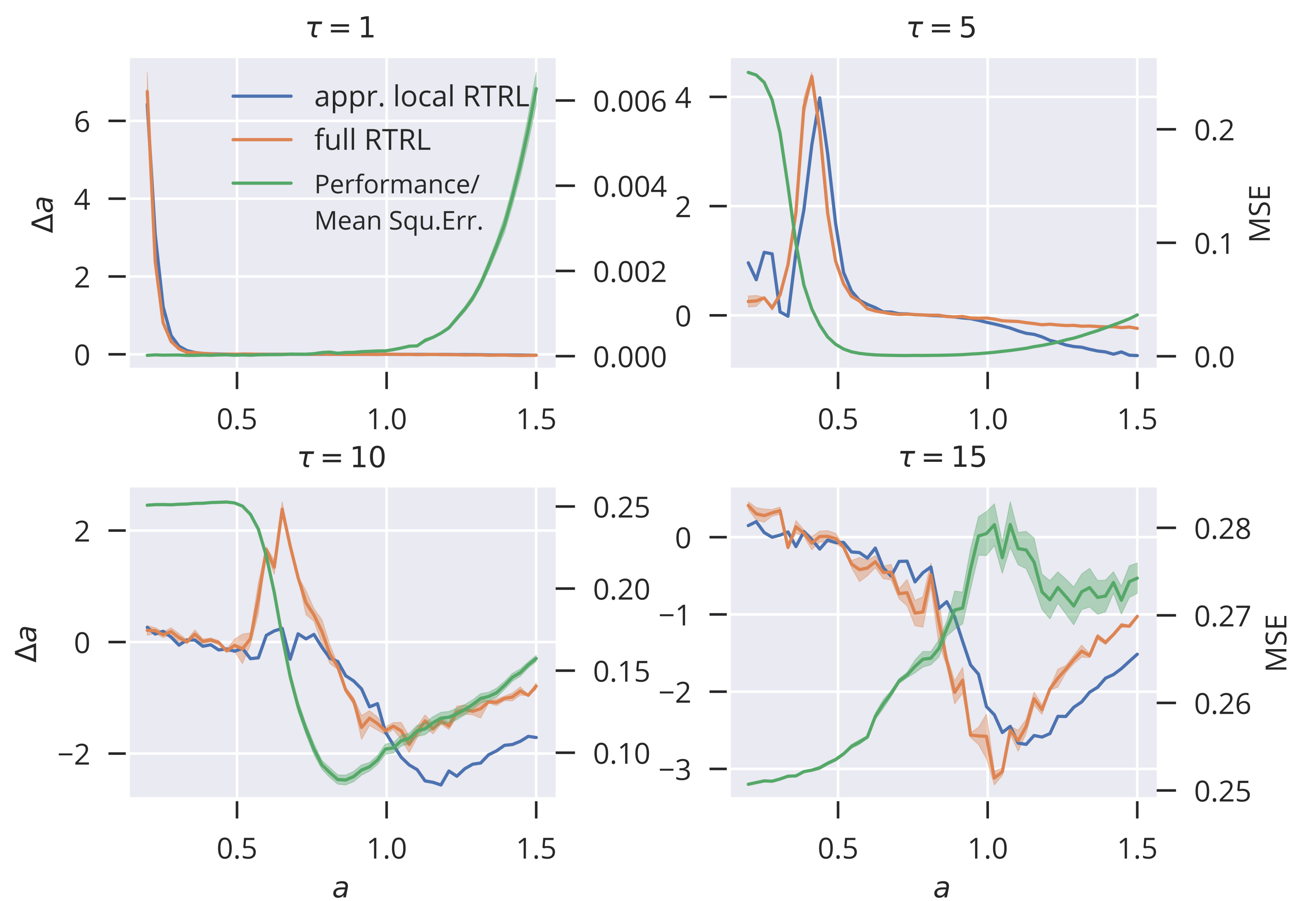


Figure 3: Gradient based update rules for the gain parameter a : Full gradient (orange) vs. local approximation (blue).

Outlook: Homeostasis and Error-Driven Gain Optimization

- Combining our homeostatic update rule with the error-driven approach might help to avoid the “trivial” fixpoint of zero gains.
- Furthermore, tuning the network into a state close to the critical transition provides a good starting point for error-driven optimization.

References

- [1] H. Jaeger. The “echo state” approach to analysing and training recurrent neural networks. GMD Report 148, GMD - German National Research Institute for Computer Science, 2001.
- [2] L. Livi, F. M. Bianchi, and C. Alippi. Determination of the edge of criticality in echo state networks through Fisher information maximization. *arXiv:1603.03685v2*, 2016.
- [3] Benjamin Schrauwen, Marion Wardermann, David Verstraeten, Jochen J. Steil, and Dirk Stroobandt. Improving reservoirs using intrinsic plasticity. *Neurocomputing*, 71(7-9):1159–1171, mar 2008.
- [4] Ken Caluwaerts, Francis Wyffels, Sander Dieleman, and Benjamin Schrauwen. 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, 2013.
- [5] James M Murray. Local online learning in recurrent networks with random feedback. *eLIFE*, 2019.