Bifurcations and loss jumps in RNN training

Lukas Eisenmann[1,2,*] , Zahra Monfared[1,*], Niclas Göring [1,2], and Daniel Durstewitz[1,2,3]

1 Department of Theoretical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany
2 Faculty of Physics and Astronomy, Heidelberg University, Heidelberg, Germany
3 Interdisciplinary Center for Scientific Computing, Heidelberg University
* These authors contributed equally

PLRNN

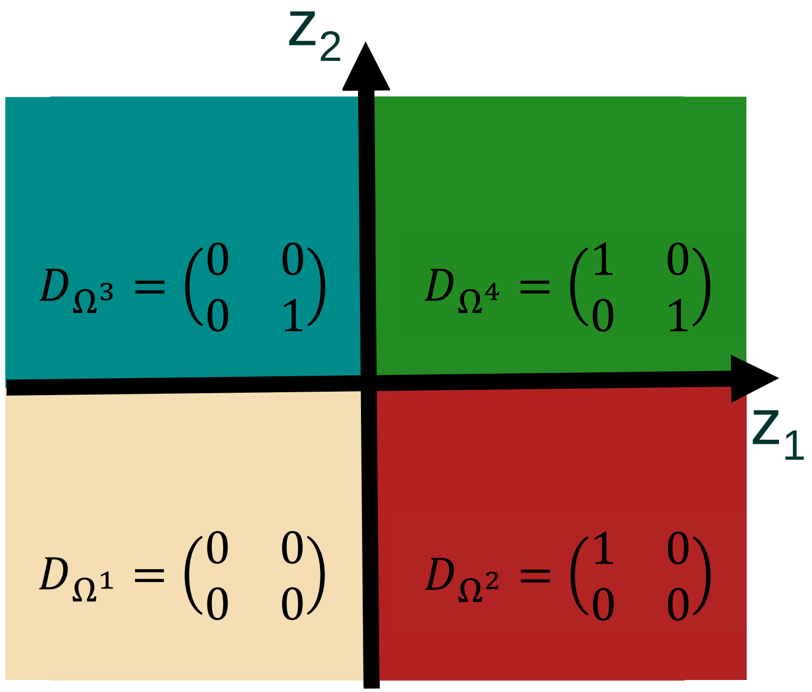
$$z_t = F(z_{t-1}) = A\, z_{t-1} + W\, ReLU(z_{t-1}) + C\, s_t + h$$

$A \in \mathbb{R}^{MxM}$ h $\in \mathbb{R}^{M}$ $W \in \mathbb{R}^{MxM}$ $C \in \mathbb{R}^{MxK}$ $s_t \in \mathbb{R}^{K}$ $z_t \in \mathbb{R}^{M}$

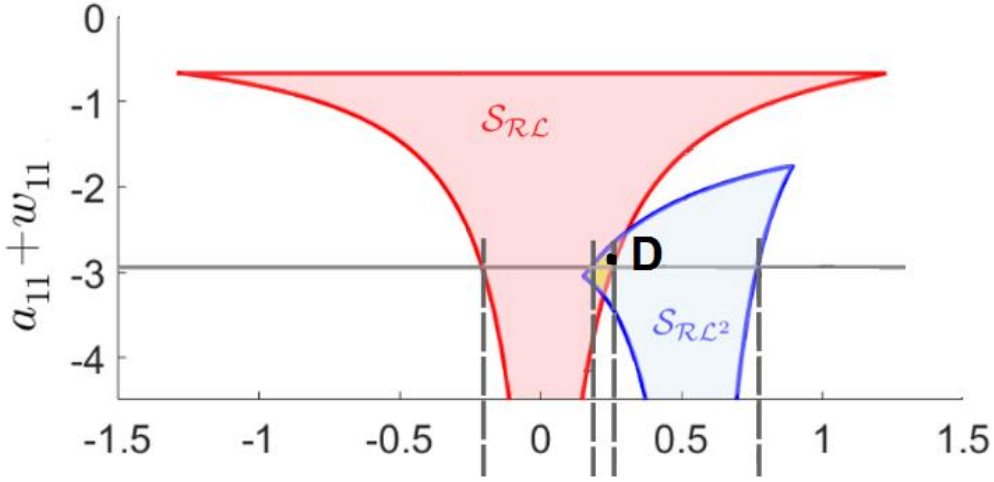
$$D_{\Omega(t)} \coloneqq diag(d_{\Omega,t}) \text{ with } d_{m,t} = \begin{cases} 1, & z_{m,t} > 0 \\ 0, & else \end{cases}$$

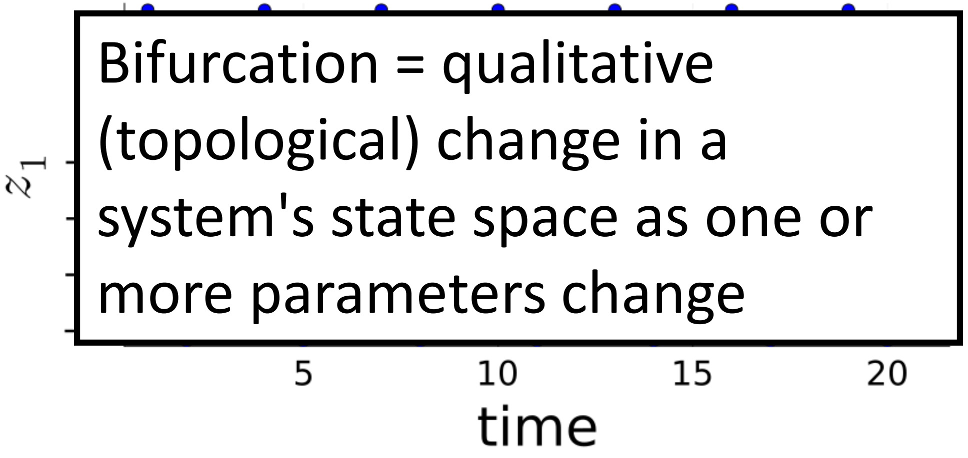
$$z_t = F(z_{t-1}) = \left(A + W D_{\Omega(t-1)}\right) z_{t-1} + C\, s_t + h$$

$$=: W_{\Omega(t-1)}\, z_{t-1} + C\, s_t + h$$



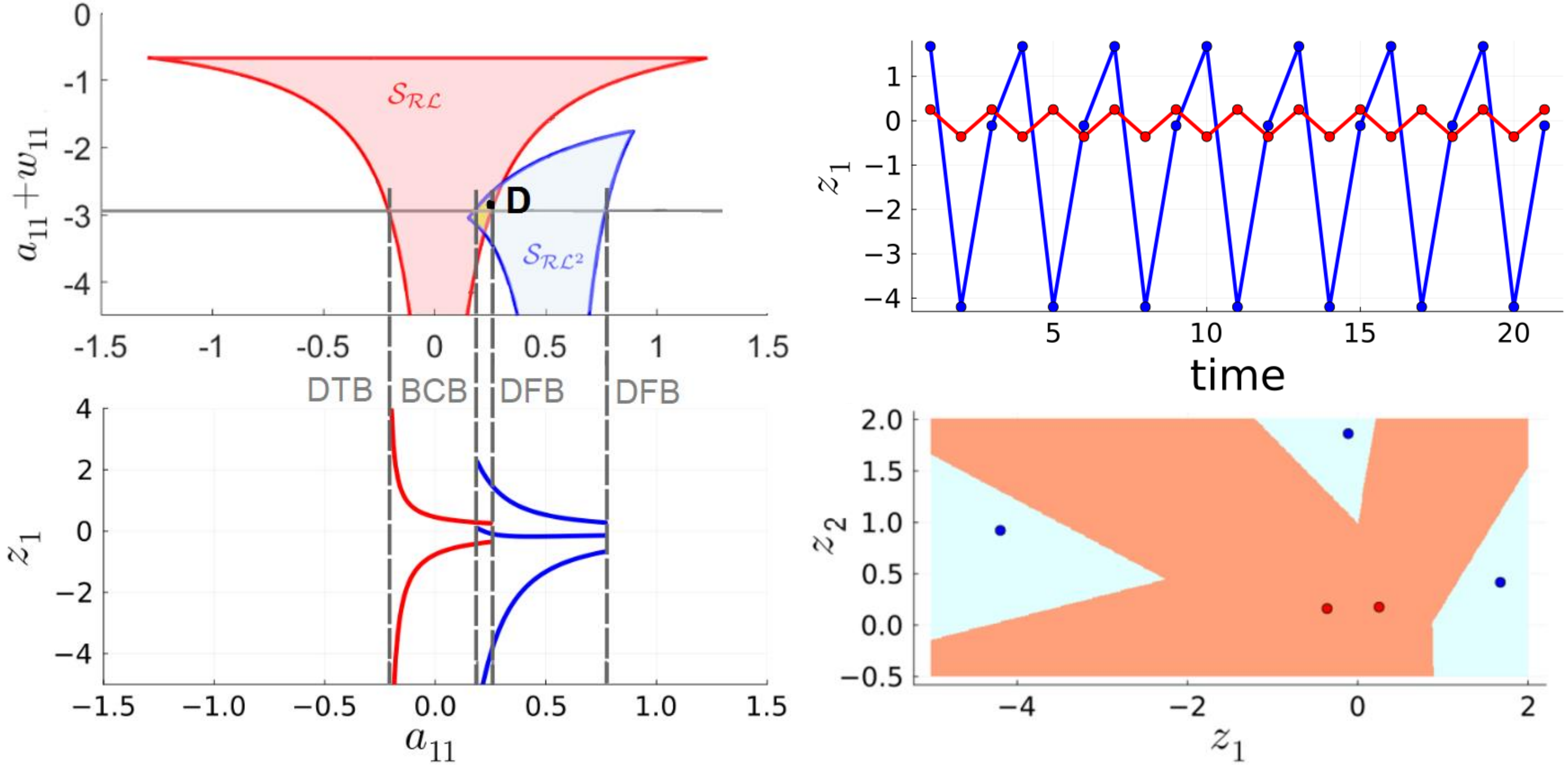
Bifurcation manifolds in PLRNN parameter space





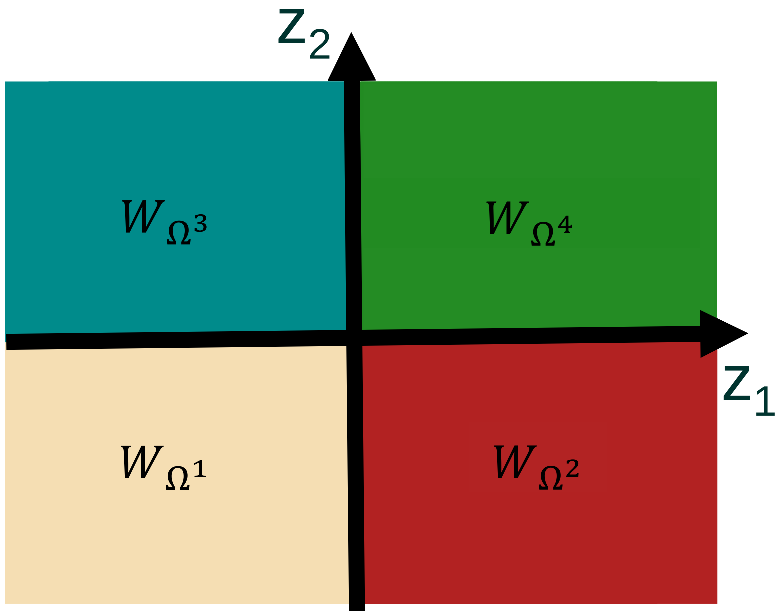
Bifurcation = qualitative (topological) change in a system's state space as one or more parameters change

Bifurcation manifolds in PLRNN parameter space



Searcher for Cycles and Fixed points: SCYFI

- Mathematically tractable: Allows for semi-analytic calculation of fixed points and cycles
- BUT: Combinatorial problem!



→ Number of linear regions: 2^{Mk}

Select initial set of linear subregions $W_{\Omega_{init}}$

Solve for cycle candidate

$$z^*_{k,candidate} = \left(\mathbb{1} - \prod_{r=0}^{k-1} W_{\Omega(k-r)}\right)^{-1} \left[\sum_{j=2}^{k-1} \prod_{r=0}^{k-j} W_{\Omega(k-r)} + \mathbb{1}\right] + h$$

Determine linear subregions corresponding to cycle candidate

$$W_{\Omega_{candidate}} = \left(A + W\, diag(d_{\Omega,t})\right) \text{ with } d_{\Omega,t} = \begin{cases} 1, & z^*_{k,candidate} > 0 \\ 0, & else \end{cases}$$

If $W_{\Omega_{candidate}} = W_{\Omega_{init}}$

Real FP/*k*-cycle found

Determine if **stable, unstable or saddle**

Done

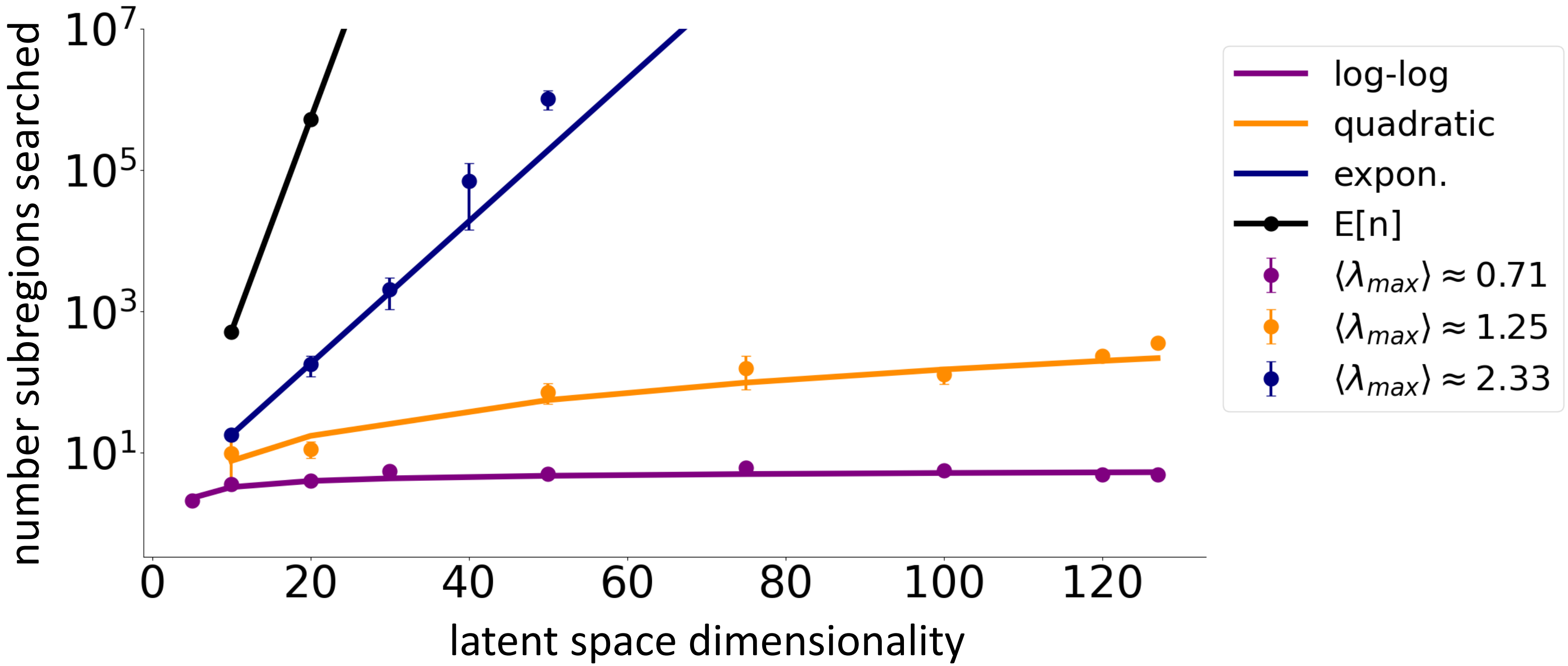
Else

virtual FP/*k*-cycle

$W_{\Omega_{init}} \rightarrow W_{\Omega_{candidate}}$

SCYFI: Scaling

Theorem 3. Under the condition $\|A\| + \|W\| < 1$, SCYFI will converge in at most linear time

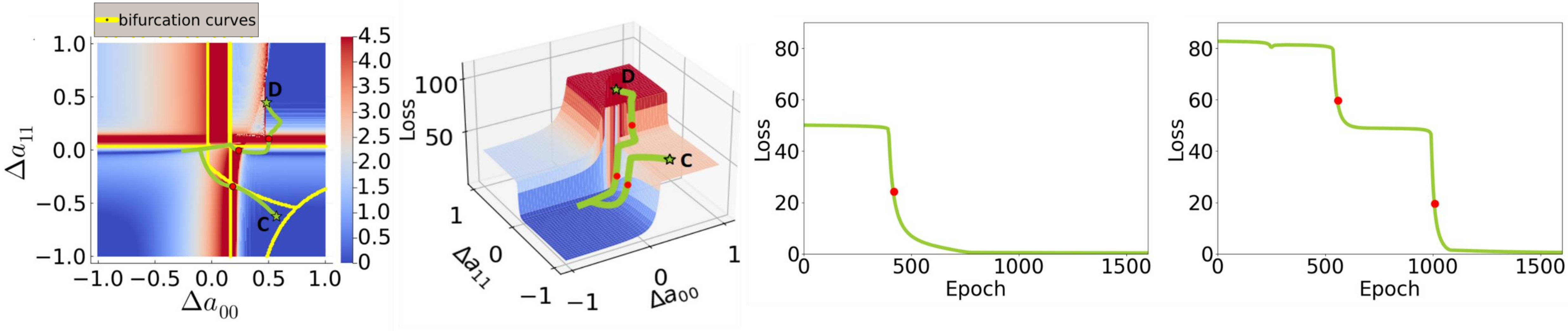


Theorems: Bifurcations and exploding/ vanishing gradients

Theorem 1. If a stable fixed point or a *k*-cycle undergoes a degenerate transcritical bifurcation, the norm of the PLRNN loss gradient tends to infinity $\lim_{t\to\infty} \left\| \frac{dL}{d\theta} \right\| = \infty$

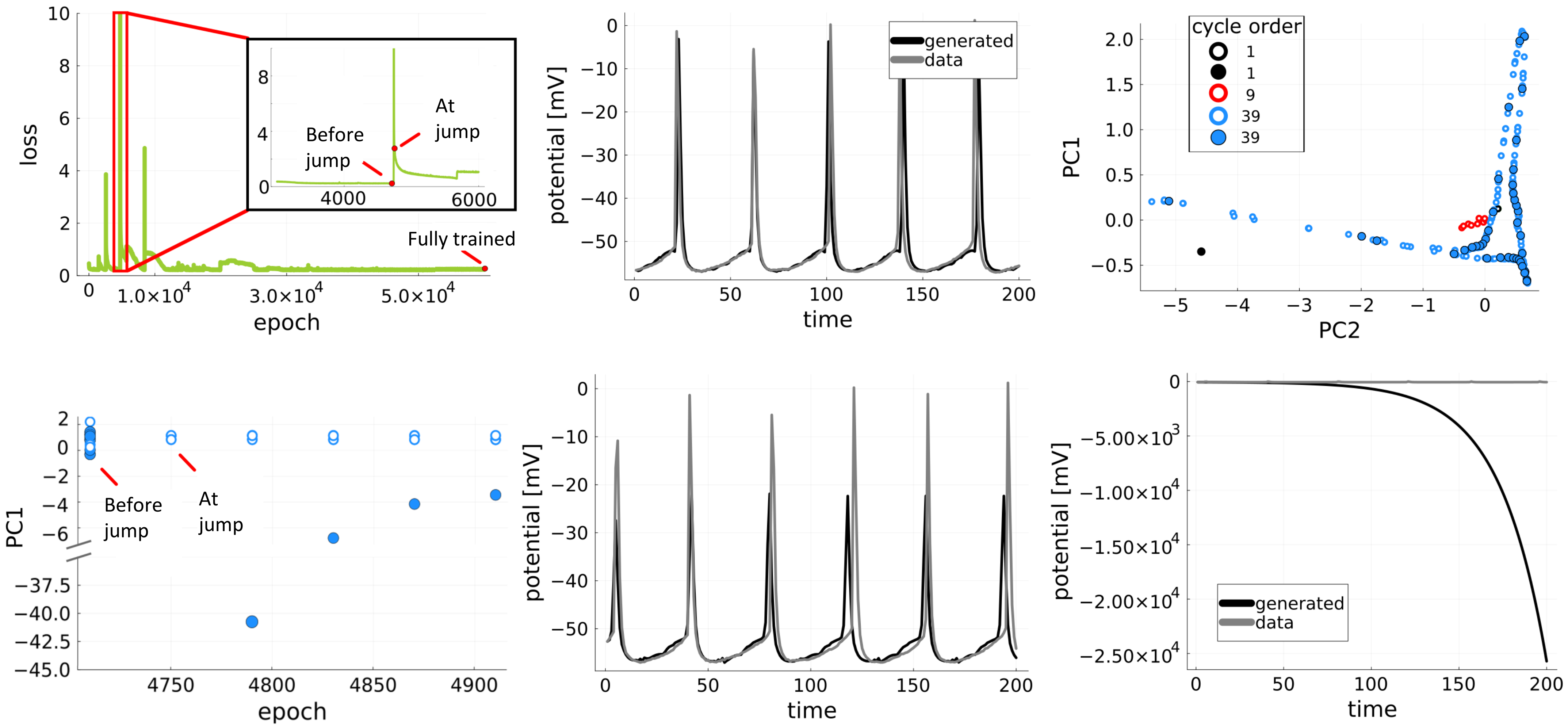
Theorem 2. If a stable fixed point or a *k*-cycle undergoes a border collision bifurcation, the norm of the PLRNN loss gradients vanishes $\lim_{t\to\infty} \left\| \frac{dL}{d\theta} \right\| = 0$

Toy example

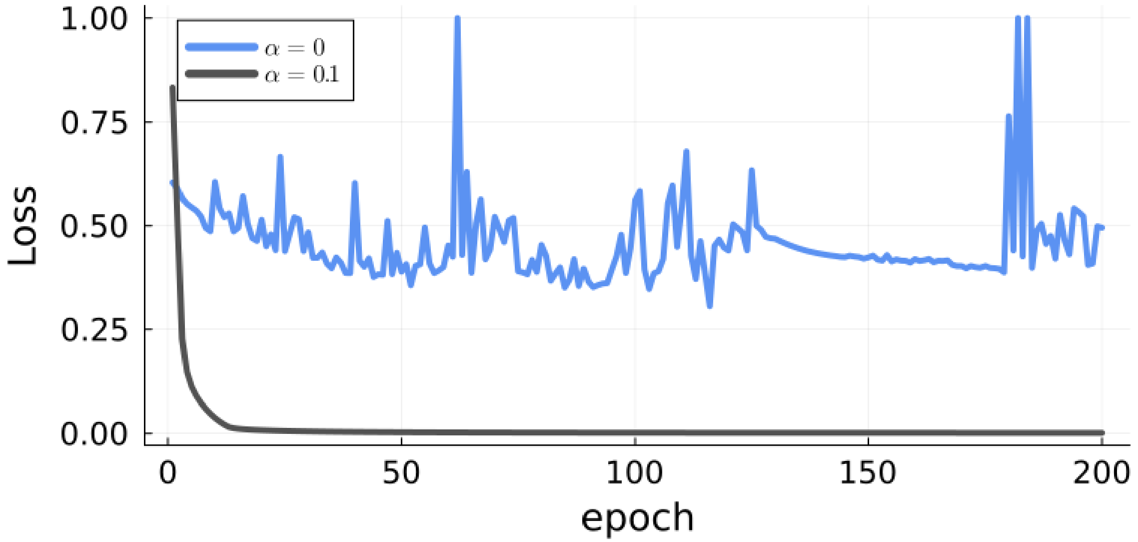


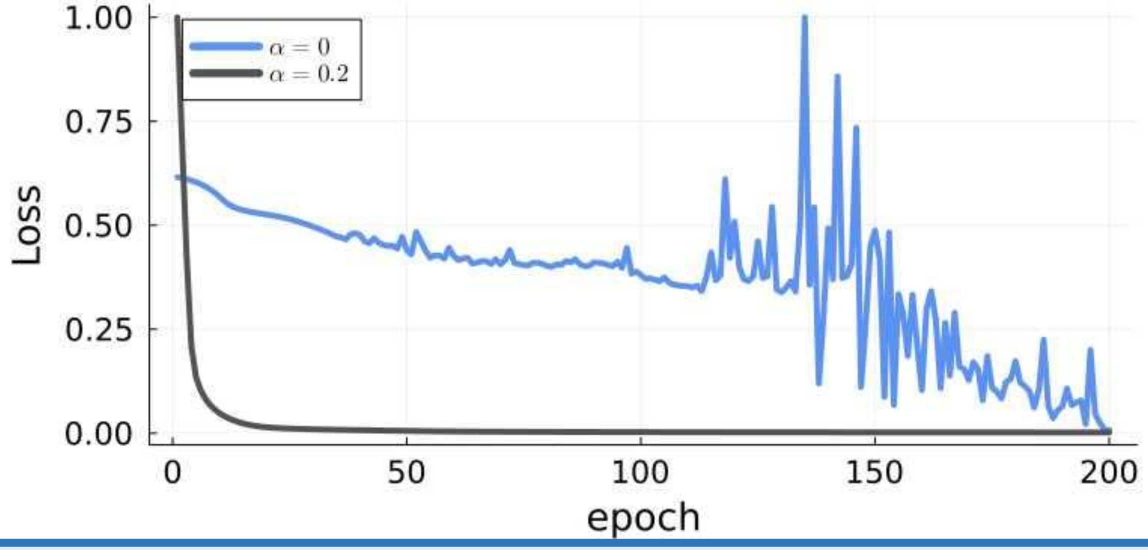
→ Bifurcations can cause jumps in the loss

Empirical example: Training PLRNN on membrane voltage traces of real cell



Generalized Teacher Forcing (GTF)[1] prevents bifurcations in training





Theorem 3.

If $||A|| + ||W|| < 1$ then for any $0 < \alpha < 1$ GTF controls the system, preventing degenerate transcritical bifurcations.

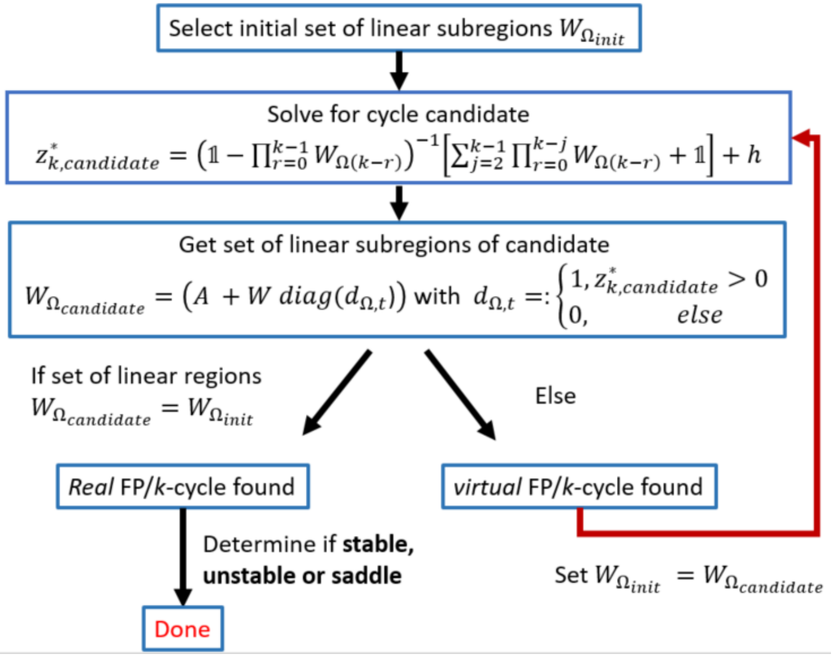
If $||A|| + ||W|| = r > 1$ then for any $1 - \frac{1}{r} < \alpha < 1$ GTF prevents degenerate transcritical bifurcations.

[1] F. Hess, Z. Monfared, M. Brenner, and D. Durstewitz. Generalized teacher forcing for learning chaotic dynamics. In Proceedings of the 40th International Conference on Machine Learning, volume 202 of Proceedings of Machine Learning Research Jul 2023

Conclusion

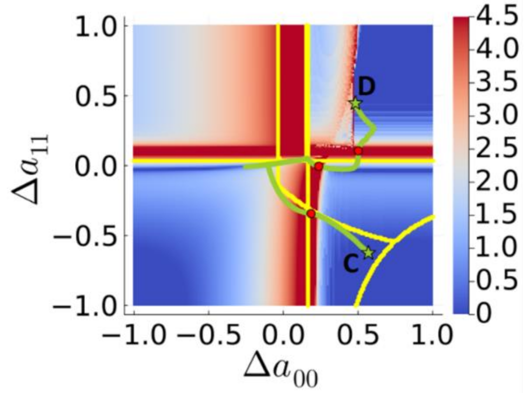
SCYFI

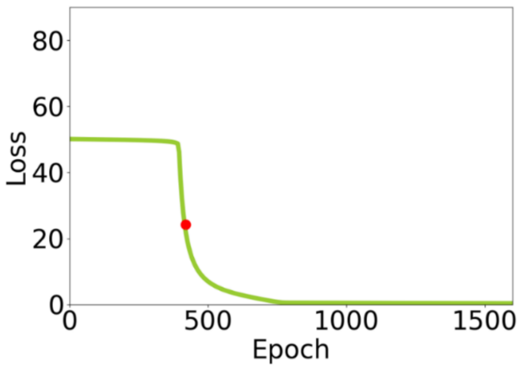
- Computes fixed points *and* k-cycles exactly
- Efficient: surprisingly good, often linear, scaling



Theorems

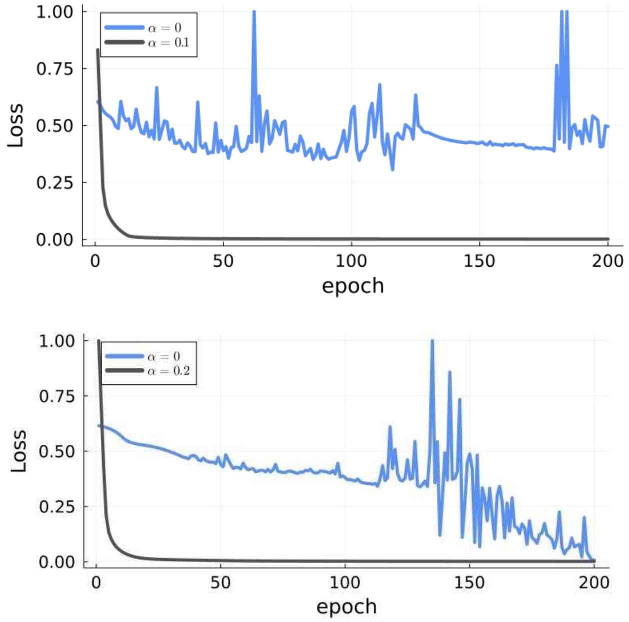
Formal connection between bifurcations and exploding or vanishing gradients





Implications for RNN training

Generalized teacher forcing provably avoids bifurcations in training



Thanks for your attention!

This work was supported by the German Research Foundation (DFG) through individual grants Du 354/10-1 & Du 354/15-1 to DD, within research cluster FOR-5159 (“Resolving prefrontal flexibility”; Du 354/14-1), and through the Excellence Strategy EXC 2181/1 – 390900948 (STRUCTURES)