

A Appendix

A.1 Analysis of bifurcation curves

A.1.1 PLRNNs

The standard PLRNN [18], given in eq. (1) in sect. 3.1, was defined by

$$\mathbf{z}_t = F_{\theta}(\mathbf{z}_{t-1}, \mathbf{s}_t) = \mathbf{A} \mathbf{z}_{t-1} + \mathbf{W} \phi(\mathbf{z}_{t-1}) + \mathbf{C} \mathbf{s}_t + \mathbf{h},$$

where $\phi(\mathbf{z}_{t-1}) = \max(\mathbf{z}_{t-1}, 0)$. There are various extensions of this basic architecture like the dendPLRNN [10] or the ‘shallow PLRNN’ (shPLRNN) [24], as used in sect. 5.1 for training on single cell membrane potentials. The latter is essentially a 1-hidden-layer version of the form

$$\mathbf{z}_t = F_{\theta}(\mathbf{z}_{t-1}, \mathbf{s}_t) = \mathbf{A} \mathbf{z}_{t-1} + \mathbf{W}_1 \phi(\mathbf{W}_2 \mathbf{z}_{t-1} + \mathbf{h}_2) + \mathbf{C} \mathbf{s}_t + \mathbf{h}_1, \quad (6)$$

with $\mathbf{W}_1 \in \mathbb{R}^{M \times L}$ and $\mathbf{W}_2 \in \mathbb{R}^{L \times M}$, $L \geq M$, connectivity matrices, $\mathbf{h}_1 \in \mathbb{R}^M$, $\mathbf{h}_2 \in \mathbb{R}^L$ bias terms, and all other parameters and variables as in eq. (1). While this formulation is beneficial for training, the shPLRNN can essentially be rewritten in standard PLRNN form (see [24]).

Assume that $\mathbf{D}_{\Omega(t)} := \text{diag}(\mathbf{d}_{\Omega(t)})$ is a diagonal matrix with an indicator vector $\mathbf{d}_{\Omega(t)} := (d_1, d_2, \dots, d_M)$ such that $d_m(z_{m,t}) =: d_m = 1$ whenever $z_{m,t} > 0$, and zero otherwise. Then eq. (1) can be rewritten as

$$\mathbf{z}_t = (\mathbf{A} + \mathbf{W} \mathbf{D}_{\Omega(t-1)}) \mathbf{z}_{t-1} + \mathbf{C} \mathbf{s}_t + \mathbf{h} =: \mathbf{W}_{\Omega(t-1)} \mathbf{z}_{t-1} + \mathbf{C} \mathbf{s}_t + \mathbf{h}.$$

Let us ignore the inputs for simplicity. There are 2^M different configurations for matrix $\mathbf{D}_{\Omega(t-1)}$ and so 2^M different forms for matrix $\mathbf{W}_{\Omega(t-1)}$ in the system

$$\mathbf{z}_t = F_{\theta}(\mathbf{z}_{t-1}) = \mathbf{W}_{\Omega(t-1)} \mathbf{z}_{t-1} + \mathbf{h}. \quad (7)$$

Thus, the phase space of the system is divided into 2^M sub-regions corresponding to the indexed matrices

$$\mathbf{W}_{\Omega^k} := \mathbf{A} + \mathbf{W} \mathbf{D}_{\Omega^k}, \quad k = 1, 2, \dots, 2^M, \quad (8)$$

see [38, 39] for more details. For $M = 2$, assuming

$$\mathbf{W} = \begin{pmatrix} w_{11} & 0 \\ w_{21} & 0 \end{pmatrix}, \quad (9)$$

in (8), we have

$$\begin{aligned} \mathbf{W}_{\Omega^1} = \mathbf{W}_{\Omega^3} &= \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} = \mathbf{A}, \\ \mathbf{W}_{\Omega^2} = \mathbf{W}_{\Omega^4} &= \begin{pmatrix} a_{11} + w_{11} & 0 \\ w_{21} & a_{22} \end{pmatrix}. \end{aligned} \quad (10)$$

Hence, for this parameter constellation, the map simplifies as there exists only one border which divides the phase space into two distinct sub-regions, such that (7) can be rewritten as a map of the form

$$\begin{aligned} \begin{pmatrix} z_{1,t} \\ z_{2,t} \end{pmatrix} &= T(z_{1,t-1}, z_{2,t-1}) \\ &= \begin{cases} T_{\mathcal{L}}(z_{1,t-1}, z_{2,t-1}) = \underbrace{\begin{pmatrix} a_l & c \\ b_l & d \end{pmatrix}}_{\mathbf{A}_{\mathcal{L}}} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}; & z_{1,t-1} \leq 0 \\ T_{\mathcal{R}}(z_{1,t-1}, z_{2,t-1}) = \underbrace{\begin{pmatrix} a_r & c \\ b_r & d \end{pmatrix}}_{\mathbf{A}_{\mathcal{R}}} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} + \begin{pmatrix} h_1 \\ h_2 \end{pmatrix}; & z_{1,t-1} \geq 0 \end{cases}, \quad (11) \end{aligned}$$

with $a_l = a_{11}$, $a_r = a_{11} + w_{11}$, $b_r = w_{21}$, $d = a_{22}$, $b_l = c = 0$. The map (11) is a PWL dynamical system whose phase space is split into left and right half-planes (sub-regions) by the borderline Σ (z_2 -axis). Note that bifurcation curves of the $2d$ PLRNN (7) in the $(a_{11}, a_{11} + w_{11})$ -parameter space can be determined analogous to those of the PWL map (11) in the (a_l, a_r) -parameter space.

Another way to simplify the PLRNN to a $2d$ ($M = 2$) PWL map with just a single border is to remove one of the ReLU nonlinearities and define $\phi(z_{t-1}) = (\phi_1(z_{1,t-1}), \beta z_{2,t-1})^\top$, where $\beta \in \mathbf{R}$ and ϕ_1 is some variant of the *ReLU* function such as the leaky or parametric *ReLU* given by

$$\phi_1(z) = \begin{cases} z; & z > 0 \\ \alpha z; & z \leq 0 \end{cases} \quad (\alpha \in \mathbf{R}). \quad (12)$$

Then $D_{\Omega(t)} := \text{diag}(d_1, \beta)$ such that

$$d_1(z_{1,t}) =: d_1 = \begin{cases} 1; & z_{1,t} > 0 \\ \alpha; & z_{1,t} \leq 0 \end{cases}, \quad (13)$$

and so

$$\begin{aligned} W_{\Omega^1} = W_{\Omega^3} &= \begin{pmatrix} a_{11} + \alpha w_{11} & \beta w_{12} \\ \alpha w_{21} & a_{22} + \beta w_{22} \end{pmatrix} =: \begin{pmatrix} a_l & c \\ b_l & d \end{pmatrix}, \\ W_{\Omega^2} = W_{\Omega^4} &= \begin{pmatrix} a_{11} + w_{11} & \beta w_{12} \\ w_{21} & a_{22} + \beta w_{22} \end{pmatrix} =: \begin{pmatrix} a_r & c \\ b_r & d \end{pmatrix}. \end{aligned} \quad (14)$$

This gives another example of a representative of $2d$ PWL maps with only one border defined in eq. (11). We are pointing this out because eq. (11) is a generic system considered more widely in the discrete dynamical systems literature [5, 6], and also was the basis for the analyses below and in Fig. 1.

A.1.2 Fixed points of the map (11) and their bifurcations

For $a_l, a_r, b_l, b_r, c, d, h_1, h_2 \in \mathbf{R}$, the map (11) has the following two fixed points

$$\mathcal{O}_{\mathcal{L}/\mathcal{R}} = (z_1^{\mathcal{L}/\mathcal{R}}, z_2^{\mathcal{L}/\mathcal{R}})^\top = \left(\frac{(1-d)h_1 + ch_2}{(1-d)(1-a_{l/r}) - b_{l/r}c}, \frac{b_{l/r}h_1 + (1-a_{l/r})h_2}{(1-d)(1-a_{l/r}) - b_{l/r}c} \right)^\top. \quad (15)$$

The fixed points $\mathcal{O}_{\mathcal{L}}$ and $\mathcal{O}_{\mathcal{R}}$ exist iff $z_1^{\mathcal{L}} < 0$ and $z_1^{\mathcal{R}} > 0$ respectively; otherwise they are virtual. Hence, the existence regions of admissible fixed points are

$$\begin{aligned} E_{\mathcal{O}_{\mathcal{L}}} &= \left\{ (h_1, h_2, a_l, b_l, c, d) \mid \frac{(1-d)h_1 + ch_2}{(1-d)(1-a_l) - b_l c} < 0 \right\}, \\ E_{\mathcal{O}_{\mathcal{R}}} &= \left\{ (h_1, h_2, a_r, b_r, c, d) \mid \frac{(1-d)h_1 + ch_2}{(1-d)(1-a_r) - b_r c} > 0 \right\}. \end{aligned} \quad (16)$$

Let $\mathcal{D}_{\mathcal{L}/\mathcal{R}}$ be the determinant and $\mathcal{T}_{\mathcal{L}/\mathcal{R}}$ the trace of $\mathbf{A}_{\mathcal{L}/\mathcal{R}}$, and

$$\mathcal{P}_{\mathcal{L}/\mathcal{R}}(\lambda) = \lambda^2 - (a_{l/r} + d)\lambda + a_{l/r}d - b_{l/r}c = \lambda^2 - \mathcal{T}_{\mathcal{L}/\mathcal{R}}\lambda + \mathcal{D}_{\mathcal{L}/\mathcal{R}}, \quad (17)$$

its characteristic polynomial. The corresponding eigenvalues are given by

$$\lambda_{1,2}(\mathcal{O}_{\mathcal{L}/\mathcal{R}}) = \frac{a_{l/r} + d}{2} \pm \frac{\sqrt{(a_{l/r} - d)^2 + 4b_{l/r}c}}{2} = \frac{\mathcal{T}_{\mathcal{L}/\mathcal{R}}}{2} \pm \frac{\sqrt{\mathcal{T}_{\mathcal{L}/\mathcal{R}}^2 - 4\mathcal{D}_{\mathcal{L}/\mathcal{R}}}}{2}, \quad (18)$$

which are always real for $b_{l/r}c \geq 0$, while for $b_{l/r}c < 0$ they are real provided that $|a_{l/r} - d| > 2\sqrt{-b_{l/r}c}$. For complex conjugate eigenvalues of $\mathbf{A}_{\mathcal{L}/\mathcal{R}}$ obviously $|\lambda|^2 = \mathcal{D}_{\mathcal{L}/\mathcal{R}}$. Thus computing the real eigenvalues, the stability condition for the fixed points is determined as

$$-(1 + \mathcal{D}_{\mathcal{L}/\mathcal{R}}) < \mathcal{T}_{\mathcal{L}/\mathcal{R}} < 1 + \mathcal{D}_{\mathcal{L}/\mathcal{R}}. \quad (19)$$

Accordingly, the stability region of the fixed points $\mathcal{O}_{\mathcal{L}}$ and $\mathcal{O}_{\mathcal{R}}$ can be obtained by $\mathcal{P}_{\mathcal{L}/\mathcal{R}}(\pm 1) = 1 \mp (a_{l/r} + d) + a_{l/r}d - b_{l/r}c > 0$ and $\mathcal{D}_{\mathcal{L}/\mathcal{R}} < 1$ as

$$\mathcal{S}_{\mathcal{L}/\mathcal{R}} = \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid a_{l/r}d - b_{l/r}c < 1, \right.$$

$$1 \pm (a_{l/r} + d) + a_{l/r} d - b_{l/r} c > 0 \}. \quad (20)$$

Note that when $\mathcal{D}_{\mathcal{L}/\mathcal{R}} < 0$, all the eigenvalues are real and so there cannot be any spiralling orbit.

Remark 1. Consider the PLRNN (2) with $M = 2$. For the parameter setting (3), i.e.

$$\mathbf{W}_{\Omega^1} = \mathbf{W}_{\Omega^3} = \begin{pmatrix} a_{11} & 0 \\ 0 & a_{22} \end{pmatrix} =: \mathbf{A}_{\mathcal{L}}, \quad \mathbf{W}_{\Omega^2} = \mathbf{W}_{\Omega^4} = \begin{pmatrix} a_{11} + w_{11} & 0 \\ w_{21} & a_{22} \end{pmatrix} =: \mathbf{A}_{\mathcal{R}}, \quad (21)$$

the two fixed points $\mathcal{O}_{\mathcal{L}/\mathcal{R}} = (z_1^{\mathcal{L}/\mathcal{R}}, z_2^{\mathcal{L}/\mathcal{R}})^\top$ are given by

$$\mathcal{O}_{\mathcal{L}} = \left(\frac{h_1}{1 - a_{11}}, \frac{h_2}{1 - a_{22}} \right)^\top, \quad \mathcal{O}_{\mathcal{R}} = \left(\frac{h_1}{1 - a_{11} - w_{11}}, \frac{w_{21} h_1 + (1 - a_{11} - w_{11}) h_2}{(1 - a_{22})(1 - a_{11} - w_{11})} \right)^\top. \quad (22)$$

Hence, the existence regions of admissible fixed points are

$$E_{\mathcal{O}_{\mathcal{L}}} = \left\{ (h_1, a_{11}, a_{22}) \mid \frac{h_1}{1 - a_{11}} < 0 \right\}, \quad E_{\mathcal{O}_{\mathcal{R}}} = \left\{ (h_1, a_{11}, a_{22}, w_{11}) \mid \frac{h_1}{1 - a_{11} - w_{11}} > 0 \right\},$$

and their stability regions can be obtained as

$$\mathcal{S}_{\mathcal{L}} = \left\{ (h_1, a_{11}, a_{22}) \in E_{\mathcal{O}_{\mathcal{L}}} \mid a_{11} a_{22} < 1, 1 \pm (a_{11} + a_{22}) + a_{11} a_{22} > 0 \right\}, \quad (23)$$

$$\mathcal{S}_{\mathcal{R}} = \left\{ (h_1, a_{11}, a_{22}, w_{11}) \in E_{\mathcal{O}_{\mathcal{R}}} \mid (a_{11} + w_{11}) a_{22} < 1, 1 \pm (a_{11} + w_{11} + a_{22}) + (a_{11} + w_{11}) a_{22} > 0 \right\}.$$

Remark 2. If $b_{l/r} c = 0$, then $\lambda_{1,2}(\mathcal{O}_{\mathcal{L}/\mathcal{R}})$ are real and the stability regions $\mathcal{S}_{\mathcal{L}/\mathcal{R}}$ become

$$\mathcal{S}_{\mathcal{L}/\mathcal{R}} = \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid b_{l/r} c = 0, -1 \leq a_{l/r} \leq 1, -1 \leq d \leq 1 \right\}. \quad (24)$$

The fixed points are regular saddles for all parameters that belong to

$$\left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid a_{l/r} + d > 1, a_{l/r} d - a_{l/r} - d + 1 < b_{l/r} c < a_{l/r} d \right\}, \quad (25)$$

and in this case $\lambda_1(\mathcal{O}_{\mathcal{L}/\mathcal{R}}) > 1$, $0 < \lambda_2(\mathcal{O}_{\mathcal{L}/\mathcal{R}}) < 1$. Furthermore, they are flip saddles (i.e., with one negative eigenvalue) if parameters are in

$$\begin{aligned} & \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid a_{l/r} + d > 1, a_{l/r} d < b_{l/r} c < a_{l/r} d + a_{l/r} + d + 1 \right\} \cup \\ & \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid d - a_{l/r} - d + 1 < b_{l/r} c < a_{l/r} d + a_{l/r} + d + 1, \right. \\ & \quad \left. 0 < a_{l/r} + d \leq 1, a_{l/r} \right\}, \quad (26) \end{aligned}$$

for which $\lambda_1(\mathcal{O}_{\mathcal{L}/\mathcal{R}}) > 1$, $-1 < \lambda_2(\mathcal{O}_{\mathcal{L}/\mathcal{R}}) < 0$, as well as in

$$\begin{aligned} & \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid a_{l/r} + d \leq -1, a_{l/r} d < b_{l/r} c < a_{l/r} d - a_{l/r} - d + 1 \right\} \cup \\ & \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid a_{l/r} d + a_{l/r} + d + 1 < b_{l/r} c < a_{l/r} d - a_{l/r} + d - 1, \right. \\ & \quad \left. -1 < a_{l/r} + d < 0 \right\} \quad (27) \end{aligned}$$

such that $0 < \lambda_1(\mathcal{O}_{\mathcal{L}/\mathcal{R}}) < 1$, $\lambda_2(\mathcal{O}_{\mathcal{L}/\mathcal{R}}) < -1$.

When $b_{l/r}c < 0$ and $|a_{l/r} - d| < 2\sqrt{-b_{l/r}c}$, the eigenvalues are complex conjugates and both $\mathcal{O}_{\mathcal{L}}$ and $\mathcal{O}_{\mathcal{R}}$ are spirally attracting (attracting focus) if $a_{l/r}d - b_{l/r}c < 1$. In this case, if $a_{l/r} + d > 0$ then they are clockwise spiral, while for $a_{l/r} + d < 0$ the spiralling motion will be counterclockwise. Moreover, for $a_{l/r}d - b_{l/r}c > 1$ they are repelling foci. Finally, for $a_{l/r}d - b_{l/r}c = 1$, the fixed points are locally centers and they undergo a CB at the following boundaries:

$$\begin{aligned}\mathcal{C}_{\mathcal{L}} &= \left\{ (a_l, b_l, c, d) \mid b_l c < 0, |a_l - d| < 2\sqrt{-b_l c}, a_l d - b_l c = 1 \right\}, \\ \mathcal{C}_{\mathcal{R}} &= \left\{ (a_r, b_r, c, d) \mid b_r c < 0, |a_r - d| < 2\sqrt{-b_r c}, a_r d - b_r c = 1 \right\}.\end{aligned}\quad (28)$$

At these boundaries, the fixed points lose their stability with a pair of complex conjugate eigenvalues crossing the unit circle. For the parameters belonging to $\mathcal{C}_{\mathcal{L}/\mathcal{R}}$, the Jacobian $J_{\mathcal{L}/\mathcal{R}}$ is a rotation matrix whose determinant is equal to 1. In this case, $J_{\mathcal{L}/\mathcal{R}}$ can be determined by a rotation number which is either rational ($\frac{p}{q}$) or irrational (ρ). Therefore, in some neighborhood of $\mathcal{O}_{\mathcal{L}/\mathcal{R}}$, there is a region filled with invariant ellipses such that they are periodic with period p (if the rotation number is a rational number $\frac{p}{q}$) or quasiperiodic (if the rotation number is an irrational number ρ); for more information see [60, 59]. For $(1-d)h_1 + ch_2 \neq 0$, at the boundary

$$\tau_{\mathcal{L}} = \left\{ (h_1, h_2, a_l, b_l, c, d) \mid 1 - a_l - d + a_l d - b_l c = 0 \right\}, \quad (29)$$

the fixed point $\mathcal{O}_{\mathcal{L}}$ undergoes a DTB, since, if the parameters tend to $\tau_{\mathcal{L}}$, then $\mathcal{O}_{\mathcal{L}} \rightarrow \pm\infty$ and $\lambda(\mathcal{O}_{\mathcal{L}}) \rightarrow 1$. Similarly, for $(1-d)h_1 + ch_2 \neq 0$, a DTB occurs for the fixed point $\mathcal{O}_{\mathcal{R}}$ at the boundary

$$\tau_{\mathcal{R}} = \left\{ (h_1, h_2, a_r, b_r, c, d) \mid 1 - a_r - d + a_r d - b_r c = 0 \right\}. \quad (30)$$

A DTB of a fixed point results in its disappearance, as in this case the fixed point becomes virtual which may lead to changes in the global dynamics [6]. Furthermore, the BCB curves are given by

$$\xi_{\mathcal{L}} = \left\{ (h_1, h_2, a_l, b_l, c, d) \mid (1-d)(1-a_l) - b_l c \neq 0, (1-d)h_1 + ch_2 = 0 \right\}, \quad (31)$$

and

$$\xi_{\mathcal{R}} = \left\{ (h_1, h_2, a_r, b_r, c, d) \mid (1-d)(1-a_r) - b_r c \neq 0, (1-d)h_1 + ch_2 = 0 \right\}. \quad (32)$$

In addition, the DFB curves for the fixed points $\mathcal{O}_{\mathcal{L}}$ and $\mathcal{O}_{\mathcal{R}}$ are

$$\begin{aligned}\mathcal{F}_{\mathcal{L}} &= \left\{ (h_1, h_2, a_l, b_l, c, d) \mid 1 + a_l + d + a_l d - b_l c = 0 \right\}, \\ \mathcal{F}_{\mathcal{R}} &= \left\{ (h_1, h_2, a_r, b_r, c, d) \mid 1 + a_r + d + a_r d - b_r c = 0 \right\}.\end{aligned}\quad (33)$$

Remark 3. The existence regions $E_{\mathcal{O}_{\mathcal{L}}}$ and $E_{\mathcal{O}_{\mathcal{R}}}$ are bounded by the BCB curves $\xi_{\mathcal{L}}$ and $\xi_{\mathcal{R}}$.

Remark 4. The stability regions $\mathcal{S}_{\mathcal{L}/\mathcal{R}}$ of fixed points (eq. (20)) are bounded by the DTB curves $\tau_{\mathcal{L}/\mathcal{R}}$ (eqn. (29) and (30)), the DFB curves $\mathcal{F}_{\mathcal{L}/\mathcal{R}}$ (eq. (33)), and the CB curves $a_{l/r}d - b_{l/r}c = 1$. For instance, for $d = 1$, $\mathcal{S}_{\mathcal{L}/\mathcal{R}}$ are illustrated in Fig. S1(a). In this case, the stability regions only exist for $b_{l/r}c < 0$ and $a_{l/r} - d < -2\sqrt{-b_{l/r}c}$. Moreover, as shown in Fig. S1(b), for $d = 0.01$, these stability regions can exist for both cases $b_{l/r}c < 0$, $a_{l/r} - d < -2\sqrt{-b_{l/r}c}$ (in blue), and $b_{l/r}c < 0$, $a_{l/r} - d > 2\sqrt{-b_{l/r}c}$ (in green), but not for $b_{l/r}c \geq 0$. Furthermore, if $c = 1$, there are stability regions $\mathcal{S}_{\mathcal{L}/\mathcal{R}}$ for the two cases $b_{l/r}c < 0$, $a_{l/r} - d < -2\sqrt{-b_{l/r}c}$ (in blue), and $b_{l/r}c > 0$ (in purple); see Fig. S2(a). Finally, when $c = 0$, as explained in Remark 2, the stability regions have the form (24), i.e.

$$\mathcal{S}_{\mathcal{L}/\mathcal{R}} = \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{L}/\mathcal{R}}} \mid c = 0, b_{l/r} \in \mathbb{R}, -1 \leq a_{l/r}, d \leq 1 \right\}. \quad (34)$$

which are displayed in Fig. S2(b).

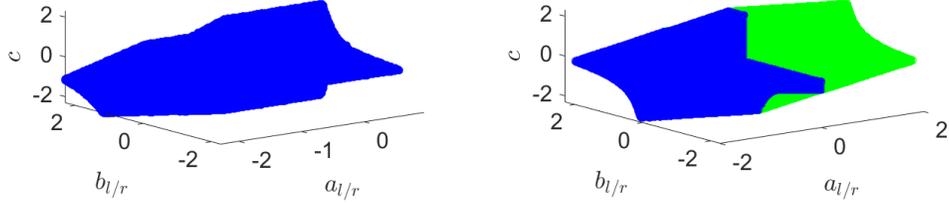


Figure S1: Stability regions $\mathcal{S}_{\mathcal{L}/\mathcal{R}}$. Left: for $d = 1$; right: for $d = 0.01$. The case $b_{l/r} c < 0$, $a_{l/r} - d < -2\sqrt{-b_{l/r} c}$ is plotted in blue, and the case $b_{l/r} c < 0$, $a_{l/r} - d > 2\sqrt{-b_{l/r} c}$ is drawn in green.

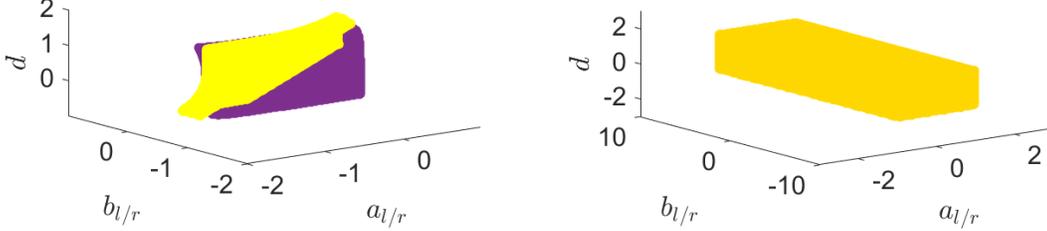


Figure S2: Stability regions $\mathcal{S}_{\mathcal{L}/\mathcal{R}}$. Left: for $c = 1$; the case $b_{l/r} c < 0$, $a_{l/r} - d < -2\sqrt{-b_{l/r} c}$ is plotted in yellow, and the case $b_{l/r} c > 0$ is drawn in purple. Right: for $c = 0$.

Since the system (11) is a linear map in each sub-region \mathcal{L} and \mathcal{R} , there cannot be any n -cycle, $n \geq 2$, with all periodic points on only one linear side. So, all period- n orbits have both letters \mathcal{L} and \mathcal{R} in their symbolic sequence.

A.1.3 2-cycles of the map (11) and their bifurcations

The 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}$ of the map (11) is determined by solving the equation $T_{\mathcal{L}} \circ T_{\mathcal{R}}(z_1, z_2) = (z_1, z_2)^{\top}$ where

$$T_{\mathcal{L}} \circ T_{\mathcal{R}}(z_1, z_2) = \begin{pmatrix} a_l a_r + b_r c & a_l c + c d \\ a_r b_l + b_r d & d^2 + b_l c \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} + \begin{pmatrix} c h_2 + h_1 (a_l + 1) \\ b_l h_1 + h_2 (d + 1) \end{pmatrix}. \quad (35)$$

In this case if $(I - J_{\mathcal{L}} J_{\mathcal{R}})$ is invertible, then the solution $(z_1, z_2)^{\top} = (z_1^{(1)}, z_2^{(1)})^{\top}$ is given by

$$(z_1^{(1)}, z_2^{(1)})^{\top} = \begin{pmatrix} \frac{((1-d)h_1 + c h_2)(a_l + d + a_l d - b_l c + 1)}{(a_r d - b_r c)(a_l d - b_l c) - c(b_l + b_r) - d^2 - a_l a_r + 1}, \\ \frac{h_2(1 + d - a_l a_r - b_r c - a_l a_r d + a_r b_l c) + h_1(b_l + a_r b_l + b_r d + a_l b_r d - b_l b_r c)}{(a_r d - b_r c)(a_l d - b_l c) - c(b_l + b_r) - d^2 - a_l a_r + 1} \end{pmatrix}. \quad (36)$$

Also $T_{\mathcal{R}}(z_1^{(1)}, z_2^{(1)}) = (z_1^{(2)}, z_2^{(2)})^{\top}$ yields

$$(z_1^{(2)}, z_2^{(2)})^{\top} = \begin{pmatrix} \frac{((1-d)h_1 + c h_2)(a_r + d + a_r d - b_r c + 1)}{(a_r d - b_r c)(a_l d - b_l c) - c(b_l + b_r) - d^2 - a_l a_r + 1}, \\ \frac{h_2(1 + d - a_l a_r - b_l c - a_l a_r d + a_l b_r c) + h_1(b_r + a_l b_r + b_l d + a_r b_l d - b_l b_r c)}{(a_r d - b_r c)(a_l d - b_l c) - c(b_l + b_r) - d^2 - a_l a_r + 1} \end{pmatrix}. \quad (37)$$

Hence, the existence region of the 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}$ is

$$E_{\mathcal{O}_{\mathcal{R}\mathcal{L}}} = \left\{ (h_1, h_2, a_l, b_l, c, d) \mid \frac{((1-d)h_1 + c h_2)(a_l + d + a_l d - b_l c + 1)}{(a_r d - b_r c)(a_l d - b_l c) - c(b_l + b_r) - d^2 - a_l a_r + 1} > 0, \right.$$

$$\left. \frac{((1-d)h_1 + c h_2)(a_r + d + a_r d - b_r c + 1)}{(a_r d - b_r c)(a_l d - b_l c) - c(b_l + b_r) - d^2 - a_l a_r + 1} < 0 \right\}. \quad (38)$$

The characteristic polynomial of $J_{\mathcal{R}\mathcal{L}} = J_{\mathcal{L}} J_{\mathcal{R}} = \begin{pmatrix} a_l a_r + b_r c & a_l c + c d \\ a_r b_l + b_r d & d^2 + b_l c \end{pmatrix}$ is given by

$$\mathcal{P}_{\mathcal{O}_{\mathcal{R}\mathcal{L}}}(\lambda) = \lambda^2 - (d^2 + a_l a_r + b_l c + b_r c)\lambda + (a_r d - b_r c)(a_l d - b_l c), \quad (39)$$

and

$$\mathcal{D}_{\mathcal{R}\mathcal{L}} = (a_r d - b_r c)(a_l d - b_l c),$$

$$\mathcal{P}_{\mathcal{O}_{\mathcal{R}\mathcal{L}}}(1) = (a_r d - b_r c)(a_l d - b_l c) - c(b_l + b_r) - d^2 - a_l a_r + 1,$$

$$\mathcal{P}_{\mathcal{O}_{\mathcal{R}\mathcal{L}}}(-1) = (a_r d - b_r c)(a_l d - b_l c) + c(b_l + b_r) + d^2 + a_l a_r + 1,$$

$$\begin{aligned} \lambda_{1,2}(\mathcal{O}_{\mathcal{R}\mathcal{L}}) &= \frac{a_l a_r + c(b_l + b_r) + d^2}{2} \\ &\pm \frac{\sqrt{(a_l a_r + b_l c)^2 + (b_r c + d^2)^2 + 2(a_l a_r - b_l c)(b_r c - d^2) + 4 c d(a_l b_r + a_r b_l)}}{2}. \end{aligned} \quad (40)$$

Thus, the stability region of $\mathcal{O}_{\mathcal{R}\mathcal{L}}$ is

$$\begin{aligned} \mathcal{S}_{\mathcal{R}\mathcal{L}} &= \left\{ (h_1, h_2, a_l/r, b_l/r, c, d) \in E_{\mathcal{O}_{\mathcal{R}\mathcal{L}}} \mid -1 < (a_r d - b_r c)(a_l d - b_l c) < 1, \right. \\ &\quad \left. - (a_r d - b_r c)(a_l d - b_l c) - 1 < c(b_l + b_r) + d^2 + a_l a_r < (a_r d - b_r c)(a_l d - b_l c) + 1 \right\}. \end{aligned} \quad (41)$$

In addition, for $((1-d)h_1 + c h_2)(a_l + d + a_l d - b_l c + 1) \neq 0$, the set

$$\begin{aligned} \tau_{\mathcal{R}\mathcal{L}} &= \left\{ (h_1, h_2, a_l, a_r, b_l, b_r, c, d) \mid a_l a_r + b_l c + b_r c + d^2 - a_l a_r d^2 - b_l b_r c^2 + a_l b_r c d \right. \\ &\quad \left. + a_r b_l c d - 1 = 0 \right\}, \end{aligned} \quad (42)$$

is the DTB curve for the 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}$. As in this case, for the parameter values belonging to $\tau_{\mathcal{R}\mathcal{L}}$, the points of the 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}$ tend to $\pm\infty$, and the corresponding eigenvalue tends to one. Moreover, for $(1-d)h_1 + c h_2 \neq 0$, the BCB curves of $\mathcal{O}_{\mathcal{R}\mathcal{L}}$ can be computed as

$$\begin{aligned} \xi_{\mathcal{R}\mathcal{L}}^1 &= \left\{ (h_1, h_2, a_l, b_l, c, d) \mid a_l + d + a_l d - b_l c + 1 = 0 \right\}, \\ \xi_{\mathcal{R}\mathcal{L}}^2 &= \left\{ (h_1, h_2, a_r, b_r, c, d) \mid a_r + d + a_r d - b_r c + 1 = 0 \right\}. \end{aligned} \quad (43)$$

Note that here the condition $(1-d)h_1 + c h_2 \neq 0$ guarantees a regular BCB in the sense that only one periodic point of $\mathcal{O}_{\mathcal{R}\mathcal{L}}$ collides with the switching boundary; for more details see [6]. Besides,

$$\begin{aligned} \mathcal{F}_{\mathcal{R}\mathcal{L}} &= \left\{ (h_1, h_2, a_l, a_r, b_l, b_r, c, d) \mid a_l a_r + b_l c + b_r c + d^2 + a_l a_r d^2 + b_l b_r c^2 \right. \\ &\quad \left. - a_l b_r c d - a_r b_l c d + 1 = 0 \right\}, \end{aligned} \quad (44)$$

is the DFB curve of the 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}$.

Remark 5. One can see that for $(1-d)h_1 + c h_2 \neq 0$ the DFB curves of the fixed points $\mathcal{O}_{\mathcal{L}}$ and $\mathcal{O}_{\mathcal{R}}$ ($\mathcal{F}_{\mathcal{L}}$ and $\mathcal{F}_{\mathcal{R}}$) and the BCB boundaries of the 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}$ ($\xi_{\mathcal{R}\mathcal{L}}^1$ and $\xi_{\mathcal{R}\mathcal{L}}^2$) are the same. In this case, the DFB of the fixed points can lead to the (attracting) 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}$.

A.1.4 3-cycles of the map (11) and their bifurcations

Here, we investigate the existence, stability and bifurcation structure of maximal or basic 3-cycles. Note that for the continuous map (11), basic n -cycles $\mathcal{O}_{\mathcal{R}\mathcal{L}^{n-1}}$ ($n \geq 3$) exist in pairs with their complementary cycles ($\mathcal{O}_{\mathcal{R}\mathcal{L}^{n-2}\mathcal{R}}$), and they appear via BCBs such that one of them may be attracting and the other repelling [6, 20, 38]. In this case, a BCB of basic cycles demonstrates a non-smooth fold bifurcation which includes a stable basic orbit and an unstable nonbasic orbit [6, 5, 19]. Furthermore, the complementary orbits can have nonempty stability regions such that, similar to the basic orbits, they are bounded by curves of BCBs, DTBs and DFBs [6, 5].

Basic 3-cycles $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ and their complementary cycles $\mathcal{O}_{\mathcal{R}^2\mathcal{L}}$. The basic 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ can be obtained from the equation $T_{\mathcal{L}} \circ T_{\mathcal{L}} \circ T_{\mathcal{R}}(z_1, z_2) = (z_1, z_2)^T$ where

$$\begin{aligned} T_{\mathcal{L}} \circ T_{\mathcal{L}} \circ T_{\mathcal{R}}(z_1, z_2) = & \begin{pmatrix} a_r(a_l^2 + b_l c) + b_r(a_l c + c d) & c(a_l^2 + b_l c) + d(a_l c + c d) \\ b_r(d^2 + b_l c) + a_r(a_l b_l + b_l d) & d(d^2 + b_l c) + c(a_l b_l + b_l d) \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \end{pmatrix} \\ & + \begin{pmatrix} h_1(b_l c + a_l(a_l + 1) + 1) + h_2(a_l c + c(d + 1)) \\ h_1(b_l d + b_l(a_l + 1)) + h_2(b_l c + d(d + 1) + 1) \end{pmatrix}. \end{aligned} \quad (45)$$

If $(I - J_{\mathcal{L}}^2 J_{\mathcal{R}})$ is invertible, then the solution $(z_1, z_2)^T = (z_1^{(1)}, z_2^{(1)})^T$ is

$$(z_1^{(1)}, z_2^{(1)})^T = \left(\frac{((1-d)h_1 + ch_2)G_1}{G}, \frac{G_2}{G} \right)^T, \quad (46)$$

where

$$\begin{aligned} G_1 &= a_l^2 d^2 + a_l^2 d + a_l^2 - 2a_l b_l c d - a_l b_l c + a_l d^2 + a_l d + a_l + b_l^2 c^2 - b_l c d + b_l c + d^2 + d + 1, \\ G &= -a_l^2 a_r - d^3 - c(a_l b_l + a_l b_r + a_r b_l + d(2b_l + b_r)) + (a_r d - b_r c)(a_l d - b_l c)^2 + 1, \\ G_2 &= h_2 + b_l h_1 + d h_2 + d^2 h_2 + a_l b_l h_1 + b_l c h_2 + b_l d h_1 - a_l^2 a_r h_2 + b_r d^2 h_1 - a_r b_l^2 c h_1 \\ &\quad - a_l^2 a_r d h_2 + a_l b_r d^2 h_1 + b_l b_r c^2 h_2 - a_r b_l^2 c^2 h_2 - a_l^2 a_r d^2 h_2 + a_l^2 b_r d^2 h_1 + b_l^2 b_r c^2 h_1 \\ &\quad + a_l a_r b_l h_1 - a_l b_r c h_2 - a_r b_l c h_2 + a_r b_l d h_1 + b_l b_r c h_1 - b_r c d h_2 + a_l a_r b_l c h_2 \\ &\quad + a_l a_r b_l d h_1 - a_l b_r c d h_2 - b_l b_r c d h_1 + 2a_l a_r b_l c d h_2 - 2a_l b_l b_r c d h_1. \end{aligned} \quad (47)$$

Further

$$\begin{aligned} T_{\mathcal{R}}(z_1^{(1)}, z_2^{(1)}) &= (z_1^{(2)}, z_2^{(2)})^T = \left(\frac{((1-d)h_1 + ch_2)K_1}{G}, \frac{K_2}{G} \right)^T, \\ T_{\mathcal{L}}(z_1^{(2)}, z_2^{(2)}) &= (z_1^{(3)}, z_2^{(3)})^T = \left(\frac{((1-d)h_1 + ch_2)H_1}{G}, \frac{H_2}{G} \right)^T, \end{aligned} \quad (48)$$

where

$$\begin{aligned} K_1 &= a_r + d + a_l a_r + b_l c + a_r d + a_r d^2 + d^2 + a_l a_r d - a_l b_r c - b_r c d + a_l a_r d^2 + b_l b_r c^2 \\ &\quad - a_l b_r c d - a_r b_l c d + 1, \\ K_2 &= h_2 + b_r h_1 + d h_2 + d^2 h_2 + a_l b_r h_1 + b_r c h_2 + b_l d h_1 - a_l^2 a_r h_2 + a_l^2 b_r h_1 + b_l d^2 h_1 \\ &\quad + a_l^2 b_r c h_2 + a_r b_l d^2 h_1 + b_l b_r c^2 h_2 - a_l^2 a_r d^2 h_2 + b_l^2 b_r c^2 h_1 - a_l b_l c h_2 - a_r b_l c h_2 + a_l b_l d h_1 \\ &\quad + b_l b_r c h_1 - b_l c d h_2 + a_l a_r b_l d h_1 - a_l b_l b_r c h_1 - a_r b_l c d h_2 - b_l b_r c d h_1 + a_l a_r b_l d^2 h_1 \\ &\quad - a_l b_l b_r c^2 h_2 - a_r b_l^2 c d h_1 + a_l^2 b_r c d h_2 + a_l a_r b_l c d h_2 - a_l b_l b_r c d h_1 - a_l^2 a_r d h_2, \\ H_1 &= a_l + d + a_l a_r + a_l d + b_r c + a_l d^2 + d^2 + a_l a_r d - a_r b_l c - b_l c d + a_l a_r d^2 + b_l b_r c^2 \\ &\quad - a_l b_r c d - a_r b_l c d + 1, \\ H_2 &= h_2 + b_l h_1 + d h_2 + d^2 h_2 + b_l^2 c^2 h_2 + a_r b_l h_1 + b_l c h_2 + b_r d h_1 - a_l^2 a_r h_2 + b_l^2 c h_1 \end{aligned}$$

$$\begin{aligned}
& + b_l d^2 h_1 - a_l^2 a_r d h_2 - b_l c d h_2 + a_l b_l d^2 h_1 + a_l^2 b_r d h_1 - b_l^2 c d h_1 - a_l^2 a_r d^2 h_2 + b_l^2 b_r c^2 h_1 \\
& + a_l a_r b_l h_1 - a_l b_l c h_2 - a_l b_r c h_2 + a_l b_r d h_1 + a_l a_r b_l c h_2 - a_l b_l b_r c h_1 - a_l b_l c d h_2 \\
& + a_l a_r b_l d^2 h_1 - a_l b_l b_r c^2 h_2 - a_r b_l^2 c d h_1 + a_l^2 b_r c d h_2 + a_l a_r b_l c d h_2 - a_l b_l b_r c d h_1. \tag{49}
\end{aligned}$$

Therefore, the existence region of the 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ is given by

$$\begin{aligned}
E_{\mathcal{O}_{\mathcal{R}\mathcal{L}^2}} = \left\{ (h_1, h_2, a_l, b_l, c, d) \mid \frac{((1-d)h_1 + c h_2)G_1}{G} > 0, \frac{((1-d)h_1 + c h_2)K_1}{G} < 0, \right. \\
\left. \frac{((1-d)h_1 + c h_2)H_1}{G} < 0 \right\}, \tag{50}
\end{aligned}$$

where G, G_1, K_1 and H_1 are defined in (47) and (49). On the other hand, the characteristic polynomial of

$$J_{\mathcal{R}\mathcal{L}^2} = J_{\mathcal{L}}^2 J_{\mathcal{R}} = \begin{pmatrix} a_r(a_l^2 + b_l c) + b_r(a_l c + c d) & c(a_l^2 + b_l c) + d(a_l c + c d) \\ b_r(d^2 + b_l c) + a_r(a_l b_l + b_l d) & d(d^2 + b_l c) + c(a_l b_l + b_l d) \end{pmatrix},$$

is

$$\mathcal{P}_{\mathcal{O}_{\mathcal{R}\mathcal{L}^2}}(\lambda) = \lambda^2 - \left(a_l^2 a_r + d^3 + c(a_l b_l + a_l b_r + a_r b_l + d(2b_l + b_r)) \right) \lambda + (a_r d - b_r c)(a_l d - b_l c)^2. \tag{51}$$

According to

$$\begin{aligned}
\mathcal{D}_{\mathcal{R}\mathcal{L}^2} &= (a_l d - b_l c)^2 (a_r d - b_r c), \\
\mathcal{P}_{\mathcal{O}_{\mathcal{R}\mathcal{L}^2}}(1) &= -a_l^2 a_r - d^3 - c(a_l b_l + a_l b_r + a_r b_l + d(2b_l + b_r)) \\
\mathcal{P}_{\mathcal{O}_{\mathcal{R}\mathcal{L}^2}}(-1) &= a_l^2 a_r + d^3 + c(a_l b_l + a_l b_r + a_r b_l + d(2b_l + b_r)) + (a_r d - b_r c)(a_l d - b_l c)^2 + 1, \tag{52}
\end{aligned}$$

the stability region of the 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ is given by

$$\begin{aligned}
\mathcal{S}_{\mathcal{R}\mathcal{L}^2} = \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \in E_{\mathcal{O}_{\mathcal{R}\mathcal{L}^2}} \mid -1 < (a_l d - b_l c)^2 (a_r d - b_r c) < 1, \right. \\
- (a_l d - b_l c)^2 (a_r d - b_r c) - 1 < a_l^2 a_r + d^3 + c(a_l b_l + a_l b_r + a_r b_l + d(2b_l + b_r)) \\
\left. < (a_l d - b_l c)^2 (a_r d - b_r c) + 1 \right\}. \tag{53}
\end{aligned}$$

Furthermore, for $(1-d)h_1 + c h_2 \neq 0$ and $G_1, G_2, K_1, K_2, H_1, H_2 \neq 0$, the DTB curve for the 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ is

$$\begin{aligned}
\tau_{\mathcal{R}\mathcal{L}^2} = \left\{ (h_1, h_2, a_l, a_r, b_l, b_r, c, d) \mid b_r a_l^2 c d^2 - a_r a_l^2 d^3 + a_r a_l^2 - 2 b_r a_l b_l c^2 d + 2 a_r a_l b_l c d^2 \right. \\
\left. + a_l b_l c + b_r a_l c + b_r b_l^2 c^3 - a_r b_l^2 c^2 d + 2 b_l c d + a_r b_l c + b_r c d + d^3 - 1 = 0 \right\}. \tag{54}
\end{aligned}$$

For $(1-d)h_1 + c h_2 \neq 0$

$$\begin{aligned}
\xi_{\mathcal{R}\mathcal{L}^2}^1 = \left\{ (h_1, h_2, a_r, b_r, c, d) \mid K_1 = a_r + d + a_l a_r + b_l c + a_r d + a_r d^2 + d^2 + a_l a_r d \right. \\
\left. - a_l b_r c - b_r c d + a_l a_r d^2 + b_l b_r c^2 - a_l b_r c d - a_r b_l c d + 1 = 0 \right\},
\end{aligned}$$

$$\xi_{\mathcal{R}\mathcal{L}^2}^2 = \left\{ (h_1, h_2, a_r, b_r, c, d) \mid H_1 = a_l + d + a_l a_r + a_l d + b_r c + a_l d^2 + d^2 + a_l a_r d \right.$$

$$- a_r b_l c - b_l c d + a_l a_r d^2 + b_l b_r c^2 - a_l b_r c d - a_r b_l c d + 1 = 0 \}, \quad (55)$$

are (regular) BCB curves of $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$. Furthermore, the set

$$\begin{aligned} \mathcal{F}_{\mathcal{R}\mathcal{L}^2} = \left\{ (h_1, h_2, a_l, a_r, b_l, b_r, c, d) \mid - b_r a_l^2 c d^2 + a_r a_l^2 d^3 + a_r a_l^2 + 2b_r a_l b_l c^2 d - 2a_r a_l b_l c d^2 \right. \\ \left. + a_l b_l c + b_r a_l c - b_r b_l^2 c^3 + a_r b_l^2 c^2 d + 2b_l c d + a_r b_l c + b_r c d + d^3 + 1 = 0 \right\}, \end{aligned} \quad (56)$$

is the DFB curve of $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$. As noted, the basic 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ exists in a pair with its complementary cycle $\mathcal{O}_{\mathcal{R}^2\mathcal{L}}$. Moreover, the existence region of $\mathcal{O}_{\mathcal{R}^2\mathcal{L}}$ can easily be found by interchanging the letters \mathcal{L} and \mathcal{R} in all notations of the equations (45)-(49) and considering

$$z_1^{(1)} < 0, \quad z_1^{(2)} > 0, \quad z_1^{(3)} > 0. \quad (57)$$

Further, the stability region of the 3-cycle $\mathcal{O}_{\mathcal{R}^2\mathcal{L}}$ for the parameter values satisfying (57) is given by

$$\begin{aligned} \mathcal{S}_{\mathcal{R}^2\mathcal{L}} = \left\{ (h_1, h_2, a_{l/r}, b_{l/r}, c, d) \mid -1 < (a_r d - b_r c)^2 (a_l d - b_l c) < 1, \right. \\ \left. - (a_r d - b_r c)^2 (a_l d - b_l c) - 1 < a_r^2 a_l + d^3 + c(a_r b_r + a_r b_l + a_l b_r + d(2b_r + b_l)) \right. \\ \left. < (a_r d - b_r c)^2 (a_l d - b_l c) + 1 \right\}. \end{aligned} \quad (58)$$

Notice that whenever the stable 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ exists, its complementary orbit $\mathcal{O}_{\mathcal{R}^2\mathcal{L}}$ also exists, but it is unstable. Furthermore, for $(1-d)h_1 + c h_2 \neq 0$ both the 3-cycles $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}$ and $\mathcal{O}_{\mathcal{R}^2\mathcal{L}}$ appear at the same BCB curves (55). On the other hand, the DTB and DFB curves of the 3-cycle $\mathcal{O}_{\mathcal{R}^2\mathcal{L}}$ are given by

$$\begin{aligned} \tau_{\mathcal{R}^2\mathcal{L}} = \left\{ (h_1, h_2, a_l, a_r, b_l, b_r, c, d) \mid b_l a_r^2 c d^2 - a_l a_r^2 d^3 + a_l a_r^2 - 2b_l a_r b_r c^2 d + 2a_l a_r b_r c d^2 \right. \\ \left. + a_r b_r c + b_l a_r c + b_l b_r^2 c^3 - a_l b_r^2 c^2 d + 2b_r c d + a_l b_r c + b_l c d + d^3 - 1 = 0 \right\}, \end{aligned} \quad (59)$$

and

$$\begin{aligned} \mathcal{F}_{\mathcal{R}^2\mathcal{L}} = \left\{ (h_1, h_2, a_l, a_r, b_l, b_r, c, d) \mid - b_l a_r^2 c d^2 + a_l a_r^2 d^3 + a_l a_r^2 + 2b_l a_r b_r c^2 d - 2a_l a_r b_r c d^2 \right. \\ \left. + a_r b_r c + b_l a_r c - b_l b_r^2 c^3 + a_l b_r^2 c^2 d + 2b_r c d + a_l b_r c + b_l c d + d^3 + 1 = 0 \right\}, \end{aligned} \quad (60)$$

respectively.

A.1.5 Multiple attractor bifurcations (MABs) of the map (11)

To detect multiple attractor bifurcations for the map (11), a straightforward way is to determine the overlapping stability regions of different periodic orbits. For instance, as shown in Fig. 1A, for $c = 0.8$, $d = 0.2$, $b_l = -0.4$, $b_r = 0.5$, two stability regions $\mathcal{S}_{\mathcal{R}\mathcal{L}}$ and $\mathcal{S}_{\mathcal{R}\mathcal{L}^2}$ overlap in the (a_l, a_r) -parameter plane (or in the $(a_{11}, a_{11} + w_{11})$ -parameter space for the $2d$ PLRNN (7)). Their overlapping region, displayed in yellow, reveals the structure of the (a_l, a_r) -parameter plane. This helps us to find various MABs. Assuming $h_2 = 0$ and varying h_1 from a negative value to a positive one, an MAB of the form

$$\mathcal{O}_{\mathcal{L}}^s \xleftrightarrow{h_1} \mathcal{O}_{\mathcal{R}\mathcal{L}}^s + \mathcal{O}_{\mathcal{R}\mathcal{L}^2}^s, \quad (61)$$

occurs in the overlapping region. An example of this kind of bifurcation is illustrated in Fig. S3A.

Moreover, Fig. S4 indicates, for $c = 0.9$, $d = 0.3$, $b_l = -0.6$, $b_r = -1.54$, there are nonempty overlapping regions $\mathcal{S}_{\mathcal{R}\mathcal{L}} \cap \mathcal{S}_{\mathcal{R}\mathcal{L}^2}$ and $\mathcal{S}_{\mathcal{R}\mathcal{L}^2} \cap \mathcal{S}_{\mathcal{R}}$. This leads to the occurrence of two different MABs given by (61) and

$$\mathcal{O}_{\mathcal{L}}^s \xleftrightarrow{h_1} \mathcal{O}_{\mathcal{R}}^s + \mathcal{O}_{\mathcal{R}\mathcal{L}^2}, \quad (62)$$

for $h_2 = 0$ and h_1 changing from negative to positive values. Both of these bifurcations are shown in Fig. S3A and Fig. S3B, associated with the points P_1 and P_2 in Fig. S4. Note that in Fig. S4, all the points P_2 , P_3 and P_4 belong to the overlapping region $\mathcal{S}_{\mathcal{R}\mathcal{L}^2} \cap \mathcal{S}_{\mathcal{R}}$ (in sky blue). These points are related to the parameter values $c = 0.9$, $d = 0.3$, $b_l = -0.6$, $b_r = -1.54$, $a_r = -1.8$, $h_2 = 0$, and they only differ in the parameter a_l . In this case, one can see that fixing all parameters and changing only the parameter a_l , from P_2 to P_4 , the basins of attraction change. The corresponding basins of attraction for these three points are demonstrated in Fig. S3B (right) and Fig. S5 for $h_1 = 0.5$ (after the bifurcation).

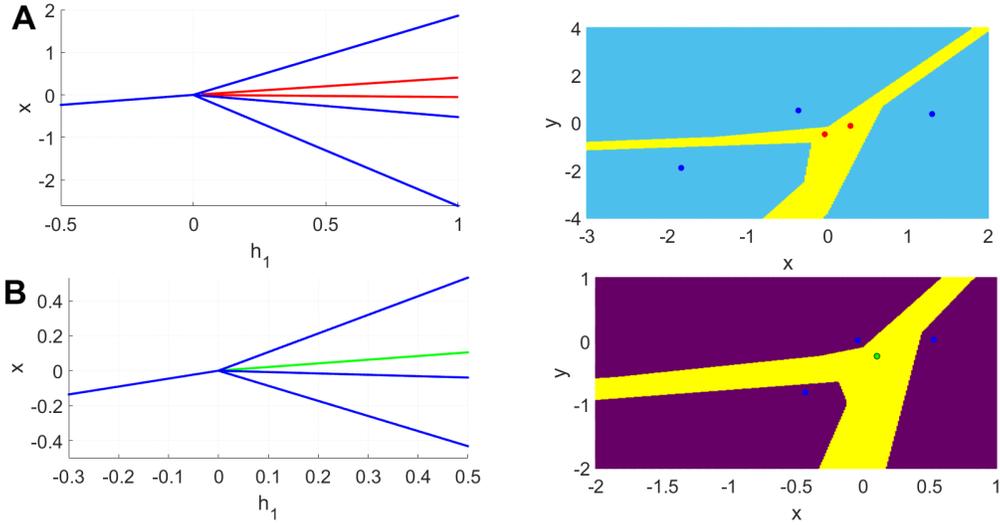


Figure S3: MAB at $c = 0.9$, $d = 0.3$, $b_l = -0.6$, $b_r = -1.54$, $h_2 = 0$. A) Left: Bifurcation diagram for $a_l = -0.44$ and $a_r = -1.8$ corresponding to the point P_1 in Fig. S4. Right: Multistability of the fixed point $\mathcal{O}_{\mathcal{R}}^s$ and the 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}^s$ after the bifurcation and their basins of attraction at $h_1 = 0.5$. B) Left: Bifurcation diagram for $a_l = -0.35$ and $a_r = -2.2$ corresponding to the point P_2 in Fig. S4. Right: Multistability of the 2-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}}^s$ and the 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}^s$ after the bifurcation and their basins of attraction at $h_1 = 0.7$.

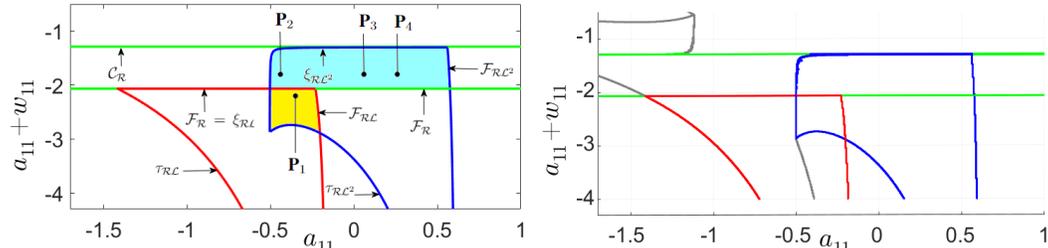


Figure S4: Analytically calculated stability regions for a different parameter setting than used in Fig. 1. Left: Analytically calculated stability regions $\mathcal{S}_{\mathcal{R}}$, $\mathcal{S}_{\mathcal{R}\mathcal{L}}$ and $\mathcal{S}_{\mathcal{R}\mathcal{L}^2}$, shown in green, red and blue, respectively, in the $(a_{11}, a_{11} + w_{11})$ -parameter plane for $a_{22} = 0.3$, $w_{21} = -1.54$. The overlapping regions $\mathcal{S}_{\mathcal{R}\mathcal{L}} \cap \mathcal{S}_{\mathcal{R}\mathcal{L}^2}$ and $\mathcal{S}_{\mathcal{R}\mathcal{L}^2} \cap \mathcal{S}_{\mathcal{R}}$, representing multi-stable regimes, are given in yellow and sky blue. Right: Bifurcation curves for the same parameter settings as determined by SCYFI. Note that SCYFI identifies additional structure (regions demarcated by gray curves) not included in our analytical derivations.

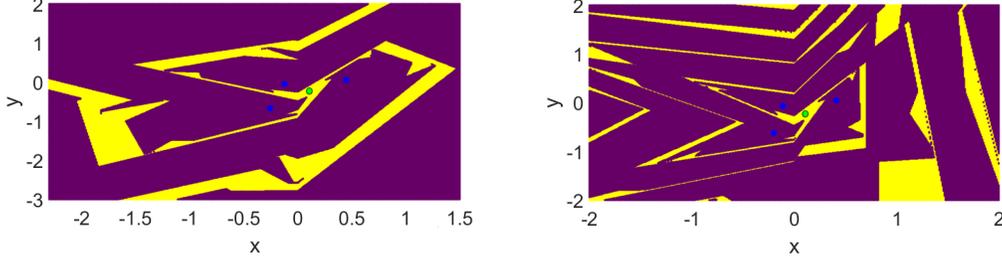


Figure S5: Multistability of the fixed point $\mathcal{O}_{\mathcal{R}}^s$ and the 3-cycle $\mathcal{O}_{\mathcal{R}\mathcal{L}^2}^s$ at $c = 0.9$, $d = 0.3$, $b_l = -0.6$, $b_r = -1.54$, $a_r = -1.8$, $h_2 = 0$ after the bifurcation and their basins of attraction at $h_1 = 0.5$. Left: $a_l = 0.06$ (point P_3 in Fig. S4); right: $a_l = 0.26$ (point P_4 in Fig. S4).

A.2 Proofs of theorems

A.2.1 Proof of theorem 1

Proof. Let $\mathcal{L}(\theta)$ be some loss function employed for PLRNN training that decomposes in time as $\mathcal{L} = \sum_{t=1}^T \mathcal{L}_t$. Then

$$\begin{aligned} \frac{\partial \mathcal{L}}{\partial \theta} &= \sum_{t=1}^T \frac{\partial \mathcal{L}_t}{\partial \theta}, \\ \frac{\partial \mathcal{L}_t}{\partial \theta} &= \frac{\partial \mathcal{L}_t}{\partial \mathbf{z}_t} \frac{\partial \mathbf{z}_t}{\partial \theta}. \end{aligned} \quad (63)$$

Denoting the Jacobian of system (2) at time t by

$$\mathbf{J}_t := \frac{\partial F_{\theta}(\mathbf{z}_{t-1})}{\partial \mathbf{z}_{t-1}} = \frac{\partial \mathbf{z}_t}{\partial \mathbf{z}_{t-1}}, \quad (64)$$

we have

$$\frac{\partial \mathbf{z}_t}{\partial \theta} = \frac{\partial \mathbf{z}_t}{\partial \mathbf{z}_{t-1}} \frac{\partial \mathbf{z}_{t-1}}{\partial \theta} + \frac{\partial^+ \mathbf{z}_t}{\partial \theta} = \mathbf{J}_t \frac{\partial \mathbf{z}_{t-1}}{\partial \theta} + \frac{\partial^+ \mathbf{z}_t}{\partial \theta}, \quad (65)$$

where ∂^+ denotes the immediate partial derivative (see [48] for more details). Assume that Γ_k is a k -cycle ($k \geq 1$) of (2). Thus, Γ_k is a set of temporally successive periodic points

$$P_k := \{\mathbf{z}_{t^*k}, \mathbf{z}_{t^*k-1}, \dots, \mathbf{z}_{t^*k-(k-1)}\} = \{\mathbf{z}_{t^*k}, F(\mathbf{z}_{t^*k}), \dots, F^{k-1}(\mathbf{z}_{t^*k})\}, \quad (66)$$

such that all of them are fixed points of

$$\mathbf{z}_{t+k} = F_{\theta}^k(\mathbf{z}_t) = F_{\theta}(F_{\theta}(F_{\theta}(\dots F_{\theta}(\mathbf{z}_t)\dots))), \quad (67)$$

and k is the smallest such positive integer (for $k = 1$, Γ_1 is a fixed point of F_{θ}). Similar to (65), the tangent vector $\frac{\partial \mathbf{z}_{t+k}}{\partial \theta}$ can be computed as

$$\frac{\partial \mathbf{z}_{t+k}}{\partial \theta} = \frac{\partial \mathbf{z}_{t+k}}{\partial \mathbf{z}_t} \frac{\partial \mathbf{z}_t}{\partial \theta} + \frac{\partial^+ \mathbf{z}_{t+k}}{\partial \theta} = \prod_{r=0}^{k-1} \mathbf{J}_{t+k-r} \frac{\partial \mathbf{z}_t}{\partial \theta} + \frac{\partial^+ \mathbf{z}_{t+k}}{\partial \theta}. \quad (68)$$

Thus, for $\mathbf{z}_{t^*k} = F_{\theta}^k(\mathbf{z}_{t^*k})$ we have

$$\frac{\partial \mathbf{z}_{t^*k}}{\partial \theta} = \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r} \frac{\partial \mathbf{z}_{t^*k}}{\partial \theta} + \frac{\partial^+ \mathbf{z}_{t^*k}}{\partial \theta}. \quad (69)$$

Accordingly

$$\frac{\partial \mathbf{z}_{t^*k}}{\partial \theta} = \left(\mathbf{I} - \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r} \right)^{-1} \frac{\partial^+ \mathbf{z}_{t^*k}}{\partial \theta} = \frac{\text{adj}\left(\mathbf{I} - \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r}\right)}{P_{\prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r}}(1)} \frac{\partial^+ \mathbf{z}_{t^*k}}{\partial \theta}, \quad (70)$$

where $P_{\prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r}}(1) = \det \left(\mathbf{I} - \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r} \right)$. Moreover, from (63) and (70) we have

$$\left\| \frac{\partial \mathcal{L}_t}{\partial \theta} \right\| = \frac{1}{P_{\prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r}}(1)} \left\| \frac{\partial \mathcal{L}_t}{\partial \mathbf{z}_{t^*k}} \text{adj} \left(\mathbf{I} - \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r} \right) \frac{\partial^+ \mathbf{z}_{t^*k}}{\partial \theta} \right\|. \quad (71)$$

Now, suppose that Γ_k undergoes a DTB, such that the fixed or cyclic points given by (66) tend to infinity and one of their eigenvalues tends to 1 for some parameter value $\theta = \theta_0$. This implies $P_{\prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r}}(1)$ becomes zero at $\theta = \theta_0$ and so, due to (70), $\left\| \frac{\partial \mathbf{z}_{t^*k}}{\partial \theta} \right\|$ goes to infinity. Therefore the norm of the loss gradient, $\left\| \frac{\partial \mathcal{L}_t}{\partial \theta} \right\|$, tends to infinity at $\theta = \theta_0$ which results in a abrupt jump in the loss function.

Let $\{\mathbf{z}_{t_1}, \mathbf{z}_{t_2}, \mathbf{z}_{t_3}, \dots\}$ be an orbit which converges to Γ_k , i.e.

$$\lim_{n \rightarrow \infty} d(\mathbf{z}_{t_n}, \Gamma_k) = 0. \quad (72)$$

Then there exists a neighborhood U of Γ_k and k sub-sequences $\{\mathbf{z}_{t_{k_m}}\}_{m=1}^{\infty}$, $\{\mathbf{z}_{t_{k_m+1}}\}_{m=1}^{\infty}$, \dots , $\{\mathbf{z}_{t_{k_m+(k-1)}}\}_{m=1}^{\infty}$ of the sequence $\{\mathbf{z}_{t_n}\}_{n=1}^{\infty}$ such that all these sub-sequences belong to U and

- a) $\mathbf{z}_{t_{k_m+s}} = F^k(\mathbf{z}_{t_{k(m-1)+s}})$, $s = 0, 1, 2, \dots, k-1$,
- b) $\lim_{m \rightarrow \infty} \mathbf{z}_{t_{k_m+s}} = \mathbf{z}_{t^*k-s}$, $s = 0, 1, 2, \dots, k-1$,
- c) for every $\mathbf{z}_{t_n} \in U$ there is some $s \in \{0, 1, 2, \dots, k-1\}$ such that $\mathbf{z}_{t_n} \in \{\mathbf{z}_{t_{k_m+s}}\}_{m=1}^{\infty}$.

This implies for every $\mathbf{z}_{t_n} \in U$ with $\mathbf{z}_{t_n} \in \{\mathbf{z}_{t_{k_m+s}}\}_{m=1}^{\infty}$, there exists some $\tilde{n} \in \mathbb{N}$ such that $\mathbf{z}_{t_n} = \mathbf{z}_{t_{k\tilde{n}+s}}$ and $\lim_{\tilde{n} \rightarrow \infty} \mathbf{z}_{t_{k\tilde{n}+s}} = \mathbf{z}_{t^*k-s}$. Consequently, there exists some $\tilde{N} \in \mathbb{N}$ such that for every $\tilde{n} \geq \tilde{N}$ both $\mathbf{z}_{t_{k\tilde{n}+s}}$ and \mathbf{z}_{t^*k-s} belong to the same sub-region and so the matrices $\mathbf{W}_{\Omega(t_{k\tilde{n}+s})}$ and $\mathbf{W}_{\Omega(t^*k-s)}$ ($s \in \{0, 1, 2, \dots, k-1\}$) are identical. Without loss of generality, let $s = 0$. Since $\mathbf{z}_{t_{k(\tilde{n}+1)}} = F^k(\mathbf{z}_{t_{k\tilde{n}}})$, so

$$\frac{\partial \mathbf{z}_{t_{k(\tilde{n}+1)}}}{\partial \theta} = \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r} \frac{\partial \mathbf{z}_{t_{k\tilde{n}}}}{\partial \theta} + \frac{\partial^+ \mathbf{z}_{t_{k(\tilde{n}+1)}}}{\partial \theta}. \quad (73)$$

On the other hand, $\lim_{\tilde{n} \rightarrow \infty} \frac{\partial \mathbf{z}_{t_{k(\tilde{n}+1)}}}{\partial \theta} = \lim_{\tilde{n} \rightarrow \infty} \frac{\partial \mathbf{z}_{t_{k\tilde{n}}}}{\partial \theta}$, which results in

$$\begin{aligned} \lim_{\tilde{n} \rightarrow \infty} \frac{\partial \mathbf{z}_{t_{k(\tilde{n}+1)}}}{\partial \theta} &= \left(\mathbf{I} - \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r} \right)^{-1} \lim_{\tilde{n} \rightarrow \infty} \frac{\partial^+ \mathbf{z}_{t_{k(\tilde{n}+1)}}}{\partial \theta} \\ &= \frac{\text{adj} \left(\mathbf{I} - \prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r} \right)}{P_{\prod_{r=0}^{k-1} \mathbf{J}_{t^*k-r}}(1)} \lim_{\tilde{n} \rightarrow \infty} \frac{\partial^+ \mathbf{z}_{t_{k(\tilde{n}+1)}}}{\partial \theta}. \end{aligned} \quad (74)$$

This means as $\tilde{n} \rightarrow \infty$, for any orbit converging to Γ_k the norm of the loss gradient tends to infinity at $\theta = \theta_0$ which completes the proof. \square

A.2.2 Proof of theorem 2

Proof. Let Γ_k be a k -cycle ($k \geq 1$) of (2) defined by periodic points (66). Suppose further that Γ_k undergoes a BCB for some parameter value $\theta = \theta_0$. Hence, one of its periodic points, e.g.

$\mathbf{z}_{t^{*k}}$, collides with one border. Therefore, $z_{m t^{*k}} = 0$ for some $1 \leq m \leq M$ by the definition of discontinuity boundaries in [38, 39]. Similar to the proof of Theorem 1, for $\boldsymbol{\theta} = \mathbf{A}, \mathbf{W}$ we have

$$\frac{\partial \mathbf{z}_{t^{*k}-1}}{\partial \theta} = \frac{\text{adj}\left(\mathbf{I} - \prod_{r=0}^{k-1} \mathbf{J}_{t^{*k}-1-r}\right)}{P_{\prod_{r=0}^{k-1} \mathbf{J}_{t^{*k}-1-r}}(1)} \frac{\partial^+ \mathbf{z}_{t^{*k}-1}}{\partial \theta}, \quad (75)$$

in which

$$\begin{aligned} \frac{\partial^+ \mathbf{z}_{t^{*k}-1}}{\partial w_{nm}} &= \mathbf{1}_{(n,m)} \mathbf{D}_{\Omega(t^{*k})} \mathbf{z}_{t^{*k}}, \\ \frac{\partial^+ \mathbf{z}_{t^{*k}-1}}{\partial a_{mm}} &= \mathbf{1}_{(m,m)} \mathbf{z}_{t^{*k}}, \end{aligned} \quad (76)$$

where $\mathbf{1}_{(n,m)}$ is an $M \times M$ indicator matrix with a 1 for the (n, m) 'th entry and 0 everywhere else. Since $z_{m t^{*k}} = 0$ at $\theta = \theta_0$, due to (76) $\frac{\partial^+ \mathbf{z}_{t^{*k}-1}}{\partial \theta}$ becomes the zero vector at $\theta = \theta_0$. Consequently, $\left\| \frac{\partial \mathbf{z}_{t^{*k}-1}}{\partial \theta} \right\|$ and so $\left\| \frac{\partial \mathcal{L}_t}{\partial \theta} \right\|$ vanishes at $\theta = \theta_0$. Now it can be shown that at $\theta = \theta_0$ the loss gradient goes to zero for every $\mathbf{z}_1 \in \mathcal{B}_{\Gamma_k}$ (the proof is similar to the last part of the proof of Theorem 1). \square

A.2.3 Proof of corollary 1

Proof. For $M = 2$, let $h_1 \neq 0$. Then the DFB curves of the fixed point Γ_1 coincide with the BCB curves of the 2-cycle $\mathcal{O}_{\mathcal{RL}}$ of the form

$$\mathcal{F}_1 = \xi_{\mathcal{RL}}^1 = \{(h_1, h_2, a_{11}, a_{22}) | 1 + a_{11} + a_{22} + a_{11}a_{22} = 0\}, \quad (77)$$

or

$$\mathcal{F}_2 = \xi_{\mathcal{RL}}^2 = \{(h_1, h_2, a_{11}, w_{11}, w_{21}, a_{22}) | 1 + a_{11} + w_{11} + a_{22} + (a_{11} + w_{11})a_{22} = 0\}. \quad (78)$$

For $M > 2$, assume that $\Gamma_1 = \{\mathbf{z}_1^*\}$ is a fixed point of the system, i.e.

$$\mathbf{z}_1^* = (\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)})^{-1} \mathbf{h} = \frac{\text{adj}(\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)})}{P_{\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)}}(1)} \mathbf{h}, \quad (79)$$

where $P_{\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)}}(1)$ is the characteristic polynomial of $\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)}$ at 1. Let us denote the first row of the adjoint matrix of $\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)}$ by $\text{adj}(\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)})_1$. If $\text{adj}(\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)})_1 \mathbf{h} \neq 0$, then we can analogously demonstrate that the DFB curves of the fixed point align with the BCB curves of the 2-cycles. This implies that, in accordance with Theorem 2, DFBs of fixed points will also lead to vanishing gradients in the loss function. \square

A.2.4 Convergence of SCYFI

To ensure that SCYFI almost surely converges, we can simply choose the number of random initializations (i.e., N_{out} in algorithm 1) large enough such that every linear subregion will have been sampled with probability almost 1. More precisely, drawing uniformly from the 2^M different \mathbf{D}_{Ω} -matrices (linear subregions) for initialization, the probability that a particular subregion has not been drawn after r repetitions is $p = (1 - \frac{1}{2^M})^r$. Hence, in order to ensure that all 2^M subregions have been visited with probability $1 - \epsilon$ we need $r \geq \lceil \frac{\ln(\epsilon)}{\ln(1 - \frac{1}{2^M})} \rceil$ iterations. Choosing $N_{out} = r$, we can thus ensure that SCYFI was initialized in each subregion with probability almost 1, and thus, in the limit, will have probed all subregions for dynamical objects. This argument extends to k -cycles by replacing 2^M by 2^{kM} above (strictly, a more precise bound for $k \geq 2$ is given by $2^{M(k-1)} \times (2^M - 1) = 2^{Mk} - 2^{M(k-1)}$, due to the fact that the PLRNN (2) is a linear map in each subregion and, hence, cannot have any k -cycles with all periodic points in only one subregion).

A.2.5 Proof of theorem 3

Proof. We examine the convergence and scaling behavior of SCYFI for fixed points. A similar argument applies to k -cycles where $k > 1$.

Let \mathbf{z}_1^* be a fixed point of the system, i.e.

$$\mathbf{z}_1^* = (\mathbf{I} - \mathbf{W}_{\Omega(t_1^*)})^{-1} \mathbf{h}. \quad (80)$$

\mathbf{z}_1^* is a true fixed point iff

$$(d_m(t_1^*) - a) \cdot z_{m,t_1^*} > 0, \quad \forall m \in \{1, 2, \dots, M\}, \quad (81)$$

where $\mathbf{D}_{\Omega(t_1^*)} = \text{diag}(d_1(t_1^*), d_2(t_1^*), \dots, d_M(t_1^*))$ and $0 < a < 1$ is a positive real constant.

For examining SCYFI's efficiency, here we focus on two scenarios that impose specific constraints on parameters θ ; other cases remain to be investigated.

Case (I) : Let \mathbf{R} be a randomly generated matrix with uniformly distributed entries in the interval $[0, 1)$, and $\mathbf{h} \neq 0$ be a random vector with all its components being non-negative. For an arbitrary $\epsilon > 0$, we set

$$\begin{aligned} \mathbf{A} &= \frac{1}{2 + \|\mathbf{R}\| + \epsilon} \text{diag}(\mathbf{R}), \\ \mathbf{W} &= \frac{1}{2 + \|\mathbf{R}\| + \epsilon} (\mathbf{R} - \text{diag}(\mathbf{R})). \end{aligned} \quad (82)$$

Then

$$\begin{aligned} \|\mathbf{A}\| &= \frac{\|\text{diag}(\mathbf{R})\|}{2 + \|\mathbf{R}\| + \epsilon} < \frac{1}{2 + \|\mathbf{R}\| + \epsilon}, \\ \|\mathbf{W}\| &= \frac{\|\mathbf{R} - \text{diag}(\mathbf{R})\|}{2 + \|\mathbf{R}\| + \epsilon} \leq \frac{\|\mathbf{R}\| + \|\text{diag}(\mathbf{R})\|}{2 + \|\mathbf{R}\| + \epsilon} < \frac{1 + \|\mathbf{R}\|}{2 + \|\mathbf{R}\| + \epsilon}, \end{aligned} \quad (83)$$

and so $\|\mathbf{A}\| + \|\mathbf{W}\| < 1$. Therefore

$$\forall t \quad \|\mathbf{W}_{\Omega(t)}\| = \|\mathbf{A} + \mathbf{W}\mathbf{D}_{\Omega(t)}\| \leq \|\mathbf{A}\| + \|\mathbf{W}\| \|\mathbf{D}_{\Omega(t)}\| \leq \|\mathbf{A}\| + \|\mathbf{W}\| < 1, \quad (84)$$

and so

$$\forall t \quad \rho(\mathbf{W}_{\Omega(t)}) \leq \|\mathbf{W}_{\Omega(t)}\| < 1. \quad (85)$$

In this case, for any $n \in \mathbb{N}$, we also have

$$\left\| \prod_{i=1}^n \mathbf{W}_{\Omega(t_i)} \right\| \leq \prod_{i=1}^n \|\mathbf{W}_{\Omega(t_i)}\| \leq (\|\mathbf{A}\| + \|\mathbf{W}\|)^n < 1. \quad (86)$$

This ensures the stability of all fixed points and k -cycles of the system.

According to (85), we have

$$(I - \mathbf{W}_{\Omega(t_1^*)})^{-1} = \sum_{n=0}^{\infty} \mathbf{W}_{\Omega(t_1^*)}^n = \mathbf{I} + \mathbf{W}_{\Omega(t_1^*)} + \mathbf{W}_{\Omega(t_1^*)}^2 + \dots \quad (87)$$

Hence, all the elements of $(I - \mathbf{W}_{\Omega(t_1^*)})^{-1}$ are positive, and so $z_{m,t_1^*} > 0$ for every t_1^* . This implies that all true and virtual fixed points exist within a singular sub-region. Additionally, only one fixed point is true, while all the other fixed points are virtual.

Case (II) : Let $\mathbf{h} = (h_1, h_2, \dots, h_M)^T$ be a random vector with all h_m uniformly distributed in $(0, 1]$ and

$$\begin{aligned} \beta_{min} &= \min \{h_m : h_m \in \mathbf{h}, 1 \leq m \leq M\} > 0, & 0 < \beta_{min} \leq 1, \\ \beta_{max} &= \max \{h_m : h_m \in \mathbf{h}, 1 \leq m \leq M\} > 0, & 0 < \beta_{max} \leq 1. \end{aligned} \quad (88)$$

Assume further that \mathbf{R}_1 is a randomly generated matrix with uniformly distributed entries in the interval $(-1, 0]$, and for $M \geq 2$

$$\mathbf{W} = \frac{\beta_{\min}}{M + \|\mathbf{R}_1\| + \epsilon} \left(\mathbf{R}_1 - \text{diag}(\mathbf{R}_1) \right). \quad (89)$$

Consider

$$\alpha_{\max} = \max \{ |w_{ij}| : w_{ij} \in \mathbf{W} \}, \quad 0 \leq \alpha_{\max} < \frac{\beta_{\min}}{M + \|\mathbf{R}\| + \epsilon} \quad (90)$$

and $S \subset \{1, 2, \dots, M\} = I$ such that $K = 2^{M - \text{card}(S)} \ll 2^M$. Suppose that $\mathbf{R}_2 = \text{diag}(r_1, \dots, r_M)$ is a randomly chosen diagonal matrix with r_m uniformly distributed in $(-1, 1)$ for $m \in I \setminus S$, and the other elements ($m \in S$) uniformly distributed in $(r^* - 1, 0)$ where $r^* = \frac{(M-1)\alpha_{\max} \beta_{\max}}{\beta_{\min}}$. Since

$$0 \leq \frac{(M-1)\alpha_{\max} \beta_{\max}}{\beta_{\min}} < \frac{(M-1)\beta_{\max}}{M + \|\mathbf{R}\| + \epsilon} \leq \frac{(M-1)}{M + \|\mathbf{R}\| + \epsilon} < \frac{(M-1)}{M} < 1, \quad (91)$$

so $-1 \leq r^* - 1 < 0$.

If

$$\mathbf{A} = \frac{1}{2 + \|\mathbf{R}_1\| + \epsilon} \mathbf{R}_2, \quad (92)$$

then

$$\begin{aligned} \|\mathbf{A}\| &= \frac{\|\mathbf{R}_2\|}{2 + \|\mathbf{R}_1\| + \epsilon} < \frac{1}{2 + \|\mathbf{R}_1\| + \epsilon}, \\ \|\mathbf{W}\| &= \frac{\beta_{\min} \|\mathbf{R}_1 - \text{diag}(\mathbf{R}_1)\|}{M + \|\mathbf{R}_1\| + \epsilon} \leq \frac{\|\mathbf{R}_1\| + \|\text{diag}(\mathbf{R}_1)\|}{M + \|\mathbf{R}_1\| + \epsilon} < \frac{1 + \|\mathbf{R}_1\|}{2 + \|\mathbf{R}_1\| + \epsilon}, \end{aligned} \quad (93)$$

which implies $\|\mathbf{A}\| + \|\mathbf{W}\| < 1$. We set $\epsilon > 0$ large enough to satisfy the condition

$$(I - \mathbf{W}_{\Omega(t_1^*)})^{-1} = \sum_{n=0}^{\infty} \mathbf{W}_{\Omega(t_1^*)}^n \approx I + \mathbf{W}_{\Omega(t_1^*)} \quad \forall t_1^*. \quad (94)$$

On the other hand, for any t we have

$$\mathbf{W}_{\Omega(t)} = \mathbf{A} + \mathbf{W} \mathbf{D}_{\Omega(t)} = \begin{pmatrix} a_{11} & w_{12}d_2(t) & w_{13}d_3(t) & \cdots & w_{1M}d_M(t) \\ w_{21}d_1(t) & a_{22} & w_{23}d_3(t) & \cdots & w_{2M}d_M(t) \\ w_{31}d_1(t) & w_{32}d_2(t) & a_{33} & \cdots & w_{3M}d_M(t) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{M1}d_1(t) & w_{M2}d_2(t) & w_{M3}d_3(t) & \cdots & a_{MM} \end{pmatrix}. \quad (95)$$

Hence

$$z_{m,t_1^*} = (1 + a_{mm}) h_m + \sum_{\substack{j=1 \\ j \neq m}}^M w_{mj} d_j(t_1^*) h_j = (1 + a_{mm}) h_m - \sum_{\substack{j=1 \\ j \neq m}}^M |w_{mj}| d_j(t_1^*) h_j. \quad (96)$$

Since for every t_1^*

$$\sum_{\substack{j=1 \\ j \neq m}}^M |w_{mj}| d_j(t_1^*) h_j \leq \sum_{\substack{j=1 \\ j \neq m}}^M |w_{mj}| h_j, \quad (97)$$

so

$$z_{m,t_1^*} \geq (1 + a_{mm}) h_m - \sum_{\substack{j=1 \\ j \neq m}}^M |w_{mj}| h_j \quad \forall m \in I. \quad (98)$$

Moreover $a_{ss} \in (r^* - 1, 0)$, for every $s \in S$, and thus

$$a_{ss} + 1 > \frac{(M-1)\alpha_{max}\beta_{max}}{\beta_{min}} = \frac{\sum_{j=1, j \neq s}^M \alpha_{max}\beta_{max}}{\beta_{min}} \geq \frac{\sum_{j=1}^M |w_{sj}| h_j}{h_s}. \quad (99)$$

Therefore, due to (98) and (99), $z_{s,t_1^*} > 0$ for every t_1^* and $s \in S$. This means that all true and virtual fixed points only exist within a relatively small number of sub-regions, denoted as $K = 2^{M-card(S)} \ll 2^M$. Given our specific initialization of θ , in both cases **(I)** and **(II)** there is a set of K different sub-regions, each associated with a unique $\mathbf{D}_{\Omega(t)}$ matrix. We refer to the entire set of these matrices as

$$\mathcal{D}_K = \{\mathbf{D}_1, \dots, \mathbf{D}_K\}. \quad (100)$$

SCYFI, by its definition, only moves within the sub-regions that have virtual and true fixed points, continuing until it discovers a true fixed point (or gets stuck in a virtual cycle). Thus, it can iterate between $J \leq K$ sub-regions

$$\mathcal{D}_J = \{\mathbf{D}_1, \dots, \mathbf{D}_J\} \subseteq \mathcal{D}_K, \quad (101)$$

or within the set of virtual fixed points

$$\mathcal{Z}_L = \{z_1, \dots, z_L\}. \quad (102)$$

In case **(I)**, there is only one true fixed point. Since all virtual fixed points are located within the same single sub-region, SCYFI's initialization will naturally position it within the correct linear region, requiring no more than 1 iteration. Hence, it needs at most 2 iterations to find the true fixed point. Consequently, SCYFI's scaling is constant.

For case **(II)**, if we suppose that SCYFI follows the virtual/true fixed point structure of the underlying system in these K sub-regions, the necessity for the probability of discovering the fixed point to be close to 1, specifically $1 - \epsilon$, is to have

$$N \geq \left\lceil \frac{\ln(\epsilon)}{\ln(1 - \frac{1}{2^{M-card(S)}})} \right\rceil = \left\lceil \frac{\ln(\epsilon)}{\ln(1 - \frac{1}{K})} \right\rceil, \quad (103)$$

iterations. Since $1 \leq card(S) \leq M - 1$, so $K \geq 2$ and $\ln(1 - \frac{1}{K}) \approx \frac{-1}{K}$. For $\epsilon^* \geq \epsilon$, let $N = \left\lceil \frac{\ln(\epsilon^*)}{\ln(1 - \frac{1}{K})} \right\rceil \geq \left\lceil \frac{\ln(\epsilon)}{\ln(1 - \frac{1}{K})} \right\rceil$, then

$$N = \left\lceil \frac{\ln(\epsilon^*)}{\ln(1 - \frac{1}{K})} \right\rceil \leq \frac{\ln(\epsilon^*)}{\ln(1 - \frac{1}{K})} \approx \ln\left(\frac{1}{\epsilon^*}\right)K := cK, \quad (104)$$

which implies the number of iterations is bounded from above. If, for every M , we choose K small enough, then the upper bound will stay within a linear growth. \square

A.2.6 Proof of theorem 4

In GTF [24], during training RNN latent states are replaced by a weighted sum of forward propagated states $z_t = F_\theta(z_{t-1})$ and data-inferred states $\tilde{z}_t = G_\phi^{-1}(x_t)$ (obtained by inversion of the decoder model G_ϕ):

$$\tilde{z}_t := (1 - \alpha)z_t + \alpha\tilde{z}_t, \quad (105)$$

where $0 \leq \alpha \leq 1$ is the GTF parameter (usually adaptively regulated in training, see [24]). This leads to the following factorization of Jacobians in PLRNN (2) training:

$$\mathbf{J}_t^{GTF} = \frac{\partial z_t}{\partial z_{t-1}} = \frac{\partial z_t}{\partial \tilde{z}_{t-1}} \frac{\partial \tilde{z}_{t-1}}{\partial z_{t-1}} = \frac{\partial \mathbf{F}_\theta(\tilde{z}_{t-1})}{\partial \tilde{z}_{t-1}} \frac{\partial \tilde{z}_{t-1}}{\partial z_{t-1}} = (1 - \alpha)\tilde{\mathbf{J}}_t = (1 - \alpha)\mathbf{W}_{\Omega(t)}. \quad (106)$$

Proof. (i) Since $\|\mathbf{A}\| + \|\mathbf{W}\| \leq 1$, we have

$$\forall t \quad \|\mathbf{W}_{\Omega(t)}\| = \|\mathbf{A} + \mathbf{W}\mathbf{D}_{\Omega(t)}\| \leq \|\mathbf{A}\| + \|\mathbf{W}\| \|\mathbf{D}_{\Omega(t)}\| \leq \|\mathbf{A}\| + \|\mathbf{W}\| \leq 1, \quad (107)$$

and so

$$\forall t \quad \rho(\mathbf{W}_{\Omega(t)}) \leq \|\mathbf{W}_{\Omega(t)}\| \leq 1, \quad (108)$$

where ρ denotes the spectral radius of a matrix. In this case, for any $n \in \mathbb{N}$, we also have

$$\rho\left(\prod_{i=1}^n \mathbf{W}_{\Omega(t_i)}\right) \leq \left\| \prod_{i=1}^n \mathbf{W}_{\Omega(t_i)} \right\| \leq \prod_{i=1}^n \|\mathbf{W}_{\Omega(t_i)}\| \leq \left(\|A\| + \|W\|\right)^n \leq 1. \quad (109)$$

Now, for any $0 < \alpha < 1$, the product of Jacobians under GTF is

$$\prod_{i=1}^n \mathbf{J}_{t_i}^{GTF} = (1 - \alpha)^n \prod_{i=1}^n \tilde{\mathbf{J}}_{t_i} = (1 - \alpha)^n \prod_{i=1}^n \mathbf{W}_{\Omega(t_i)}, \quad (110)$$

and

$$\begin{aligned} \rho\left(\prod_{i=1}^n \mathbf{J}_{t_i}^{GTF}\right) &= \rho\left((1 - \alpha)^n \prod_{i=1}^n \mathbf{W}_{\Omega(t_i)}\right) = (1 - \alpha)^n \rho\left(\prod_{i=1}^n \mathbf{W}_{\Omega(t_i)}\right) \\ &\leq (1 - \alpha)^n < 1. \end{aligned} \quad (111)$$

Hence $\rho\left(\prod_{i=1}^n \mathbf{J}_{t_i}^{GTF}\right) < 1$ which implies for any $n \in \mathbb{N}$ and $0 < \alpha < 1$, the product $\prod_{i=1}^n \mathbf{J}_{t_i}^{GTF}$ has no eigenvalue equal to 1 and so no DTB can occur (see definition of DTB in sect. 3).

(ii) Let $\|A\| + \|W\| = r > 1$, then for any $n \in \mathbb{N}$ we have

$$\rho\left(\prod_{i=1}^n \tilde{\mathbf{J}}_{t_i}\right) \leq r^n, \quad (112)$$

and thus

$$\rho\left(\prod_{i=1}^n \mathbf{J}_{t_i}^{GTF}\right) = (1 - \alpha)^n \rho\left(\prod_{i=1}^n \mathbf{W}_{\Omega(t_i)}\right) \leq [(1 - \alpha)r]^n. \quad (113)$$

Since $0 < 1 - \frac{1}{r} < 1$, inserting $1 - \frac{1}{r} < \alpha = \alpha^* < 1$ into the r.h.s. of (113) again gives $\rho\left(\prod_{i=1}^n \mathbf{J}_{t_i}^{GTF}\right) < 1$ for any $n \in \mathbb{N}$, implying that no DTB can occur. \square

A.3 Additional results

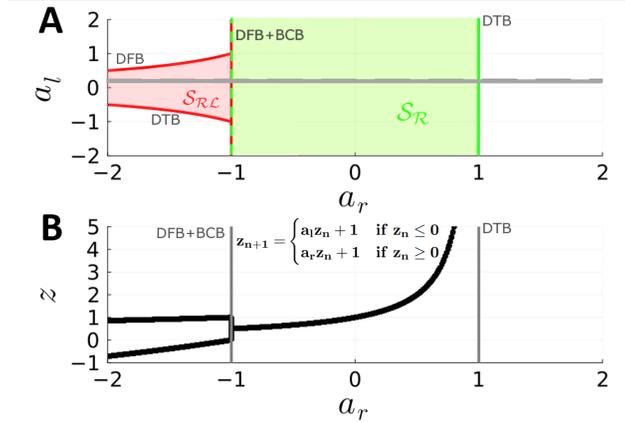


Figure S6: A) Analytically calculated stability regions for a 2-cycle $\mathcal{S}_{\mathcal{R}\mathcal{L}}$ (red), and a fixed point $\mathcal{S}_{\mathcal{R}}$ (green) for the $1d$ skew tent map, as defined in the figure, in the parameter plane given by (a_r, a_l) . B) Bifurcation diagram along the cross section indicated by the gray line in A, showing a BCB and DFB occurring simultaneously at $a_r = -1$ and a DTB occurring at $a_r = 1$.

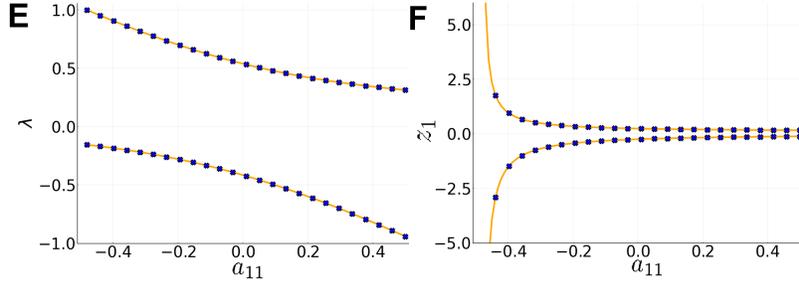


Figure S7: Results from SCYFI (blue) versus analytical results (orange) for the 2-cycle in Fig. 1 ($a_{11} + w_{11} = -2$). Eigenvalues (left) and location in state space (right) for one of the cyclic points. This confirms that fixed point locations and eigenvalues computed in closed-form and via SCYFI exactly agree, as they should.

A.3.1 Scaling analysis

Although the results presented in Fig. 2 suggest that SCYFI's scaling behavior is much better than theoretically expected, the fact that it is hard to obtain ground truth comparisons for high-dimensional systems (because of the combinatorial explosion) generally makes an extensive empirical analysis difficult. For Fig. 2 we therefore focused on scenarios for which we can also provide analytical curves for an exhaustive search strategy (eq. (5)) and where we then either examined scaling with cycle-order k for rather low-dimensional systems, or where we explicitly embedded fixed points to search for which allowed us to move to very high dimensionality M . In general we observed that the scaling behavior also depended on the PLRNN's matrix norms and the eigenspectrum of the embedded fixed points, so we constructed different scenarios where we varied these factors as well.

To construct a fairly well behaved case with low matrix norms, we randomly generated matrices \mathbf{R} with uniformly distributed entries in the interval $[-1, 1]$ and then normalized by its maximal eigenvalue: We set PLRNN parameters $\mathbf{A} = \frac{1}{\lambda_{max}} \text{diag}(\mathbf{R})$ and $\mathbf{W} = \frac{1}{\lambda_{max}} (\mathbf{R} - \text{diag}(\mathbf{R}))$, and chose \mathbf{h} uniformly in the interval $[-50, 50]$. For each of 10 different such systems, we fixed the number of outer loops and inner loops (N_{out} , N_{in} in Algorithm 1) such that a fixed point would be detected in at least 50/75 independent runs of the algorithm, and then determined the total number n of linear regions (i.e., across all N_{out} different initializations) the algorithm needed to cycle through

to detect a stable fixed point. We also ensured that across all different runs this stable fixed point would be the same, in accordance with our assumptions. The resulting scaling behavior was well fitted by a doubly-logarithmic curve of the form $c_1 \ln(\ln(M)) + c_2$ ($R^2 \approx 0.913, p < 10^{-4}$). This low-matrix norm scenario with a stable fixed point may be seen as a kind of lower bound on the scaling.

To embed a specific fixed point z^* , we again start with a matrix R as described above and take $A = \text{diag}(R)$ and $W = (R - \text{diag}(R))$. We then minimize

$$\min_{A, W, h} |z^* - ((A + W \cdot D_{\Omega(t^*)}) \cdot z^* + h)|, \quad (114)$$

subject to A staying diagonal and W off-diagonal (we observed that adding a small Gaussian noise term to the right appearance of z^* in eq. 114 which decayed proportionally to the learning rate improved numerical stability in the optimization process). The such constructed PLRNNs generally have several fixed points, but to compute n we only search for the inserted fixed point z^* (making eq. (5) directly applicable). This way we produced 5 – 10 systems, initializing R with values in $[-0.2, 0.2]$ (orange curve in Fig. 2B) or $[-1.0, 1.0]$ (blue curve in Fig. 2B), thus effectively restricting the eigenspectrum of the fixed point as well as the matrix norms of the PLRNN to a certain range. However, since matrix norms may change during optimization, eq. (114), our procedure is not strictly guaranteed to produce eigenspectra and matrix norms within a desired range, which is crucial especially for the first scenario where we wanted to keep norms within a ‘typical range’ (see below). So here, to ensure consistency among drawn systems and with our assumptions, the mean absolute eigenvalue of the embedded fixed points was kept close to 0.31 ± 0.05 and the mean maximum absolute eigenvalue close to 1.25 ± 0.13 . For $> 75\%$ of the resulting systems spectral matrix norms were within the range $[1.0, 3.0]$. While this produced matrix spectra typical for trained PLRNNs ($> 95\%$ out of 361 PLRNNs trained on various benchmarks and data had spectral matrix norms within $[1.0, 3.0]$), the second initialization range resulted in unnaturally large matrix norms and hence may be seen as providing a kind of upper bound on SCYFI’s scaling behavior. Fig. S8 shows the best case (left; purple curve in Fig. 2B) and typical (right; orange curve in Fig. 2B) scaling scenarios on linear scale to better expose the scaling behavior and function fits.

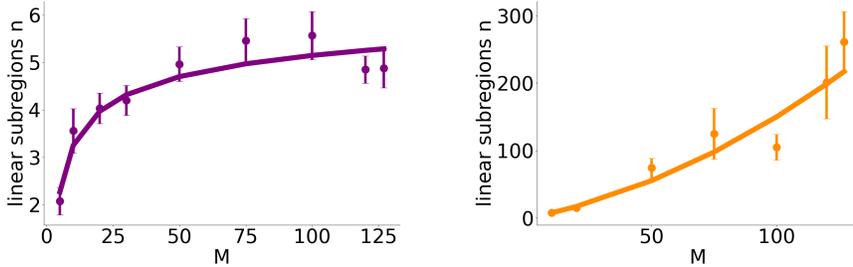


Figure S8: Zoom-ins on linear axes of the scenarios with doubly-logarithmic (left; $R^2 \approx 0.913, p < 10^{-4}$) and quadratic (right; $R^2 \approx 0.925, p < 10^{-5}$) scaling behaviors from Fig. 2B.

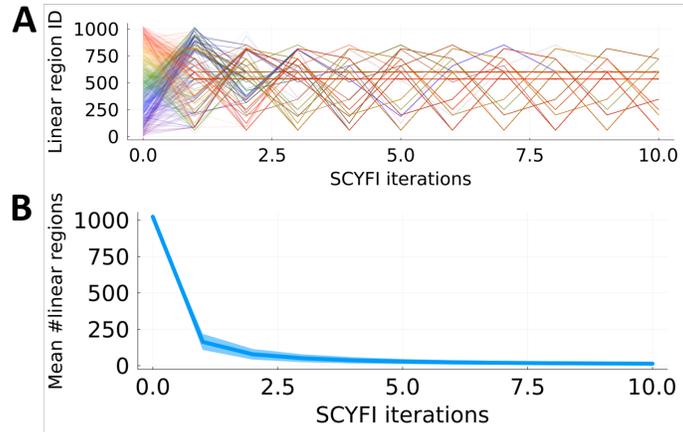


Figure S9: A) Initializing SCYFI in a wide array of different subregions (different colors), it quickly converges – within just a few iterations – to the same set of linear subregions which contain the dynamical objects of interest (fixed points in this case). B) The number of different subregions explored by SCYFI when started from different initializations shrinks exponentially fast with the number of iterations. Shown are means (\pm stdv) from 10 different systems with $M = 10$.

A.3.2 Loss jumps & bifurcations in PLRNN training on biophysical model simulations

Here we provide an additional illustration of how SCYFI can be used to dissect bifurcations in model training. For this, we produced time series of membrane voltage and a gating variable from a biophysical neuron model [17], on which we trained a dendPLRNN [10] using BPTT [52] with sparse teacher forcing (STF) [37]. The dendPLRNN used ($M = 9$ latent states, $B = 2$ bases) has 2^{18} different linear subregions and $|\theta| = 124$ parameters. Fig. S10A gives the MSE loss as a function of training epoch (i.e., single SGD updates). The loss curve exhibits several steep jumps. Zooming into these points and examining the transitions in parameter space using SCYFI, we find they are indeed produced by bifurcations, with an example given in Fig. S10B. As we had done for Fig. 4 in the main text, since the state and parameter spaces are very high dimensional, for the bifurcation diagram in Fig. S10B all extracted k -cycles ($k \geq 1$), including fixed points, were projected onto a line given by the PCA-derived maximum eigenvalue component, and plotted as a function of training epoch. For the example in Fig. S10B, we found that a BCB (Theorem 2) underlies the transition in the qualitative dynamics of the PLRNN as training progresses. Fig. S10C illustrates the dendPLRNN dynamics just before (left) and right after (right) the bifurcation point highlighted in Fig. S10B, together with time series from the true system.

More generally, whether a bifurcation associated with *vanishing* gradients produces a loss jump depends on the system’s dynamics before and after the bifurcation point. In the case of BCBs, one possible scenario involves a change in stability, as illustrated in Fig. S10. During a BCB, a stable fixed point (or cycle) can lose stability as it passes through the bifurcation point. The maximum Lyapunov exponent of an unstable fixed point (or cycle) is positive, resulting in exploding gradients right after the bifurcation point [37], and consequently to a very steep slope in the loss function near the bifurcation point as in Fig. S10.

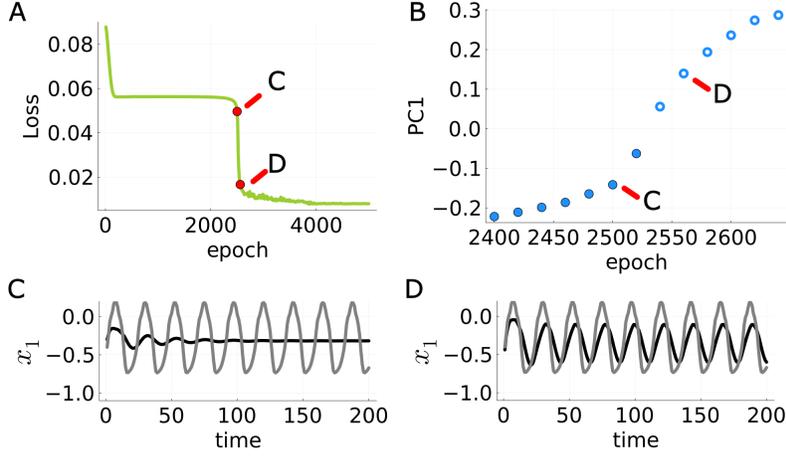


Figure S10: A) Loss across training epochs for a dendPLRNN ($M = 9$ states, $B = 2$ bases) trained on a biophysical neuron model in a limit cycle (spiking) regime. Red dots indicate training epochs just before and after a loss jump for which time graphs are given in C and D. B) Bifurcation diagram of the dendPLRNN as a function of training epoch, with all state space locations of stable (filled circles) and unstable (open circles) objects projected onto the first principle component. The loss jump in A is accompanied by a bifurcation from fixed point to cyclic behavior. C) Time series of the voltage variable (x_1) of the biophysical model (gray) and that predicted by the dendPLRNN (black) before the bifurcation event indicated in B. D) Same directly after the bifurcation event.

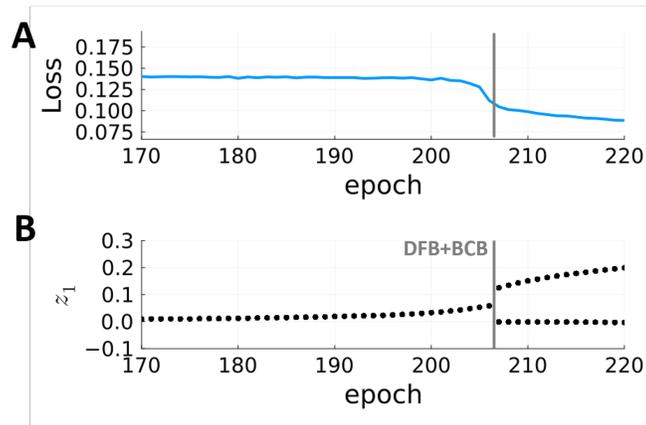


Figure S11: Loss jump induced by a degenerate flip bifurcation (DFB). A) Loss during a training run of a PLRNN ($M = 5$) on a 2-cycle. The gray line indicates a loss jump corresponding to a DFB and a simultaneously occurring border collision bifurcation (BCB). B) Bifurcation diagram of the PLRNN, with the DFB and BCB leading to the destruction of the fixed point and the emergence of a 2-cycle as indicated by the gray line.

A.3.3 Dealing with bifurcations in RNN training

Here are some additional thoughts on how RNN training algorithms could possibly be modified to deal with bifurcations. If the algorithm finds itself during training in a parameter regime which does not exhibit the right topological structure, it does not make sense to further dwell within that regime, or possibly anywhere within the vicinity of the current parameter estimate. Unlike standard SGD, the algorithm should therefore perhaps take large leaps in parameter space as soon as it gets stuck in a non-suitable dynamical regime. One possibility to implement this is through a ‘look-ahead’ mechanism that probes for topological properties of regions not visited so far. While fully fleshing out this idea is beyond this paper, a proof of concept that this may speed up convergence is provided in Fig. S12. Along similar lines, if we knew the model’s full bifurcation structure in parameter space ahead of time, we could simply pick a parameter set which corresponds to the right dynamics

describing patterns in the data best. While of course it will in general not be feasible to chart the whole bifurcation structure before training (this is in a sense the whole point of a training algorithm), it may be possible to design smart initialization procedures based on this insight, e.g. probing topological regimes at randomly selected points in parameter space before starting training and initializing with parameters that produce a desired type of dynamics (e.g., cyclic behavior) to begin with.

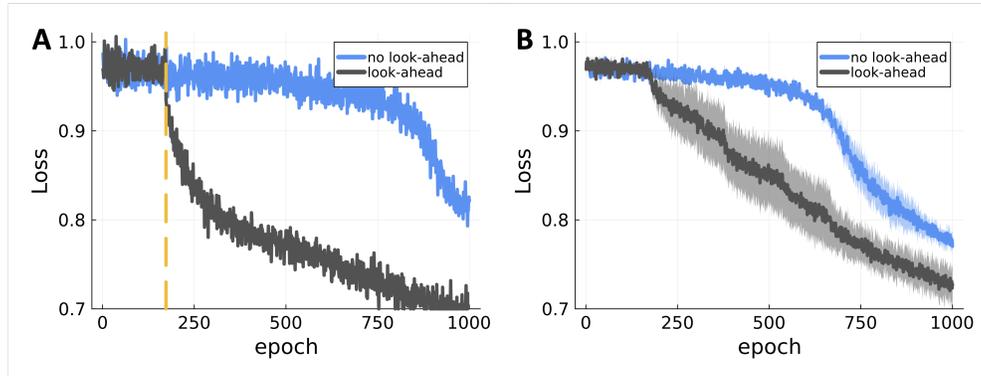


Figure S12: A) Example loss curves for RNNs trained on electrophysiological recordings by BPTT without (blue) vs. with (black) 'look-ahead' (the look-ahead function checks whether there would be a bifurcation away from a stable fixed point when taking $10\times$ the current gradient step). Dashed yellow line indicates the epoch at which the look-ahead step was executed. B) Average across 6 loss curves of RNNs trained without (blue) vs. with (black) look-ahead. Error bands = SEM.