nature communications

Article

https://doi.org/10.1038/s41467-026-75924-7

Data-driven reduced modeling of neural dynamics

Received: 1 October 2025

Accepted: 15 July 2026

Published online: 28 July 2026

Check for updates

A. Marraffa [1,2,5] ✉, R. Krause [3,4], V. Mante [3,4] & G. Haller [1]

Neural ordinary differential equations (ODEs) are widely used in neuroscience to model the collective activity of neurons during behavioral tasks. The high dimensionality of their parameter and activity spaces, however, often make it challenging to infer and interpret the fundamental features of their dynamics. In this study, we employ recent nonlinear dynamical system techniques to uncover the core dynamics of several Neural ODEs used in contemporary neuroscience. Specifically, using a data-driven approach, we identify Spectral Submanifolds (SSMs), i.e., low-dimensional attracting invariant manifolds tangent to the eigenspaces of fixed points. The internal dynamics of SSMs serve as nonlinear models that reduce the dimensionality of the full RNNs by orders of magnitude. Through low-dimensional, SSM-reduced models, we give mathematically precise definitions of line and ring attractors, which are intuitive concepts commonly used to explain decision-making and working memory. This unprecedented level of understanding of Neural ODEs obtained from SSM reduction enables the interpretation of mathematically well-defined and robust structures in neuronal dynamics, leading to predictions about the neural computations underlying behavior.

Behavior emerges from concerted activity across large populations of neurons. In the study of neural population activity, two primary questions emerge: How do neurons collectively create an internal representation of the external world based on the inputs they receive, and how do neural computations on these representations lead to behavior?

Recent advances in neuronal modeling and analysis have made it possible to address these questions, providing insights into the nature of computations implemented at the level of neural populations. To establish a connection between the firing patterns of neural populations and low-dimensional, interpretable behavioral outputs, individual neurons can be interpreted as degrees of freedom within a dynamical system. Computations can then be understood in terms of the evolution of this dynamical system when driven by the inputs it receives from the senses or from other areas[1,2]. Dynamical systems modeling naturally lends itself to analyses based on systems identification, whereby neural computations reflect the properties and interaction of low-dimensional, attracting structures, even when single-neuron responses are not directly interpretable.

Accumulating evidence from such approaches suggests that specific classes of dynamical motifs are consistently involved in cognitive and motor operations. For instance, saddle-point dynamics have been associated with decision-making processes[3,4], and rotational dynamics have been proposed as a hallmark of population activity underlying movement preparation and execution[5,6]. Some other heuristics are also helpful in characterizing observed dynamics in more complex tasks: "ring attractors" have been proposed as substrates for working memory[7–12], and "line attractors" have been discussed for tasks that require the accumulation of evidence[13–16]. Importantly, more involved and structured behavioral tasks could be understood as the composition of such dynamical primitives[17,18].

Quantitatively identifying and characterizing such dynamical motifs requires accurate models of neural population dynamics. To this end, a wide range of methods has been developed in recent years

[1]Institute for Mechanical Systems, ETH Zurich, Zurich, Switzerland. [2]VIB Center for AI and Computational Biology, VIB, Leuven, Belgium. [3]University of Zurich & ETH Zurich, Zurich, Switzerland. [4]Neuroscience Center Zurich, University of Zurich & ETH Zurich, Zurich, Switzerland. [5]Present address: Department of Electrical Engineering, KU Leuven, Leuven, Belgium. ✉e-mail: alice.marraffa@kuleuven.be

to infer dynamics from population-level recordings. Many techniques rely on fitting a model of dynamics to the data (see ref. 19 for an exception), but generally suffer from a tradeoff between complexity, expressive power, and the interpretability of the fitted models. At one extreme, machine-learning-based methods such as LFADS[20] and CEBRA[21] (see also refs. 22,23) can be used to fit highly nonlinear, even chaotic dynamics, but often at the expense of mechanistic insight and predictive power. At the other end are models based on linear dynamical systems, either a single one[24], or switching[25], piecewise linear[26], or time-dependent linear systems[27]. Since the possible dynamics of linear systems are well understood, the resulting descriptions of dynamics are highly interpretable, but cannot capture intrinsically nonlinear behavior, such as coexisting attractors and chaotic dynamics.

Here, we propose an approach to estimate neural dynamics based on a recent nonlinear model reduction technique that strikes a balance between the ability to capture intrinsically nonlinear dynamics and interpretability. Central to our approach is the recent theory of Spectral Submanifolds (SSMs)[28], which are ubiquitous and robust low-dimensional, attracting invariant manifolds that originate from steady state solutions along eigenspaces of the locally linearized dynamics (see[29] for a general introduction, applications, and further references). Invariant manifolds are sets in the phase space that have a manifold structure and are mapped to themselves by the system's evolution map, the flow map. These manifolds are attracting if the trajectories converge to them asymptotically, while they are robust if they persist as an invariant set under small perturbations of the system, and smoothly depend on them. The internal dynamics of SSMs serve as low-dimensional, explicit, polynomial models that explain the dominant slow nonlinear dynamics of high-dimensional dynamical systems. Based on these reduced models, dynamical motifs in neural networks can be analyzed and explained in a mathematically rigorous way.

We demonstrate the power of data-driven SSM-based modeling for neural dynamics using simulation output data from neural ordinary differential equations, also referred to as recurrent neural networks (RNNs). We will use the term RNN in the rest of the paper, as it is clear from the context that we mean the specific class of RNNs identifiable with neural ODEs. RNNs trained on specific behavioral tasks have emerged not only as a key tool for reproducing the features of neural dynamics in brain recordings, but also for generating hypotheses about the neural computations implemented by the underlying neural circuits. The goal is to abstract the essence of the computations, focusing on the interactions between inputs and recurrent dynamics[30],[13,31–37]. However, obtaining a complete understanding of the dynamics of RNNs is often challenging, even when all their underlying parameters are fully known. Commonly used techniques for identifying dynamical motifs in RNNs rely on the search for fixed points and "slow points"[13,18,38], i.e., locations in neural state space where dynamics do not change or vary slowly. The geometry of slow points is largely conserved across different RNNs solving the same tasks and is therefore considered a useful heuristic of the computations they implement[30]. Further insights into computations can be obtained through linear approximations of the dynamics around the identified fixed points.

Such standard approaches to understanding RNN computations have several downsides, which are overcome by SSM-based modeling. First, slow points are typically found numerically based on subjectively chosen thresholds on the rate of change of neural dynamics. Second, linear approximations of non-linear systems are valid only only locally around fixed points, but not around slow points. Third, even when dynamics can be linearized, around hyperbolic fixed points the linear approximations are only valid in domains with simple, linearizable dynamics. These domains cannot contain multiple isolated fixed points, periodic orbits, quasiperiodic tori, chaotic attractors, or transitions among such invariant sets. In contrast, we show that data-driven algorithms to estimate low-dimensional SSM and their reduced dynamics can provide a complete, yet simple description of both linear and nonlinear dynamics across all computationally relevant activity domains. The resulting mathematical descriptions of the global dynamics[39] provide the basis for more precise hypotheses about the nature of neural computations in biological neural networks.

Results

We model recurrent neural networks with evolution rules of the form:

$$\frac{dx}{dt} = R_\theta(x(t), u(t)) \quad (1)$$

where $x(t) \in \mathbb{R}^N$ is the network state variable, $u(t) \in \mathbb{R}^{N_{in}}$ is the input vector and R_θ is a function depending on the parameter vector θ that defines the network dynamics.

Our approach to analyzing autonomous, high-dimensional RNN dynamics begins by assuming the existence of a known steady-state solution, which in practice can be identified as zeros of the right-hand-side of Eq. (1). Around such a steady state, we can deduce the existence of attracting invariant SSMs.

SSMs can be technically defined as the smoothest invariant manifolds that are tangent to dominant eigenspaces of steady states. The existence of these manifolds is ensured under nonresonance conditions on the spectrum of the linearized dynamical system at the steady state[29]. A spectrum is nonresonant if no eigenvalue can be written as a linear combination of the other eigenvalues with positive integer coefficients. Once the existence of an SSM is established, it can be approximated through a polynomial expansion over the dominant spectral subspace used in its identification. While the existence of SSMs is deduced from the linearized spectrum at a steady state, these invariant manifolds do extend beyond the domain of linearization and hence can carry characteristically nonlinear dynamics. The internal dynamics of attracting SSMs serve as a mathematically exact reduced-order model with which nearby trajectories synchronize exponentially fast.

We employ the SSMLearn algorithm[39] to find explicit polynomial parametrizations for low-dimensional, attracting, invariant SSMs attached to steady-state solutions. SSMLearn then approximates the reduced polynomial ordinary differential equation (ODE) that governs the system evolution restricted to the SSM, providing a low-dimensional, polynomial model that captures the core dynamics of the original RNN.

A key advantage of this approach is the substantial reduction in model dimensionality, which converts a high-dimensional RNN into a low-dimensional polynomial ODE. Therefore, SSM-reduced models allow for the explicit identification of phase space structures, such as fixed points, limit cycles, and bifurcations, that determine the global asymptotic behavior of the network. To assess the effect of system parameters on dynamics (1), we can construct a slow manifold for each value of a given system parameter and view it as a section of a global slow manifold in the extended phase space that includes the system parameters as well. This, in turn, enables us to build parameter-dependent polynomial models for the SSM-reduced dynamics. As a result, parameters are explicitly present in the SSM and reduced dynamics equations, enabling us to clearly understand their role.

A model for context-dependent decision-making

As a first example, we analyze the 100D RNN performing the context-dependent decision-making task introduced by ref. 13, as described in the Methods section "RNN for context-dependent decision-making". This network receives as input two sensory cues, s_1 and s_2, whose values can be positive or negative, and two context cues, (c_1, c_2) with values in {(0, 1), (1, 0)}. The network's output is expressed through a

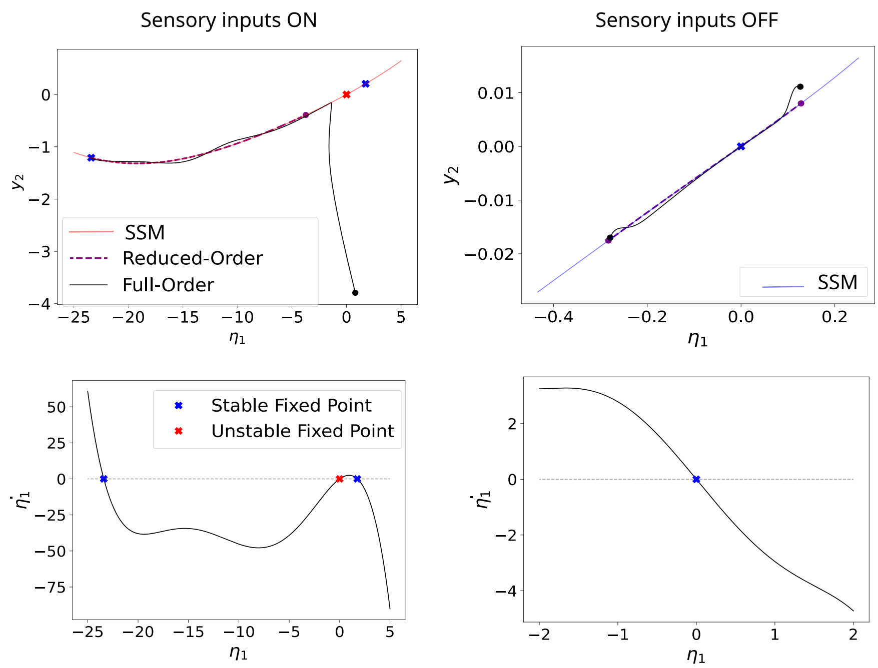

Fig. 1 | SSMs carrying the RNN reduced dynamics. (Upper left) Unstable manifold at order five in coordinates (η_1, y_2), where η_1 parametrizes the spectral subspace E_1 and $y_i = x_i - x_{i,0}$ are the original RNN coordinates centered around the unstable fixed point. We depict a full-order test trajectory and the corresponding reduced trajectory on the SSM (after $t \approx \frac{1}{\lambda_2}$, when transients have decayed, purple dot), converging to the fixed point with the larger domain of attraction along the unstable manifold. Black dots correspond to initial conditions. The fixed points in the plot (crosses) are the fixed points of the full model, lying on the unstable manifold and correctly reproduced by the reduced-order model. The Manifold Fitting Error (MFE, see Eq. (14)), the mean distance between observed trajectories and trajectories projected on the manifold, has magnitude ≈0.007, and the Normalized Mean Trajectory Error (NMTE, see Eq. (15)), the mean reconstruction error of the reduced model, is ≈0.03. (Upper right) The slowest SSM at order three and test full-order and reduced-order trajectories plotted in coordinates (η_1, y_2), with y_i now centered around the stable fixed point. The MFE (see Eq. (14)) and NMTE (see Eq. (15)) have magnitude ≈0.05. (Bottom left) Graph of the right-hand sides of the reduced-order model on the 1D unstable manifold at order five. We observe convergence to the stable fixed point. (Bottom right) Right-hand sides of the reduced-order model on the 1D slowest SSM at order five. The unstable fixed point (red cross) serves as a boundary for the domains of attraction of the two stable fixed points (blue crosses) on the unstable manifold.

scalar readout function that will have the same sign of s_1, if $(c_1, c_2) = (1, 0)$, or s_2, $(c_1, c_2) = (0, 1)$.

The authors of ref. 13 explain the network dynamics underlying flexible decision-making using the concept of a *line attractor*, which they view as a curve in phase space consisting of *slow points*[13,38,40]. Although not an attractor in a strict mathematical sense, the line attractor is a heuristic used to explain what is observed in the first principal component space: trajectories are attracted to this curve when sensory inputs are off and diverge from it once sensory inputs are turned on.

Within this 1D structure,[13] finds two stable fixed points that represent the two possible decisions in the task. The final state of the network, i.e., the fixed point to which it converges, is then argued to depend on the direction of perturbation away from the line attractor, determined by the direction and magnitude of the input vector (see Fig. 3 in ref. 13 and Fig. 2 in ref. 41).

Here we revisit this interpretation through SSM theory, reinterpreting the RNN behavior as reduced dynamics on a 1D invariant slow manifold. We begin by analyzing the system under fixed context and sensory input, considering two conditions: the sensory input-off regime, corresponding to the cue period before sensory integration in animal experiments, and the sensory input-on regime, where active sensory processing occurs.

In both cases, we identify at least one fixed point in the system's phase space, which ensures the existence of SSMs for the flow map. In the sensory input-on case, the network has an unstable fixed point and two stable fixed points, each corresponding to a distinct choice in the task. Importantly, these two stable fixed points are present for every input parameter value in the range explored in this experiment. Linearization around the unstable fixed point reveals a single unstable eigenvalue, indicating a 1D unstable subspace. We use SSMLearn to find the expansion coefficients of the corresponding unstable manifold of the nonlinear system (Fig. 1, upper left) up to order five and derive the associated reduced-order model

$$\dot{\eta}_1 = \sum_{i=1}^{5} a_i \eta_1^i + \mathcal{O}(|\eta_1|^6) \tag{2}$$

whose right-hand side is plotted in Fig. 1, bottom left. The SSM-reduced model (2) at this order of approximation enables accurate prediction of the full RNN dynamics from simulations of the reduced model.

In fact, the 1D model (2) is sufficient to reproduce the RNN behavior during decision making. On the unstable manifold, the size of the domains of attraction of the two stable fixed points determines the global phase space geometry: random initial conditions are more likely to evolve toward the fixed point associated with the larger domain of attraction, which corresponds to the correct choice in the task. We estimate these domains of attraction in the full phase space using Finite-Time Lyapunov Exponent (FTLE) analysis in the Supplementary Materials. FTLE ridges identify regions of phase space where initial conditions close to each other have a high rate of separation. In this context, these initial conditions mark the boundaries of the domains of attraction of the two fixed points.

In the sensory input-off case, the dynamics in a small neighborhood of the origin (where initial conditions are set in the experiments) are governed by a stable fixed point with one slowest eigenmode in its linearized dynamics. Using SSMLearn, we find the expansion

https://doi.org/10.1038/s41467-026-75924-7

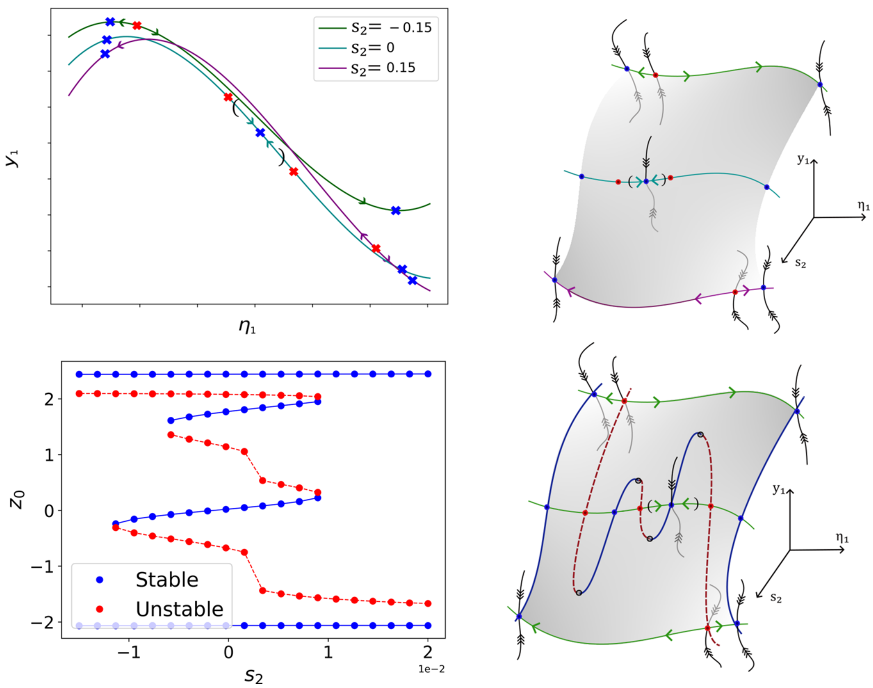

Fig. 2 | Bifurcation diagram in the extended phase space of the context-dependent model. (Upper left) Plot of the 1D slow manifolds in a section of the phase space with coordinates (η_1, y_1) for selected s_2 values. (Upper right) Sketch of the robust (normally hyperbolic) 2D invariant manifold (*slow manifold)* in the extended phase space constructed including the s_2-parameter direction, in coordinates (η_1, s_2, y_1). The 1D latent dynamics underlying flexible decision-making change when varying the relevant sensory input s_2 from negative to positive values, resulting in a change in the preferred choice for the network. When $s_2 = 0$, there is a stable fixed point that attracts initial conditions around zero and, locally, the slow manifold coincides with its 1D slowest SSM (in round brackets). (Bottom left) Bifurcation diagram for the fixed points of the RNN, when sensory input 2 s_2 (selected by context input 2) is varied in the range $s_2 \in [-0.15, 0.15]$ and sensory input 1 is kept fixed at $s_1 = 0.036$. Four saddle-node bifurcations occur in the plotted critical parameter range $[-0.015, 0.020]$ around zero. Through this cascade of saddle-node bifurcations, the system adapts by converging to a fixed point that yields the correct readout: negative when $s_2 < 0$, and positive when $s_2 > 0$. We plot a one-dimensional readout defined by $z_0 = \langle \mathbf{Y}, \mathbf{x}_0 \rangle$. (Bottom right) Sketch of the cascade of saddle-node bifurcations on the robust, normally hyperbolic 2D slow manifold in the extended phase space. When $s_2 = 0$, the stable fixed point around zero appeared via a saddle-node bifurcation attracts trajectories along its 1D slowest SSM (in round brackets).

coefficients of the corresponding 1D slowest SSM (see Fig. 1, upper right). The equations for the reduced dynamics on the SSM have the form (2). The graph of the right-hand side is shown in Fig. 1 (bottom right).

These results lead to a re-interpretation of the line attractor in ref. 13. The set of slow points identified in ref. 13 during the sensory input off-period corresponds to the slowest SSM anchored to the stable fixed point. When sensory inputs are off, this manifold attracts nearby trajectories and carries stable, near-equilibrium dynamics. These dynamics are perceived as slow once the trajectories are close enough to the stable fixed point, to which they converge in infinite time.

When sensory inputs are activated, the dynamics transition: trajectories converge to the decision-related stable fixed point along the 1D unstable manifold of a now unstable fixed point (in Fig. 3 of ref. 13, trajectories labeled as sensory input on). The reduced dynamics restricted to this SSM exhibit a bistable structure, wherein the unstable fixed point acts as a boundary separating the domains of attraction of the two stable fixed points.

Importantly, as depicted in Fig. 2 (upper right and upper left), a 1D slow manifold is present throughout the explored range of sensory inputs and carries the relevant dynamics throughout the decision-making process. Specifically, when sensory inputs are zero, the slow manifold coincides locally with the slowest SSM attached to a stable fixed point. As the relevant sensory input s_i (with $i = 1$ for context 1 and $i = 2$ for context 2) is varied from zero, we retrieve the slow manifold as the unstable manifold of a (different for different s_i signs) unstable fixed point that separates the basins of attraction of the two stable equilibria (see Fig. 2, upper left and right).

We note that, even though the two stable fixed points corresponding to the two choices remain fixed, the number and location of the other fixed points change with the parameter. In particular, four saddle-node bifurcations occur within a small interval around $s_2 = 0$, as illustrated in the bifurcation diagram in Fig. 2 (bottom left and right). It is through this cascade of saddle-node bifurcations that the system adapts to favor the convergence to the stable fixed point that yields the correct readout (positive for positive values of s_2 and vice versa) along the branches of the 1D unstable manifold of an unstable fixed point (see Supplementary Material S2 for a more detailed discussion).

This cascade of bifurcations occurs, however, for sensory inputs close to zero. In the original experiment of ref. 13, the network would be, for the sensory input off-condition, always initialized close to zero. Therefore, trajectories would converge always to the same fixed point (in brackets in Fig. 2), which coincides with the stable fixed point on the right hand side of Fig. 1.

As we show in the Supplementary Methods SM1, we can assess the robustness of the 1D slow manifold as sensory inputs change, and we know that its geometry and associated reduced dynamics deform smoothly. In other words, in the extended phase space (constructed including the parameter axis, as in Fig. 2, bottom left and upper left), we find a 2D, at least C^1 attracting slow manifold that carries the latent dynamics underlying flexible decision-making. Around persisting fixed points, we can explicitly determine the dependencies on parameters through parameter-dependent polynomial expansions on their SSMs (see Supplementary Materials S2 for a complete discussion).

Our analysis also clarifies the role of the input vector and the recurrent dynamics in determining the outcome of the decision-

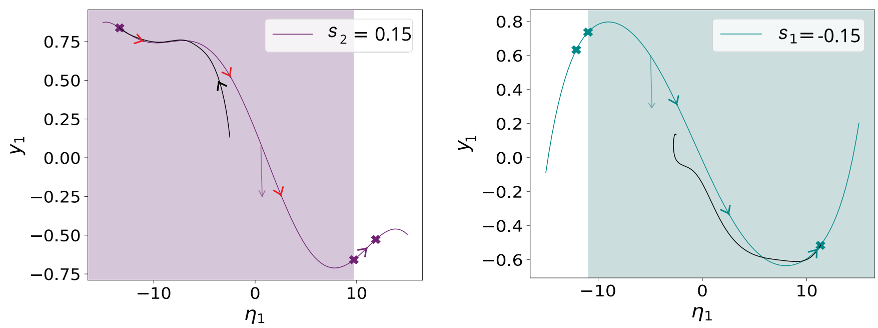

Fig. 3 | One-dimensional slow manifolds (unstable manifolds) of the system for different values of the relevant sensory input (s_2 and s_1, respectively). The way the RNN realizes the correct input-output relationship is by changing location to the unstable fixed point (central cross) along the slow manifold: the unstable fixed point separates the domains of attraction of the two stable fixed points in the reduced dynamics, so its location determines the asymptotic behavior of trajectories, given their initial conditions. The crosses correspond to the fixed points, the two stable one divided by the unstable one. We display the arrows indicating effect of the input vector on the reduced dynamics (projection onto the SSM). This picture shows that the projection of the input vector (or its mean, in the noisy case) on the reduced dynamics does not determine the asymptotic behavior (choice) of the RNN, as we can see that, for example, it would lead to the wrong direction when context 2 is selected and s_2 is positive. However, whenever the perturbed initial conditions lie within the correct domain of attraction, the convergence to the correct choice is guaranteed, as one can notice from the true trajectories (black lines).

making task, allowing us, in particular, to distinguish among the three hypotheses proposed by ref. 41. Notably, we find that the direction of the input vector in the reduced dynamics does not, by itself, determine the result of sensory integration (see Fig. 3). Moreover, as illustrated in Fig. 2 the geometry of the SMMs remains largely constant across context and sensory inputs. As the input vector changes orientation, the tangent spaces of the SSMs do not align with it to facilitate integration (as shown in Fig. 3). Instead, convergence is determined by the location of initial conditions relative to the two domains of attraction. The width of these domains (and the location of the separating unstable fixed point on the slow manifold) varies parametrically with the relevant sensory input (or with its mean in the noisy case), and with the context input values that select it, to achieve the desired outcome. In other words, mean input vectors are parameters that determine the geometry of the phase space of the RNN.

In summary, based on the above findings we propose an alternative to the line attractor interpretation of these RNN, namely a mathematically rigorous 1D slow manifold explanation, and provide a method to systematically find these slow manifolds emanating from known steady-state solutions.

More generally, 1D slow manifolds may play a central role in decision-making tasks, not only in RNNs but also in biological neural systems. For instance,[42] shows that a linear Vanilla RNN trained on mouse behavioral data from a sensory discrimination task learns a 1D slowest mode during training. This observation suggests that SSM theory has the potential to aid the analysis of neural computations in both artificial and biological settings. Importantly, the derivation of SSMs extends to cases where activity is corrupted by noise. We discuss the effects of adding noise to the RNN dynamics in the Supplementary Materials S3, where we treat uniformly bounded Gaussian noise as a small-amplitude, time-dependent perturbation to the RNN studied so far.

A model for oscillations with input-dependent frequency

Our second example, the 100D *Sine Wave Generator* RNN in ref. 43 (see Method section "Sine-wave generator RNN"), is trained to produce a sinusoidal scalar readout $z(t)$, with oscillation frequency selected by a scalar input u (see Eq. (16)).

In the phase space of the constant-input RNN, there is an unstable fixed point with two complex-conjugate unstable eigenvalues widely separated in their real parts from the stable ones. As a consequence, a 2D slow SSM exists in this problem and coincides with the unstable manifold of the fixed point. This SSM carries the observable oscillatory dynamics of the system. We find the SSM equations and its reduced dynamics with SSMLearn up to order three and six, which are the lowest expansion orders that give low MFE (see Eq. (14)) (distance of trajectories from the SSM) and NMTE (see Eq. (15)) (distance between full and reduced trajectories), respectively. Specifically, the SSM-reduced dynamics takes the form

$$\dot{\eta}_\ell = \sum_{i=0}^{6}\sum_{j=0}^{6-i} a_{ij}^{(\ell)}\,\eta_1^i\eta_2^j + \mathcal{O}(7), \quad \ell = 1, 2. \tag{3}$$

Analyzing the SSM-reduced dynamics (3), we find a globally attracting limit cycle that is responsible for the periodic oscillations with the desired frequency. This type of dynamics is exactly the simplest possible realization of a dynamical system that oscillates with a single frequency.

The unstable manifold, the limit cycle, and full and reduced trajectories are plotted in Fig. 4 (upper left), while contour lines of the vector field defining the reduced-order model found through SSMLearn at order seven, projected to the unstable subspace E_2 can be seen in Fig. 4 (bottom left). Hence, we obtain a 2D invariant manifold and reduced-order model which provide a complete and detailed description of the dynamics of this high-dimensional, oscillatory RNN. When we choose an input u with frequency $f = 1.9$ Hz, the frequency of the resulting limit cycle captured by the SSM-reduced model is $f_{RED} = 1.92$ Hz, which is very close to the measured frequency $f_{FULL} = 1.91$ Hz in the full RNN model.

To analyze the dependence of the limit cycle frequency on the input parameter u, we recall that unstable manifolds depend smoothly on parameters, as long as their underlying unstable spectral subspaces persist under the variation of those parameters (see, for example, ref. 44). As a consequence, we have a smooth, u-dependent family of 2D unstable manifolds in the phase space, as illustrated in Fig. 4 (upper right).

Along members of this SSM family, we have an explicit, u-dependent polynomial expansion for a 2D, parametric, SSM-reduced model of the form

$$\begin{aligned} \dot{\eta}_1 &= \sum_{i+j+k\le 4} a_{ijk}^{(1)}\,\eta_1^i\eta_2^j u^k + \mathcal{O}(5), \\ \dot{\eta}_2 &= \sum_{i+j+k\le 4} a_{ijk}^{(2)}\,\eta_1^i\eta_2^j u^k + \mathcal{O}(5). \end{aligned} \tag{4}$$

The limit cycle frequencies in the parametrized model (4) agree with the measured frequencies of the full trajectories up to an error of 10^{-1}.

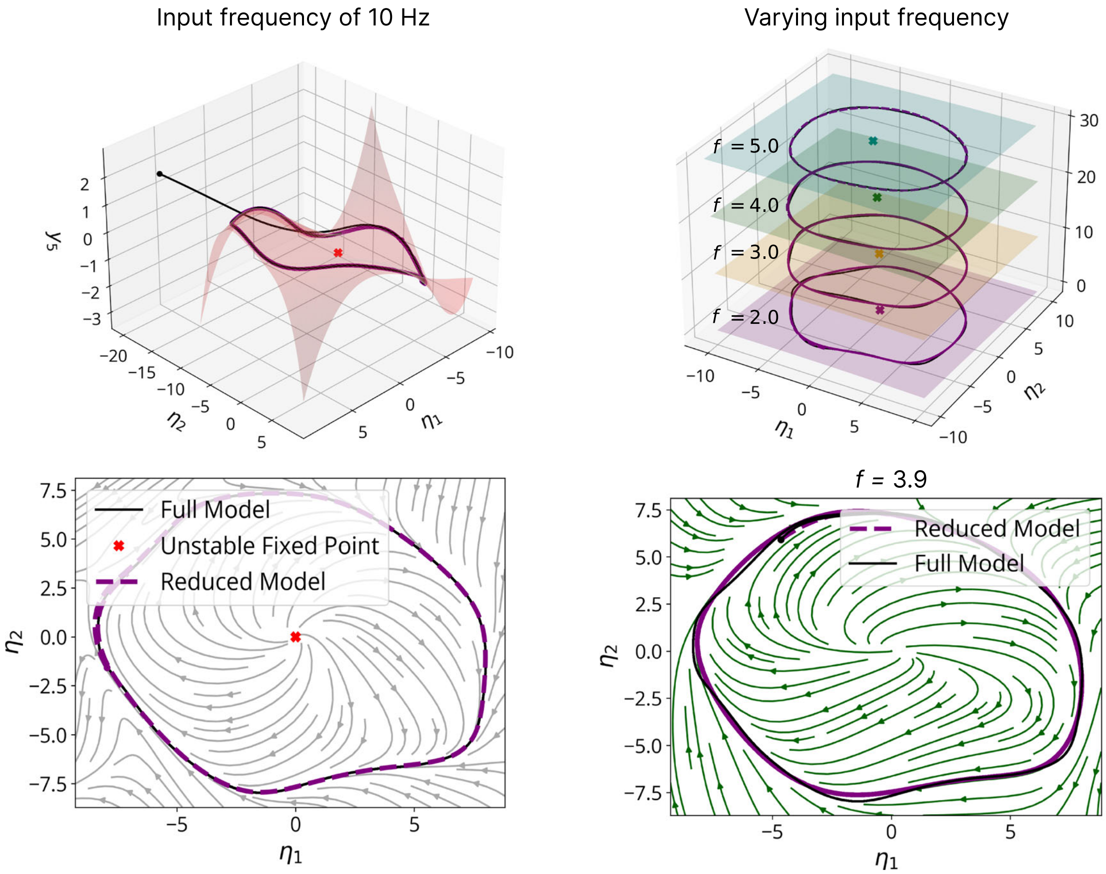

Fig. 4 | SSM carrying the core RNN dynamics and parameter-dependent SSMs with their reduced dynamics projected to the spectral subspaces. (Upper left) Unstable manifold at order three in coordinates (η_1, η_2, y_5), where (η_1, η_2) are the coordinates of the unstable subspace. The stable limit cycle on the unstable manifold is globally attracting: in the picture, one full and one reduced trajectory are converging to the stable limit cycle. The MFE (see Eq. (14)) and the NMTE (see Eq. (15)) calculated on test trajectories have magnitude ≈0.02 and ≈0.5, respectively. (Bottom left) Projections onto the unstable subspace of the streamlines of the right-hand side of the reduced-order model on the two-dimensional unstable manifold up to order six, with projected full and reduced trajectories. (Upper right) Two-dimensional unstable manifolds for different values of the parameter u (and corresponding output frequency values f) plotted in the three-dimensional space with coordinates (η_1, η_2, y_5). The manifolds are shifted on the y_5-axis for visualization purposes. (Bottom right) Projection onto the unstable subspace of the streamlines of the right-hand side of the reduced-order model on the two-dimensional unstable manifold up to order six for parameter u corresponding to a frequency of 3.9 Hz, with projected full and reduced trajectories.

This example illustrates how our reduction approach can uncover the specific solution that the training process of an RNN selects from many potential solutions to perform a given task. Rather than viewing RNNs as black box systems that replicate a given behavior, SSM-reduced modeling uncovers the low-dimensional, interpretable core dynamics (3) of RNNs. This in turn enables us to identify a very specific dynamical structure, a limit cycle, that is directly responsible for producing the output that the RNN was trained to produce.

A model for a memory-pro task

As our third example, we now describe the SSM-based reduction of the multitasking RNN in ref. 18 performing a Memory-Pro Task described in the Method section "Multitasking RNN: memory-pro task". This RNN receives as input a fixation input u_1 with values in {0, 1}, and two 2D angular sensory inputs $u_2 = \cos\theta_1$, $u_3 = \sin\theta_1$, $u_4 = \cos\theta_2$, and $u_5 = \sin\theta_2$. In the setting considered here, the network has to respond according to the angle indicated by the first stimulus when the fixation input is off (u_1 = 0). This is expressed through a three-dimensional readout vector $\mathbf{z}$: $z_1 = u_1$ (fixation) and $z_2 = \cos\theta_1, z_3 = \sin\theta_1$.

The authors of[18] interpret the network dynamics underlying the memory period within the task through a 1D structure composed of slow points[13,38,40] located along a circle, which they call a *ring attractor*. In this structure, the position that a trajectory reaches during the memory period is interpreted as the persistent pattern that the network maintains over time, encoding the correct stimulus it has to respond to when the fixation input is off.

For this reason, we focus on the input parameter setting corresponding to the memory and context periods, when fixation and context inputs are on and sensory inputs are off. For this configuration, we seek to clarify what dynamical structure is behind what has been empirically labeled as a ring attractor. Such clarification is necessary, as the attractor topology closest to a ring would be that of a torus. There is, however, no observational evidence for sustained quasiperiodic behavior, which would be the hallmark of a toroidal attractor.

Our analysis reveals an unstable fixed point with two unstable, real eigenvalues and an asymptotically stable fixed point. Based on the spectrum of the linearized dynamics around the unstable fixed point, we conclude that the 2D SSM embodied by the unstable manifold of the unstable fixed point will carry the observable core dynamics once the transients have died out.

We find the unstable manifold using SSMLearn up to polynomial order three (see Fig. 5, upper left).

The reduced-order model on this SSM is of the form

$$\dot{\eta}_\ell = \sum_{i=0}^{3}\sum_{j=0}^{3-i} a_{ij}^{(\ell)}\,\eta_1^i\eta_2^j + \mathcal{O}(4), \qquad \ell = 1, 2. \tag{5}$$

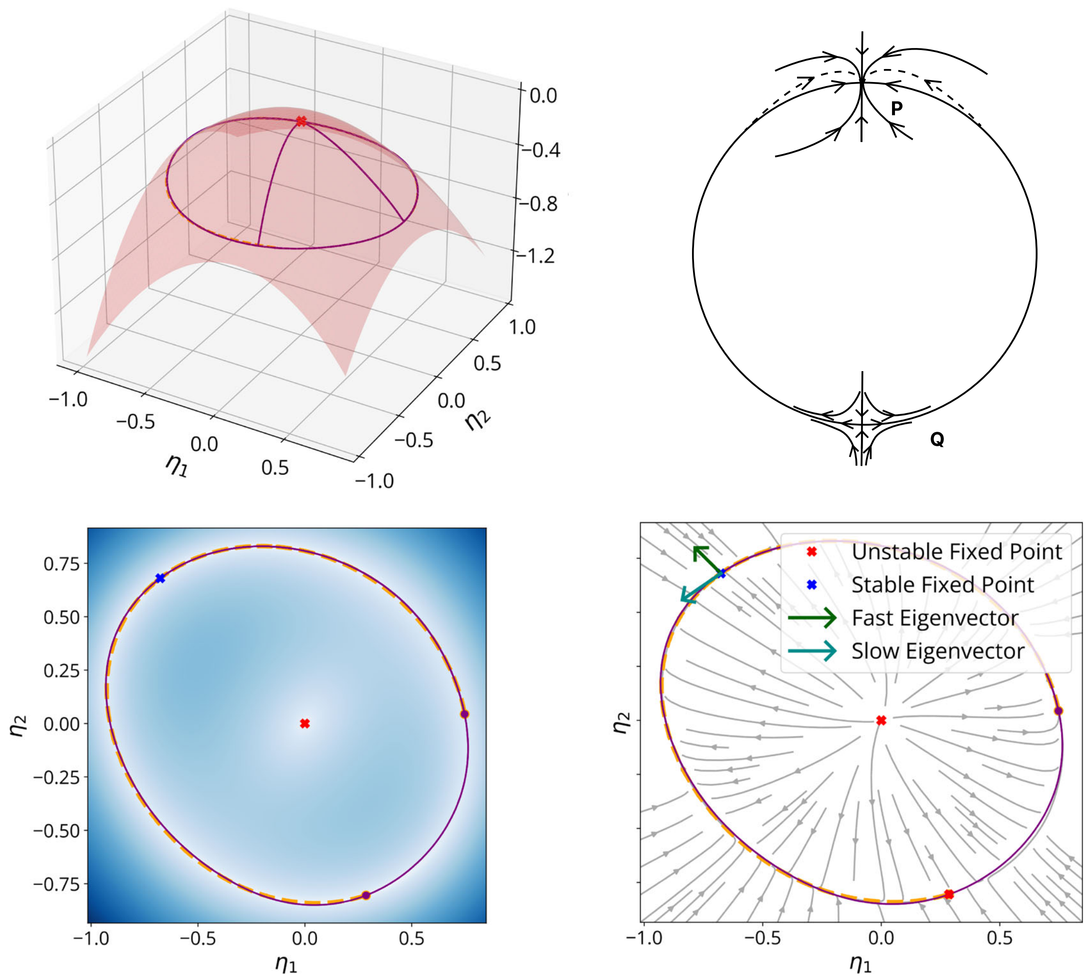

Fig. 5 | The heteroclinic orbit on the 2D SSM. (Upper left) The two-dimensional unstable manifold attached to the unstable fixed point carries the reduced dynamics. Trajectories (full model trajectories in purple and reduced model in orange) disposed on a circular curve converge to the stable fixed point in infinite time. (Bottom left) Norm of the right-hand side of the reduced model on the unstable manifold plotted on the unstable subspace with coordinates (η_1, η_2), with darker blues corresponding to larger norm values. The MFE (see Eq. (14)) and the NMTE (see Eq. (15)) calculated on test trajectories have magnitude 10^{-2} and 10^{-1}, respectively. (Bottom right) Heteroclinic orbits connecting the unstable fixed point on the bottom right and the stable fixed point on the upper left, projected onto the unstable subspace with coordinates (η_1, η_2). Streamlines of the right-hand side of the reduced model are projected onto the unstable subspace (in gray). The structure is normally hyperbolic, as we can see from the directions of the slow and the fast eigenvectors of the linearization around the stable fixed point of the full model. (Upper right) Pictorial representation of the normally hyperbolic heteroclinic structure, as an example of normally hyperbolic invariant manifold in ref. 45. The unstable manifold of the unstable fixed point Q coincides with the slow stable manifold of the stable fixed point P, and the rate of normal attraction to P is bigger than the rate of tangential compression.

Solving for the zeros of the right-hand side of (5), we find an additional unstable fixed point. This fixed point is connected to the stable fixed point on the SSM via a pair of heteroclinic orbits that form a structurally stable heteroclinic loop (see Fig. 5, bottom). The structural stability (i.e., smooth persistence under perturbations of the RNN) of this heteroclinic loop follows from the fact that it is normally hyperbolic, i.e., the attraction rates normal to it dominate the attraction rates inside it (refs. 45,46).

This strict result from the SSM-reduced model (5) clarifies that the true phase space structure behind the empirically observed slow circular structure (ring attractor) is an attracting, 1D invariant manifold. Diffeomorphic to a circle, this manifold is formed by a structurally stable heteroclinic loop between two fixed points inside a 2D slow SSM.

The structural stability of normally hyperbolic heteroclinic loops explains the ubiquitous presence of slow, circular attractors in RNNs performing working memory tasks (see refs. 7–10). If, in fact, working memory corresponds to the maintenance of persistent neuronal firing patterns, represented in our framework as fixed points in phase space, it is tempting to conclude (see refs. 11,12,14–16,47–49) that encoding a continuous variable would require a continuum of such fixed points, each parameterized by a possible value of the variable.

A continuum of fixed points, however, is inherently non-robust and can be destroyed by small perturbations to the neuronal dynamics like, for example, the ones coming from the biological variability of the neurons, or small variations in experimental conditions. Instead, robust circular attractors that are not limit cycles are necessarily formed by chains of heteroclinic connections among fixed points. Here, we have observed the simplest such chain, one that is formed just by two heteroclinic orbits.

Discussion

We have shown how analyzing RNNs based on recent results from nonlinear dynamical systems theory reveals the core dynamics of those networks via considerably lower-dimensional, polynomial models on attracting SSMs. Our results firm up previous, mostly heuristic interpretations of RNN dynamics. Specifically, the low dimension of the resulting SSM-reduced models allows us to develop a detailed mathematical understanding of the global RNN behavior. We describe these dynamics in terms of attracting invariant manifolds, the invariant sets they carry, and the transitions among those sets.

In particular, we have obtained a one-dimensional (1D) parameter-dependent model on a *slow manifold* for the 100D, context-dependent decision-making RNN in ref. 13. In a similar, data-driven fashion, we have constructed 2D models on slow SSMs for the 100D sine wave generator RNN in ref. 43, and the 256D multitasking RNN in ref. 18 performing a memory-pro task. As opposed to slow manifolds arising in geometric singular perturbation studies (see ref. 50), SSMs exist without the assumption of a global time scale separation that divides the phase space into slow and fast variables. Therefore, SSMs exist in a more general setting in which the core dynamics settle on a low-dimensional manifold once the faster transient dynamics have died out.

Specifically, for the RNN performing the context-dependent discrimination task, we have explained the "line attractor" observed by ref. 13 as a robust, parameter-dependent 1D slow manifold. For every configuration of the input parameters, we have identified the corresponding 1D slow manifold (as a section of the parameter-dependent slow manifold in the extended phase space) that captures the reduced dynamics of the full RNN. The resulting behavior following sensory integration arises from the dynamics on a 1D unstable manifold, along which the system converges to the stable fixed point associated with the correct choice.

We have also explained the flexibility of the decision-making process in terms of parameter-dependent SSMs and reduced models. In particular, context and sensory inputs determine which of the two fixed points on the 1D unstable manifold has the larger domain of attraction.

For the sine-wave generator RNN in ref. 43, we have uncovered the dynamics responsible for the input-dependent sinusoidal readout with a simple 2D model on a 2D unstable manifold carrying an attracting periodic orbit. This explains how the frequency of this orbit varies with the scalar input through a parameter-dependent model on the parameter-dependent unstable manifold.

Finally, we have found that the empirically described "ring attractor" in the memory-pro task in ref. 18 is a normally hyperbolic invariant slow manifold that coincides with two heteroclinic orbits connecting a stable and an unstable fixed point on a 2D SSM. The stable fixed point on the heteroclinic circle attracts nearby trajectories, and the heteroclinic structure carries slow, converging dynamics.

The relationship between RNNs trained to perform specific readout functions and the actual computations carried out by neurons is unclear, as many solutions are available for a high-dimensional system to produce a low-dimensional readout. However, an accurate description of the RNNs' phase space geometry enables the generation and verification of hypotheses on the dynamical systems underlying the observed dynamics in the data (see ref. 32,51). Interestingly, in the case of the 100D network trained for a binary choice task (context-dependent decision-making), the solution is a simple 1D bistable system. Similarly, we have found that a limit cycle with input-dependent frequency is responsible for the input-tuned oscillations of the sine-wave generator network. Moreover, the heteroclinic structure explaining the memory-pro task is a robust manifold that appears to be the common mechanism shared by all RNNs trained to reproduce short-term memory mechanisms[7,8].

The data-driven SSM-reduction approach used here is general enough to be applied to RNNs performing more complex tasks, whether or not they are related to neural computations. Indeed, RNNs are high-dimensional, nonlinear dynamical systems, whose use extends beyond neuroscience to describe any system with rich dynamics. We rely on simulations of RNN trajectories to find the reduced models of their dynamics through SSMLearn[39]. This is a fully data-driven algorithm that is applicable to systems whose equations of motion are either unknown or do not display any particular symmetry. Our only assumption is that the RNN has at least one fixed point whose location is at least approximately known. Importantly, the dimension of the SSM-reduced model (typically one or two) is independent of the RNN dimension for generic parameter choices. Therefore, even for more complex RNNs not considered in this study, the potential for dimensionality reduction using SSMs is virtually unlimited. For RNNs used as models for neural population dynamics, the key advantage of model reduction is interpretability: it provides insight into the structural organization of the learned dynamics. Reduced models help identify abstract mechanisms, namely, the most parsimonious low-dimensional description of the dynamics that still captures the underlying computation[14]. This perspective is becoming relevant in emerging areas such as computational psychiatry[52], where such reduced descriptions may have practical diagnostic value. For RNNs used in broader contexts (e.g., control circuits), reduced-order models can serve as less expensive surrogates to reduce computational cost and enable faster simulations.

The SSM theory was originally developed for model reduction in mechanical engineering, to reduce the dimensionality of high-dimensional mechanical systems, first in an equation-driven fashion[53], and later extended to fully data-driven problems[39]. However, neural systems are inherently different from mechanical systems, which generally have well-understood governing equations that are consistent with observations. The existence of a steady-state solution, a key assumption in SSM reduction, is well established for most mechanical systems. When modeling biological networks, instead, we can only make assumptions on the existence of (unstable) steady-state solutions to attach SSMs to. For this reason, the application of SSM theory to neural data is promising but not straightforward. Recent extensions of SSM theory cover forced systems[54] and random dynamical systems with small, uniformly bounded noise[55]. This setting[56] preserves a nonlinear deterministic core while explicitly accounting for biologically plausible fluctuations. The special solutions we can attach SSMs to are then not exact steady states. Instead, they become random, time-dependent anchor states arising from uniformly bounded perturbations to the exact steady states of the autonomous dynamics. Therefore, a starting point for SSM reduction carried out directly on experimental neural data will require a feasible model of noise in neuronal dynamics.

Methods

The theory of spectral submanifolds

To identify robust, smooth, and unique invariant manifolds in high-dimensional Vanilla RNNs, we employ the theory of SSMs, first formulated by Haller and Ponsioen in ref. 28 to find invariant manifolds in high-dimensional nonlinear systems around an attracting steady-state solution. The reader can consult[29] for a complete and detailed introduction to the theory and its applications.

SSM theory is based on a fundamental observation about linear dynamical systems of the form

$$\dot{x} = Ax, \quad x \in \mathbb{R}^N, \quad A \in \mathbb{R}^{N\times N}. \tag{6}$$

In this setting, the eigenspaces of A, as well as their algebraic direct sums, are invariant under the flow map and therefore constitute linear and smooth invariant manifolds of the system. We refer to any subspace of the form $E = E_1 \oplus \cdots \oplus E_k$, $k \le n$, where each E_j is the eigenspace associated with a particular eigenvalue λ_j, as a *spectral subspace*. Important examples include the stable, center, and unstable subspaces, spanned respectively by the generalized eigenvectors corresponding to eigenvalues with negative, zero, and positive real parts.

From the eigendecomposition of A, we can always find solutions to the linear equations (6) and gather them together to construct infinitely many invariant manifolds tangent to each spectral subspace at the fixed point at the origin. Among these, the spectral subspaces are the smoothest (analytic), whereas all others have limited smoothness governed by the ratios of the real parts of the eigenvalues of A. One can easily check this fact by looking at the degree of smoothness at zero of trajectories of a linear system that do not coincide with the eigenspaces.

SSM theory exploits the structure of linear systems and uses linearization theorems to infer the existence, smoothness, and uniqueness of invariant manifolds, SSMs, in nonlinear systems around a fixed point. Although the existence of these manifolds is obtained from linearization, they extend over the domain of linearizability of the system and can carry nonlinear reduced dynamics. We look for SSMs as smooth invariant hypersurfaces tangent to the spectral subspaces of the linearized dynamics, which, in the general case, are not invariant manifolds of the system themselves and hence not suited for model reduction (see ref. 57 for further discussion).

In particular, consider the nonlinear autonomous system:

$$\dot{x} = Ax + f_0(x), \quad x \in U \subset \mathbb{R}^N, \tag{7}$$

where $f_0 \in C^r$ for some $r \in \mathbb{N} \cup \{\infty, a\}$ (where C^a indicates that the function is analytic) and satisfies $f_0(x) = \mathcal{O}(\| x\|^2)$ as $x \to 0$. The fixed point at the origin is assumed to be hyperbolic and attracting, meaning that all eigenvalues of A have strictly negative real parts.

Given a spectral subspace E for the linearized dynamics, we define its *spectral quotient* as the integer ratio

$$\sigma(E) = \mathrm{Int}\left[\frac{\max_{\lambda_k \notin \mathrm{spect}(A|_E)} \mathrm{Re}\,\lambda_k}{\min_{\lambda_i \in \mathrm{spect}(A|_E)} \mathrm{Re}\,\lambda_i}\right], \tag{8}$$

which compares the fastest decay rate outside E to the slowest decay rate within E. We can then define a *external nonresonance condition* for E (see ref. 28): no low-order (with coefficients m_j such that $2 \le \sum_{i=1}^{k} m_k \le \sigma(E)$) integer linear combination of eigenvalues $\sum_{j=1}^{k} m_k \lambda_k$ within E should coincide with an eigenvalue outside of E.

Under the above assumptions, for a nonresonant spectral subspace E of the linearized dynamics of system (7) there exists an invariant manifold, $\mathcal{W}_E(0)$, tangent to E at the origin. This manifold, the SSM associated with E, is the smoothest among all invariant manifolds tangent to E, and unique in the smoothest class $C^{\sigma(E)+1}$. Furthermore, $\mathcal{W}_E(0)$ inherits smooth dependence on coordinates and parameters from the right-hand side of (7).

In cases where the fixed point is unstable but has a nonempty stable manifold, a similar result holds, provided stronger nonresonance conditions are satisfied, specifically, that no eigenvalue of A is a nonnegative integer linear combination (with $m \ge 2$) of all eigenvalues. Under these assumptions, one still obtains an invariant SSM, $\mathcal{W}_E(0)$, that is unique in its smoothness class and tangent to the spectral subspace, as shown in ref. 58.

In summary, attached to a hyperbolic fixed point with a nonresonant spectral subspace of the linearized dynamics, there is a unique invariant manifold, the SSM, distinguished by its highest smoothness. Since the SSM $\mathcal{W}_E(0)$ is tangent to its underlying spectral subspace E at the fixed point, it is convenient to locally represent it as a graph over that spectral subspace.

As an example, let $E = \mathrm{span}\{E_{j_1}, E_{j_2}\}$ be a 2D spectral subspace for A in (7), with $f_0 \in C^a$ (analytic), for which the external nonresonance conditions are met. We assume that we have chosen a basis such that E is parameterized by the reduced coordinates (η_1, η_2), and the full coordinate system is given by $x = (\eta_1, \eta_2, y_1, ..., y_{N-2})$. We look for the Taylor expansion of the SSM, $\mathcal{W}_E(0)$, expressed as a graph over E

$$\begin{pmatrix} y_1 \\ \vdots \\ y_{N-2} \end{pmatrix} = h(\eta_1, \eta_2) = \begin{pmatrix} h_1(\eta_1, \eta_2) \\ \vdots \\ h_{N-2}(\eta_1, \eta_2) \end{pmatrix} = \begin{pmatrix} \sum_{d=2}^{\infty} \sum_{j=0}^{d} a_{1,d,j} \eta_1^{d-j} \eta_2^{j} \\ \vdots \\ \sum_{d=2}^{\infty} \sum_{j=0}^{d} a_{N-2,d,j} \eta_1^{d-j} \eta_2^{j} \end{pmatrix} \tag{9}$$

Since $\mathcal{W}_E(0)$ is invariant, we can write the *invariance PDE*

$$\begin{aligned} \dot{x}|_{\mathcal{W}_E(0)} &= A \begin{pmatrix} \eta_1 \\ \eta_2 \\ h_1(\eta_1, \eta_2) \\ \vdots \\ h_{N-2}(\eta_1, \eta_2) \end{pmatrix} + f_0((\eta_1\, \eta_2\, h(\eta_1, \eta_2))) \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ \partial_{\eta_1} h_1(\eta_1, \eta_2) & \partial_{\eta_2} h_1(\eta_1, \eta_2) \\ \vdots & \vdots \\ \partial_{\eta_1} h_{N-2}(\eta_1, \eta_2) & \partial_{\eta_2} h_{N-2}(\eta_1, \eta_2) \end{pmatrix} \begin{pmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \end{pmatrix} \end{aligned} \tag{10}$$

with boundary conditions at zero $h(0) = Dh(0) = 0$, where $Dh(0)$ is the differential of h at 0 with respect to (η_1, η_2). Substitution of the Taylor expansion (9) into (10) yields a family of linear equations for the Taylor coefficients $a_{i,j}$, which can be solved recursively under the asumed nonresonance conditions.

If the right-hand-side of (7) is not analytic but C^r, $r \in \mathbb{N} \cup \{\infty\}$, increasing the order of the expansion will not necessarily improve the accuracy of the approximation, even on smaller domains. If f_0 is C^r with r finite, its expansion would not exist to orders higher than r. One should choose accurately the order to truncate the expansion, taking care not to exceed the order at which the approximation ceases to improve on the desired domain.

This theoretical basis for explicitly constructing invariant manifolds over arbitrary non-resonant spectral subspaces of the linearized system can be conveniently used for the reduction of high-dimensional nonlinear systems like RNNs. In particular, as a model for the slow, dominant dynamics, we can choose a *slow spectral subspaceE* such that $E = \{E_1 \oplus E_2 \oplus \cdots \oplus E_q\}$ is the direct sum of the q slowest eigenspaces whose corresponding eigenvalues have the largest real parts (they represent the lowest decaying dynamics in the linearized system if the fixed point is stable but also growing dynamics if the fixed point is unstable).

We point out that the dimensional parameter q in the choice of the slow spectral subspace E is a free parameter. It can be chosen so that the slow SSM becomes the low-dimensional attractor that prevails in the available data. For this purpose, we can examine the largest spectral gaps (difference in the real parts) between consecutive (ordered) eigenvalues and pick a slow spectral subspace that is separated from the others on a time scale suitable for the system under study, or simply the slow subspace corresponding to the largest gap.

Back to our example, once a functional expression for the SSM is obtained, we can find the reduced dynamics on the SSM by substituting the y variables with $y|_{\mathcal{W}_E(0)} = h(\eta_1, \eta_2)$

$$\begin{pmatrix} \dot{\eta}_1 \\ \dot{\eta}_2 \end{pmatrix} = \begin{pmatrix} f_{0,1}(\eta_1, \eta_2, h_1(\eta_1, \eta_2), \ldots, h_{N-2}(\eta_1, \eta_2)) \\ f_{0,2}(\eta_1, \eta_2, h_1(\eta_1, \eta_2), \ldots, h_{N-2}(\eta_1, \eta_2)) \end{pmatrix} \tag{11}$$

thus obtaining an explicit polynomial expansion for the 2D reduced-order model on the manifold. Since we have enslaved all the other dynamical variables to η_1 and η_2 through the SSM equation (9), we can retrieve the dynamics of the full system by inserting the solutions of the reduced equations (11) into the SSM equation.

The latest version of the algorithm that performs these equation-driven calculations is available in the open source MATLAB live script package SSMTool https://github.com/haller-group/SSMTool-2.4(see ref. 59).

Parameter-dependent and time-dependent models on SSMs. To understand the input-output relationship in Vanilla RNNs that resemble the sensory input-behavioral output connections in the animal tasks, we must consider how parameter changes affect the dynamics of the RNNs and their SSMs.

In the absence of noise or other types of explicit time dependence (which we may add later as perturbations), the input vector collects the parameters on which the system depends (see the discussion in section “RNN for context-dependent decision-making”). Although altering sensory inputs and context inputs (see refs. 13,40,41) may affect the geometry of the phase space differently, both inputs keep the RNN an autonomous dynamical system.

Specifically, varying the additive parameters in the input vector can modify the phase space geometry by shifting the fixed point locations and altering the sizes of their domains of attraction. Additionally, when parameter values are close to *bifurcation values* (see ref. 44), this can lead to changes in the stability of the fixed points or to their disappearance.

If the dependence on the parameters of the right-hand side of the equations of motion is smooth, their SSMs will also display smooth dependency on parameters[28,29]. However, whenever a fixed point disappears due to a bifurcation, we can no longer rely on the existence of its SSMs as invariant manifolds for the system.

In our examples, fixed points generally do not undergo bifurcations that lead to their disappearance, except in the context of the decision-making RNN when the relevant sensory inputs are varied. In that case, though, the reduced dynamics are carried by a Normally Hyperbolic Invariant Manifold (NHIM)[45,46]. Therefore, the NHIMs theory and the asymptotic properties, uniqueness, and smoothness of unstable manifolds guarantee the persistence of a 1D invariant slow manifold for the whole parameter range considered, even if the fixed points undergo a cascade of saddle-node bifurcations (for details, see the Supplementary Methods SM1).

In all the other cases, we can rely on the smooth dependence of SSMs on parameters to find parameter-dependent reduced-order models for the RNN dynamics. In fact, we can extend the phase space of parameter-dependent dynamical systems $\dot{x} = f(x; \mu)$, with $\mu \in \mathbb{R}^p$ the parameter vector, by treating the parameters as dummy variables for the system, which have trivial dynamics $\dot{\mu} = 0$. We can therefore find polynomial parameter-dependent expansions for SSMs and reduced-order models by solving the invariance equations in the extended phase space, including the parameter directions in the spectral subspaces.

We can also use a strategy to exploit the smooth dependence of SSMs on parameters to obtain a reduced-order model of the transition to chaotic dynamics. The key requirement for this strategy to work is the existence of a persistent, unstable fixed point in the phase space of the system that contains the bifurcating attractor. The unstable fixed point has an SSM that coincides with its unstable manifold, and mixed-mode SSMs consisting of the unstable manifold and slow, stable SSMs of different dimensions. Based on the Hausdorff dimension of the chaotic attractor, we can choose a high-dimensional enough, parameter-dependent SSM attached to the unstable fixed point. This SSM will carry the bifurcating attracting set transitioning to a chaotic attractor; the parameter-dependent reduced model on the SSM will account for the transition to the chaotic dynamics. See ref. 60 for examples of SSM-reduced models of chaotic systems with an unstable fixed point in their phase space. Similarly, the existence of an unstable steady state solution, the edge state, is used in ref. 61 to describe the transition to turbulence in periodic pipe flows.

On the other hand, to examine the impact of bounded noise on the dynamics of RNNs, we apply an extension of SSM theory to non-autonomous systems subjected to weak aperiodic forcing, as discussed in ref. 54. We treat uniformly bounded noise as a form of discontinuous-in-time, weak forcing and derive time-dependent SSMs and reduced-order models that capture the noisy dynamics (refer to Supplementary Methods SM3 for the general discussion on non-autonomous SSMs). These models will maintain the same phase space geometry as the autonomous system at any given moment, but they will evolve over time due to the influence of noisy perturbations.

In particular, fixed points for the autonomous system perturb, in the noisy setting, into noisy *anchor trajectories* with the same stability properties. We can think of such a trajectory as the perturbation of the line that would be present in the extended phase space with coordinates $X = (x, t) \in \mathbb{R}^{N+1}$ with dynamics $\dot{X} = (\dot{x}, 1)$, $L = \{(x, t) \in \mathbb{R}^{N+1} \,|\, x = 0,\ t \geq 0\}$.

Any autonomous SSM perturbs into a time-dependent SSM that evolves over time along the anchor trajectory perturbing from the fixed point the autonomous SSM was attached to. The time-dependent SSM can be seen as the result of perturbing with additive, small noise terms the autonomous SSM in the extended phase space $\mathbb{R}^{N+1}$ with coordinates (y, t). A more detailed discussion and the calculations specific for the noisy RNN systems can be found in the Supplementary Methods SM3.

Data-driven computations of SSMs. As we have seen, SSMs can be found locally as invariant hypersurfaces over their tangent spaces at the fixed point and approximated through Taylor expansions. This makes it possible to use a simple data-driven algorithm to obtain low-dimensional, predictive models for high-dimensional, nonlinear systems for which the equations of motions are not known (see ref. 39). The data-driven algorithm that finds SSMs and reduced models on them, SSMLearn, is open-source and available at https://github.com/haller-group/SSMLearn.

To ensure the correct use of the SSMLearn package, several online resources are available for practical guidance. In particular, the reader can find fully dedicated tutorial slides at https://github.com/haller-group/SSMLearn/blob/main/docs/SSMLearn_tutorial.pdfand practical considerations slides at https://github.com/haller-group/SSMLearn/blob/main/docs/Practical_considerations.pdf. These serve as support for the worked-out examples in the SSMLearn repository and for the book that summarizes all SSM theory and related applications[29].

In this study, we combine the equation and data-driven approaches, since we use simulated RNN trajectories to find optimal polynomial SSMs and reduced-order model approximations through SSMLearn.

More specifically, SSMLearn is an algorithm designed to find the optimal polynomial coefficient matrix H for the SSM equations

$$y_{SSM} = V_E V_E^T y + H\phi_{d,2:M}(V_E^T y); \tag{12}$$

where d is the dimension of the SSM and M is the order of approximation of the polynomial expansion. The function $\phi_{d,2:M}(q)$ outputs the vector of all m_M monomials from order two up to order M given a d-vector q, H is the $N \times m_M$ coefficient matrix, V_E^T is the projection matrix to the spectral subspace, $\eta := V_e^T y$ is the d-dimensional coordinate vector projected to the spectral subspace.

The optimization problem for H, and possibly the $N \times d$-matrix V_e, is formulated as minimizing a quadratic cost function that measures the distance between train trajectories and the SSM. When an estimate of the spectral subspace is already available (V_e is known), the optimization problem is convex.

It is then possible to compute the polynomial expansion for the reduced equations of motions $\dot{\eta}$, $\eta = V_e^T y$. In particular, SSMLearn finds the optimal polynomial coefficient matrix W_r for the equation

$$\dot{\eta} = W_r \phi(\eta)_{d,1:M_r} \tag{13}$$

minimizing the distance between trajectories $\eta(t)$ and the train trajectories projected to the spectral subspace. Finally, full-order trajectories can be obtained by lifting the reduced-order trajectories through the SSM equations.

There is currently no automated pipeline for selecting the order of the Taylor expansion approximating the SSM or the reduced dynamics. For analytic SSMs, within the radius of convergence of the Taylor expansion around the fixed point, higher-order approximations should yield better local results. Since, however, we generally don't know the convergence domain, we stop at the lowest order of the expansion that captures the core dynamics. We usually evaluate the approximations through the mean-square distance with test trajectories: we use the manifold fitting error (MFE) for the SSM

$$\mathrm{MFE} = \frac{1}{N_{\mathrm{test}}} \sum_{i=1}^{N_{\mathrm{test}}} \frac{1}{N_{\mathrm{timepoints}}} \sum_{j=1}^{N_{\mathrm{timepoints}}} ||y_j^i - y_{SSM,j}^i|| \frac{1}{\max_j y_j^i} \tag{14}$$

where y_{SSM} is calculated according to (12) and the normalized mean trajectory error (NMTE) for the reduced dynamics

$$\mathrm{NMTE} = \frac{1}{N_{\mathrm{test}}} \sum_{i=1}^{N_{\mathrm{test}}} \frac{1}{N_{\mathrm{timepoints}}} \sum_{j=1}^{N_{\mathrm{timepoints}}} ||y_j^i - y_{\mathrm{pred},j}^i|| \frac{1}{\max_j y_j^i} \tag{15}$$

where $y_{\mathrm{pred}} = V_E \eta + H\phi_{d,2:M}(\eta)$, and ηs are solutions of (13). The MFE measures the mean distance between the lifts of reduced test trajectories and the SSM and the actual test trajectories. If there were no noise and the SSM model were exact, this distance should be zero. Similarly, the NMTE measures the distance between the reduced model's predictions, given test initial conditions, and the actual test trajectories.

The Vanilla RNN models

The RNNs described in refs. 18,[43] are trained to obtain specific readout dynamics associated with behavioral experiments. These RNNs are designed to produce the correct output values based on a given set of input parameters, enabling them to replicate the target behaviors exhibited by trained animals. In this context, the optimal RNN parameters are identified to accurately reproduce a N_o-dimensional readout function $\mathbf{z}$ that depends on the specific task.

The network units are coupled through a connectivity matrix $\mathbf{W}$ and have nonlinear dynamics due to a bounded, smooth, nonlinear activation function. They have equations

$$\begin{aligned} &\tau\dot{\mathbf{x}} = -\mathbf{x} + \mathbf{W}\,\mathbf{r}(\mathbf{x}) + \mathbf{B}\,\mathbf{u}(t) + \boldsymbol{\sigma}(t), \\ &\mathbf{x} \in \mathbb{R}^N,\ N = 100,\ \mathbf{r}(\mathbf{x}) = (\tanh(x_1), \ldots \tanh(x_N))^T \in \mathbb{R}^N, \\ &\mathbf{W} \in \mathbb{R}^{N\times N},\ \mathbf{B} \in \mathbb{R}^{N\times N_i},\ \mathbf{u} \in \mathbb{R}^{N_i} \end{aligned} \tag{16}$$

where $\mathbf{u}$ is the input vector of dimension N_i and $\boldsymbol{\sigma} \approx A\,\mathcal{N}(0,1) \in \mathbb{R}^N$ is a vector with entries drawn at each time step from a Normal distribution with zero mean and unit standard deviation. In our treatment, we will reject drawings outside ± 3 standard deviations from the mean to make the Gaussian noise bounded.

The matrices $\mathbf{W}$, $\mathbf{B}$ and $\mathbf{Y} \in \mathbb{R}^{N_o\times N}$ (N_o is the readout vector dimension) are selected by training to produce a read-out function $z(t) = \langle \mathbf{Y}, \mathbf{r}(\mathbf{x}(t))\rangle$ with specific dynamics adapted to each task.

We first remove the noisy components from equations (16) and choose constant input vectors (which, in the experiments in refs. 13,18,43, are only varied among trials). We then vary the input parameters to obtain manifolds and reduced ODEs that explicitly depend on them and describe the parametric changes to the geometry of the phase space. Finally, we insert noise back into the equation to obtain a noisy, time-dependent model.

RNN for context-dependent decision-making. We now describe the specific features of the Vanilla RNN described in Eq. (16) that adapt it to the context-dependent decision-making task of Mante et al. in ref. 13.

In the experiment reported in ref. 13, monkeys were trained to distinguish between the direction of motion (left or right) or the color (red or green) of a random-dot display based on the given context cue. Meanwhile, extracellular responses from the monkey's neurons were recorded in and around the frontal eye field and analyzed in the space of the first three principal components.

At each trial, neurons in the RNN receive two independent sensory inputs, each with a positive or negative value, that mimic evidence for motion and color in a single random dot stimulus, and the context inputs indicate which of the two sensory inputs will determine the sign of the final scalar readout $z \in \mathbb{R}$ ($N_o = 1$).

The 100D network evolves according to Eq. (16), with a 4D ($N_i = 4$) input vector $\mathbf{u} = (u_1 u_2 u_3 u_4)^T$ that accounts for both context and sensory stimuli in the experiment. In particular, its first two entries $u_1 = s_1$ and $u_2 = s_2$ (represent sensory stimuli (color and direction of motion, respectively)). The network was trained on the set of values $\{-0.15, -0.036, -0.009, 0.009, 0.036, 0.15\}$. The last two entries in the input vector u_3 and u_4 are one-hot-encoded and represent context 1 when $u_3 = c_1 = 1$ and $u_4 = c_2 = 0$, and context 2 when $u_4 = c_2 = 1$ and $u_3 = c_1 = 0$.

The input vector $\mathbf{u}$ is piecewise constant during one trial: in the first part of the trial, before the burnlength time T_B, the sensory inputs are switched off $s_1 = s_2 = 0$ and only one of the context inputs is on. During this period, the target readout $z^*(t) = const$. Instead, when $t > T_B$, the sensory inputs are switched on $u_1 = s_1$, $u_2 = s_2$ and the sensory integration process begins. The target readout is a scalar function converging to either $+1$ or -1, depending on the selected sensory input.

Sine-wave generator RNN. The sine-wave generator network is a 100D Vanilla RNN with equations of motions of the same structure as in Eq. (16). The behavior implemented in this case is not linked to a specific experiment but rather a general model of oscillatory responses in a neuronal network. It is relevant to the study of the primary motor cortex in primates, where a high-dimensional network of interconnected neurons appears to generate oscillatory dynamics at different frequencies during movement onset and muscle contraction (see ref. 43).

The input u is, in this case, a scalar ($N_i = 1$), whose values represent the desired oscillation frequencies. The network is trained so that the scalar readout $z(t) = \langle \mathbf{Y}, \mathbf{r}(\mathbf{x}(t))\rangle$ evolves in time as a sine-wave with the frequency indicated by the scalar input.

Multitasking RNN: memory-pro task. The multitasking RNN implemented in ref. 18 is

$$\tau\dot{\mathbf{x}} = -\mathbf{x} + \mathbf{r}(\mathbf{W}_{rec}\mathbf{x} + \mathbf{W}_{in}\mathbf{u} + \mathbf{B}_{rec}), \tag{17}$$

a variant of the Vanilla RNNs described in Eq. (16) with $\mathbf{x} \in \mathbb{R}^{256}$. The matrices $\mathbf{W}_{rec}$, $\mathbf{W}_{in}$, $\mathbf{B}_{rec}$ are trained so that the RNN can perform multiple tasks, depending on the values of the 20D input vector $\mathbf{u}$. The one-hot-encoded last fifteen entries u_i, $i = 6, \ldots, 20$ indicate which task the network has to perform in each trial. The first entry of the input vector plays the role of a fixation input, taking values 0 or 1 depending on whether the network, similarly to animals in an experiment, has to respond to sensory inputs or not. Finally, the two 2D angular sensory inputs are represented in the entries $u_2 = \cos\theta_1$, $u_3 = \sin\theta_1$, $u_4 = \cos\theta_2$, and $u_5 = \sin\theta_2$.

When the fixation input is off ($u_1 = 0$), the network has to respond according to the angle indicated by the stimulus that is relevant to the current task. The three-dimensional readout vector $\mathbf{z}$ ($N_o = 3$) encodes whether the network is fixating or not in its first entry $z_1 = 1$ or $z_1 = 0$, and for the output angle through the values of $z_2 = \cos\theta_{out}$ and $z_3 = \sin\theta_{out}$.

In a Memory-Pro task, the RNN is exposed to piecewise constant changes in the input vector. In the context period $[t_0, t_s]$, the sensory inputs are off, while the fixation and context inputs are on. Then, in the stimulus period t_s, t_m, sensory inputs are switched on, while all the other inputs remain constant. Sensory inputs are switched off again in the memory period $[t_m, t_r]$, after which the fixation stimulus is switched off, and the network has to output the correct angle θ_{out}.

This task is designed to understand possible dynamical motifs that underlie working memory in animals. In our analysis, we will focus on the configuration during context and memory periods, with context and fixation input on and sensory inputs off.

Data availability

The data is available at this URL https://github.com/alicemarr/RNN-paper under the MIT license. https://doi.org/10.5281/zenodo.20623670.

Code availability

The code is available at this URL https://github.com/alicemarr/RNN-paper under the MIT license.

References

1. Vyas, S., Golub, M. D., Sussillo, D. & Shenoy, K. V. Computation through neural population dynamics. *Annu. Rev. Neurosci.* **43**, 249–275 (2020).
2. Brody, C. D. & Hanks, T. D. Neural underpinnings of the evidence accumulator. *Curr. Opin. Neurobiol.* **37**, 149–157 (2016).
3. Machens, C. K., Romo, R. & Brody, C. D. Flexible control of mutual inhibition: a neural model of two-interval discrimination. *Science* **307**, 1121 – 1124 (2005).
4. Wang, X.-J. Decision making in recurrent neuronal circuits. *Neuron* **60**, 215–234 (2008).
5. Shenoy, K. V., Sahani, M. & Churchland, M. M. Cortical control of arm movements: a dynamical systems perspective. *Annu. Rev. Neurosci.* **36**, 337–59 (2013).
6. Churchland, M. M. et al. Neural population dynamics during reaching. *Nature* **487**, 51 – 56 (2012).
7. Ságodi, Á., Mart'in-S'anchez, G., Sok'ol, P. & Park, I. M. Back to the continuous attractor. *Proceedings of the 38th International Conference on Neural Information Processing Systems* (NeurIPS, 2024).
8. Beirán, M., Meirhaeghe, N., Sohn, H., Jazayeri, M. & Ostojic, S. Parametric control of flexible timing through low-dimensional neural manifolds. *Neuron* **111**, 739–753 (2021).
9. Chaudhuri, R. & Fiete, I. R. Computational principles of memory. *Nat. Neurosci.* **19**, 394–403 (2016).
10. Khona, M. & Fiete, I. R. Attractor and integrator networks in the brain. *Nat. Rev. Neurosci.* **23**, 744 – 766 (2021).
11. Romo, R., Brody, C. D., Hernández, A. & Lemus, L. Neuronal correlates of parametric working memory in the prefrontal cortex. *Nature* **399**, 470–473 (1999).
12. Renart, A., Song, P. & Wang, X.-J. Robust spatial working memory through homeostatic synaptic scaling in heterogeneous cortical networks. *Neuron* **38**, 473–485 (2003).
13. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *Nature* **503**, 78 – 84 (2013).
14. Vinograd, A., Nair, A., Kim, J., Linderman, S. W. & Anderson, D. J. Causal evidence of a line attractor encoding an affective state. *Nature* **634**, 910 – 918 (2024).
15. Liu, M., Nair, A., Coria, N., Linderman, S. W. & Anderson, D. J. Encoding of female mating dynamics by a hypothalamic line attractor. *Nature* **634**, 901 – 909 (2024).
16. Seung, H. S. & Hopfield, J. J. How the brain keeps the eyes still. *Proc. Natl. Acad. Sci. USA* **93 23**, 13339–44 (1996).
17. Riveland, R. & Pouget, A. Natural language instructions induce compositional generalization in networks of neurons. *Nat. Neurosci.* **27**, 988 – 999 (2024).
18. Driscoll, L. N., Shenoy, K. & Sussillo, D. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nat. Neurosci.* **27**, 1349 – 1363 (2024).
19. Yu, B. M. et al. Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity. *J. Neurophysiol.* **102**, 614–35 (2008).
20. Pandarinath, C. et al. Inferring single-trial neural population dynamics using sequential auto-encoders. *Nat. Methods* **15**, 805–815 (2017).
21. Schneider, S., Lee, J. H. & Mathis, M. W. Learnable latent embeddings for joint behavioural and neural analysis. *Nature* **617**, 360–368 (2022).
22. DePasquale, B., Brody, C. D. & Pillow, J. W. Neural population dynamics underlying evidence accumulation in multiple rat brain regions. *eLife* **13**, e84955 (2021).
23. Gosztolai, A., Peach, R. L., Arnaudon, A., Barahona, M. & Vandergheynst, P. Marble: interpretable representations of neural population dynamics using geometric deep learning. *Nat. Methods* **22**, 612–620 (2023).
24. Buesing, L., Macke, J. H. & Sahani, M. Learning stable, regularised latent models of neural population dynamics. *Netw. Comput. Neural Syst.* **23**, 24 – 47 (2012).
25. Linderman, S. W. et al. Bayesian learning and inference in recurrent switching linear dynamical systems. *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (PMLR **54**, 914–922, 2017).
26. Smith, J., Linderman, S. W. & Sussillo, D. Reverse engineering recurrent neural networks with Jacobian switching linear dynamical systems *Proceedings of the 34th International Conference on Neural Information Processing Systems* (NeurIPS, 2021).
27. Galgali, A. R., Sahani, M. & Mante, V. Residual dynamics resolves recurrent contributions to neural computation. *Nat. Neurosci.* **26**, 326–338 (2021).
28. Haller, G. & Ponsioen, S. Nonlinear normal modes and spectral submanifolds: existence, uniqueness and use in model reduction. *Nonlinear Dyn.* **86**, 1493–1534 (2016).
29. Haller, G. *Modeling Nonlinear Dynamics from Equations and Data - with Applications to Solids, Fluids, and Controls* (Society for Industrial and Applied Mathematics, 2025).
30. Maheswaranathan, N., Williams, A. H., Golub, M. D., Ganguli, S. & Sussillo, D. Universality and individuality in neural dynamics across large populations of recurrent networks. *Adv. Neural Inf. Process. Syst.* **32**, 15629 – 15641 (2019).
31. Kanwisher, N. G., Khosla, M. & Dobs, K. Using artificial neural networks to ask 'why' questions of minds and brains. *Trends Neurosci.* **46**, 240–254 (2023).
32. Mastrogiuseppe, F. & Ostojic, S. Linking connectivity, dynamics, and computations in low-rank recurrent neural networks. *Neuron* **99**, 609–623 (2017).
33. Laje, R. & Buonomano, D. V. Robust timing and motor patterns by taming chaos in recurrent neural networks. *Nat. Neurosci.* **16**, 925 – 933 (2012).
34. Hennequin, G., Vogels, T. P. & Gerstner, W. Optimal control of transient dynamics in balanced networks supports generation of complex movements. *Neuron* **82**, 1394–1406 (2014).
35. Sussillo, D., Churchland, M. M., Kaufman, M. T. & Shenoy, K. V. A neural network that finds a naturalistic solution for the production of muscle activity. *Nat. Neurosci.* **18**, 1025–1033 (2015).

36. Durstewitz, D., Koppe, G. & Thurm, M. I. Reconstructing computational system dynamics from neural data with recurrent neural networks. *Nat. Rev. Neurosci.* **24**, 693 – 710 (2023).
37. Song, H. F., Yang, G. R., Wang, X.-J. & Wang, X.-J. Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. *PLoS Comput. Biol.* **12**, e1004792 (2016).
38. Sussillo, D. & Barak, O. Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks. *Neural Comput.* **25**, 626–649 (2013).
39. Cenedese, M., Axås, J., Bäuerlein, B., Avila, K. & Haller, G. Data-driven modeling and prediction of non-linearizable dynamics via spectral submanifolds. *Nat. Commun.* **13**, 872 (2022).
40. Pagan, M., Valente, A., Ostojic, S. & Brody, C. D. Brief technical note on linearizing recurrent neural networks (rnns) before vs after the pointwise nonlinearity. *ArXiv Preprint* https://doi.org/10.48550/arXiv.2309.04030 (2023).
41. Pagan, M. et al. Individual variability of neural computations underlying flexible decisions. *Nature* **639**, 421 – 429 (2024).
42. Chadwick, A. et al. Learning shapes cortical dynamics to enhance integration of relevant sensory input. *Neuron* **111**, 106 – 120 (2021).
43. Krause, R., Cook, M., Kollmorgen, S., Mante, V. & Indiveri, G. Operative dimensions in unconstrained connectivity of recurrent neural networks. *Proceedings of the 35th International Conference on Neural Information Processing Systems*(NeurIPS, 2022).
44. Guckenheimer, J. M. & Holmes, P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields* (Springer New York, 1983).
45. Fenichel, N. Persistence and smoothness of invariant manifolds for flows. *Indiana Univ. Math. J.* **21**, 193–226 (1971).
46. Wiggins, S. *Normally Hyperbolic Invariant Manifolds in Dynamical Systems* (Springer Science & Business Media, 1994).
47. Zhang, K. Representation of spatial orientation by the intrinsic dynamics of the head-direction cell ensemble: a theory. *J. Neurosci.* **16**, 2112–2126 (1996).
48. Seelig, J. D. & Jayaraman, V. Neural dynamics for landmark orientation and angular path integration. *Nature* **521**, 186 – 191 (2015).
49. Stroud, J. P., Watanabe, K., Suzuki, T., Stokes, M. G. & Lengyel, M. Optimal information loading into working memory explains dynamic coding in the prefrontal cortex. *Proc. Natl. Acad. Sci. USA* **120**, e2307991120 (2021).
50. Fenichel, N. Geometric singular perturbation theory for ordinary differential equations. *J. Differ. Equ.* **31**, 53–98 (1979).
51. Langdon, C. & Engel, T. A. Latent circuit inference from heterogeneous neural responses during cognitive tasks. *Nat. Neurosci.* **28**, 665 – 675 (2022).
52. Murray, J. D., Demirtaş, M. & Anticevic, A. Biophysical modeling of large-scale brain dynamics and applications for computational psychiatry. *Biol. Psychiatry Cogn. Neurosci. Neuroimaging* **3**, 777–787 (2018).
53. Ponsioen, S., Pedergnana, T. & Haller, G. Automated computation of autonomous spectral submanifolds for nonlinear modal analysis. *J. Sound Vib.* **420**, 269–295 (2017).
54. Haller, G. & Kaundinya, R. S. Nonlinear model reduction to temporally aperiodic spectral submanifolds. *Chaos* **34**, 043152 (2024).
55. Xu, Z., Kaundinya, R., Jain, S. & Haller, G. Nonlinear model reduction to random spectral submanifolds in random vibrations. *J. Sound Vib.* **600**, 118923 (2025).
56. Arnold, L. *Random Dynamical Systems* (Springer Monographs in Mathematics, 1998).
57. Haller, G. & Kaszás, B. Data-driven linearization of dynamical systems. *Nonlinear Dyn.* **112**, 18639 – 18663 (2024).
58. Haller, G., Kaszas, B., Liu, A. & Axås, J. Nonlinear model reduction to fractional and mixed-mode spectral submanifolds. *Chaos* **33**, 063138 (2023).
59. Jain, S. & Haller, G. How to compute invariant manifolds and their reduced dynamics in high-dimensional finite element models. *Nonlinear Dyn.* **107**, 1417 – 1450 (2021).
60. Liu, A., Axås, J. & Haller, G. Data-driven modeling and forecasting of chaotic dynamics on inertial manifolds constructed as spectral submanifolds. *Chaos* **34**, 033140 (2023).
61. Kaszás, B. & Haller, G. Capturing the edge of chaos as a spectral submanifold in pipe flows. *J. Fluid Mech.* **979**, A48 (2023).

Acknowledgements
We are grateful to Laura N. Driscoll for sharing the code to simulate the multitask RNN.

Author contributions
A.M.: Formal analysis (lead); Methodology (lead); Software (lead); Visualization (lead); Writing—original draft (lead); Writing—review and editing (lead). R.K.: Data curation (supporting); Writing—review and editing (supporting). V.M.: Conceptualization (supporting); Supervision (supporting); Writing—review and editing (equal). G.H.: Conceptualization (lead); Methodology (equal); Project administration (lead); Supervision (lead); Writing—review and editing (equal).

Funding
Open access funding provided by Swiss Federal Institute of Technology Zurich.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-026-75924-7.

Correspondence and requests for materials should be addressed to A. Marraffa.

Peer review information *Nature Communications* thanks the anonymous reviewers for their contribution to the peer review of this work. A peer review file is available.