Dynamics on the manifold: Identifying computational dynamical activity from neural population recordings

Lea Duncker[1,2], **Maneesh Sahani**[1,†]

[1] Gatsby Computational Neuroscience Unit, University College London, London, UK

[2] Howard Hughes Medical Institute, Stanford University, Stanford, CA 94305, USA

† Correspondence should be addressed to: M.S. (maneesh@gatsby.ucl.ac.uk)

Highlights

- Neural activity is often confined to low-dimensional activity manifolds
- Dynamical systems can connect activity manifolds to underlying circuit computation
- We highlight data-analysis approaches to study computation through dynamics
- Future directions include linking data-analysis to artificial network models

Abstract

The question of how the collective activity of neural populations gives rise to complex behaviour is fundamental to neuroscience. At the core of this question lie considerations about how neural circuits can perform computations that enable sensory perception, decision making, and motor control. It is thought that such computations are implemented by the dynamical evolution of distributed activity in recurrent circuits. Thus, identifying dynamical structure in neural population activity is a key challenge towards a better understanding of neural computation. At the same time, interpreting this structure in light of a computation of interest is essential for linking the time-varying activity patterns of the neural population to ongoing computational processes. Here, we review methods that aim to quantify structure in neural population recordings through a dynamical system defined in a low-dimensional latent variable space. We discuss advantages and limitations of different modelling approaches and address future challenges for the field.

Introduction

Survival often depends on computations that are explicitly dynamical in structure. Consider navigating to a remembered shelter, integrating ambiguous or noisy cues to detect a predator, or planning and executing a precise movement to catch or avoid becoming prey. In each case an estimate of current state—reflecting relative location, partial information about the threat, or pose of the body in relation to the target—must be retained and updated to determine the best course of future action.

Dynamical computations, and the associated maintenance of dynamical state, are likely to be supported by the exuberantly recurrent circuitry found in much of the central nervous system. Even rapid responses to brief stimuli may depend in part on recurrent dynamics. One signature of dynamical computation is that it takes time. Although observers can assign a broad category to the visual object that appears in a flashed image with a speed that seems consistent with feedforward processing

alone, more detailed categorisation requires longer processing of the same briefly-presented stimulus [1]. Indeed, almost all sensory decisions benefit (to a point) from greater processing time, even when the duration of the sensory stimulus itself remains fixed. This is the ubiquitous phenomenon of speed-accuracy tradeoff [2]. The hypothesis that this gain in accuracy results from greater time for dynamical computation is supported by studies of backward masking, in which performance may be limited by 'stopping' such processing prematurely [3].

An emphasis on dynamics in neural computation is increasingly bringing to the fore empirical methods that characterise the temporal evolution of neural activity in recurrent circuits. Many such methods focus on visualisation, for example using dimensionality reduction techniques, but by themselves such methods cannot identify the dynamical rule that dictates how current state and inputs combine to shape the evolution of activity. It is this dynamical rule that embodies the computational algorithm implemented by the network, and thus it is important for analytic methods to move beyond mere visualisation of temporal patterns to direct statistical identification of dynamics.

Here, we review recent work that seeks to combine characteristion of a low-dimensional manifold explored by population activity with estimation of the dynamical rules responsible for generating the relevant activity patterns. We begin with a brief overview of dimensionality reduction, connecting those ideas to a dynamical systems framework for studying neural computation. Next, we explore methods that seek to capture dynamical laws of increasing complexity on the low-dimensional manifolds underlying population data. We close by considering limitations and opportunities for future work in the field.

Characterising neural activity manifolds

The time-varying activity of a population of neurons can be described in terms of a high-dimensional coordinate system with axes corresponding to the firing rate of each neuron. The collective activity of all neurons in the population determines a single point in this high-dimensional coordinate system—the *population state*. As the activity unfolds over time, the population state traces out a *trajectory* in the high-dimensional space. In some situations, the dimensionality of the space explored by these population trajectories for a given computation may be much smaller than the number of neurons in the circuit that maintains the relevant representation. In this case, computational activity will be confined to a low-dimensional subset or *manifold* contained within this high-dimensional *ambient* space (Figure 1) [4]. Here, each state value corresponds to one (or few) pattern(s) of population activity, and the set of all activity patterns encountered during computation will, up to intrinsic neuronal noise, reflect this dimensionality rather than the size of the network.

Computational manifolds with a range of geometries have been observed in neural activity. These include linear or affine subspaces (Figure 1a) observed in cortical activity before and during deliberate reaching [5–8], decision making [9, 10] and motor timing [11, 12], where the activity manifold lies almost flat. Nonlinear examples include representations of space in the hippocampus and entorhinal cortex [13–15], such as a ring topology in the head-direction cell system [16, 17] (Figure 1b), or more generally curved manifolds [18]. In other settings, the manifold may be star-shaped (Figure 1c), as when sparse groups of neurons encode different state values [19], or comprise a union of disjoint submanifolds as in the collection of synchronous firing states of a synfire chain [20].

In many of these examples, the dimensionality of the computational manifold is closely related to the dimensionality of an encoded state variable, such as location in two or three dimensional space or head

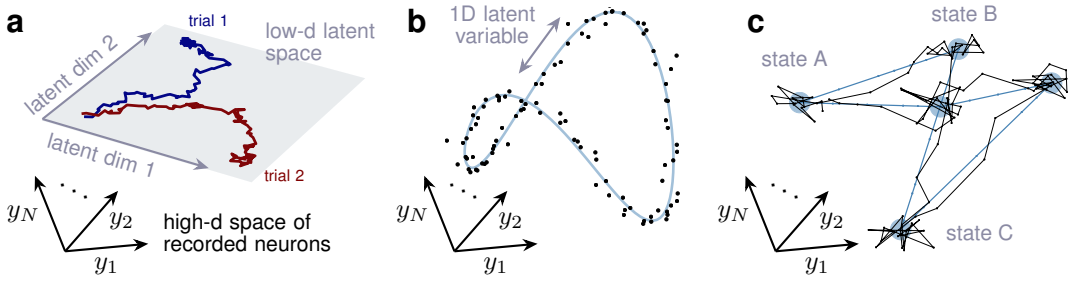

Figure 1: **a:** Population activity evolves in a high dimensional coordinate space, where each axis corresponds to the firing rate of a recorded unit. Even though this space is high-dimensional, neural activity may only explore a low-dimensional subspace. Inter-trial differences in task outcome may be reflected in inter-trial differences in the evolution of the neural population state over time. Dimensionality reduction techniques can be applied to obtain descriptions of the high-dimensional population activity in terms of a low-dimensional latent variable. **b:** Schematic illustration of a one-dimensional manifold, embedded nonlinearly in the high-dimensional space of recorded neurons. Similar topologies have been identified in the head-direction cell system [16, 17]. **c:** Illustration of a manifold representing different state values, where each state engages different groups of neurons.

direction. The dynamical compution that updates this state variable in response to self motion, sensory input, or perhaps as part of memory or action planning, will then appear to move activity along the neutrally stable manifold of valid encodings giving rise to a computation that is largely confined within this manifold. More generally, if the complexity of the underlying computation is low—as is often the case in tightly controlled experimental tasks—then only a few computational variables need to be encoded in the circuit and the dimensionality of the associated activity patterns may hence be low as well [21]. Where representational demands are more elaborate, for instance when extracting features from a high-dimensional set of natural stimuli, then neural activity reflects these added dimensions [22]. Even in this case, however, the temporal evolution of activity in response to any one image may remain confined to a lower dimensional space [23].

When relevant, the low-dimensional manifold of activity, as well as the trajectories by which activity evolves along that manifold, may be found by applying dimensionality reduction techniques to population data recorded while the computation is performed [24]. In effect, these methods use instantaneous relationships between the activity in different neurons [5, 16, 19, 25–28] or between that activity and externally measureable quantities [9, 29, 30], to estimate the low-dimensional structure in the data. Often, these statistical relationships in observed activity are captured by assuming dependence on a shared but unobserved *latent* variable. In such cases, the latent variable provides a direct estimate of the representation of the underlying computational state. Otherwise a further step is necessary to establish a coordinate system for the recovered manifold and so provide a variable that represents state [16].

The manifold of population states is one part of the link between neural circuits and dynamical computation. However, activity manifolds are only the signature of an ongoing computational process and do not by themselves provide a description of the relevant computation. Although at least one recent approach seeks to identify low-dimensional projections most compatible with dynamics [31], a full connection between the observed population activity patterns and the underlying circuit computation depends on explicit identifcation of the dynamical rule that dictates how neural activity evolves in time.

From data to dynamics to computation

The way in which temporal evolution of population activity implements computation is formalised by viewing the network as a dynamical system [32–34], defined by a state and a dynamical law—the system *dynamics*—that governs the future evolution of that state based on its current value and inputs. This dynamics, ultimately implemented by the connections and integration properties of the neurons in the network, then captures the computational transformation of inputs to outputs over time. Its features, such as the location and stability of fixed points or limit cycles, their basins of attraction, and properties of input-driven orbits in the state space, provide insight into the structure of the computation.

Consider, for example, a network involved in a simple perceptual decision-making task, such as reporting the direction of motion of a random dots stimulus [35]. To solve this task, the network would need to update a decision variable representing left- or rightward motion, based on perceptual evidence arriving as inputs to the system. We may hence observe that activity in this network evolves gradually towards a different pattern of firing for each choice made (Figure 1a). Such trajectories are consistent with a simple dynamical system in which two (or more) attractor states are positioned such that transient inputs that reflect sensory evidence nudge the state from a un- or marginally stable initial point towards one or the other attractor (Figure 2c). Indeed, from a theoretical point of view, spiking neural network models that achieve such integrative decision making [36] can be reduced to just such a two-dimensional dynamical description in terms of decision-related variables [37] (Figure 3c). In this case, the link between circuit computation and dynamics is made explicit.

Is it possible to extract an analogous description in terms of low-dimensional latent variables directly from data recorded during a perceptual decision making task? Data analysis methods that facilitate such a description could offer ways to refine and test hypotheses around computation in recurrent circuits (Figure 3a), and thus represent a path towards a data-driven understanding of the biological computational strategy.

Identifying dynamical systems from data

To study neural computation, we are interested in estimating and analysing the dynamics that govern the temporal evolution of the underlying computational state of the network. A general class of latent dynamical, or *state-space* models, facilitate this through the description of the evolution of neural activity in terms of a low-dimensional latent computational state (here written $\boldsymbol{x}(t)$ or $\boldsymbol{x}_t$ in discretised time) that relates linearly or nonlinearly to recorded neural activity and evolves in time with dynamics of an assumed functional form. Given recorded neural activity, the models can be fit to data to infer a distribution over the path of the latent variable, and estimate the dynamics that describes this evolution.

Linear models

The simplest class of latent dynamical models incorporates a linear dynamical system, modelling the evolution of the latent state as

$$\boldsymbol{x}_{t+1} = A\,\boldsymbol{x}_t + B\,\boldsymbol{u}_{t+1} + \boldsymbol{\eta}_t \tag{1}$$

in terms of a dynamics matrix A, inputs $B\,\boldsymbol{u}_t$, and (typically random Gaussian) state innovations $\boldsymbol{\eta}_t$.

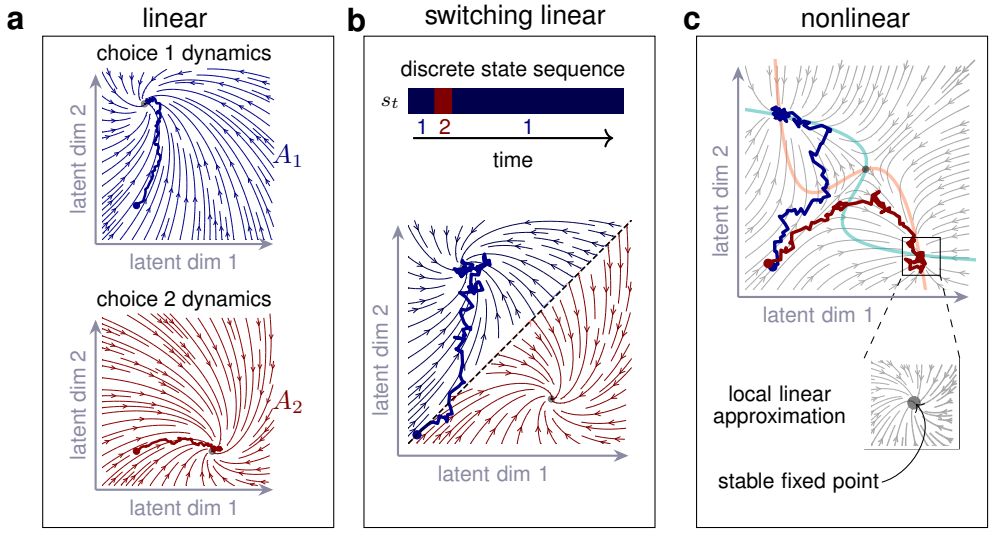

Figure 2: Description of the temporal evolution of the population activity in Figure 1a in terms of different dynamical systems defined on the low-dimensional latent variable. The dynamics are illustrated through a flow field. **a:** Linear dynamics with a single fixed point fit to trials corresponding to the same choice. **b:** Switching dynamical system composed of two linear sub-systems and a discrete variable indicating which system dictates the evolution of the latent variable at each time point. When the discrete state variable is dependent on the current value of the latent variable, one subsystem is more likely to be active in a given region of state-space. **c:** Nonlinear dynamical system with three fixed points. Coloured lines show the system's nullclines, where the temporal derivative of one of the latent dimensions is zero. The path of the latent variables converges to one of the two stable fixed points. The local behaviour of the system can be analysed via linear approximations.

Linear systems are made particularly attractive by the fact that their asymptotic behaviour can be understood in terms of the eigenvalues and eigenvectors of A. For stable systems and in the absence of inputs, activity will eventually decay towards a single fixed point at zero. Depending on the orientation of the eigenvectors of A, the system may also exhibit transient, non-normal amplification on shorter timescales [38, 39]. Analysis of control-theoretic objects like controllability and observability Gramians provide further insight into the system's behaviour in response to inputs [40].

Linear dynamical systems have been applied to brain-machine interface control [41], and used to study mechanisms underlying contextual evidence integration in prefrontal cortex [42]. They have also provided a theoretical framework within which to understand features of neural responses during motor preparation and execution [43, 44], in primary visual cortex [39] and working memory [38]. These applications have provided insight even though linear dynamical transformations of input are unlikely to capture the full richness of neural dynamics. In part this is because linear systems provide parsimonious *local* descriptions (for a single experiment condition, a single task epoch, or for restricted ranges of neural activity) of potentially more complex circuit dynamics (Figure 2a). Local linear approximations also represent an important building block for the analysis of the behaviour of more complex, nonlinear dynamical systems (Figure 2c) [45].

Switching linear models

Switching linear dynamical systems mix multiple linear updates under the control of a discrete switching variable. The simplest model takes the form:

$$\begin{aligned} s_{t+1} &\sim p(s_{t+1}|s_t)\,, \qquad s_t \in \{1,\ldots,M\} \\ \boldsymbol{x}_{t+1} &= A_{s_t}\boldsymbol{x}_t + B_{s_t}\boldsymbol{u}_{t+1} + \boldsymbol{\eta}_t\,, \end{aligned} \tag{2}$$

with the discrete Markov variable s_t selecting the linear evolution A_i and input mapping B_i ($i = 1,\ldots,M$) that act on the computational state at time t. Switching models are more expressive than a single linear model: switches in dynamics may reflect the influence of multiple fixed points, or capture changes in dynamics with time, or under the control of unobserved inputs. More generally, arbitrary non-linear dynamics may be approximated by a piecewise linear form (Figure 2b). However, while the simple form of eq. (2) allows transitions between dynamical regimes to be estimated from neural data, it lacks the capacity to select linear dynamics based on the value of the computational state. Thus, a richer model includes dependence of s_t (and thus the local linear dynamics) on the current computational state $\boldsymbol{x}_t$ [46–49]. One advantage to such an approximation is that by retaining explicit linearity in each subsystem, dynamics within each region can still be analysed and understood using classic tools from linear systems analysis.

Applications of switching linear dynamical systems include the analysis of neural population activity during motor planning and execution [49, 50], and the identification of latent activity patterns in recordings from C. elegans [47, 49]. In each case, models were able to identify relationships between behavioral states and the discrete state sequence inferred in the model.

Nonlinear models

Theoretical studies have suggested essential roles for nonlinear network dynamics in computation [36, 51, 52], with outputs depending on multiple fixed points, limit cycles or line attractors, or bifurcations with changes in inputs. Many of these features cannot be expressed in linear systems, and may be difficult to approximate well with switching linear systems without overwhelming complexity. Consequently, there has also been substantial interest in identifying nonlinear models of dynamics in a circuit from neural data. In these models, the dynamics are expressed very generally through a parametrised function $\boldsymbol{f}(\cdot)$:

$$\boldsymbol{x}_{t+1} = \boldsymbol{f}(\boldsymbol{x}_t, \boldsymbol{u}_{t+1}) + \boldsymbol{\eta}_t \tag{3}$$

which describes the state evolution as a function of both current latent state $\boldsymbol{x}$ and any inputs arriving to the system $\boldsymbol{u}$.

Choosing flexible function classes for $\boldsymbol{f}(\cdot)$ makes it possible to describe richer sets of dynamics than linear or piecewise-linear models. Such flexible choices include nonparametric approaches like Gaussian Process State Space Models [53–55], representation of the dynamics in terms of eigenfunctions of the Hermitian operator of the dynamical system [56], or parametrisations in terms of a neural network [57, 58].

The local behaviour of nonlinear dynamics can still be understood in terms of linear approximations given by the Jacobian matrix evaluated at fixed points of the system (Figure 2c) [45]. However, identifying fixed-points and Jacobians of the nonlinear system typically requires a second stage of estimation [59], or the specification of models to include these features directly as additional parameters [55].

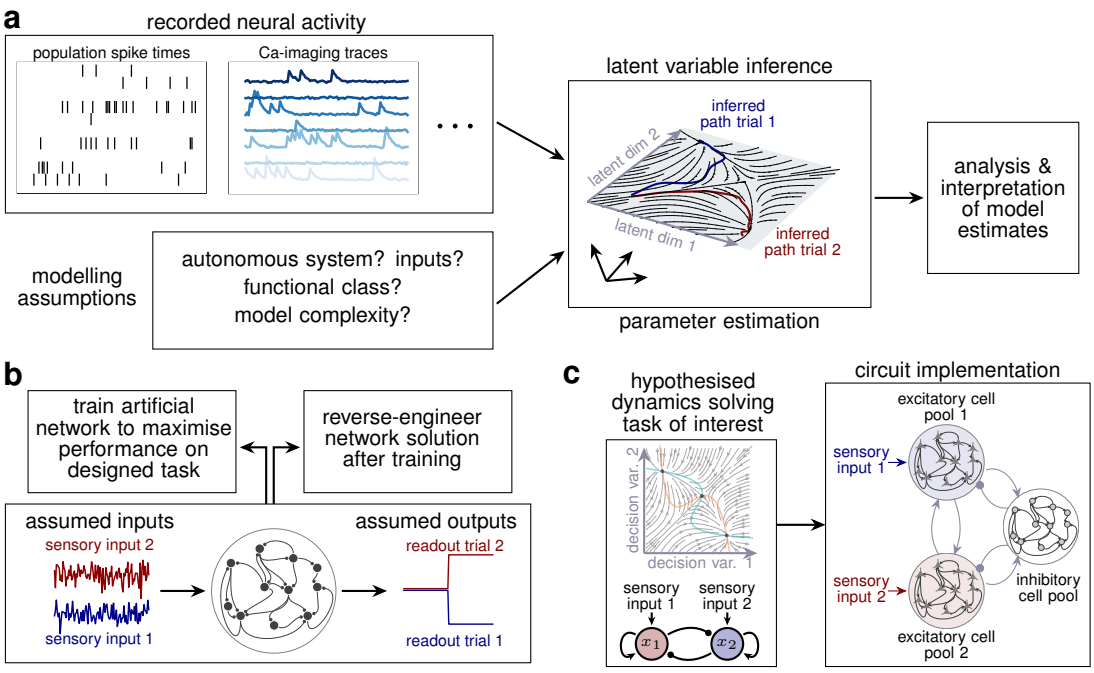

Figure 3: Different approaches towards studying neural population dynamics and computation. **a:** Data analysis approaches require making assumptions to specify a parametric model of the underlying populations dynamics. Different modalities of recorded neural activity can be used for model fitting to obtain a description of low-dimensional structure in the neural population and estimates of the model parameters. The recovered dynamical system can then be analysed in light of the task and computation of interest. **b:** Task-optimised artificial network models learn a pre-specified input-output transformation. Everything about the system is observable and can be analysed in detail to extract insight about the resulting dynamical system solving the task. However, the solution will depend on assumptions about task and network design, as well as the cost function used for optimisation. **c:** Mechanistic models are designed to implement a hypothesised solution to a given task, often taking biological constraints like cell-type structure into account. The design is often set up to capture selected qualitative features of neural data.

Outstanding challenges and future directions

While dynamical systems offer a compelling framework for studying computation in recurrent circuits, connecting such ideas to neural data is met with a number of challenges in practice.

These include challenges associated with fitting dynamical models to data, like choosing good criteria for model selection when the goal is interpretation and scientific discovery [56], and adapting statistical models to accommodate neural response features, which can lead to intractabilities in performing statistical inference.

Furthermore, inputs to the local population are generally unknown, which makes it difficult to determine which features of neural activity reflect the dynamics of the local circuit, or dynamics of inputs from other areas. Recent work has begun to investigate the use of single trial activity to tease apart these contributions [60]. Still, existing approaches for studying the dynamics of a local population generally have to make assumptions about these components in order to arrive at a model description that is

identifiable from data (Figure 3a). So far, it has been difficult to validate such assumptions, yet their impact on the inferred dynamics may be large.

A further challenge relates to the interpretability of latent dynamical models with respect to the goal of better understanding computation in biological circuits. The low-dimensional dynamics evolve in an abstract latent space, which is typically identified in an unsupervised way and may not easily be related to variables relevant to the computation or task of interest. Deriving explicit statements about mechanistic circuit computation from the estimated model fits is hard, since any inference about characteristics of dynamics in latent space do not directly translate to dynamics defined through a connectivity matrix in neural space. Known constraints on the dynamics of the biological circuit, such as excitatory and inhibitory cell-type structure and Dale's law, are typically not taken into account in methods that learn dynamics from data. However, such properties might influence features of the neural responses used for fitting. This discrepancy in levels of analysis also makes it difficult to derive testable experimental predictions based on any insights data analysis methods may provide.

To address these challenges, it is important to highlight that there are other, complementary approaches towards studying network dynamics and computation. These include reverse-engineering task-optimised artificial neural network models [45, 61] (Figure 3b) and designing mechanistic network implementations of a particular computation [36] (Figure 3c). Such network models are fully observable and can incorporate biological constraints, yet their connection to data is often only qualitative in nature. Furthermore, they too rely on model assumptions, such as network architecture, cost function, and structure in network inputs and target outputs.

Drawing connections between mechanistic models of network computations, task-optimized aritifical neural networks, and data analysis will help to better inform model assumptions, and provide more quantitative links between theoretical network models and observed neural data. Such connections will also aid model interpretability, and ultimately inform the design of experiments to test and refine the hypotheses different models may generate. Indeed, recent work has begun to formulate statistical models in ways to establish more direct connections between theory and data analysis [55, 62], or use experimentally observed responses to constrain activity in recurrent network network models [63]. In addition to this, establishing connections between latent variable models and mechanistic modelling approaches might aid the interpretability of model parameters that are learned from data. Going forward, it will be important to establish clearer links between theoretical descriptions of network computation and data-driven latent dynamical models.

Acknowledgements

This work was supported by the Gatsby Charitable Foundation and the Simons Collaboration on the Global Brain. The authors would like to thank Laura Driscoll, Joana Soldado Magraner and Jorge Menendez for feedback on the manuscript and insightful discussions contributing to the content presented here.

References

• of special interest
•• of outstanding interest

[1] Michèle Fabre-Thorpe. **The characteristics and limits of rapid visual categorization**. *Frontiers in Psychology*, **2**:243, 2011. doi: 10.3389/fpsyg.2011.00243.

[2] Wayne A. Wickelgren. **Speed-accuracy tradeoff and information processing dynamics**. *Acta Psychologica*, **41**(1):67–85, 1997.

[3] Adam Reeves. **An analysis of visual masking, with a defense of 'stopped processing'**. *Advances in Cognitive Psychology*, **3**(1–2):57–65, 2008.

[4] Peiran Gao and Surya Ganguli. **On simplicity and complexity in the brave new world of large-scale neuroscience**. *Current opinion in neurobiology*, **32**:148–155, 2015.

[5] Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. **Gaussian-process factor analysis for low-dimensional single-trial analysis of neural population activity**. *Journal of Neurophysiology*, **102**(1):614–635, 2009.

[6] Krishna V. Shenoy, Maneesh Sahani, and Mark M. Churchland. **Cortical Control of Arm Movements: a Dynamical Systems Perspective**. *Annual Review of Neuroscience*, **36**(1): 337–359, 2013.

[7] Mark M Churchland, John P Cunningham, Matthew T Kaufman, Justin D Foster, Paul Nuyujukian, Stephen I Ryu, and Krishna V Shenoy. **Neural population dynamics during reaching**. *Nature*, **487**(7405):51–56, 2012.

[8] Gamaleldin F Elsayed, Antonio H Lara, Matthew T Kaufman, Mark M Churchland, and John P Cunningham. **Reorganization between preparatory and movement population responses in motor cortex**. *Nature communications*, **7**(1):1–15, 2016.

[9] Valerio Mante, David Sussillo, Krishna V Shenoy, and William T Newsome. **Context-dependent computation by recurrent dynamics in prefrontal cortex**. *nature*, **503**(7474):78–84, 2013.

[10] Christian K Machens, Ranulfo Romo, and Carlos D Brody. **Functional, but not anatomical, separation of "what" and "when" in prefrontal cortex**. *Journal of Neuroscience*, **30**(1): 350–360, 2010.

[11] Evan D Remington, Devika Narain, Eghbal A Hosseini, and Mehrdad Jazayeri. **Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics**. *Neuron*, **98**(5): 1005–1019, 2018.

[12] Hansem Sohn, Devika Narain, Nicolas Meirhaeghe, and Mehrdad Jazayeri. **Bayesian computation through cortical latent dynamics**. *Neuron*, **103**(5):934–947, 2019.

[13] John O'Keefe and Jonathan Dostrovsky. **The hippocampus as a spatial map: Preliminary evidence from unit activity in the freely-moving rat.** *Brain research*, 1971.

[14] Jeffrey S Taube, Robert U Muller, and James B Ranck. **Head-direction cells recorded from the postsubiculum in freely moving rats. i. description and quantitative analysis**. *Journal of Neuroscience*, **10**(2):420–435, 1990.

[15] Marianne Fyhn, Sturla Molden, Menno P Witter, Edvard I Moser, and May-Britt Moser. **Spatial representation in the entorhinal cortex**. *Science*, **305**(5688):1258–1264, 2004.

[16] Rishidev Chaudhuri, Berk Gerçek, Biraj Pandey, Adrien Peyrache, and Ila Fiete. **The intrinsic attractor manifold and population dynamics of a canonical cognitive circuit across waking and sleep**. *Nature Neuroscience*, **22**(9):1512–1520, 2019.

[17] Kristopher Jensen, Ta-Chu Kao, Marco Tripodi, and Guillaume Hennequin. **Manifold GPLVMs for discovering non-euclidean latent structure in neural data**. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 22580–22592. Curran Associates, Inc., 2020.

• Jensen et al. develop an novel approach to infer a latent variable specified on a manifold of a given geometry, such as a sphere, ring or torus. The approach recovers improved latent state estimates when the underlying manifold is taken into account, rather than assuming a Euclidean latent space. In cases where the manifold topology is unknown, cross-validation can be used for model selection across candidate topologies.

[18] Ryan J Low, Sam Lewallen, Dmitriy Aronov, Rhino Nevers, and David W Tank. **Probing variability in a cognitive map using manifold inference from neural dynamics**. *bioRxiv*, page 418939, 2018.

[19] Alon Rubin, Liron Sheintuch, Noa Brande-Eilat, Or Pinchasof, Yoav Rechavi, Nitzan Geva, and Yaniv Ziv. **Revealing neural correlates of behavior without behavioral measurements**. *Nature communications*, **10**(1):1–14, 2019.

[20] Moshe Abeles. *Local Cortical Circuits An Electrophysiological Study*. Springer, 1982.

[21] Peiran Gao, Eric Trautmann, Byron Yu, Gopal Santhanam, Stephen Ryu, Krishna Shenoy, and Surya Ganguli. **A theory of multineuronal dimensionality, dynamics and measurement**. *BioRxiv*, page 214262, 2017.

[22] Carsen Stringer, Marius Pachitariu, Nicholas Steinmetz, Matteo Carandini, and Kenneth D Harris. **High-dimensional geometry of population responses in visual cortex**. *Nature*, **571**(7765): 361–365, 2019.

[23] Jeffrey S Seely, Matthew T Kaufman, Stephen I Ryu, Krishna V Shenoy, John P Cunningham, and Mark M Churchland. **Tensor analysis reveals distinct population structure that parallels the different computational roles of areas m1 and v1**. *PLoS computational biology*, **12**(11): e1005164, 2016.

[24] John P Cunningham and Byron M Yu. **Dimensionality reduction for large-scale neural recordings**. *Nature neuroscience*, **17**(11):1500–1509, 2014.

[25] Saul Kato, Harris S Kaplan, Tina Schrödel, Susanne Skora, Theodore H Lindsay, Eviatar Yemini, Shawn Lockery, and Manuel Zimmer. **Global brain dynamics embed the motor command sequence of caenorhabditis elegans**. *Cell*, **163**(3):656–669, 2015.

[26] Anqi Wu, Stan Pashkovski, Sandeep R Datta, and Jonathan W Pillow. **Learning a latent manifold of odor representations from neural responses in piriform cortex**. In *Advances in Neural Information Processing Systems*, pages 5378–5388, 2018.

[27] Juan A Gallego, Matthew G Perich, Lee E Miller, and Sara A Solla. **Neural manifolds for the control of movement**. *Neuron*, **94**(5):978–984, 2017.

[28] Stephen Keeley, Mikio Aoi, Yiyi Yu, Spencer Smith, and Jonathan W Pillow. **Identifying signal and noise structure in neural population activity with gaussian process factor models**. *Advances in Neural Information Processing Systems*, **33**, 2020.

[29] Dmitry Kobak, Wieland Brendel, Christos Constantinidis, Claudia E Feierstein, Adam Kepecs, Zachary F Mainen, Xue-Lian Qi, Ranulfo Romo, Naoshige Uchida, and Christian K Machens. **Demixed principal component analysis of neural population data**. *Elife*, **5**:e10989, 2016.

[30] Mikio C Aoi, Valerio Mante, and Jonathan W Pillow. **Prefrontal cortex exhibits multidimensional dynamic encoding during decision-making**. *Nature Neuroscience*, pages 1–11, 2020.

[31] Virginia M. S. Rutten, Alberto Bernacchia, Maneesh Sahani, and Guillaume Hennequin. **Non-reversible Gaussian processes for identifying latent dynamical structure in neural data**. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 9622–9632. Curran Associates, Inc., 2020.

[32] Krishna V Shenoy, Maneesh Sahani, and Mark M Churchland. **Cortical control of arm movements: a dynamical systems perspective**. *Annual review of neuroscience*, **36**, 2013.

[33] Laura N Driscoll, Matthew D Golub, and David Sussillo. **Computation through cortical dynamics**. *Neuron*, **98**(5):873–875, 2018.

[34] Saurabh Vyas, Matthew D Golub, David Sussillo, and Krishna V Shenoy. **Computation through neural population dynamics**. *Annual Review of Neuroscience*, **43**:249–275, 2020.

[35] Jamie D Roitman and Michael N Shadlen. **Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task**. *Journal of neuroscience*, **22**(21):9475–9489, 2002.

[36] Xiao-Jing Wang. **Probabilistic decision making by slow reverberation in cortical circuits**. *Neuron*, **36**(5):955–968, 2002.

[37] Kong-Fatt Wong and Xiao-Jing Wang. **A recurrent network mechanism of time integration in perceptual decisions**. *Journal of Neuroscience*, **26**(4):1314–1328, 2006.

[38] Mark S Goldman. **Memory without feedback in a neural network**. *Neuron*, **61**(4):621–634, 2009.

[39] Brendan K Murphy and Kenneth D Miller. **Balanced amplification: a new mechanism of selective amplification of neural activity patterns**. *Neuron*, **61**(4):635–648, 2009.

[40] Ta-Chu Kao and Guillaume Hennequin. **Neuroscience out of control: control-theoretic perspectives on neural circuit dynamics**. *Current opinion in neurobiology*, **58**:122–129, 2019.

[41] Jonathan C Kao, Paul Nuyujukian, Stephen I Ryu, Mark M Churchland, John P Cunningham, and Krishna V Shenoy. **Single-trial dynamics of motor cortex and their applications to brain-machine interfaces**. *Nature communications*, **6**(1):1–12, 2015.

[42] Joana Soldado Magraner. *Linear Dynamics of Evidence Integration in Contextual Decision Making*. PhD thesis, UCL (University College London), 2018.

[43] Guillaume Hennequin, Tim P Vogels, and Wulfram Gerstner. **Optimal control of transient dynamics in balanced networks supports generation of complex movements**. *Neuron*, **82** (6):1394–1406, 2014.

[44] Ta-Chu Kao, Mahdieh S Sadabadi, and Guillaume Hennequin. **Optimal anticipatory control as a theory of motor preparation: a thalamo-cortical circuit model**. *Neuron*, 2021.

[45] David Sussillo and Omri Barak. **Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks**. *Neural Computation*, **25**(3):626–649, 2013.

[46] Scott Linderman, Matthew Johnson, Andrew Miller, Ryan Adams, David Blei, and Liam Paninski. **Bayesian learning and inference in recurrent switching linear dynamical systems**. In *Artificial Intelligence and Statistics*, pages 914–922, 2017.

[47] Scott Linderman, Annika Nichols, David Blei, Manuel Zimmer, and Liam Paninski. **Hierarchical recurrent state space models reveal discrete and continuous dynamics of neural activity in c. elegans**. *bioRxiv*, page 621540, 2019.

• Linderman et al. introduce a hierarchical switching dynamical system to model neural activity in C. elegans. Their hierarchical approach allows them to use data from different partial recordings and different animals to estimate model structure that is shared or unique to each animal. They demonstrate that the interred time-course of discrete states in the model is behaviourally meaningful.

[48] Josue Nassar, Scott Linderman, Monica Bugallo, and Il Memming Park. **Tree-structured recurrent switching linear dynamical systems for multi-scale modeling**. In *International Conference on Learning Representations*, 2019.

[49] Joshua Glaser, Matthew Whiteway, John P Cunningham, Liam Paninski, and Scott Linderman. **Recurrent switching dynamical systems models for multiple interacting neural populations**. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 14867–14878. Curran Associates, Inc., 2020.

•• Glaser et al. use a switching linear dynamical system to model the dynamics of interacting neural populations. The model allows for switches in the dynamics of each populations, as well as in the interactions across populations. The authors apply their method to data from motor cortical areas in nonhuman primate and to neural activity in C. elegans.

[50] Biljana Petreska, Byron M Yu, John P Cunningham, Gopal Santhanam, Stephen I Ryu, Krishna V Shenoy, and Maneesh Sahani. **Dynamical segmentation of single trials from population neural data**. In *Advances in neural information processing systems*, pages 756–764, 2011.

[51] Christian K Machens, Ranulfo Romo, and Carlos D Brody. **Flexible control of mutual inhibition: a neural model of two-interval discrimination**. *Science*, **307**(5712):1121–1124, 2005.

[52] Francesca Mastrogiuseppe and Srdjan Ostojic. **Linking connectivity, dynamics, and computations in low-rank recurrent neural networks**. *Neuron*, **99**(3):609–623, 2018.

[53] Jack Wang, Aaron Hertzmann, and David J Fleet. **Gaussian Process Dynamical Models**. In *Advances in Neural Information Processing Systems 18*, pages 1441–1448, 2006.

[54] Ryan Turner, Marc Deisenroth, and Carl Rasmussen. **State-Space Inference and Learning with Gaussian Processes**. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 868–875, 2010.

[55] Lea Duncker, Gergo Bohner, Julien Boussard, and Maneesh Sahani. **Learning interpretable continuous-time models of latent stochastic dynamical systems**. In Kamalika Chaudhuri and Ruslan Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 1726–1734, Long Beach, California, USA, 09–15 Jun 2019. PMLR.

•• The authors develop an approach to learn a low-dimensional nonlinear dynamical system directly from the spike times of simultaneously recorded neurons. The nonlinear dynamics are described via a Gaussian Process, conditioned on fixed point and local Jacobian matrices around them. These interpretable features of the dynamics are hence directly accessible for further analysis, instead of requiring a second stage of estimation to identify fixed points of the dynamics.

[56] Mikhail Genkin and Tatiana A. Engel. **Moving beyond generalization to accurate interpretation of flexible models**. *Nature Machine Intelligence*, 2020.

• Genkin et al. develop an approach for maximum likelihood learning of a latent nonlinear dynamical system from spike train observations. The authors also discuss issues relating to model selection based on generalisation performance and the interpretation of the resulting dynamics. They propose a new complexity metric to select interpretable models across a range of models with similar predictive performance.

[57] Chethan Pandarinath, Daniel J O'Shea, Jasmine Collins, Rafal Jozefowicz, Sergey D Stavisky, Jonathan C Kao, Eric M Trautmann, Matthew T Kaufman, Stephen I Ryu, Leigh R Hochberg, et al. **Inferring single-trial neural population dynamics using sequential auto-encoders**. *Nature Methods*, page 1, 2018.

[58] Daniel Hernandez, Antonio K Moretti, Shreya Saxena, Ziqiang Wei, John Cunningham, and Liam Paninski. **Nonlinear evolution via spatially-dependent linear dynamics for electrophysiology and calcium data**. *Neurons, Behavior, Data analysis, and Theory*, **3**(3):13476, 2020.

[59] Matt Golub and David Sussillo. **Fixedpointfinder: A Tensorflow toolbox for identifying and characterizing fixed points in recurrent neural networks**. *Journal of Open Source Software*, **3**(31):1003, 2018.

[60] Aniruddh R Galgali, Maneesh Sahani, and Valerio Mante. **Residual dynamics resolves recurrent contributions to neural computation**. *bioRxiv*, 2021.

•• Galgali et al. develop a method for identifying local recurrent contribution to observed neural activity patterns by fitting time-varying linear dynamics to single-trial residuals around a condition average trajectory. The authors apply their approach to population activity from macaque prefrontal cortex to study its role in perceptual decision making.

[61] Omri Barak. **Recurrent neural networks as versatile tools of neuroscience research**. *Current opinion in neurobiology*, **46**:1–6, 2017.

[62] David M Zoltowski, Jonathan W Pillow, and Scott W Linderman. **A general recurrent state space framework for modeling neural dynamics during decision-making**. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine*

Learning Research, pages 11680–11691, Virtual, 13–18 Jul 2020. PMLR.

• Zoltowski et al. express different theoretical models of decision-making as specific instantiations of a switching dynamical system. This allows them to quantitatively evaluate and compare the different theoretical models in neural data. The inferred latent states have a direct theoretical interpretation in their modelling framework.

[63] Matthew G Perich, Charlotte Arlt, Sofia Soares, Megan E Young, Clayton P Mosher, Juri Minxha, Eugene Carter, Ueli Rutishauser, Peter H Rudebeck, Christopher D Harvey, and Kanaka Rajan. **Inferring brain-wide interactions using data-constrained recurrent neural network models**. *bioRxiv*, 2020.