Neuroscience out of control: control-theoretic perspectives on neural circuit dynamics

Ta-Chu Kao[1] **and Guillaume Hennequin**[@1]

[1]Computational and Biological Learning Lab, Dept. of Engineering, University of Cambridge, Cambridge, UK

[@] Corresponding author (g.hennequin@eng.cam.ac.uk)

September 6, 2019

Summary

A major challenge in systems neuroscience is to understand how the dynamics of neural circuits give rise to behaviour. Analysis of complex dynamical systems is also at the heart of control engineering, where it is central to the design of robust control strategies. Although a rich engineering literature has grown over decades to facilitate the analysis of such systems, little of it has percolated into neuroscience so far. Here, we give a brief introduction to a number of core control-theoretic concepts that provide useful perspectives on neural circuit dynamics. We introduce important mathematical tools related to these concepts, and establish connections to neural circuit analysis, focusing on a number of themes that have arisen from the modern "state-space" view on neural population dynamics.

Highlights

Control theory offers unique, geometric perspectives on neural circuit dynamics.

Observability and controllability characterize the input/output behaviour of circuits.

Observability reveals input sensitivity, nullspaces, and communication subspaces.

Controllability defines intrinsic manifolds, illuminating recent BCI experiments.

Model reduction helps understand and interpret dynamics in high-dimensional spaces.

Introduction

Behaviour arises from dynamics that unfold in recurrently connected neural circuits, both locally within specialized regions and globally across multiple brain areas. To understand the principles that govern these dynamics, computational neuroscientists build model networks in a variety of ways: i) models that directly implement a number of known physiological features of the brain area(s) of interest [1], ii) low-dimensional latent dynamical models fitted to neural population recordings [2–5], and iii) artificial neural networks trained to perform complex tasks [6]. In all cases, a critical research step is to understand the behaviour of the model through an analysis of its dynamics [7, 8].

Basic linear algebra provides useful geometric intuitions for the structure of population activity patterns, often represented as points in a so-called "state-space" (Figure 1; [9–13]). Concepts such as projections, subspaces, nullspaces, etc, lie at the core of many linear dimensionality reduction techniques that are widely used to explore neural datasets [14–16], and help us reason geometrically about the computations carried out by neural circuits [17–25]. However, similar intuitions are more difficult to obtain for population *dynamics*, i.e. for the temporal evolution of neural activity in state space (Figure 1). This challenge is especially relevant given the recent shift in focus from static analyses to dynamical descriptions of circuit computations [4, 10, 17, 26••, 20, 24]. Thus, an important goal in computational neuroscience is to develop an "algebra of dynamics", i.e. a set of conceptual, algebraic, and numerical tools for the study of neural circuits.

Control theorists have long been developing such tools [27]. Here, we review the key theoretical concepts of controllability, observability, model reduction and stability, which we find particularly relevant to the analysis of neural circuits. We briefly introduce these concepts in the control engineering context in which they were developed, and discuss the new perspectives that they provide on a number of recent results concerning the dynamics of neural computation.

State space models

A large part of classical control theory is dedicated to the control of linear state-space models of the form:

$$\begin{aligned} \frac{d\mathbf{x}}{dt} &= \mathbf{A}\mathbf{x}(t) + \mathbf{B}\mathbf{u}(t) + \text{noise} \\ \mathbf{y}(t) &= \mathbf{C}\mathbf{x}(t) + \text{noise}. \end{aligned} \tag{1}$$

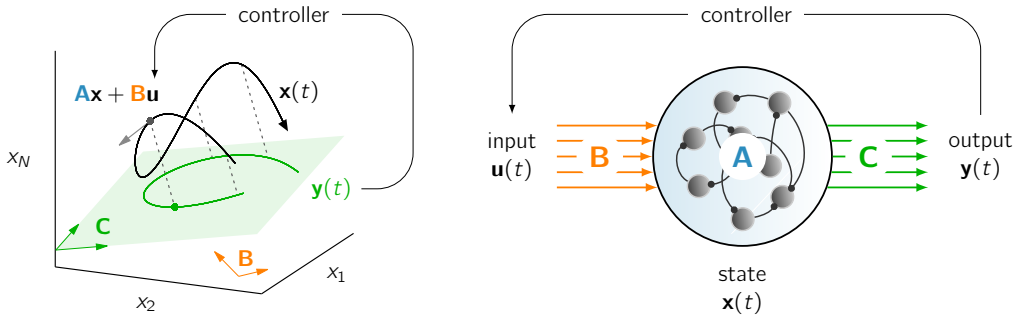

Figure 1: **State space & control perspective on neural dynamics.** The population activity vector $\mathbf{x}(t)$ traces out trajectories in state space (solid black, left diagram), following a flow (gray arrow) determined by the state matrix $\mathbf{A}$ as well as external inputs $\mathbf{u}(t)$ (right diagram; cf. Equation (1)). In a standard feedback control scenario, inputs are computed based on some measurements $\mathbf{y}(t) = \mathbf{C}\mathbf{x}(t)$ (green) of the state vector, modifying the flow of activity along a few select "input channels" $\mathbf{B}$ (orange).

Here, $\mathbf{A}$ is the so-called "state matrix" that governs the temporal evolution of the activity vector $\mathbf{x}(t)$ in the state space, possibly subject to process noise, and $\mathbf{B}$ is an "input matrix" whose columns define the directions along which time-varying inputs $\mathbf{u}(t)$ actuate the system (Figure 1). To control the state of the system (e.g. steer it along desired trajectories or towards desired end points), inputs are often derived as feedback from some (noisy) linear measurements $\mathbf{y}(t)$ of the full state, involving a readout matrix $\mathbf{C}$.

Equation (1) will look familiar to many neuroscientists. It lies at the heart of many statistical models, where it describes the dynamics of a set of "latent factors" that recapitulate the spatiotemporal structure in population recordings [4, 28, 29•]. Due to their analytical tractability, linear(ized) models have also shed light on a variety of circuit computations, including short-term memory [30–32], selective amplification and surround suppression in primary visual cortex [33••, 34], movement generation[26••, 20], attention and decision-making [35], and probabilistic inference [36, 37]. Equation (1) has also been the workhorse of numerous studies on the origin, stimulus-dependence, and attentional-modulation of noise correlations [38, 39]. In many of these contexts, $\mathbf{x}(t)$ represents the vector of momentary firing rates in the network, $\mathbf{A}$ encapsulates recurrent connectivity and single-neuron leak, and $\mathbf{B}$ is a matrix of input synaptic weights.

Control theory textbooks contain a wealth of results that summarize important dynamical properties of systems governed by Equation (1), using matrices that are obtained through algebraic manipulations of $\mathbf{A}$, $\mathbf{B}$ and $\mathbf{C}$. We begin our review with two important such matrices that can be used to describe the input-output behaviour of a neural network: the observability and controllability Gramians.

Network observability

Control-theoretic context

Before even attempting to design a feedback control strategy, a control engineer will first establish feasibility. In particular, do the available partial observations $\mathbf{y}(t)$ of the full state $\mathbf{x}(t)$ provide enough information for efficiently controlling the system? Answers are found in the observability Gramian [27], a positive definite matrix associated with $\mathbf{A}$ and $\mathbf{C}$ and defined as

$$\mathbf{Q} = \int_0^\infty \exp\left(t\mathbf{A}^T\right) \mathbf{C}^T\mathbf{C} \exp\left(t\mathbf{A}\right) dt. \tag{2}$$

This matrix characterizes how well the initial state $\mathbf{x}(0)$, and therefore any subsequent state, can be inferred from observations of $\mathbf{y}(t)$ and $\mathbf{u}(t)$ – systems that make such inferences are often called "observers" [40]. Determining where $\mathbf{x}(0)$ was positioned along a given direction $\mathbf{v}$ in state space requires the quantity $\mathcal{E}(\mathbf{v}) = \mathbf{v}^T\mathbf{Q}\mathbf{v}$ to be strictly positive. If it is zero, one simply cannot resolve $\mathbf{x}(0)$ along $\mathbf{v}$, and such ambiguity makes controlling $\mathbf{x}$ more difficult. More generally, $\mathcal{E}(\mathbf{v})$ quantifies an ideal observer's confidence in this estimate. Conveniently, $\mathbf{Q}$ is also the solution to a simple "Lyapunov equation",

$$\mathbf{A}^T\mathbf{Q} + \mathbf{Q}\mathbf{A} + \mathbf{C}^T\mathbf{C} = 0, \tag{3}$$

with efficient solvers available for most programming languages.

Geometry of input sensitivity in neural circuits

The observability Gramian is a very useful tool for the analysis of neural circuits, as it provides information about a network's sensitivity to specific input patterns, or initial conditions. Specifically, $\mathcal{E}(\mathbf{v})$ above can be re-interpreted as the amount of "energy" in the output $\mathbf{y}(t)$ evoked by a pulse of input along direction $\mathbf{v}$ in state space [26••] (Figure 2A and B). Input that momentarily pushes (or initializes) the state along a highly observable direction typically gives rise to strong and/or long output transients (Figure 2A, bottom left), especially in networks known as "nonnormal" which include all physiologically realistic models of brain circuits with balanced excitation and inhibition [26••, 32, 33••, 41]. In contrast, weakly observable initial conditions cause $\mathbf{y}(t)$ to decay away rapidly, or outright not to respond (Figure 2A, bottom right).

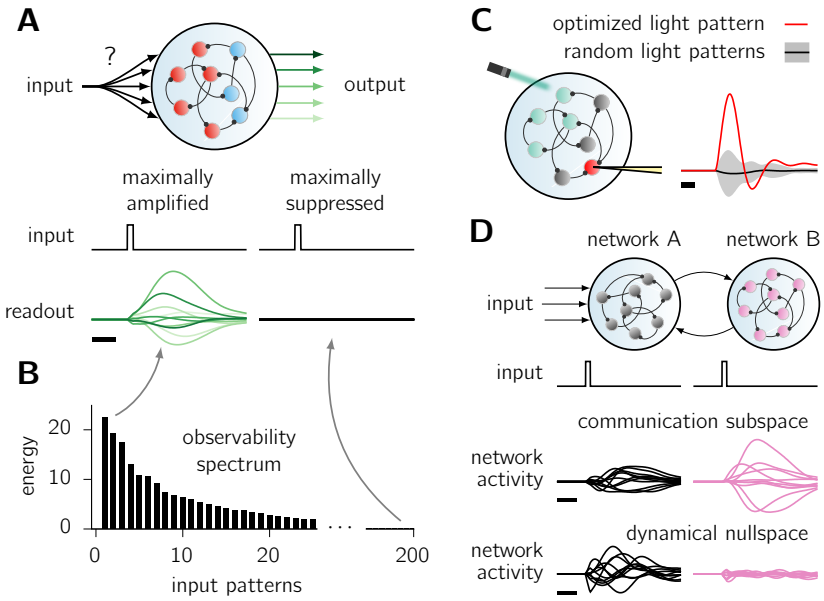

Figure 2: **Network observability. (A)** Interpretation of network observability as input sensitivity. Using simple algebraic techniques based on a neural network's observability Gramian **Q**, one can identify directions in state space (i.e. sets of input weights, black arrows) along which input pulses trigger large transients (bottom left) in the available readouts (shades of green), or, on the contrary, generate no output response at all (bottom right). **(B)** More generally, the (eigenbasis of the) observability Gramian defines a full set of orthogonal initial conditions that can be sorted by how much energy they evoke, from the most to the least. **(C)** Using the observability Gramian, one can determine the optimal way of stimulating a subset of neurons (green) to elicit the strongest possible response in another neuron (red). The optimal response (red) here is much larger than the typical response obtained through random stimulation patterns of the same energy (black; gray shading shows $\pm$ std.). **(D)** Illustration of communication subspaces and dynamical nullspaces, explained in the main text. Scale bars denote 20 ms in all panels.

This connection was first exploited to examine the response properties of high-dimensional, inhibition-stabilized network models of primary motor cortex[26••, 42]. It is also highly relevant to the current view of motor cortical networks as near-autonomous dynamical systems, where initial conditions (a.k.a. preparatory states) largely determine the activity trajectories that follow [10, 24, 43–45]. The critical role of initial states is further corroborated by the success of modern data analysis methods such as LFADS [5••], which use trial-specific initial conditions to accurately predict subsequent neural population activity through near-autonomous, nonlinear latent dynamics.

Optimal drive, robustness to perturbations

Knowing a circuit's sensitivity to inputs can lead to efficient strategies for interacting with it, e.g. to most efficiently drive or suppress some target cells. Figure 2C shows an example where the concept of observability is used to derive the optimal way of stimulating a subset of neurons, so as to elicit maximum response in another, non-stimulated neuron.

Conversely, emergent technologies for optical, random-access perturbations of neural dynamics [46–48] will soon enable the identification of observability spectra in neural circuits. This could be done using empirical techniques [49] similar to the classical reverse correlation-based mapping of receptive fields in sensory cortices. Alternatively, input sensitivity could be inferred by fitting dynamic latent-variable models, which — beyond explaining variance in the recorded data — incorporate the effect of external perturbations. Recently, by fitting such a model to monkey M1 recordings, Duncker et al. could explain the surprising robustness of M1 dynamics to optogenetic perturbations [50]. Their analysis suggested that optogenetic inputs transiently excite a set of weakly observable state-space directions, whose contributions to the momentary activity state tend to decay rapidly. This points to the existence of a "dynamical nullspace" in the circuit.

While weak observability can result from rapidly decaying directions in the full state space (as discussed above), it can also result from state-space trajectories evolving orthogonally to the subspace that defines the network readout (green plane in Figure 1, left). Such output-null dynamics may underlie movement preparation [18] or learning [23] in monkey, where preparatory cortical activity is largely orthogonal to movement-related activity [51]. An analogous phenomenon has also been observed in the premotor cortex of mice, where orthogonality seems achieved through anatomical segregation of preparatory neurons and output-related neurons. Indeed, this brain area contains a class of projection neurons that have a direct influence on movement, and which show little selectivity for the upcoming movement during preparation; these neurons are genetically distinct from other neurons that show strong preparatory selectivity, and are coupled to the thalamus [52].

Interacting brain areas

Beyond local circuits, dynamical nullspaces could also emerge from interactions between distant brain areas. For example, preparatory activity in mouse premotor cortex recovers from unilateral — but not bilateral — optical silencing during a delayed sensory discrimination task [53]. This suggests that it is the interactions between the two hemispheres that underpin the system's weak sensitivity to unilateral perturbation. More generally, input sensitivity may dictate how multiple, interconnected brain areas influence each other. Consider the two coupled networks of Figure 2D. Network B responds strongly when the activity of network A traverses their "communication subspace", i.e. when A's input to B spans B's most observable modes. On the other hand, it is also possible for network A to produce "private" activity fluctuations that do not influence B, so long as A's input falls in B's dynamical nullspace. A recent analysis of the joint dynamics of macaque areas V1 and V2 point to such signal propagation motifs [25]: a small "communication" subspace of V1 activity was found to predict V2 activity, with other "private" dimensions having little influence on V2 activity yet accounting for non-negligible variance in V1 activity. Mechanistically, selective propagation of signals across brain areas could — in theory — arise from a form of detailed excitation/inhibition balance, which only allows specific balance-breaking activity transients to propagate[54, 26●●].

Network controllability

Control-theoretic context

To establish control feasibility, an engineer will examine not only the observability of the system (c.f. above), but also its controllability. Can the state $\mathbf{x}(t)$ of the network be steered along any direction $\mathbf{v}$ in state space, using control inputs $\mathbf{u}(t)$ that can only actuate the network along a restricted set of directions (defined by the columns of $\mathbf{B}$; Figure 1)? Answers can be found in the controllability Gramian [27], a positive definite matrix associated with $\mathbf{A}$ and $\mathbf{B}$ and defined as:

$$\mathbf{P} = \int_0^\infty \exp(t\mathbf{A})\,\mathbf{B}\mathbf{B}^T \exp\left(t\mathbf{A}^T\right) dt \qquad (4)$$

(similar to the observability Gramian in Equation (2), $\mathbf{P}$ can be obtained by solving $\mathbf{A}\mathbf{P} + \mathbf{P}\mathbf{A}^T + \mathbf{B}\mathbf{B}^T = 0$). In particular, the minimum amount of input "energy" required to move the state of the system some distance along direction $\mathbf{v}$ is proportional to $\mathbf{v}^T\mathbf{P}^{-1}\mathbf{v}$; if this quantity is infinite, no clever control strategy will ever succeed in moving $\mathbf{x}$ along $\mathbf{v}$, whereas if it is small, control is easy and cheap. Figure 3A illustrates these ideas in a toy three-dimensional system.

Intrinsic manifolds and control costs

The controllability Gramian $\mathbf{P}$ has a useful interpretation for the analysis of neural circuits: whereas the observability Gramian encodes a network's input sensitivity, $\mathbf{P}$ encodes the "intrinsic manifold" of the network's dynamics, i.e. the directions in state space that the network activity is most inclined to visit. This intrinsic manifold is a reflection of the network's connectivity ($\mathbf{A}$) and input channels ($\mathbf{B}$; Equation (4)). More formally, stimulating each of the available input channels individually results in a collection of state-space trajectories (Figure 3B) whose covariance matrix is, in fact, $\mathbf{P}$. Thus, the quantity $\sigma(\mathbf{v}) = \mathbf{v}^T\mathbf{P}\mathbf{v}$ is the average energy in these trajectories along state-space direction $\mathbf{v}$. In particular, directions with large $\sigma(\mathbf{v})$ (often referred to as "principal components") form the "natural repertoire" of the network, as they contribute a lot to state-space trajectories elicited by unspecific inputs (Figure 3B,C). In contrast, directions with small $\sigma(\mathbf{v})$ are almost never visited, unless the network is specifically driven by strong input patterns (Figure 3A).

The above connection between intrinsic manifold and control gives an interesting perspective on the results of recent brain-computer interface (BCI) experiments, in which monkeys modulate neural activity in a subset of M1 neurons to move a cursor on a screen [19, 21, 22]. As these studies show, how well a monkey can actuate a BCI strongly depends on the mapping from recorded M1 activity to BCI state variables (e.g. cursor velocity) — in other words, performance depends on the state-space directions that M1 activity needs to visit. In particular, monkeys struggle to learn new mappings that require M1 to produce activity outside its intrinsic manifold [19, 55]. In the toy illustration of Figure 3D, where the green plane depicts the intrinsic manifold, that would correspond to cursor velocity being suddenly associated with neural activity along the pink direction. In fact, even when the new mapping requires no such difficult excursions (for example, velocity originally defined by the yellow direction and now defined by the blue direction), the cloud of activity patterns generated within the intrinsic manifold under the new mapping appears similar to the one generated under the previous mapping [21, 22]. This suggests that learning occurs by repurposing a fixed distribution of within-manifold activity patterns, by reassociating each pattern with a new intended cursor movement.

These phenomenona emerge naturally in networks such as the one discussed in Figure 3A, which has a set of highly controllable directions (forming the intrinsic manifold) and a set of poorly controllable ones (orthogonal to the intrinsic manifold). Modulation of neural activity within the intrinsic manifold is straightforward, as it can exploit the natural flow of activity without the need for large control inputs. In contrast, control of activity outside the intrinsic manifold requires much larger inputs, perhaps larger than is physiologically feasible, thus limiting learnability of outside-manifold mappings. Moreover, if inputs to the network are energy-limited, controlling the state of the network along any direction within the intrinsic manifold will likely produce a fixed, common distribution of activity patterns that is determined by the controllability spectrum of the network [56].

Structural controllability

A recent study of the nervous system of C. elegans used an extension of the concept of controllability to investigate the

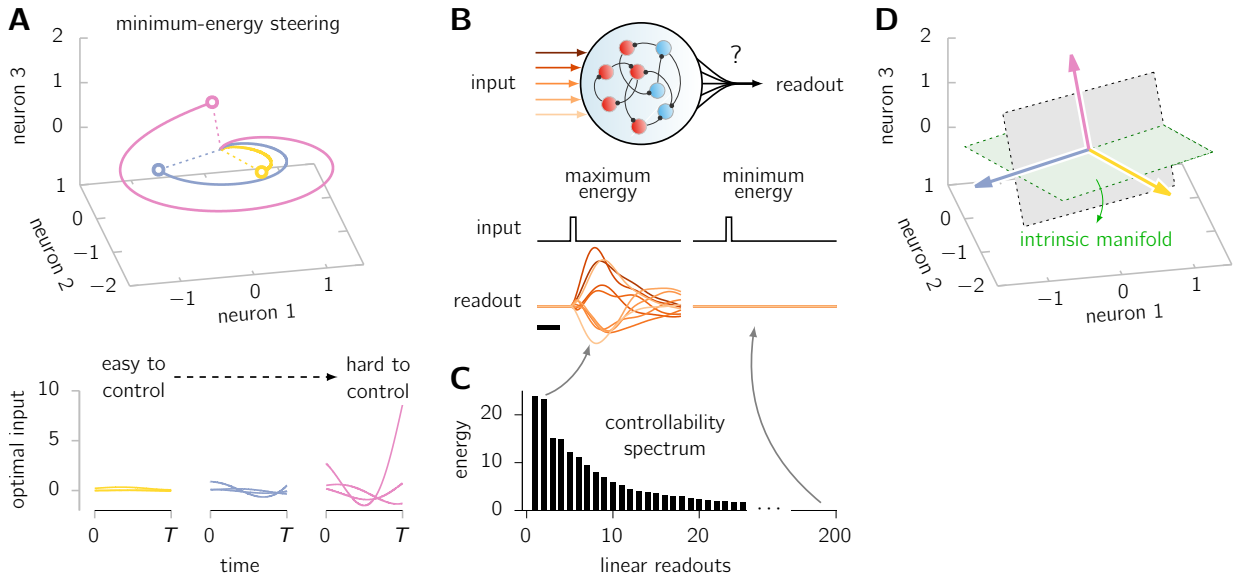

Figure 3: **Network controllability. (A)** Minimum-energy control of a 3-neuron network, which can be easily steered along two directions in state space (orange and blue), but much less easily along a third direction (pink). Solid lines (top) represent state-space trajectories under the action of optimal control inputs (bottom, same color code), ie. those of least "energy" that achieve the target end points (open circles) in finite time T. **(B)** Illustration of controllability for the same inhibition-stabilized network as used in Figure 2A [26]. The controllability Gramian contains information about the amount of variance (or "energy") in the activity of any readout, across all possible input channels (shades of orange) through which the network can be stimulated. In particular, it tells us which readout captures the most (bottom left) or least (right) state variance. The scale bar denotes 20 ms. **(C)** More generally, the (eigenbasis of the) controllability Gramian defines a full set of orthogonal state-space directions that can be sorted by how controllable they are, from the most to the least. **(D)** In BCI experiments, tasks that can be solved by modulating neural activity within the intrinsic manifold (green) can be learned more rapidly than tasks in which activity must span other, non-intrinsic directions (gray).

role of various neuron classes in controlling the nematode's behaviour [57•]. The authors considered the theoretical impact of individually ablating every possible class of neurons on the so-called "structural controllability" of the worm's muscles. They predicted that ablating certain classes of neurons would result in a reduction in the number of controllable muscles. Interestingly, while most of these neuron classes had already been causally implicated in the control of locomotion in previous studies, one neuron class had not been previously investigated. Ablation of these neurons resulted in small but significant motor impairments. Analogous predictions were also made and validated for the ablation of individual neurons within a class of motor neurons, illustrating the promise of applying control-theoretic methods to make causal predictions regarding the neural circuit basis of behaviour.

Model reduction

In control engineering, examination of observability and controllability often culminates in model reduction: the substitution of the original large-scale model with a more tractable, lower-dimensional model. Principled methods exist for performing dimensionality reduction while preserving the dynamic mapping from inputs to outputs. One such method is "balanced truncation" [27], whereby the state space is trimmed to eliminate directions that are both weakly controllable and weakly observable (Figure 4).

Critically, a reduced model is cheaper to simulate, enables computationally efficient control strategies, and is often easier to understand qualitatively. In neuroscience, model reduction has been applied to linearized models of signal propagation along active cables [58]. We speculate it will be useful for reverse engineering brain computations, by analyzing model networks that are either learned from data [5], or trained [6, 7] or hand-crafted [59] to perform a task. For example, trained models often need to be large enough to find solutions to their assigned task, but theses solutions often involve low-dimensional dynamics [17, 20]. Model reduction could help reveal and interpret these dynamics.

Stability and homeostasis

Control theory and neuroscience can also cross-fertilize in seeking to understand how complex, inherently unstable dynamical processes can be controlled and stabilized. Brain circuits contain major sources of dynamical instability that are kept in check by specific regulatory processes. Two prominent examples include i) positive feedback due to the presence of recurrent excitatory connections, which tend to cause runaway activity, and ii) positive feedback in Hebbian learning, causing runaway synaptic potentiation [60]. The control-theoretic methods aimed at stabilizing unstable systems could shed light on how neural circuits achieve stable

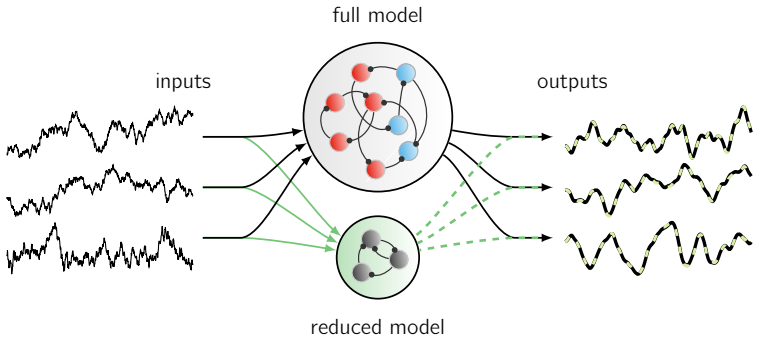

Figure 4: **Model reduction.** An inhibition-stabilized network with 200 neurons (from [26••]) is reduced to a much smaller linear model of dimension 8. Here, the reduced model responds to time-varying inputs in almost exactly the same way as the full model does (compare black and dashed green outputs).

behaviour. For example, even simple control rules such as homeostatic scaling of inhibitory synapses based on post-synaptic activity can restore E/I balance in networks with structural heterogeneity [61]. A more sophisticated form of Hebbian plasticity at inhibitory synapses [41] establishes and maintains stability in networks with strong excitatory feedback [62], and can be understood as an approximation to optimal robust stabilization algorithms [26••].

On a more technical level, elements of systems theory have inspired useful parameterizations of (linear) latent dynamical models that guarantee stability, even when only limited data is available to constrain the system [29•, 36]. Recently, systems theory has also been elegantly combined with statistical approaches to characterize the stability and other aspects of the closed-loop behaviour of networks trained to have multiple fixed points [63•].

Conclusions

The brain solves a variety of hard control problems, which are naturally studied in the framework used by engineers to tackle similar challenges. Control theory has long been applied to study motor control [64, 65], providing insights into the computational and algorithmic principles of internal models [66], error-corrective feedback [67], and planning [68]. Here, we have reviewed concepts that form the foundations of classical control theory and are specifically applicable to the analysis of neuronal network dynamics. They offer simple geometric descriptions of the dynamic input/output behaviour of neural circuits, along with a set of accessible analytical and numerical tools.

While the methods discussed here apply primarily to linear state-space models, they can be adapted to nonlinear systems in various ways, from standard linearization to more advanced operator-theoretic techniques [69]. Further extensions may be necessary for the analysis of recurrent neural networks that may be operating in highly nonlinear regimes. As we have noted, the observability and controllability Gramians afford several equivalent definitions in the linear case; nonlinear extensions will likely not capture all of them [49], and the specific needs of neuroscientists could in fact inspire the development of relevant extensions by the control community.

In reviewing the most basic properties of linear systems (i.e. those covered in the first few chapters of most control theory textbooks [27]), we have only scratched the surface. Control theory is primarily the science of feedback, and will continue to provide unique insights into how feedback is used in the brain to support the functions of single neurons, circuits, and systems. More generally, dissecting the circuit basis of complex computations such as motor control or reinforcement learning will benefit from a deeper integration of control theory, machine learning, and neuroscience.

Acknowledgments

This work was supported by the Wellcome Trust (Seed Award, grant number 202111/Z/16/Z).

Declarations of interest: none

References

[1] X.-J. Wang. "Neural dynamics and circuit mechanisms of decision-making". In: *Cur. Op. Neurobiol.* 22.6 (2012), pp. 1039–1046.

[2] L. Paninski et al. "A new look at state-space models for neural data". In: *J. Comput. Neurosci.* 29.1-2 (2010), pp. 107–126.

[3] Y. Gao et al. "High-dimensional neural spike train analysis with generalized count linear dynamical systems". In: *Adv. Neural Inf. Process. Syst.* 2015, pp. 2044–2052.

[4] M. M. Churchland et al. "Neural population dynamics during reaching". In: *Nature* 487.7405 (2012), pp. 51–56.

[5] C. Pandarinath et al. "Inferring single-trial neural population dynamics using sequential auto-encoders". In: *Nat. Methods* 15 (2018), pp. 805–815. •• A powerful new method for discovering latent dynamics that may underlie neural population activity, and thereby obtaining smooth and compact descriptions of single-trial population spike rasters. The authors show that behavioural variables can be accurately predicted on a trial by trial basis by specifying initial conditions for the learned dynamics.

[6] O. Barak. "Recurrent neural networks as versatile tools of neuroscience research". In: *Cur. Op. Neurobiol.* 46 (2017), pp. 1–6.

[7] D. Sussillo and O. Barak. "Opening the black box: low-dimensional dynamics in high-dimensional recurrent neural networks". In: *Neural Comput.* 25 (2013), pp. 626–649.

[8] D. Sussillo. "Neural circuits as computational dynamical systems". In: *Cur. Op. Neurobiol.* 25 (2014), pp. 156–163.

[9] R. W. Friedrich and G. Laurent. "Dynamic optimization of odor representations by slow temporal patterning of mitral cell activity". In: *Science* 291 (2001), pp. 889–894.

[10] K. V. Shenoy, M. Sahani, and M. M. Churchland. "Cortical control of arm movements: a dynamical systems perspective". In: *Ann. Rev. Neurosci.* 36 (2013), pp. 337–359.

[11] J. A. Gallego et al. "Neural manifolds for the control of movement". In: *Neuron* 94 (2017), pp. 978–984.

[12] A. A. Russo et al. "Motor cortex embeds muscle-like commands in an untangled population response". In: *Neuron* 97 (2018), pp. 953–966.

[13] S. Saxena and J. P. Cunningham. "Towards the neural population doctrine". In: *Cur. Op. Neurobiol.* Machine Learning, Big Data, and Neuroscience 55 (2019), pp. 103–111.

[14] J. P. Cunningham and B. M. Yu. "Dimensionality reduction for large-scale neural recordings". In: *Nat. Neurosci.* 17.11 (2014), pp. 1500–1509.

[15] J. P. Cunningham and Z. Ghahramani. "Linear dimensionality reduction: Survey, insights, and generalizations". In: *J. Mach. Learn. Res.* 16 (2015), pp. 2859–2900.

[16] D. Kobak et al. "Demixed principal component analysis of neural population data". In: *eLife* 5 (2016), e10989.

[17] V. Mante et al. "Context-dependent computation by recurrent dynamics in prefrontal cortex". In: *Nature* 503 (2013), pp. 78–84.

[18] M. T. Kaufman et al. "Cortical activity in the null space: permitting preparation without movement". In: *Nat. Neurosci.* 17.3 (2014), pp. 440–448.

[19] P. T. Sadtler et al. "Neural constraints on learning". In: *Nature* 512 (2014), pp. 423–426.

[20] D. Sussillo et al. "A neural network that finds a naturalistic solution for the production of muscle activity". In: *Nat. Neurosci.* 18 (2015), pp. 1025–1033.

[21] M. D. Golub et al. "Learning by neural reassociation." In: *Nat. Neurosci.* 21 (2018), pp. 607–616.

[22] J. A. Hennig et al. "Constraints on neural redundancy". In: *eLife* 7 (2018), pp. 1–34.

[23] M. G. Perich, J. A. Gallego, and L. E. Miller. "A neural population mechanism for rapid learning". In: *Neuron* 100.4 (2018), pp. 964–976.

[24] E. D. Remington et al. "Flexible sensorimotor computations through rapid reconfiguration of cortical dynamics". In: *Neuron* 98.5 (2018), pp. 1005–1019.

[25] J. D. Semedo et al. "Cortical areas interact through a communication subspace". In: *Neuron* 102 (2019), pp. 1–11.

[26] G. Hennequin, T. P. Vogels, and W. Gerstner. "Optimal control of transient dynamics in balanced networks supports generation of complex movements". In: *Neuron* 82 (June 2014), pp. 1394–1406. •• The authors used methods from control theory to construct inhibition-stabilized networks, optimizing the network's inhibitory connections to optimally stabilize strong recurrent excitatory feedback. A study of observability in these networks revealed initial conditions that evoke rich activity transients reminiscent of monkey M1 population activity during reaching.

[27] S. Skogestad and I. Postlethwaite. *Multivariable feedback control: analysis and design*. Vol. 2. Wiley New York, 2007.

[28] S. Roweis and Z. Ghahramani. "A unifying review of linear Gaussian models". In: *Neural Comput.* 11 (1999), pp. 305–345.

[29] L. Buesing, J. H. Macke, and M. Sahani. "Learning stable, regularised latent models of neural population dynamics". In: *Netw. Comput. Neural Syst.* 23 (2012), pp. 24–47. • A new way of parameterizing linear dynamical systems based on their controllability Gramians, which ensures that these systems do not become unstable when fit to experimental recordings — a common pitfall of many previous approaches.

[30] O. L. White, D. D. Lee, and H. Sompolinsky. "Short-term memory in orthogonal neural networks". In: *Phys. Rev. Lett.* 92 (2004), pp. 1–4.

[31] S. Ganguli, D. Huh, and H. Sompolinsky. "Memory traces in dynamical systems". In: *Proceedings of the National Academy of Sciences* 105 (2008), pp. 18970–18975.

[32] M. S. Goldman. "Memory without feedback in a neural network". In: *Neuron* 61 (2009), pp. 621–634.

[33] B. K. Murphy and K. D. Miller. "Balanced amplification: a new mechanism of selective amplification of neural activity patterns". In: *Neuron* 61.4 (2009), pp. 635–648. •• Beyond providing a simple, mechanistic explanation for the emergence of spatially structured spontaneous fluctuations in cat V1, this paper showed that balanced excitation/inhibition (E/I) networks are generically nonnormal, and selectively amplify small patterns of E/I imbalance into large, balanced activity. In the language of this review, the most observable modes of network activity are orthogonal to the most controllable ones.

[34] H. Ozeki et al. "Inhibitory stabilization of the cortical network underlies visual surround suppression". In: *Neuron* 62.4 (2009), pp. 578–592.

[35] S. Ganguli et al. "One-dimensional dynamics of attention and decision making in LIP". In: *Neuron* 58 (2008), pp. 15–25.

[36] G. Hennequin, L. Aitchison, and M. Lengyel. "Fast sampling-based inference in balanced neuronal networks". In: *Adv. Neural Inf. Process. Syst.* 2014, pp. 2240–2248.

[37] L. Aitchison and M. Lengyel. "The Hamiltonian brain: efficient probabilistic inference with excitatory-inhibitory neural circuit dynamics". In: *PLoS Comput. Biol.* 12 (2016), e1005186.

[38] G. Hennequin et al. "The dynamical regime of sensory cortex: stable dynamics around a single stimulus-tuned attractor account for patterns of noise variability". In: *Neuron* 98 (2018), pp. 846–860.

[39] C. Huang et al. "Circuit models of low-dimensional shared variability in cortical networks". In: *Neuron* 101 (2019), pp. 337–348.

[40] S. Haykin. *Kalman filtering and neural networks*. Vol. 47. John Wiley & Sons, 2004.

[41] G. Hennequin, E. J. Agnes, and T. P. Vogels. "Inhibitory plasticity: balance, control, and codependence". In: *Ann. Rev. Neurosci.* 40 (2017), pp. 557–579.

[42] J. P. Stroud et al. "Motor primitives in space and time via targeted gain modulation in cortical networks". In: *Nat. Neurosci.* 21 (2018), pp. 1774–1783.

[43] M. M. Churchland et al. "Cortical preparatory activity: representation of movement or first cog in a dynamical machine?" In: *Neuron* 68 (2010), pp. 387–400.

[44] J. Wang et al. "Flexible timing by temporal scaling of cortical responses". In: *Nature Neuroscience* 21 (2018), pp. 102–110.

[45] J. J. Paton and D. V. Buonomano. "The neural basis of timing: distributed mechanisms for diverse functions". In: *Neuron* 98 (2018), pp. 687–705.

[46] A. M. Packer et al. "Simultaneous all-optical manipulation and recording of neural circuit activity with cellular resolution in vivo". In: *Nature methods* 12 (2014), pp. 140–146.

[47] V. Emiliani et al. "All-optical interrogation of neural circuits". In: *J. Neurosci.* 35 (2015), pp. 13917–13926.

[48] S. N. Chettih and C. D. Harvey. "Single-neuron perturbations reveal feature-specific competition in V1". In: *Nature* 567 (2019), pp. 334–340.

[49] J. Hahn and T. F. Edgar. "An improved method for nonlinear model reduction using balancing of empirical gramians". In: *Comput. Chem. Eng.* 26 (2002), pp. 1379–1397.

[50] L. Duncker et al. "Low-rank non-stationary population dynamics can account for robustness to optogenetic stimulation". In: *Cosyne, Salt Lake City, UT. T-29.* 2017.

[51] G. F. Elsayed et al. "Reorganization between preparatory and movement population responses in motor cortex". In: *Nature Communications* 7 (2016), pp. 1–15.

[52] M. Economo et al. "Distinct descending motor cortex pathways and their roles in movement". In: *Nature* 563 (2018), pp. 79–84.

[53] N. Li et al. "Robust neuronal dynamics in premotor cortex during motor planning". In: *Nature* 532 (2016), pp. 459–464.

[54] T. P. Vogels and L. Abbott. "Gating multiple signals through detailed balance of excitation and inhibition in spiking networks". In: *Nat. Neurosci.* 12 (2009), pp. 483–491.

[55] E. Wärnberg and A. Kumar. "Perturbing low dimensional activity manifolds in spiking neuronal networks". In: *PLoS Comput. Biol.* 15 (2019), e1007074.

[56] T.-C. Kao and G. Hennequin. "Null ain't dull: new perspectives on motor cortex". In: *Trends Cog. Sci.* 22 (2018), pp. 1069–1071.

[57] G. Yan et al. "Network control principles predict neuron function in the Caenorhabditis elegans connectome". In: *Nature* 550.7677 (2017), pp. 519–523. • The authors used the concept of (structural) controllability to examine the role of specific neurons in controlling locomotion in C. elegans. In particular, they were able to successfully predict whether ablating certain neurons would reduce the number of controllable muscles, or motorneurons, thus leading to behavioural impairment.

[58] F. Gabbiani and S. J. Cox. *Mathematics for neuroscientists*. Academic Press, 2017.

[59] F. Mastrogiuseppe and S. Ostojic. "Linking connectivity, dynamics, and computations in low-rank recurrent neural networks". In: *Neuron* 99 (2018), pp. 609–623.

[60] F. Zenke, W. Gerstner, and S. Ganguli. "The temporal paradox of Hebbian learning and homeostatic plasticity". In: *Cur. Op. Neurobiol.* 43 (2017), pp. 166–176.

[61] I. D. Landau et al. “The impact of structural heterogeneity on excitation-inhibition balance in cortical networks”. In: *Neuron* 92 (2016), pp. 1106–1121.

[62] T. P. Vogels et al. “Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks”. In: *Science* 334 (2011), pp. 1569–1573.

[63] A. Rivkind and O. Barak. “Local dynamics in trained recurrent neural networks”. In: *Phys. Rev. Lett.* 118.25 (2017), p. 258101. • The authors combine mean-field and control theory techniques to study various aspects of closed-loop dynamics in trained recurrent neural networks, including stability and resonance around training-induced fixed points.

[64] D. M. Wolpert and Z. Ghahramani. “Computational principles of movement neuroscience”. In: *Nat. Neurosci.* 3 (2000), pp. 1212–1217.

[65] S. H. Scott. “Optimal feedback control and the neural basis of volitional motor control”. In: *Nature Reviews Neuroscience* 5.7 (2004), pp. 534–546.

[66] M. D. Golub, B. M. Yu, and S. M. Chase. “Internal models for interpreting neural population activity during sensorimotor control”. In: *eLife* 4 (2015), e10015.

[67] E. Todorov and M. I. Jordan. “Optimal feedback control as a theory of motor coordination”. In: *Nat. Neurosci.* 5.11 (2002), pp. 1226–1235.

[68] E. Todorov. “Optimality principles in sensorimotor control”. In: *Nat. Neurosci.* 7.9 (2004), p. 907.

[69] B. Lusch, J. N. Kutz, and S. L. Brunton. “Deep learning for universal linear embeddings of nonlinear dynamics”. In: *Nat. Commun.* 9 (2018), pp. 1–10.