17 Linear Dynamics and Control of Brain Networks

Jason Z. Kim and Danielle S. Bassett

Abstract

The brain is an intricately structured organ responsible for the rich emergent dynamics that support the complex cognitive functions we enjoy as humans. With around 10^{11} neurons and 10^{15} synapses, understanding how the human brain works has proven to be a daunting endeavor, requiring concerted collaboration across traditional disciplinary boundaries. In some cases, that collaboration has occurred between experimentalists and technicians, who offer new physical tools to measure and manipulate neural function. In other contexts, that collaboration has occurred between experimentalists and theorists, who offer new conceptual tools to explain existing data and inform new directions for empirical research. In this chapter, we offer an example of the latter. Specifically, we focus on the simple but powerful framework of linear systems theory as a useful tool both for capturing biophysically relevant parameters of neural activity and connectivity and for analytical and numerical study. We begin with a brief overview of state-space representations and linearization of neural models for non-linear dynamical systems. We then derive core concepts in the theory of linear systems such as the impulse and controlled responses to external stimuli, achieving desired state transitions, controllability, and minimum energy control. Afterward, we discuss recent advances in the application of linear systems theory to structural and functional brain data across multiple spatial and temporal scales, along with methodological considerations and limitations. We close with a brief discussion of open frontiers and our vision for the future.

Keywords

Linear systems theory · Control theory · Dynamical brain networks · Controllability · Simple models

J. Z. Kim
Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA

D. S. Bassett (✉)
Departments of Bioengineering, Electrical and Systems Engineering, Physics and Astronomy, Neurology, and Psychiatry, University of Pennsylvania, Philadelphia, PA, USA

Santa Fe Institute, Santa Fe, NM, USA

17.1 Emergence in the Structure and Function of Complex Systems

In the observable world, some of the most beautiful and most puzzling phenomena arise in physical and biological systems characterized by heterogeneous interactions between constituent elements. For example, in materials physics, heterogeneous interactions between particles in

B. He (ed.), *Neural Engineering*, https://doi.org/10.1007/978-3-030-43395-6_17

granular matter (such as a sand pile) constrain whether the matter acts as a liquid (flowing with gravity) or a solid (supporting load-bearing) [1, 2]. In sociology, heterogeneous interactions between humans in a society are thought to be responsible for surges in online activity, peaks in book sales, traffic jams, and correlated spikes in demand for emergency services [3]. In biology, heterogeneous interactions between computational units in the brain are thought to support a divergence of the correlation length, an anomalous scaling of correlation fluctuations, and the manifestation of mesoscale structure in patterns of functional coupling between units, all features that allow for a diversity of dynamics underlying a diversity of cognitive functions [4, 5]. The feature of these systems that often drives our fascination is the capacity for heterogeneous interactions to produce suprising dynamics, in the form of drastic state transitions, spikes of collective activity, and multiple accessible dynamical regimes.

Because element-element interactions are heterogeneous in such systems, traditional approaches from statistical mechanics – such as continuum models and mean-field approximations – fail to offer satisfying explanations for system function. There exists a critical need to develop alternative approaches to understand how interactions map to emergent behavior. The need is particularly salient in the context of neural systems, where such an understanding could directly inform models of neurological disease and psychiatric disorders [6, 7]. Moreover, gaining such an understanding is a prerequisite for the well-reasoned development of interventions [8], whether in the form of brain stimulation [9, 10], pharmacological agents [11, 12], or other therapies [13]. Technically, such interventions in systems characterized by heterogeneous interactions can be parsimoniously considered as forms of network control, thus motivating extensive recent interest in the utility of network control theory for neural systems [8].

Despite the generic importance of understanding how interactions map to emergent properties, and the specific importance of understanding that mapping in the human brain, progress toward that understanding has remained surprisingly slow. Some efforts have sought to develop detailed multiscale computational models [14]. Yet such efforts are faced with the ever-present quandary that, in point of fact, "the best material model of a cat is another, or preferably the same, cat" [15]. Detailed models are difficult to construct and intractable to analytic approaches, require extensive time to simulate, contain parameters that are frequently underconstrained by experimental data, and in the end produce dynamics that are themselves difficult to understand or to explain from any specific choices in the model. In contrast, approaches from physics consider natural phenomena as if dynamics at macroscopic length scales were almost independent of the underlying, shorter length scale details [16]. A hallmark of effective physical theories is a marked compression of the full parameter space into a few governing variables that are sufficient to describe the observables of interest at the scale of interest. Interestingly, recent theoretical work demonstrates that such simple models are the natural culmination of processes maximizing the information learned from finite data [17].

Here we embrace simplicity by considering the utility of linear systems theory for the understanding and control of neural systems comprised of computational units coupled by heterogeneous interactions. We begin by placing our remarks within the context of quantitative dynamical models of neurons and their interactions, as well as the spatial and temporal considerations inherent in choosing such models. We will then turn to a discussion of approximations to those dynamical models, the incorporation of exogeneous control input, and model linearization. Our treatment then naturally brings us to a discussion of the theory of linear systems, as well as their response to perturbative impulses, and to explicit control strategies. We lay out the formalism for probing state transitions, controllabilty, and the minumum control energy needed for a given state transition. After completing our formal treatment, we discuss the application of linear systems theory to neural systems, and efforts to map network architecture to control properties. We close with a description of several particularly pertinent methodological considerations and limitations, before outlining emerging frontiers.

17.2 Quantitative Dynamical Models of Neural Systems and Interactions

Historically, many neural behaviors and mechanisms have been successfully modeled quantitatively. Here we briefly describe several illustrative examples of such models. The classic fundamental biophysical model of a single neuron (Fig. 17.1, left) was developed by Alan Hodgkin and Andrew Huxley in 1952 (see [18] for details). The model is now known as the *Hodgkin-Huxley* model. It treats a segment of a neuron as an electrical circuit, where the membrane (capacitor) and voltage-gated ion channels (resistors) are parallel circuit elements. The time evolution of membrane voltage, V_m, between the inside and the outside of the neuron is given by

$$C_m \dot{V}_m(t) = \bar{g}_K n^4(t)(V_K - V_m) + \bar{g}_{Na} m^3(t) h(t)(V_{Na} - V_m) + \bar{g}_l (V_l - V_m) + I(t),$$

where C_m is the membrane capacitance; $\bar{g}_K$, $\bar{g}_{Na}$, and $\bar{g}_l$ are maximum ion conductances for potassium, sodium, and passive leaking ions; and I is an external stimulus current, all per unit area. In addition, V_K, V_{Na}, and V_l represent the reversal potential of these ions. The variables n, m, and h vary between 0 and 1 and model the ion channel gate kinetics to determine the fraction of open sodium (m, h) and potassium (n) channels:

$$\dot{n}(t) = \alpha_n(V_m(t))(1 - n(t)) - \beta_n(V_m(t))n(t)$$
$$\dot{m}(t) = \alpha_m(V_m(t))(1 - m(t)) - \beta_m(V_m(t))m(t)$$
$$\dot{h}(t) = \alpha_h(V_m(t))(1 - h(t)) - \beta_h(V_m(t))h(t),$$

where the functions $\alpha_i(V_m)$ and $\beta_i(V_m)$ are empirically determined. These segments are then spatially connected together, such that the propagation of an action potential across a neuron is modeled by a set of partial differential equations. Due to the biophysical realism of variables and parameters, this model can make powerful and accurate predictions of neuron activity in different environments and stimulation regimes [19–21]. Simplified versions of this model, such as the FitzHugh-Nagumo model [22], can also produce many of the same neuronal dynamics.

However, many complex behaviors of neural systems arise from *interactions* between multiple neurons. With four variables (membrane voltage, gates) and even more parameters to model the behavior of a single neuron, the space of models to explore interacting neurons quickly becomes intractable to both analytical and numerical interrogation. An alternative approach is to capture the simplest aspects of neural interactions that are crucial for the phenomenon of interest. Such was the approach taken by Warren McCulloch and Walter Pitts [23], who developed what would later become a canonical model of an artificial neuron. In this model, each neuron i at any point in time t exists in one of two states: firing $x_i(t) = 1$ or not firing $x_i(t) = 0$. The state of the neuron is determined by a weighted sum of inputs from connected neurons j at the previous time step. Then, neuron i in a system of N neurons evolves in time as

$$x_i(t+1) = f_i\left(\sum_{j=1}^{N} w_{ij} x_j(t)\right),$$

where w_{ij} is the strength of excitation ($w_{ij} > 0$) or inhibition ($w_{ij} < 0$) from neuron j to neuron i and function f_i is typically a thresholding function (Fig. 17.1, center). Instantiations and extensions of this model are used to study associative memory (Hopfield [24]), machine learning (perceptron [25]), and cellular automata [26].

In many cases, the sheer number of neurons and interactions renders even these simple models difficult to study. A typical solution is to instead model the average activity of a *population* of neurons. This is the approach taken by Hugh Wilson and Jack Cowan [27] in the *Wilson-Cowan* model. Here, a group of neurons is separated into excitatory and inhibitory populations, where the fraction of cells firing at time t in each population

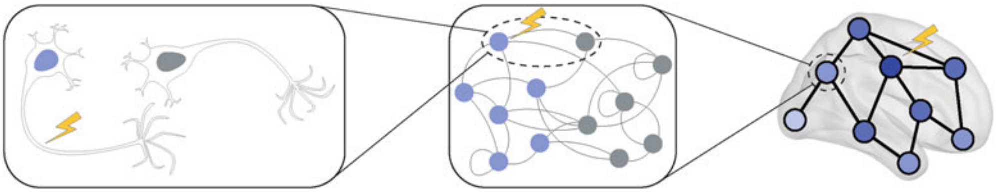

Fig. 17.1 **Schematic of neural models and controlling perturbations at different scales.** Here, the Hodgkin-Huxley model describes the biophysical behavior of single neurons (*left*) that may be excitatory (blue) or inhibitory (gray). The artificial neuron models describe the simplified weighted connections and binary states of many neurons (*center*). The Wilson-Cowan model describes the activity of large neural populations in a region (*right*) or in a cortical column by modeling the excitatory and inhibitory connections of each population. In each case, a controlling perturbation (yellow) can affect the neural system at different scales

is $E(t)$ and $I(t)$, respectively, that evolve in time as

$$\begin{aligned}\tau_e \dot{E}(t) &= -E(t) + (k_e - r_e E(t)) S_e \left(c_1 E(t)\right.\\ &\quad \left. - c_2 I(t) + P(t)\right)\\ \tau_i \dot{I}(t) &= -I(t) + (k_i - r_i I(t)) S_i \left(c_3 E(t)\right.\\ &\quad \left. - c_4 I(t) + Q(t)\right).\end{aligned}$$

Here, $c_1, c_2 > 0$ represent connection strength into the excitatory population, and $c_3, c_4 > 0$ represent connection strength into the inhibitory population, r_e, r_i are the refractory periods, and S_e, S_i are sigmoid functions from the distribution of neuron input thresholds for firing. Such models produce oscillations such as those observed in noninvasive measurements of large-scale brain activity (Fig. 17.1, right) in patients with epilepsy [28].

In these and many other models, a common theme is the tradeoff between realism and tractability. We desire sufficient realism to study crucial features of neural systems such as the activity of each unit, the interaction strength between units, the connection topology, and the effect of external stimulation. We also desire sufficient tractability (either to analytical or numerical interrogation) to make consistent and meaningful predictions about our neural system by understanding relations between the model parameters and the model behavior. In this chapter, we will discuss one such model from the theory of linear dynamical systems.

17.2.1 Spatial and Temporal Considerations

When modeling neural systems, an immediately salient consideration is the vast range of spatial and temporal scales at which nontrivial – and thus quite interesting – dynamics occur. It stands to reason that the most relevant type of model for understanding a given phenomenon depends on the spatiotemporal scale at which that phenomenon is observed. For example, consider the fact that while it is generally known that certain sensory regions such as the visual cortex are both anatomically linked to and functionally responsible for sensory inputs, it is more difficult to assign a set of neurons that are necessary for distributed cognitive processes such as attention and cognitive control. Thus, biophysical models at the level of single neurons may be viable for simulating receptive fields in visual processing, but may be less useful for studies of task-switching or gating. Similarly, consider the fact that a single neuron may fire every few milliseconds, while human reaction times are on the order of hundreds of milliseconds, and brain-wide fluctuations in activity on the order of seconds. Thus, the form of the model considered should match the temporal scales of the behavior to be studied.

From a modeling perspective, balancing these considerations of spatial and temporal scales with model realism impacts the category of model that has the greatest utility. If one wishes to consider small spatial scales, then a rather simplistic neuron-level model such as the McCulloch-Pitts may be particularly useful, where each neural unit has *discrete states* such that each neuron i is either firing $x_i(t) = 1$ or not $x_i(t) = 0$. In contrast, if one wishes to consider larger spatial scales characteristic of distributed cognitive processes, it may be more appropriate to consider models in which each neural unit reflects the average population activity of a brain region as a *continuous state*, where $x_i(t)$ is a real number. Similar considerations are relevant and important in the time domain. For models that assume fairly uniform delays in neuronal interactions such as the McCulloch-Pitts, a *discrete time* model where time evolves in integer increments may be appropriate. In contrast, if the timing of interactions between neural units such as myelinated *versus* unmyelinated axons is heterogeneous, a *continuous time* model may be more suitable, where time t is a real number.

In addition to affecting the definition of neural activity and the nature of its propagation, these considerations also affect the meaning of interactions between units. In a neuron-level model whose units reflect neurons, the unit-to-unit interactions may represent structural synapses between neurons. In contrast, in a population model whose units reflect average neural activity of a brain region, unit-to-unit interactions may represent a summary measure of the collective strength or extent of structural connections between regions. Both types of connections can be empirically measured using either invasive (staining, flourescence imaging, tract tracing [29]) or non-invasive (tractography [30]) methods. The specific type of interaction studied constrains the sorts of inferences that one can draw from the subsequent model, as well as the types of model-generated hypotheses that one can test in new experiments.

17.2.2 Dynamical Model Approximations

Both here and in the following sections, we will consider systems with both continuous state and time. However, we note that the theory of linear systems extends naturally to discrete time systems as well. We begin our formulation with a set of N neural units, where each unit has an associated level of activity $x_i(t)$ that is a real number at some time $t \geq 0$ that is also a real number. Then the collection of activity for all units into column vector $\boldsymbol{x}(t) = [x_1(t); x_2(t); \cdots ; x_N(t)]$ is called the *state* of our system at time t. For example, in the Hodgkin-Huxley equations, our state vector is $\boldsymbol{x} = [V; n; m; h]$. In many models including Hodgkin-Huxley, the time evolution of the system states can be written as a vector differential equation:

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} = \underbrace{\begin{bmatrix} f_1(\boldsymbol{x}(t)) \\ f_2(\boldsymbol{x}(t)) \\ \vdots \\ f_N(\boldsymbol{x}(t)) \end{bmatrix}}_{\boldsymbol{f}(\boldsymbol{x}(t))},$$

where $\boldsymbol{f}$, the vector of functions f_i, determines how the system states change, $\dot{\boldsymbol{x}}$, at every particular state $\boldsymbol{x}$. We can think of these equations as generating a vector field, where at each point $\boldsymbol{x}$, we draw an arrow with magnitude and direction equal to $\boldsymbol{f}(\boldsymbol{x})$. As an example, consider the following two neuron system x_1, x_2 that evolves in time as:

$$\dot{x}_1(t) = 2x_2(t) - \sin(x_1(t))$$
$$\dot{x}_2(t) = x_1^2(t) - x_2(t),$$

where the vector field and example trajectory from initial state $\boldsymbol{x}(0) = [-0.3; -0.4]$ are shown (Fig. 17.2, top). Note how at every point x_1, x_2 the above equation determines a vector of motion $\dot{\boldsymbol{x}}$ that the system traces from the initial point. This quantitative modeling of neural dynamics allows us to study and predict the response of our neural system to changes in interaction strength or external stimulation.

17.2.3 Incorporating Exogenous Control

While modeling intrinsic system behavior is already a broad topic of current research, there is an increasing need for the principled study of therapeutic interventions to correct dysfunctional neural activity. These interventions may take the form of targeted invasive (deep bran stimulation) or noninvasive (transcranial magnetic stimulation) inputs, or more diffusive drug treatments. Hence, in our modeling efforts, we also often desire to incorporate the effect of some external stimuli $u_1(t), \cdots, u_k(t)$. We collect these stimuli into a vector $\boldsymbol{u}(t) = [u_1(t); u_2(t); \cdots; u_k(t)]$ and include their effect on the rates of change of system states in our function:

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} = \underbrace{\begin{bmatrix} f_1(\boldsymbol{x}(t), \boldsymbol{u}(t)) \\ f_2(\boldsymbol{x}(t), \boldsymbol{u}(t)) \\ \vdots \\ f_N(\boldsymbol{x}(t), \boldsymbol{u}(t)) \end{bmatrix}}_{\boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t))}.$$

As an example in our two-unit system, we can apply an input to the first unit

$$\dot{x}_1(t) = 2x_2(t) - \sin(x_1(t)) + u(t)$$
$$\dot{x}_2(t) = x_1^2(t) - x_2(t),$$

thereby changing our system of equations. We plot the vector field and trajectory of our system under some constant input $u(t) = 0.5$ (Fig. 17.2, bottom). Notice how the control input changes the trajectory and final state of our system by modifying the vector field. Also notice that our input only shifts the x_1 component of our vectors because we only stimulate x_1. These abilities to map neural interactions $\boldsymbol{f}$ to the full trajectory of activity $\boldsymbol{x}(t)$ and to find control inputs $\boldsymbol{u}(t)$ that

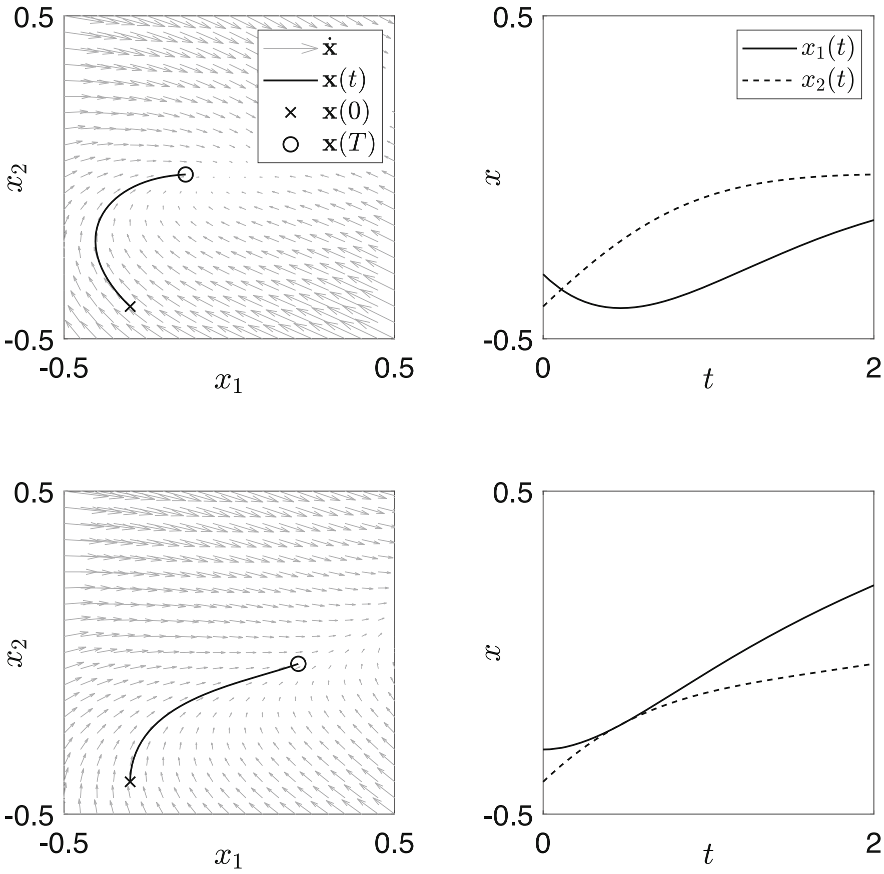

Fig. 17.2 Vector fields and trajectories, with and without control inputs. Example simple vector field of two states with a particular trajectory from initial condition $\boldsymbol{x}(0) = [-0.3; -0.4]$ (*top left*) in state space, with the corresponding plot of each state over time (*top right*) and the corresponding vector field and trajectory with control input $u(t) = 0.5$ (*bottom left*) with corresponding states over time (*bottom right*)

drive our neural system to a desired final state $\boldsymbol{x}(T)$ are among the core contributions of linear systems theory.

17.2.4 Model Linearization

While we have a quantitative framework for the evolution of a controlled neural system, there are no general principles for determining the full trajectory $\boldsymbol{x}(t)$ or control input $\boldsymbol{u}(t)$ to reach a desired final state for a general nonlinear system. In systems of only a few neural units, there exist several powerful numerical and analytic tools. However, the study and control of large neural systems is made difficult by our inability to know how a stimulus will affect our system without first simulating the full trajectory. Further, for multiple stimuli, the number of possible stimulus patterns grows exponentially.

A special class of simplified systems called *linear systems* circumvents this issue. In our state representation, a linear system is described by

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \vdots \\ \dot{x}_N(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} = \underbrace{\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1N} \\ a_{21} & a_{22} & \cdots & a_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ a_{N1} & a_{N2} & \cdots & a_{NN} \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_N(t) \end{bmatrix}}_{\boldsymbol{x}(t)} + \underbrace{\begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1k} \\ b_{21} & b_{22} & \cdots & b_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ b_{N1} & b_{N2} & \cdots & b_{Nk} \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} u_1(t) \\ u_2(t) \\ \vdots \\ u_k(t) \end{bmatrix}}_{\boldsymbol{u}(t)}, \tag{17.1}$$

that is characterized by the time evolution of any state $\dot{x}_i(t)$ being a weighted sum of current states $\sum_{j=1}^N a_{ij}x_j(t)$ and external inputs $\sum_{j=1}^k b_{ij}u_j(t)$. Here, a_{ij} is a real number that determines how activity in state x_j influences the rate of change of state x_i and b_{ij} is a real number that determines how external input u_j influences the rate of change of state x_i. We see that our example two-unit system is *not* linear, because the first state $\dot{x}_1(t)$ depends on $\sin(x_1(t))$, and the second state $\dot{x}_2(t)$ depends on $x_1^2(t)$, and is therefore a *nonlinear* system.

To transform the nonlinear system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u})$, into a linear system $\dot{\boldsymbol{x}} = A\boldsymbol{x} + B\boldsymbol{u}$, we can create an approximate model of our vector field about a particular constant operating state $\boldsymbol{x}^*$ and input $\boldsymbol{u}^*$. We first evaluate the dynamics at this operating point, $\boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{u}^*)$. Then we approximate the vector field along small deviations from this point by computing the derivative of $\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u})$ with respect to the states to get matrix A and with respect to control inputs to get matrix B:

$$A = \left.\begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_N} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_2}{\partial x_N} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial x_1} & \frac{\partial f_N}{\partial x_2} & \cdots & \frac{\partial f_N}{\partial x_N} \end{bmatrix}\right|_{\boldsymbol{x}=\boldsymbol{x}^*, \boldsymbol{u}=\boldsymbol{u}^*} \qquad B = \left.\begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} & \cdots & \frac{\partial f_1}{\partial u_k} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} & \cdots & \frac{\partial f_2}{\partial u_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_N}{\partial u_1} & \frac{\partial f_N}{\partial u_2} & \cdots & \frac{\partial f_N}{\partial u_k} \end{bmatrix}\right|_{\boldsymbol{x}=\boldsymbol{x}^*, \boldsymbol{u}=\boldsymbol{u}^*}.$$

Then, for states near $\boldsymbol{x}^*$ and inputs near $\boldsymbol{u}^*$, the vector field is approximately

$$\dot{\boldsymbol{x}}(t) = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}) \tag{17.2}$$

$$\approx \boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{u}^*) + A(\boldsymbol{x}(t) - \boldsymbol{x}^*) + B(\boldsymbol{u}(t) - \boldsymbol{u}^*). \tag{17.3}$$

A typical operating point for the input is $\boldsymbol{u}^* = 0$ corresponding to no input, because neural stimulation is viewed as a perturbation to the natural and unstimulated dynamics. A typical operating point for the state $\boldsymbol{x}^*$ is a *fixed point* where $\boldsymbol{f}(\boldsymbol{x}^*, \boldsymbol{u}^*) = 0$, because then the evolution of our system Eq. 17.2 only depends on deviations from the point, and not on its actual value. Finally, we can write the linearized equation explicitly as a function of these deviations through a change of variables $\boldsymbol{y}(t) = \boldsymbol{x}(t) - \boldsymbol{x}^*$:

$$\dot{\boldsymbol{y}}(t) = \dot{\boldsymbol{x}}(t) \approx A\boldsymbol{y}(t) + \boldsymbol{B}\boldsymbol{u}(t).$$

We will continue to use variable $\boldsymbol{x}$ instead of $\boldsymbol{y}$ with the understanding that it represents deviations from the fixed point. For example, in our two-unit system, we can linearize about $x_1^* = 0, x_2^* = 0$, and $u^* = 0$ to yield

$$\underbrace{\begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \end{bmatrix}}_{\dot{\boldsymbol{x}}(t)} \approx \underbrace{\begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}}_{\boldsymbol{x}(t)} + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{B} u(t).$$

We show the vector fields and trajectories for both the nonlinear and linear equations without control where $u(t) = 0$ (Fig. 17.3, top) and with control where $u(t) = 0.5$ (Fig. 17.3, bottom) from the same initial condition, and we notice that in the neighborhood of $x_1^* = 0, x_2^* = 0$, the field and trajectories are similar. Hence, by linearizing our neural dynamics about $\boldsymbol{x}^*, \boldsymbol{u}^*$, we

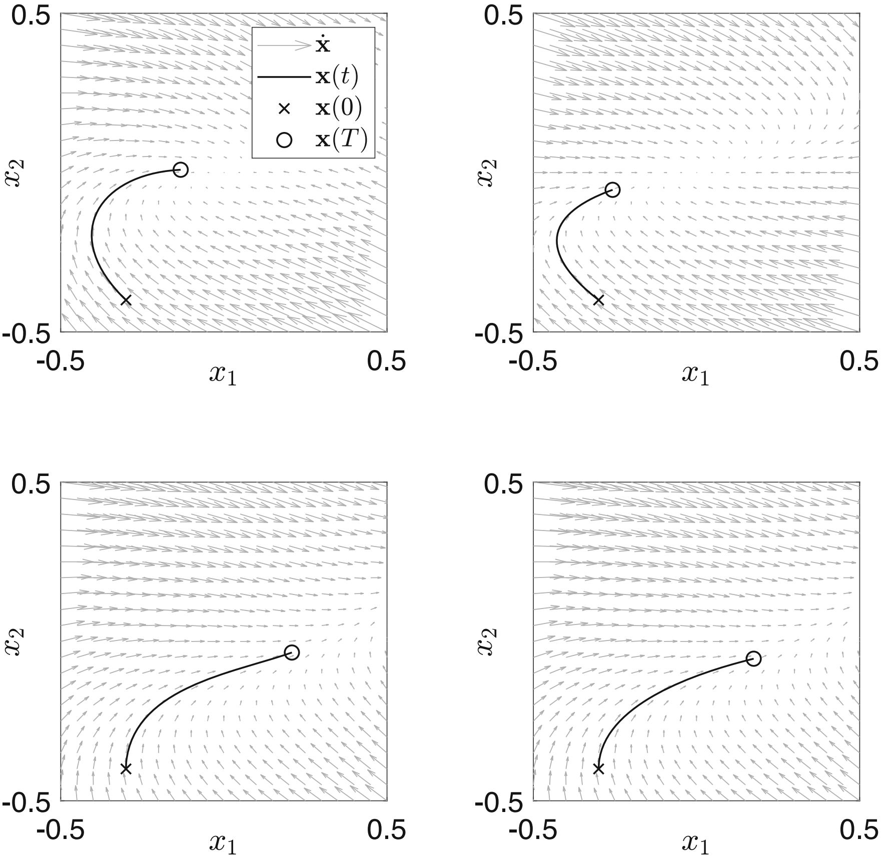

Fig. 17.3 Vector fields and trajectories for a nonlinear system and its linearized form. Example vector field of two states with a particular trajectory from initial condition $\boldsymbol{x}(0) = [-0.3; -0.4]$ for the uncontrolled nonlinear system (*top left*), the uncontrolled linear system (*top right*), the controlled nonlinear system (*bottom left*), and the controlled linear system (*bottom right*)

can preserve the behavior of our neural system at state $\boldsymbol{x}(t)$ and inputs $\boldsymbol{u}(t)$ near this point while enabling the use of powerful tools developed in the next section.

17.3 Theory of Linear Systems

A useful model for therapeutic intervention in a neural system should capture both how the activity over time depends on the connections between neural units and how to change the activity in a desired way through stimulation. Now that we have a model that captures features of neural activity and connectivity in a linearized form, we will develop equations that yield precisely these features. Specifically, we will first determine the system's response to control through mathematical relations as opposed to simulations. Then we will use these principles to design stimuli that optimally guide our system from some initial state $\boldsymbol{x}(0)$ to some final state $\boldsymbol{x}(T)$.

17.3.1 Impulse Response

First, we find the natural evolution of system states from some initial neural state $\boldsymbol{x}(0)$ without any external input. This task amounts to finding the state trajectory $\boldsymbol{x}(t)$ that solves our dynamic equation $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t)$. For scalar systems where $x(t)$ is not a vector, we are reminded of the solution to $\dot{x} = ax$:

$$\begin{aligned}
\frac{dx}{dt} &= ax && \text{differential equation,}\\
\frac{1}{x}dx &= adt && \text{divide by } x,\\
\int \frac{1}{x}dx &= \int adt + c && \text{integrate both sides,}\\
\ln|x| &= at + c && \\
x(t) &= Ce^{at} && \text{solution to differential equation,}
\end{aligned}$$

where the constant is the initial condition $C = x(0)$. We can prove that this solution satisfies $\dot{x} = ax$ by using a Taylor series of the exponential function $e^{at} = \sum_{k=0}^{\infty} \frac{(at)^k}{k!}$. Taking the time derivative of $x(t) = e^{at}$, we see $\dot{x} = ax$:

$$\begin{aligned}
\frac{d}{dt}e^{at} &= \frac{d}{dt}\left(1 + \frac{at}{1!} + \frac{a^2t^2}{2!} + \frac{a^3t^3}{3!} + \cdots + \frac{a^kt^k}{k!} + \cdots\right) && \text{Taylor series of } e^{at},\\
&= 0 + \frac{a}{1!} + 2\frac{a^2t}{2!} + 3\frac{a^3t^3}{3!} + \cdots + k\frac{a^kt^{k-1}}{k!} + \cdots && \text{differentiate each term,}\\
&= a\left(1 + \frac{at}{1!} + \frac{a^2t^2}{2!} + \cdots + \frac{a^kt^k}{k!} + \cdots\right) && \text{factor out scalar } a,\\
&= ae^{at} && \text{substitute Taylor series.}
\end{aligned}$$

A *matrix exponential* is defined exactly the same as above with $e^{At} = \sum_{k=0}^{\infty} \frac{(At)^k}{k!}$, and we again show that the time derivative satisfies the vector relation $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t)$:

$$\begin{aligned}
\frac{d}{dt}e^{At} &= \frac{d}{dt}\left(I + \frac{At}{1!} + \frac{A^2t^2}{2!} + \frac{A^3t^3}{3!} + \cdots + \frac{A^kt^k}{k!} + \cdots\right) && \text{Taylor series of } e^{At}, \\
&= 0 + \frac{A}{1!} + 2\frac{A^2t}{2!} + 3\frac{A^3t^3}{3!} + \cdots + k\frac{A^kt^{k-1}}{k!} + \cdots && \text{differentiate each term,} \\
&= A\left(I + \frac{At}{1!} + \frac{A^2t^2}{2!} + \cdots + \frac{A^kt^k}{k!} + \cdots\right) && \text{factor out matrix } A, \\
&= Ae^{At} && \text{substitute Taylor series.}
\end{aligned}$$

Hence, we see that the following solution

$$\boldsymbol{x}(t) = e^{At}\boldsymbol{x}(0) \tag{17.4}$$

satisfies our dynamic equation. Here, the matrix exponential e^{At} is called the *state transition matrix*, and Eq. 17.4 is called the *impulse response* of our system. Hence, we can find the state at any time T without solving for intermediate states $0 < t < T$.

As an example in our linearized two-unit model, to find the state of our system at $T = 2$ given an initial start at $\boldsymbol{x}(0) = [-0.3; -0.4]$, we can use a software to numerically compute the matrix exponential at time $t = 2$ and multiply by our initial state Eq. 17.4

$$\boldsymbol{x}(2) = e^{2A}\boldsymbol{x}(0) = \begin{bmatrix} 0.1353 & 0.5413 \\ 0 & 0.1353 \end{bmatrix}\begin{bmatrix} -0.3 \\ -0.4 \end{bmatrix} = \begin{bmatrix} -0.2571 \\ -0.0541 \end{bmatrix},$$

which agrees with the simulation results (Fig. 17.3).

17.3.2 Control Response

Next, we derive the system response from an initial state $\boldsymbol{x}(0)$ to some controlling input $\boldsymbol{u}(t)$ through some algebraic manipulation and calculus. We begin with our system equations $\dot{\boldsymbol{x}}(t) - A\boldsymbol{x}(t) = B\boldsymbol{u}(t)$ and multiply both sides by a matrix exponential

$$e^{-At}\dot{\boldsymbol{x}}(t) - e^{-At}A\boldsymbol{x}(t) = e^{-At}B\boldsymbol{u}(t).$$

Next, we see that the left-hand side is the result of a product rule where $\frac{d}{dt}(e^{-At}\boldsymbol{x}(t)) = e^{-At}\dot{\boldsymbol{x}}(t) - Ae^{-At}\boldsymbol{x}(t)$, recalling that functions of matrices can switch orders of multiplication, such that $Ae^{-At} = e^{-At}A$. Hence, we can write our equation as

$$\frac{d}{dt}(e^{-At}\boldsymbol{x}(t)) = e^{-At}B\boldsymbol{u}(t),$$

and integrate both sides from $t = 0$ to $t = T$ to yield

$$e^{-AT}\boldsymbol{x}(T) - \boldsymbol{x}(0) = \int_0^T e^{-At}B\boldsymbol{u}(t)dt.$$

We note the matrix exponential at $t = 0$ becomes $e^{-A\cdot 0} = I$ from the Taylor series. Next, we move the initial state $\boldsymbol{x}(0)$ to the right-hand side and multiply by e^{AT}:

$$e^{AT}e^{-AT}\boldsymbol{x}(T) = e^{AT}\boldsymbol{x}(0) + e^{AT}\int_0^T e^{-At}B\boldsymbol{u}(t)dt.$$

Finally we use the fact that e^{AT} and e^{-AT} are inverses of each other where $e^{AT}e^{-AT} = I$, and we bring e^{AT} into the integral to derive the system's response to control input:

$$\boldsymbol{x}(T) = \underbrace{e^{AT}\boldsymbol{x}(0)}_{\text{natural}} + \underbrace{\int_0^T e^{A(T-t)} B\boldsymbol{u}(t)dt}_{\text{controlled}}. \tag{17.5}$$

Intuitively, we see that the first part of the response, $e^{AT}\boldsymbol{x}(0)$, is just the natural evolution of our system from an initial state and that the second part of the response is a convolution of our mapped inputs, $B\boldsymbol{u}(t)$, with the impulse response. We will next take advantage of the convolution's property of linearity to draw powerful relations between the state evolution, control input, and system structure.

17.3.3 Linear Relation Between the Convolution and Control Input

Previously, we focused on the evolution of a neural system in response to a known control input $\boldsymbol{u}(t)$ in Eq. 17.5. However, our goal is to design a control input that drives our neural system to some desired final state that may stabilize an epileptic seizure [31], or aid in memory recall [32]. In this scenario, we fix the initial state $\boldsymbol{x}(0) = \boldsymbol{x}_0$ and the final state $\boldsymbol{x}(T) = \boldsymbol{x}_T$ as constants and rewrite Eq. 17.5 to move the variables $\boldsymbol{u}(t)$ to the left-hand side and the constants to the right-hand side:

$$\int_0^T e^{A(T-t)} B \underbrace{\boldsymbol{u}(t)}_{\text{variable}} dt = \underbrace{\boldsymbol{x}(T) - e^{AT}\boldsymbol{x}(0)}_{\text{constant}}.$$

This formulation is a linear equation with a structure that is similar to a typical system of linear equations used in regression, $M\boldsymbol{v} = \boldsymbol{b}$, where $\boldsymbol{v}$ is the variable, $\boldsymbol{b}$ is a constant vector, and matrix M is the linear function acting on $\boldsymbol{v}$. Here, the control input $\boldsymbol{u}(t)$ is the variable, $\boldsymbol{x}(T) - e^{AT}\boldsymbol{x}(0)$ is the constant vector, and the convolution

$$\mathcal{L}(\boldsymbol{u}(t)) = \int_0^T e^{A(T-t)} B\boldsymbol{u}(t)dt$$

is the linear function acting on our control inputs. By linear function, we mean that for two control inputs $\boldsymbol{u}_1(t)$ and $\boldsymbol{u}_2(t)$, if $\mathcal{L}(\boldsymbol{u}_1(t)) = \boldsymbol{c}_1$, and $\mathcal{L}(\boldsymbol{u}_2(t)) = \boldsymbol{c}_2$, then a weighted sum of inputs yields the same weighted sum of outputs, such that

$$\mathcal{L}(a\boldsymbol{u}_1(t) + b\boldsymbol{u}_2(t)) = a\boldsymbol{c}_1 + b\boldsymbol{c}_2. \tag{17.6}$$

This linearity allows us to treat solutions to our control function problem the same as solutions to our linear system of equations. Specifically, suppose the control input $\boldsymbol{u}^*(t)$ is a *particular solution* to our control problem such that $\mathcal{L}(\boldsymbol{u}^*(t)) = \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0$. Further, suppose that inputs $\boldsymbol{u}_1(t), \boldsymbol{u}_2(t), \cdots$ are *homogeneous solutions* such that $\mathcal{L}(\boldsymbol{u}_i(t)) = \boldsymbol{0}$. If we construct a control input that is the particular solution added to a weighted sum of homogeneous solutions

$$\boldsymbol{u}(t) = \underbrace{\boldsymbol{u}^*(t)}_{\text{particular}} + \sum_i a_i \underbrace{\boldsymbol{u}_i(t)}_{\text{homogeneous}},$$

then the convolution of this combined input yields the desired output:

$$\begin{aligned}
\mathcal{L}(\boldsymbol{u}(t)) &= \mathcal{L}\left(\boldsymbol{u}^*(t) + \sum_i a_i\boldsymbol{u}_i(t)\right) \\
&= \mathcal{L}(\boldsymbol{u}^*(t)) + \sum_i \mathcal{L}(a_i\boldsymbol{u}_i(t)) \\
&= \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0 + \sum_i a_i\boldsymbol{0} \\
&= \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0.
\end{aligned}$$

Hence, if we have a particular control input $\boldsymbol{u}^*(t)$ that drives our system to a desired final state, then the homogeneous control inputs $\boldsymbol{u}_i(t)$ give us the flexibility to design less costly, more energy-efficient inputs.

17.3.4 Controllability

For any system, we would first like to know if a particular solution exists to the control problem described above. A system is *controllable* if there is a control input that brings our system from any initial state to any final state in finite time. For nonlinear systems, if we know that the input $\boldsymbol{u}^*(t)$

brings our system from the initial state $\mathbf{0}$ to some final state $\boldsymbol{x}_T$, there is in general no way to know what input will take our system to a scaled final state $a\boldsymbol{x}_T$.

In contrast, due to the linearity of our convolution operator, we know that a scaled input $a\boldsymbol{u}^*(t)$ will produce a scaled output $\mathcal{L}(a\boldsymbol{u}^*(t)) = a\boldsymbol{x}_T$. Further, any N-dimensional vector can be written as a weighted sum of N *linearly independent* vectors $\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_N$. Here, linear independence means that no vector $\boldsymbol{v}_i$ in the set can be written as a weighted sum of the remaining vectors $\boldsymbol{v}_{j\neq i}$. For example, a column vector $\boldsymbol{a} = [a_1; a_2; \cdots; a_N]$ can be written as the weighted sum

$$\underbrace{\begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{bmatrix}}_{\boldsymbol{a}} = a_1 \underbrace{\begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{v}_1} + a_2 \underbrace{\begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}}_{\boldsymbol{v}_2} + \cdots + a_N \underbrace{\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}}_{\boldsymbol{v}_N},$$

where none of the vectors $\boldsymbol{v}_i$ can be written as a weighted sum of remaining vectors $\boldsymbol{v}_{j\neq i}$. Hence, our system is controllable if we can find input functions $\boldsymbol{u}_1(t), \cdots, \boldsymbol{u}_N(t)$ that reach N linearly independent vectors $\mathcal{L}(\boldsymbol{u}_1(t)), \cdots, \mathcal{L}(\boldsymbol{u}_N(t))$, because then we can always reach any final state from any initial state through the weighted sum

$$\underbrace{\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0}_{\boldsymbol{a}} = a_1 \underbrace{\mathcal{L}(\boldsymbol{u}_1(t))}_{\boldsymbol{v}_1} + a_2 \underbrace{\mathcal{L}(\boldsymbol{u}_2(t))}_{\boldsymbol{v}_2} + \cdots + a_N \underbrace{\mathcal{L}(\boldsymbol{u}_N(t))}_{\boldsymbol{v}_N},$$

through the control input $\boldsymbol{u}(t) = a_1\boldsymbol{u}_1(t) + a_2\boldsymbol{u}_2(t) + \cdots + a_N\boldsymbol{u}_N(t)$. This information of reachable states is encoded in the *controllability matrix*

$$C = \left[B,\ AB,\ A^2B,\ \cdots,\ A^{N-1}B\right], \quad (17.7)$$

where the *rank* of this matrix (given by the number of linearly independent columns of C) tells us how many of these N independent vectors can be reached using control input. If this rank $= N$, then the system is controllable and can reach all states. However, even if the rank $< N$, there still exists a control input that drives the system from $\boldsymbol{x}_0$ to $\boldsymbol{x}_T$ if the vector $\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0$ can be written as a weighted sum of the columns of C. This set of vectors spanned by the columns of C is called the *controllable subspace* and the remaining set of vectors the *uncontrollable subspace*.

As an example in our linearized two-unit system, A, B, and C are written as

$$A = \begin{bmatrix} -1 & 2 \\ 0 & -1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 \\ 0 \end{bmatrix},$$

$$C = \left[B, AB\right] = \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix},$$

which is *not* controllable, because the rank of C is 1. To consider the controllable subspace, notice that the columns of C only have non-zero entry in the first row. Hence, the controllable subspace contains any desired value of $x_1(T)$, but excludes all values of $x_2(T)$. Intuitively, this loss of controllability arises because x_2 does not receive an input, nor is it affected by x_1. Hence, there is no way to influence the activity of x_2 in a desired way.

17.3.5 Minimum Energy Control

Once we know a system is controllable, we would like to determine the control input function $\boldsymbol{u}(t)$ that transitions our system from initial $\boldsymbol{x}_0$ to final $\boldsymbol{x}_T$ states. However, there are often limitations on the input magnitude such as electrical and thermal damage of neural tissue or battery life of chronic implanted stimulators. Due to the system's linearity, we can find not only an input function but an optimal one $\boldsymbol{u}^*(t)$ that minimizes input cost.

First, we must define a measure of the *size* of our control input functions $\boldsymbol{u}(t)$. In many applications of electrical stimulation, the cost of control scales quadratically with the input, such as with resistive heating. This quadratic measure of size is mathematically and intuitively defined using the *inner product*. For N-dimensional column vectors of numbers, $\boldsymbol{a}$, the inner product is the well known *dot product*

$$< \boldsymbol{a}, \boldsymbol{a} >= a_1^2 + a_2^2 + \cdots + a_N^2 = \boldsymbol{a}^\top \boldsymbol{a},$$

where $\boldsymbol{a}^\top$ is the transpose that turns column vector $\boldsymbol{a}$ into a row vector. We see that doubling $\boldsymbol{a}$ will quadruple the inner product. For k-dimensional column vectors of functions $\boldsymbol{a}(t)$ from time $t = 0$ to $t = T$, the inner product is similarly defined as

$$< \boldsymbol{a}(t), \boldsymbol{a}(t) >= \int_0^T a_1^2(t) + a_2^2(t) + \cdots + a_N^2(t) dt = \int_0^T \boldsymbol{a}(t)^\top \boldsymbol{a}(t) dt$$

that has the same quadratic relation. Hence, we define the *control energy* as

$$E =< \boldsymbol{u}(t), \boldsymbol{u}(t) > . \qquad (17.8)$$

Now that we have a measure of how large an input is, we wish to find a minimal input $\boldsymbol{u}^*(t)$ that minimizes the control energy. This task is analogous to a typical linear system of equations, $M\boldsymbol{v} = \boldsymbol{b}$, where we want to find $\boldsymbol{v}^*$ that solves the equation with the smallest cost $< \boldsymbol{v}^*, \boldsymbol{v}^* >$. Here, if M has full row rank where the rows of M are linearly independent, then the minimum solution is given by the equation for least squares $\boldsymbol{v}^* = M^\top (MM^\top)^{-1}\boldsymbol{b}$. Here, $M^\top$ is the transpose, or *adjoint* of M.

This same principle holds for our linear system $\mathcal{L}(\boldsymbol{u}(t)) = \boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0$, where we want to find $\boldsymbol{u}^*(t)$ that solves the equation with the smallest cost $< \boldsymbol{u}^*(t), \boldsymbol{u}^*(t) >$. However, while matrix M inputs a vector of numbers $\boldsymbol{v}$ and outputs a vector of numbers $\boldsymbol{b}$, our linear function $\mathcal{L}$ inputs a vector of *functions* and outputs a vector of numbers. Hence, we need to carefully define the adjoint of $\mathcal{L}$; because $\mathcal{L}$ is not a finite matrix, we cannot use $\mathcal{L}^\top$ to denote the adjoint. Instead, we will use $\mathcal{L}^*$ to denote the adjoint of $\mathcal{L}$. In the case of matrix M, the adjoint preserves the inner product between inputs and outputs such that

$$< M\boldsymbol{v}, \boldsymbol{b} > =< \boldsymbol{v}, M^\top \boldsymbol{b} >$$
$$(M\boldsymbol{v})^\top \boldsymbol{b} = \boldsymbol{v}^\top (M^\top \boldsymbol{b}).$$

Identically, for state transition $\boldsymbol{x} = e^{AT}\boldsymbol{x}_0 - \boldsymbol{x}_T$, the adjoint of $\mathcal{L}$ preserves the inner product between the vectors of input functions $\boldsymbol{u}(t)$ and output numbers $\boldsymbol{x}$ as

$$< \mathcal{L}(\boldsymbol{u}(t)), \boldsymbol{x} > =< \boldsymbol{u}(t), \mathcal{L}^*(\boldsymbol{x}) >$$
$$\left(\int_0^T e^{A(T-t)} B\boldsymbol{u}(t) dt\right)^\top \boldsymbol{x} = \int_0^T \boldsymbol{u}^\top(t)(B^\top e^{A^\top (T-t)}\boldsymbol{x}) dt.$$

Notice that the inner product on the left is over vectors of numbers, while the inner product on the right is over vectors of functions. Then, we see that our adjoint is

$$\mathcal{L}^*(\boldsymbol{x}) = B^\top e^{A^\top(T-t)}\boldsymbol{x}$$

and takes as input a vector of numbers and outputs a vector of functions. Then, just as our system $M\boldsymbol{v} = \boldsymbol{b}$, the minimum input $\boldsymbol{u}^*(t)$ is given by

$$\boldsymbol{u}^*(t) = \mathcal{L}^*(\mathcal{L}\mathcal{L}^*)^{-1}(\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0). \quad (17.9)$$

Finally, through substitution into Eq. 17.8, we can write the minimum control energy as

$$E_{\min} = (\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0)^\top(\mathcal{L}\mathcal{L}^*)^{-1}(\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0). \quad (17.10)$$

In conclusion, we point out the crucially important term of the minimum energy, $\mathcal{L}\mathcal{L}'$, as the *controllability Gramian* written as

$$W_c(T) = \mathcal{L}\mathcal{L}^* = \int_0^T e^{A(T-t)}BB^\top e^{A^\top(T-t)}dt. \quad (17.11)$$

First, we notice that this Gramian is only a function of the underlying neural relationships, A; the matrix determining where the inputs are placed, B; and time T. Next, we notice that $W_c(T)$ is actually an $N \times N$ matrix and can therefore be numerically evaluated and analytically studied. Finally, we see that if our system begins at an initial state of $\boldsymbol{x}_0 = \boldsymbol{0}$, then the minimum energy can be written as

$$E_{\min} = \boldsymbol{x}_T^\top W_c^{-1}(T)\boldsymbol{x}_T,$$

where the role of neural interactions and stimulation parameters on our ability to control the system is fully encapsulated in the Gramian. This ability to decouple the states $\boldsymbol{x}_T$ from the neural interactions and stimulation parameters A, B, T is a powerful tool for studying and designing control properties of neural systems.

17.4 Mapping Network Architecture to Control Properties

By formulating our neural system in a linear way, we can solve difficult problems such as predicting the system's response to control, finding the set of states that the system can reach, and designing efficient input stimuli, without the need to try every control input and simulate every trajectory. Further, by directly mapping control properties to neural activity and network architecture in an algebraic way, we can study how features of interaction patterns impact our ability to control neural activity [8]. As an active area of research, the variety of questions being asked and systems being studied is very large, and require simultaneous innovations in experiment, computation, and theory. In this section, we will describe a few recent applications and advances.

17.4.1 Neuronal Control in Model Organisms

While most neural systems are too large to empirically measure activity and connectivity or to analyze numerically, there do exist a few sufficiently simple model organisms. Among these is the worm *Caenorhabditis elegans* [33] with several hundred neurons that can be recorded from simultaneously [34]. Even for such a small system, it is difficult to map the functional form of how activity in neuron i affects the activity in neuron j. However, the presence or absence of connections between neurons in this organism, and by consequence the presence or absence of elements in the connectivity matrix A, is well known.

Advances in the study of *structural controllability* [35] allow us to ask questions about our ability to control a system given only the binary presence or absence of edges. Colloquially, this framework focuses on connectivity matrices A where non-zero entries can only exist in the presence of binary edges, and can be used to determine whether the system is controllable for *most* values where an edge is present. Using this framework, recent work has sought to determine whether the removal of certain neurons in *C. elegans* will reduce structural controllability [36]. Specifically, the modeling involves input to the sensory receptor neurons as the control input that is mapped to the system through a matrix B and the connectivity between neurons and muscle cells through a matrix A. Further, instead of recording the activity of each neuron, the motion of muscles was recorded. This framework involves the appended control framework

$$\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$$
$$\boldsymbol{y}(t) = C\boldsymbol{x}(t),$$

where $\boldsymbol{y}(t)$ represents the states (muscles) that are measured and C is the map from neurons and muscles $\boldsymbol{x}(t)$ to the measured output [37]. Here, the authors find that the ablation of a neuron not previously implicated in motion, PDB, decreased structural controllability, significantly reducing ventral bias in deep body bends in *C. elegans*.

17.4.2 State Transitions in the Human Brain

While neuron-level structural synapses map most directly to functional relationships between neurons, there are also well-characterized structural connections between larger-scale brain regions. These connections contain thick bundles of myelinated axonal fibers that run throughout the brain and are thought to play a crucial role in coupling the activity of distant brain regions [38]. These fibers are resolved by measuring water diffusion throughout the brain using magnetic resonance [39] and tracing fibers along this diffusion field using computational algorithms [30]. The whole brain is typically divided into hundreds to thousands of discrete brain regions using a variety of parcellation schemes [40, 41], and the strength of fibers between these regions comprises the connectivity matrix A [42].

Such region-level study of brain dynamics has led to the discovery of macroscopic functional organization in the human brain at rest [43] and during various cognitively demanding tasks [44]. Here, brain activity can be empirically measured through methods such as magnetic resonance imaging (blood oxygen level dependent) or electrophysiology (aggregate electrical activity). Of particular interest are large-scale functional brain networks that display stereotyped changes in activity patterns during tasks that demand certain cognitive or sensorimotor processes [45]. Here, it is thought that the brain uses underlying structural connections to support circuit-level coordination, as well as to guide itself to specific patterns of activity using *cognitive control* [46, 47].

Recent work has begun formulating cognitive control as a linear systems problem [46, 48–51], where matrix A is the network of white matter connections between brain regions, B represents the regions that were chosen to be responsible for control, and $\boldsymbol{x}(t)$ represents the activity of each region over time. Specifically in [48, 50], the authors quantify cognitive states as vectors corresponding to activity in the brain regions during cognitive tasks and compute the minimum control energy Eq. 17.9 to transition between cognitive states for various sets of control regions. Colloquially, if a set of regions requires less input energy to transition between cognitive states, then those regions may easily transition the whole brain between these states along an optimal trajectory given they are responsible for cognitive control. Moreover, individual differences in the minimal control energy are correlated with individual differences in performance on cognitive control tasks [52]. In complementary studies, individual differences in controllability statistics calculated for distinct regions of the brain are correlated with individual differences in measures of cognitive control assessed with common neuropsychological test batteries [49, 51].

17.5 Methodological Considerations and Limitations

While the theory of linear systems is a powerful quantitative framework for studying and controlling dynamical neural systems, there are several important caveats. Here we mention three: dimensionality and numerical stability, model validation and experimental data, and the assumption of linearity.

17.5.1 Dimensionality and Numerical Stability

The benefit of studying linear systems is that we take difficult and largely intractable questions of controllability and control input design and greatly simplify them into algebraic problems of computing objects like the controllability matrix Eq. 17.7 and the controllability Gramian Eq. 17.11. However, these matrices scale quadratically with the number of neural units, and numerical calculations and manipulations using these matrices quickly face computational issues.

Most viable approaches to dealing with these issues involve numerically representing the elements of our matrices and performing algebraic operations. However, these representations are imperfect, as it is impossible to completely represent irrational numbers such as π. Hence, the matrices are truncated to *numerical precision*, and this truncation error propagates with each computation. Further, the propagation of error tends to scale faster than the number of dimensions. This issue is prevalent in the computation of the state-transition matrix [53], as well as in the calculation of the controllability Gramian and its inverse. With the application of this theory to high-dimensional neural systems, the study of useful controllability metrics is an active area of research [54].

17.5.2 Model Validation and Experimental Data

A fundamental limitation for modeling any neural system is the ability to empirically and accurately measure model parameters and variables. A crucial parameter is the network of connectivity encoded by our adjacency matrix A, where the element in the i-th column and j-th row models the effect of unit i on the rate of change of unit j. While we typically use the structural connections in synapses between neurons, or bundles of axons between brain regions as a proxy for A, it is very difficult to measure the true functional effect that activity in unit i has on activity in unit j, particularly for large systems. This problem is exacerbated by further methodological limitations such as the inability to resolve directionality of connections in diffusion tractography. Along these lines, many statistical and autoregressive methods have been developed to infer functional relationships from recordings of neural activity [55–59] and to use that inferred activity to better understand control [60]. However, the degree of causality in these methods as measured by true response to external stimuli remains controversial.

Another such fundamental limitation is our inability to fully measure every state of the system. The state-space representation of our model requires that every state is observed. However, it is impossible to simultaneously record the activity of every neuron in almost all biological systems, although this recording has been achieved in sufficiently simple organisms [34]. As a result of only being able to observe a small subset of the full state-space, these models of interactions may become largely descriptive and phenomenological in nature. In response, there is a continuing effort to improve the spatial and temporal resolution of neuroimaging methods [61].

17.5.3 Assumption of Linearity

An inherent limitation is the lack of generality in our linear approximation of the full nonlinear neural dynamics. In response, there is a sizable quantity of research studying the control properties of nonlinear dynamical systems [62]. An interesting bridge between these two disciplines exists in the theory of the Koopman or composition operator [63]. The underlying benefit of this theory is that, while our system of equations may evolve nonlinearly in time given the current set of N states, there may exist a higher-dimensional set of $M > N$ state variables in which the dynamical system does evolve linearly [64]. While the extension of linear systems theory to actually controlling this higher-dimensional system may be limited, it remains a promising future area of research.

17.6 Open Frontiers

Many exciting and open frontiers exist in the study of brain network dynamics using linear systems theory. Here we constrain our remarks to three main topic areas, but freely admit that this discussion is far from comprehensive. First, we describe opportunities in the further development of useful controllability statistics as well as in the development of foundational theory linking control profiles to the system's underlying network architecture. Second, we underscore the need for a better understanding of how control is implemented in the brain, how control strategies might depend on context, and how control processes could facilitate the effective manipulation of information. Third, we describe the relevance of the modeling efforts we discussed here for our understanding of neurological disease and psychiatric disorders as well as the development of personalized and targeted therapeautic interventions for alterations in mental health.

17.6.1 Theory and Statistics

Linear systems theory has its basis in a rich literature stemming from now well-developed areas of mathematics, physics, and engineering [65]. Yet, much is still unknown about exactly how the network topology of a given unit-to-unit interaction pattern impacts the capacity for control, the trajectories accessible to the systems, and the minimum control energy. Some preliminary efforts have begun to make headway by using linear network control theory to derive accurate closed-form expressions that relate the connectivity of a subset of structural connections (those linking driver nodes to non-driver nodes) to the minimum energy required to control networked systems [66]. Further work is needed to gain an intuition for the role of higher-order structures (e.g., cycles) in the control of the networked system and any dependence on edge directionality [67]. Moreover, it would be fruitful in the future to further develop a broader set of controllablity statistics, extending beyond node controllability [54], and edge controllability [68], to the control of motifs [69]. Finally, throughout such investigations, it will be useful to understand which features of control are shared across networks with various topologies, versus those features which are specific to networks with a particular topology [70–72].

17.6.2 Context, Computations, and Information Processing

Despite the emerging appreciation that linear systems theory has considerable utility in the study of cognitive function, we still know very little about exactly how control is implemented in the brain, across spatial scales, and capitalizing on the unit-to-unit interaction patterns at each of those scales. Some initial evidence suggests that features of synaptic connectivity – and particularly autaptic connections – can serve to tune the excitability of the neural circuit, altering its controllability profile and propensity to display synchronous bursts of activity [73]. Complementary evidence also at the cellular scale demonstrates how intrinsic network structure and exogeneous stimulus patterns together determine the manner in which a stimulus propagates through the network, with important implications for cognitive faculties that require persistent activation of neuronal patterns such as working memory and attention [74]. There are interesting similarities between these observations and evidence at larger spatial scales, which suggests that the architecture of white matter tracts connecting brain areas can be used to infer the probability with which the brain persists in certain states [75]. Such conceptual similarities motivate concerted efforts to better understand how the architecture of brain networks across spatial scales supports information processing and cognitive computations and how those processes and computations might depend on the context in which the brain is placed. Formally, it would be interesting to consider context as a form of exogeneous input to the system, in a manner reminiscent of how we currently consider brain stimulation [8]. We speculate that such a formulation of the problem could help to explain a range of observations, such as the ability of cognitive effort to suppress epileptic activity [76].

17.6.3 Disease and Intervention

The fact that controllability can depend on network topology [66, 70] and can be altered by edge pruning [77] suggests that it might also be a useful biomarker in some neurological diseases and psychiatric disorders, many of which are associated with changes in the structural topology of neural circuitry at various spatial scales [6, 7]. Indeed, recent studies have reported differences in controllability statistics estimated in brain networks of patients with bipolar disorder [78], temporal lobe epilepsy [79], and mild traumatic brain injury [50]. In a complementary line of work, studies are beginning to ask whether the altered controllability profiles of brain networks in these patients could help to inform the development of more targeted interventions for their illness, in the form of brain stimulation [31, 80], pharmacological agents, or cognitive behavioral therapy. Other efforts have begun to consider symptoms of a given disease as a network and to identify symptoms predicted to have high impulse response in the patient's daily life [81]. It would be interesting in the future to determine whether the linear systems approach could be useful in more carefully formalizing that problem as a network control problem, which in turn could be used to determine which symptom to treat in order to move the entire symptom network toward a healthier state [82].

Homework

1. Linearize the following system about point $x_1^* = 1, x_2^* = -1, x_3^* = 0$,
$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\end{bmatrix} = \begin{bmatrix}-x_1^2(t) - 2x_2(t) + x_3(t) - 1\\ 2x_1(t) - 2x_2^2(t) + 2x_3(t)\\ x_1(t)x_2(t) - x_3(t) + 1\end{bmatrix}.$$
and demonstrate that this point is a fixed point where $\dot{x}_1 = \dot{x}_2 = \dot{x}_3 = 0$.
2. Prove that the matrix exponential of $A = \begin{bmatrix}a & 0\\ 0 & b\end{bmatrix}$ is
$$e^A = \begin{bmatrix}e^a & 0\\ 0 & e^b\end{bmatrix},$$
using the Taylor series of the scalar and matrix exponentials.
3. Prove that the system response to control
$$\boldsymbol{x}(t) = e^{At}\boldsymbol{x}_0 + \int_0^t e^{A(t-\tau)}B\boldsymbol{u}(\tau)d\tau$$
satisfies the dynamical equation $\dot{\boldsymbol{x}}(t) = A\boldsymbol{x}(t) + B\boldsymbol{u}(t)$ by substitution.
4. Prove that the convolution operator
$$\mathcal{L}(\boldsymbol{u}(t)) = \int_0^T e^{A(T-\tau)}B\boldsymbol{u}(\tau)d\tau$$
is linear according to Eq. 17.6; that is, if $\mathcal{L}(\boldsymbol{u}_1(t)) = \boldsymbol{c}_1$, and $\mathcal{L}(\boldsymbol{u}_2(t)) = \boldsymbol{c}_2$, then demonstrate that $\mathcal{L}(a\boldsymbol{u}_1(t) + b\boldsymbol{u}_2(t)) = a\boldsymbol{c}_1 + b\boldsymbol{c}_2$.
5. Determine if the following system is controllable
$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\end{bmatrix} = \begin{bmatrix}0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\\ x_3(t)\end{bmatrix} + \begin{bmatrix}1\\ 0\\ 0\end{bmatrix}u(t),$$
by constructing the controllability matrix.
6. Determine for what value of a the system is not controllable
$$\begin{bmatrix}\dot{x}_1(t)\\ \dot{x}_2(t)\\ \dot{x}_3(t)\end{bmatrix} = \begin{bmatrix}0 & 0 & 0\\ 1 & 1 & 0\\ 1 & 0 & a\end{bmatrix}\begin{bmatrix}x_1(t)\\ x_2(t)\\ x_3(t)\end{bmatrix} + \begin{bmatrix}1\\ 0\\ 0\end{bmatrix}u(t),$$
by constructing the controllability matrix.
7. Derive the minimum energy equation Eq. 17.10
$$E_{\min} = (\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0)^\top(\mathcal{L}\mathcal{L}^*)^{-1}(\boldsymbol{x}_T - e^{AT}\boldsymbol{x}_0),$$

by substituting the minimum input $\boldsymbol{u}^*(t)$ into the control energy Eq. 17.8

$$E =< \boldsymbol{u}(t), \boldsymbol{u}(t) > .$$

8. Show that the controllability Gramian can be written as

$$W_C(T) = \int_0^T e^{A(T-t)} B B^\top e^{A^\top (T-t)} dt$$
$$= \int_0^T e^{A\tau} B B^\top e^{A^\top \tau} d\tau,$$

using the substitution $\tau = T - t$.

9. Show that the controllability Gramian for system

$$A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

is

$$W_C(T) = \begin{bmatrix} \frac{1}{2a}\left(e^{2aT} - 1\right) & 0 \\ 0 & \frac{1}{2b}\left(e^{2bT} - 1\right) \end{bmatrix}$$

10. Compute the minimum energy required for the system

$$A = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & 2 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix},$$

to transition from initial state $\boldsymbol{x}(0) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ to final state $\boldsymbol{x}(T) = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ in time $T = 1$.

Acknowledgements We gratefully acknowledge comments and feedback from Arian Ashourvan, Ann E. Sizemore, Melody X. Lim, Jennifer A. Stiso, Erin G. Teich, Teresa Karrer, Zhixin Lu, Harang Ju, and Eli J. Cornblath. We also thank Ann E. Sizemore for generous assistance with and input on schematic figure construction. JZK acknowledges support from the NIH T32-EB020087, PD: Felix W. Wehrli, and the National Science Foundation Graduate Research Fellowship No. DGE-1321851. DSB acknowledges support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the Paul G. Allen Foundation, the Army Research Laboratory through contract number W911NF-10-2-0022, the Army Research Office through contract numbers W911NF-14-1-0679 and W911NF-16-1-0474, the National Institute of Health (2-R01-DC-009209-11, 1R01HD086888-01, R01-MH107235, R01-MH107703, R01MH109520, 1R01NS099348 and R21-M MH-106799), the Office of Naval Research, and the National Science Foundation (BCS-1441502, CAREER PHY-1554488, BCS-1631550, and CNS-1626008).The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

References

1. M. Maier, A. Zippelius, M. Fuchs, Emergence of long-ranged stress correlations at the liquid to glass transition. Phys. Rev. Lett. **119**(26), 265701 (2017). https://doi.org/10.1103/PhysRevLett.119.265701
2. S. Kivelson, S.A. Kivelson, Defining emergence in physics. Quantum Mater. **1**, 16024 (2016). https://doi.org/10.1038/npjquantmats.2016.24
3. C.W. Lynn, L. Papadopoulos, D. Lee, D.S. Bassett, Surges of collective human activity emerge from simple pairwise correlations. Phys. Rev. X **9**, 011022-1–011022-19 (2018, in Press)
4. D.S. Bassett, M.S. Gazzaniga, Understanding complexity in the human brain. Trends Cogn. Sci. **15**(5), 200–209 (2011). https://doi.org/10.1016/j.tics.2011.03.006
5. A. Haimovici, E. Tagliazucchi, P. Balenzuela, D.R. Chialvo, Brain organization into resting state networks emerges at criticality on a model of the human connectome. Phys. Rev. Lett. **110**(17), 178101 (2013). https://doi.org/10.1103/PhysRevLett.110.178101
6. U. Braun, A. Schaefer, R.F. Betzel, H. Tost, A. Meyer-Lindenberg, D.S. Bassett, From maps to multi-dimensional network mechanisms of mental disorders. Neuron **97**(1), 14–31 (2018). https://doi.org/10.1016/j.neuron.2017.11.007
7. C.J. Stam, Modern network science of neurological disorders. Nat. Rev. Neurosci. **15**(10), 683–695 (2014). https://doi.org/10.1038/nrn3801
8. E. Tang, D.S. Bassett, Control of dynamics in brain networks. Rev. Mod. Phys. **90**, 031003 (2018). https://doi.org/10.1103/RevModPhys.90.031003
9. J. Downar, J. Geraci, T.V. Salomons, K. Dunlop, S. Wheeler, M.P. McAndrews, N. Bakker, D.M. Blumberger, Z.J. Daskalakis, S.H. Kennedy, A.J. Flint, P. Giacobbe, Anhedonia and reward-circuit connectivity distinguish nonresponders from responders to dorsomedial prefrontal repetitive transcranial magnetic stimulation in major depression. Biol. Psychiatry **76**(3), 176–85 (2014). https://doi.org/10.1016/j.biopsych.2013.10.026

10. J.D. Medaglia, D.Y. Harvey, N. White, A. Kelkar, J. Zimmerman, D.S. Bassett, R.H. Hamilton, Network controllability in the inferior frontal gyrus relates to controlled language variability and susceptibility to TMS. J. Neurosci. **38**(28), 6399–6410 (2018). https://doi.org/10.1523/JNEUROSCI.0092-17.2018
11. N. Gass, R. Becker, M. Sack, A.J. Schwarz, J. Reinwald, A. Cosa-Linan, L. Zheng, C.C. von Hohenberg, D. Inta, A. Meyer-Lindenberg, W. Weber-Fahr, P. Gass, A. Sartorius, Antagonism at the NR2B subunit of NMDA receptors induces increased connectivity of the prefrontal and subcortical regions regulating reward behavior. Psychopharmacology (Berl). **235**(4), 1055–1068 (2018). https://doi.org/10.1007/s00213-017-4823-2
12. U. Braun, A. Schafer, D.S. Bassett, F. Rausch, J.I. Schweiger, E. Bilek, S. Erk, N. Romanczuk-Seiferth, O. Grimm, L.S. Geiger, L. Haddad, K. Otto, S. Mohnke, A. Heinz, M. Zink, H. Walter, E. Schwarz, A. Meyer-Lindenberg, H. Tost, Dynamic brain network reconfiguration as a potential schizophrenia genetic risk mechanism modulated by NMDA receptor function. Proc. Natl. Acad. Sci. U. S. A. **113**(44), 12568–12573 (2016). https://doi.org/10.1073/pnas.1608819113
13. Z. Yang, S. Gu, N. Honnorat, K.A. Linn, R.T. Shinohara, I. Aselcioglu, S. Bruce, D.J. Oathes, C. Davatzikos, T.D. Satterthwaite, D.S. Bassett, Y.I. Sheline, Network changes associated with transdiagnostic depressive symptom improvement following cognitive behavioral therapy in MDD and PTSD. Mol. Psychiatry **23**(12), 2314–2323 (2018). https://doi.org/10.1038/s41380-018-0201-7
14. H. Markram, E. Muller, S. Ramaswamy, M.W. Reimann, M. Abdellah, C.A. Sanchez, A. Ailamaki, L. Alonso-Nanclares, N. Antille, S. Arsever, G.A. Kahou, T.K. Berger, A. Bilgili, N. Buncic, A. Chalimourda, G. Chindemi, J.D. Courcol, F. Delalondre, V. Delattre, S. Druckmann, R. Dumusc, J. Dynes, S. Eilemann, E. Gal, M.E. Gevaert, J.P. Ghobril, A. Gidon, J.W. Graham, A. Gupta, V. Haenel, E. Hay, T. Heinis, J.B. Hernando, M. Hines, L. Kanari, D. Keller, J. Kenyon, G. Khazen, Y. Kim, J.G. King, Z. Kisvarday, P. Kumbhar, S. Lasserre, J.V. Le Be, B.R. Magalhaes, A. Merchan-Perez, J. Meystre, B.R. Morrice, J. Muller, A. Munoz-Cespedes, S. Muralidhar, K. Muthurasa, D. Nachbaur, T.H. Newton, M. Nolte, A. Ovcharenko, J. Palacios, L. Pastor, R. Perin, R. Ranjan, I. Riachi, J.R. Rodriguez, J.L. Riquelme, C. Rossert, K. Sfyrakis, Y. Shi, J.C. Shillcock, G. Silberberg, R. Silva, F. Tauheed, M. Telefont, M. Toledo-Rodriguez, T. Trankler, W. Van Geit, J.V. Diaz, R. Walker, Y. Wang, S.M. Zaninetta, J. DeFelipe, S.L. Hill, I. Segev, F. Schurmann, Reconstruction and simulation of neocortical microcircuitry. Cell. **163**(2), 456–492 (2015). https://doi.org/10.1016/j.cell.2015.09.029
15. A. Rosenblueth, N. Wiener, The role of models in science. Philos. Sci. **12**(4), 316–321 (1945)
16. B.B. Machta, R. Chachra, M.K. Transtrum, J.P. Sethna, Parameter space compression underlies emergent theories and predictive models. Science **342**(6158), 604–607 (2013) https://doi.org/10.1126/science.1238723
17. H.H. Mattingly, M.K. Transtrum, M.C. Abbott, B.B. Machta, Maximizing the information learned from finite data selects a simple model. Proc. Natl. Acad. Sci. U. S. A. **115**(8), 1760–1765 (2018)
18. A.L. Hodgkin, A.F. Huxley, A quantitative description of membrane current and its application to conduction and excitation in nerve. J. Physiol. (1952). https://doi.org/10.1113/jphysiol.1952.sp004764
19. G. Cano, R. Dilao, Intermittency in the Hodgkin-Huxley model. J. Comput. Neurosci. **43**(2), 115–125 (2017). https://doi.org/10.1007/s10827-017-0653-9
20. J.H. Goldwyn, E. Shea-Brown, The what and where of adding channel noise to the Hodgkin-Huxley equations. PLoS Comput. Biol. **7**(11), e1002247 (2011). https://doi.org/10.1371/journal.pcbi.1002247
21. W. Teka, D. Stockton, F. Santamaria, Power-law dynamics of membrane conductances increase spiking diversity in a Hodgkin-Huxley model. PLoS Comput. Biol. **12**(3), e1004776 (2016). https://doi.org/10.1371/journal.pcbi.1004776
22. R. FitzHugh, Impulse and physiological states in theoretical models of nerve membrane. Biophys. J. (1961). https://doi.org/10.1016/S0006-3495(61)86902-6
23. W.S. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity. Bull. Math. Sci. (1943). https://doi.org/10.1007/BF02478259
24. J.J. Hopfield, Neural networks and physical systems with emergent collective computational abilities. Proc. Natl. Acad. Sci. (1982). https://doi.org/10.1073/pnas.79.8.2554
25. F. Rosenblatt, The perceptron: a probabilistic model for information storage and organization in the brain. Psychol. Rev. (1958). https://doi.org/10.1037/h0042519
26. G.A. Hedlund, Math. Syst. Theory **3**, 320 (1969). https://doi.org/10.1007/BF01691062
27. H.R. Wilson, J.D. Cowan, Excitatory and inhibitory interactions in localized populations of model neurons. Biophys. J. (1972). https://doi.org/10.1016/S0006-3495(72)86068-5
28. V. Shusterman, W.C. Troy, From baseline to epileptiform activity: a path to synchronized rhythmicity in large-scale neural networks. Phys. Rev. E. Stat. Nonlinear Soft. Matter. Phys. (2008). https://doi.org/10.1103/PhysRevE.77.061911
29. S.W. Oh et al., A mesoscale connectome of the mouse brain. Nature (2014). https://doi.org/10.1038/nature13186
30. P.J. Basser, S. Pajevic, C. Pierpaoli, J. Duda, A. Aldroubi, In vivo fiber tractography using DT-MRI data. Magn. Reson. Med. (2000). https://doi.org/10.1002/1522-2594(200010)44:4<625::AID-MRM17>3.0.CO;2-O

31. P.N. Taylor, J. Thomas, N. Sinha, J. Dauwels, M. Kaiser, T. Thesen, J. Ruths, Optimal control based seizure abatement using patient derived connectivity. Front Neurosci. **9**, 202 (2015). https://doi.org/10.3389/fnins.2015.00202
32. Y. Ezzyat et al., Direct brain stimulation modulates encoding states and memory performance in humans. Curr. Biol. (2017). https://doi.org/10.1016/j.cub.2017.03.028
33. J.G. White, E. Southgate, J.N. Thomson, S. Brenner, The structure of the nervous system of the nematode Caenorhabditis elegans. Philos. Trans. R. Soc. Lond. B. Biol. Sci. (1986). https://doi.org/10.1098/rstb.1986.0056
34. J.P. Nguyen et al., Whole-brain calcium imaging with cellular resolution in freely behaving Caenorhabditis elegans. Proc. Natl. Acad. Sci. (2016). https://doi.org/10.1073/pnas.1507110112
35. C.T. Lin, Structural controllability. IEEE Trans. Autom. Control (1974). https://doi.org/10.1109/TAC.1974.1100557
36. G. Yan et al., Network control principles predict neuron function in the *Caenorhabditis elegans* connectome. Nature (2017). https://doi.org/10.1038/nature24056
37. E.K. Towlson et al., *Caenorhabditis elegans* and the network control framework–FAQs. Philos. Trans. R. Soc. B. (2018). https://doi.org/10.1098/rstb.2017.0372
38. A. Avena-Koenigsberger, B. Misic, O. Sporns, Communication dynamics in complex brain networks. Nat. Rev. Neurosci. **19**(1), 17–33 (2017). https://doi.org/10.1038/nrn.2017.149
39. D.G. Taylor, M.C. Bushell, The spatial mapping of translational diffusion coefficients by the NMR imaging technique. Phys. Med. Biol. (1985). https://doi.org/10.1088/0031-9155/30/4/009
40. P. Hagmann et al., Mapping the structural core of human cerebral cortex. PLoS Biol. (2008). https://doi.org/10.1371/journal.pbio.0060159
41. J.D. Power et al., Functional network organization of the human brain. Neuron (2011). https://doi.org/10.1016/j.neuron.2011.09.006
42. D.S. Bassett, P. Zurn, J.I. Gold, On the nature and use of models in network neuroscience. Nat. Rev. Neurosci. **19**(9), 566–578 (2018). https://doi.org/10.1038/s41583-018-0038-8
43. M.E. Raichle et al., A default mode of brain function. PNAS (2001). https://doi.org/10.1073/pnas.98.2.676
44. O. Sporns, R.F. Betzel, Modular brain networks. Annu. Rev. Psychol. **67**, 613–640 (2016). https://doi.org/10.1146/annurev-psych-122414-033634
45. S.L. Bressler, V. Menon, Large-scale brain networks in cognition: emerging methods and principles. Trends. Cogn. Sci. (2010). https://doi.org/10.1016/j.tics.2010.04.004
46. S. Gu, F. Pasqualetti, M. Cieslak, Q.K. Telesford, A.B. Yu, A.E. Kahn, J.D. Medaglia, J.M. Vettel, M.B. Miller, S.T. Grafton, D.S. Bassett, Controllability of structural brain networks. Nat. Commun. **6**, 8414 (2015). https://doi.org/10.1038/ncomms9414
47. J.D. Medaglia, W. Huang, E.A. Karuza, A. Kelkar, S.L. Thompson-Schill, A. Ribeiro, D.S. Bassett, Functional alignment with anatomical networks is associated with cognitive flexibility. Nat. Hum. Behav. **2**(2), 156–164 (2018). https://doi.org/10.1038/s41562-017-0260-9
48. R.F. Betzel et al., Optimally controlling the human connectome: the role of network topology. Sci. Rep. (2016). https://doi.org/10.1038/srep30770
49. E. Tang, C. Giusti, G.L. Baum, S. Gu, E. Pollock, A.E. Kahn, D.R. Roalf, T.M. Moore, K. Ruparel, R.C. Gur, R.E. Gur, T.D. Satterthwaite, D.S. Bassett, Developmental increases in white matter network controllability support a growing diversity of brain dynamics. Nat. Commun. **8**(1), 1252 (2017)
50. S. Gu, R.F. Betzel, M.G. Mattar, M. Cieslak, P.R. Delio, S.T. Grafton, F. Pasqualetti, D.S. Bassett, Optimal trajectories of brain state transitions. Neuroimage. **148**, 305–317 (2017). https://doi.org/10.1016/j.neuroimage.2017.01.003
51. E.J. Cornblath, E. Tang, G.L. Baum, T.M. Moore, A. Adebimpe, D.R. Roalf, R.C. Gur, R.E. Gur, F. Pasqualetti, T.D. Satterthwaite, D.S. Bassett, Sex differences in network controllability as a predictor of executive function in youth. Neuroimage **188**, 122–134 (2018). https://doi.org/10.1016/j.neuroimage.2018.11.048
52. Z. Cui, J. Stiso, G.L. Baum, J.Z. Kim, D.R. Roalf, R.F. Betzel, S. Gu, Z. Lu, C.H. Xia, R. Ciric, T.M. Moore, R.T. Shinohara, K. Ruparel, C. Davatzikos, F. Pasqualetti, R.E. Gur, R.C. Gur, D.S. Bassett, T.D. Satterthwaite, Optimization of energy state transition trajectory supports the development of executive function during youth. bioRxiv 424929; https://doi.org/10.1101/424929
53. C. Moler, C.V. Loan, Nineteen dubious ways to compute the exponential of a matrix, twenty-five years later. SIAM Rev. (2003). https://doi.org/10.1137/S00361445024180
54. F. Pasqualetti, S. Zampiere, F. Bullo, Controllability metrics, limitations and algorithms for complex networks, in *2014 American Control Conference* (2014). https://doi.org/10.1109/ACC.2014.6858621
55. C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods. Econometrica (1969). https://doi.org/10.2307/1912791
56. A.K. Seth, A.B. Barrett, L. Barnett, Granger causality analysis in neuroscience and neuroimaging. J. Neurosci. **35**(8), 3293–3297 (2015). https://doi.org/10.1523/JNEUROSCI.4399-14.2015
57. L. Barnett, A.B. Barrett, A.K. Seth, Misunderstandings regarding the application of Granger causality in neuroscience. Proc. Natl. Acad. Sci. U. S. A. **115**(29), E6676–E6677 (2018). https://doi.org/10.1073/pnas.1714497115

58. K.J. Friston, Functional and effective connectivity: a review. Brain Connect. **1**(1), 13–36 (2011). https://doi.org/10.1089/brain.2011.0008
59. A.R. McIntosh, Tracing the route to path analysis in neuroimaging. Neuroimage. **62**(2), 887–890 (2012). https://doi.org/10.1016/j.neuroimage.2011.09.068
60. C.O. Becker, D.S. Bassett, V.M. Preciado, Large-scale dynamic modeling of task-fMRI signals via subspace system identification. J. Neural Eng. **15**(6), 066016 (2018). https://doi.org/10.1088/1741-2552/aad8c7
61. C. Stosiek, O. Garaschuk, K. Holthoff, A. Konnerth, In vivo two-photon calcium imaging of neuronal networks. Proc. Natl. Acad. Sci. (2003). https://doi.org/10.1073/pnas.1232232100
62. A.E. Motter, Networkcontrology. Chaos **25**(9), 097621 (2015). https://doi.org/10.1063/1.4931570
63. B.O. Koopman, Hamiltonian systems and transformations in Hilbert space. Proc. Natl. Acad. Sci. (1931). https://doi.org/10.1073/pnas.17.5.315
64. S.L. Brunton, B.W. Brunton, J.L. Proctor, J.N. Kutz, Koopman invariant subspaces and finite linear representations of nonlinear dynamical systems for control. PLoS One (2016). https://doi.org/10.1371/journal.pone.0150171
65. T. Kailath, *Linear Systems* (Prentice-Hall, Englewood Cliffs, 1980)
66. J.Z. Kim, J.M. Soffer, A.E. Kahn, J.M. Vettel, F. Pasqualetti, D.S. Bassett, Role of graph architecture in controlling dynamical networks with applications to neural systems. Nat. Phys. **14**, 91–98 (2018). https://doi.org/10.1038/nphys4268
67. Y. Xiao, S. Lao, L. Hou, M. Small, L. Bai, Effects of edge directions on the structural controllability of complex networks. PLoS One **10**(8), e0135282 (2015). https://doi.org/10.1371/journal.pone.0135282
68. S.P. Pang, W.X. Wang, F. Hao, Y.C. Lai, Universal framework for edge controllability of complex networks. Sci. Rep. **7**(1), 4224 (2017). https://doi.org/10.1038/s41598-017-04463-5
69. A.J. Whalen, S.N. Brennan, T.D. Sauer, S.J. Schiff, observability and controllability of nonlinear networks: the role of symmetry. Phys. Rev. X **5**, 011005 (2015).
70. E. Wu-Yan, R.F. Betzel, E. Tang, S. Gu, F. Pasqualetti, D.S. Bassett, Benchmarking measures of network controllability on canonical graph models. J. Nonlinear Sci. 1–39 (2018). https://doi.org/10.1007/s00332-018-9448-z
71. C. Tu, R.P. Rocha, M. Corbetta, S. Zampieri, M. Zorzi, S. Suweis, Warnings and caveats in brain controllability. Neuroimage **176**, 83–91 (2018). https://doi.org/10.1016/j.neuroimage.2018.04.010
72. T. Menara, D.S. Bassett, F. Pasqualetti, Structural controllability of symmetric networks. IEEE Trans. Autom. Control **64**(9), 3740–3747 (2019). https://ieeexplore.ieee.org/document/8533416
73. L. Wiles, S. Gu, F. Pasqualetti, B. Parvesse, D. Gabrieli, D.S. Bassett, D.F. Meaney, Autaptic connections shift network excitability and bursting. Sci. Rep. **7**, 44006 (2017). https://doi.org/10.1038/srep44006
74. H. Ju, J.Z. Kim, D.S. Bassett, Network topology of neural systems supporting avalanche dynamics predicts stimulus propagation and recovery (2018). arXiv:1812.09361
75. E.J. Cornblath, A. Ashourvan, J.Z. Kim, R.F. Betzel, R. Ciric, G.L. Baum, X. He, K. Ruparel, T.M. Moore, R.C. Gur, R.E. Gur, R.T. Shinohara, D.R. Roalf, T.D. Satterthwaite, D.S. Bassett, Temporal sequences of brain activity at rest are constrained by white matter structure and modulated by cognitive demands. Commun. Biol. (2020, In Press).
76. S.F. Muldoon, J. Costantini, W.R.S. Webber, R. Lesser, D.S. Bassett, Locally stable brain states predict suppression of epileptic activity by enhanced cognitive effort. Neuroimage Clin. **18**, 599–607 (2018). https://doi.org/10.1016/j.nicl.2018.02.027
77. S.A. Mengiste, A. Aertsen, A. Kumar, Effect of edge pruning on structural controllability and observability of complex networks. Sci. Rep. **5**, 18145 (2015)
78. J. Jeganathan, A. Perry, D.S. Bassett, G. Roberts, P.B. Mitchell, M. Breakspear, Fronto-limbic dysconnectivity leads to impaired brain network controllability in young people with bipolar disorder and those at high genetic risk. Neuroimage Clin. **19**, 71–81 (2018). https://doi.org/10.1016/j.nicl.2018.03.032
79. B.C. Bernhardt, M. Liu, R. Vos de Wael, J. Smallwood, E. Jefferies, S. Gu, D.S. Bassett, A. Bernasconi, N. Bernasconi, Hippocampal pathology modulates white matter connectome topology and controllability in temporal lobe epilepsy. Neurology. **92**(19), e2209-e2220 (2019).
80. S.F. Muldoon, F. Pasqualetti, S. Gu, M. Cieslak, S.T. Grafton, J.M. Vettel, D.S. Bassett, Stimulation-based control of dynamic brain networks. PLoS Comput. Biol. **12**(9), e1005076 (2016). https://doi.org/10.1371/journal.pcbi.1005076
81. X. Yang, N. Ram, S.D. Gest, D.M. Lydon-Staley, D.E. Conroy, A.L. Pincus, P.C.M. Molenaar, Socioemotional dynamics of emotion regulation and depressive symptoms: a person-specific network approach. Complexity **2018**, pii: 5094179 (2018) https://doi.org/10.1155/2018/5094179
82. D.M. Lydon-Staley, I. Barnett, T.D. Satterthwaite, D.S. Bassett, Digital phenotyping for psychiatry: Accommodating data and theory with network science methodologies. Curr. Opin. Biomed. Eng. **9**, 8–13 (2019). https://doi.org/10.1016/j.cobme.2018.12.003