Geometric framework to predict structure from function in neural networks

Tirthabir Biswas[1,2,*] and James E. Fitzgerald [1,†]

[1]*Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, Virginia 20147, USA*
[2]*Department of Physics, Loyola University, New Orleans, Louisiana 70118, USA*

This article is part of the Physical Review Research *collection titled Physics of Neuroscience.*

Neural computation in biological and artificial networks relies on the nonlinear summation of many inputs. The structural connectivity matrix of synaptic weights between neurons is a critical determinant of overall network function, but quantitative links between neural network structure and function are complex and subtle. For example, many networks can give rise to similar functional responses, and the same network can function differently depending on context. Whether certain patterns of synaptic connectivity are required to generate specific network-level computations is largely unknown. Here we introduce a geometric framework for identifying synaptic connections required by steady-state responses in recurrent networks of threshold-linear neurons. Assuming that the number of specified response patterns does not exceed the number of input synapses, we analytically calculate the solution space of all feedforward and recurrent connectivity matrices that can generate the specified responses from the network inputs. A generalization accounting for noise further reveals that the solution space geometry can undergo topological transitions as the allowed error increases, which could provide insight into both neuroscience and machine learning. We ultimately use this geometric characterization to derive certainty conditions guaranteeing a nonzero synapse between neurons. Our theoretical framework could thus be applied to neural activity data to make rigorous anatomical predictions that follow generally from the model architecture.

I. INTRODUCTION

Structure-function relationships are fundamental to biology [1–3]. In neural networks, the structure of synaptic connectivity critically shapes the functional responses of neurons [4,5], and large-scale techniques for measuring neural network structure and function provide exciting opportunities for examining this link quantitatively [6–15]. The ellipsoid body in the central complex of *Drosophila* is a beautiful example where modeling showed how the structural pattern of excitatory and inhibitory connections enables a persistent representation of heading direction [16–19]. Lucid structure-function links have also been found in several other neural networks [20–23]. However, it is generally hard to predict either neural network structure or function from the other [5,24]. For example, functionally inferred connectivity can capture neuronal response correlations without matching structural connectivity [25–28], and network simulations with structural constraints do not automatically reproduce function [29–31]. Two broad modeling difficulties hinder the establishment of robust structure-function links. First, models with too much

detail are difficult to adequately constrain and analyze. Second, models with too little detail may poorly match biological mechanisms, the model mismatch problem. Here we propose a rigorous theoretical framework that attempts to balance these competing factors to predict components of network structure required for function.

Neural network function probably does not depend on the exact strength of every synapse. Indeed, multiple network connectivity structures can generate the same functional responses [32,33], as illustrated by structural variability across individual animals [24,34] and artificial neural networks [29,35–37]. Such redundancy may be a general feature of emergent phenomena in physics, biology, and neuroscience [38–40]. Nevertheless, some important details may be consistent despite this variability, and here we find well-constrained structure-function links by characterizing all connectivity structures that are consistent with the desired functional responses [24]. We also account for ambiguities caused by measurement noise. Our goal is not to find degenerate networks that perform equivalently in all possible scenarios. We instead seek a framework that finds connectivity required for specific functional responses, independently of whatever else the network might do.

The model mismatch problem has at least two facets. First, neurons and synapses are incredibly complex [41–44], but which complexities are needed to elucidate specific structure-function relationships is unclear [5,45,46]. This issue is very hard to address in full generality, and here we seek a theoretical framework that makes clear experimental predictions that can adjudicate candidate models empirically. In

---

*biswast@janelia.hhmi.org

†fitzgeraldj@janelia.hhmi.org
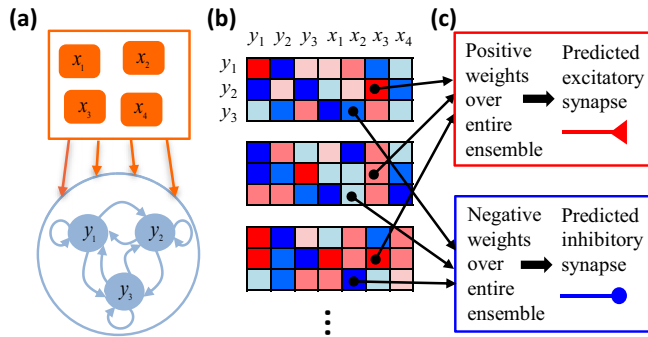
(a) **(b)** **(c)**

FIG. 1. Cartoon of theoretical framework. (a) We first specify some steady-state responses of a recurrent threshold-linear neural network receiving feedforward input. (b) We then find all synaptic weight matrices that have fixed points at the specified responses. Red (blue) matrix elements are positive (negative) synaptic weights. (c) When a weight is consistently positive (or consistently negative) across all possibilities, then the model needs a nonzero synaptic connection to generate the responses. We therefore make the experimental prediction that this synapse must exist. We also predict whether the synapse is excitatory or inhibitory.

particular, we predict neural network structure only when it occurs in all networks generating the functional responses. This high bar precludes the analysis of biophysically-detailed network models, which require numerical exploration of the connectivity space that is typically incomplete [24,32,47–49]. We instead focus on recurrent firing rate networks of threshold-linear neurons, which are growing in popularity because they strike an appealing balance between biological realism, computational power, and mathematical tractability [12,16,18,20,22,23,29,30,37,50–55].

The second facet of the model mismatch problem is hidden variables, such as missing neurons, neuromodulator levels, and physiological states [5,56–58]. Here we take inspiration from whole-brain imaging in small organisms [15], such as *Caenorhabditis elegans* [9], larval zebrafish [8,12,57], and larval *Drosophila* [11], and assume access to all relevant neurons. Our model neglects neuromodulators and other state variables, which would be interesting to consider in the future. Furthermore, many experiments indirectly assess neuronal spiking activity, such as by calcium florescence [58–61] or hemodynamic responses [25,62–64]. We restrict our analysis to steady-state responses to mitigate mismatch between fast firing rate changes and these inherently slow measurement techniques.

Our analysis begins with an analytical characterization of synaptic weight matrices that realize specified steady-state responses as fixed points of neural network dynamics [Figs. 1(a) and 1(b)]. A key insight is that asymmetrically constrained dimensions appear as a consequence of the threshold nonlinearity. Synaptic weight components in these semiconstrained dimensions are completely uncertain in one half of the dimension but well-constrained in the other. We then compute error surfaces by finding weight matrices with fixed points near the desired ones. This error landscape has a continuum of local and global minima, and constant-error surfaces exhibit topological transitions that add semiconstrained dimensions as the error increases. This may help explain the importance

of weight initialization in machine learning, as poorly initialized models can get stuck in semiconstrained dimensions that abruptly vanish at nonzero error. By studying the geometric structure of the neural network ensemble that can approximate the functional responses, we derive analytical formulas that pinpoint a subset of connections, which we term certain synapses, that must exist for the model to work [Fig. 1(c)]. These analytical results are especially useful for studying high-dimensional synaptic weight spaces that are otherwise intractable. Since the presence of a synapse is readily measurable, our theory generates accessible experimental predictions [Fig. 1(c)]. Tests of these predictions assess the utility of the modeling framework itself, as the predictions hold across model parameters. Their successes and failures can thus move us forward toward identifying the mechanistic principles governing how neural networks implement brain computations.

The rest of the paper begins in Sec. II with a toy problem that concretely demonstrates the approach illustrated in Fig. 1 and relates the geometry of the solution space (all synaptic weight matrices that realize a given set of response patterns) to the concept of a certain synapse. In Sec. III, we explain how the solution space for a limited number of response patterns can be calculated for an arbitrarily large threshold-linear recurrent neural network. Section IV is devoted to three simple toy problems that provide additional insights into how the geometry of the solution space can help us to identify certain synapses. This is followed by Sec. V, where we explain and numerically test the precise algebraic relation that must be satisfied for a synapse to be certain when the response patterns are orthonormal. Section VI generalizes our analyses to include noise, including numerical tests via simulation. Finally, Sec. VII concludes the paper by summarizing our main results and discussing important future directions.

II. AN ILLUSTRATIVE TOY PROBLEM

To gain intuition on how robust structure-function links can be established, including the effects of nonlinearity, we begin by analyzing the structural implications of functional responses in a very simple threshold-linear feedforward network [Fig. 2(a)]. We assume that two input neurons, x_1 and x_2, provide signals to a single driven neuron, y, via synaptic weights, w_1 and w_2. The weights are unknown, and we constrain their possible values using two neuronal response patterns, labeled $\mu = +$ and $\mu = -$. We suppose that steady-state activities of the input neurons and driven neuron are nonlinearly related according to

$$y = \Phi(w_1 x_1 + w_2 x_2), \tag{1}$$

where x_1, x_2, and y denote firing rates of the corresponding neurons, and

$$\Phi(s) = \max(0, s) \tag{2}$$

is the threshold-linear transfer function. The driven neuron responds ($y = 1$) when $x_1 = x_2 = 1$ in the $\mu = +$ pattern. In contrast, the driven neuron does not respond ($y = 0$) when $x_1 = -x_2 = 1$ in the $\mu = -$ pattern. If the transfer function were linear, then it is easy to see that there is a unique set
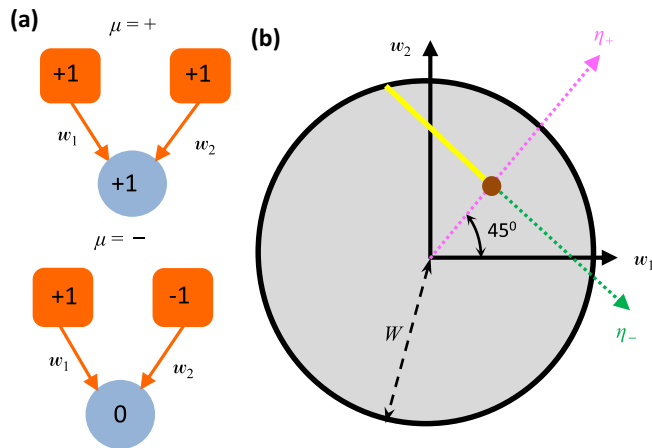
FIG. 2. An illustrative two-dimensional problem. (a) Cartoon depicting two stimulus response patterns in a simple feedforward network with two input neurons and one driven neuron. (b) Since the driven neuron in panel (a) responds in one condition but not the other, we have one constrained dimension (magenta axis) and one semiconstrained dimension (green axis). The yellow ray depicts the space of weights, (w_1, w_2), that generate the stimulus transformation. The weight vector $(\frac{1}{2}, \frac{1}{2})$ (brown dot) would uniquely generate the neural responses in a linear network. We assume that the magnitude of the weight vector is bounded by W, such that all candidate weight vectors lie within a circle of that radius. A nonzero synapse $x_2 \rightarrow y$ exists in all solutions, but the $x_1 \rightarrow y$ synapse can be zero because the yellow ray intersects the $w_1 = 0$ axis.

of weights, $w_1 = w_2 = \frac{1}{2}$, that produces these driven neuron responses, the brown dot in Fig. 2(b).

How does the nonlinearity change the solution space of weights that reproduce the driven neuron responses? To answer this question, we define two linear combinations of weights,

$$\eta_\pm = w_1 \pm w_2, \qquad (3)$$

which correspond to the driven neuron's input drive in patterns $\mu = \pm$. Equation (1) now yields rather simple algebraic constraints for the two patterns:

$$y_+ = 1 = \Phi(\eta_+) \Rightarrow \eta_+ = 1, \qquad (4)$$

$$y_- = 0 = \Phi(\eta_-) \Rightarrow \eta_- \leqslant 0. \qquad (5)$$

Note that η_- would have had to be zero if Φ were linear, but because the threshold-linear transfer function turns everything negative into a null response, η_- can now also be any negative number. However, sufficiently negative values of η_- correspond to implausibly large weight vectors, and hence we focus on solutions with norm bounded above by some value, W. The nonlinearity thus turns the unique linear solution [brown dot in Fig. 2(b)] into a continuum of solutions [yellow line segment in Fig. 2(b)]. This continuum lies along what we will refer to as a semiconstrained dimension. Indeed, this will turn out to be a generic feature of threshold-linear neural networks: every

time there is a null response, a semiconstrained dimension emerges in the solution space.[1]

Although we found infinitely many weight vectors that solve the problem, all solutions to the problem have a synaptic connection $x_2 \rightarrow y$, and this connection is always excitatory [Fig. 2(b)]. Positive, negative, or zero connection weights are all possible for $x_1 \rightarrow y$. However, this reveals why the value of the synaptic weight bound, W, has important implications for the solution space. For example, all solutions in Fig. 2(b) with $|\vec{w}| < 1$ have $w_1 > 0$, whereas larger magnitude weight vectors have $w_1 \leqslant 0$. Therefore, one would be certain that an excitatory $x_1 \rightarrow y$ synapse exists if the weight bound were biologically known to be less than $W_{\rm cr} = 1$. We refer to this weight bound as W-critical. Looser weight bounds raise the possibility that the synapse is absent or inhibitory. Note that too tight weight bounds, here less than $W_{\min} = 1/\sqrt{2}$, can exclude all solutions.

The example of Fig. 2 concretely illustrates the general procedure diagramed in Fig. 1. First, we specified a network architecture and steady-state response patterns [Figs. 1(a) and 2(a)]. Second, we found all synaptic weight vectors that can implement the nonlinear transformation [Figs. 1(b) and 2(b)]. Finally, we determined whether individual synaptic weights varied in sign across the solution space [Figs. 1(c) and 2(b)]. Section III will generalize the first two parts of this procedure to characterize the solution space of any threshold-linear recurrent neural network, assuming that the number of response patterns is at most the dimensionality of the weight vectors. Sections IV and V will then generalize the final part of this procedure to pinpoint synaptic connections that are critical for generating any specified set of orthonormal responses.

III. SOLUTION SPACE GEOMETRY

A. Neural network structure and dynamics

Consider a neural network of $\mathcal{I}$ input neurons that send signals to a recurrently connected population of $\mathcal{D}$ driven neurons [Fig. 3(a)]. We compactly represent the network connectivity with a matrix of synaptic weights, w_{im}, where $i = 1, \ldots, \mathcal{D}$ indexes the driven neurons, and $m = 1, \ldots, \mathcal{D} + \mathcal{I}$ indexes presynaptic neurons from both the driven and input populations. We suppose that activity in the population of driven neurons dynamically evolves according to

$$\tau_i \frac{dy_i}{dt} = -y_i + \Phi\left(\sum_{m=1}^{\mathcal{D}} w_{im} y_m + \sum_{m=\mathcal{D}+1}^{\mathcal{D}+\mathcal{I}} w_{im} x_{m-\mathcal{D}}\right), \qquad (6)$$

where y_i is the firing rate of the ith driven neuron, x_m is the firing rate of the mth input neuron, and τ_i is the time constant that determines how long the ith driven neuron integrates its presynaptic signals. It is possible that prior biological knowledge dictates that certain synapses appearing in Eq. (6) are absent. For notational convenience, in this paper we will assume that the number of synapses onto each driven neuron remains

---

[1]Assuming that the number of patterns does not exceed the dimensionality of the synaptic weight vector.
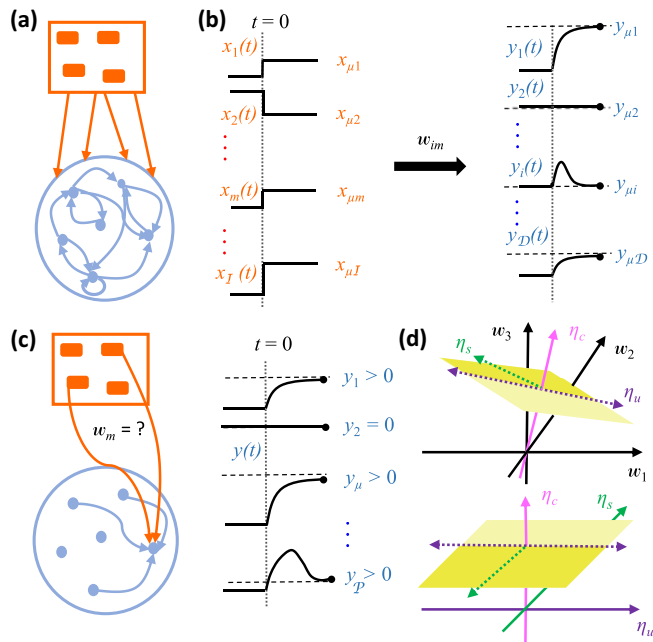
FIG. 3. Finding network structure that implements functional responses. (a) Cartoon depicting a recurrent network of *driven* neurons (blue) receiving feedforward input from a population of *input* neurons (orange). (b) The μth pattern of input neuron activity ($x_{\mu m}$) appears at $t = 0$ and drives the recurrent neurons to approach the steady-state response pattern ($y_{\mu i}$) via feedforward and recurrent network connectivity (w_{im}). (c) (*Left*) We focus on one driven neuron at a time, referred to henceforth as the *target* neuron, to determine its possible incoming synaptic weights, w_m. (*Right*) These weights must reproduce the target neuron's $\mathcal{P}$ steady-state responses from the steady-state activity patterns of all $\mathcal{N}$ presynaptic neurons. (d) The yellow planes depict the subspace of incoming weights that can exactly reproduce all nonzero responses of the target neuron, and the subregion shaded dark yellow indicates weights that also reproduce the target neuron's zero responses. The top graph depicts the weight space parametrized by physically meaningful w coordinates, but the solution space is more simply parametrized by abstract η coordinates (bottom). The η coordinates depend on the specified stimulus transformation ($x_{\mu m} \to y_{\mu i}$), and η_c, η_s, and η_u are coordinates in $\mathcal{C}$-dimensional constrained, $\mathcal{S}$-dimensional semiconstrained, and $\mathcal{U}$-dimensional unconstrained subspaces, respectively.

the same,[2] and we will denote this number of the incoming synapses as $\mathcal{N}$. Note that $\mathcal{N} = \mathcal{I} + \mathcal{D}$ for a general recurrent network, $\mathcal{N} = \mathcal{I} + \mathcal{D} - 1$ for recurrent networks without self-synapses, and $\mathcal{N} = \mathcal{I}$ for feedforward networks. We suppose that the network functionally maps input patterns, $x_{\mu m}$, to steady-state driven signals, $y_{\mu i} \geqslant 0$, where $\mu = 1, \ldots, \mathcal{P}$ labels the patterns [Fig. 3(b)]. We assume throughout that $\mathcal{P} \leqslant \mathcal{N}$, as the number of known response patterns is typically small, and the number of possible synaptic inputs is large. Experimentally, different response patterns often correspond

---

[2] It will become progressively evident that our construction of the solution space and certainty condition can be trivially adapted to the case where the number of presynaptic neurons changes from one driven neuron to another.

to different stimulus conditions, so we will often refer to μ as a stimulus index and $x_{\mu m} \to y_{\mu i}$ as a stimulus transformation.

B. Decomposing a recurrent network into $\mathcal{D}$ feedforward networks

Our goal is to find features of the synaptic weight matrix that are required for the stimulus transformation discussed above. For notational simplicity, let us consider the case where we potentially have all-to-all connectivity, so that $\mathcal{N} = \mathcal{D} + \mathcal{I}$, but we will later explain how our arguments generalize. Since all time-derivatives are zero at steady-state, the response properties provide $\mathcal{D} \times \mathcal{P}$ nonlinear equations for $\mathcal{D} \times \mathcal{N}$ unknown parameters[3]:

$$y_{\mu i} = \Phi\left(\sum_{m=1}^{\mathcal{D}} w_{im} y_{\mu m} + \sum_{m=\mathcal{D}+1}^{\mathcal{D}+\mathcal{I}} w_{im} x_{\mu, m-\mathcal{D}}\right). \quad (7)$$

Inspection of the above equation, however, reveals that each neuron's steady-state activity depends only on a single row of the connectivity matrix [Fig. 3(c)]; the responses of the ith driven neuron, $\{y_{\mu i}, \mu = 1, \ldots, \mathcal{P}\}$, are only affected by its incoming synaptic weights, $\{w_{im}, m = 1, \ldots, \mathcal{N}\}$. Thus, the above equations separate into $\mathcal{D}$ independent sets of equations, one for each driven neuron. In other words, we now have to solve $\mathcal{D}$ feedforward problems, each of which will characterize the incoming synaptic weights of a particular driven neuron, which we term the target neuron. Note that since a generic target neuron receives signals from both the input and the driven populations, the activities of both input and driven neurons serve to produce the presynaptic input patterns that drive the responses of the target neuron in the reduced feedforward problem.

C. Solution space for feedforward networks

We have just seen how we can solve the problem of finding synaptic weights consistent with steady-state responses of a recurrent population of neurons, provided we know how to solve the equivalent problem for feedforward networks. Accordingly, we will now focus on a feedforward network, where a single target neuron, y, receives inputs from $\mathcal{N}$ neurons $\{x_m; m = 1, \ldots, \mathcal{N}\}$, to find the ensemble of synaptic weights that reproduce this target neuron's observed responses. The constraint equations are

$$y_\mu = \Phi\left(\sum_{m=1}^{\mathcal{N}} x_{\mu m} w_m\right), \quad (8)$$

where y_μ now stands for the activity of the target neuron driven by the μth input pattern, and $\bar{w}$ is the $\mathcal{N}$-vector of

---

[3] A slightly different rate equation,

$$\tau_i \frac{dv_i}{dt} = -v_i + \sum_{m=1}^{\mathcal{D}} w_{im} r_m + \sum_{m=\mathcal{D}+1}^{\mathcal{D}+\mathcal{I}} w_{im} x_{m-\mathcal{D}},$$

with $r_i = \Phi(v_i)$, is also in vogue. While the dynamics of this model are slightly different from Eq. (6), at steady state they reduce to the same form as Eq. (7). In particular, $r_i = \Phi(\sum_{m=1}^{\mathcal{D}} w_{im} r_m + \sum_{m=\mathcal{D}+1}^{\mathcal{D}+\mathcal{I}} w_{im} x_{m-\mathcal{D}})$.

synaptic weights onto the target neuron. Assuming that the $\mathcal{P} \times \mathcal{N}$ matrix x is rank $\mathcal{P}$, we let the $\mathcal{N} \times \mathcal{N}$ matrix X be rank $\mathcal{N}$ with $X_{\mu m} = x_{\mu m}$ for $\mu = 1, \dots, \mathcal{P}$. This implies that the last $\mathcal{N} - \mathcal{P}$ rows of X span the null space of x, and X defines a basis transformation on the weight space,

$$\eta_\mu = \sum_{m=1}^{\mathcal{N}} X_{\mu m} w_m \Leftrightarrow w_m = \sum_{\mu=1}^{\mathcal{N}} X_{m\mu}^{-1} \eta_\mu. \tag{9}$$

The $\mathcal{N}$ linearly independent columns of X^{-1} define the basis vectors corresponding to the η coordinates,

$$X^{-1} = (\vec{\varepsilon}_1 \cdots \vec{\varepsilon}_\mu \cdots \vec{\varepsilon}_\mathcal{N}). \tag{10}$$

In other words,

$$\vec{\varepsilon}_\mu = \sum_{m=1}^{\mathcal{N}} \widehat{e}_m X_{m\mu}^{-1}, \tag{11}$$

where $\{\widehat{e}_m\}$ is the *physical* orthonormal basis whose coordinates, $\{w_m\}$, correspond to the material substrates of network connectivity. These basis vectors can be obtained from $\{\vec{\varepsilon}_\mu\}$ by an inverse basis transformation:

$$\widehat{e}_m = \sum_{\mu=1}^{\mathcal{N}} \vec{\varepsilon}_\mu X_{\mu m}. \tag{12}$$

We can thus write any vector of incoming weights as

$$\vec{w} = \sum_{m=1}^{\mathcal{N}} w_m \widehat{e}_m = \sum_{\mu=1}^{\mathcal{N}} \eta_\mu \vec{\varepsilon}_\mu. \tag{13}$$

In terms of η coordinates, the nonlinear constraint equations take a rather simple form:

$$y_\mu = \Phi(\eta_\mu) \quad \text{for } \mu = 1, \cdots, \mathcal{P}. \tag{14}$$

Accordingly, η coordinates succinctly parametrize the solution space of all weight matrices that support the specified fixed points [Fig. 3(d)]. Each η dimension can be neatly categorized into one of three types. First, for each stimulus condition μ where $y_\mu > 0$, we must have $\eta_\mu > 0$. This in turn implies that $\Phi(\eta_\mu) = \eta_\mu = y_\mu$. Because the coordinate η_μ must adopt a specific value to generate the transformation, we say that μ defines a *constrained* dimension. We denote the number of constrained dimensions as $\mathcal{C} \leqslant \mathcal{P}$. Second, note that the threshold in the transfer function implies that $\Phi(a) = 0$ for all $a \leqslant 0$. Therefore, for any stimulus condition such that $y_\mu = 0$, we have a solution whenever $\eta_\mu \leqslant 0$. Because positive values of η_μ are excluded but all negative values are equally consistent with the transformation, we say that μ defines a *semiconstrained* dimension. We denote the number of semiconstrained dimensions as $\mathcal{S} = \mathcal{P} - \mathcal{C}$. Finally, we have no constraint equations for η_μ if $\mu = \mathcal{P} + 1, \cdots, \mathcal{N}$. Because all positive or negative values of η_μ are equally consistent with the stimulus transformation, we say that μ defines an *unconstrained* dimension. We denote the number of unconstrained dimensions as $\mathcal{U} = \mathcal{N} - \mathcal{P}$. Altogether, the stimulus transformation is consistent with every incoming

weight vector that satisfies

$$\begin{aligned} \eta_\mu &= y_\mu & &\text{if } y_\mu > 0, \mu \leqslant \mathcal{P} \\ -\infty &< \eta_\mu \leqslant 0 & &\text{if } y_\mu = 0, \mu \leqslant \mathcal{P} \\ -\infty &< \eta_\mu < \infty & &\text{if } \mu > \mathcal{P}. \end{aligned} \tag{15}$$

Note that one can enumerate the solutions in the physically meaningful w coordinates by simply applying the inverse basis transformation in Eq. (9) to any solution found in η coordinates.

Going forward, it will be convenient to extend the $\mathcal{P}$-dimensional vector of target neuron activity to an $\mathcal{N}$-dimensional vector whose components along the unconstrained dimensions are equal to zero, because this will allow us to compactly write equations in terms of dot products between the activity vector and vectors in the $\mathcal{N}$-dimensional weight space. Rather than introducing a new notation for this extended $\mathcal{N}$-dimensional vector, we simply write $\vec{y}$ with $y_\mu = 0$ for $\mu = \mathcal{P} + 1, \dots, \mathcal{N}$. It is critical to remember that this is merely a notational convenience, and the solution space distinguishes between semiconstrained dimensions and unconstrained dimensions according to Eq. (15). In particular, $y_\mu = 0$ is a constraint equation for semiconstrained dimensions, but $y_\mu = 0$ is a notational convenience for unconstrained dimensions.

D. Back to the recurrent network

To understand how the solution space geometry of the feedforward network can be translated back to the recurrent network, it is useful to group together the steady-state activities of all input and driven neurons that are presynaptic to the ith driven neuron as a $\mathcal{P} \times \mathcal{N}$ input pattern matrix, $z^{(i)}$.[4] The entries of the matrix, $z_{\mu m}^{(i)}$, correspond to the responses of the mth presynaptic neuron to the μth stimulus. At this point it is easy to see that when biological constraints dictate that some of the synapses are absent, then one should just exclude those presynaptic neurons when constructing $z^{(i)}$, such that the m index excludes those presynaptic neurons. Similarly, by a suitable reordering, which will depend on the driven neuron, we can always ensure that $m = 1, \dots, \mathcal{N}$ runs only over the neurons that are presynaptic to the given driven neuron.

Once the input patterns feeding into the ith neuron are known, we can follow the steps outlined in the previous subsection to define the $\mathcal{N} \times \mathcal{N}$ full rank extension of $z^{(i)}$, $Z^{(i)}$, and the $\eta^{(i)}$ coordinates via

$$\eta_\mu^{(i)} = \sum_{m=1}^{\mathcal{N}} Z_{\mu m}^{(i)} w_{im}. \tag{16}$$

The nature of the $\eta_\mu^{(i)}$ coordinates, that is whether they are constrained, semiconstrained, or unconstrained, is determined by how the ith neuron responded to the stimulus conditions, as in Eq. (15). Repeating this process for all driven neurons

---

[4]In fact, one can easily incorporate the case when the number of presynaptic partners differs from one driven neuron to another. This just means that the $z^{(i)}$ matrices will have dimensions $\mathcal{P} \times \mathcal{N}_i$, where $\mathcal{N}_i$ represents the number of presynaptic partners of the ith neuron.

provides a geometric characterization of the entire recurrent network solution space, which involves all elements of the synaptic weight matrix, w_{im}.

An important special case is all-to-all network connectivity. In this case, the $Z^{(i)}$ matrices are the same for all driven neurons, and therefore the directions corresponding to the η coordinates are also preserved.[5] In particular, the orientation of the unconstrained subspace with respect to the physical basis does not change from one driven neuron to another. However, how a given driven neuron responds to a particular stimulus determines whether the corresponding η direction is going to be constrained or semiconstrained for the feedforward network associated with that driven neuron.

IV. CERTAIN SYNAPSES IN ILLUSTRATIVE 3D EXAMPLES

Although we have found infinitely many weight matrices that produce a given stimulus transformation, it is nevertheless possible that the solutions imply firm anatomical constraints (e.g., Sec. II). In this paper we focus on finding synapses that must be nonzero in order for the response patterns to be fixed points of the neural network dynamics. We refer to such synapses as certain, because the synapse must exist in the model, and its sign is identifiable from the response patterns. It is clear from the geometry of the solution space that the relative orientations between the η coordinates and the physical w coordinates are significant determinants of synapse certainty. To build quantitative intuition for how the solution space geometry precisely determines synapse certainty, we begin by first analyzing a few illustrative toy problems. In the next section we will describe the more general treatment of high-dimensional networks. Importantly, we select and parametrize each toy problem to introduce concepts and notations that will reappear in the general solution.

More specifically, we first consider three feedforward examples with $\mathcal{N} = 3$ [Fig. 4(a)]. The first two examples have $\mathcal{P} = 3$, and the third has $\mathcal{P} = 2$. In the first example, we will assume that the driven neuron does not respond to the first two stimulus patterns, but responds positively to the third pattern. So we have two semiconstrained and one constrained dimension,

$$\eta_1 \leqslant 0, \quad \eta_2 \leqslant 0, \quad \text{and} \quad \eta_3 = y_3 > 0. \tag{17}$$

In contrast, in the second example we will have two constrained and one semiconstrained dimension,

$$\eta_1 = y_1 > 0, \quad \eta_2 = y_2 > 0, \quad \text{and} \quad \eta_3 \leqslant 0. \tag{18}$$

The final example will feature one unconstrained, one semiconstrained, and one constrained dimension,

$$\eta_1 \leqslant 0, \quad \eta_2 = y_2 > 0, \quad \text{and} \quad -\infty < \eta_3 < \infty. \tag{19}$$

---

[5]Nevertheless, the vector spaces of synaptic weights are fundamentally distinct for different driven neurons, as these vector spaces pertain to the incoming synapses onto different driven neurons. The fact that the $Z^{(i)}$ matrices are the same for all i means that the relative orientation of the η directions, with respect to the physical w-coordinate axes (labeled by the presynaptic indices), remains the same for all the driven neurons.

For technical simplicity we will consider orthonormal input patterns, $X^{-1} = X^T$, which implies that

$$\sum_{m=1}^{\mathcal{N}} X_{\mu m} X_{\nu m} = \delta_{\mu\nu} = \vec{\varepsilon}_\mu \cdot \vec{\varepsilon}_\nu, \tag{20}$$

where $\delta_{\mu\nu}$ is the Kronecker δ function, which equals 1 if $\mu = \nu$ and 0 if $\mu \neq \nu$, so $\hat{\varepsilon}_\mu = \vec{\varepsilon}_\mu$. This trivially implies that the η coordinates are related to the synaptic coordinates via a rotation, so the spherical biological bound on the physical coordinates transforms to an identical spherical bound on the η coordinates:

$$\sum_{\mu=1}^{\mathcal{N}=3} \eta_\mu^2 = \sum_{m=1}^{\mathcal{N}=3} w_m^2 \leqslant W^2. \tag{21}$$

A. Problem 1

Let us first focus on the example with two semiconstrained and one constrained dimension, whose solution space is depicted in deep yellow in Fig. 4(b). Suppose we are interested in assessing whether the w_1 synapse is certain. Since the $w_1 = 0$ plane divides the weight space into the positive and the negative halves, the synapse will be certain if this plane does not intersect with the solution space, which clearly depends on the orientation of the plane relative to the various η directions [Fig. 4(b)]. It is thus useful to consider how the $w_1 = 0$ plane's unit normal vector pointing toward positive weights, $\hat{e} \equiv \hat{e}_1$, is oriented relative to the η directions. For ease of graphical illustration, here we assume the specific orientation diagramed in Figs. 4(b) and 4(c). Using Eq. (12) and the orthogonality of X, we can parametrize $\hat{e}$ as

$$\hat{e} = \sum_{\mu=1}^{\mathcal{N}=3} X_{\mu 1} \hat{\varepsilon}_\mu = \cos\theta \hat{c} + \sin\theta \hat{s}, \tag{22}$$

where

$$\hat{c} = -\hat{\varepsilon}_3, \quad \text{and} \quad \hat{s} = -\cos\gamma \hat{\varepsilon}_1 + \sin\gamma \hat{\varepsilon}_2 \tag{23}$$

[Figs. 4(b) and 4(c)]. Geometrically, $\hat{c}$ and $\hat{s}$ are unit vectors along the projections of $\hat{e}$ onto the constrained and semiconstrained subspaces [Fig. 4(b)]. Thus, $\cos\theta \geqslant 0$ and $\sin\theta \geqslant 0$, making θ an acute angle. In this example, γ is also an acute angle, as depicted in Fig. 4(c).

Note that all solutions lie within the two-dimensional semiconstrained subspace having $\eta_3 = y_3$. The $w_1 = 0$ plane intersects this semiconstrained subspace as a line [Figs. 4(b) and 4(c)], and its equation in η coordinates is

$$w_1 = \hat{e} \cdot \vec{w} = \sin\theta(-\cos\gamma \eta_1 + \sin\gamma \eta_2) - \cos\theta y_3 = 0. \tag{24}$$

From the geometry of the problem [Fig. 4(c)], it is clear that if the perpendicular distance, d_s, from the origin to this line is large enough, then it will not intersect the all-negative quadrant of the semiconstrained subspace within the weight bound. According to simple trigonometry, this occurs when

$$d_s > \widetilde{W} \cos\gamma = \sqrt{W^2 - y_3^2} \cos\gamma, \tag{25}$$

where $\widetilde{W} = \sqrt{W^2 - y_3^2}$ is the radius of the semiconstrained subspace containing the solutions. The perpendicular distance
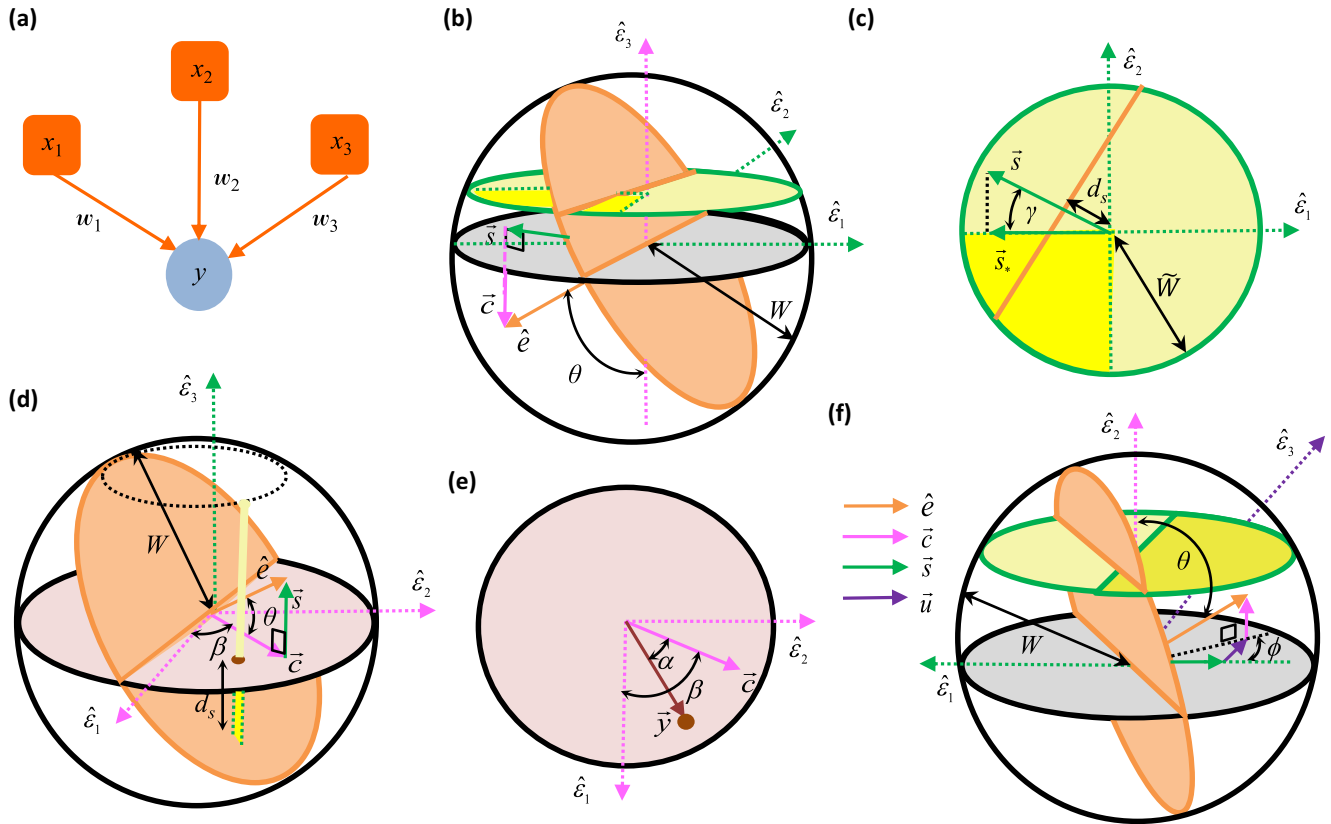
FIG. 4. Geometric quantities determining whether neurons must be synaptically connected in several three-dimensional toy problems. (a) Cartoon depicting the $\mathcal{N} = 3$ feedforward network corresponding to the toy problems. (b), (c) Geometrically determining whether a synapse is nonzero when the target neuron responds to one input pattern but does not to two other patterns. A synapse can only vanish if the $w_1 = 0$ plane (orange circle) intersects the solution space (dark yellow wedge) within the weight bounds (bounding sphere). For example, this intersection occurs in panel (b), so the synapse is not required for the responses. For every synapse one can associate a direction in synaptic weight space (orange arrow) that is normal to the planes with constant synaptic weight. This synapse vector can be decomposed into its projections into the semiconstrained subspace (green arrow, $\vec{s}$) and along the constrained dimension (pink arrow, $\vec{c}$). In this example, whether the synapse is certain is determined by the size of the bounding synapse space, W [see panel (b)], the angle θ between the synapse direction (orange arrow) and the closest axis of the constrained dimension ($-\hat{\varepsilon}_3$) [see panel (b)], and the angle γ between $\vec{s}$ and its closest vector in the solution space ($\vec{s}_*$) [see panel (c)]. In panel (c), d_s depicts the perpendicular distance from the origin of the yellow semiconstrained plane in panel (b) to its intersection line with the $w_1 = 0$ orange plane. If this distance is sufficiently large, then the orange line will not intersect the solution space within the yellow plane's circular bound of radius $\widetilde{W}$. (d), (e) Geometrically determining whether a synapse is nonzero when the target neuron responds to two input patterns but not the third pattern. In panel (d), the orange $w_1 = 0$ plane intersects the solution space (deep yellow line) within the bounding sphere, so the synapse is not certain. In this example, the factors that determine synapse certainty are W [see panel (d)], the angle θ that the synapse vector (orange arrow) makes with its projection along the constrained subspace (pink arrow) [see panel (d)], and the angle α between the target response vector (brown arrow) and the pink arrow [see panel (e)]. The angle β does not ultimately matter, but it is included in the diagrams to aid the derivation. Here d_s is the distance from the brown dot to the point of intersection between the yellow line and the orange plane. Again this point will lie outside the bounding sphere if d_s is large enough, and this signals a certain synapse. (f) Geometrically determining whether a synapse is nonzero when the target neuron responds to one input pattern but does not to a second pattern. In the figure shown, the $w_1 = 0$ orange plane intersects the solution space (deep yellow semicircle) within the bounding sphere, so the synapse is not certain. In this example, apart from W, what determines synapse certainty are the angles θ and ϕ, which encode how the synapse vector (orange arrow) can be decomposed into its projections along the constrained direction (pink arrow), semiconstrained direction (green arrow) and unconstrained direction (purple arrow).

can be identified from Eq. (24) as

$$d_s = y_3 \cot \theta. \tag{26}$$

Substituting this expression for d_s into Eq. (25), one finds through simple algebra that the $w_1 = 0$ hyperplane does not intersect the solution space, and hence the synapse is certain,

if the response magnitude exceeds a critical value,

$$y_3 > y_{\text{cr}} = W \sqrt{\frac{\sin^2 \theta \cos^2 \gamma}{\cos^2 \theta + \sin^2 \theta \cos^2 \gamma}}, \tag{27}$$

which we generally refer to as y-critical.

Notice that if θ increases in Fig. 4(b), then the orange line in Fig. 4(c) comes closer to the origin, making it intersect with

the solution space for more γ angles. Therefore, the synapse is more difficult to identify, and indeed Eq. (27) shows that y_{cr} increases. However, if γ increases, then the orange line in Fig. 4(c) rotates away from the solution space, making the synapse easier to identify with small d_s. Accordingly, y_{cr} decreases.

It will turn out that the concept of y-critical is general, and y_{cr} can always be expressed in terms of projections of $\hat{e}$ along several specific directions. In this example, if we define e_{s*} and e_y to be projections of $\hat{e}$ along $\hat{s}_* = -\vec{\varepsilon}_1$ and $\hat{y} = \vec{\varepsilon}_3$, respectively, then it is easy to check that one can re-express y_{cr} as

$$y_{cr} = W \sqrt{\frac{e_{s*}^2}{e_y^2 + e_{s*}^2}}. \qquad (28)$$

We will later discover that these projections are closely related to correlations between pre-synaptic and postsynaptic neuronal activity patterns. Thus, the expressions in Eq. (28) will provide a deeper understanding of the determinants of synapse certainty.

B. Problem 2

Having identified two key angles, θ and γ, that play a role in synapse certainty, let us look at the example of two constrained and one semiconstrained dimensions to uncover other important geometric quantities. In this case, the solution space is a ray defined by $\eta_1 = y_1$, $\eta_2 = y_2$, and $-\infty < \eta_3 \leqslant 0$, and the magnitude of η_3 is at most

$$\widetilde{W} = \sqrt{W^2 - y_1^2 - y_2^2} \qquad (29)$$

for solutions within the weight bound [Fig. 4(d)]. Figure 4(d) shows a geometry where the $w_1 = 0$ plane intersects the solution space at the point

$$\vec{w}_{int} = y_1\vec{\varepsilon}_1 + y_2\vec{\varepsilon}_2 + \eta_3\vec{\varepsilon}_3. \qquad (30)$$

Now we must have

$$\hat{e} \cdot \vec{w}_{int} = 0, \qquad (31)$$

as the intersection point lies on the $w_1 = 0$ plane by definition, where we have defined $\hat{e} \equiv \hat{e}_1$ as in the previous toy problem. The projection directions of $\hat{e}$ onto the constrained and semiconstrained subspaces are given by

$$\hat{c} = \cos\beta\hat{\varepsilon}_1 + \sin\beta\hat{\varepsilon}_2, \quad \text{and} \quad \hat{s} = \hat{\varepsilon}_3, \qquad (32)$$

[Fig. 4(d)]. Then combining Eqs. (22) and (32), we can find an equation to determine η_3 at the intersection point

$$\hat{e} \cdot \vec{w}_{int} = \cos\theta(y_1\cos\beta + y_2\sin\beta) + \eta_3\sin\theta = 0. \qquad (33)$$

We next introduce α to represent the angle between $\hat{c}$ and $\vec{y}$ [Fig. 4(e)], such that

$$y_1 = y\cos(\beta - \alpha) \quad \text{and} \quad y_2 = y\sin(\beta - \alpha), \qquad (34)$$

where $y = |\vec{y}|$. The first two terms in Eq. (33) can then be trigonometrically combined with a difference of angles identity to arrive at

$$y\cos\theta\cos\alpha + \eta_3\sin\theta = 0 \implies \eta_3 = -y\cot\theta\cos\alpha. \qquad (35)$$

To be able to identify the sign of w_1, this intersection point must lie beyond the weight bounds of the solution line segment, so $\eta_3 < -\widetilde{W}$. After some straightforward algebra we obtain the certainty condition as

$$y > y_{cr} \equiv W\sqrt{\frac{\sin^2\theta}{\cos^2\theta\cos^2\alpha + \sin^2\theta}}. \qquad (36)$$

From the geometry of the problem in Figs. 4(d) and 4(e), one sees that as θ or α increases, the point where the orange hyperplane intersects the yellow line is closer to the origin. Indeed, y_{cr} increases, making it more difficult to identify the synapse sign. Again, one can re-express y_{cr} as Eq. (28) in terms of projections, with the role of $\hat{s}_*$ being played by $\vec{\varepsilon}_3$.

C. Problem 3

Through the two above examples we found three angles, θ, α, and γ, that determine how large the response of the driven neuron has to be in order for a given synapse to be certain. However in both examples the number of patterns were equal to the number of synapses, $\mathcal{P} = \mathcal{N}$. When $\mathcal{P} < \mathcal{N}$, we have unconstrained dimensions, and the projection of the $\hat{e} \equiv \hat{e}_1$ vector into the unconstrained subspace will also matter, because it relates to how much we do not know about the response properties of the driven neuron.

Here we consider a $\mathcal{N} = 3$ example with one constrained, one semiconstrained, and one unconstrained dimension [Fig. 4(f)]. In this case, we can express the $\hat{e}$ synaptic direction as a linear combination of its projections along the constrained, semiconstrained and unconstrained dimension as

$$\hat{e} = \sum_{\mu=1}^{\mathcal{N}} X_{\mu 1}\vec{\varepsilon}_\mu = \cos\theta\hat{c} + \sin\theta\cos\phi\hat{s} + \sin\theta\sin\phi\hat{u}, \qquad (37)$$

where we can always choose the directions of the unit vectors to make θ and ϕ acute angles. For the example shown in Fig. 4(f), this is achieved by choosing

$$\hat{c} = \vec{\varepsilon}_2, \hat{s} = -\vec{\varepsilon}_1, \quad \text{and} \quad \hat{u} = \vec{\varepsilon}_3. \qquad (38)$$

Obtaining the certainty condition again involves ascertaining whether the $w_1 = 0$ hyperplane intersects the deep yellow solution space [Fig. 4(f)]. In the example of Fig. 4(f), one can see that increasing the driven neuron response moves the yellow plane up, and there will come a critical point when the orange $w_1 = 0$ plane just touches the solution space at the corner $(\eta_1 = 0, \eta_2 = y_{cr}, \eta_3)$. Thus,

$$\vec{w}_{int} = y_{cr}\vec{\varepsilon}_2 + \eta_3\vec{\varepsilon}_3. \qquad (39)$$

Since this corner point has a negative η_3 component and lies on the bounding sphere, we must also have

$$\eta_3 = -\sqrt{W^2 - y_{cr}^2}, \qquad (40)$$

[Fig. 4(f)]. Substituting $\vec{w}_{int}$ in the $w_1 = 0$ plane equation,

$$\hat{e} \cdot \vec{w}_{int} = -\sin\theta\sin\phi\sqrt{W^2 - y_{cr}^2} + \cos\theta\, y_{cr} = 0, \qquad (41)$$

we can then determine y_{cr} through simple algebra as

$$y_{cr} = W\sqrt{\frac{\sin^2\theta\sin^2\phi}{\cos^2\theta + \sin^2\theta\sin^2\phi}}. \qquad (42)$$

The final result now depends on the two acute orientation angles, θ and ϕ. By inspection of Fig. 4(f) or Eq. (42), it is clear that y_{cr} increases if either θ or ϕ increases toward $\pi/2$. One therefore needs a larger response (y_3) to make the synapse certain. We can again express y_{cr} in terms of projections

$$y_{cr} = W \sqrt{\frac{e_u^2}{e_y^2 + e_u^2}}, \qquad (43)$$

where e_u is the projection of $\hat{e}$ along $\hat{u}$, and e_{s*} does not appear because the intersection occurred at the origin of the semiconstrained subspace.

V. CERTAIN SYNAPSES, THE GENERAL TREATMENT

A. High-dimensional feedforward networks

We have seen in the previous section how geometric considerations can identify synapses that must be present to generate observed response patterns in small networks. One can similarly ask when a synapse is required in high-dimensional networks [Fig. 5(a)].

Although the rigorous derivation is intricate, this certainty condition is remarkably simple for orthonormal X (Appendix A). Quantitatively, orthonormal X imply that only a few parameters matter for the certainty condition, each illustrated in the previous section and abstractly summarized in Fig. 5(b).

For any given synapse, its physical basis vector, $\hat{e}_m$, can always be written as a sum of components in the constrained, semiconstrained, and unconstrained subspaces,

$$\hat{e}_m = \sum_{\mu=1}^{\mathcal{N}} X_{\mu m} \vec{\varepsilon}_\mu = \vec{c}_m + \vec{s}_m + \vec{u}_m, \qquad (44)$$

where $\vec{c}_m$, $\vec{s}_m$, and $\vec{u}_m$ denote the partial sums over μ in the constrained, semiconstrained, and unconstrained subspaces, respectively. Note that $\{\vec{\varepsilon}_\mu\}$ are orthogonal unit vectors if and only if X is an orthogonal matrix. In this case, the decomposition of $\hat{e}_m$ is a sum of three orthogonal vectors that can be parameterized by two angles,

$$\hat{e}_m = \cos\theta \, \hat{c}_m + \sin\theta \cos\phi \, \hat{s}_m + \sin\theta \sin\phi \, \hat{u}_m, \qquad (45)$$

where $\hat{c}_m$, $\hat{s}_m$, and $\hat{u}_m$ are unit vectors in the constrained, semiconstrained, and unconstrained subspaces, and (θ, ϕ) are spherical coordinates[6] specifying the orientation of $\hat{e}_m$ with respect to these subspaces [e.g., Fig. 4(f)]. In particular,

$$\cos\theta = \sqrt{\sum_{\{\mu | y_\mu > 0\}}^{\mathcal{P}} X_{\mu m}^2},$$

$$\sin\theta \cos\phi = \sqrt{\sum_{\{\mu | y_\mu = 0\}}^{\mathcal{P}} X_{\mu m}^2},$$

$$\sin\theta \sin\phi = \sqrt{\sum_{\mu = \mathcal{P}+1}^{\mathcal{N}} X_{\mu m}^2}. \qquad (46)$$

---

[6]The angles also depend on the synapse but we have dropped the m index for brevity.
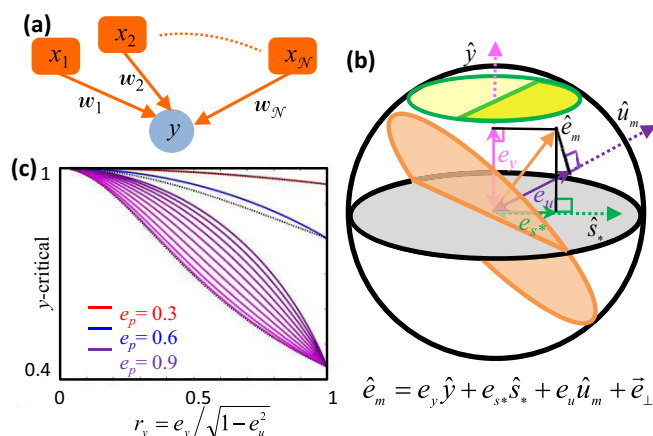

FIG. 5. Identifying certain synapses in high-dimensional networks. (a) Cartoon depicting the high-dimensional feedforward network under consideration. (b) Geometrically determining whether a synapse is nonzero throughout a high-dimensional solution space. A synapse can only vanish if the $w_m = 0$ hyperplane (orange circle) intersects the solution space (dark yellow wedge) within the weight bounds (bounding sphere). In the example shown, this intersection does not occur, so the synapse must be present. For orthonormal neural responses, only a few parameters determine whether this intersection occurs (Appendix A). First, the magnitude of the weight bound, W, controls the extent of the solution space. Second, there are three projections of the synapse direction (orange arrow) whose lengths are important determinants of the certainty condition: e_y, the length of projection along the target response vector (pink arrow); e_{s*}, the length of projection along the closest boundary vector in the semiconstrained solution subspace [green arrow, see also $\vec{s}_*$ in Fig. 4(c)]; and e_u, the length of projection into the unconstrained subspace (purple arrow). Note that the shown example would have had an intersection if the solution space (dark yellow wedge) were moved down (along $\hat{c}$) to lie below the hyperplane (orange circle). The solution space's height is proportional to the magnitude of the postsynaptic responses, y. Thus, the solution space does not intersect the hyperplane only if y exceeds a critical value, y_{cr}. (c) Plots of the certainty condition, Eq. (57), for $W = 1$. The red, blue, and purple curves plot y_{cr} as a function of $r_y = e_y / e_p$ for $e_p = 0.3$, 0.6, and 0.9, respectively. Different purple shades correspond to different values of $r_{s*} = e_{s*} / \sqrt{e_p^2 - e_y^2}$. As this ratio increases, nonlinear effects increase y_{cr} and make the sign harder to determine. The red and blue curves are for the maximally nonlinear case when $r_{s*} = 1 \Rightarrow e_{s*} = \sqrt{e_p^2 - e_y^2}$. The dashed black curves represent y_{cr} in a linear model, which cannot exceed the nonlinear y_{cr}.

As we have seen in the toy examples, these two orientation angles heavily influence whether the synapse is certain.

Additionally, because the solution space's height along $\hat{c}_m$ [e.g., Fig. 4(b)] is controlled by the angle between $\hat{c}_m$ and $\hat{y}$, the equation for the $w_m = 0$ hyperplane that divides the positive and negative synaptic regions in the solution space depends on

$$\vec{y} \cdot \hat{c}_m = y \cos\alpha, \qquad (47)$$

where y is the length of $\vec{y}$ and α is the angle between $\vec{y}$ and $\hat{c}_m$ [Fig. 4(e)]. Finally, there is another critical angle, which we call γ, that encodes how $\hat{s}_m$ is oriented with respect to the solution space in the semiconstrained subspace. Using

a more convenient direction, $\widehat{s}'_m \equiv -\text{Sgn}(\cos\alpha)\widehat{s}_m$, which is either along or opposite to the $\widehat{s}_m$ direction, we define γ to be the minimal angle between $\widehat{s}'_m$ and the solution space [e.g., Fig. 4(b)]. It is generally given by

$$\cos\gamma = \sqrt{\sum_{\{\mu|\widehat{s}'_\mu < 0\}} \widehat{s}'^2_\mu} \qquad (48)$$

(Appendix A), where $\widehat{s}'_\mu$ is the μth component of $\widehat{s}'_m$, and we have suppressed m to avoid cluttered notation. Although this definition and equation for γ may initially appear opaque, we soon clarify its meaning in terms of interpretable projections of the synapse vector.

Putting all the pieces together, we find that the mth synapse must be present, and its sign is unambiguous, if and only if y exceeds the critical value

$$y_{\text{cr}} = W\sqrt{\frac{\cos^2\gamma \sin^2\theta \cos^2\phi + \sin^2\theta \sin^2\phi}{\cos^2\alpha \cos^2\theta + \cos^2\gamma \sin^2\theta \cos^2\phi + \sin^2\theta \sin^2\phi}} \qquad (49)$$

(Appendix A). Intuitively, W bounds the magnitude of weight vectors, and large W increase y_{cr} by admitting more solutions. Note that a synapse is certain, for a given y, when the weight bound is less than a critical value,

$$W_{\text{cr}} = y\sqrt{\frac{\cos^2\alpha \cos^2\theta + \cos^2\gamma \sin^2\theta \cos^2\phi + \sin^2\theta \sin^2\phi}{\cos^2\gamma \sin^2\theta \cos^2\phi + \sin^2\theta \sin^2\phi}}. \qquad (50)$$

Finally, we note that we must have $W \geqslant y$ for any solutions to exist. One can straightforwardly obtain the special cases Eqs. (27), (36), and (42), by substituting $\alpha = \phi = 0$, $\gamma = \phi = 0$, and $\alpha = \cos\gamma = 0$ in the general expression given by Eq. (49).

The geometric description of Eq. (49) can be written more intuitively as

$$y_{\text{cr}} = W\sqrt{\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}} = W\sqrt{\frac{1}{1 + e_y^2/(e_{s*}^2 + e_u^2)}} \qquad (51)$$

(Appendix A), where $\widehat{s}_*$ is the unit vector in the solution space that is most aligned with $\widehat{s}'_m$ [e.g., Fig. 4(b)], and e_y, e_{s*}, and e_u are the projections of $\widehat{e}_m$ onto $\widehat{y} = \vec{y}/y$, $\widehat{s}_*$, and $\widehat{u}_m$ [Fig. 5(b)]. Indeed, Eqs. (28) and (43) can be readily recognized as special cases of the above general expression.

Each of these projections is interpretable in light of the fact that $x_{\mu m}$ represents the activity level of the mth presynaptic neuron in the μth response pattern. Most simply,

$$e_y = \widehat{e}_m \cdot \widehat{y} = \frac{\sum_{\mu=1}^{\mathcal{P}} y_\mu x_{\mu m}}{\sqrt{\sum_{\nu=1}^{\mathcal{P}} y_\nu^2}} \qquad (52)$$

is a normalized correlation of the pre- and postsynaptic activity (note that $\sum_{\rho=1}^{\mathcal{N}} X_{\rho m}^2 = 1$). As expected, synapse certainty is aided by large magnitudes of e_y. Moreover, the sign of a certain synapse is the sign of this correlation, or equivalently the sign of e_y. Synapse sign identifiability is hindered by large

values of

$$e_u = \widehat{e}_m \cdot \widehat{u}_m = \sqrt{1 - \sum_{\mu=1}^{\mathcal{P}} x_{\mu m}^2}, \qquad (53)$$

which effectively measures the weakness of the presynaptic neuron's activity, as it is the amount of presynaptic drive for which we do not have any information on the target neuron's response. The more subtle quantity is

$$e_{s*} = \widehat{e}_m \cdot \widehat{s}_* = -\text{Sgn}(\cos\alpha)\sqrt{\sum_{\{\mu|s'_\mu < 0\}}^{\mathcal{P}} x_{\mu m}^2}$$

$$\implies e_{s*}^2 = \sum_{\{\mu|s'_\mu < 0\}}^{\mathcal{P}} x_{\mu m}^2. \qquad (54)$$

The condition that $s'_\mu < 0$ selects for patterns where the sign of the presynaptic activity is $\text{Sgn}(\cos\alpha) = \text{Sgn}(e_y)$, but the postsynaptic neuron does not respond. In other words, presynaptic activity should have promoted a response in the target neuron according to the observed activity correlation. That it does not generates uncertainty in the sign of the synapse. See Appendix A for a heuristic derivation of y_{cr} based on this argument.

We can gain more useful intuition by interpreting our result in relation to what we would obtain in a linear neural network. In the linear problem, there are only constrained and unconstrained dimensions; every dimension that was semiconstrained in the nonlinear problem becomes constrained, with all solutions having $\eta_\mu = y_\mu$ for $\mu = 1, \ldots, \mathcal{P}$. This implies that

$$y_{\text{cr,lin}} = W\sqrt{\frac{e_u^2}{e_y^2 + e_u^2}}. \qquad (55)$$

Returning to the nonlinear problem, recall that the certainty condition finds the largest y for which the solution space and $w_m = 0$ hyperplane intersect within the weight bound, and this intersection is simply a point when $y = y_{\text{cr}}$. Importantly, each semiconstrained dimension can either behave like a linear constrained dimension with $\eta_\mu = y_\mu = 0$ at this intersection point (toy problems 1 and 3), or like an unconstrained dimension with $\eta_\mu < 0$ at the intersection point (toy problems 1 and 2).[7] The first case occurs when $\widehat{e}_m \cdot \vec{\varepsilon}_\mu$ and $e_y = \widehat{e}_m \cdot \widehat{y}$ have opposite signs and $s'_\mu > 0$; the second case occurs when they have the same sign and $s'_\mu < 0$. This means that one could compute the nonlinear theory's y-critical from $y_{\text{cr,lin}}$ by appending the second class of semiconstrained dimensions onto the unconstrained dimensions. Mathematically, this corresponds to the replacement

$$e_u^2 \to e_u^2 + e_{s*}^2 = \sum_{\mu=\mathcal{P}+1}^{\mathcal{N}} x_{\mu m}^2 + \sum_{\{\mu|s'_\mu < 0\}}^{\mathcal{P}} x_{\mu m}^2, \qquad (56)$$

---

[7]Since this intersection point depends on m, the semiconstrained dimension indexed by μ can behave as constrained for some synapses and unconstrained for others.

which indeed transforms Eq. (55) to Eq. (51). The role of e_{s*}^2 is to quantify the uncertainty introduced by the subset of semiconstrained dimensions that do not behave as constrained at the intersection point.

Since the parameters e_y, e_{s*}, and e_u cannot be set independently, it is convenient to reparameterize Eq. (51) as

$$y_{\text{cr}} = W \sqrt{\frac{1}{1 + r_y^2 e_p^2 / \{1 - e_p^2 [1 - r_{s*}^2 (1 - r_y^2)]\}}}, \quad (57)$$

where $e_p^2 = 1 - e_u^2$, $r_y^2 = e_y^2 / e_p^2$, $r_{s*}^2 = e_{s*}^2 / (e_p^2 - e_y^2)$, and all three composite parameters can be independently set between 0 and 1. Conceptually, r_y and r_{s*} merely normalize e_y and e_{s*} by their maximal values, and e_p is the projection of $\hat{e}_m$ into the activity-constrained subspace spanned by both constrained and semiconstrained dimensions. One could also interpret r_{s*} as quantifying the effect of threshold nonlinearity. For instance, $r_{s*} = 0$ describes the case where all semiconstrained dimensions are effectively constrained, but r_{s*} increases as some of the semiconstrained dimensions start to behave like unconstrained dimensions. As expected, y_{cr} is a decreasing function of r_y^2 and e_p^2 and an increasing function of r_{s*}^2 [Fig. 5(c)].

B. Regarding nonorthogonal input patterns

While a complete treatment of the certainty condition for generally correlated input patterns is beyond the scope of this paper, we could find a conservative bound for y-critical that may be useful when patterns are close to being orthogonal. The details of the derivation are discussed in the final subsection of Appendix A.

The major challenge caused by nonorthogonal patterns is that the spherical weight space becomes elliptical in terms of the η coordinates. Thus, the main idea behind the bound is that one can always find the sphere that just encompasses this ellipse. We can then use our formalism to obtain a conservative y-critical, such that if the norm of $\vec{y}$ is larger than this value then all solutions within the encompassing sphere have a consistent sign for the synapse under consideration. An interesting insight that emerges from our analysis is that the relative orientations between

$$\vec{n}_m \equiv \sum_{\mu=1}^{\mathcal{N}} X_{m\mu}^{-1} \vec{\varepsilon}_\mu \quad (58)$$

and the various important η directions play the role of θ, ϕ, α and γ (Appendix A). Note that $\vec{n}_m = \hat{e}_m$ when X is an orthogonal matrix. We anticipate that $\vec{n}$ will also be an important player in a more comprehensive treatment of nonorthogonal patterns.

C. Application to recurrent networks

As we explained in Sec. III, to find the ensemble of all incoming weight vectors onto the ith driven neuron, one can use the results obtained for the feedforward network and just substitute X with the $Z^{(i)}$ matrix. Consequently, identifying certain synapses onto the ith neuron can follow the route outlined for the feedforward scenario as long as $Z^{(i)}$ is orthogonal. So for example, if we want to ascertain whether any
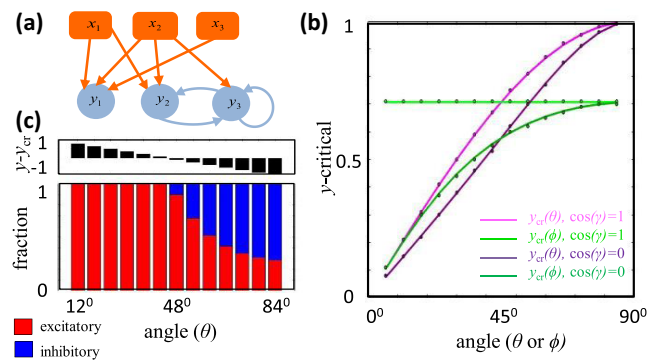

FIG. 6. Testing the certainty condition with exhaustive low-dimensional simulations. (a) A simple recurrent network with three input neurons and three driven neurons (Appendix E). (b) We plot the theoretically derived y_{cr} for feedforward synapses to y_1 as we vary θ (green curves) or ϕ (magenta curves), keeping the other angle fixed at 45°. The lighter shades correspond to $\cos \gamma = 1 \Rightarrow r_{s*} = 1$. The darker shades correspond to $\cos \gamma = 0 \Rightarrow r_{s*} = 0$, where the predictions from the nonlinear network match those of a linear network. The dots represent y_{cr} estimated through simulations, and they agree well with the theory. (c) (Bottom) Bar graph of the fraction of solutions with positive (red) and negative (blue) self-couplings ($y_3 \to y_3$) as a function of θ. (Top) As predicted, all solutions have positive w_{y_3, y_3} when $y - y_{\text{cr}} > 0$.

incoming synapse to the ith neuron is certain, we have to replace $x_{\mu m} \to z_{\mu m}^{(i)}$ and $y_\mu \to y_{\mu i}$ in Eqs. (51)–(54) to compute y_{cr}.

D. Numerical illustration of the certainty condition

To illustrate and test the theory numerically, we first considered a small neural network of three input neurons and three driven neurons [Fig. 6(a)]. This small number of synapses meant that we could comprehensively scan the entire spherical weight space without relying on a numerical algorithm to find solutions.[8] This is important because numerical techniques, such as gradient descent learning, potentially find a biased set of solutions that incompletely test the theory. We supposed that each driven neuron has three inputs, and we constrained weights with two orthonormal stimulus responses. We set $W = 1$ for all simulations and numerically screened weights randomly. See Appendix F for complete simulation details.

The first driven neuron in Fig. 6(a), y_1, receives only feedforward drive, and we suppose that it responds to one stimulus condition with response y ($\mu = 2$), but it does not respond to the other ($\mu = 1$). Its synapses thus have one constrained, one semiconstrained, and one unconstrained dimension, and all of the terms in Eq. (49) contribute to y-critical. We could thus use y_1 to verify Eq. (49). Moreover, this scenario includes the illustrative example of Fig. 4(f) as a special case, so we could also use y_1 to verify Eq. (42).

---

[8]Our results for the certainty condition hold for network ensembles that exactly generate the desired responses. For numerical tests, we had to allow for small deviations from the desired responses, but our predictions proved robust.

To these ends, we decided to focus on a two-parameter family of input patterns,

$$x = \begin{pmatrix} -\sin\psi\cos\chi & \cos\psi\cos\chi & \sin\chi \\ \cos\psi & \sin\psi & 0 \end{pmatrix}, \quad (59)$$

where rows correspond to different input patterns and columns correspond to different input neurons, as usual, and we extend x to the full-rank orthogonal matrix

$$X = \begin{pmatrix} -\sin\psi\cos\chi & \cos\psi\cos\chi & \sin\chi \\ \cos\psi & \sin\psi & 0 \\ \sin\psi\sin\chi & -\cos\psi\sin\chi & \cos\chi \end{pmatrix}. \quad (60)$$

By Eq. (44), the physical basis vector corresponding to the synapse from the first input neuron is thus

$$\widehat{e}_1 = \cos\psi\,\vec{\varepsilon}_2 + \sin\psi\cos\chi(-\vec{\varepsilon}_1) + \sin\psi\sin\chi\,\vec{\varepsilon}_3, \quad (61)$$

and it has the same general form as Eqs. (37) and (45), where $\vec{\varepsilon}_3$ plays the role of $\widehat{u}$. If ψ and χ are both acute, then one can identify them with θ and ϕ in Fig. 4(f), and the roles of $\widehat{c}$ and $\widehat{s}$ are played by $\vec{\varepsilon}_2$ and $-\vec{\varepsilon}_1$, respectively. In this case $\alpha = 0$, $\cos\gamma = 0$, and the theoretical dependencies of y_{cr} on θ and ϕ are given by Eq. (42). Figure 6(b) illustrates these dependencies as the purple and dark green curves. If ψ is acute, but χ is obtuse, then according to our conventions, $\theta = \psi$, $\phi = \pi - \chi$, $\widehat{c} = \vec{\varepsilon}_2$, and $\widehat{s} = \vec{\varepsilon}_1$. Now $\alpha = 0$ and $\cos\gamma = 1$, and our general formula, Eq. (49), implies

$$y_{\text{cr}} = W\sin\theta = W\sin\psi. \quad (62)$$

These dependencies are plotted as the pink and the light green curves in Fig. 6(b). We do not plot cases where ψ is obtuse, because obtuse and acute ψ result in equivalent y_{cr} formulas. Whether ψ is acute or obtuse nevertheless matters because it determines the sign of the w_1 synapse when it is certain.

The black dots in Fig. 6(b) show the largest response magnitude, y, for which we numerically found solutions with both positive and negative w_1 (see Appendix F for numerical methods), thereby providing a numerical estimate of y_{cr}. The theoretical curves and numerical points precisely aligned in all cases. The differences between the light and dark theoretical curves illustrates the effect of nonlinearity. When χ is obtuse, the semiconstrained dimension effectively behaves as unconstrained, and the mixing angle between the semiconstrained and unconstrained dimension is irrelevant to y-critical. When χ is acute, the semiconstrained dimension effectively behaves as constrained, as if its coordinate were set to zero. Moreover, these results confirmed that stronger responses were needed to make synapses fixed sign when the synaptic direction was less aligned with the constrained dimension [Fig. 6(b), purple and pink]. Furthermore, smaller y-critical values occurred when the synaptic direction anti-aligned with the semiconstrained dimension [Fig. 6(b), purple versus pink, dark green versus light green].

We next wanted to check the validity of our results for the recurrently connected neurons in Fig. 6(a). We therefore needed to tailor the steady-state activity levels of the recurrent network to result in orthogonal presynaptic input patterns for each driven neuron. In mathematical terms, $Z^{(i)}$ must be an orthogonal matrix for $i = 1, 2, 3$. We achieved this by considering a two-parameter family of driven neuronal responses

in which the activity patterns of y_2 and y_3 were matched to those of x_1 and x_3, respectively. This construction means that all three driven neurons receive the same input patterns. To ensure positivity of driven neuronal responses, we set χ as an acute angle and ψ as the negative of an acute angle.

Although y_2 has both feedforward and recurrent inputs, we can analyze its connectivity in exactly the same way as y_1. Recurrence only complicates the analysis for neurons that synapse onto themselves, like y_3, since changing the output activity also changes the input drive. So $|\vec{y}_3|$ and y_{cr} are not independent. Here we focused on the certainty condition for the self-synapse, w_{y_3, y_3}, for which $y_{\text{cr}} = \cos\chi$, and $|\vec{y}_3| = \sin\chi$. Therefore, the synapse should be certain if $45^0 < \chi \leqslant 90^0$. Since $\theta = \pi/2 - \chi$ according to our conventions,[9] this is equivalent to $0 \leqslant \theta < 45^0$ [Fig. 6(c), top]. Our numerical results precisely recapitulated these theoretical expectations [Fig. 6(c), bottom], as the self-connection was consistently positive across all simulations whenever this condition on θ was met. See Appendix E for certainty condition analyses for other synapses onto y_3 and Appendix F for complete simulation details.

VI. ACCOUNTING FOR NOISE

A. Finding the solution space in the presence of noise

So far we have only considered exact solutions to the fixed point equations. However, it is also important to determine weights that lead to fixed points near the specified ones. For example, biological variability and measurement noise generally make it infeasible to specify exact biological responses. Furthermore, numerical optimization typically produces model networks that only approximate the specified computation. We therefore define the $\mathcal{E}$-*error surface* as those weights that generate fixed points a distance $\mathcal{E}$ from the specified ones,

$$\mathcal{V}_{\mathcal{E}} = \left\{ w \,\middle|\, \sum_{\mu=1}^{\mathcal{P}} \sum_{i=1}^{\mathcal{D}} (y_{\mu i} - \widetilde{y}_{\mu i}(w))^2 = \mathcal{E}^2 \right\}, \quad (63)$$

where $y_{\mu i}$ is the specified activity of the ith driven neuron in the μth fixed point, and $\widetilde{y}_{\mu i}$ is the corresponding activity level in the fixed point approached by the model network when it is initialized as $y_i(t = 0) = y_{\mu i}$. If the network dynamics do not approach a fixed point, perhaps oscillating or diverging instead [54], we say $\mathcal{E} = \infty$.

Each $\mathcal{E}$-error surface can be found exactly for feedforward networks. For illustrative purposes, let us first consider the $\mathcal{D} = 1$ feedforward scenario in which the driven neuron is active in every response pattern. This means that $y_\mu > 0$ for all $\mu = 1, \dots, \mathcal{P}$, and we can reorder the μ indices to sort the driven neuron responses in ascending order, $0 < y_1 < y_2 < \cdots < y_{\mathcal{P}}$. Here we assumed that no two response levels are exactly equal, as is typical of noisy responses. Since all responses are positive, the zero-error solution space has

---

[9] For y_3, $\mu = 1, 2$ are constrained and semiconstrained, respectively. Accordingly,

$$\widehat{e}_{y_3 y_3} = \sin\chi\,\vec{\varepsilon}_1 + \cos\chi\,\vec{\varepsilon}_3 = \cos\theta\,\vec{\varepsilon}_1 + \sin\theta\,\vec{\varepsilon}_3 = \cos\theta\,\widehat{c} + \sin\theta\,\widehat{u}.$$
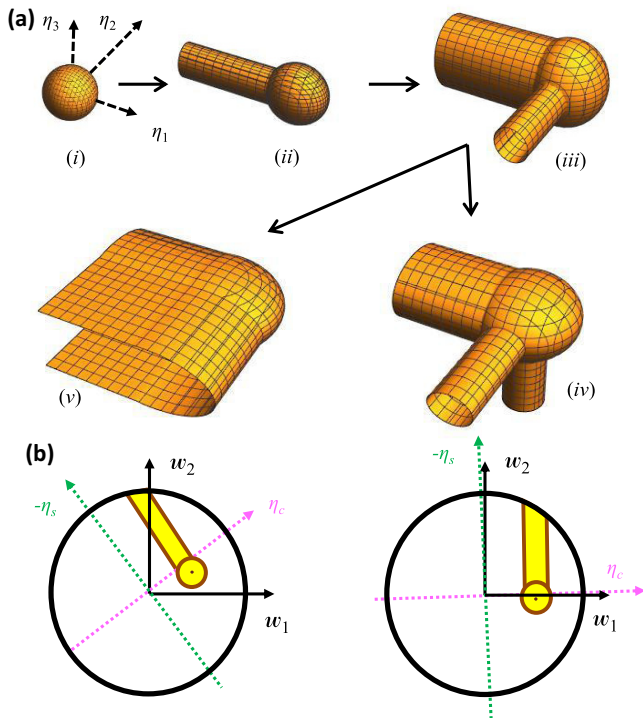
FIG. 7. The solution space geometry changes as the allowed error increases. (a) Error surface contours in a three-dimensional subspace corresponding to η_1, η_2, and η_3. Several topological transitions occur as the error increases. (i) We consider the case where all responses are positive, so the contours are spherical for small errors, just like in a linear neural network. (ii), (iii) Two cylindrical dimensions sequentially open up when the error is large enough for some η coordinates to become negative. (iv), (v) After that, either a third cylindrical dimension can open up, or the two cylindrical axes can join to form a plane. Which transition occurs at lower error depends on the pattern of neural responses. (b) (Left) We illustrate a case where there is a unique exact solution to the problem (brown dot). Allowing error but neglecting topological transitions would expand the solution space to an ellipse (here, brown circle), but the signs of w_1 and w_2 remains positive. Including topological transitions in the error surface can cap the ellipse with a cylinder (full yellow solution space). Now we can say with certainty that the sign of w_2 is positive, but negative values of w_1 become possible. (Right) Graphical conventions are the same. However, in this case all solutions inside the cylinder have $w_2 > 0$. Therefore, the topological transition breaks a near symmetry between positive and negative weights.

no semiconstrained dimensions, and the only freedom for choosing w is in the $\mathcal{U} = \mathcal{N} - \mathcal{P}$ unconstrained dimensions. Therefore, the zero-error surface of exact solutions, $\mathcal{V}_0$, is a $\mathcal{U}$-dimensional linear subspace, and $\mathcal{V}_0$ is a point in the $\mathcal{P}$-dimensional activity-constrained subspace.

How does this geometry change as we allow error? For $0 < \mathcal{E} < y_1$, we must have $\widetilde{y}_\mu > 0$ for all μ. Therefore, the nonlinearity is irrelevant, and $\mathcal{E}$-error surfaces are spherical in the activity-constrained η coordinates [Eq. (63), Fig. 7(a(i))]. However, once $\mathcal{E} = y_1$ it becomes possible that $\widetilde{y}_1 = 0$, and suddenly a semi-infinite line of solutions appears with $\eta_1 \leqslant 0$. As $\mathcal{E}$ further increases, this line dilates to a high-dimensional

cylinder [Fig. 7(a(ii))]. A similar transition happens at $\mathcal{E} = y_2$, whereafter two cylinders cap the sphere [Fig. 7(a(iii))]. Things get more interesting as $\mathcal{E}$ increases further because two transitions are possible. A third cylinder appears at $\mathcal{E}' = y_3$. However, at $\mathcal{E}'' = \sqrt{(y_1)^2 + (y_2)^2}$ it is possible for both $\widetilde{y}_1$ and $\widetilde{y}_2$ to be zero, and the two cylindrical axes merge into a semi-infinite hyperplane defined by $\eta_1 \leqslant 0$, $\eta_2 \leqslant 0$. Thus, when $\mathcal{E}' < \mathcal{E}''$ the error surface grows to attach a third cylinder [Fig. 7(a(iv))], and when $\mathcal{E}'' < \mathcal{E}'$ the two cylindrical surfaces merge to also include planar surfaces in between [Fig. 7(av)]. These topological transitions continue by adding new cylinders and merging existing ones, and the sequence is easily calculable from $\{y_\mu\}$. Note that we use the terminology "topological transition" to emphasize that the structure of the error surface changes discontinuously at these values of error. The geometric transitions we observe here also relate to topological changes in a formal mathematical sense. For instance, while there are no incontractible circles in Fig. 7(a(ii)), one develops as we transition to Fig. 7(a(iii)).

In general, y_μ may also be zero or negative in the presence of noise. Whenever $y_\mu = 0$, the μth response pattern generates a semiconstrained dimension in $\mathcal{V}_0$. However, if some response levels are negative, then there are no exact solutions at all. However, it becomes possible to find solutions when $\mathcal{E} = \sqrt{\sum_{\{\mu | y_\mu < 0\}} y_\mu^2}$, and each response pattern associated with a negative y_μ acts as a semiconstrained dimension in $\mathcal{V}_\mathcal{E}$. As illustrated above, more semiconstrained dimensions open up as more error is allowed in each of these cases.

This geometry only approximates $\mathcal{E}$-error surfaces for recurrent networks (Appendix C). For instance, displacing $y_{\mu i}$ from its specified value changes the input pattern that define the $\vec{\varepsilon}_\mu$ directions for downstream driven neurons, but this effect is neglected here. We will nevertheless find that this feedforward approximation to $\mathcal{E}$-error surfaces is practically useful for predicting synaptic connectivity in recurrent networks as well.

B. Predicting connectivity in the presence of noise

The threshold nonlinearity and error-induced topological transitions can have a major impact on synapse certainty [Fig. 7(b)]. For example, one might model a neuronal dataset with a linear neural network and find that models with acceptably low error consistently have positive signs for some synapses. However, if measured neural activity was sometimes comparable to the noise level, then semiconstrained dimensions could open up that suddenly make some of these synapse signs ambiguous [Fig. 7(b), left]. Although semiconstrained dimensions can never make an ambiguous synapse fully unambiguous, semiconstrained dimensions can heavily affect the distribution of synapse signs across the model ensemble by providing a large number of solutions that have consistent anatomical features [Fig. 7(b), right].

We therefore generalized the certainty condition to include the effects of error, including topological transitions in the error surface (Appendix C). As before, finding the certainty condition amounts to determining when the $w_m = 0$ hyperplane intersects the solution space within the weight bound, but to account for noise of magnitude ε, we must now check

whether an intersection occurs with any $\mathcal{E}$-error surface with $\mathcal{E} \leqslant \varepsilon$. No intersections will occur if and only if every nonnegative $\vec{\bar{y}}$ within ε of the provided $\mathcal{P}$-vector of noisy target neuron activity [Fig. 3(c)] satisfies its zero-error certainty condition, and each $\vec{\bar{y}}$ is a possible denoised version of it [Eq. (63)]. We thus define y-critical in the presence of noise as the maximal y_{cr} [Eq. (51)] among this set of $\vec{\bar{y}}$.

Although we lack an exact expression for y-critical in the presence of noise, we derived several useful bounds and approximations (Appendix C). We usually focus on a theoretical upper bound for y-critical, $y_{\mathrm{cr,max}}$. Note that this upper bound suffices for making rigorous predictions for certain synapses, because $y > y_{\mathrm{cr,max}} \implies y > y$-critical. In the absence of topological transitions, this formula is

$$y_{\mathrm{cr,max}} = W\left[\sqrt{\left(\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}\right) + \frac{\varepsilon^2}{W^2}\left(1 + \frac{e_y^2(e_p^2 - e_y^2)}{(e_y^2 + e_{s*}^2 + e_u^2)^2}\right)} + \frac{\varepsilon}{W}\sqrt{1 + \frac{e_y^2(e_p^2 - e_y^2)}{(e_y^2 + e_{s*}^2 + e_u^2)^2}} \right]. \tag{64}$$

We also computed a lower bound, $y_{\mathrm{cr,min}}$, to assess the tightness of the upper bound. This bound is

$$y_{\mathrm{cr,min}} = W\left[\sqrt{\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}} + \frac{\varepsilon}{W} \right] \tag{65}$$

without topological transitions. Both bounds increase with error and should be considered to be bounded above by W. As expected, both expressions reduce to Eq. (51) as $\varepsilon/W \to 0$. We also note that the two bounds coincide, to leading order in ε/W, if $e_y \ll \max(e_{s*}, e_u)$ and $e_p/\max(e_{s*}, e_u) = \mathcal{O}(1)$, and we argue in Appendix B that this is typical when the network size is large.

The effect of topological transitions is that $y_{\mathrm{cr,max}}$ and $y_{\mathrm{cr,min}}$ become the maximums of several terms, each corresponding to a way that constrained dimensions could behave as semiconstrained within the error bound (Appendix C). We compute each term from generalizations of Eqs. (64) and (65) that account for the amount of error needed to open up semiconstrained dimensions.

C. Testing the theory with simulations

To examine our theory's validity, we assessed its predictions with numerical simulations of feedforward and recurrent networks [Fig. 8(a)]. Each assessment used gradient descent learning to find neural networks whose late time activity approximated some specified orthogonal configuration of input neuron activity and driven neuron activity (Appendix F). We then used our analytically derived certainty condition with noise to identify a subset of synapses that were predicted to not vary in sign across the model ensemble ($W = 1$), and we checked these predictions using the numerical ensemble. We similarly checked predictions from simpler certainty conditions that ignored the nonlinearity or neglected topological transitions in the error surface (Appendix C). Note that we expected gradient descent learning to often fail at finding good solutions in high dimensions, as our theory predicts that each semiconstrained dimension induces local minima in the error surface [Fig. 7(a)]. Since we did not want the theory to bias our numerical verification of it, we focused our simulations on small to moderately sized networks, where we could reasonably sample the initial weight distribution randomly. Future work will consider more realistic neural network applications.

We first considered feedforward network architectures, for which our analytical treatment of noise is exact. To illustrate how nonlinearity and noise affect synapse certainty, we calcu-

lated the magnitude of postsynaptic activity needed to make a particular synapse sign certain [Fig. 8(b)]. We specifically considered 102 random input-output configurations of a small feedforward network with 6 input neurons ($\mathcal{P} = 5$, $\mathcal{C} = 2$), which were tailored to have orthonormal input patterns and generate one topological error surface transition at small errors. In particular, we generated random orthogonal matrices by exponentiating random antisymmetric matrices, we set one element of $\hat{y}$ to a small random value to encourage the topological transition, and we ensured that the other nonzero random element of $\hat{y}$ was large enough to preclude additional transitions (Appendices C and F). For each input-output configuration, we then systematically varied the magnitude of driven neuron activity, y, finding 10^5 synaptic weight matrices with moderate error, $\mathcal{E}^2 \approx \varepsilon^2$, for each magnitude y. Since randomly screening a six-dimensional synaptic weight space is not numerically efficient, we applied gradient descent learning. Nevertheless, the small network size meant that we could comprehensively sample the solution space and numerically probe the distinct predictions made by each bound or approximation used to estimate y-critical.

As expected, the maximum value of y that produced numerical solutions with mixed synapse signs [Fig. 8(b), black dots] was always below the theoretical upper bound for y-critical [Fig. 8(b), black line]. In contrast, mixed-sign numerical ensembles were often found above theoretical y-critical values that neglected topological transitions in the error surface [Fig. 8(b), yellow line] or that neglected the nonlinearity entirely [Fig. 8(b), cyan line]. This means that these simplified calculations for estimating y-critical make erroneous predictions, because the synapse sign is supposed to be exclusively positive or negative whenever y exceeds y-critical, by definition. Therefore, we were able to accurately assess synapse certainty, and this generally required us to include both the nonlinearity and noise-induced topological transitions in the error surface.

We next asked how often we could identify certain synapses in larger networks. For this purpose, we generated 25 random input-output configurations in the feedforward setting (Appendix F), again with orthonormal input patterns, but this time we increased the number of input neurons from 4 to 100 across the configurations [Fig. 8(c)]. As we increased the size of the network, we kept $\mathcal{C}/\mathcal{N}$ fixed at 0.25 and $\mathcal{P}/\mathcal{N}$ fixed at 1 [Fig. 8(c), brown] or 0.5 [Fig. 8(c), purple]. These scaling relationships put our simulations in the setting of high-dimensional statistics [65], where both the number
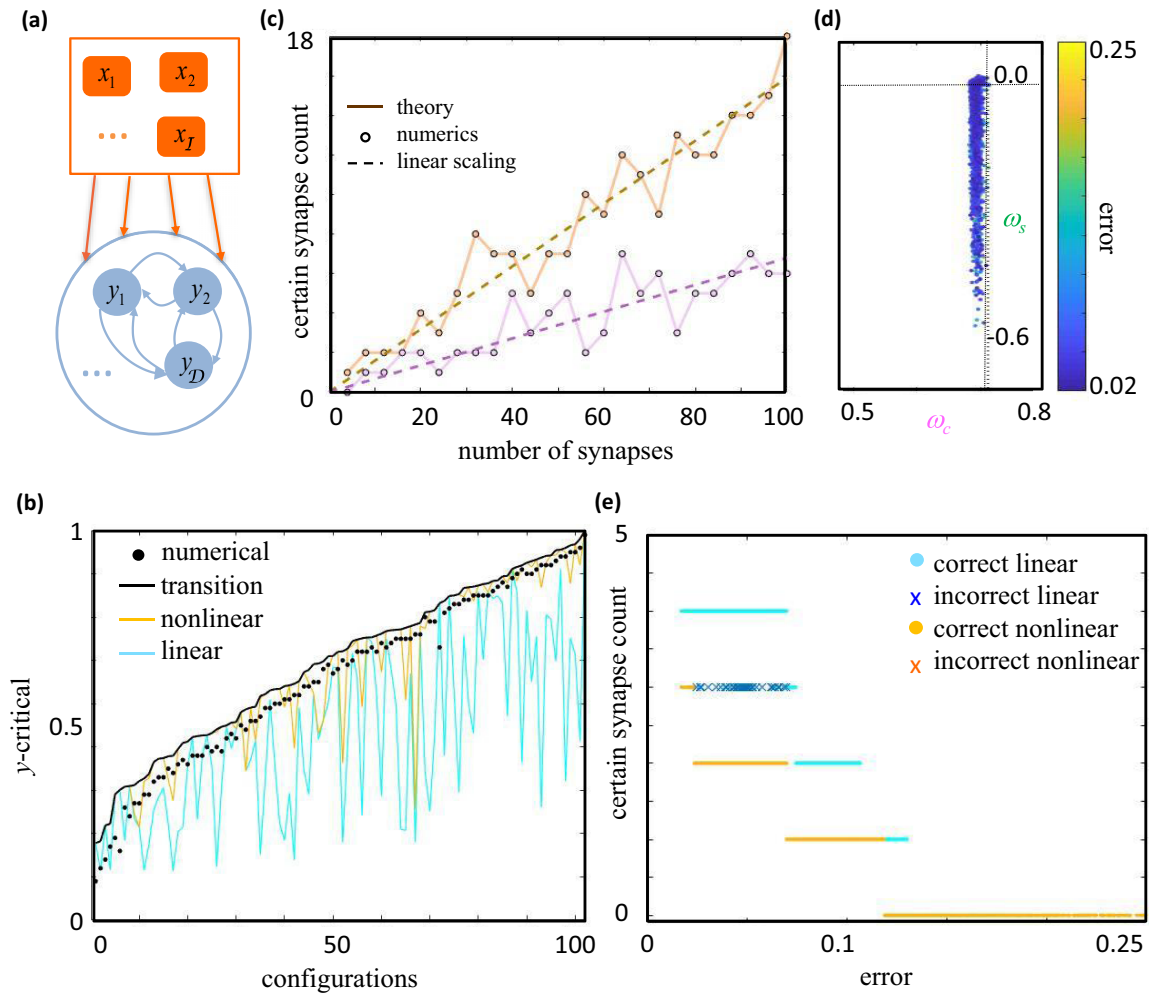
FIG. 8. The theory accounting for error explains numerical ensembles of feedforward and recurrent networks. (a) Cartoon of a recurrent neural network. We disallow recurrent connectivity of neurons onto themselves throughout this figure. $\mathcal{D} = 1$ corresponds to the feedforward case, and $W = 1$ for all panels. (b) Comparison of numerical and theoretical y-critical values for 102 random configurations of input-output activity (Appendix F). We considered a feedforward network with $\mathcal{I} = 6$, $\mathcal{P} = 5$, $\mathcal{C} = 2$. For each configuration and postsynaptic activity level y, we used gradient descent learning to numerically find many solutions to the problem with $\mathcal{E} \approx 0.1$. The black dots correspond to the maximal value of y in our simulations that resulted in an inconsistent sign for the synaptic weight under consideration. The continuous curves show theoretical values for y-critical that upper bound the true y-critical ($y_{\text{cr,max}}$, black), that neglect topological transitions in the error surface (yellow), or that neglect the threshold nonlinearity (cyan). Only the black curve successfully upper bounded the numerical points. Configurations were sorted by the $y_{\text{cr,max}}$ value predicted by the black curve. (c) The number of certain synapses increased with the total number of synapses in feedforward networks. Purple and brown correspond to $\mathcal{N} = 2\mathcal{P} = 4\mathcal{C}$ and $\mathcal{N} = \mathcal{P} = 4\mathcal{C}$, respectively. The solid lines plot the predicted number of certain synapses. The circles represent the number of correctly predicted synapse signs in the simulations. The dashed brown and purple lines are best-fit linear curves with slopes 0.16($\pm$0.01) and 0.07($\pm$0.01) at 95% confidence level, significantly less than the zero error theoretical estimates of 0.28 and 0.18 (Appendix B). (d), (e) Testing the theory in a recurrent neural network with $\mathcal{N} = 10$, $\mathcal{I} = 7$, $\mathcal{D} = 4$, $\mathcal{P} = 8$, and $\mathcal{C} = 3$. Each dot shows a model found with gradient descent learning. (d) x and y axes show two η coordinates predicted to be constrained and semiconstrained, respectively, and the color axis shows the model's root-mean-square error over neurons, $\mathcal{E}/\sqrt{\mathcal{D}}$. Although our theory for error surfaces is approximate for recurrent networks, the solution space was well explained by the constrained and semiconstrained dimensions. Note that the numerical solutions tend to have constrained coordinates smaller than the theoretical value (vertical line) because the learning procedure is initialized with small weights and stops at nonzero error. (e) The x axis shows the model's error, and the y axis shows the number of synapse signs correctly predicted by the nonlinear theory (yellow dots or red crosses) or linear theory (cyan dots or blue crosses). Dots denote models for which every model prediction was accurate, and crosses denote models for which some predictions failed.

of parameters and the number of constraints increase with the size of the network. In this high-dimensional regime, a simple heuristic argument suggests that the number of zero-error certain synapses should scale linearly with the number of synapses (Appendix B), because y_{cr} and the typical magnitude of y scale equivalently with $\mathcal{N}$. Here we tested this prediction by setting $\hat{y}$ randomly, setting $y = 1 - \ln 2/\mathcal{C}$ to approximate the median norm of vectors in the unit $\mathcal{C}$-ball (Appendix B), and numerically finding a small error solution for each configuration ($\mathcal{E}^2/\mathcal{P} \approx 10^{-6}$).

As expected, we empirically found that the number of certain synapses predicted by the theory [Fig. 8(c), solid lines] scaled with the network size linearly [Fig. 8(c), dashed lines]. The jaggedness of the solid curves reflect the fact that each point is specific to the random input-output configuration constructed for that value of $\mathcal{N}$. The purple curve corresponds to the case when $\mathcal{N} = 2\mathcal{P} = 4\mathcal{C}$ and the brown curve when $\mathcal{N} = \mathcal{P} = 4\mathcal{C}$. Furthermore, for every certain synapse predicted, we verified that its predicted sign was realized in the numerical solution we found [Fig. 8(c), circles]. These results suggest that the theory will predict many synapses to be certain in realistically large neural systems.

Finally, we empirically tested our theory for a recurrent network [Figs. 8(d) and 8(e)], where our treatment of noise is only approximate. For this purpose, we considered networks without the self-coupling terms, $\mathcal{N} = \mathcal{I} + \mathcal{D} - 1$. We constructed a single random configuration with nonnegative driven neuron responses and orthogonal presynaptic patterns for one of the driven neurons[10] (Appendix F). This driven neuron could thus serve as the target neuron for our analyses. Note that it is sometimes possible to orthogonalize the input patterns for more than one driven neuron, but this is irrelevant to our analysis and is not pursued here. We then used gradient descent learning to find around 4500 networks that approximated the desired fixed points with variable accuracy. For technical simplicity, we first found connectivity matrices using a proxy cost function that treated the network as if it were feedforward. We then simulated the neural network dynamics with these weights and correctly evaluated the model's error as prescribed by Eq. (63).

This network ensemble revealed that constrained and semi-constrained dimensions accurately explained the structure of the solution space for recurrent networks with nonzero error. Figure 8(d) shows the projection of the corresponding solution space along two η directions, one predicted to be constrained by the feedforward theory and the other predicted to be semiconstrained. As predicted, the extension of the solution space along the negative semiconstrained direction was clearly discernible. However, recurrence implies that the exact solution space is not perfectly cylindrical around the semiconstrained axes (Appendix C), because the driven neuron inputs to the target neuron can themselves vary due to noise. Here this effect was empirically insignificant, and the geometric structure of the solution space conformed rather well to our feedforward prediction. One might have expected the error [color in Fig. 8(d)] to increase monotonically as one moves away from the center of semiconstrained cylinder, but this expectation is incorrect for two reasons. First, we are visualizing the error surface as a projection along two dimensions, yet variations in other η coordinates add variation to the error.[11] Second, we are visualizing the solution space for one target neuron, but other driven neurons in the recurrent network contribute to the summed error represented by the color.

Moreover, the theory correctly predicted how the number of certain synapses would decrease as a function of ε [Fig. 8(e)], and we never found a numerical violation of the theoretical certainty condition that included nonlinearity and noise. In Fig. 8(e), the yellow circles represent the number of certain synapses that were predicted by the theory and verified to have synapse signs that agreed with the theoretical prediction. Here accurate predictions did not require us to account for topological error surface transitions. In contrast, although our simulations usually agreed with the predictions of the linear theory [Fig. 8(e), cyan circles], they could also disagree. In Fig. 8(e), the blue crosses indicate configurations where the linear theory incorrectly predicted some synapse signs. The absence of red crosses reiterates the consistency of predictions coming from the nonlinear treatment.

VII. DISCUSSION

In summary, we enumerated all threshold-linear recurrent neural networks that generate specified sets of fixed points, under the assumption that the number of candidate synapses onto a neuron is at least the specified number of fixed points. We found that the geometry of the solution space was elegantly simple, and we described a coordinate transformation that permits easy classification of weight-space dimensions into constrained, semiconstrained, and unconstrained varieties. This geometric approach also generalized to approximate error-surfaces of model parameters that imprecisely generate the fixed points. We used this geometric description of the error surface to analyze structure-function links in neural networks. In particular, we found that it is often possible to identify synapses that must be present for the network to perform its task, and we verified the theory with simulations of feedforward and recurrent neural networks.

Rectified-linear units are also popular in state of the art machine learning models [29,66–68], so the fundamental insights we provide into the effects of neuronal thresholds on neural network error landscapes may have practical significance. For example, machine learning often works by taking a model that initially has high error and gradually improving it by modifying its parameters in the gradient direction [69]. However, error surfaces with high error can have semiconstrained dimensions that abruptly vanish at lower errors (Fig. 7). Local parameter changes typically cannot move the model through these topological transitions, because models that wander deeply into semiconstrained dimensions are far from where they must be to move down the error surface. The model has continua of local and global minima, and the network needs to be initialized correctly to reach its lowest possible errors. This could provide insight into deep learning theories that view its success as a consequence of weight subspaces that happen to be initialized well [70,71].

The geometric simplicity of the zero-error solution space provides several insights into neural network computation. Every time a neuron has a vanishing response, half of a dimension remains part of the solution space, which the network could explore to perform other tasks. In other words, by replacing an equality constraint with an inequality constraint, simple thresholding nonlinearities effectively increase the computational capacity of the network [72,73]. The flexibility

---

[10]We performed this numerical experiment with several random configurations to confirm that the results did not qualitatively depend on the random sample.

[11]For example, imagine projecting the 3D surfaces in Fig. 7(a) along two dimensions.

afforded by vanishing neuronal responses thereby provides an intuitive way to understand the impressive computational power of sparse neural representations [50,74–76]. Furthermore, the brain could potentially use this flexibility to set some synaptic strengths to zero, thereby improving wiring efficiency. This would link sparse connectivity to sparse response patterns, both of which are observed ubiquitously in neural systems.

Our theory could be extended in several important ways. First, we only derived the certainty condition to identify critical synapses from orthonormal sets of fixed points. Although our orthogonal analysis also provides a conservative bound for a general set of fixed points (Appendix A), a more precise analysis will be needed to pinpoint synapses in realistic biological settings where stimulus-induced activity patterns may be strongly correlated. Since our error surface description made no orthonormality assumptions, this analysis will only require more complicated geometrical calculations to discern whether the synapse sign is consistent across the space of low-error models. Furthermore, we could use the error surfaces to identify multisynapse anatomical motifs that are required for function, or to estimate the fraction of models in which an uncertain synapse is excitatory versus inhibitory. It would also be interesting to relax the assumption that the number of fixed points is small. This would allow us to consider scenarios where the fixed points can only be generated nonlinearly. We could also consider cases where no exact solution exists at all. Here we assumed that we knew the activity level of every neuron in the circuit. This is not always the case, and it will be important to determine how unobserved neurons alter the error landscape for synaptic weights connecting the observed neurons. The error landscape geometry will also be affected by recurrent network effects that we ignored here (Appendix C). It will be interesting to see whether the geometric toolbox of theoretical physics can provide insights into the nontrivial effects of unobserved neurons and recurrent network dynamics. Finally, we note that it will sometimes be important to analyze networks with alternate nonlinear transfer functions. Our analyses already apply exactly to recurrent networks with arbitrary threshold-monotonic nonlinear transfer functions (Appendix D). Moreover, our analyses can approximate any nonlinearity by treating its departures from threshold-linearity as noise (Appendix D). An extension to capped rectified linear units [67], which saturate above a second threshold, would also be straightforward. In particular, semiconstrained dimensions would emerge from any condition where the target neuron is inactive or saturated.

Our primary motivation for undertaking this study was to find rigorous theoretical methods for predicting neural circuit structure from its functional responses. This identification can be used to corroborate or broaden circuit models that posit specific connectivity patterns, such as center-surround excitation-inhibition in ring attractors [16–18] or contralateral relay neuron connectivity in zebrafish binocular vision [12,77]. More generally, if an experimental test violates the certainty conditions we derived using our ensemble modeling approach, it will suggest that some aspect of model mismatch is important. We could then move on to the development of qualitatively improved models that might modify neuronal nonlinearities, relax weight bounds, incorporate subcellular processes or neuromodulation, or hypothesize hidden cell

populations. However, we hope that our focus on predictions that follow with certainty from simple network assumptions will enable predictions that are relatively insensitive to minor mismatches between our abstract model and the real biological brain. More nuanced predictions may require more nuanced models.

An important parameter of the theory is the weight bound. In particular, W bounds the magnitude of synaptic weight vectors in biological networks, and our certainty condition declares a synapse to be necessary when the ratio y/W exceeds a critical value. It is not *a priori* clear how to set this scale parameter without additional biological data. Nevertheless, one could use the neuronal activity data to compute each synapse's W-critical value, below which the certainty condition is satisfied, and rank-order the synapses according to decreasing W-critical values. Until we know the value of W, we do not know where to draw the line between certain synapses and uncertain synapses. However, our theory predicts that all of the certain synapses will be at the top of the list, which specifies a sequence of experimentally testable predictions and may already provide biological insights into the important synaptic connections. Testing these predictions can help constrain the theory's biological bound parameter.

Our theory describes function at the level of neural representations. This description is useful because many systems neuroscience experiments measure representations directly, and it is important to build mechanistic models that explain these data in terms of neural network interactions [12,15,18,77]. However, it would also be interesting to link structure to function at the higher levels of behavior and cognition. This is a significantly different problem because multiple representations can support the same high-level functions, and both neural network structure and representation can change over time [78–85]. Consequently, experimental tests of our current framework must measure network structure and representation on timescales shorter than the network's representational dynamics, and certain synapses may be most biologically meaningful in innate circuits with limited plasticity. Extensions to our framework may also be useful for relating structural and representational dynamics in circuits for learning [86].

An exciting prospect is to explore how our ensemble modeling framework can be combined with other theoretical principles and biological constraints to obtain more refined structure-function links. For instance, we could refine our ensemble by restricting to stable fixed points. Alternatively, once the sign of a given synapse is identified, Dale's principle might allow us to fix the signs of all other synapses from this neuron [87]. This would restrict the solution space and could make other synapses certain. Utilizing limited connectomic data to impose similar restrictions might also be a fruitful way to benefit from large-scale anatomical efforts [7,10,13,14]. Finally, rather than restricting the magnitude of the incoming synaptic weight vector, we could consider alternate biologically relevant constraints, such as limiting the number of synapses, minimizing the total wiring length, or positing that the network operates at capacity [88,89]. These changes would modify the certainty conditions in our framework, as well as our experimental predictions. We could therefore assess candidate optimization principles and biological priors

experimentally. While the base framework developed here was designed to identify crucial network connections required for function, we hope that our approach will eventually allow us to assess theoretical principles that determine how neural network structure follows from function.

APPENDIX A: A CERTAINTY CONDITION TO PINPOINT SYNAPSES REQUIRED FOR SPECIFIED RESPONSE PATTERNS

1. Preliminaries

For completeness, we begin by briefly reviewing a few central concepts from the main manuscript.

a. From recurrent to feedforward networks

Let us consider a neural network of $\mathcal{I}$ input neurons that send signals to an interconnected population of $\mathcal{D}$ driven neurons governed by dynamical equations (6), as described in the main manuscript. At steady-state, since all time-derivatives are zero, Eq. (6) yields

$$
\begin{aligned}
y_{\mu i} &= \Phi\left(\sum_{m=1}^{\mathcal{D}} w_{im} y_{\mu m} + \sum_{m=\mathcal{D}+1}^{\mathcal{D}+\mathcal{I}} w_{im} x_{\mu, m-\mathcal{D}}\right) \\
&= \Phi\left(\sum_{m=1}^{\mathcal{N}} w_{im} z_{\mu m}\right),
\end{aligned}
\tag{A1}
$$

where, as prescribed in the main manuscript, $y_{\mu i}$ and $x_{\mu m}$ denote steady-state activity levels of the driven and input neurons to the μth stimulus, which we have combined into $z_{\mu m}$, and $\mathcal{N}$ is the number of incoming synapses onto each of the driven neurons. Equation (A1) provides $\mathcal{D} \times \mathcal{P}$ nonlinear equations for $\mathcal{D} \times \mathcal{N}$ unknown parameters. However, we immediately notice that the steady-state activity of neuron i depends only on the ith row of the connectivity matrix, so these equations separate into $\mathcal{D}$ independent sets of $\mathcal{P}$ equations with $\mathcal{N}$ unknowns, the weights onto a given driven neuron. In other words, the recurrent network involving $\mathcal{D}$ driven and $\mathcal{I}$ input neurons decomposes into $\mathcal{D}$ feedforward networks with $\mathcal{N} = \mathcal{D} + \mathcal{I}$ feedforward inputs. The steady-state equations for these feedforward networks are given by

$$
y_\mu = \Phi\left(\sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m\right),
\tag{A2}
$$

where we have now suppressed the i index in $y_{\mu i}$ and in w_{im}. For this feedforward network we will refer the ith neuron as the target neuron, and it is as if that all the neurons (driven and input) are providing feedforward inputs to it. As long as we only consider exact solutions to the fixed point equations, the problem of identifying synaptic connectivity in a recurrent network reduces to solving the problem for feedforward networks. Thus, in the rest of this Appendix we will focus on identifying w_m's satisfying Eq. (A2).

Note that the main text used the notation $z_{\mu m}^{(i)}$ to emphasize that the set of presynaptic neurons may depend on the target neuron, but we simply write $z_{\mu m}$ throughout the Appendices with the understanding that the formalism applies to a specified target neuron whose index is suppressed. Furthermore, for conceptual simplicity the main text first stated many results in a feedforward setting with a single driven neuron, but the Appendices immediately treat the general case where presynaptic partners may come from either the input or driven populations of neurons.

b. A convenient set of variables

In all our discussions in this section the input neuronal response matrix, $z_{\mu m}$, will be assumed to be fixed. Note that $z_{\mu m}$ connects synaptic weight vectors to the target response vector and can be used to define $\mathcal{P}$ weight combinations, the η coordinates. Each η coordinate controls the target response to a single stimulus condition:

$$
y_\mu = \Phi(\eta_\mu), \quad \text{where} \quad \eta_\mu \equiv \sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m.
\tag{A3}
$$

It is rather convenient to extend this set of $\mathcal{P}$ η coordinates to a basis set of $\mathcal{N}$ η coordinates, such that all synaptic weights can be uniquely expressed as a linear combination of these η coordinates, and vice versa. To see how this can be done, we will henceforth make the simplifying assumption that the $\mathcal{P} \times \mathcal{N}$ matrix has the maximal rank, $\mathcal{P}$, although we anticipate that much of our framework, results, and insights will apply more generally. If z has maximal rank, then its kernel will be an $(\mathcal{N} - \mathcal{P})$-dimensional linear subspace spanned by $(\mathcal{N} - \mathcal{P})$ orthogonal basis vectors, denoted by $\vec{\varepsilon}_\mu$ for $\mu = \mathcal{P} + 1 \ldots \mathcal{N}$. We can now extend z to an $\mathcal{N} \times \mathcal{N}$ matrix, Z, as follows:

$$
\begin{aligned}
Z_{\mu m} &= z_{\mu m} \quad \text{for } \mu = 1 \ldots \mathcal{P}, \text{ and } \forall\, m, \\
Z_{\mu m} &= \varepsilon_{\mu m} \quad \text{for } \mu = \mathcal{P} + 1 \ldots \mathcal{N}, \text{ and } \forall\, m,
\end{aligned}
\tag{A4}
$$

where $\varepsilon_{\mu m}$ is the mth component of the null vector $\vec{\varepsilon}_\mu$. With this construction, it is easy to see that the new η coordinates,

$$
\eta_\mu \equiv \sum_{m=1}^{\mathcal{N}} Z_{\mu m} w_m, \quad \text{for } \mu = \mathcal{P} + 1, \ldots, \mathcal{N},
\tag{A5}
$$

remain completely unconstrained by the specified response patterns, as these linear combinations do not contribute to any of the target responses. In contrast, the original η coordinates,

$$
\eta_\mu = \sum_{m=1}^{\mathcal{N}} Z_{\mu m} w_m = \sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m, \quad \text{for } \mu = 1, \ldots, \mathcal{P},
\tag{A6}
$$

are all constrained by the data:

$$
\eta_\mu
\begin{cases}
= y_\mu & \text{for } \mu = 1, \ldots, \mathcal{C}, \text{ the constrained dimensions,} \\
\leqslant 0 & \text{for } \mu = \mathcal{C} + 1, \ldots, \mathcal{P}, \text{ the semiconstrained dimensions,}
\end{cases}
\tag{A7}
$$

where for notational simplicity we have ordered the response patterns such that $y_\mu \neq 0$ only for $\mu = 1, \ldots, \mathcal{C}$. Also, we extend the y_μ's to an $\mathcal{N}$-dimensional vector, $\vec{y}$, by assigning $y_\mu = 0$ for $\mu = \mathcal{P} + 1 \ldots \mathcal{N}$.

The extended response matrix Z defines a basis transformation connecting physical synaptic directions, $\widehat{e}_m$, with directions

$$\vec{\varepsilon}_\mu \equiv \sum_{m=1}^{\mathcal{N}} \widehat{e}_m Z_{m\mu}^{-1}, \tag{A8}$$

along which the η coordinates change. These $\vec{\varepsilon}$ vectors clearly differentiate directions in the weight space that are activity-constrained by neuronal responses ($\mu = 1, \ldots, \mathcal{P}$) from those that are not ($\mu = \mathcal{P} + 1, \ldots, \mathcal{N}$). We can express any weight vector in either the $\{\widehat{e}_m\}$ basis or the $\{\vec{\varepsilon}_\mu\}$ basis:

$$\vec{w} = \sum_{m=1}^{\mathcal{N}} w_m \widehat{e}_m = \sum_{\mu=1}^{\mathcal{N}} \eta_\mu \vec{\varepsilon}_\mu,$$

where $w_m = \sum_{\mu=1}^{\mathcal{N}} Z_{m\mu}^{-1} \eta_\mu, \quad \eta_\mu = \sum_{m=1}^{\mathcal{N}} Z_{\mu m} w_m,$

$$\vec{\varepsilon}_\mu \equiv \sum_{m=1}^{\mathcal{N}} \widehat{e}_m Z_{m\mu}^{-1}, \quad \widehat{e}_m = \sum_{\mu=1}^{\mathcal{N}} \vec{\varepsilon}_\mu Z_{\mu m}. \tag{A9}$$

For later convenience we also define the number of semi-constrained and unconstrained dimensions as, $\mathcal{S} = \mathcal{P} - \mathcal{C}$ and $\mathcal{U} = \mathcal{N} - \mathcal{P}$, respectively.

2. Derivation of the certainty condition for orthogonal input patterns

Our goal here is to use the solution space (i.e., ensemble of weights that are precisely able to recover the specified target responses) to derive a condition for when we can be certain that a given synapse must be nonzero. For technical simplicity, we will specialize to the case when all the response patterns are orthonormal, i.e.,

$$\sum_{m=1}^{\mathcal{N}} z_{\mu m} z_{\nu m} = \delta_{\mu\nu} \Leftrightarrow zz^T = I, \tag{A10}$$

where I is the identity matrix. Then we can always choose the extended Z matrix to be an $\mathcal{N} \times \mathcal{N}$ orthogonal matrix, such that $Z^{-1} = Z^T$ and the $\vec{\varepsilon}_\mu$ vectors now form an orthonormal basis. Motivated by biological constraints, we will impose a bound on the magnitude of the synaptic weight vector. For orthonormal response patterns, this translates into a spherical bound on η coordinates as well [see Fig. 5(b)]:

$$|\vec{w}|^2 = \sum_{m=1}^{\mathcal{N}} w_m^2 = \sum_{\mu=1}^{\mathcal{N}} \eta_\mu^2 \leqslant W^2. \tag{A11}$$

We refer to this $\mathcal{N}$-dimensional ball, in which all admissable synaptic weights reside, as the weight space.

a. A heuristic argument for y-critical

Before diving into the rigorous and technical derivation, in this subsection we first try to intuitively understand how the certainty condition (51) can arise. For this purpose, let us

start with a linear theory with no unconstrained dimension, so $\mathcal{S} = \mathcal{U} = 0$. In this case, there is a unique set of weights that can precisely reproduce the observed responses:

$$w_m = \sum_{\mu=1}^{\mathcal{N}} Z_{m\mu}^{-1} y_\mu = \sum_{\mu=1}^{\mathcal{N}} Z_{\mu m} y_\mu. \tag{A12}$$

Since $Z_{\mu m} = z_{\mu m}$ represents the responses of the mth presynaptic neuron, the solution for the mth synaptic weight (A12) is simply the correlation between the pre and post synaptic activity. In a linear theory, the sign of the synapse is thus dictated by the sign of the correlation between the pre and post synaptic neuron.

Let us now allow a single ($\mathcal{N}$th) unconstrained direction. One can think of this situation as if we do not have the information on how the target neuron would respond to the unconstrained stimulus pattern. If we knew that this response was say, y_u, then we would have been able to determine the sign of w_m:

$$\text{Sgn}(w_m) = \text{Sgn}\left(\sum_{\mu=1}^{\mathcal{N}-1} Z_{\mu m} y_\mu + Z_{\mathcal{N}m} y_u\right). \tag{A13}$$

However, since we do not know what the last term is, if it can cancel the first term for some allowed value of y_u then the overall sign becomes ambiguous. Conversely, w_m becomes certain if

$$\left|\sum_{\mu=1}^{\mathcal{N}-1} Z_{\mu m} y_\mu\right| > |Z_{\mathcal{N}m} y_u| \; \forall \; y_u. \tag{A14}$$

Now, it is easy to recognize that the first term is just $\widehat{e} \cdot \vec{y} = y e_y$, where we have suppressed the m index on $\widehat{e}_m$ here to reduce notational clutter and will continue to do so while referring to the synapse direction whose sign we are considering.[12] $e_y = \widehat{e} \cdot \widehat{y}$ refers to the projection of $\widehat{e}$ along $\widehat{y}$. Also, note that in this simple case with one unconstrained direction, the projection of $\widehat{e}$ along the unconstrained subspace is just given by $e_u = \widehat{e} \cdot \vec{\varepsilon}_\mathcal{N} = Z_{\mathcal{N}m}$. Further, since Z is orthogonal, to have any solution at all

$$y^2 + y_u^2 \leqslant W^2. \tag{A15}$$

Substituting the maximum $|y_u|$ from Eq. (A15) into Eq. (A14), after some algebra we get the condition for sign certainty as

$$y > y_{\text{cr}} = W \sqrt{\frac{e_u^2}{e_u^2 + e_y^2}}. \tag{A16}$$

The same argument applies if the $\mathcal{N}$th direction is semiconstrained instead of unconstrained, with one notable difference. If the $\mathcal{N}$th pattern was semiconstrained that means $y_\mathcal{N} = 0$, and the nonlinear thresholding is masking how the target neuron would have responded in a linear model.[13] However, the ambiguity in sign can only arise if the second term has a

---

[12]We do want to point out that in the main manuscript since we were introducing the various concepts and relevant quantities, for clarity we did explicitly keep track of the m index.

[13]Note that our relation between the sign of the synapse and the sign of the correlation is based on a linear response.

sign opposite to the first term, or a sign opposite to $\text{Sgn}(e_y)$. Moreover, for the thresholding to act, the target response for the semiconstrained pattern must be negative in the linear theory, so $Z_{\mathcal{N}m}$ has to have the same sign as e_y to generate the ambiguity. And, if it is indeed so, then we obtain a certainty condition that is identical to Eq. (A16) except that $e_u \rightarrow e_s$, the projection of $\widehat{e}$ along the semiconstrained direction:

$$y > y_{\text{cr}} = W\sqrt{\frac{e_s^2}{e_s^2 + e_y^2}}. \qquad \text{(A17)}$$

If $Z_{\mathcal{N}m}$ and e_y have opposite signs, then the synapse always has the same sign throughout the solution space.

While this derivation of y_{cr} is heuristic and only deals with a single semiconstrained or unconstrained dimension, it provides intuition for the general result (51). Essentially, whether the sign of a given synapse is constant across the solution space depends on two competing quantities: the correlation between the pre- and postsynaptic responses; and the strength of the postsynaptic drive for patterns where the target response is either unknown or masked by the thresholding nonlinearity.

b. Hyperplane dividing excitatory and inhibitory synaptic regions

Having gained some intuition about the certainty condition, let us now proceed to a rigorous derivation of the result. Since the constrained coordinates are fixed for the weight vectors that belong to the solution space [the deep yellow wedge in Fig. 5(b)], we must have

$$y^2 \equiv \sum_{\mu=1}^{\mathcal{C}} y_\mu^2 = \sum_{\mu=1}^{\mathcal{C}} \eta_\mu^2, \qquad \text{(A18)}$$

so that the solution space resides within an $(\mathcal{U} + \mathcal{S})$-dimensional ball with radius

$$\overline{W} \equiv \sqrt{W^2 - y^2}, \qquad \text{(A19)}$$

as depicted in Fig. 5(b), by the yellow region. We refer to this semiconstrained plus unconstrained subspace as the flexible subspace.

Now, the synaptic direction of interest, $\widehat{e}$, can be decomposed into its projections along constrained, semiconstrained and unconstrained subspaces. For notational simplicity, let us denote $e_\mu \equiv (\widehat{e} \cdot \vec{\varepsilon}_\mu) = Z_{\mu m}$ as the component of $\widehat{e}$ along $\vec{\varepsilon}_\mu$. Note that the second equality follows from the orthogonality assumption and Eq. (A9). In general, in this manuscript we will use subscripts on e to denote projections of $\widehat{e}$ along different directions or subspaces. We can now write[14]

$$\widehat{e} = \sum_{\mu=1}^{\mathcal{N}} e_\mu \vec{\varepsilon}_\mu = \sum_{\mu=1}^{\mathcal{C}} e_\mu \vec{\varepsilon}_\mu + \sum_{\mu=\mathcal{C}+1}^{\mathcal{P}} e_\mu \vec{\varepsilon}_\mu + \sum_{\mu=\mathcal{P}+1}^{\mathcal{N}} e_\mu \vec{\varepsilon}_\mu = \cos\theta\widehat{c} + \sin\theta\cos\phi\widehat{s} + \sin\theta\sin\phi\widehat{u}, \qquad \text{(A20)}$$

where

$$\widehat{c} \equiv \frac{\sum_{\mu=1}^{\mathcal{C}} e_\mu \vec{\varepsilon}_\mu}{\sqrt{\sum_{\mu=1}^{\mathcal{C}} e_\mu^2}}, \quad \widehat{s} \equiv \frac{\sum_{\mu=\mathcal{C}+1}^{\mathcal{P}} e_\mu \vec{\varepsilon}_\mu}{\sqrt{\sum_{\mu=\mathcal{C}+1}^{\mathcal{P}} e_\mu^2}}, \quad \widehat{u} \equiv \frac{\sum_{\mu=\mathcal{P}+1}^{\mathcal{N}} e_\mu \vec{\varepsilon}_\mu}{\sqrt{\sum_{\mu=\mathcal{P}+1}^{\mathcal{N}} e_\mu^2}}, \qquad \text{(A21)}$$

are unit vectors that lie within the constrained, semiconstrained and unconstrained subspaces, and

$$e_c \equiv \widehat{c} \cdot \widehat{e} = \cos\theta = \sqrt{\sum_{\mu=1}^{\mathcal{C}} e_\mu^2} \geq 0, \quad e_s \equiv \widehat{s} \cdot \widehat{e} = \sin\theta\cos\phi = \sqrt{\sum_{\mu=\mathcal{C}+1}^{\mathcal{P}} e_\mu^2} \geq 0,$$

$$e_u \equiv \widehat{u} \cdot \widehat{e} = \sin\theta\sin\phi = \sqrt{\sum_{\mu=\mathcal{P}+1}^{\mathcal{N}} e_\mu^2} \geq 0 \qquad \text{(A22)}$$

are the projections of $\widehat{e}$ along these directions. One could think of θ, ϕ as representing a spherical coordinate system where the role of x, y, and z axes are played by $\widehat{s}, \widehat{u}$ and $\widehat{c}$ respectively, and our definitions (A20)–(A22) imply the convention, $0 \leq \{\theta, \phi\} < \pi/2$. For later convenience, let us also introduce the projection of $\widehat{e}$ onto the activity-constrained subspace:

$$e_p \equiv \widehat{p} \cdot \widehat{e} = \sqrt{\cos^2\theta + \sin^2\theta\cos^2\phi} = \sqrt{\sum_{\mu=1}^{\mathcal{P}} e_\mu^2} \geq 0, \quad \text{where} \quad \widehat{p} \equiv \frac{\sum_{\mu=1}^{\mathcal{P}} e_\mu \vec{\varepsilon}_\mu}{\sqrt{\sum_{\mu=1}^{\mathcal{P}} e_\mu^2}}. \qquad \text{(A23)}$$

We would also like to emphasize that we can compute θ, ϕ just from the knowledge of the neuronal responses, $z_{\mu m} = e_\mu$, which is particularly useful for numerical calculations:

$$\theta \equiv \cos^{-1}\left(\sqrt{\sum_{\mu=1}^{\mathcal{C}} z_{\mu m}^2}\right), \quad \text{and} \quad \phi \equiv \cos^{-1}\left(\sqrt{\sum_{\mu=\mathcal{C}+1}^{\mathcal{P}} z_{\mu m}^2} \Big/ \sqrt{1 - \sum_{\mu=1}^{\mathcal{C}} z_{\mu m}^2}\right). \qquad \text{(A24)}$$

---

[14]Again, we remind the readers that in the main manuscript these projected vectors were denoted by $\widehat{c}_m, \widehat{s}_m$ and $\widehat{u}_m$.
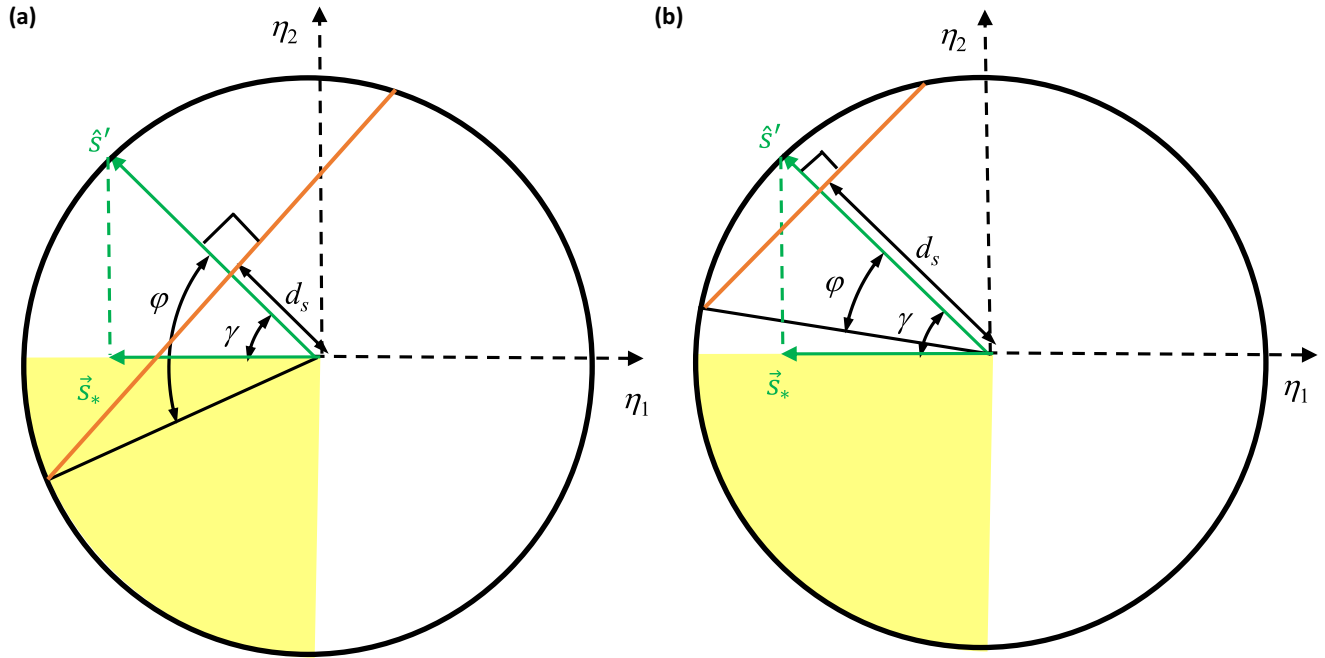
FIG. 9. Cartoons depicting the orientation of the semiconstrained projection of a given synaptic weight direction ($\widehat{s}'$) within the semiconstrained subspace and its impact on determining the sign of the given weight. In these plots, the yellow wedges represent the solution space, $\eta_1, \eta_2 \leqslant 0$. d_s is the distance of the $w_m = 0$ orange line (hyperplane in higher dimension) from the origin. If d_s is small, as in the left plot (a), then the projection angle γ is smaller than φ, half of the angle subtended by the orange line to the origin, and therefore the orange line and the yellow cone intersect. This means that solutions with both positive and negative w's are present. In the right plot (b), d_s is sufficiently large such that $\gamma > \varphi$ and consequently, all the solutions must have consistent sign.

Now, any weight vector in the solution space can be written as

$$\vec{w} = \vec{y} + \vec{w}_s + \vec{w}_u, \tag{A25}$$

where $\vec{w}_s$ and $\vec{w}_u$ are the projections of $\vec{w}$ onto the semiconstrained and unconstrained subspaces, and the constrained part of $\vec{w}$ is fixed at $\vec{y}$. Using Eqs. (A20) and (A25), one then finds that the $w_m = 0$ hyperplane dividing the excitatory and inhibitory regions in the flexible subspace satisfies the equation

$$
\begin{aligned}
w &= \vec{w} \cdot \widehat{e} \\
&= y \cos\theta \cos\alpha + \sin\theta \cos\phi \, \widehat{s} \cdot \vec{w}_s \\
&\quad + \sin\theta \sin\phi \, \widehat{u} \cdot \vec{w}_u = 0,
\end{aligned}
\tag{A26}
$$

where we have now also suppressed the index m in w_m. Also, we have defined $\alpha \in [0, \pi]$ to be the angle between $\vec{y}$ and $\widehat{c}$. We now notice that the origin of the flexible subspace, $\vec{w}_s = \vec{w}_u = 0$, is in the solution space and the sign of w for this solution point is given by

$$\mathrm{Sgn}(w) = \mathrm{Sgn}(\cos\alpha) = \mathrm{Sgn}(e_y) = \mathrm{Sgn}\left(\sum_{\mu=1}^{\mathcal{C}} z_{\mu m} y_\mu\right). \tag{A27}$$

In other words, if the sign of the synapse is certain, this certain sign must be $\mathrm{Sgn}(\cos\alpha)$, which corresponds to the sign of the correlation between the target neuron and the presynaptic neuron. Intuitively, positive correlations point to an excitatory connection, and negative correlations point to an inhibitory connection.

c. Special case without unconstrained dimensions

To derive the certainty condition, let us start by looking at the case when $\mathcal{P} = \mathcal{N}$, so that there are no unconstrained directions, or equivalently, $\phi = 0$. In this case, the solution space is just the all-negative orthant in the $\mathcal{S}$-dimensional semiconstrained hypersphere (Fig. 9), and the equation for the $w = 0$ hyperplane can be written as

$$\sin\theta \, (\widehat{s}' \cdot \vec{w}_s) = y \cos\theta |\cos\alpha|, \tag{A28}$$

where the right-hand side (RHS) is positive, and we have introduced

$$\widehat{s}' \equiv -\mathrm{Sgn}(\cos\alpha)\widehat{s}, \tag{A29}$$

which flips the direction of $\widehat{s}$ if $\cos\alpha > 0$, or equivalently, if $e_y > 0$. Now, if the $w = 0$ hyperplane (orange lines in Fig. 9) is far enough along $\widehat{s}'$ from the origin that it does not intersect with the all-negative orthant within the weight bounds, then we can be certain that w is nonzero and always has a consistent sign. To check this, we need to compare the cone angle that the orange hyperplane subtends at the center, φ, with the minimum angle, γ, that the $\widehat{s}'$ vector makes with the all-negative orthant.

First, φ can easily be inferred from trigonometry:

$$\cos\varphi = \frac{d_s}{W} = \frac{y \cot\theta |\cos\alpha|}{\sqrt{W^2 - y^2}}, \tag{A30}$$

where $d_s = y \cot \theta |\cos \alpha|$ represents the distance from the center of the semiconstrained sphere to the hyperplane. The expression for d_s follows from the general mathematical result that if

$$\vec{\beta} \cdot \vec{x} + \beta_0 = 0 \qquad (A31)$$

is an equation for a hyperplane, where $\vec{x}$ denotes the coordinate vector and $\vec{\beta}$, β_0 are constants, then the perpendicular distance, $d_\perp$, to it from a point $\vec{x} = \vec{\zeta}$ is given by

$$d_\perp = \frac{|\vec{\beta} \cdot \vec{\zeta} + \beta_0|}{|\vec{\beta}|}. \qquad (A32)$$

Note, we are interested in the distance from the origin, $\vec{\zeta} = 0$, to the hyperplane satisfying Eq. (A28), so $\vec{\beta} = \sin \theta \, \widehat{s}' \Rightarrow |\vec{\beta}| = \sin \theta$ and $\beta_0 = -y \cos \theta |\cos \alpha|$.

To provide a geometric intuition for γ, let us first assume that $\widehat{s}'$ does not point into the all-negative orthant. If we can find the projection of $\widehat{s}'$ on the correct boundary of the solution space, then γ will be given by the angle between $\widehat{s}'$ and the appropriate semiconstrained boundary vector, $\vec{s}_*$ (Fig. 9). Since all the components in the solution space (all-negative orthant) have to be negative or zero, to find the appropriate projection vector of $\widehat{s}'$ onto the boundary of solution space, we essentially have to set all the positive components to zero:

$$\vec{s}_* = \sum_{\mu = \mathcal{C}+1}^{\mathcal{P}} s'_\mu \Theta(-s'_\mu)\vec{\varepsilon}_\mu = \mp \sum_{\mu = \mathcal{C}+1}^{\mathcal{P}} s_\mu \, \Theta(\pm s_\mu)\vec{\varepsilon}_\mu, \quad (A33)$$

depending upon whether $\mathrm{Sgn}(e_y) = \pm$. Here s'_μ, s_μ are just the μth components of $\widehat{s}'$ and $\widehat{s}$ vectors, and $\Theta(x)$ is the Heaviside step function, which is one if x is positive and zero otherwise. Then, γ is given by

$$\cos \gamma = \widehat{s}' \cdot \widehat{s}_* = \frac{\widehat{s}' \cdot \vec{s}_*}{|\vec{s}_*|} = \sqrt{\sum_{\mu = \mathcal{C}+1}^{\mathcal{P}} s'^2_\mu \, \Theta(-s'_\mu)}$$

$$= \sqrt{\sum_{\mu = \mathcal{C}+1}^{\mathcal{P}} s^2_\mu \, \Theta(\pm s_\mu)} = |\vec{s}_*|, \qquad (A34)$$

where again the sign in Θ is determined by the sign of e_y.

A formal way to see that γ is indeed given by Eq. (A34) is to start with any unit vector, $\vec{w}_s$, lying in the solution space. Then, the angle, γ, between $\widehat{s}'$ and $\vec{w}_s$ is given by

$$\cos \gamma = \sum_{\mu = \mathcal{C}+1}^{\mathcal{P}} s'_\mu w_{s\mu} = \sum_{\mu \in A_+} s'_\mu w_{s\mu} + \sum_{\mu \in A_-} s'_\mu w_{s\mu}, \quad (A35)$$

where we have defined $A_\pm$ as the set of all μ indices for which s'_μ is positive/negative, respectively. Since $\vec{w}_s$ is in the solution space, $w_{s\mu} \leqslant 0$, and therefore the second term sums positive quantities while the first term subtracts. Thus,

$$\cos \gamma \leqslant \sum_{\mu \in A_-} s'_\mu w_{s\mu} = \vec{s}_* \cdot \vec{w}_s \leqslant |\vec{s}_*|, \qquad (A36)$$

where both the equalities are achieved when $\vec{w}_s$ is aligned with the boundary semiconstrained vector, $\vec{s}_*$, or $\vec{w}_s = \widehat{s}_*$, as argued previously. Note also, that this formal proof did not assume any restrictions on $\widehat{s}'$ direction and thus Eq. (A34) turns out

to be a general result that also holds if $\widehat{s}'$ points into the all-negative orthant.

Combining Eqs. (A34) and (A30), the certainty condition now reads

$$\varphi < \gamma \Rightarrow \frac{y^2 \cot^2 \theta \cos^2 \alpha}{W^2 - y^2} > \cos^2 \gamma$$

$$\Rightarrow y > y_{\mathrm{cr}} \equiv W \sqrt{\frac{\cos^2 \gamma \sin^2 \theta}{\cos^2 \alpha \cos^2 \theta + \cos^2 \gamma \sin^2 \theta}}. \quad (A37)$$

d. General case with unconstrained dimensions

We can extend the above analysis to the case when we have unconstrained dimensions by noting that, for a given set of unconstrained coordinates, the solution space is again the all-negative orthant in a semiconstrained hypersphere. Isometry along unconstrained dimensions ensures that it is always possible to make one of the null directions, lets say $\vec{\varepsilon}_\mathcal{N}$, align with $\widehat{u}$. Then, the $w = 0$ hyperplane Eq. (A26) reads

$$w = \sin \theta \cos \phi \, \widehat{s} \cdot \vec{w}_s + y \cos \theta \cos \alpha + \eta_u \sin \theta \sin \phi = 0, \qquad (A38)$$

which can be rewritten as

$$\sin \theta \cos \phi \, \widehat{s}' \cdot \vec{w}_s = y \cos \theta |\cos \alpha| - \eta'_u \sin \theta \sin \phi, \quad (A39)$$

where we have introduced $\eta'_u = -\mathrm{Sgn}(\cos \alpha)\eta_u$. To have a certain synapse, the $w = 0$ hyperplane cannot intersect the solution space for any allowed value of η'_u.

The direction of $\widehat{s}'$ is independent of the unconstrained coordinates and hence the value of γ remains unchanged. However, the cone-angle, φ, does depend on the unconstrained coordinates in two ways. First, the radius, $\widetilde{W}$, of the $\mathcal{S}$-dimensional spherical subspace containing admissible solutions is now

$$\widetilde{W} = \sqrt{W^2 - y^2 - \eta_\perp^2 - \eta'^2_u}, \qquad (A40)$$

where $\eta_\perp$ is the magnitude of the weight-vector in the $(\mathcal{U} - 1)$-dimensional subspace that is perpendicular to $\vec{\varepsilon}_\mathcal{N} = \widehat{u}$. We note in passing that Eq. (A40) implies $\eta_\perp, \eta'_u \leqslant \sqrt{W^2 - y^2} = \overline{W}$. Second, the distance of the hyperplane from the origin that follows from Eq. (A39) is now a function of η'_u:

$$d_s = \frac{y|\cos \theta \cos \alpha| - \eta'_u \sin \theta \sin \phi}{\sin \theta \cos \phi}. \qquad (A41)$$

Strictly speaking, this expression for the distance is only valid as long as the numerator in the d_s expression stays positive. However, if there exists an allowed $\eta'_u \leqslant \overline{W}$ (let us call it η_{u0}) for which the numerator can vanish, then the synapse cannot have a certain sign, because at that point $d_s = 0$, the hyperplane intersects the origin, and the weight can vanish even for a linear theory. In fact, the $d_s = 0$ condition provides us with the y-critical value below which the synapse sign

becomes uncertain in a linear theory:

$$\eta_{u0} = \frac{y \cos\theta |\cos\alpha|}{\sin\theta \sin\phi} \leqslant \overline{W} \Rightarrow y_{\text{cr,lin}} = W \sqrt{\frac{\sin^2\theta \sin^2\phi}{\cos^2\theta \cos^2\alpha + \sin^2\theta \sin^2\phi}}. \tag{A42}$$

So we will now look into cases when $y \geqslant y_{\text{cr,lin}}$ which also means that Eq. (A41) will remain valid.

Combining Eqs. (A40) and (A41) we get

$$\cos\varphi = \frac{d_s}{\widetilde{W}} = \frac{y|\cos\theta \cos\alpha| - \eta'_u \sin\theta \sin\phi}{\sin\theta \cos\phi \sqrt{W^2 - y^2 - \eta_\perp^2 - \eta'^2_u}}. \tag{A43}$$

For us to be certain that w is nonzero, we have to make sure that even the largest φ that one can obtain by varying $\eta_\perp$ and η'_u is still smaller than γ. Clearly, to make φ large it is best to make $\eta_\perp = 0$. Also, it is clear from inspection that $\cos\varphi$ starts to initially decrease as η'_u increases from zero, being dominated by the linear term. However, as the quadratic term in η'_u in the denominator becomes more and more important, $\cos\varphi$ reaches a minimum and starts to increase. Imposing $d \cos\varphi/d\eta'_u = 0$, we can find that this minimum is reached at

$$\eta'_u = \frac{\sin\theta \sin\phi \, (W^2 - y^2)}{y \cos\theta |\cos\alpha|} = \overline{W} \sqrt{\frac{(W/y)^2 - 1}{(W/y_{\text{cr,lin}})^2 - 1}} \leqslant \overline{W}, \tag{A44}$$

where we substituted $y_{\text{cr,lin}}$ from Eq. (A42) and used the fact that $W \geqslant y \geqslant y_{\text{cr,lin}}$ to obtain the inequality. This proves that the minimum $\cos\varphi$ indeed occurs at an allowed positive value of $\eta'_u \leqslant \overline{W}$. Substituting the above η'_u in Eq. (A43), after some algebra we find that this minimum value of $\cos\varphi$, or equivalently the maximum φ, is given by

$$\cos\varphi = \sqrt{\frac{y^2 \cos^2\theta \cos^2\alpha - (W^2 - y^2)\sin^2\theta \sin^2\phi}{(W^2 - y^2)\cos^2\phi \sin^2\theta}}. \tag{A45}$$

The certainty condition then requires

$$\cos^2\varphi = \frac{y^2 \cos^2\theta \cos^2\alpha - (W^2 - y^2)\sin^2\theta \sin^2\phi}{(W^2 - y^2)\cos^2\phi \sin^2\theta} > \cos^2\gamma, \tag{A46}$$

which can be recast as

$$y > y_{\text{cr}} \equiv W \sqrt{\frac{\cos^2\gamma \sin^2\theta \cos^2\phi + \sin^2\theta \sin^2\phi}{\cos^2\alpha \cos^2\theta + \cos^2\gamma \sin^2\theta \cos^2\phi + \sin^2\theta \sin^2\phi}}, \tag{A47}$$

It is illuminating to express y-critical in terms of the projections, e_y, e_u, e_{s*}, of the synaptic direction, $\widehat{e}$, respectively along the data vector, $\widehat{y}$, the unconstrained unit vector, $\widehat{u}$, and the semiconstrained boundary vector, $\vec{s}_*$:

$$y_{\text{cr}} = W \sqrt{\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}}, \tag{A48}$$

where

$$e_y \equiv \widehat{e} \cdot \widehat{y} = \frac{\sum_{\mu=1}^{\mathcal{C}} y_\mu e_\mu}{\sqrt{\sum_{\mu=1}^{\mathcal{C}} y_\mu^2}} = \cos\theta \cos\alpha, \quad e_{s*} \equiv \widehat{e} \cdot \widehat{s}_* = \sqrt{\sum_{\mu \in A_-} e_\mu^2} = -\text{Sgn}(\cos\alpha)\sin\theta \cos\phi \cos\gamma, \tag{A49}$$

and e_u is given by Eq. (A22). We note that setting $\phi = 0$ precisely reproduces the correct limit with no unconstrained directions (A37).

3. Regarding orthogonal input patterns in recurrent networks

While our analysis of the solution space and the certainty condition (A48) translate directly to recurrent networks, the requirement of orthogonality for the derivation of our certainty condition imposes certain technical restrictions on its scope when it comes to recurrent neural networks.

The certainty condition we derived for feedforward networks can be applied to two different recurrent neural network set ups. First, let us consider networks where neurons have

self-couplings. A consequence of having orthogonal response patterns in this case is that the certainty condition can only be satisfied for self-couplings w_{ii}, as long as $W \geqslant 1$. This is because the imposition of orthogonality in response patterns also restricts the correlation between the target neuron and the other neurons:

$$\sum_{m=1}^{\mathcal{N}} Z_{\mu m} Z_{\nu m} = \delta_{\mu\nu} \Rightarrow \sum_{\mu=1}^{\mathcal{N}} Z_{\mu m} Z_{\mu n} = \delta_{mn}. \tag{A50}$$

However, for the synapse-sign to be certain, the responses of pre and postsynaptic neurons need to be correlated. To see the problem more quantitatively, suppose we are interested in constraining the synapse from the mth neuron onto the

ith neuron, as before. Now, the first $\mathcal{P}$ elements of the unit vectors, $\widehat{e}_i$ and $\widehat{e}_m$ contain the responses of the ith and the mth neuron in the $\mathcal{P}$ patterns. We have already derived a decomposition of $\widehat{e}_m$ in terms of its projections onto the constrained, semiconstrained and unconstrained subspaces (A20). Similarly, $\widehat{e}_i$ can be decomposed as

$$\widehat{e}_i = y\widehat{y} + \sqrt{1-y^2}\,\widehat{y}_\perp, \tag{A51}$$

where $\widehat{y}$ lies entirely along the constrained directions, and $\widehat{y}_\perp$ is orthogonal to it and only has components along unconstrained directions. Then, orthogonality implies

$$\widehat{e}_i \cdot \widehat{e}_m = y\cos\theta\cos\alpha + \sqrt{1-y^2}\sin\theta\sin\phi\,(\widehat{y}_\perp \cdot \widehat{u})$$

$$= 0 \Rightarrow \sin\theta\sin\phi = -\frac{y\cos\theta\cos\alpha}{(\widehat{y}_\perp \cdot \widehat{u})\sqrt{1-y^2}}. \tag{A52}$$

Starting from the certainty condition (A48), we can now go through a sequence of (in)equalities:

$$y^2 > W^2 \frac{\sin^2\theta\sin^2\phi + \sin^2\theta\cos^2\phi\cos^2\gamma}{\sin^2\theta\sin^2\phi + \sin^2\theta\cos^2\phi\cos^2\gamma + \cos^2\theta\cos^2\alpha}$$

$$\geqslant \frac{W^2\sin^2\theta\sin^2\phi}{\sin^2\theta\sin^2\phi + \cos^2\theta\cos^2\alpha}$$

$$= \frac{W^2 y^2\cos^2\theta\cos^2\alpha}{y^2\cos^2\theta\cos^2\alpha + (1-y^2)(\widehat{y}_\perp \cdot \widehat{u})^2\cos^2\theta\cos^2\alpha}$$

$$= \frac{W^2 y^2}{y^2 + (1-y^2)(\widehat{y}_\perp \cdot \widehat{u})^2} \geqslant W^2 y^2, \tag{A53}$$

where we substituted $\sin\theta\sin\phi$ from Eq. (A52). Note that the RHS is minimized when $\widehat{u}$ and $\widehat{y}_\perp$ are either aligned or antialigned. Even in this case, RHS $= W^2 y^2$, and thus the certainty condition cannot be satisfied if $W \geqslant 1$. One can check that when $i = m$, because the RHS in the first equation of (A52) is one and not zero, no similar constraints appear. Indeed, the certainty condition may be satisfied depending upon the specific response patterns.

As a second possibility, suppose that no self-couplings are present. Then to be able to apply our framework and determine the couplings w_{im} for a given i, we only need the truncated row vectors of z whose ith column entry is absent, to be orthonormal. Therefore, the response of the ith driven neuron, which consists of the entries of the ith column, can now be chosen independently from the responses of its input neurons. In other words, $\widehat{e}_i$ and $\widehat{e}_m$, $m \neq i$, no longer need to satisfy orthogonality constraint of Eq. (A52). Consequently, the w_{im} weights can indeed satisfy the certainty condition, just as in the feedforward case.

4. Implied conservative bound on the certainty condition for nonorthogonal input patterns

A complete treatment of the certainty conditions for nonorthogonal fixed point patterns is beyond the scope of this work. However, here we provide some preliminary results and insights by explaining how our formalism for analyzing orthogonal fixed-point patterns can be simply adapted to derive an exact, but conservative, upper bound for y-critical that applies to general sets of patterns.

Conceptually speaking, deriving the certainty condition amounts to determining when the $w_m = 0$ hyperplane intersects the solution space within the sphere of weight vectors with norm at most W. Because the solution space is exceedingly simple in η coordinates, our orthogonal analysis used η coordinates to conveniently recast the equations for the bounding sphere and $w_m = 0$ hyperplane. To adapt this analysis to the nonorthogonal case, it is important to account for three important changes to the geometry of the problem. Most fundamentally, the $\vec{\varepsilon}_\mu$ directions corresponding to η coordinates are no longer orthonormal. However, the mathematical notions of orthogonality and normality are implicitly defined with respect to the inner-product structure imposed on the vector space, and our geometrical calculations from the orthogonal case easily carry over to the nonorthogonal case if we redefine the inner product structure of the weight space to give an orthonormal coordinate system with respect to the η coordinates rather than the physical coordinates.[15] In practice, all that this will entail is interpreting the η coordinates as if they define coordinates along orthogonal axes, and we will never need to explicitly write down the associated inner product. Second, the equation for the weight bound in the orthogonal system of η coordinates is elliptical around the origin, rather than spherical. However, for any ellipse one can find a sphere that just encompasses it. If we can find the radius of this bounding sphere, then one can look for an intersection anywhere within this sphere and our geometrical approach for deriving Eq. (A48) will carry over and provide a conservative bound for y-critical. This bound will poorly approximate the true y-critical when some axes of the ellipse are much longer than others. Third, the normal vector to the $w_m = 0$ hyperplane is no longer $\widehat{e}_m$ in the orthogonal system of η coordinates. Therefore, the projections of $\widehat{e}_m$ in Eq. (A48) must be generalized to become projections of the hyperplane's normal vector.

To obtain the radius of the bounding sphere, consider the SVD decomposition of the data-matrix:

$$z = L\Lambda R^T, \tag{A54}$$

where L and R are $\mathcal{P} \times \mathcal{P}$ and $\mathcal{N} \times \mathcal{N}$ orthogonal rotation matrices, and Λ is a $\mathcal{P} \times \mathcal{N}$ rectangular diagonal matrix whose only nonzero entries are given by

$$\Lambda_{aa} = \lambda_a \geqslant 0, \quad a = 1 \dots \mathcal{P}. \tag{A55}$$

Note that $\lambda_1, \cdots, \lambda_\mathcal{P}$ are called the singular values of z. We can now define rotated coordinates:

$$\eta' = L^T\eta, \quad \text{and} \quad w' = R^T w, \tag{A56}$$

so that

$$\eta' = L^T L\Lambda R^T w = \Lambda w' \Rightarrow \eta'_a = \lambda_a w'_a \,\forall\, a = 1 \dots \mathcal{P}. \tag{A57}$$

---

[15]In particular, for an orthogonal Z matrix, the two inner product structures defined via $\widehat{e}_m \cdot \widehat{e}_n = \delta_{mn}$ and $\vec{\varepsilon}_\mu \cdot \vec{\varepsilon}_\nu = \delta_{\mu\nu}$ are equivalent, but this is not the case when Z is nonorthogonal. Although one would conventionally adopt the first inner product structure, both the derivation and interpretation of the conservative y-critical formula is easier in terms of the latter inner product structure, which makes all of the response pattern directions orthonormal by definition.

Note that η and η' are $\mathcal{P}$ vectors in the current notation. We also note that since w' is just a rotation of the original synaptic coordinates, the biological bound does not change as we go from w to w' coordinates:

$$\sum_a w_a'^2 = \sum_m w_m^2 \leqslant W^2. \tag{A58}$$

This makes it possible to find an inequality in terms of the η' coordinates:

$$
\begin{aligned}
W^2 &\geqslant \sum_{a=1}^{\mathcal{N}} w_a'^2 = \sum_{a=1}^{\mathcal{P}} \frac{\eta_a'^2}{\lambda_a^2} + \sum_{a=\mathcal{P}+1}^{\mathcal{N}} w_a'^2 \\
&\geqslant \frac{1}{\lambda_{\max}^2} \sum_{a=1}^{\mathcal{P}} \eta_a'^2 + \sum_{a=\mathcal{P}+1}^{\mathcal{N}} w_a'^2 \\
&= \frac{1}{\lambda_{\max}^2} \sum_{\mu=1}^{\mathcal{P}} \eta_\mu^2 + \sum_{a=\mathcal{P}+1}^{\mathcal{N}} w_a'^2,
\end{aligned}
\tag{A59}
$$

where in the last step we have used the fact that the orthogonal matrix L does not change the L2-norm as one goes from η' to η coordinates, and $\lambda_{\max}$ is defined as the maximal singular value. To obtain spherical symmetry, we thus define unconstrained coordinates via

$$\eta_\mu = \lambda_{\max} w_\mu', \quad \forall \, \mu > \mathcal{P}, \tag{A60}$$

such that

$$W'^2 \equiv \lambda_{\max}^2 W^2 \geqslant \sum_{\mu=1}^{\mathcal{N}} \eta_\mu^2. \tag{A61}$$

We now realize that the problem of finding y-critical using this conservative bound can be recast into the problem of the orthogonal case: As we just described, the η coordinates satisfy the conservative spherical bound. Our goal can then be to find the minimum value of y for which the hyperplane satisfying, $w_m = 0$, does not intersect the solution space. Now, if we define Z to be, as in the orthogonal case, a full rank extension[16] of z:

$$Z_{\mu m} = \begin{cases} z_{\mu m}, & \forall \, \mu \leqslant \mathcal{P}, \\ \lambda_{\max} R_{\mu m}^T, & \forall \, \mu > \mathcal{P}, \end{cases} \tag{A63}$$

so that for all value of μ,

$$\eta_\mu = \sum_{m=1}^{\mathcal{N}} Z_{\mu m} w_m, \tag{A64}$$

---

[16]One can equivalently obtain Z as

$$Z = \widetilde{L}\widetilde{\Lambda}R, \tag{A62}$$

where $\widetilde{\Lambda}$ is now an $\mathcal{N} \times \mathcal{N}$ diagonal matrix with same entries as Λ and $\widetilde{\Lambda}_{\mu\mu} = 1/\lambda_{\max}$ for $\mu > \mathcal{P}$, and $\widetilde{L}$ is the an $\mathcal{N}$-dimensional extension of the rotation matrix L where the $(\mathcal{N} - \mathcal{P})$-dimensional block is an identity and there is no mixing between the unconstrained and the activity-constrained coordinates.

then the hyperplane equation can be rewritten as

$$w_m = \sum_{\mu=1}^{\mathcal{N}} Z_{m\mu}^{-1} \eta_\mu = 0. \tag{A65}$$

From Eq. (A65) it is clear that

$$\vec{n}_m = \sum_{\mu=1}^{\mathcal{N}} Z_{m\mu}^{-1} \vec{\varepsilon}_\mu \tag{A66}$$

is perpendicular to the $w_m = 0$ hyperplane, where we remind the readers that $\vec{\varepsilon}_\mu$'s were defined by Eq. (A9). $\hat{n}_m$ thus plays the role of $\widehat{e}_m$ whose orientation with respect to the constrained, semiconstrained and unconstrained dimensions determines the certainty condition. Specifically,

$$y_{\mathrm{cr}} = W' \sqrt{\frac{n_{s*}^2 + n_u^2}{n_y^2 + n_{s*}^2 + n_u^2}} = W\lambda_{\max} \sqrt{\frac{n_{s*}^2 + n_u^2}{n_y^2 + n_{s*}^2 + n_u^2}}, \tag{A67}$$

where the various projections of $\hat{n}_m$ are given by

$$n_y \equiv \frac{\sum_{\mu=1}^{\mathcal{C}} y_\mu Z_{m\mu}^{-1}}{\sqrt{\sum_{\mu=1}^{\mathcal{C}} y_\mu^2} \sqrt{\sum_{\mu=1}^{\mathcal{N}} \left(Z_{m\mu}^{-1}\right)^2}},$$

$$n_{s*} \equiv \frac{\sqrt{\sum_{\mu \in A_-} \left(Z_{m\mu}^{-1}\right)^2}}{\sqrt{\sum_{\mu=1}^{\mathcal{N}} \left(Z_{m\mu}^{-1}\right)^2}},$$

$$\text{and} \quad n_u \equiv \frac{\sqrt{\sum_{\mu=\mathcal{P}+1}^{\mathcal{N}} \left(Z_{m\mu}^{-1}\right)^2}}{\sqrt{\sum_{\mu=1}^{\mathcal{N}} \left(Z_{m\mu}^{-1}\right)^2}}. \tag{A68}$$

We remind the readers that in the orthogonal case the set A_- contained each semiconstrained μ index along which the component of $\widehat{e}_m$ had the same sign as e_y. Similarly, here A_-, contains those semiconstrained μ indices along which the component of $\vec{n}_m$ have the same sign as n_y. To summarize, the above analysis suggests that both Z and Z^{-1} will play an important role in generalizing the certainty condition to nonorthogonal patterns, and especially the relative orientation of $\hat{n}_m$ (defined by Z^{-1}) with respect to the η directions.

APPENDIX B: ESTIMATING THE PROBABILITY THAT A SYNAPSE IS CERTAIN IN LARGE FEEDFORWARD NETWORKS

For given values of $\mathcal{N}(=\mathcal{I}), \mathcal{C}, \mathcal{P}$ in a feedforward setting we will here try to assess how likely it is that noiseless orthonormal neuronal responses require a given synapse to be nonzero. As we have seen in Eq. (49), whether a synapse is certain to exist depends on six parameters, $\theta, \phi, \gamma, \alpha, W$, and y. The first four quantities depend on how $\widehat{e}$ is oriented with respect to various directions in the weight space. Since $\widehat{e}$ is a unit vector, typically we expect its component along any given direction to be $\sim \mathcal{O}(1/\sqrt{\mathcal{N}})$. Thus, we typically expect

$$e_y^2 = \cos^2\theta \cos^2\alpha \sim \frac{1}{\mathcal{N}}; \quad e_s^2 = \sin^2\theta \cos^2\phi \sim \frac{\mathcal{S}}{\mathcal{N}};$$

$$e_u^2 = \sin^2\theta \sin^2\phi \sim \frac{\mathcal{U}}{\mathcal{N}} \quad \text{and} \quad e_{s*}^2 = \cos^2\gamma \, e_s^2 \sim \frac{\mathcal{S}}{2\mathcal{N}}. \tag{B1}$$

Hence, we approximate the typical y_{cr} as

$$y_{cr} = W\sqrt{\frac{\mathcal{S}/(2\mathcal{N}) + \mathcal{U}/\mathcal{N}}{1/\mathcal{N} + \mathcal{S}/(2\mathcal{N}) + \mathcal{U}/\mathcal{N}}} = W\sqrt{\frac{\mathcal{S} + 2\mathcal{U}}{2 + \mathcal{S} + 2\mathcal{U}}}. \tag{B2}$$

Let us now suppose that all the dimensions scale with the network size, such that

$$\mathcal{S} = \sigma\mathcal{N} \quad \text{and} \quad \mathcal{U} = \upsilon\mathcal{N}. \tag{B3}$$

Then, we find that as the network size increases y_{cr} behaves as

$$\frac{y_{cr}}{W} \approx 1 - \left(\frac{1}{\sigma + 2\upsilon}\right)\frac{1}{\mathcal{N}}, \tag{B4}$$

and y_{cr} is essentially pushed up toward W.

However, the typical scale of y behaves similarly as the dimensions increase. To see this concretely, let us define $\vec{y}_{cons} \equiv \{y_\mu | \mu = 1 \ldots \mathcal{C}\}$ as a $\mathcal{C}$-dimensional vector, and assume that every possible $\vec{y}_{cons}$ is equally likely within a sphere of radius W (larger activity levels of the target neuron admit no solutions). Then the average and median values of $y = |\vec{y}_{cons}|$ are given by

$$\langle y \rangle = \frac{\int_0^W dy\, y^{\mathcal{C}}}{\int_0^W dy\, y^{\mathcal{C}-1}} = W\left(\frac{\mathcal{C}}{\mathcal{C}+1}\right)$$

$$\Rightarrow \frac{\langle y \rangle}{W} = \frac{1}{1 + 1/(\eta\mathcal{N})} \approx 1 - \frac{1}{\eta\mathcal{N}}, \quad \text{where } \eta \equiv \frac{\mathcal{C}}{\mathcal{N}},$$

$$\text{and} \quad \frac{\int_0^{y_M} dy\, y^{\mathcal{C}-1}}{\int_0^W dy\, y^{\mathcal{C}-1}} = \frac{1}{2} \Rightarrow \frac{y_M}{W} = \left(\frac{1}{2}\right)^{\frac{1}{\mathcal{C}}}$$

$$\Rightarrow \frac{y_M}{W} \approx 1 - \frac{\ln 2}{\eta\mathcal{N}}, \tag{B5}$$

respectively. Since y and y_{cr} scale similarity as one increases the network size, the probability of a synapse being certain should not change as the network size increases. In Fig. 8(c), we show that if we choose, $y = 1 - \ln 2/\mathcal{C}$, as the approximate median value in simulations with random input-output configurations (see Appendix F for details), then the number of certain synapses does indeed increase linearly with $\mathcal{N}$.

To quantitatively estimate the probability of finding a certain synapse, we can compute the fraction of volume of $\vec{y}_{cons}$'s for which the synapse is certain for the typical projections (B2), as compared to the volume of $\vec{y}_{cons}$'s for which solutions to the steady-state equations exist. We know that $\vec{y}_{cons}$ has to lie within a $\mathcal{C}$-dimensional sphere of radius W in order for there to be any solutions to the problem.[17] However, for the synapse sign to be certain, we need $W \geqslant |\vec{y}_{cons}| = y > y_{cr}$, where y_{cr} is given by Eq. (B2). So, we need to compare the spherical shell volume, $V_{W \geqslant y > y_{cr}}$, with the volume of the $\mathcal{C}$-dimensional sphere, $V_{y \leqslant W}$. To find $V_{W \geqslant y > y_{cr}}$, we have to subtract the $\mathcal{C}$-dimensional spherical volume with radius y_{cr} from the spherical volume with radius W. Since

---

[17]The allowed $\vec{y}_{cons}$ must also lie in the all positive orthant, but as we will compute the ratio of two spherical volumes the reduction factor will cancel out.

n-dimensional spherical volumes scale as the nth power of the radius, the probability, P, of ascertaining the sign of the synapse is approximately given by

$$P \approx \frac{V_{W \geqslant y > y_{cr}}}{V_{y \leqslant W}} = \frac{W^{\mathcal{C}} - y_{cr}^{\mathcal{C}}}{W^{\mathcal{C}}} = 1 - \left(\frac{y_{cr}}{W}\right)^{\mathcal{C}}. \tag{B6}$$

Now, when $\mathcal{S}, \mathcal{U} \gg 1$ we can approximately evaluate the RHS as follows:

$$\left(\frac{y_{cr}}{W}\right)^{\mathcal{C}} = \left(\frac{\mathcal{S} + 2\mathcal{U}}{2 + \mathcal{S} + 2\mathcal{U}}\right)^{\mathcal{C}/2} = \left(1 + \frac{2}{\mathcal{S} + 2\mathcal{U}}\right)^{-\mathcal{C}/2}$$

$$\Rightarrow \ln\left[\left(\frac{y_{cr}}{W}\right)^{\mathcal{C}}\right] = -\frac{\mathcal{C}}{2}\ln\left(1 + \frac{2}{\mathcal{S} + 2\mathcal{U}}\right)$$

$$= \frac{-\mathcal{C}}{\mathcal{S} + 2\mathcal{U}}\left[1 + \mathcal{O}\left(\frac{1}{\mathcal{S} + 2\mathcal{U}}\right)\right]. \tag{B7}$$

Thus, we get

$$P \approx 1 - e^{-\frac{\mathcal{C}}{\mathcal{S} + 2\mathcal{U}}}. \tag{B8}$$

The most prominent feature of Eq. (B8) is that the probability only depends on the ratios of the various dimensions. Hence, it does not change as we increase the size of the network as long as the ratios are kept constant.

For the purpose of illustration and numerically testing this feature we assessed how certainty predictions changed when the network size is increased while holding the ratios between $\mathcal{C}, \mathcal{S}$, and $\mathcal{U}$ fixed. In Fig. 8(c) we have plotted the number of certain synapses in simulations generated from random data as we scale up $\mathcal{N}$ maintaining the ratios between $\mathcal{C}, \mathcal{S}$, and $\mathcal{U}$ (see Appendix F for more details). We illustrate two cases. In the first example, no unconstrained directions were present, and $\mathcal{S} = 3\mathcal{C}$. Then $P = 1 - e^{-1/3} \approx 0.28$, so one has a 28% chance of being able to determine the sign of the connections. This answer incidentally is the same as an example with $\mathcal{C} = \mathcal{S} = \mathcal{U}$. As another example, Fig. 8(c) considered the case when $\mathcal{U} = 2\mathcal{C} = 2\mathcal{S}$. According to Eq. (B8), then $P = 1 - e^{-1/5} \approx 0.18$, so the chance of determining the sign drops to about 18%. We only expect these numbers to be approximate. For example, our arguments relied on the assumption that all target responses admitting solutions are equally likely, an assumption that definitely needs to be revisited for realistic networks. However, the scaling behavior should hold for other probabilistic distributions as long as the scale of $\vec{y}_{cons}$ behaves similar to Eq. (B5) with increasing $\mathcal{N}$.

APPENDIX C: NONZERO-ERROR CERTAINTY CONDITIONS

There are various reasons why we may want to not only consider weights that exactly reproduce the specified neuronal responses, but also weights that do so approximately. For instance, we are always limited by the accuracy of the measurement apparatus. More importantly, there are various sources of biological noise that typically lead to uncertainties in observed values of neuronal responses. For the purpose of this paper we will consider any set of weights to be part of the ε-error solution space if it is able to reproduce the specified neuronal responses with an error $\mathcal{E} \leqslant \varepsilon$ [see Eq. (63) for definition of $\mathcal{E}$]. We will neglect uncertainties in the input

responses to the target neuron, but we will comment on their possible effects toward the end of this Appendix.

1. Errors in feedforward networks

Let us first focus on feedforward networks. Allowing for error increases the value of y_{cr} by expanding the solution space. One way to think about this is to realize that we have to now make sure that Eq. (51) is satisfied for any nonnegative $\widetilde{\vec{y}} = \vec{y} + \vec{\delta}$, where $\vec{y}$ and $\vec{\delta}$ are vectors in the $\mathcal{P}$-dimensional activity-constrained subspace, the former representing the observed responses, and the latter coming from noise. We will initially assume that all the observed responses are nonnegative, so a zero-error solution is possible and the noise is bounded by $|\vec{\delta}| \leqslant \varepsilon$. Our strategy will be to first seek the minimum y needed to have a certain synapse for a given $\vec{\delta}$. We then find the maximum among these y-critical values as we let $\vec{\delta}$ vary within the ε-ball. Since this procedure will guarantee that the $w = 0$ hyperplane does not intersect the entire solution space with $\mathcal{E} \leqslant \varepsilon$, this means that the synapse must exist for the network to generate the specified responses patterns. The synapse's sign will match the zero-error analysis. We will first estimate y-critical when the error is small enough to not induce topological transitions in the error surface. In the subsequent sections, we will include the effects of topological transitions, as well as explain how to deal with situations where some of the observed responses are negative, which is possible due to noise.

a. When all observed responses are nonnegative and no topological transitions occur

To understand how errors affect the certainty conditions, let us consider the case where the observed responses are nonnegative and the allowed error satisfies $0 < \varepsilon < \min\{y_\mu\}_{\mu=1,...,\mathcal{C}}$, so that no topological transitions can occur. If some responses that were zero in $\vec{y}$ are now nonzero in $\widetilde{\vec{y}}$, then both e_y and e_{s*} can change due to the noise. Without loss of generality, let us assume that $\vec{\delta}$ only has nonzero components along $\mu = \mathcal{C} + 1, \ldots, \mathcal{Q}$ semiconstrained dimensions,[18] as well as along some (or all) of the constrained dimensions. Then e_y changes to

$$\widetilde{e}_y = \frac{\widehat{e} \cdot (\vec{y} + \vec{\delta})}{|\vec{y} + \vec{\delta}|}. \tag{C1}$$

Furthermore, if some previously semiconstrained components that contributed to $\widehat{s}_*$ have now become constrained,[19] then $\widehat{s}_*$ no longer has those components. This means that we have to subtract these components from e_{s*}:

$$\widehat{\widetilde{s}}_* = \widehat{s}_* - \sum_{\mu=\mathcal{C}+1}^{\mathcal{Q}} A_\mu e_\mu \vec{\varepsilon}_\mu \Rightarrow \widetilde{e}_{s*}^2 = e_{s*}^2 - \sum_{\mu=\mathcal{C}+1}^{\mathcal{Q}} A_\mu e_\mu^2, \tag{C2}$$

where A_μ is 1 if $(\widehat{y} \cdot \widehat{c})$ and e_μ have the same sign and 0 otherwise. This follows from the definition of the boundary

---

[18]The components of $\vec{\delta}$ along these semiconstrained directions are all positive since $\vec{y}$ must be nonnegative.

[19]This will happen if a component of $\vec{y}$ that was zero now has a nonzero component.

projection vector (A33) and e_{s*} (A49). Thus, for a given $\vec{\delta}$ the certainty condition (A48) yields

$$|\widetilde{\vec{y}}|^2 = |\vec{y} + \vec{\delta}|^2 > W^2 \left(\frac{\widetilde{e}_{s*}^2 + e_u^2}{\widetilde{e}_y^2 + \widetilde{e}_{s*}^2 + e_u^2} \right)$$

$$\Rightarrow |\vec{y} + \vec{\delta}|^2 \left(\frac{\widetilde{e}_y^2}{\widetilde{e}_{s*}^2 + e_u^2} + 1 \right) > W^2$$

$$\Rightarrow \frac{|\widehat{e} \cdot (\vec{y} + \vec{\delta})|^2}{\widetilde{e}_{s*}^2 + e_u^2} + |\vec{y} + \vec{\delta}|^2 > W^2. \tag{C3}$$

As before, one can interpret the above inequality as equivalently specifying either y-critical or W-critical. For a fixed $\vec{y}$, one can obtain a minimum value of the left-hand side (LHS) of the latter inequality by varying $\vec{\delta}$ within the ε ball. The square root of this is W-critical. Then as long as W is less than W-critical, we will have a certain synapse. Inverting the relation, one finds y-critical as the minimum y needed to make the synapse sign certain for all $\vec{\delta}$ and given $\widehat{y}$ and W. More explicitly, equating the two sides of the inequality for any given $\widehat{y}$, $\vec{\delta}$, and W, we get a minimal y that depends on $\vec{\delta}$. To find y-critical, we have to take the maximum of the minimal y as we vary over all possible $\vec{\delta}$ in the ε-ball.

Let us first obtain a lower bound on y-critical. By inspection of the LHS of the above inequality, it is clear that the more the $\vec{\delta}$-dependent terms can cancel the $\vec{y}$-dependent terms, the harder it is to satisfy the certainty condition. We observe that in Eq. (C3), the second term is minimized when $\vec{\delta} = -\varepsilon\widehat{y}$.[20] Accordingly, one can obtain a lower bound on y-critical by substituting $\delta = -\varepsilon\widehat{y}$ in Eq. (C3):

$$(\widehat{e} \cdot \widehat{y})^2 (y - \varepsilon)^2 + (e_{s*}^2 + e_u^2)(y - \varepsilon)^2 > W^2(e_{s*}^2 + e_u^2), \text{ or,}$$

$$y > y_{cr,min,0} \equiv W \sqrt{\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}} + \varepsilon, \tag{C4}$$

where we have used $e_y = \widehat{e} \cdot \widehat{y}$ and $\widetilde{e}_{s*}^2 = e_{s*}^2$, since $\vec{\delta}$ has no components along the semiconstrained directions. We will see later that this simple lower bound can approximate the actual y-critical very well in many situations. Notice that we used a subscript "0" to denote this lower bound. This is because, as we will soon see, when noise allows for topological transitions, one may be able to obtain stricter lower bounds by allowing some constrained dimensions to behave as semiconstrained. This "0" emphasized that no constrained indices behave as semiconstrained. Next, we can find an upper bound for y-critical by noting

$$\frac{|\widehat{e} \cdot (\vec{y} + \vec{\delta})|^2}{\widetilde{e}_{s*}^2 + e_u^2} + |\vec{y} + \vec{\delta}|^2$$

$$\geqslant \frac{|\widehat{e} \cdot (\vec{y} + \vec{\delta})|^2}{e_{s*}^2 + e_u^2} + |\vec{y} + \vec{\delta}|^2$$

$$\geqslant \frac{(\widehat{e} \cdot \vec{y})^2 + 2(\widehat{e} \cdot \vec{y})(\widehat{e} \cdot \vec{\delta})}{e_{s*}^2 + e_u^2} + y_0^2 + 2\,\vec{y} \cdot \vec{\delta}, \tag{C5}$$

---

[20]This assumes that $y > \varepsilon$. Smaller values of y permit $\vec{y} = \vec{0}$ and all weights can be set to zero.

where the first inequality is true because $\widetilde{e}_{s*}^2 \leqslant e_{s*}^2$, and the second inequality because we have dropped positive $\mathcal{O}(\delta^2)$ terms. Then we can obtain an upper bound on y-critical by finding a y such that even the last expression on the RHS is greater than W^2. Specifically,

$$(\widehat{e} \cdot \vec{y})^2 + 2(\widehat{e} \cdot \vec{y})(\widehat{e} \cdot \vec{\delta}) + (e_{s*}^2 + e_u^2)(y_0^2 + 2\,\vec{y} \cdot \vec{\delta})$$
$$> W^2(e_{s*}^2 + e_u^2). \tag{C6}$$

So, let us try to find the $\vec{\delta}$ that minimizes the LHS:

$$\text{LHS} = y^2 e_y^2 + 2y e_y(\widehat{e} \cdot \vec{\delta}) + (e_{s*}^2 + e_u^2)(y^2 + 2y\,\widehat{y} \cdot \vec{\delta})$$
$$= y^2(e_y^2 + e_{s*}^2 + e_u^2) + 2y(e_y\widehat{e} + (e_{s*}^2 + e_u^2)\widehat{y}) \cdot \vec{\delta}$$
$$= y^2(e_y^2 + e_{s*}^2 + e_u^2) + 2y\,\vec{\xi} \cdot \vec{\delta}, \tag{C7}$$

where $\vec{\xi} \equiv e_y \sum_{\mu=1}^{\mathcal{P}} e_\mu \vec{\varepsilon}_\mu + (e_{s*}^2 + e_u^2)\widehat{y}$, and we have noted that $\widehat{u} \cdot \vec{\delta} = 0$ because $\vec{\delta}$ must be in the activity-constrained subspace. It is now clear that LHS is minimized if $\vec{\delta}$ anti-aligns with $\vec{\xi}$. Then Eq. (C6) yields

$$y^2(e_y^2 + e_{s*}^2 + e_u^2) - 2y|\vec{\xi}|\varepsilon > W^2(e_{s*}^2 + e_u^2). \tag{C8}$$

Equating the two sides of Eq. (C8) and solving for y,[21] we now get an upper bound for y-critical:

$$y_{\text{cr,max,0}} \equiv \sqrt{W^2\left(\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}\right) + \frac{\varepsilon^2 \xi^2}{(e_y^2 + e_{s*}^2 + e_u^2)^2}}$$
$$+ \frac{\varepsilon \xi}{e_{s*}^2 + e_u^2 + e_y^2}, \tag{C9}$$

where ξ is the norm of $\vec{\xi}$ and can be simplified as

$$\xi^2 = \sum_{\mu=1}^{\mathcal{P}} \left[e_y e_\mu + (e_{s*}^2 + e_u^2)\widehat{y}_\mu\right]^2$$
$$= e_y^2 \sum_{\mu=1}^{\mathcal{P}} e_\mu^2 + (e_{s*}^2 + e_u^2)^2 \sum_{\mu=1}^{\mathcal{C}} \widehat{y}_\mu^2 + 2(e_{s*}^2 + e_u^2)e_y \sum_{\mu=1}^{\mathcal{C}} \widehat{y}_\mu e_\mu$$
$$= e_y^2 e_p^2 + (e_{s*}^2 + e_u^2)^2 + 2(e_{s*}^2 + e_u^2)e_y^2$$
$$= (e_y^2 + e_{s*}^2 + e_u^2)^2 + e_y^2(e_p^2 - e_y^2). \tag{C10}$$

Thus, we have

$$y_{\text{cr,max,0}}$$

$$\equiv \sqrt{W^2\left(\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}\right) + \varepsilon^2\left(1 + \frac{e_y^2(e_p^2 - e_y^2)}{(e_y^2 + e_{s*}^2 + e_u^2)^2}\right)}$$

$$+ \varepsilon\sqrt{1 + \frac{e_y^2(e_p^2 - e_y^2)}{(e_y^2 + e_{s*}^2 + e_u^2)^2}}. \tag{C11}$$

As with lower bound, we will see that to obtain the correct upper bound in presence of topological transitions, one has to maximize over several upper bounds. Hence, we refer the above upper bound that does not include any effects from topological transitions with an index "0."

Finally, we would like to point out that for small errors one can also obtain an approximate correction to y-critical that lies in between $y_{\text{cr,min,0}}$ and $y_{\text{cr,max,0}}$. To obtain this estimate, let us first write down the bound on y that one would obtain from Eq. (C3) as $\delta \to 0$:

$$y > W\sqrt{\frac{\widetilde{e}_{s*}^2 + e_u^2}{\widetilde{e}_y^2 + \widetilde{e}_{s*}^2 + e_u^2}}. \tag{C12}$$

If $\vec{\delta}$ has components along any semiconstrained direction that contributes toward the original $\widehat{e}_{s*}$ vector, then $\widetilde{e}_{s*}^2 < e_{s*}^2$, and comparing Eqs. (A48) and (C12) we see that, as $\delta \to 0$, the bound on y will be less than the zero-error y_{cr}. In other words, for sufficiently small errors, if $\vec{\delta}$ explores directions that contribute to e_{s*}, then the corresponding bound on y is going to be smaller than even the zero-error y_{cr}. Thus, for these small errors the leading order corrections to Eq. (A48) is obtained only if $\vec{\delta}$ do not have any components along these semiconstrained directions. This means $\widetilde{e}_{s*}^2 = e_{s*}^2$, and we can reorder the indices such that the semiconstrained directions along which excursions of $\vec{\delta}$ will be considered range from $\mathcal{C} + 1, \ldots, \mathcal{Q}$; i.e., $\text{Sgn}(\widehat{s}'_\mu)$ is negative for these, and only these, semiconstrained indices. To obtain the certainty condition, one can then follow steps (C5)[22] through (C9) except that $\vec{\delta}$ is restricted to only have nonzero components along constrained directions and those semiconstrained directions that do not contribute to $\widehat{e}_{s*}$, i.e., for $\mu = 1 \ldots \mathcal{Q}$. In other words, it can at the most anti-align with a truncated $\vec{\xi}$,

$$\vec{\xi}_{\text{trunc}} \equiv \sum_{\mu=1}^{\mathcal{Q}} \xi_\mu \vec{\varepsilon}_\mu = \sum_{\mu=1}^{\mathcal{Q}} \left[e_y e_\mu + (e_{s*}^2 + e_u^2)\widehat{y}_\mu\right]\vec{\varepsilon}_\mu, \quad \text{and}$$

$$(\vec{\xi} \cdot \vec{\delta})_{\text{min}} = -\varepsilon \xi_{\text{trunc}}, \tag{C13}$$

where ξ_{trunc} is the norm of $\vec{\xi}_{\text{trunc}}$ and can be simplified as

$$\xi_{\text{trunc}}^2 = \sum_{\mu=1}^{\mathcal{Q}} \left[e_y e_\mu + (e_{s*}^2 + e_u^2)\widehat{y}_\mu\right]^2$$

$$= e_y^2 \sum_{\mu=1}^{\mathcal{Q}} e_\mu^2 + (e_{s*}^2 + e_u^2)^2 \sum_{\mu=1}^{\mathcal{C}} \widehat{y}_\mu^2$$

$$+ 2(e_{s*}^2 + e_u^2)e_y \sum_{\mu=1}^{\mathcal{C}} \widehat{y}_\mu e_\mu$$

$$= e_y^2(e_p^2 - e_{s*}^2) + (e_{s*}^2 + e_u^2)^2 + 2(e_{s*}^2 + e_u^2)e_y^2$$

$$= (e_y^2 + e_{s*}^2 + e_u^2)^2 + e_y^2(e_p^2 - e_{s*}^2 - e_y^2). \tag{C14}$$

---

[21]This quadratic equation obviously has two solutions. The correct one can easily be identified, for instance, by taking the $\varepsilon \to 0$ limit.

[22]Since we are only interested in the leading order correction, we could also drop the $\mathcal{O}(\delta^2)$ terms needed to arrive at an expression such as the RHS of Eq. (C5).
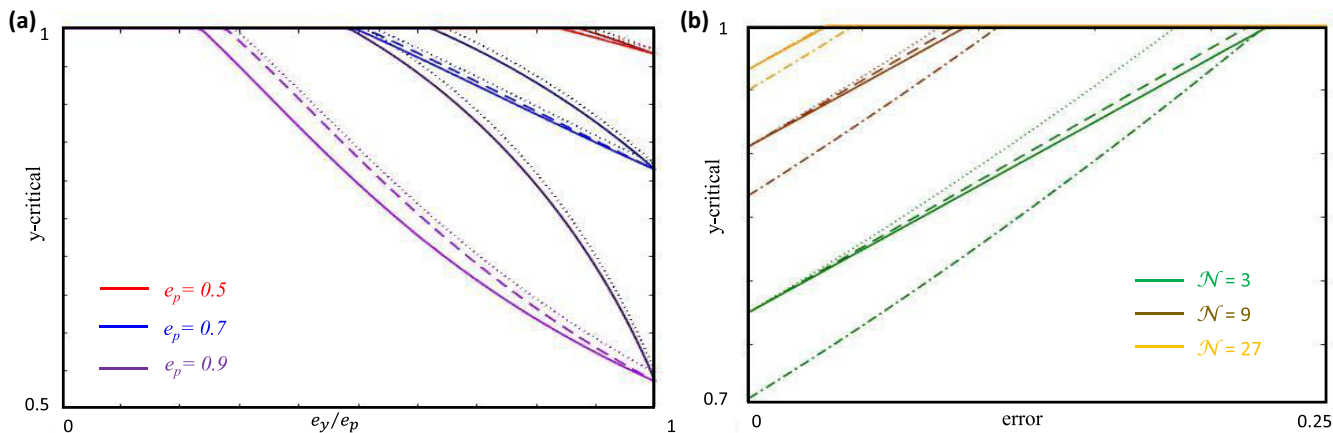
FIG. 10. Dependence of y-critical on various parameters for nonzero errors. (a) The red, blue, and purple curves track y-critical as a function of e_y/e_p for $e_p = 0.5, 0.7$, and 0.9, respectively. The dotted, dashed and bold curves represent the lower bound, leading-order and upper bound y-critical curves for a fixed error, $\varepsilon = 0.1W$. The darker shade correspond to the most nonlinear case when $e_{s*}/\sqrt{e_p^2 - e_y^2} = 1$, while the lighter shade correspond to $e_{s*} = 0$. These latter curves are also the ones that one obtains in a linear theory. Clearly, the difference between the linear and nonlinear theory increases as e_p increases. In all these cases y-critical decreases with increase of e_p, and for a given e_p, as e_y/e_p increases. Also, as e_{s*} increases and the semiconstrained dimensions become more important, it becomes harder to constrain the synapse sign, and therefore y-critical increases. (b) The green, brown, and orange curves again track y-critical, but this time as a function of ε, for networks with $\mathcal{N} = 3, 9$, and 27 input neurons, respectively. The dotted, dashed and bold curves plot the lower bound, leading-order and upper bound on y-critical for typical values of e_p, e_y, and e_{s*} that one expects in these networks (B1). We see that these curves come closer together as the network size increases. The dot-dashed curves correspond to the linear theory ($e_{s*} = 0$), which remains clearly separated from the nonlinear curves. In each of these networks, $\mathcal{P}/\mathcal{N} = 2/3$ and $\mathcal{C}/\mathcal{P} = 1/2$.

Substituting $\xi = \xi_{\text{trunc}}$ into the counterpart of Eq. (C9), and keeping only the linear terms in ε, we thus get the leading order correction to Eq. (A48):

$$
\begin{aligned}
y_{\text{cr,appr,0}} &\approx W\sqrt{\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}} + \frac{\varepsilon\xi_{\text{trunc}}}{e_y^2 + e_{s*}^2 + e_u^2} \\
&= W\sqrt{\frac{e_{s*}^2 + e_u^2}{e_y^2 + e_{s*}^2 + e_u^2}} + \varepsilon\sqrt{1 + \frac{e_y^2\left(e_p^2 - e_{s*}^2 - e_y^2\right)}{\left(e_y^2 + e_{s*}^2 + e_u^2\right)^2}}.
\end{aligned}
$$

$$(C15)$$

We will see later how $y_{\text{cr,appr,0}}$ can be generalized to provide an approximation, $y_{\text{cr,appr}}$, to y-critical that accounts for topological transitions.

Reassuringly, we see that at $\varepsilon = 0$, $y_{\text{cr,appr,0}}$, $y_{\text{cr,max,0}}$, and $y_{\text{cr,min,0}}$, all reduce to the zero-error y_{cr} Eq. (A48). Also it is obvious that the coefficient of ε in $y_{\text{cr,appr,0}}$ is greater than that of $y_{\text{cr,min,0}}$ but less than that of $y_{\text{cr,max,0}}$. Finally, note that $y_{\text{cr,appr,0}}$ coincides with $y_{\text{cr,min}}$ in the maximally nonlinear case where $e_{s*}^2 = e_p^2 - e_y^2$. In Fig. 10(a), we have plotted how these different quantities depend on e_p, e_y and e_{s*}. In particular we note that as the network size increases, these curves typically come closer together [Fig. 10(b)], so that they provide a good approximation for y-critical. Finally, for future reference we point out that for a given set of input patterns, $z_{\mu m}$, the various y-criticals that we have computed above depend on the orientation of the target response vector, or $\widehat{y}$, and the total noise budget, ε. In other words, $y_{\text{cr,min,0}} = y_{\text{cr,min,0}}(\widehat{y}, \varepsilon)$, $y_{\text{cr,max,0}} = y_{\text{cr,max,0}}(\widehat{y}, \varepsilon)$, and $y_{\text{cr,appr,0}} = y_{\text{cr,appr,0}}(\widehat{y}, \varepsilon)$.

b. Comparing predictions from linear and nonlinear models

To assess the effects of nonlinearity it is useful to compare the predictions for certain-synapses between the linear and nonlinear theory. In a linear theory, there are no semiconstrained directions, and therefore, a lower bound, leading order and upper bound on y-critical can be obtained from Eqs. (C4), (C15), and (C11), respectively, by setting $e_{s*} = 0$:

$$
y_{\text{cr,min,lin}} = W\sqrt{\frac{e_u^2}{e_y^2 + e_u^2}} + \varepsilon, \tag{C16}
$$

$$
y_{\text{cr,appr,lin}} = W\sqrt{\frac{e_u^2}{e_y^2 + e_u^2}} + \varepsilon\sqrt{1 + \frac{e_y^2\left(e_p^2 - e_y^2\right)}{\left(e_y^2 + e_u^2\right)^2}}, \tag{C17}
$$

$$
\begin{aligned}
y_{\text{cr,max,lin}} &= \sqrt{W^2\left(\frac{e_u^2}{e_y^2 + e_u^2}\right) + \varepsilon^2\left(1 + \frac{e_y^2\left(e_p^2 - e_y^2\right)}{\left(e_y^2 + e_u^2\right)^2}\right)} \\
&\quad + \varepsilon\sqrt{1 + \frac{e_y^2\left(e_p^2 - e_y^2\right)}{\left(e_y^2 + e_u^2\right)^2}}.
\end{aligned}
\tag{C18}
$$

We note that since all these quantities are increasing function of e_{s*}, the linear values are always less than or equal to the nonlinear counterparts. Since no topological transitions are possible in a linear theory, these expressions do not need a qualifying "0" index. In Fig. 10, we show a comparison between the upper bound on y-critical obtained in the linear and the nonlinear theories.

c. $y_{cr,max}$, when one and only one nonzero response is smaller than noise

In the previous section we have considered responses which are either zero, or positive and greater than the noise bound, ε. In this section, we consider a situation where one and only one of the observed responses is smaller than the noise, $|y_1| < \varepsilon$.

Note that once we admit noise, it is possible for the small observed response to be negative. In this case, there is no zero-error solution as $\Phi(\eta_1)$ cannot be negative, and therefore we need a minimum noise, and incur a minimum error:

$$\delta_1 = -y_1 \Rightarrow \mathcal{E}_{min} = y_1^2. \tag{C19}$$

In fact, since the noise for this observation has to be positive, we must have

$$\Phi(\eta_1) = y_1 + \delta_1 \equiv \delta_1' \geqslant 0. \tag{C20}$$

Then using

$$\delta_1^2 = \delta_1'^2 + y_1^2 - 2y_1\delta_1', \tag{C21}$$

we obtain a modified bound on the noise:

$$\sum_{\mu=1}^{\mathcal{P}} \delta_\mu^2 = \sum_{\mu=2}^{\mathcal{P}} \delta_\mu^2 + \delta_1'^2 + y_1^2 - 2y_1\delta_1' \leqslant \varepsilon^2$$

$$\Rightarrow \delta_1'^2 + \sum_{\mu=2}^{\mathcal{P}} \delta_\mu^2 \leqslant \varepsilon^2 - y_1^2 + 2y_1\delta_1' \leqslant \varepsilon^2 - y_1^2, \tag{C22}$$

since $y_1\delta_1' < 0$. Or,

$$\delta_1'^2 + \sum_{\mu=1}^{\mathcal{P}} \delta_\mu^2 \leqslant \varepsilon'^2 \equiv \varepsilon^2 - y_1^2. \tag{C23}$$

Let us now introduce a new reduced response vector whose response to the first pattern is set to zero:

$$\vec{y}\,' \equiv \sum_{\mu=2}^{\mathcal{C}} y_\mu \vec{\mathcal{E}}_\mu. \tag{C24}$$

We can then identify δ_1' to be the noise associated with the $\mu = 1$ response in this new feedforward problem, while the other δ_μ's can continue to represent the noise associated with all the other responses. Thus, a sufficient condition for a given synapse to be certain is

$$|\vec{y}\,'| > y_{cr,max,0}(\widehat{y}', \varepsilon'). \tag{C25}$$

A very similar condition arises if y_1 is positive but small enough to admit a topological transition. To see how, let us first remember that in order for a synapse to be certain, the solution space should not intersect with the $w = 0$ hyperplane. Now, let us look at the solution space coming from denoised $\tilde{\tilde{y}}$'s that have $\tilde{y}_1 > 0$. Since, the solution space corresponding to these $\tilde{\tilde{y}}$'s do not have any additional semiconstrained dimension as compared to the observed response, $\vec{y}$, the condition for no intersection with this part of the solution space is simply given by

$$|\vec{y}| > y_{cr,max,0}(\widehat{y}, \varepsilon), \tag{C26}$$

a condition that guarantees a certain synapse when no topological transitions are considered. Next consider the solution space for denoised $\tilde{\tilde{y}}$'s with $\tilde{y}_1 = 0$. The solution space for these $\tilde{\tilde{y}}$'s have an additional semiconstrained dimension corresponding to the first pattern. We can therefore use the reduced response vector, $\vec{y}\,'$ (C24), so that the solution space corresponding this new response vector with the error bound, ε' (C23), along with the solution space with $\tilde{y}_1 > 0$ accounts for the full solution space of $\vec{y}$ with error ε. Note, that the noise budget is again reduced according to Eq. (C23) since we are committed to making at least an error of y_1 to convert the first response to a semiconstrained dimension. To ensure that there is no intersection of the $w = 0$ hyperplane with the $\vec{y}\,'$ solution space we must therefore also satisfy Eq. (C25). We note that to calculate the right-hand side using Eq. (C11), the various projections have to be recalculated according to

$$e_{y'} = \frac{\sum_{\mu=2}^{\mathcal{C}} y_\mu e_\mu}{\sqrt{\sum_{\mu=2}^{\mathcal{C}} y_\mu^2}} \quad \text{and}$$

$$e_{s'*} = \sqrt{\sum_{\mu \in A_-} e_\mu^2 + \Theta(\text{Sgn}(e_{y'})e_1)e_1^2}. \tag{C27}$$

The condition (C25) on $|\vec{y}\,'|$ translates to a condition on $|\vec{y}|$:

$$|\vec{y}| > \sqrt{y_{cr,max,0}^2(\widehat{y}', \varepsilon') + y_1^2} \Rightarrow |\vec{y}| > \frac{y_{cr,max,0}(\widehat{y}', \varepsilon')}{\sqrt{1 - \widehat{y}_1^2}}, \tag{C28}$$

where we have defined $\widehat{y}_\mu$'s to be the μth component of $\widehat{y}$.

We note that this is a nonlinear inequality as the right-hand side depends on $|\vec{y}|$ through its implicit dependence on ε'. When we have a negative y_1, only Eq. (C28) needs to be satisfied to guarantee a certain synapse, but if y_1 is positive, both Eqs. (C26) and (C28) have to be satisfied. It is not hard to see how this process should be continued if one has more than one topological transition within the allowed error. Since we know the precise sequence of topological transitions, all the sequential certainty-conditions can in principle be obtained. A synapse is certain if all of its certainty-conditions are satisfied.

So far, we have obtained a way to check whether a synapse is certain given the response data, $\vec{y}$. We also have an upper bound of y-critical, $y_{cr,max,0}$, ignoring effects from topological transitions when all the observed responses are nonnegative. We will now investigate how topological transitions can change this upper bound. We will start by quantifying effects from a single topological transition by finding potentially a new upper bound for y-critical, $y_{cr,max}$, such that we can say that if $|\vec{y}| > y_{cr,max}$, then the synapse is certain. Suppose we start out with a data vector whose norm is so large that there are no topological transitions. Then as we decrease the norm, but keep its orientation, $\widehat{y}$, fixed, eventually a semiconstrained dimension will open up in the solution space, in our example, the first direction. If we keep decreasing further, then at some point another response dimension will become semiconstrained due to the presence of noise. Let us however consider the situation where $y_{cr,max}$ (that is yet to be computed) is going to turn out to be larger than the norm when the second transition occurs. In this case, we do not have to consider this possibility (and any other transitions) because

then if $|\vec{y}| > y_{\text{cr,max}}$ the second transition cannot occur. We will later find a condition that guarantees this. Since we are trying to find the smallest value of $y_{\text{cr,max}}$ that we can find, what all this means is that at $\vec{y}_{\text{cr,max}} = y_{\text{cr,max}}\widehat{y}$, one of the two inequalities, (C26) or (C25), becomes an equality. While the first equality is trivial to solve as the right-hand side does not depend on $y_{\text{cr,max}}$, the second equation is highly nonlinear[23]:

$$y_{\text{cr,max,1}} = \frac{y_{\text{cr,max,0}}(\widehat{y}', \varepsilon')}{\sqrt{1 - \widehat{y}_1^2}}, \tag{C29}$$

where

$$\varepsilon' = \sqrt{\varepsilon^2 - y_{\text{cr,max,1}}^2 \widehat{y}_1^2}. \tag{C30}$$

In particular, we notice that there are two competing effects that ultimately determine $y_{\text{cr,max,1}}$. The numerator depends on ε', which decreases as $y_{\text{cr,max,1}}$ increases and therefore has an overall effect of decreasing $y_{\text{cr,max,1}}$. However, the presence of $\widehat{y}_1$ in the denominator within the square root tends to increase $y_{\text{cr,max,1}}$. To determine the correct upper bound for y-critical one has to compare the $y_{\text{cr,max,1}}(\widehat{y}', \widehat{y}_1, \varepsilon')$ obtained from Eq. (C29) with $y_{\text{cr,max,0}}(\widehat{y}, \varepsilon)$, and then choose the maximum because then both the inequalities (C26) and (C25) will be satisfied. For the negative response case, we simply need to solve Eq. (C29) to obtain $y_{\text{cr,max,1}}$.

Now, determining $y_{\text{cr,max,1}}$ from Eq. (C28) involves solving a quartic equation leading to expressions that are not particularly insightful. However, we can obtain a relatively simple conservative estimate bypassing the nonlinearity if we have a lower bound on y-critical, $y_{\text{cr,min}}$ because we can use this bound to overestimate ε':

$$y_{\text{cr,max}} \equiv \max \left\{ \frac{y_{\text{cr,max,0}}(\widehat{y}', \varepsilon'')}{\sqrt{1 - \widehat{y}_1^2}}, y_{\text{cr,max,0}}(\widehat{y}, \varepsilon) \right\} \quad \text{where}$$

$$\varepsilon'' \equiv \sqrt{\varepsilon^2 - y_{\text{cr,min}}^2 \widehat{y}_1^2} \geqslant \varepsilon'. \tag{C31}$$

Before we describe how we can obtain $y_{\text{cr,min}}$, let us note that if

$$y_{\text{cr,min}} \widehat{y}_\mu > \varepsilon, \quad \forall \, \mu > 1, \tag{C32}$$

then the second transition occurs at a magnitude that is lower than $y_{\text{cr,max,1}}$, and therefore does not need to be incorporated in the $y_{\text{cr,max,1}}$ calculation. Indeed, Eq. (C32) is a sufficient condition but not a necessary one.

d. $y_{\text{cr,min}}$, when one and only one nonzero response is smaller than noise

When we have a small response, $|y_1| < \varepsilon$, we have seen that we have to consider solution space around a reduced response vector, $\vec{y}'$ (C24), with a smaller error budget, ε' (C23). Accordingly, we can obtain an equation for a lower

bound on y-critical using Eq. (C4),[24]

$$y_{\text{cr,min,1}} = y_{\text{cr}}(\widehat{y}') + \varepsilon', \tag{C33}$$

where $\mu = 1$ along with $\mu = \mathcal{C} + 1 \dots \mathcal{P}$ are all treated as semiconstrained. As before, since the norms along $\vec{y}$ and $\vec{y}'$ are related via

$$|\vec{y}|\sqrt{1 - \widehat{y}_1} = |\vec{y}'|, \tag{C34}$$

we get an equation for $y_{\text{cr,min,1}}$ very similar to Eq. (C29) for $y_{\text{cr,max,1}}$:

$$y_{\text{cr,min,1}} = \frac{y_{\text{cr}}(\widehat{y}') + \varepsilon''}{\sqrt{1 - \widehat{y}_1^2}}, \quad \text{where } \varepsilon'' = \sqrt{\varepsilon^2 - y_{\text{cr,min,1}}^2 \widehat{y}_1^2}. \tag{C35}$$

Although nonlinear, the above equation reduces to a quadratic equation for $y_{\text{cr,min,1}}$,

$$y_{\text{cr,min,1}}^2 - 2y_{\text{cr,min,1}} y_{\text{cr}} \sqrt{1 - \widehat{y}_1^2} + y_{\text{cr}}^2 - \varepsilon^2 = 0, \tag{C36}$$

solving which we get[25]

$$y_{\text{cr,min,1}} = y_{\text{cr}}(\widehat{y}')\sqrt{1 - \widehat{y}_1^2} + \sqrt{\varepsilon^2 - y_{\text{cr}}^2(\widehat{y}')\widehat{y}_1^2}. \tag{C37}$$

For positive y_1 the above expression provides another lower bound along with the one obtained without the transition (C4). To ensure we have the tightest possible lower bound we thus maximize:

$$y_{\text{cr,min}} \equiv \max\{y_{\text{cr,min,0}}, y_{\text{cr,min,1}}\}. \tag{C38}$$

e. When more than one nonzero responses are smaller than allowed error

It is not difficult to see how the arguments above generalize if we have more than one small ($< \varepsilon$) observed response. We have to consider cases where all the negative observed responses, and different possible combinations of the positive responses, are set to zero. Let us denote T to be one such possible set of μ indices. As before, we define a reduced response vector, which is now indexed by T:

$$\vec{y}_T \equiv \sum_{\mu \notin T} y_\mu \vec{\varepsilon}_\mu, \tag{C39}$$

so $y_{T,\mu} = 0$ for all $\mu \in T$. Then, essentially following the same algebraic manipulations as above we obtain a lower bound according to

$$y_{\text{cr,min},T} = y_{\text{cr}}(\widehat{y}_T)\sqrt{1 - \sum_{\mu \in T} \widehat{y}_\mu^2} + \sqrt{\varepsilon^2 - y_{\text{cr}}^2(\widehat{y}_T)\sum_{\mu \in T} \widehat{y}_\mu^2}. \tag{C40}$$

To reiterate, whenever an observed response is negative, which is inconsistent with a threshold linear transfer function,

---

[23]Here the "1" in the subscript indicates that this possibility for y-critical is computed by only considering the first topological transition.

[24]To remind the readers, the expression for $y_{\text{cr,min,0}}$ was obtained by computing y_{cr} for $\widetilde{\vec{y}} = (1 - \varepsilon)\vec{y}$, a denoised point that is allowed because of the noise. In this case, the corresponding point is $\widetilde{\vec{y}} = (1 - \varepsilon')\vec{y}'$.

[25]The second root gives a negative result, and accordingly does not reduce to the correct $\varepsilon \to 0$ limit.
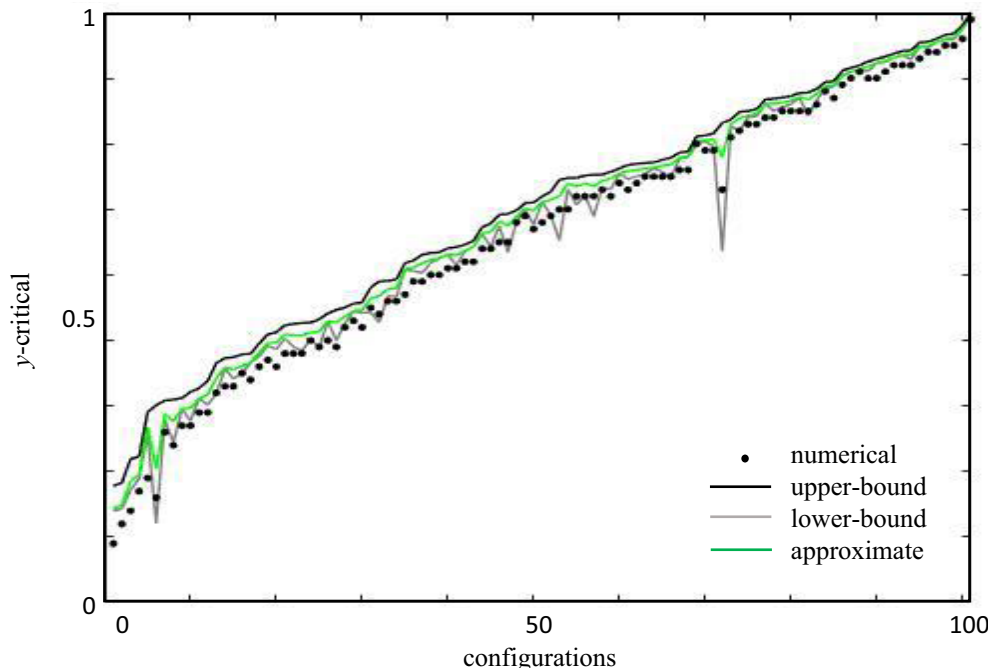
FIG. 11. Testing bounds on y-critical for solutions with error. We show the same 102 random configurations of input-output activity as Fig. 8(b). The bold black, green, and gray curves represent the upper bound $y_{\mathrm{cr,max}}$, approximate $y_{\mathrm{cr,appr}}$, and lower bound y-critical $y_{\mathrm{cr,min}}$, values, respectively. The black dots correspond to the maximum value of y in our simulations that resulted in mixed signs for the synaptic weights under consideration.

this means that some of the noise budget has to be used up to bring this response up to zero, and the same noise reduction occurs if one wants to consider topological transitions. Each $y_{\mathrm{cr,min},T}$ evaluated this way provides us with a lower bound, and hence we have to take a maximum over all these to find the tightest lower bound, $y_{\mathrm{cr,min}}$. To make things explicit, let us also enumerate the new expressions for the various projections of $\hat{e}$ that one needs to calculate $y_{\mathrm{cr}}(\widehat{y}_T)$:

$$e_{y,T} = \left(\sum_{\mu \notin T} y_\mu e_\mu\right) \bigg/ \sqrt{\sum_{\mu \notin T} y_\mu^2} \quad \text{and}$$

$$e_{s*,T} = \sqrt{\sum_{\mu \in T} \Theta(e_\mu e_{y,T}) e_\mu^2}. \tag{C41}$$

Once we have a lower bound, we can obtain conservative upper bounds analogous to Eq. (C31) for each T:

$$y_{\mathrm{cr,max},T} \equiv \frac{y_{\mathrm{cr,max},0}(\widehat{y}_T, \varepsilon_T)}{\sqrt{1 - \sum_{\mu \in T} \widehat{y}_\mu^2}}, \quad \text{where}$$

$$\varepsilon_T \equiv \sqrt{\varepsilon^2 - y_{\mathrm{cr,min}}^2 \sum_{\mu \in T} \widehat{y}_\mu^2}. \tag{C42}$$

As before, for a consistent upper bound for y-critical, we need to take the maximum over all $y_{\mathrm{cr,max},T}$'s.

Finally, we note that we do not need to consider all possible transitions. While going through the sequence of transitions, as soon as we find a T such that $y_{\mathrm{cr,min},T}\widehat{y}_\mu > \varepsilon$ for all $\mu \notin T$ we can stop as this means that by the time $|\vec{y}|$ is small enough

that any additional y_μ's can be set to zero, the synapse is already uncertain.

f. Numerical simulation

To illustrate the behavior of the various y-critical functions and check their utility, in Fig. 11 we have plotted $y_{\mathrm{cr,min}}$ (light gray curve) and $y_{\mathrm{cr,max}}$ (black curve) for the same 102 configurations as the ones depicted in Fig. 8(b) involving a feedforward simulation with $\mathcal{N} = 6$, $\mathcal{P} = 5$, $\mathcal{C} = 2$, and $\mathcal{E} < \varepsilon = 0.1$. We also defined, $y_{\mathrm{cr,appr}}$, as a maximum over different approximations, $y_{\mathrm{cr,appr},T}$'s, that incorporate topological transitions and are defined as natural generalizations of Eq. (C15):

$$y_{\mathrm{cr,appr,T}} = W\sqrt{\frac{e_{s*,T}^2 + e_u^2}{e_{y,T}^2 + e_{s*,T}^2 + e_u^2}}$$
$$+ \varepsilon_T\sqrt{1 + \frac{e_{y,T}^2\left(e_p^2 - e_{s*,T}^2 - e_{y,T}^2\right)}{\left(e_{y,T}^2 + e_{s*,T}^2 + e_u^2\right)^2}}. \tag{C43}$$

We have plotted $y_{\mathrm{cr,appr}}$, the approximation of y-critical, in green in Fig. 11. As in Fig. 8(b), the black dots here denote the maximum value of y in our simulations that still admitted mixed signs for the synapse under consideration, for details on the simulations, please see Appendix F. As one can see, most of the black dots seem to closely track the $y_{\mathrm{cr,min}}$ curve, but some of the dots lie between the $y_{\mathrm{cr,appr}}$ and $y_{\mathrm{cr,min}}$ curves.

2. New sources of corrections in recurrent neural networks

It is clear that recurrent neural networks inherit error corrections to y-critical that were already present in the feedforward case. There are two additional sources of error that one could consider as one moves from feedforward to recurrent networks. However, our numerical simulations of recurrent networks suggest that these are sometimes small effects, and we leave their systematic study for the future.

First, we could account for the fact that the $\vec{\varepsilon}_\mu$ directions themselves can change. This is because the inputs driving any given driven neuron can no longer be assumed to be fixed at $z_{\mu m}$ if the other driven neurons suffer from noise. However, these activity patterns define the $\vec{\varepsilon}_\mu$ directions and η_μ coordinates. Allowing noise in input neurons would lead to similar corrections.

Second, the total error in Eq. (63) may be unevenly distributed across the driven neurons. If the total squared error summed over all responses and neurons is $\varepsilon_{\tt tot}^2$, then on average, the root-mean-square error associated with each driven neuron is $\varepsilon_{\tt tot}/\sqrt{\mathcal{D}}$. We can thus hope that a substitution of $\varepsilon = \varepsilon_{\tt tot}/\sqrt{\mathcal{D}}$ in the various y-critical formulas will provide a good approximation. However, it is also possible that a few neurons will incur most of the error (up to $\varepsilon_{\tt tot}$), potentially leading to violation of the certainty conditions computed from the root-mean-square error over neurons.

APPENDIX D: BEYOND THRESHOLD-LINEAR TRANSFER FUNCTIONS

So far, we have always modeled the firing rate as a threshold-linear function applied to the input drive. Here we will explain how our analyses of y-critical with noise also provide a formalism to analyze a much more general class of nonlinear transfer functions.

1. Bounded deviations from the threshold-linear function

Let us start by considering transfer functions with bounded differences from the threshold-linear function:

$$\Psi(x) = \Phi(x) + \Delta(x), \ \text{with} \ |\Delta(x)| < \Delta_0 \ \forall \, x. \quad \text{(D1)}$$

In this case, the fixed-point equations become

$$y_\mu = \Psi\left(\sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m\right) = \Phi\left(\sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m\right) + \delta_\mu, \quad \text{where}$$

$$\delta_\mu \equiv \Delta\left(\sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m\right). \quad \text{(D2)}$$

Since $|\Delta(x)|$ is bounded by Δ_0, we have a bound on the squared norm of $\vec{\delta}$:

$$|\vec{\delta}|^2 < \mathcal{P}\Delta_0^2. \quad \text{(D3)}$$

It is therefore clear that we can estimate y-critical for the Ψ nonlinearity with exactly the same formalism that we used to estimate y-critical for the threshold nonlinearity in the presence of noise. In particular, all the y-critical estimates [(64), (65), (C15)] are valid with the substitution, $\varepsilon = \sqrt{\mathcal{P}}\Delta_0$. Moreover, one can account for other sources of noise

(bounded by ε_0) by instead substituting

$$\varepsilon = \sqrt{\varepsilon_0^2 + \mathcal{P}\Delta_0^2} \quad \text{(D4)}$$

to obtain estimates and bounds on y-critical.

2. Bounded departures from any threshold-monotonic nonlinearity

Let us now consider transfer functions, $\Psi(x)$, that are close to a function, $\Xi(x)$, that monotonically increases above a threshold, x_T:

$$\Xi(x) = 0 \, , \text{if} \ x \leqslant x_T,$$
$$\Xi(x) > \Xi(y) > 0, \text{if} \ x > y > x_T, \ \text{and}$$
$$\Psi(x) = \Xi(x) + \Delta(x), \ \text{with} \ |\Delta(x)| < \Delta_0 \ \forall \, x. \quad \text{(D5)}$$

Accordingly, we find

$$y_\mu = \Psi\left(\sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m\right) = \Xi(\eta_\mu) + \Delta(\eta_\mu) \Rightarrow \Xi(\eta_\mu)$$
$$= y_\mu - \Delta(\eta_\mu). \quad \text{(D6)}$$

Since the monotonicity condition ensures that Ξ^{-1} is well defined above threshold, and $|\Delta(\eta_\mu)| < \Delta_0$, we then have the upper bound,

$$0 \leqslant \Xi(\eta_\mu) < y_\mu + \Delta_0 \Rightarrow \eta_\mu < \Xi^{-1}(y_\mu + \Delta_0). \quad \text{(D7)}$$

Additionally, if $y_\mu > \Delta_0$, then we also have a lower bound:

$$0 < y_\mu - \Delta_0 < \Xi(\eta_\mu) \Rightarrow \eta_\mu > \Xi^{-1}(y_\mu - \Delta_0). \quad \text{(D8)}$$

Thus, combining the upper and lower bounds, we find

$$\Xi^{-1}(y_\mu + \Delta_0) > \eta_\mu > \Xi^{-1}(y_\mu - \Delta_0) > 0 \quad \text{(D9)}$$

However, if $y_\mu \leqslant \Delta_0$, then there is no lower bound, and any η_μ satisfying the upper bound (D7) is allowed. Now, we can introduce effective responses, representing the midpoint of possible superthreshold input drives,

$$\bar{y}_\mu \equiv \begin{cases} \frac{1}{2}[\Xi^{-1}(y_\mu + \Delta_0) + \Xi^{-1}(y_\mu - \Delta_0)], & \text{if} \ y_\mu > \Delta_0, \\ \frac{1}{2}\Xi^{-1}(y_\mu + \Delta_0), & \text{if} \ y_\mu \leqslant \Delta_0, \end{cases} \quad \text{(D10)}$$

and effective noise limits,

$$\bar{\varepsilon}_\mu \equiv \begin{cases} \frac{1}{2}[\Xi^{-1}(y_\mu + \Delta_0) - \Xi^{-1}(y_\mu - \Delta_0)], & \text{if} \ y_\mu > \Delta_0, \\ \frac{1}{2}\Xi^{-1}(y_\mu + \Delta_0), & \text{if} \ y_\mu \leqslant \Delta_0, \end{cases} \quad \text{(D11)}$$

which allow η_μ to span the full allowed range. By inspection, we now see that the solution space is equivalent to the solution space of a threshold-linear problem:

$$\Phi(\eta_\mu) = \bar{y}_\mu + \bar{\delta}_\mu, \quad \text{with} \ |\bar{\delta}_\mu| \leqslant \bar{\varepsilon}_\mu. \quad \text{(D12)}$$

Thus, again all the y-critical estimates [(64), (65), (C15)] will be valid with the substitution $y_\mu \to \bar{y}_\mu$ and a conservative
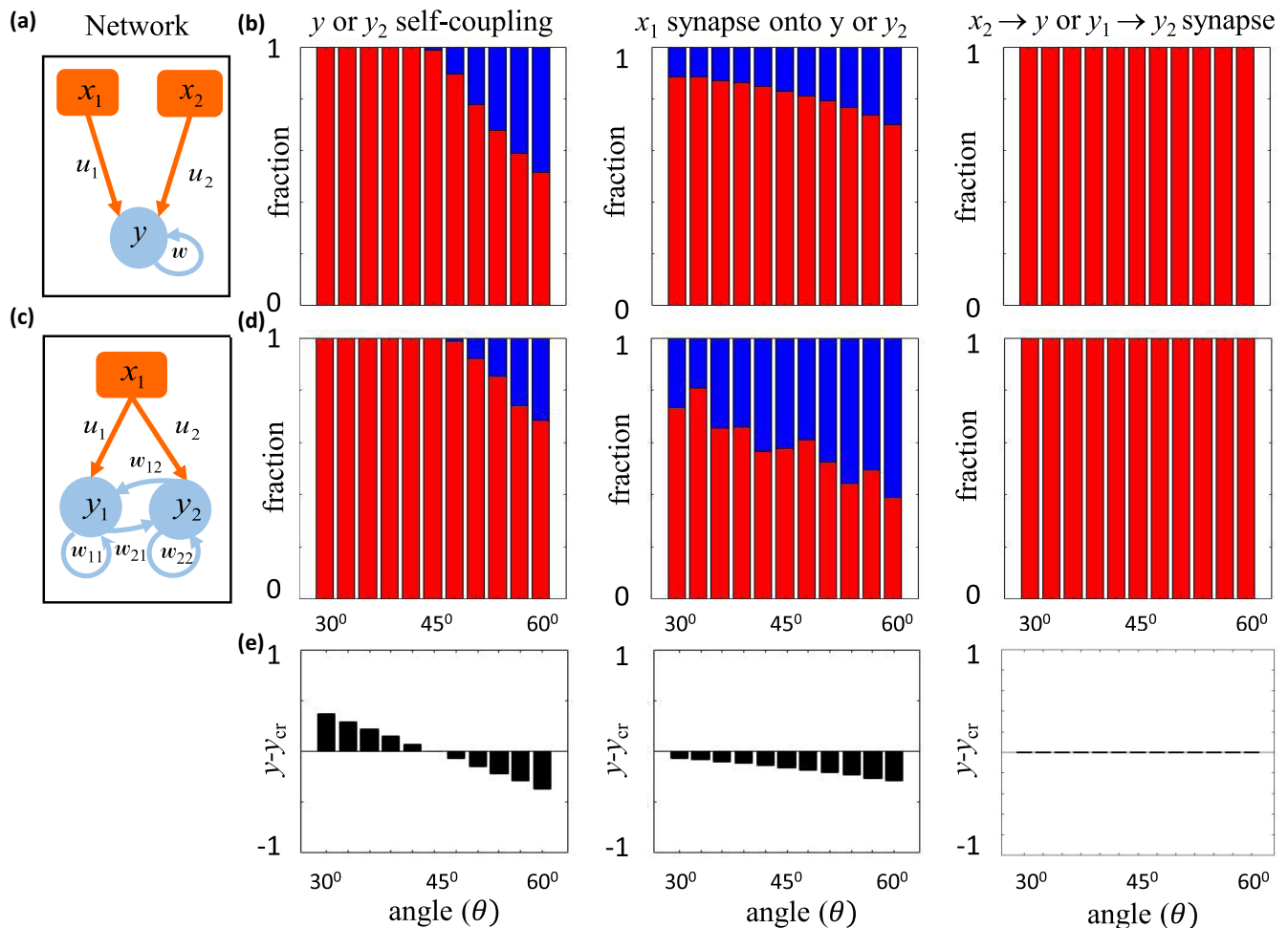
FIG. 12. Comparing simulation and theoretical results in $\mathcal{N} = 3$ recurrent network. (a) A simple $\mathcal{N} = 3$ recurrent neural network with one driven and two input neurons. Note that the y_1 neuron shown here maps onto the y_3 neuron in Fig. 6(a) by interpreting the x_1 and x_2 neurons shown here as the x_3 and y_2 neurons shown in Fig. 6(a). (b) Bar graphs depicting the fraction of positive (red) and negative (blue) weights from the network depicted in panel (a). (c) Another $\mathcal{N} = 3$ recurrent neural network, this time with two driven and one input neuron. (d) Bar graphs depicting the fraction of positive (red) and negative (blue) weights from the network depicted in panel (c). (e) The black bars depict $y - y_{\text{cr}}$ for the corresponding synapses.

error bound

$$\varepsilon^2 = \sum_{\mu=1}^{\mathcal{P}} \bar{\varepsilon}_\mu^2. \tag{D13}$$

APPENDIX E: CERTAIN SYNAPSES IN LOW-DIMENSIONAL RECURRENT NETWORKS WITH SELF-CONNECTIONS

As discussed in Appendix A, when one moves from feedforward to recurrent neural networks with self-synapses, the input patterns can no longer be considered independent from the target neuron responses. How then does one assess synapse certainty for driven neurons with self-synapses, such as y_3 in Fig. 6(a), y in Fig. 12(a), y_1 in Fig. 12(c), and y_2 in Fig. 12(c)?

We begin by concretely analyzing neuron y in Fig. 12(a), because this is the conceptually simplest example, and fundamentally the mathematical analyses are the same for the other examples. In particular, to test our formalism and analytical

results using this neuron, we performed low-dimensional simulations where the $\mathcal{N} \times \mathcal{N}$ extended input pattern matrix was

$$Z = \begin{pmatrix} -\sin\psi\cos\chi & \cos\psi\cos\chi & \sin\chi \\ \cos\psi & \sin\psi & 0 \\ \sin\psi\sin\chi & -\cos\psi\sin\chi & \cos\chi \end{pmatrix}, \tag{E1}$$

which is the same as X in Eq. (60), except that the role of x_3 is now played by y itself.[26] The third column of Eq. (E1) corresponds to the responses of the driven neuron, but it also provides a self-input. The two input neuron responses are given by the first two columns. Equation (E1) is meant to correspond to the case where $\mathcal{P} = 2$ and $\mathcal{C} = 1$, such that $\mu = 1, 2, 3$ correspond to the constrained, semiconstrained,

_____

[26]In the context of Fig. 6(a), y, x_1, and x_2 can be identified with y_3, y_2, and x_2, respectively.

and unconstrained response patterns, respectively. For the purpose of numerical testing, we assumed that $\chi \in (0°, 90°)$ and $\psi \in (-90°, 0°)$, as this range of angles ensures that driven neuron responses were nonnegative.[27] We set $W = 1$. See Appendix F for numerical simulation details.

This example problem has one self-coupling, w, and two feedforward couplings, u_1 and u_2. For this response structure, one can use Eq. (51) to calculate y_{cr} for each of these three couplings. We find

$$y_{cr,w} = W \cos \chi,$$

$$y_{cr,u_1} = W \sqrt{\cos^2 \psi + \cos^2 \chi \sin^2 \psi},$$

$$y_{cr,u_2} = W \sin \chi. \quad (E2)$$

We assess synapse certainty by checking whether these formulas for y_{cr} are smaller than the magnitude of $\vec{y}$,

$$y = \sin \chi. \quad (E3)$$

Since $W = 1$, we see that the self-coupling becomes certain only if

$$\sin \chi > \cos \chi, \quad \text{or,} \quad 0° < \theta \equiv \pi/2 - \chi < 45° \quad (E4)$$

[Fig. 12(b), left], where θ has been defined to be the angle between $\widehat{e}_{yy}$ and $\widehat{c} = \widehat{y}$,[28] consistent with the conventions of Eq. (37). Next, basic trigonometric manipulations tell us that the certainty condition can never be satisfied for u_1 [Fig. 12(b), middle]. Finally, we see that the condition for certainty is always *just* not satisfied for u_2. Here this implies that all solutions have $u_2 \geqslant 0$ [Fig. 12(b), right]. This is because it is a very special case where $\cos \gamma = e_{s*} = 0$ and $\widehat{y}_\perp$ in Eq. (A51) is aligned with $\widehat{u}$, so that both the inequalities in Eq. (A53) turn into equalities. In each panel of Fig. 12(b), we tracked the fraction of positive and negative synapse signs across the simulations, as we varied θ. In particular, we see that w had a unique sign as long as $\theta < 45°$, u_1 always had mixed signs, and u_2 was nonnegative.[29]

The same exact response matrix (E1) can also be used to consider neurons y_3 in Fig. 6(a) and y_2 in Fig. 12(c). The only difference is that the three columns respectively encode the responses of the second driven neuron, second input neuron, and third driven neuron in Fig. 6(a) and the responses of the single input neuron and two driven neurons, y_1 and y_2, in Fig. 12(c). This correspondence can be seen by comparing the numerical results in Figs. 12(b) and 12(d). We use this correspondence to avoid having to simulate the full network in Fig. 6(a), and the numerical results in Fig. 6(c) are the same as those in Fig. 12(b).

---

[27]This range also allowed us to ensure nonnegativity for the other example recurrent circuits in Figs. 6(a) and 12(c).

[28]Note that

$$\widehat{e}_{yy} = \sin \chi \, \vec{\varepsilon}_1 + \cos \chi \, \vec{\varepsilon}_3 = \cos \theta \, \vec{\varepsilon}_1 + \sin \theta \, \vec{\varepsilon}_3, \quad (E5)$$

where we have identified $\widehat{c}$ and $\widehat{u}$ with $\vec{\varepsilon}_1$ and $\vec{\varepsilon}_3$, respectively.

[29]As we explained before, $y_{cr,u_2} = y$ for all values of θ. Hence, the certainty condition is not satisfied because u_2 may be zero. The fact that u_2 can vanish is not discernible from our simulations because the weight magnitudes were generated randomly, and weights where $u_2 = 0$ comprise a zero measure set.

APPENDIX F: NUMERICAL METHODS

1. Low-dimensional numerical methods

Here we detail the numerical methods relevant for Figs. 6 and 12.

a. Feedforward analysis

To test the analytic dependence in Fig. 6(b), we wanted to simulate solutions without biasing ourselves by the particular search algorithm used to find solutions. Accordingly, to find solutions to the fixed point equations (A2) with very small error ($\mathcal{E} < \varepsilon = 0.01$) we performed a random screen where each weight was chosen randomly from a uniform distribution between -1 and $+1$. For feedforward circuits, given the synaptic weights, one can obtain the fixed point responses of the target neuron by direct substitution of the known input responses in Eq. (A2) and then comparing these simulated target responses with the known target responses. We varied ψ, χ in the response data (E1) systematically in steps of $6°$.[30] For the light and dark green curves, ψ was fixed at $45°$, and χ was varied between $(0°, 90°)$ and $(90°, 180°)$, respectively, while for the pink and purple curves χ was fixed at $45°$, $135°$, respectively, and ψ was varied between $(0°, 90°)$. Finally, for a given choice of ψ, χ, we systematically varied y between 0 and 1, in intervals of $\Delta y = 0.01$. For each value of ψ, χ, and y, we obtained $\sim \mathcal{O}(10^2 - 10^4)$ solutions[31] satisfying the error and the biological bound (A11) from five to ten million different trial weight vectors. We then identified the maximal value of y for which the solutions had both positive and negative w_1's. This simulation point should lie beneath the theoretical y_{cr} if no error is allowed. However, since the error is small but nonzero, occasionally the y-criticals determined from simulations did slightly exceed the theoretical value. Also, since we vary y by small amounts $\Delta y = 0.01$, we expected the simulated y-criticals to be discrete but close to the theoretical predictions, which is exactly what we found in Fig. 6(b).

b. Recurrent analysis

Because the recurrent network solution space separates into several feedforward solution spaces at zero error, we numerically treated the driven neurons one at a time. To find solutions for the recurrent neurons in Figs. 6(a), 12(a), and 12(c), we fixed χ, ψ and then performed screens with random weights, selected in the same manner as the feedforward simulations discussed earlier. For each set of weights, and for each $\mu = 1, 2$, we obtained the late time values of y by solving the time evolution equation [Eq. (6) with $\tau_i = 20$ ms] using Euler's method starting with initial conditions $y_i(0) = y_{\mu i}$, for $\mu = 1, 2$. We used a time step of $\Delta t = 0.2$ ms. The $\widetilde{y}_\mu$'s obtained from the simulation at late times, $t \sim 600$ ms, were

---

[30]Since here we were primarily interested in the zero-error result, we restricted ourselves to a range of ψ, χ where no topological transitions can occur due to the small but finite error we had to allow for numerical simulations.

[31]The number of solutions varied between 200 and 40 000 depending primarily on the value of y, the higher the value, typically the more difficult it was to find solutions.

then compared with y_μ to obtain $\mathcal{E}$. If the weights satisfied, $\mathcal{E} < 0.05\sqrt{\mathcal{D}}$ and the biological bound (A11), then we considered the weights as solutions and checked the sign of the synaptic weights. For every value of ψ, χ, we found at least 50 solutions[32] to test the certainty predictions.

2. High-dimensional numerical methods

Here we detail the numerical methods relevant for Figs. 8 and 11.

a. Generating random orthogonal matrices

In several simulations we had to generate orthogonal response matrices. This meant that we had to obtain $\mathcal{P}$ orthonormal $\mathcal{N}$-dimensional vectors. This was done by first generating an $\mathcal{N} \times \mathcal{N}$ matrix, G, where each of its entries was randomly selected from a uniform distribution between -1 and $+1$. We then antisymmetrized the matrix, $G \rightarrow (G - G^T)/2$, and a random orthogonal response matrix was then obtained via matrix exponentiation, $Z = e^G$, where the matrix exponential is defined by substituting the matrix G into the power series expansion of the exponential function and is distinct from the simple exponentiation of individual matrix elements. The first $\mathcal{P}$ rows of Z could then correspond to the $\mathcal{P}$ orthogonal patterns, and Z can be interpreted as the orthogonal extension of z as discussed before.

b. Generating random orthogonal matrices with nonnegativity constraints

In recurrent networks, all driven neurons must have nonnegative responses for all patterns. Accordingly, when the input response pattern includes responses of driven neurons, we follow a different procedure for generating the response matrix, which works as long as $\mathcal{I} \geqslant \mathcal{P} - 1$. We started by choosing a $(\mathcal{P} \times \mathcal{N})$-dimensional matrix, z, containing the responses of $\mathcal{D}$ driven neurons and $\mathcal{I} \geqslant \mathcal{P} - 1$ input neurons. The first columns of z corresponded to driven neuron responses, and the last columns to input neuron responses, such that $z = (y \quad x)$. We made sure that the responses of the driven neurons were all nonnegative, as the threshold nonlinearity dictates, by choosing them to lie randomly between 0 and 1. To mimic a sparse response pattern, we set driven responses to 0 with 50% probability. The feedforward inputs, however, were randomly selected between -1 and 1. We then orthogonalized the input responses to the target neuron as follows. We start by normalizing the $\nu = 1$ pattern:

$$z_{1m} \rightarrow \frac{z_{1m}}{\sqrt{\sum_{n=1}^{\mathcal{N}} z_{1n}^2}}. \tag{F1}$$

Then, for each row, $\nu = 2 \ldots \mathcal{P}$, in a sequential order we performed the following operations:

(1) We started by defining a $(\nu - 1)$-dimensional square matrix, x':

$$x'_{\mu m} \equiv x_{\mu m} \quad \text{for } \mu, m = 1 \ldots (\nu - 1). \tag{F2}$$

(2) We next changed the first $m = 1 \ldots \nu - 1$ elements of the νth row of x

$$x_{\nu m} \equiv -\sum_{\mu=1}^{\nu-1} x'^{-1}_{m\mu} \left(\sum_{i=1}^{\mathcal{D}} y_{\mu i} y_{\nu i} + \sum_{n=\nu}^{\mathcal{I}} x_{\mu n} x_{\nu n} \right), \tag{F3}$$

and thus z. The other elements of the νth row of z were left unchanged. In particular, none of the driven neuron responses changed during this step.

(3) Finally, we rescaled all the elements of the νth row of z for normalization:

$$z_{\nu m} \rightarrow \frac{z_{\nu m}}{\sqrt{\sum_{n=1}^{\mathcal{N}} z_{\nu n}^2}}. \tag{F4}$$

This algorithm essentially uses the responses of the input neurons to the νth stimulus to ensure that the full νth response pattern involving both the driven and input neurons is orthogonal to all $\mu \leqslant \nu - 1$ patterns.

c. Generating target responses and response directions

To generate $\mathcal{P}$ target responses with $\mathcal{S}$ null responses, we simply randomly selected numbers between 0 and 1 for the $\mathcal{C} = \mathcal{P} - \mathcal{S}$ nonzero responses.

In some simulations, we wanted to consider situations where one has to account for a single topological transition to compute y-critical. Accordingly, we tailored the responses as follows. First, we set one nonzero response of the target neuron to a small value, 0.1ε. $\widehat{y}$ was then obtained by dividing the response vector by its norm. We then only considered those $\widehat{y}$'s whose other entries were large enough to prevent additional topological transitions from affecting y-critical. This was done by: evaluating $y_{\mathrm{cr,min}}$, the theoretical lower bound for y-critical that includes the first topological transition (Appendix C); constructing $\vec{y} = y_{\mathrm{cr,min}}\hat{y}$, which approximates the activity vector right below y-critical; and ensuring that all the other entries of $\vec{y}$ were greater than the allowed error, which guarantees that no other constrained dimensions can become semiconstrained in between $y_{\mathrm{cr,min}}$ and the true y-critical. This way, typically one and only one constrained direction became semiconstrained when we allowed solutions with errors $\lesssim \varepsilon$.

d. Finding solutions using gradient descent learning in feedforward networks

In all the high-dimensional simulations, we had to find solutions to the fixed point equations (A2). Since scanning a high-dimensional synaptic weight space randomly is not numerically efficient, we applied gradient descent learning[33] to obtain solutions. For feedforward networks, this meant using

---

[32]The number of solutions varied between 10^4 and 10^5 for the $\mathcal{D} = 1$ simulation in Fig. 12(b) and between 56 and 110 for the $\mathcal{D} = 2$ simulation in Fig. 12(d). Note that for the $\mathcal{D} = 2$ simulation we have a six-dimensional weight space, which makes it a lot harder to find solutions through random scanning. Also, for this latter case we only checked that the biological constraint is satisfied by the incoming weights to y_1.

[33]Typically with learning rate ~ 0.01.

the loss function

$$\mathcal{E}^2 = \sum_{\mu=1}^{\mathcal{P}} (\widetilde{y}_\mu - y_\mu)^2, \quad \text{where } \widetilde{y}_\mu = \Phi\left(\sum_{m=1}^{\mathcal{N}} z_{\mu m} w_m\right). \quad \text{(F5)}$$

We performed gradient descent optimization until we reached the desired error bound, $\mathcal{E} < \varepsilon$. The initial weights were first chosen randomly from a uniform distribution between -1 and 1. The initial weight vector was then rescaled to have a norm between 0 and $W = 1$, chosen uniformly.

e. Finding solutions using gradient descent learning in recurrent networks

To find solutions for recurrent neural networks, we used the modified loss function,

$$\bar{\mathcal{E}}^2 \equiv \sum_{i=1}^{\mathcal{D}} \sum_{\mu=1}^{\mathcal{P}} (\widetilde{y}_{\mu i} - y_{\mu i})^2 \equiv \sum_{i=1}^{\mathcal{D}} \bar{\mathcal{E}}_i^2, \quad \text{where}$$

$$\widetilde{y}_{\mu i} = \Phi\left(\sum_{m=1}^{\mathcal{N}} z_{\mu m} w_{im}\right), \quad \text{(F6)}$$

to perform gradient descent, instead of Eq. (63). Since the responses of the driven neurons can vary for nonzero errors, the two loss functions, $\mathcal{E}$ and $\bar{\mathcal{E}}$, differ. However, it is numerically a lot quicker to obtain solutions via gradient descent with $\bar{\mathcal{E}}$ as compared to using back-propagation through time to consider the entire time evolution of the network. Thus, the strategy we adopted to find solutions with $\mathcal{E} \lesssim \varepsilon$ was to first find weights satisfying $\bar{\mathcal{E}} \lesssim \bar{\varepsilon} = \varepsilon/10$. Also, the gradient descent was done in two stages. In the first stage we minimized the error associated with each individual driven neuron, $\bar{\mathcal{E}}_i$ treating

it as a feedforward problem. Once each of these errors were less than $\bar{\varepsilon}$, we performed a second stage of gradient descent to minimize $\bar{\mathcal{E}}$ down to $\bar{\varepsilon}$. Next, we obtained the late time values of y_i's by solving the time evolution equations (6) with $\tau_i = 20$ ms using Euler's method with step time $\Delta t = 0.2$ ms for the weights obtained via gradient descent and starting with initial conditions $y_i(0) = y_{\mu i}, \forall \mu, i$. The $\widetilde{y}_{\mu i}$'s obtained at late times, $t \sim 600$ ms, this way were compared with $y_{\mu i}$ to obtain $\mathcal{E}$. Finally, we checked that the weights satisfied the biological bound[34] (A11).

f. Other minor simulation details

In Fig. 8(b), we show results from a simulation with $\mathcal{N} = 6$, $\mathcal{P} = 5$, $\mathcal{C} = 2$, and $\mathcal{E} < \varepsilon = 0.1$. The solutions were obtained using a gradient descent learning rate of 0.02. We varied the norm of the target response vector, $\vec{y} = y\widehat{y}$, systematically by $\Delta y = 0.01$ in a manner similar to the low-dimensional simulations.

In Fig. 8(c), we considered a single input-output configuration for a given value of $\mathcal{N}$ and $\mathcal{P}$, and we found a single solution with $\mathcal{E} < 0.001\sqrt{\mathcal{P}}$ using a gradient descent learning rate of 0.005.

In Figs. 8(d)–8(e), we show results of a simulation for a $\mathcal{N} = 10$, $\mathcal{D} = 4$, $\mathcal{I} = 7$, $\mathcal{P} = 8$, and $\mathcal{C} = 3$ network where the norm of $\vec{y}$ was fixed to 0.79, which approximates the median value of y for the given values of $\mathcal{N}, \mathcal{P},$ and $\mathcal{C}$ [Eq. (B5)]. Our solutions were obtained using a gradient descent learning rate of 0.004 and the overall error satisfied $0.017\sqrt{\mathcal{D}} \lesssim \mathcal{E} \lesssim 0.25\sqrt{\mathcal{D}}$.

---

[34]We only checked that the target weights satisfied the weight bound, as that is what matters for the certainty conditions. Since we initialized weights amongst the nontarget neurons to be between -7 and 7, it is likely that other components of the weight matrix were large.

[1] J. D. Watson and F. H. C. Crick, Molecular structure of nucleic acids, Nature (London) **171**, 737 (1953).

[2] R. Milo *et al.*, Network motifs: Simple building blocks of complex networks, Science **298**, 824 (2002).

[3] P. Hunter and P. Nielsen, A strategy for integrative computational physiology, Physiology **20**, 316 (2005).

[4] H. S. Seung, Reading the book of memory: Sparse sampling versus dense mapping of connectomes, Neuron **62**, 17 (2009).

[5] C. I. Bargmann and E. Marder, From the connectome to brain function, Nat. Methods **10**, 483 (2013).

[6] D. D. Bock *et al.*, Network anatomy and *in vivo* physiology of visual cortical neurons, Nature (London) **471**, 177 (2011).

[7] L. R. Varshney, B. L. Chen, E. Paniagua, D. H. Hall, and D. B. Chklovskii, Structural properties of the *Caenhabditis elegans* neuronal network, PLoS Comput. Biol. **7**, e1001066 (2011).

[8] M. B. Ahrens *et al.*, Brain-wide neuronal dynamics during motor adaptation in zebrafish, Nature (London) **485**, 471 (2012).

[9] T. Schrödel, R. Prevedel, K. Aumayr, M. Zimmer, and A. Vaziri, Brain-wide 3D imaging of neuronal activity in *Caenorhabditis elegans* with sculpted light, Nat. Methods **10**, 1013 (2013).

[10] T. Ohyama *et al.*, A multilevel multimodal circuit enhances action selection in Drosophila, Nature (London) **520**, 633 (2015).

[11] W. C. Lemon *et al.*, Whole-central nervous system functional imaging in larval Drosophila, Nat. Commun. **6**, 7924 (2015).

[12] E. A. Naumann *et al.*, From whole-brain data to functional circuit models: The zebrafish optomotor response, Cell **167**, 947 (2016).

[13] D. G. C. Hildebrand *et al.*, Whole-brain serial-section electron microscopy in larval zebrafish, Nature (London) **545**, 345 (2017).

[14] L. K. Scheffer *et al.*, A connectome and analysis of the adult *Drosophila* central brain, eLife **9**, e57443 (2020).

[15] T. Biswas, W. E. Bishop, and J. E. Fitzgerald, Theoretical principles for illuminating sensorimotor processing with brain-wide neuronal recordings, Curr. Opin. Neurobiol. **65**, 138 (2020).

[16] R. Ben-Yishai, R. L. Bar-Or, and H. Sompolinsky, Theory of orientation tuning in visual cortex, Proc. Natl. Acad. Sci. USA **92**, 3844 (1995).

[17] W. E. Skaggs, J. J. Knierem, H. S. Kudrimoti, and B. L. McNaughton, A model of the neural basis of the rat's sense of direction, Adv. Neural Inf. Process. Syst. **7**, 173 (1995).

[18] S. S. Kim, H. Rouault, S. Druckmann, and V. Jayaraman, Ring attractor dynamics in the Drosophila central brain, Science **356**, 849 (2017).

[19] D. B. Turner-Evans *et al.*, The neuroanatomical ultrastructure and function of a biological ring attractor, Neuron **108**, 145 (2020).

[20] J. S. Kim *et al.*, Space-time wiring specificity supports direction selectivity in the retina, Nature (London) **509**, 331 (2014).

[21] J. Kornfeld *et al.*, EM connectomics reveals axonal driven variation in a sequence-generating network, eLife **6**, e24364 (2017).

[22] A. A. Wanner and R. W. Friedrich, Whitening of odor representations by the wiring diagram of the olfactory bulb, Nat. Neurosci. **23**, 433 (2020).

[23] A. Vishwanathan *et al.*, Predicting modular functions and neural coding of behavior from a synaptic wiring diagram, 10.1101/2020.10.28.359620 (2020).

[24] E. Marder and A. L. Taylor, Multiple models to capture the variability in biological neurons and networks, Nat. Neurosci. **14**, 133 (2011).

[25] K. Friston, L. Harrison, and W. Penny, Dynamic causal modeling, NeuroImage **19**, 1273 (2003).

[26] E. Schneidman II, M. J. Berry, R. Segev, and W. Bialek, Weak pairwise correlations imply strongly correlated network states in a neural population, Nature (London) **440**, 1007 (2006).

[27] J. W. Pillow *et al.*, Spatio-temporal correlations and visual signalling in a complete neuronal population, Nature (London) **454**, 995 (2008).

[28] H. Huang and M. Ding, Linking functional connectivity and structural connectivity quantitatively: A comparison of methods, Brain Connect. **6**, 99 (2016).

[29] F. D. Tschopp, M. B. Reiser, and S. C. Turaga, A connectome based hexagonal lattice convolutional network model of the Drosophila visual system, arXiv:1806.04793.

[30] A. A. Zarin, B. Mark, A. Cardona, A. Litwin-Kumar, and C. Q. Doe, A multilayer circuit architecture for the generation of distinct locomotor behaviors in Drosophila, eLife **8**, e51781 (2019).

[31] A. Litwin-Kumar and S. C. Turaga, Constraining computational models using electron microscopy wiring diagrams, Curr. Opin. Neurobiol. **58**, 94 (2019).

[32] A. A. Prinz, D. Bucher, and E. Marder, Similar network activity from disparate circuit parameters, Nat. Neurosci. **7**, 1345 (2004).

[33] D. Fisher, I. Olasagasti, D. W. Tank, E. R. Aksay, and M. S. Goldman, A modeling framework for deriving the structural and functional architecture of a short-term memory microcircuit, Neuron **79**, 987 (2013).

[34] J. M. Goaillard, A. L. Taylor, D. J. Schulz, and E. Marder, Functional consequences of animal-to-animal variation in circuit parameters, Nat. Neurosci. **12**, 1424 (2009).

[35] P. Baldi and K. Hornik, Neural networks and principal component analysis: Learning from examples without local minima, Neural Networks **2**, 53 (1989).

[36] Y. N. Dauphin *et al.*, Identifying and attacking the saddle point problem in high-dimensional optimization, in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, 2014).

[37] K. Kawaguchi, Deep learning without poor local minima, in *Advances in Neural Information Processing Systems* (MIT Press, Cambridge, MA, 2016).

[38] B. B. Machta, R. Chachra, M. K. Transtrum, and J. P. Sethna, Parameter space compression underlies emergent theories and predictive models, Science **342**, 604 (2013).

[39] M. K. Transtrum *et al.*, Perspective: Sloppiness and emergent theories in physics, biology, and beyond, J. Chem. Phys. **143**, 010901 (2015).

[40] T. O'Leary, A. C. Sutton and E. Marder, Computational models in the age of large datasets, Curr. Opin. Neurobiol. **32**, 87 (2015).

[41] L. F. Abbott and W. G. Regehr, Synaptic computation, Nature (London) **431**, 796 (2004).

[42] N. Spruston, Pyramidal neurons: dendritic structure and synaptic integration, Nat. Rev. Neurosci. **9**, 206 (2008).

[43] H. Zeng and J. R. Sanes, Neuronal cell-type classification: challenges, opportunities and the path forward, Nat. Rev. Neurosci. **18**, 530 (2017).

[44] S. G. N. Grant, Synapse molecular complexity and the plasticity behaviour problem, Brain Neurosci. Adv. **2**, 239821281881068 (2018).

[45] C. Curto and K. Morrison, Relating network connectivity to dynamics: Opportunities and challenges for theoretical neuroscience, Curr. Opin. Neurobiol. **58**, 11 (2019).

[46] Y. N. Billeh *et al.*, Systematic integration of structural and functional data into multiscale models of mouse primary visual cortex, Neuron **106**, 388 (2020).

[47] M. Almog and A. Korngreen, Is realistic neuronal modeling realistic? J. Neurophysiol. **116**, 2180 (2016).

[48] S. R. Bittner *et al.*, Interrogating theoretical models of neural computation with emergent property inference, eLife **10**, e56265 (2021).

[49] P. J. Gonçalves *et al.*, Training deep neural density estimators to identify mechanistic models of neural dynamics, eLife **9**, e56261 (2020).

[50] A. Treves and E. T. Rolls, What determines the capacity of autoassociative memories in the brain? Netw. Comput. Neural Syst. **2**, 371 (1991).

[51] E. Salinas and L. F. Abbott, A model of multiplicative neural responses in parietal cortex, Proc. Natl. Acad. Sci. USA **93**, 11956 (1996).

[52] R. L. T. Hahnloser, On the piecewise analysis of networks of linear threshold neurons, Neural Netw. **11**, 691 (1998).

[53] R. H. Hahnloser, H. S. Seung, and J. J. Slotine, Permitted and forbidden sets in symmetric threshold-linear networks, Neural Comput. **15**, 621 (2003).

[54] K. Morrison, A. Degeratu, V. Itskov, and C. Curto, Diversity of emergent dynamics in competitive threshold-linear networks: a preliminary report, arXiv:1605.04463.

[55] C. Curto, J. Geneson, and K. Morrison, Fixed points of competitive threshold-linear networks, Neural Comput. **31**, 94 (2019).

[56] E. Marder, Neuromodulation of neuronal circuits: back to the future, Neuron **76**, 1 (2012).

[57] Y. Mu *et al.*, Glia accumulate evidence that actins are futile and suppress unseccessful behavior, Cell **178**, 27 (2019).

[58] Aitchison *et al.*, Model-based Bayesian inference of neural activity and connectivity from all-optical interrogation of a neural circuit, in *Proceedings of the 31st Conference on Neural Information Processing Systems (NIPS'17)*, Long Beach, CA (2017).

[59] C. Grienberger and A. Konnerth, Imaging calcium in neurons, Neuron **73**, 862 (2012).

[60] B. A. Wilt, J. E. Fitzgerald, and M. J. Schnitzer, Photon shot noise limits on optical detection of neuronal spikes and estimation of spike timing, Biophys. J. **104**, 51 (2013).

[61] L. Theis *et al.*, Benchmarking spike rate inference in population calcium imaging, Neuron **90**, 471 (2016).

[62] N. K. Logothetis and J. Pfeuffer, On the nature of the BOLD fMRI contrast mechanism, Magn. Reson. Imaging **22**, 1517 (2004).

[63] M. J. Bartolo *et al.*, Stimulus-induced dissociation of neuronal firing rates and local field potential gamma power and its relationship to the resonance blood oxygen level-dependent signal in macaque primary visual cortex, Eur. J. Neurosci. **34**, 1857 (2011).

[64] J. Heinzle, P. J. Koopmans, H. E. M. den Ouden, S. Raman, and K. E. Stephan, A hemodynamic model for layer BOLD signals, NeuroImage **125**, 556 (2016).

[65] M. Advani, S. Lahiri, and S. Ganguli, Statistical mechanics of complex neural systems and high dimensional data, J. Stat. Mech. (2013) P03014.

[66] V. Nair and G. E. Hinton, Rectified linear units improve restricted Boltzmann machines, in *Proceedings of the International Conference on Machine Learning (ICML)* (ACM Press, New York, NY, 2010).

[67] A. Krizhevsky, Convolutional deep belief networks on CIFAR-10 (2010), https://www.cs.toronto.edu/~kriz/conv-cifar10-aug2010.pdf.

[68] B. Xu, N. Wang, T. Chen, and M. Li, Empirical evaluation of rectified activations in convolutional network, in *Deep Learning Workshop* (ICML, 2015).

[69] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Learning representations by back-propagating errors, Nature (London) **323**, 533 (1986).

[70] J. Frankle and M. Carbin, The lottery ticket hypothesis: Finding sparse, trainable neural networks, arXiv:1803.03635.

[71] H. Zhou, J. Lan, R. Liu, and J. Yosinski, Deconstructing lottery tickets: Zeros, signs, and the supermask, in *Advances in Neural Information Processing Systems*, edited by H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (Curran Associates, 2019), Vol. 32.

[72] T. M. Cover, Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition, IEEE Trans. Electr. Comput. **14**, 326 (1965).

[73] E. Gardner, The space of interactions in neural network models, J. Phys. A: Math. Gen. **21**, 257 (1988).

[74] D. Marr, A theory of cerebellar cortex, J. Physiol. **202**, 437 (1969).

[75] B. A. Olshausen and D. J. Field, Sparse coding with an overcomplete basis set: A strategy employed by V1? Vision Res. **37**, 3311 (1997).

[76] X. Glorot, A. Bordes, and Y. Bengio, Deep sparse rectifier neural networks, in *Proceedings of the 14th International Conference on Artificial Intelligence and Statistics (AISTATS)* (ACM Press, New York, NY, 2011).

[77] F. Kubo *et al.*, Functional architecture of an optic flow-responsive area that drives horizontal eye movements in zebrafish, Neuron **81**, 1344 (2014).

[78] J. T. Trachtenberg *et al.*, Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex, Nature (London) **420**, 788 (2002).

[79] Y. Ziv *et al.*, Long-term dynamics of CA hippocampal place codes, Nat. Neurosci. **16**, 264 (2013).

[80] A. Attardo, J. E. Fitzgerald, and M. J. Schnitzer, Impermanence of dendritic spines in live adult CA1 hippocampus, Nature (London) **523**, 592 (2015).

[81] L. N. Driscoll, N. L. Pettit, M. Minderer, S. N. Chettih, and C. D. Harvey, Dynamic reorganization of neuronal activity patterns in parietal cortex, Cell **170**, 986 (2017).

[82] M. E. Rule, T. O'Leary, and C. D. Harvey, Causes and consequences of representational drift, Curr. Opin. Neurobiol. **58**, 141 (2019).

[83] C. E. Schoonover, S. N. Ohashi, R. Axel, and A. J. P. Fink, Representational drift in primary olfactory cortex, Nature (London) **594**, 541 (2021).

[84] T. D. Marks and M. J. Goard, Stimulus-dependent representational drift in primary visual cortex, Nat. Commun. **12**, 5169 (2021).

[85] D. Deitch, A. Rubin, and Y. Ziv, Representational drift in the mouse visual cortex, Curr. Biol. **31**, 4327 (2021).

[86] D. Kappel, R. Legenstein, S. Habenschuss, M. Hsieh, and W. Maass, A dynamic connectome supports the emergence of stable computational function of neural circuits through reward-based learning, eNeuro **5**, ENEURO.0301-17.2018 (2018).

[87] G. Burnstock, Cotransmission, Curr. Opin. Pharmacol. **4**, 47 (2004).

[88] B. L. Chen, D. H. Hall, and D. B. Chklovskii, Wiring optimization can relate neuronal structure and function, Proc. Natl. Acad. Sci. USA **103**, 4723 (2006).

[89] N. Brunel, V. Hakim, P. Isope, J. Nadal, and B. Barbour, Optimal information storage and the distribution of synaptic weights: Perceptron versus Purkinje cell, Neuron **43**, 745 (2004).