Estimating Saturation Volumes in CTRNNs

Bryan C. Daniels

August 18, 2006

1 Introduction

Our goal is to estimate the probability that we have a certain number of nodes M in their "active" regions, so that the network as a whole produces M-dimensional dynamics.

2 Finding average θ boundaries

One way to view the estimation of the probability of a random network having M-dimensional dynamics is by looking at the parameter-space volumes of the asymptotic saturation regions found in Ref. [1]. The idea is to find the total volume of the regions in parameter space that produce M active nodes, and divide by the volume of the whole parameter space (specified by the limiting boundaries placed on each parameter). This gives the probability that a set of parameters picked randomly from within the limiting boundaries will produce M active nodes.

The parameter space for CTRNNs consists of N time constants (τ_i), N biases (θ_i), N self-weights (W_{ii}), and $N^2 - N$ cross-weights ($W_{ij}, i \neq j$). We will ignore the time constants, since they do not affect saturation (see [1]). Based on the specific set of θ_i, W_{ii}, and W_{ij}, we can classify each node as being in one of three different "regimes": saturated off, in which case its output is always near 0, regardless of the states of the other nodes in the circuit; saturated on, in which case its output is always near 1; or active, in which case the node can actively respond to the input of the other nodes. In the asymptotic saturation approximation, we assume that a node saturated on (off) always has an output of *exactly* 1 (0).

With this approximation, given the weights W and the regimes of the other nodes in the network, it is possible to solve for the boundaries on θ necessary for a given node i to be in one of the three regimes; we call the lower bound l_i and the upper bound u_i. Since these boundaries will change with the weights, we can integrate over the weights to obtain "average" boundaries for θ_i:

$$\begin{aligned} \langle u_i \rangle &= \frac{1}{V_T} \int_{\mathcal{W}_i} [u_i]_{\theta_{\min}}^{\theta_{\max}} \\ \langle l_i \rangle &= \frac{1}{V_T} \int_{\mathcal{W}_i} [l_i]_{\theta_{\min}}^{\theta_{\max}}, \end{aligned} \tag{1}$$

	Saturated Off	Active	Saturated On
u_i	$I_L(w_{\rm self}) - \max \mathcal{I}^M - \sum_{j=1}^U w_j$	$I_R(w_{\rm self}) - \min \mathcal{I}^M - \sum_{j=1}^U w_j$	∞
l_i	$-\infty$	$I_L(w_{\rm self}) - \max \mathcal{I}^M - \sum_{j=1}^U w_j$	$I_R(w_{\rm self}) - \min \mathcal{I}^M - \sum_{j=1}^U w_j$

Table 1: The boundaries on θ for getting different node "regimes," from [1]. Note that these are written in an arbitrary form that works due to symmetry in the integrals over $\mathcal{W}_i$.

where $\mathcal{W}_i$ represents the N-dimensional range of weights, and V_T is the total integration volume

$$V_T = (w_{\rm max}^{\rm self} - w_{\rm min}^{\rm self})(w_{\rm max} - w_{\rm min})^{N-1}. \tag{2}$$

In this paper, we will write l_i and u_i in a form that uses the symmetry that occurs since we are always integrating over the same range of Ws for each node. These boundaries are shown in Table 1. In the table, M is the number of active nodes and U is the number of nodes saturated on. The notation $\mathcal{I}^M$ represents the input that the node receives from the M other active nodes in the network.[1] We have taken symmetry into account in that the integral will always give the same value for a given N, M, and U — for instance, we would get the same result whether node 1 is active and node 2 is on or vice versa. Thus we can write the self-weight as an arbitrary $w_{\rm self}$, and have U cross-weights that we arbitrarily label from 1 to U.

Notice that, besides ∞ and $-\infty$, there are only two distinct values in the table. We therefore lose no information if we only calculate the boundaries for the active regime:

$$\begin{aligned}
\langle u_i(\text{active})\rangle &= \frac{1}{V_T}\int_{w_{\rm min}^{\rm self}}^{w_{\rm max}^{\rm self}}\int_{w_{\rm min}}^{w_{\rm max}}\cdots\int_{w_{\rm min}}^{w_{\rm max}}\left[I_R(w_{\rm self}) - \left(\min \mathcal{I}^M + \sum_{j=1}^U w_j\right)\right]_{\theta_{\rm min}}^{\theta_{\rm max}} dw_{\rm self}\, dw_1 \ldots dw_{N-1}, \\
\langle l_i(\text{active})\rangle &= \frac{1}{V_T}\int_{w_{\rm min}^{\rm self}}^{w_{\rm max}^{\rm self}}\int_{w_{\rm min}}^{w_{\rm max}}\cdots\int_{w_{\rm min}}^{w_{\rm max}}\left[I_L(w_{\rm self}) - \left(\max \mathcal{I}^M + \sum_{j=1}^U w_j\right)\right]_{\theta_{\rm min}}^{\theta_{\rm max}} dw_{\rm self}\, dw_1 \ldots dw_{N-1}.
\end{aligned} \tag{3}$$

Furthermore, since the probabilities we are interested in will always be written in terms of differences between the average θ boundaries and $\theta_{\rm min}$ and $\theta_{\rm max}$, we can shift all of the average boundaries by a constant amount. It will be convenient to shift and rescale the average boundaries so they are always between 0 and 1. We therefore make the following definitions:

$$\begin{aligned}
{}^N\chi_D^U &= \frac{\langle u_i(\text{active})\rangle - \theta_{\rm min}}{\theta_{\rm max} - \theta_{\rm min}}, \\
{}^N\bar{\chi}_D^U &= \frac{\langle l_i(\text{active})\rangle - \theta_{\rm min}}{\theta_{\rm max} - \theta_{\rm min}},
\end{aligned} \tag{4}$$

where U and D are the number of nodes saturated on and off, respectively, and the number of active nodes is $M = N - U - D$.

As an example of the use of these definitions, we can find the *probability* of the three possible regimes for

[1] Note that this is a slightly different definition than $\mathcal{I}^N$ in [1], which indicated $N-1$ active nodes.

a single node, *given* that there are D total nodes saturated off and U total saturated on:

$$P_D^U(\text{active}) = ({}^N\chi_D^U - {}^N\bar{\chi}_D^U) \tag{5}$$

$$P_D^U(\text{on}) = (1 - {}^N\chi_D^{U-1}) \tag{6}$$

$$P_D^U(\text{off}) = {}^N\bar{\chi}_{D-1}^U. \tag{7}$$

So, for example, the probability of choosing parameters within $\mathcal{R}_N^N$, so that all N nodes are within their active regions, is[2]

$$P(\mathcal{R}_N^N) = \left[P_0^0(\text{active})\right]^N = \left(\chi_0^0 - \bar{\chi}_0^0\right)^N; \tag{8}$$

The more general case of finding $P(\mathcal{R}_M^N)$ involves overlapping regions, and becomes much more complicated; we will discuss this in Section 3. It is important to note, however, that the calculation of $P(\mathcal{R}_M^N)$ also involves only differences of χ and $\bar{\chi}$; this is ultimately the motivation for defining them.

2.1 Estimating χs

We will now discuss a way to view the calculation of χ and $\bar{\chi}$ in a different way from Eq. (3) and Eq. (4) that is more amenable to invoking helpful approximations. First, ignoring the effects of the other nodes (the weights w_1 through w_{N-1}), we view the integration in Eq. (3) of the clipped integrand over w_{self} as defining a two-dimensional area (see Figure 1). We will call the percentage of "filled" area F (or $\bar{F}$ for the corresponding integral involving I_L)[3]. Ignoring for the moment the other nodes, we have then

$$F = \frac{1}{A_T}\int_{w_{\min}^{\text{self}}}^{w_{\max}^{\text{self}}} \left[I_R(w_{\text{self}})\right]_{\theta_{\min}}^{\theta_{\max}} dw_{\text{self}}, \tag{9}$$

where we have again factored out the total area A_T of the region to get a percentage for F. Now depending on the other nodes' weights, we will also need add some value inside the square brackets [see Eq. (3)]. Calling this number x, we write F as a function of x:

$$F(x) = \frac{1}{A_T}\int_{w_{\min}^{\text{self}}}^{w_{\max}^{\text{self}}} \left[I_R(w_{\text{self}}) + x\right]_{\theta_{\min}}^{\theta_{\max}} dw_{\text{self}}. \tag{10}$$

We now view the ramaining integrals in Eq. (3) as defining the probability of getting different values for x. More specifically, we *define* a normalized distribution $\rho(x)$ such that $\Delta x\rho(x)$ gives the probability of finding $-(\min \mathcal{I}^M + \sum_{j=1}^U w_j)$ between x and $x + \Delta x$ as the ws range through their different values. With these

[2]Note that we have dropped the N superscripts, which we will do from now on where N is apparent from the context.

[3]In the remainder of this section, we will often omit the barred case for simplicity, as it is usually a straightforward generalization of the unbarred case.

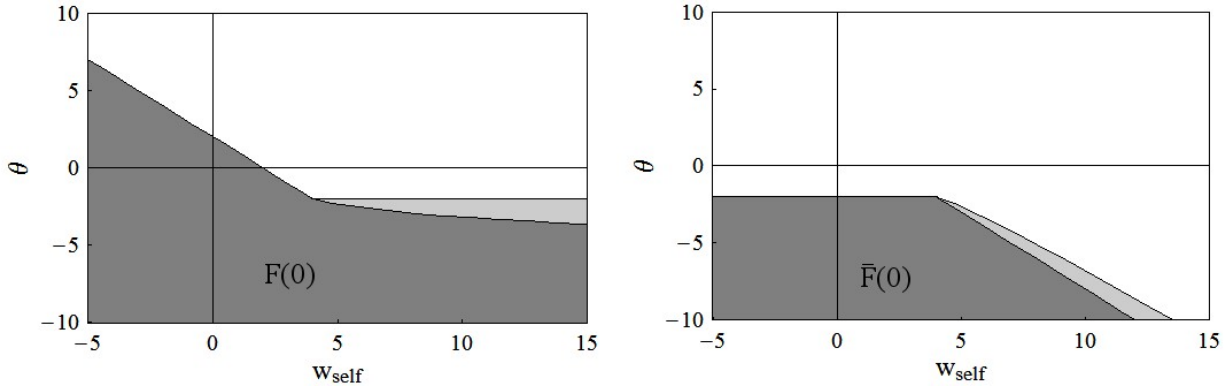

Figure 1: The function $F(x)$ gives the percentage of the total area that is under the curve $\theta = I_R(w_{\text{self}}) + x$. $\bar{F}(\bar{x})$ gives the corresponding area under the curve $\theta = I_L(w_{\text{self}}) - \bar{x}$. This area is represented by the dark shading, for $w_{\text{self}} \in (-5, 15), \theta \in (-10, 10), x = 0$. The light area plus the dark area corresponds to the simplified version of $F(x)$ and $\bar{F}(\bar{x})$.

definitions, χ can be written rather simply in terms of F and ρ:

$$ {}^{N}\chi_D^U = \int F(x)\rho(x)\, dx, \tag{11} $$

or, writing the dependence of F and ρ explicitly,

$$ {}^{N}\chi_D^U = \int F(x, w_{\text{min}}^{\text{self}}, w_{\text{max}}^{\text{self}}, \theta_{\text{min}}, \theta_{\text{max}})\ \rho(x, w_{\text{min}}, w_{\text{max}}, M, U)\ dx. \tag{12} $$

Unfortunately, there is no way to write an exact formula for either $F(x)$ or $\rho(x)$ that does not itself involve integration. The key fact, however, is that both $F(x)$ and $\rho(x)$ are well-approximated by simple functions under many circumstances.

To see this, we will first look at approximations of $F(x)$. This function gives the percentage of the region $(w_{\text{min}}^{\text{self}}, w_{\text{max}}^{\text{self}}) \times (\theta_{\text{min}}, \theta_{\text{max}})$ that lies under the curve $\theta = I_R(w_{\text{self}}) + x$. As seen in Figure 1, using the extended definitions of I_L and I_R, F and $\bar{F}$ will generally involve the area under a straight line when $w_{\text{self}} < 4$, and under a curve when $w_{\text{self}} > 4$. The first approximation we will make is shown as light shading in Figure 1: we will extend the linear definitions of I_L and I_R into the region where $w_{\text{self}} > 4$. This turns our calculation of $F(x)$ into a simple sum of rectangles and triangles of varying sizes. Appendix A gives explicit formulas for this approximation.

An example of this approximation, compared with the "exact" result obtained numerically from Eq. (10), is shown in Figure 2. For this set of ranges, the simplified version approximates $F(x)$ quite well. Note, though, that if we were to choose w_{self} and θ ranges inside which the area between I_R and its linearized version is significant, the approximation would be worse.

A further possible simplification would be to assume $F(x)$ is simply linear in the region where it travels between zero and one. This will be useful in finding an approximate functional form for χ, but the validity

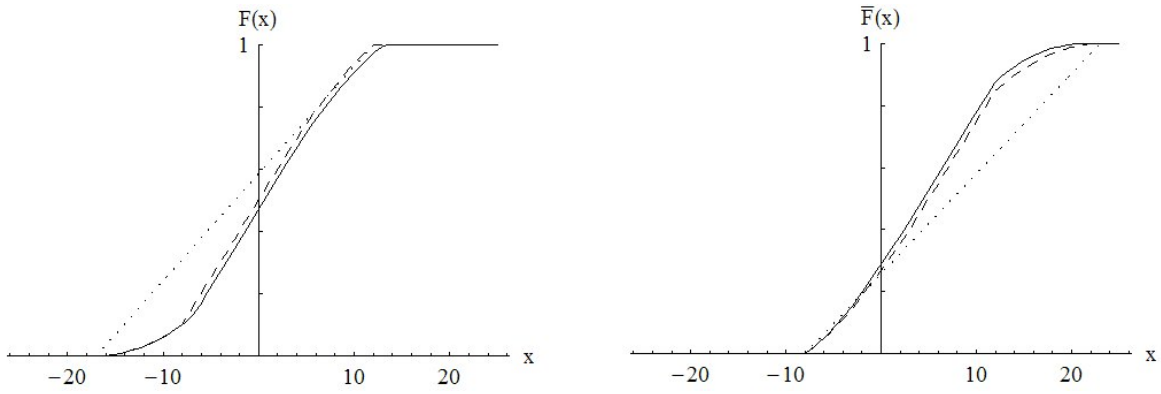

Figure 2: Solid curve: The function $F(x)$ for $w_{\text{self}} \in (-5, 15), \theta \in (-10, 10)$. $F(0)$ corresponds to the percentage of darkly shaded area in Figure 1. Dashed curve: An approximation to $F(x)$, corresponding to the straight line approximation shown in Figure 1. This approximation is defined piecewise, and has parts that are linear and quadratic in x. Dotted line: A further simplification which takes $F(x)$ to be linear in x in the middle region.

of the approximation is more questionable. See the dotted lines in Figure 2 for this second approximation of $F(x)$.

The second half of the problem for approximating χ is to find the probability distribution $\rho(x)$. Recall that $\rho(x)$ is the probability distribution corresponding to $x = (\min \mathcal{I}^M + \sum_{j=1}^{U} w_j)$, as each of the weights range from $w_{\min}$ to $w_{\max}$. Or, we can think of the problem probabilistically; we are picking M weights from the given range and computing the minimum of $\mathcal{I}^M$, and then adding this to the sum of U weights picked from the same range.

Suppose first that $M = 0$, so that we have only U other nodes saturated on. Then we just need the probability distribution corresponding to the sum of U numbers picked from uniform distributions. For large U, this approaches a normal distribution with mean μ_{on} and variance σ_{on}^2 given by

$$\mu_{on} = -\frac{U}{2}(w_{\max} + w_{\min}) \tag{13}$$

$$\sigma_{on}^2 = \frac{U}{3}\frac{{w_{\max}}^3 - {w_{\min}}^3}{w_{\max} - w_{\min}} - \frac{{\mu_{on}}^2}{U}. \tag{14}$$

This is what we will use for our approximation of $\rho(x)$ when $M = 0$.

Now we consider the case where $U = 0$, but $M \neq 0$. Then $x = \min \mathcal{I}^M$. From our probabilistic viewpoint, this is telling us to first choose M values between $w_{\min}$ and $w_{\max}$, and then to add up only the *negative* values from that set. To see what distribution this corresponds to, first suppose that $w_{\min} = -w_{\max}$. Then, with each w that we pick, we have a 50% chance of picking a positive number, which we will ignore, and a 50% chance of getting a negative number, which we will add to the total. Then with each w, we are drawing from a distribution that we will call $\rho_1(x)$, which corresponds to a "half delta function" at zero plus a "half

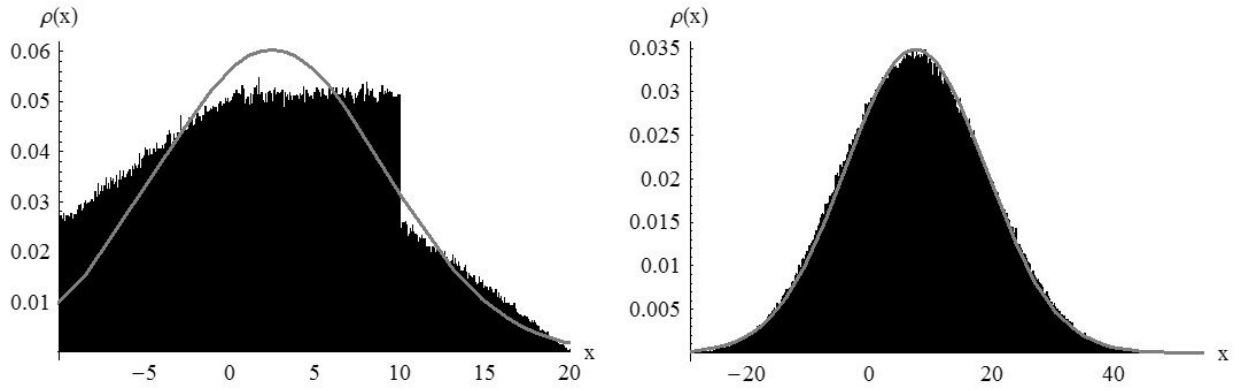

Figure 3: Comparison of empirically-generated $\rho(x)$ distributions and the normal distribution approximation used in this paper, for $w \in (-10, 10)$. The left plot is for $U = M = 1$, and the right plot is for $U = M = 3$. The agreement improves as U and M increase. In each case, 10^6 sets of random ws were generated, and x was calculated for each to construct a histogram.

uniform distribution" from $w_{\min}$ to zero. Or, if we generalize to let $w_{\min}$ and $w_{\max}$ have any values, then

$$\rho_1(x) = \frac{b}{|w_{\min}|} + (1-b)\delta(x), \qquad w_{\min} \le x \le 0, \tag{15}$$

and $\rho_1(x) = 0$ outside that range. The parameter b is the percentage of the range of w values that is negative, or (as long as $w_{\min} \le 0$)

$$b = \frac{-w_{\min}}{w_{\max} - w_{\min}}. \tag{16}$$

In our example where $w_{\min} = -w_{\max}$, $b = 1/2$. Now if M is large, the overall distribution $\rho(x)$ [which would in general be a convolution of M copies of $\rho_1(x)$] also approaches a normal distribution, with mean μ_{active} and variance σ^2_{active} given by (see appendix?)

$$\mu_{active} = -M\ b\ \frac{w_{\min}}{2} \tag{17}$$

$$\sigma^2_{active} = M\ b(4-3b)\ \frac{{w_{\min}}^2}{12}. \tag{18}$$

Finally, we need to combine these two distributions for the case where both U and M may be nonzero. Our general approximation for $\rho(x)$ will then be a convolution of the two normal distributions we have just found, which is simply another normal distribution with mean and variance

$$\mu = \mu_{on} + \mu_{active} \tag{19}$$

$$\sigma^2 = \sigma^2_{on} + \sigma^2_{active}. \tag{20}$$

Figure 3 compares this normal distribution to the actual $\rho(x)$, calculated by randomly generating many sets of ws and finding the distribution of x. The approximation is very close to the true distribution for large U

and M.

With these approximations for $F(x)$ and $\rho(x)$, it is possible to estimate ${}^N\chi_D^U$ using Eq. (11) by doing a one-dimensional integral over x. This greatly improves the speed with which calculations of $P(\mathcal{R}_M^N)$ can be made.

3 Estimating $P(\mathcal{R}_M^N)$

Using Eq. (11), we now have a way of efficiently estimating χ and $\bar{\chi}$, giving us the "average" boundaries in θ space that can be used to calculate the relative volumes of all the different $\mathcal{R}_M^N$ regions. While we already have enough information to begin efficiently estimating the probability of hitting $\mathcal{R}_N^N$ [using Eq. (8) and our approximate χ and $\bar{\chi}$], we will now introduce a recursive approach for finding $P(\mathcal{R}_M^N)$ for general M.

This problem reduces naturally to finding the probabilities of different numbers of nodes being active (M), saturated on (U), and saturated off (D). For a given N and M, we want

$$U + D = N - M, \tag{21}$$

but there are in general many ways of choosing U and D to get this total. Also, for any particular U and D, there are many ways of picking which specific nodes are on and off out of the N total. Taking all of this into account, we can write

$$P(\mathcal{R}_M^N) = \sum_{U=0}^{N-M} \binom{N}{U}\binom{N-U}{D} A_D^U S_D^U, \tag{22}$$

where A_D^U is the probability of having $M = N - U - D$ active nodes when there are U other nodes on and D nodes off, and S_D^U is the probability of having U nodes saturated on and D saturated off.

To find expressions for A_D^U and S_D^U, we first look back at Eqs. (5)-(7). These give probabilities for the different regimes of a given node based on the regimes of all the other nodes. Based on these equations, we would say[4]

$$A_D^U = \left[P_D^U(\text{active})\right]^M = (\chi_D^U - \bar{\chi}_D^U)^M, \tag{23}$$

and for S_D^U, we might guess

$$\left[P_D^U(\text{on})\right]^U \left[P_D^U(\text{off})\right]^D.$$

The problem is, this method is picking the set of θ_is correctly only if all the other nodes are saturated the way we want them to be. This in general gives a result that is too large. For example, in calculating S_2^4, we want 4 nodes saturated on and 2 saturated off. So we pick 4 nodes to have θ such that each is saturated on, *given* that 3 others are on and 2 others are off; and we pick 2 nodes to have θ such that each is saturated off, *given* that 4 others are on and 1 other is off. But it could happen that we accidentally choose the set of θs such that a subset of two or more of the nodes are actually active, since all we know for sure is that we

[4] Again, we are suppressing the N superscript on the χs and $\bar{\chi}$s.

have chosen things so that no single node can be active by itself.

A key fact, however, is that once a node's θ is chosen within its *active* region for some U and D, it always remains active even if some of the saturated nodes "become" active. So our Eq. (23) is correct as written, and we just have to worry about correctly calculating S_D^U.

One way to calculate S_D^U is to first find the probability of choosing all of the θs correctly individually, as described above, and then subtract the overlapping regions of θ space where some number M_a of the nodes are actually active.[5] We end up with a recursive equation (see Appendix B for a derivation):

$$\begin{aligned} S_D^U &= (1-\chi_D^{U-1})^U(\bar{\chi}_{D-1}^U)^D \\ &\quad - \sum_{M_a=2}^{U+D} \sum_{i=\max(M_a-D,0)}^{\min(M_a,U)} \binom{U}{i}\binom{D}{M_a-i} \left(\chi_{D_a}^{U_a}-\chi_D^{U-1}\right)^i \left(\bar{\chi}_{D-1}^U-\bar{\chi}_{D_a}^{U_a}\right)^{M_a-i} S_{D_a}^{U_a}, \end{aligned} \tag{24}$$

where $U_a = U - i$, $D_a = D - (M_a - i)$. This, combined with Eq. (22), Eq. (23), and the estimates of χ and $\bar{\chi}$ from the last section, gives a way of calculating $P(\mathcal{R}_M^N)$ for any given N, M, and parameter ranges. This is implemented in the *Mathematica* notebook.

4 An approximate formula for χ and $\bar{\chi}$

Due to the form of the approximate versions of $F(x)$ and $\rho(x)$ that we use, it is possible to write an approximate formula for χ that does not involve integration. Here we will write a formula that assumes $F(x)$ to be purely linear in x (see the dotted line in Figure 2). If we define x_1 and x_2 as the values of x between which our approximation to $F(x)$ increases linearly from 0 to 1, then[6]

$${}^N\chi_D^U \approx \frac{1}{2}\left(\frac{1}{x_1-x_2}\left(\sqrt{\frac{2\sigma^2}{\pi}}\left(e^{-\frac{(x_2-\mu)^2}{2\sigma^2}}-e^{-\frac{(x_1-\mu)^2}{2\sigma^2}}\right)+(x_1-\mu)\left(\mathrm{erf}(\frac{x_2-\mu}{\sqrt{2\sigma^2}})-\mathrm{erf}(\frac{x_1-\mu}{\sqrt{2\sigma^2}})\right)\right)+1-\mathrm{erf}(\frac{x_2-\mu}{\sqrt{2\sigma^2}})\right), \tag{25}$$

with μ and σ given by Eq. (19) and Eq. (20).

A Explicit forms for estimating $F(x)$ and $\bar{F}(\bar{x})$

$$F(x) \approx \frac{F_1(x)+F_2(x)}{(w_{\max}^{\mathrm{self}}-w_{\min}^{\mathrm{self}})(\theta_{\max}-\theta_{\min})}, \tag{26}$$

[5] A way of seeing this visually (at least up to N=3) is by looking at the pictures of the asymptotic saturation regions in Ref. [1]—we can start off with a rectangular region that is too large, and then subtract off the "corners" where there are active nodes.

[6] The transcendental function erf is the so-called error function, which shows up because we are integrating normal distributions.

where

$$F_1(x) = \begin{cases} 0 & x \le (\theta_{\min}+2) + (w^{\text{self}}_{\min} - 4), \\ \frac{1}{2}\big(x - (\theta_{\min}+2) - (w^{\text{self}}_{\min}-4)\big)^2 & (\theta_{\min}+2) + (w^{\text{self}}_{\min}-4) \le x \le \theta_{\min}+2, \\ \frac{1}{2}(w^{\text{self}}_{\min}-4)^2 + (w^{\text{self}}_{\max} - w^{\text{self}}_{\min})\big(x - (\theta_{\min}+2)\big) & \theta_{\min}+2 \le x \le \theta_{\max}+2, \\ (w^{\text{self}}_{\max} - w^{\text{self}}_{\min})(\theta_{\max} - \theta_{\min}) & x \ge \theta_{\max}+2 \end{cases} \tag{27}$$

$$F_2(x) = \begin{cases} 0 & x \le (\theta_{\max}+2) + (w^{\text{self}}_{\min} - 4), \\ -\frac{1}{2}\Big(x - \big((\theta_{\max}+2) + (w^{\text{self}}_{\min}-4)\big)\Big)^2 & (\theta_{\max}+2) + (w^{\text{self}}_{\min}-4) \le x \le \theta_{\max}+2, \\ 0 & x \ge \theta_{\max}+2. \end{cases} \tag{28}$$

$\bar{F}(\bar{x})$ is equal to 1 minus $F(x)$ with all the maxs and mins swapped and the sense of each inequality flipped.

B Deriving S^U_D

Here we derive Eq. (24) for S^U_D, the probability of finding U nodes saturated on and D nodes saturated off. We are assuming that we already know the necessary average θ boundaries (the various χ and $\bar{\chi}$).

As noted in the text, we first assume that every node is in its correct regime, and pick each node's θ from the correct range. The corresponding probability is

$$\big[P^U_D(\text{on})\big]^U \big[P^U_D(\text{off})\big]^D,$$

which we can easily compute using Eq. (6) and Eq. (7). Now we need to subtract regions where this "naive" approach overlaps with regions with some number of active nodes M_a. We will in general have overlap contributions from regions with M_a between 2 and $U+D$ (where all the nodes are actually active).

For each M_a, we need to account for all the different ways in which overlapping could occur. First, some number i of the nodes we picked to be saturated *on* will actually be active, contributing to M_a. Then the actual number of nodes saturated on is $U_a = U - i$. This will happen with a probability given by the range shared by the ranges that tell us where we're saturated on and where we're active in the two cases. Thinking of each P^U_D as specifying a range, we will write the overlap as an intersection of the ranges, so that the probability of hitting this overlap is

$$\Big[P^U_D(\text{on}) \cap P^{U_a}_{D_a}(\text{active})\Big]^i.$$

Next, there must be $(M_a - i)$ active nodes coming from the nodes that we initially picked to be saturated

off. Analogously, this happens with probability

$$\left[P_D^U(\text{off}) \cap P_{D_a}^{U_a}(\text{active})\right]^{M_a - i}.$$

To account for all the overlap, we must allow i to take on all possible values. With a little thought, we see that the lowest i could be is $M_a - D$, unless $D > M_a$, in which case we start i at zero. The highest it could be is U, unless $U > M_a$, in which case it can only go up to M_a.

Then we must also include the probability associated with picking all the rest of the saturated nodes. This will always overlap completely with our original region (why?), so it is just $S_{D_a}^{U_a}$. Finally, for each M_a and i, there are multiple ways to pick the individual nodes that are active, giving us two combinatorial factors. Putting this all together,

$$\begin{aligned} S_D^U &= \left[P_D^U(\text{on})\right]^U \left[P_D^U(\text{off})\right]^D \\ &\quad - \sum_{M_a=2}^{U+D} \sum_{i=\max(M_a-D,0)}^{\min(M_a,U)} \binom{U}{i} \binom{D}{M_a - i} \left[P_D^U(\text{on}) \cap P_{D_a}^{U_a}(\text{active})\right]^i \left[P_D^U(\text{off}) \cap P_{D_a}^{U_a}(\text{active})\right]^{M_a - i} S_{D_a}^{U_a}, \end{aligned} \tag{29}$$

with $U_a = U - i$, $D_a = D - (M_a - i)$.

To find the intersections of the different ranges, we need to know some relationships among the various χs and $\bar{\chi}$s. It turns out (explain?) that the following relationships always hold:

$$\chi_j^i \geq \chi_J^i, \quad \chi_j^i \geq \chi_j^I; \tag{30}$$

$$\bar{\chi}_j^i \leq \bar{\chi}_J^i, \quad \bar{\chi}_j^i \leq \bar{\chi}_j^I; \tag{31}$$

$$\bar{\chi}_j^i \leq \chi_l^k; \tag{32}$$

for any i, j, k, l, I, J where $I > i$ and $J > j$, and all the relationships are for the same total number of nodes N.

We make use of these relationships in computing the overlapping regions. For example, in the case of the "on" overlap range in Eq. (29), we have [using Eq. (5) and Eq. (6)]

$$P_D^U(\text{on}) \cap P_{D_a}^{U_a}(\text{active}) = \left[\chi_D^{U-1}, 1\right] \cap \left[\bar{\chi}_{D_a}^{U_a}, \chi_{D_a}^{U_a}\right]. \tag{33}$$

First, any $\bar{\chi}$ is less than any χ [Eq. (32)], so we know the lower bound of the intersection will be either χ_D^{U-1} or $\chi_{D_a}^{U_a}$. Second, we know that no χ is larger than 1, so the upper bound must be one of these same two. Finally, we know that $D_a = D - (M_a - i) \leq D$ and $U_a = U - i \leq U - 1$ in this case (note that if $i = 0$, we are not using this part at all). From Eq. (30), then, we see that $\chi_D^{U-1} \leq \chi_{D_a}^{U_a}$, so we do in general have a nonzero intersection. We get

$$P_D^U(\text{on}) \cap P_{D_a}^{U_a}(\text{active}) = \left(\chi_{D_a}^{U_a} - \chi_D^{U-1}\right). \tag{34}$$

Similar logic provides the intersection for the "off" part, and Eq. (29) becomes the final result of Eq. (24):

$$\begin{aligned} S_D^U &= (1-\chi_D^{U-1})^U(\bar{\chi}_{D-1}^U)^D \\ &\quad - \sum_{M_a=2}^{U+D} \sum_{i=\max(M_a-D,0)}^{\min(M_a,U)} \binom{U}{i}\binom{D}{M_a-i} \left(\chi_{D_a}^{U_a}-\chi_D^{U-1}\right)^i \left(\bar{\chi}_{D-1}^U-\bar{\chi}_{D_a}^{U_a}\right)^{M_a-i} S_{D_a}^{U_a}. \end{aligned} \tag{35}$$

References

[1] Randall D. Beer, *Parameter Space Structure of Continuous-Time Recurrent Neural Networks*. Accepted for publication in Neural Computation.