

# Mathematical Intuition for Criticality of Stochastic Processes

(Dated: June 30, 2018)

## I. MATHEMATICAL FRAMEWORK

Consider a network represented by directed graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ , where  $\mathcal{V} = 1, \dots, n$ , and  $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$  are the sets of network vertices and directed edges. Let  $A = [a_{ij}]$  be the weighted and directed adjacency matrix of  $\mathcal{G}$ . At each time point  $t \in \mathbb{Z}_{\geq 0}$ , we associate each node  $i$  with a discrete non-negative random variable  $X_i^t$ .

The evolution of the dynamics in our system follow a hierarchical stochastic process. Starting at some non-random initial state  $\mathbf{x}^0$ , we define discrete non-negative random variable  $X_j^t$  that represents the number of successful transmissions received by node  $j$  at time  $t$ . Initially, we define

$$X_j^{t+1} \sim \sum_{i=1}^n B(X_i^t, a_{ji}),$$

Because  $X_j$  at any point in time must be a non-negative discrete integer, we can actually enumerate all possible states of our system. We define a *state* to be any one set of spikes across all neurons  $\mathbf{x}$ . Some examples of states are

$$\mathbf{x}_i \in \left\{ \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots \right\} \in \mathbb{Z}_{\geq 0}^n.$$

Suppose we look at all states for neurons that spike up to  $m$  times. Then for  $n$  neurons, we have  $n^m$  possible states.

### A. Simplified Case: Homogeneous In-Strength

which is the sum of binomial distributions conditioned on random variables that are the states at time  $t$ . Suppose we have a system where all of the inputs into any node  $j$  are equal to each other, such that  $a_{j1} = a_{j2} = \dots = a_{jm}$ . Then we can rewrite this probability as

$$X_j^{t+1} \sim B\left(\sum_{i=1}^n X_i^t, a_j\right).$$

Finally, we realize that given  $\mathbf{X}^t$ , the random variables  $X_1^{t+1}, \dots, X_n^{t+1}$  are independent, so the probability of observing any particular state at  $t+1$  is

$$P(\mathbf{X}^{t+1} = \mathbf{x}^{t+1} | \mathbf{X}^t = \mathbf{x}^t) = \prod_{i=1}^n \binom{n_i}{x_i^{t+1}} a_i^{x_i^{t+1}} (1 - a_i^{x_i^{t+1}})^{n_i - x_i^{t+1}}, \quad (1)$$

where  $n_i$  is the sum of spikes from all neurons at time  $t$  feeding into neuron  $i$  at time  $t+1$ . Hence, we have a closed (albeit ugly) form for the probability of transitioning from any one state to any other state. This form can be generalized to arbitrarily heterogeneous edge weights, but looks a *lot* worse.

### B. Probability Matrices

Next, we realize this formulation implies that the only determinant of where to go in the next state is the previous state, and thereby construct a *Markov Process*, where state transitions from  $\mathbf{x}_i$  to  $\mathbf{x}_j$  occur with probability  $p_{ji}$  according to equation 1. Hence, we can construct probability matrix and vectors

$$\mathbb{P} = \begin{bmatrix} p_{11} & p_{21} & \cdots & p_{N1} \\ p_{12} & p_{22} & \cdots & p_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ p_{1N} & p_{2N} & \cdots & p_{NN} \end{bmatrix}, \quad \mathbf{p}^t = \begin{bmatrix} p(\mathbf{x}_1^t) \\ p(\mathbf{x}_2^t) \\ \vdots \\ p(\mathbf{x}_N^t) \end{bmatrix}$$

where  $N = n^m$ ,  $\mathbb{P}$  is the Markov *probability matrix*, and  $\mathbf{p}^t$  has entry  $i$  as the probability of the system being in state  $\mathbf{x}_i$  at time  $t$ . Then, we have the evolution of probabilities

$$\mathbf{p}^k = \mathbb{P}^k \mathbf{p}^0,$$

which is again a linear map representing the evolution of probabilities of being in particular states.

### C. Probability Matrix Spectral Properties

There are certain crucial properties of probability matrices. For one, because  $\mathbb{P}$  is a non-negative matrix, we can use the *Perron-Frobenius Theorem*, which indicates that  $\mathbb{P}$  has a real, positive largest eigenvalue  $\lambda_1$  with a non-negative eigenvector  $\mathbf{e}_1$ . Because the columns of  $\mathbb{P}$  also sum to 1, we also have that  $\lambda_1 = 1$ , and that all other eigenvalues  $|\lambda_i| < 1$  have magnitudes less than 1. Finally, a special property of probability matrices is that if  $\sum_i p_i^t = 1$ , then  $\sum_i p_i^{t+1} = 1$ , which makes intuitive sense as probabilities must always sum to 1, but will become an incredibly useful property later on. Finally, we note an interesting property specific to systems of our type, which is that if we start at the zero state  $\mathbf{x} = \mathbf{0}$ , then the probability to transition to the zero state is equal to 1. Hence,  $\lambda_1 = 1$  with  $\mathbf{e}_1 = [1; 0; \dots; 0]$ . To summarize:

- $\lambda_1 = 1$  with  $\mathbf{e}_1 = [1; 0; \dots; 0]$
- $|\lambda_i| < 1$  for  $i \neq 1$
- $\sum_i p_i^t = \sum_i p_i^{t+1} = 1$

### D. Implications of Spectral Properties

These properties are incredibly revealing about our system because the forward progression of state probabilities evolves linearly with respect to these eigenvalues and eigenvectors. Suppose any initial probability vector

$$\mathbf{p}^0 = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + \dots + c_N \mathbf{e}_N.$$

Then we simply have

$$\mathbf{p}^k = c_1 \lambda_1^k \mathbf{e}_1 + c_2 \lambda_2^k \mathbf{e}_2 + \dots + c_N \lambda_N^k \mathbf{e}_N.$$

The implication is that as time progresses forward and  $k$  becomes large, all probabilities eventually reach the Perron-Frobenius eigenvalue-eigenvector pair, which in our case is just the zero state. However, another implication is that the rate at which our probabilities are *not* the zero state decay according to a power law that is predominantly determined by  $\lambda_2$ . This is because for  $\lambda_1 > \lambda_2 > \dots > \lambda_N$ , as  $k$  becomes large,  $\lambda_2$  becomes the substantially largest decaying component. Hence, we hypothesize that  $\lambda_2$  is the exponent that determines the duration of avalanching phenomena. More generally, for any stochastic process by which the zero state an attractor state (probability from  $\mathbf{x}_0 \rightarrow \mathbf{x}_0 = 1$ ),  $\lambda_2$  will always be the dominant exponent for avalanche duration.



## II. UNIFYING FRAMEWORK FOR TRUNCATED POWER LAWS IN STOCHASTIC PROCESSES

Now that we've laid the groundwork of stochastic systems and probability matrices, let us revisit the most general case, and demonstrate the formation of truncated power laws.

Suppose we have a system of  $N$  states, where each state is represented by some vector  $\mathbf{x}_i$ , and that the probability  $p(\mathbf{x}_i \rightarrow \mathbf{x}_j) = p_{ji}$ , which forms probability matrix  $[\mathbb{P}]_{ji} = p_{ji}$ . Further assume that the only deterministic state is the zero state  $\mathbf{x}_1$ , such that  $p(\mathbf{x}_1 \rightarrow \mathbf{x}_1) = 1$ . Then from before, we have

- $\lambda_1 = 1$  with  $\mathbf{e}_1 = [1; 0; \dots; 0]$
- $|\lambda_i| < 1$  for  $i \neq 1$

and for any initial probability vector

$$\mathbf{p}^0 = c_1 \mathbf{e}_1 + c_2 \mathbf{e}_2 + \dots + c_N \mathbf{e}_N,$$

we have

$$\mathbf{p}^k = c_1 \lambda_1^k \mathbf{e}_1 + c_2 \lambda_2^k \mathbf{e}_2 + \dots + c_N \lambda_N^k \mathbf{e}_N.$$

Now, notice that  $\mathbb{P}$  is fixed and static, such that  $\lambda_1, \dots, \lambda_N$  never change. Also notice that  $c_1, \dots, c_N$  are determined by the initial probability, and are also fixed. Then the probability vector evolves forward in time as a sum of decaying exponentials.

### A. Cumulative Avalanche Duration

Suppose we care about the duration of our avalanche. Because  $\mathbf{e}_1$  is the eigenvector corresponding to the zero state  $\mathbf{0}$ , all other eigenvectors and eigenvalues correspond to probability modes that have components corresponding to non-zero states. In this case, we can write the probability of not being in the zero state as

$$P(\mathbf{X}^k \neq \mathbf{0}) = c'_2 \lambda_2^k + c'_3 \lambda_3^k + \dots + c'_N \lambda_N^k.$$

At this point, the reader is confused, because we were originally talking about avalanching phenomena that are intimately related to power law distributions, yet these avalanche durations unapologetically obey the decay of exponentials. To reconcile these notions, we imagine a contrived scenario where the ordered eigenvalues  $\lambda_2, \dots, \lambda_N$  are serendipitously arranged as a power law distribution. That is,

$$\lambda_i = c \left( \frac{i}{N+1} \right)^m. \quad (2)$$

Now, while this decision seems arbitrary, let's see what happens. If we arbitrarily select coefficients  $c'_1 = c'_2 = \dots = c'_N = 1$ , and pick eigenvalues from equation 2, we have

$$P(\mathbf{X}^k \neq \mathbf{0}) = \lambda_2^k + \lambda_3^k + \dots + \lambda_N^k.$$

Plotting  $\log_{10}(P(\mathbf{X}^k \neq \mathbf{0}))$  versus  $\log_{10}(k)$ , we get

and we'll even do that thing where we put a straight line next to it, which, I mean, you know, looks kind of like a power law for around 3 orders of magnitude, which is apparently what you "need" to be a valid power law, but that's neither here nor there. What's important here is that the mechanism that determines this very convincing power law shape is actually a power law decay of eigenvalues. So what if we want to change the exponent of this distribution?

Well, we simply have to cleverly pick the coefficients in front of the eigenvalues in order to make the power law more or less steep.

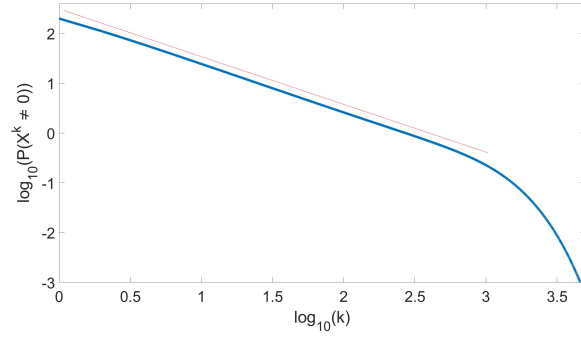


FIG. 1. Power law over three orders of magnitude using a clever sum of exponentials.

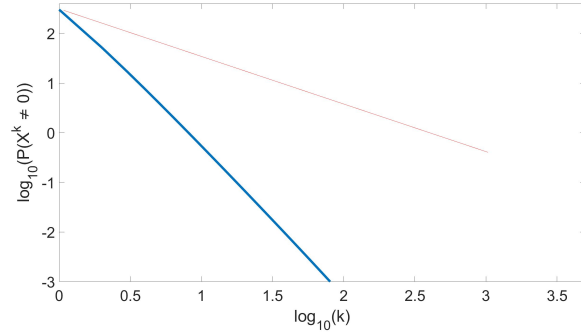


FIG. 2. Power law over three orders of magnitude using a clever sum of cleverly weighted exponentials.

### B. Non-Cumulative Avalanche Duration

Okay okay, so maybe technically the avalanche duration may be given by avalanches that are *exactly* duration  $k$ , not greater than or equal to  $k$ . Fine, so we just define a new quantity

$$P(\mathbf{X}^k \neq \mathbf{0}) = c'_2(\lambda_2^k - \lambda_2^{k+1}) + c'_3(\lambda_3^k - \lambda_3^{k+1}) + \cdots + c'_N(\lambda_N^k - \lambda_N^{k+1}),$$

which yields the probability of an avalanche that is *exactly* duration  $k$ .

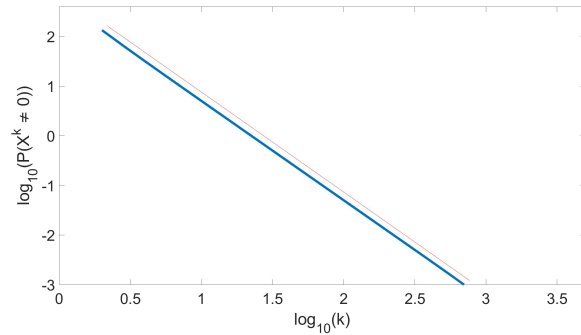


FIG. 3. Power law over three orders of magnitude using a clever sum of weighted exponentials.

and of course, picking clever weights,

Turns out, we can draw our eigenvalues from something other than a power law and still get the same phenomena. What seems to be important is that the eigenvalues decay, that some are close to (but not equal to) 1, and there actually already exists decent literature on approximating power laws with exponential functions. Apparently, "only a handful of exponentials are required in order to approximate a power law up to a given order of magnitude."

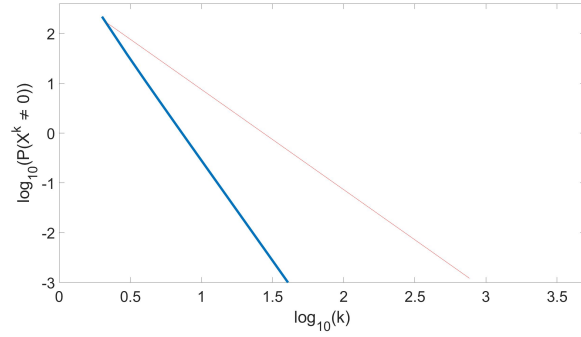


FIG. 4. Power law over three orders of magnitude using a clever sum of cleverly weighted exponentials.

Ultimately, what we demonstrate here is that power laws in stochastic systems that permit a probability matrix with a zero attractor state most likely arise from a clever selection of eigenvalues and weights.

### C. Basic Validation Using Computational Model

To demonstrate the math is true, we run a simple simulation of stochastic binomial neuronal avalanches for 3 neurons, and plot the probability of survival *versus* avalanche duration. Then, we will take the eigenvalues of the probability matrix and fit this data. As can be seen, and as expected, the probabilities propagated by the eigenvalue powers of the probability matrix match the simulated avalanche duration lengths (over 50000 trials) quite well. The

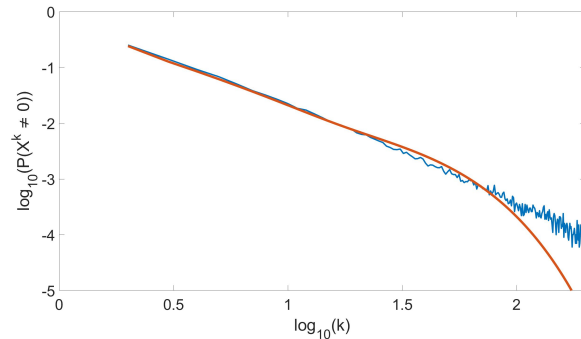


FIG. 5. Stochastic binomial spiking model avalanche probability *versus* duration against predicted.

discrepancy at the tail end arises as a result of... well, probably that we can't construct an infinite by infinite state matrix, or we're estimating the coefficients incorrectly.

### D. Constructing Systems of Arbitrarily Good Power Law Avalanche Duration

So now that we have some intuition for what features of a probability matrix's eigenvalues yields power laws in avalanche durations, we next seek to contrive systems that yield arbitrarily good power laws. As a demonstration, we

create a 9 state Markov process, with probability update rules governed by

$$\mathbb{P} = \begin{bmatrix} 1 & 0.0000 & 0.0000 & 0.0000 & 0.0001 & 0.0000 & 0.0000 & 0.0008 & 0.3541 \\ 0 & 0.9140 & 0.0000 & 0.0000 & 0.0004 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0 & 0.0001 & 0.0000 & 0.0000 & 0.0452 & 0.0001 & 0.0000 & 0.0014 & 0.0000 \\ 0 & 0.0000 & 0.0000 & 0.9540 & 0.0000 & 0.0000 & 0.0003 & 0.0046 & 0.0000 \\ 0 & 0.0000 & 0.0000 & 0.0000 & 0.9526 & 0.0000 & 0.0000 & 0.0000 & 0.0000 \\ 0 & 0.0000 & 0.1209 & 0.0000 & 0.0000 & 0.9992 & 0.0076 & 0.0000 & 0.0001 \\ 0 & 0.0385 & 0.0045 & 0.0457 & 0.0000 & 0.0000 & 0.2838 & 0.0000 & 0.0000 \\ 0 & 0.0475 & 0.0413 & 0.0002 & 0.0018 & 0.0007 & 0.0000 & 0.9932 & 0.0003 \\ 0 & 0.0000 & 0.8333 & 0.0001 & 0.0000 & 0.0000 & 0.7084 & 0.0000 & 0.6455 \end{bmatrix}.$$

Yes, I understand that looks a bit large, but bear with me. This matrix was chosen to have eigenvalues best approximating a power law, namely

$$\lambda = [0.1353 \ 0.6065 \ 0.8825 \ 0.9692 \ 0.9922 \ 0.9980 \ 0.9995 \ 0.9999 \ 1].$$

Here's the payoff

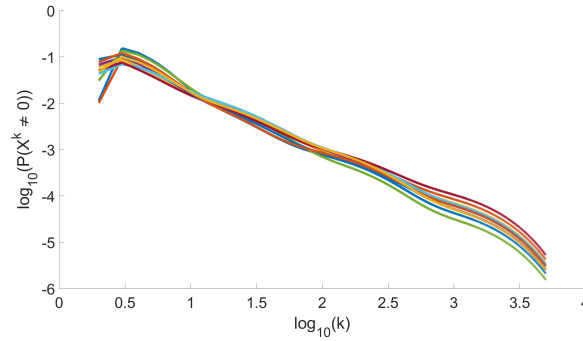


FIG. 6. Loglog plot of avalanche probability *versus* avalanche duration over 3 decades, 10 initial conditions.

As we can see, our system of 9 states has a pretty beautiful power law distribution of avalanche duration over 3 decades, which is apparently "enough." What's particularly interesting is that our system *literally* has 9 states, one which we consider the zero attractor state indicating the "end" of the avalanche. This power law is also fairly robust to the initial starting probability vector.

### E. Summary Statements

- All stochastic processes with one zero attractor state *must* permit a probability matrix  $\mathbb{P}$ .
- Any such system fundamentally evolves in time as a sum of weighted exponentials.
- Power laws in avalanche duration *must* arise as clever sum of cleverly weighted exponentials.
- Small, clever systems (e.g. 9 states) demonstrate arbitrarily good power laws over more than 3 decades.





### III. HOW TO POORLY CONSTRUCT $\mathbb{P}$

Now, the reader is thinking, “this is all fine and good, author, but how can I reliably construct these probability matrices?” Great question, reader, great question. And you’re not really going to like the answer.

#### A. Really Badly

Suppose you want your probability matrix to have the following eigenvalues

$$\lambda = [1 \ 0.9991 \ 0.9973 \ 0.9918 \ 0.9756 \ 0.9286 \ 0.8007 \ 0.5134],$$

which are eigenvalues that will yield a nice power law for 3 decades. Well, just make these elements the diagonals of your probability matrix, and let the remaining probabilities decay to zero...

$$\mathbb{P}_{bad} = \begin{bmatrix} 1 & 1 - \lambda_2 & 1 - \lambda_3 & \cdots & 1 - \lambda_N \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ 0 & 0 & \lambda_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_N \end{bmatrix}.$$

Great. So let’s push this concept a little further by now considering a system with only 5 states. Not neurons, states.

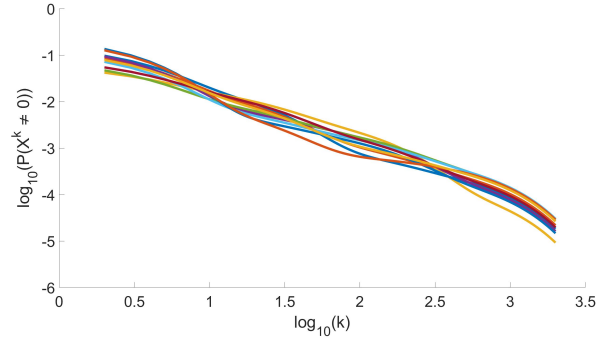


FIG. 7. Loglog plot of avalanche probability *versus* avalanche duration for 5 states, 10 initial conditions.

#### B. Not As Badly

It can easily be shown that an upper triangular matrix has eigenvalues equal to the diagonal entries. Hence,

$$\mathbb{P}_{less-bad} = \begin{bmatrix} 1 & p_{21} & p_{31} & \cdots & p_{N1} \\ 0 & \lambda_2 & p_{32} & \cdots & p_{N2} \\ 0 & 0 & \lambda_3 & \cdots & p_{N3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_N \end{bmatrix}, \quad \sum_{i=1}^N p_{ji} = 1 - \lambda_j.$$

This is much better than what we had before. in fact, we’ve filled in half of the probability matrix, which is more than we can say about the bipartite networks we’ve dealt with in the past... the figure looks the same as the previous case with just diagonal and zero state entries. So at the end of the day, not the worst possible transition matrices. we can be seeing how we can construct  $\mathbb{P}$  with desired eigenvalues using local probability rules that only define probabilities between states.

### C. Okay, But With More Difficulty

It can be similarly easily shown that for

$$\mathbb{P}_{okay} = \begin{bmatrix} P_{11} & P_{21} & P_{31} & \cdots & P_{N1} \\ 0 & P_{22} & P_{32} & \cdots & P_{N2} \\ 0 & 0 & P_{33} & \cdots & P_{N3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & P_{NN} \end{bmatrix},$$

the eigenvalues of  $\mathbb{P}_{okay}$  are just the set of eigenvalues for  $P_{ii}$ . So, we capture more entries of our matrix, but the eigenvalues of  $\mathbb{P}_{okay}$  are no longer determined by single probabilities, but instead by matrices. We will now see why this extension is necessary to study any meaningful stochastic system.



#### IV. POORLY LINKING LOCAL DYNAMICAL UPDATES TO GLOBAL PROBABILITY

The reader mumbles “what a load of mumbo jumbo, who cares about these global probability properties, I’m not God, I can only change a few parameters, not individual entries in the whole probability matrix!” We hear you, reader, we hear you. Unfortunately, we are also not deities, so here’s some contrived mortal intuition instead.

##### A. Ring Lattice

Suppose we have a very simple system of  $n$  neurons that spike some integer number of times, and are connected cyclically in a directed lattice. Then the update rules are exceedingly simple:

$$\begin{aligned} X_j^{t+1} &\sim B(X_{j-1}^t, a_{ji}) \\ X_1^{t+1} &\sim B(X_n^t, a_{1n}). \end{aligned}$$

This simplification yields an exceedingly useful property: at any time  $t$ , if there are  $\sum_i x_i^t = s^t$  spikes, then  $s^{t+1} \leq s^t$ . This is because you can never flip  $m$  coins and get *more* than  $m$  successes. Why is this property useful? Well, let us again enumerate all possible states of our system (of which there are an infinite number), but this time partition them according to the total number of spikes.

$$S_0 = \left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\}, \quad S_1 = \left\{ \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix} \right\}, \quad S_2 = \left\{ \begin{bmatrix} 1 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \dots, \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 2 \end{bmatrix} \right\}, \quad \dots$$

and we see that  $P(\mathbf{x}^{k+1} \in S_{m+1} | \mathbf{x}^k \in S_m) = 0$ . Hence, we have our block diagonal matrix  $\mathbb{P}_{okay}$ , where each block matrix  $P_{ii}$  corresponds to the probability of going from any state in  $S_i$  to any other state in  $S_i$ . All transitions from  $S_i \rightarrow S_{i-1}$  don’t contribute to the eigenvalues due to the block triangular matrix property outlined previously.

Now, in order for  $\mathbf{x}^t, \mathbf{x}^{t+1} \in S_i$ , we actually require that every spike successfully transmits. We also realize that given a successful outcome, there is only one possible next state in  $S_i$ , which is simply the rotation matrix  $R$  acting on  $\mathbf{x}_i$  such that

$$\mathbf{x}^{t+1} = \begin{bmatrix} 0 & 0 & \dots & 0 & 1 \\ 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \mathbf{x}^t.$$

Hence, we see that we can further divide  $S_i$  into sets of period  $n$ . If  $n$  is prime, then we can avoid having smaller groups with harmonic rotation periods. Hence,  $P_{ii}$  becomes a block diagonal matrix

$$P_{ii} = \begin{bmatrix} Q_{11} & 0 & \dots & 0 \\ 0 & Q_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & Q_{mm} \end{bmatrix},$$

where each diagonal block  $Q_{ii}$  is an  $n \times n$  matrix of states that are all rotational transformations of each other. These matrices  $Q_{ii}$  now become the diagonal blocks of  $\mathbb{P}$ , such that the eigenvalues of  $\mathbb{P}$  are the set of eigenvalues of all  $Q_{jj}$  for all  $P_{ii}$ . Finally, we realize that each  $Q_{ii}$  is just the same rotation matrix  $R$  with entries weighted by the probability of successfully transmitting all spikes. The eigenvalues of such a rotation matrix all have the same magnitude, given by the  $n$ th root of the product of its entries. In other words,

$$|\lambda(Q_{ii})| = (q_{12}q_{23} \dots q_{n-1n}q_{n1})^{1/n} = (a_{21}^m a_{32}^m \dots a_{nn-1}^m a_{1n}^m)^{1/n}.$$

So what are these  $q_{ij}$ ? For  $\mathbf{q}_i, \mathbf{q}_j \in Q_{kk}$ ,  $Q_{kk} \in P_{mm}$ , we have  $q_{ij} = P(\mathbf{x}^{t+1} = \mathbf{q}_j | \mathbf{x}^t = \mathbf{q}_i)$ . Because these are rotation sets, every neuron is hit  $m$  times for states with a total of  $m$  spikes, so  $|\lambda(Q_{ii})| = (a_{21}^m a_{32}^m \dots a_{nn-1}^m a_{1n}^m)^{1/n} = (a_{21} \dots a_{1n})^{m/n}$  with multiplicity  $|S_m|$ , which, well, can be determined using ideas from mathematics under “integer partitions”.

### 1. Simple Case: 2 Neurons

In a system of two neurons that are connected to each other (with  $a = a_{12}a_{21}$ ) by

$$A = \begin{bmatrix} 0 & a_{21} \\ a_{12} & 0 \end{bmatrix},$$

we have states

$$S_0 = \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}, S_1 = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right\}, S_2 = \left\{ \begin{bmatrix} 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ 2 \end{bmatrix} \right\}, S_3 = \left\{ \begin{bmatrix} 3 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 3 \end{bmatrix} \right\}, \dots,$$

with eigenvalues determined by

$$\lambda = [1 \ a^{1/n} \ a^{1/n} \ a^{2/n} \ a^{2/n} \ a^{2/n} \ a^{3/n} \ a^{3/n} \ a^{3/n} \ \dots],$$

which is, let's be honest, kind of tricky to match to a specific distribution of eigenvalues. This fact also says something very intimate about the nature of avalanche sizes in ring lattices, which is that because  $a = \prod_{i=1}^n a_{ii+1}$ , the size of the lattice doesn't really change the underlying exponents of system evolution, it simply changes the number of exponents.

### 2. Less Simple Case: $n$ Neurons

Just to explore this concept a little further, suppose we now have a system of  $n$  neurons. What are the possible enumeration of states? Well,  $|S_0| = 1$  always.

$$\begin{aligned} |S_1| &= n \\ |S_2| &= \binom{n}{2} + n \\ |S_3| &= \binom{n}{3} + 2 \binom{n}{2} + n, \dots, \end{aligned}$$

etc. As can be seen, the number of eigenvalues representing states of size  $1, 2, \dots$  become combinatorially huge. All that to say, we can design the eigenvalues of  $\mathbb{P}$  for ring lattices, but we only really have two tuning parameters ( $a$  and  $n$ ), which don't easily permit careful eigenvalue distribution tuning. A super interesting caveat of this phenomena is that these eigenvalues are the *true* eigenvalues of the probability matrix  $\mathbb{P}$ . Hence, these results hold true for very very large rings.

## B. Feed-Forward Network

The reader will now exasperatedly admit that yes indeed we have found a matrix for which we can directly relate the matrix entries to the eigenvalues of the probability matrix. Out of sheer hubris and desire to publish (and, of course, the altruistic motive of intellectual curiosity), the reader demands another example.

### 1. Trivial Case: 1 Layer

Unsurprisingly, a feed-forward network of 1 layer does not really display any interesting dynamics. At any initial state at any time, all activity decays to zero. What does this mean? This means that if the first row of  $\mathbb{P}$  corresponds to the zero state, then  $\mathbb{P}$  is a zero matrix, with all ones in the first row

$$\mathbb{P}_1 = \begin{bmatrix} 1 & \dots & 1 \\ 0 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 0 \end{bmatrix}.$$

## 2. Less Trivial Case: 2 (and eventually $n$ ) Layers

With two layers, we can now find an interesting partitioning of states. Intuitively, in any state where the first layer of neurons fired 0 times, any future state cannot have any firing in the first layer. In this way, we define

$$S_2 = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}, \quad S_1 = \begin{bmatrix} \mathbf{0}_1 \\ z_2 \end{bmatrix}, \quad S_0 = \begin{bmatrix} \mathbf{0}_1 \\ \mathbf{0}_2 \end{bmatrix}$$

as our partitions, where  $z_i$  represents the number of spikes of neurons in layer  $i$  excluding the zero state  $\mathbf{0}_i \notin z_i$ . We note a very interesting property here, where the probabilities  $P(\mathbf{x}^{k+1} \in S_i | \mathbf{x}^k \in S_{j \leq i}) = 0$ . This means that even the diagonal block elements of  $\mathbb{P}$  are also zero.

$$\begin{aligned} \mathbb{P} &= \begin{bmatrix} P(S_0 \rightarrow S_0) & P(S_1 \rightarrow S_0) & P(S_2 \rightarrow S_0) \\ P(S_0 \rightarrow S_1) & P(S_1 \rightarrow S_1) & P(S_2 \rightarrow S_1) \\ P(S_0 \rightarrow S_2) & P(S_1 \rightarrow S_2) & P(S_2 \rightarrow S_2) \end{bmatrix} \\ &= \begin{bmatrix} 1 & P(S_1 \rightarrow S_0) & P(S_2 \rightarrow S_0) \\ 0 & 0 & P(S_2 \rightarrow S_1) \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

Here, we have stumbled upon a most unfortunate situation, where this system is actually composed of nontrivial *Jordan Blocks*. Jordan blocks are super bad news bears for us, and can be directly interpreted as divine retribution for our hubris.





## V. STATE-FEEDBACK FOR DETERMINING GOOD INITIAL PROBABILITY DISTRIBUTIONS

So this might go a little fast, but bear with me. Recall that we have some rule

$$\mathbf{X}^{k+1} \sim D(\mathbf{X}^k, A),$$

where the states at the next time step are stochastically drawn from some distribution with some adjacency matrix parameters  $A$ , and given an enumeration of system states  $\mathbf{x}_i \in X$ , where  $X$  is the set of all states, and  $|X| = N$ , we can construct probability matrix  $\mathbb{P} \in \mathbb{R}^{N \times N}$  with some zero attractor state  $\mathbf{x}_0$  such that  $p(\mathbf{x}_0 \rightarrow \mathbf{x}_0) = 1$ . Then given some initial probability distribution of states  $\mathbf{p}_0$ , we have

$$\mathbf{p}^k = \mathbb{P}^k \mathbf{p}_0,$$

and as always, we care about the eigenvalue distribution of  $\mathbb{P}$ . Suppose we consider the new following system

$$\mathbf{q}^k = \mathbb{P}^{kT} \mathbf{q}_0,$$

such that the first *row* of  $\mathbb{P}$  has a 1 at the first entry, and zeros for the rest of the entries. Next, consider the following state feedback problem

$$\mathbf{q}^{k+1} = \mathbb{P}^T \mathbf{q}^k + B \mathbf{b}^k,$$

where  $\mathbf{b}^k$  is some control input. If we pick  $\mathbf{b}^k = -K \mathbf{q}^k$ , such that our input is a linear combination of system states, then we get

$$\mathbf{q}^{k+1} = (\mathbb{P}^T - BK) \mathbf{q}^k,$$

which is our classic state feedback problem. It turns out that given this state feedback, given our system  $(\mathbb{P}^T, B)$  is controllable, we can arbitrarily place the eigenvalues of  $\mathbb{P}^T - BK$ . Hence, if we pick  $B$  with all zeros except for the very first entry, then  $BK$  only modifies the first row of  $\mathbb{P}^T$ . This modification of the first row of  $\mathbb{P}^T$  corresponds directly to picking the transition probabilities *leaving* the zero state, which can be thought of as the picking the initial probability vector  $\mathbf{p}_0$ . This basically changes our global probability propagation dynamics to

$$\mathbf{p}^k = (\mathbb{P} - (BK)^T) \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where we can place the eigenvalues of  $\mathbb{P} - (BK)^T$  where we want using state-feedback.