COMPUTING LIMITING STATIONARY DISTRIBUTIONS OF SMALL NOISY NETWORKS

Fred Richman and Katarzyna Winkowska-Nowak
Florida Atlantic University
Boca Raton, FL 33431

Abstract

The dynamics of opinion transformation is modeled by a neural network with a nonnegative matrix of connections. Noise is introduced at each site, and the limit of the stationary distributions of the resulting Markov chains as the noise goes to zero is taken as an indication of what configurations will be seen. An algorithm for computing this limit is given, and a number of examples are worked out. Some of the mathematical ideas developed, such as visible states, time scales, and a calculus of indexed probabilities, are of independent interest.

1 Introduction

The mathematics developed in this paper was inspired by models of the statistical mechanics of social interaction [6]. One considers a finite group of individuals who can influence each other's opinion depending on how close they are to each other and on how persuasive they are. An individual's opinion at any time will depend on his opinion, and that of his neighbors, at the previous time. For simplicity we limit the possible opinions of the individuals to two. The interest is in long term collective phenomena. In particular, the possible persistence of minority opinion in spite of the fact that the most stable configurations occur when everyone in the group holds the same opinion.

Our models bear a superficial resemblance to voter models [1]. In both the individuals are in one of two states and are influenced by the states of their neighbors. In both one is interested in the existence of inhomogeneous stationary distributions. However, voter models normally involve infinite homogeneous networks, fixed probabilistic transitions, and continuous time. Our networks are finite, often small, and usually inhomogeneous. The transitions are based on a deterministic discrete stepping which is modified by injecting noise, and we analyze the behavior of the systems as the noise goes to zero.

Of course it is easy to construct deterministic models where minority opinion persists. If the influence of each individual is equal to zero, then any distribution of opinion will persist throughout all time. Similar things happen for models that include limited noise if there are configurations that cannot be altered by that noise. For this reason, and because a natural way to insert noise is to add a normally distributed random variable to the influences determining an individual's opinion, we consider noisy models where, for any level of noise, it is possible to go from each configuration to any other.

What does it mean for the majority opinion always to prevail in such a situation? Because of the nature of the noise, there will always be times when there is some minority opinion. Moreover, the majority opinion itself will constantly change, if for no other reason than that all individuals will be struck by a revelation, in the form of an extremely rare occurrence of noise, that will change all their opinions simultaneously. The first problem is to formulate a mathematical condition that accurately reflects the statement that majority opinion always prevails.

For a fixed level of noise, the configurations of opinion are states in a Markov chain. Because of the nature of the noise, this Markov chain is **irreducible** (recurrent, ergodic), that is, starting at any state, there is a positive probability of being in any other state at some time in the future—indeed there is a positive probability of being in any other state at the next step. Such a Markov chain has a unique stationary distribution, independent of what its initial state might be. The stationary distribution is the set of probabilities that the system will be in any given state after it has been running for a long time. Because the chain is irreducible, these probabilities are all nonzero. We focus attention on what happens to these probabilities as the level of noise goes to zero. The statement that majority opinion always prevails may be interpreted as the condition that the limit of the stationary distributions assigns zero probability to every state except the two where all the individuals have the same opinion.

If a state has a nonzero probability in the limiting stationary distribution we say that it is **visible**. What that means is that if you wait a long time and then look at the distribution of opinion, the probability that you will see that particular state is bounded away from zero for all noise levels. The persistence of minority opinion corresponds to the condition that there be visible states other than the homogeneous ones.

Our specific examples are much like cellular automata except that they need not be homogeneous: the individuals can have different powers of persuasion, different **strengths**. They fall under the general rubric of neural networks without negative influences. More generally they can be thought of as finite discrete dynamical systems with noise. These determine Markov processes which can be analyzed in the zero-noise limit. Here we must distinguish between the limiting process, which is usually quite easy to determine, along with its stationary distributions, and the limiting stationary distribution.

Different time scales play a big role in understanding what is going on. As the noise level goes to zero, various critical events happen infrequently. These events can be classified by how long you have to wait, on average, to see them, measured by the reciprocals of their probabilities. Certain events are rare, but others will happen only after thousands of those rare events occur. Of course the rarity depends on the actual noise level. To handle this phenomenon we develop a calculus of threshold expressions (Section 5) which stand for the probability that the noise variables exceed certain thresholds. They are not simply probabilities, but probabilities indexed by the noise level, and they are sorted into equivalence classes depending on their limiting behavior for small noise. Two (equivalence classes of) indexed probabilities represent the same time scale if each is a positive real multiple of the other.

Sections 2 to 6 are devoted to developing tools to calculate limiting stationary distributions, while the remaining sections present the calculations for a number of specific examples. In Section 6.1 we outline an algorithm, which has been implemented on a computer, that will perform these calculations. The implementation is limited to networks with at most 12 nodes. The last section lists a few unanswered questions.

2 Finite discrete dynamical systems with noise

Consider a finite discrete dynamical system, that is, a finite set S of states and a function $T : S \to S$. We can think of this as a Markov chain whose transition matrix $\mathbf{T}$ consists of 0's and 1's, with exactly one 1 in each row. Noise is introduced into the system by means of a finite number of independent identically distributed random variables X_i. The common density function of the X_i is of the form

$$\frac{1}{\sigma} f\left(\frac{x}{\sigma}\right)$$

where f is continuous, everywhere nonzero, symmetric around zero, and, for each $\theta > 1$,

$$\lim_{x\to\infty} \frac{f(\theta x)}{f(x)} = 0. \qquad (1)$$

The parameter σ represents the strength of the noise. If $\int_{-\infty}^{\infty} x^2 f(x)dx$ is finite, then we might as well assume that $\int_{-\infty}^{\infty} x^2 f(x)dx = 1$, in which case σ^2 is the variance of X_i.

The main example of such an f is the **standard normal density function**

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$

In this case X_i is a normal random variable with mean 0 and variance σ^2. An example of a function f that doesn't satisfy Condition 1 is $f(x) = \frac{\pi}{1+x^2}$ where the limit is $1/\theta^2$. For a distribution with finite variance, consider $f(x) = \frac{\sqrt{2}/\pi}{1+x^4}$ which has variance 1 and the limit is $1/\theta^4$.

Another example is the **logistic distribution** $1/(1+\exp(-2\beta h))$ which has been prominent in the study of neural networks (see [4, page 27]). The variance of this distribution is $\sigma^2 = \pi^2/(12\beta^2)$. Its density function, normalized to have variance 1, is

$$f(x) = \frac{\pi/\sqrt{12}}{2\cosh^2(\pi x/\sqrt{12})}$$

where we have set the parameter β equal to $\pi/\sqrt{12}$.

The noise changes the transition matrix $\mathbf{T}$ by replacing the 0's by (typically small) positive numbers, and changing the 1's so that the matrix remains stochastic (the sum of the elements in each row is 1). Denote the resulting transition matrix by $\mathbf{A}(\sigma)$. We will require the entries of $\mathbf{A}(\sigma)$ that replace the 0's of $\mathbf{T}$ to be products of terms of the form $P(X_i > r_i)$ and $P(X_i < r_i)$. A simple, and typical, example is provided by a bounded cellular automaton on a rectangular grid where each site can take either the value $+1$ or the value -1. Here is a state of a 5×7 example.

$$\begin{array}{ccccccc}
+ & - & + & + & + & - & + \\
- & + & + & + & - & + & + \\
- & - & - & + & + & - & + \\
+ & + & - & + & - & - & - \\
+ & - & + & - & - & + & -
\end{array}$$

The state space S consists of the 2^{35} different patterns. The stepping rule is that the value of the site at time $t+1$ is determined by the values of it and its (up to four) nearest neighbors at time t in accordance with majority rule—we simply add the values at those sites, and use the resulting sign. In case of a tie, where the sum is zero, there is no change. Noise is added by letting the random variable X_i contribute to the vote at the i-th site. Thus the threshold r_i will be the negative of the vote at site i without the noise.

Let $P_r = \frac{1}{\sigma}\int_r^\infty f\left(\frac{x}{\sigma}\right) dx$ be the probability that the noise exceeds r. Because f is symmetric, $P_r + P_{-r} = 1$. The crucial condition for our analysis is that, if $0 \le r < s$, then

$$\frac{P_s}{P_r} = \frac{\int_s^\infty f\left(\frac{x}{\sigma}\right) dx}{\int_r^\infty f\left(\frac{x}{\sigma}\right) dx} = \frac{\int_{\frac{s}{\sigma}}^\infty f(x)\, dx}{\int_{\frac{r}{\sigma}}^\infty f(x)\, dx} \to 0 \tag{$*$}$$

as $\sigma \to 0$. That is to say, whatever the difference in noise thresholds required for two transitions, if the noise is sufficiently weak, then the probability is overwhelming in favor of the more likely transition. In particular, taking $r = 0$, we see that P_s converges to 0, 1 or 1/2, as σ goes to 0, depending on whether $s > 0$, $s < 0$, or $s = 0$. So there is a well-defined limiting transition matrix $\mathbf{A}(0^+)$, whose nonzero entries are powers of $1/2$.

It's easy to see that if f satisfies Condition 1, then f satisfies this condition. Indeed if $\theta = s/r$ and $n = r/\sigma$, then

$$\frac{\int_{\frac{s}{\sigma}}^\infty f(x)\, dx}{\int_{\frac{r}{\sigma}}^\infty f(x)\, dx} = \frac{\int_{\theta n}^\infty f(x)\, dx}{\int_n^\infty f(x)\, dx} = \frac{\theta \int_n^\infty f(\theta x)\, dx}{\int_n^\infty f(x)\, dx} \le \theta \sup\left\{\frac{f(\theta x)}{f(x)} : x \ge n\right\}.$$

For the normal distribution we can get a more explicit estimate of P_s/P_r. If $f(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$, then

$$f(x)\left(\frac{1}{x} - \frac{1}{x^3}\right) < \int_x^\infty f(\xi)d\xi < f(x)\frac{1}{x}$$

for $x > 0$ (see [2, page 175]) so

$$\frac{P_s}{P_r} \sim \exp\left(-\frac{s^2 - r^2}{2\sigma^2}\right)\frac{r}{s}$$

For the logistic distribution we get

$$\frac{P_s}{P_r} \sim \exp\left(-\frac{s - r}{\sigma\sqrt{3}/\pi}\right)$$

We thus have a family of Markov processes indexed by $\sigma \in (0, \infty)$, which we may identify with a family of stochastic matrices $\mathbf{A}(\sigma)$. The matrix $\mathbf{T}$ associated with the dynamical system may be identified with $\mathbf{A}(0)$. If some of the noise thresholds are 0, the limit $\mathbf{A}(0^+)$ of $\mathbf{A}(\sigma)$ as σ approaches 0 from above need not be the same as the matrix $\mathbf{A}(0)$. For example, take the 1×2 case of the cellular automaton above. This has four states $--$, $-+$, $+-$, and $++$ which are all fixed points (by our tie convention). So the matrix $\mathbf{A}(0)$ is the identity,

$$\mathbf{A}(0) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

while

$$\mathbf{A}(0^+) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 1/4 & 1/4 & 1/4 & 1/4 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

We will denote the entries of a matrix $\mathbf{A}$ by a_{ij}, and the entries of $\mathbf{A}^m$ by $a_{ij}^{(m)}$. A state i is called **stable** if $a_{ii}(0^+) = 1$, so the states $--$ and $++$ above are stable. A state i is called **transient** for $\mathbf{A}(0^+)$ if there exists a state j and a positive integer m such that $a_{ij}^{(m)}(0^+) > 0$, and $a_{ji}^{(n)}(0^+) = 0$ for all n. So the states $-+$ and $+-$ are transient states for $\mathbf{A}(0^+)$.

As each $\mathbf{A}(\sigma)$ is irreducible (in fact, all of its entries are nonzero), each has a unique stationary distribution $\mathbf{p}(\sigma)$, a probability distribution on the states such that $\mathbf{p}(\sigma)\mathbf{A}(\sigma) = \mathbf{p}(\sigma)$. We want to calculate the limit $\mathbf{p}(0^+)$ of $\mathbf{p}(\sigma)$ as $\sigma \to 0$. In this limiting distribution, the number p_i represents the probability that, if you look at the system after it has been running for a long time at a low noise level, you will see the state i. We say that the state i is **visible** if $p_i > 0$, **invisible** if $p_i = 0$. It's pretty clear (and is a consequence of Theorem 1 below) that transient states are invisible. We shall see that stable states can also be invisible.

3 The influence matrix

We are interested in dynamical systems that represent distributions of opinions among m individuals (nodes, sites). The **states** of the system are m-vectors, indexed by those individuals, whose entries are ± 1, representing the two different opinions that an individual might have. The dynamics are specified by an **influence matrix**, an $m \times m$ matrix $\mathbf{C}$ whose entry c_{ij} indicates the influence exerted by the j-th node on the i-th node. The matrix $\mathbf{C}$ operates on a state vector $\mathbf{v}$ by transforming it into a vector $\mathbf{v}'$ according to the stepping equation

$$\mathbf{v}' = \operatorname{sgn} \mathbf{C}\mathbf{v}$$

where we set $v_i' = v_i$ if this calculation results in $v_i' = 0$. In the 1×2 case of the cellular automaton above, the influence matrix is simply

$$\mathbf{C} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

and

$$v_i' = \operatorname{sgn}(v_1 + v_2)$$

if $v_1 + v_2 \neq 0$, so, in fact, $v_i' = v_i$ always. The 1×3 case is a little more interesting. Here the influence matrix is

$$\mathbf{C} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 1 \end{pmatrix}.$$

We think of $\mathbf{Cv}$ as representing the **local opinion field** at each site i, the individual assuming the sign of that field, or remaining unchanged if the field is zero. The behavior when the field is zero turns out to be unimportant because it will be an event of probability zero. In fact, were we to allow probabilities to enter already at this stage, we would be inclined to flip a coin to determine the value of v_i' in case the local opinion field at i is zero.

This is a description of a neural network, the matrix $\mathbf{C}$ being the **matrix of connections**. In a general neural network, there is also a vector $\mathbf{b}$ of thresholds, and the stepping equation is $\mathbf{v}' = \operatorname{sgn}(\mathbf{Cv} - \mathbf{b})$. For us, $\mathbf{b} = 0$.

To add noise to the system, we change the stepping equation to

$$\mathbf{v}' = \operatorname{sgn}(\mathbf{X} + \mathbf{Cv})$$

where $\mathbf{X}$ is a noise vector. Notice that the negative mean of the noise $\mathbf{X}$ could serve as a vector of thresholds. We will only consider noise with mean 0.

Normally all the entries of $\mathbf{C}$ are nonnegative, and the nonzero entries in column i are all equal—the common value is called the **strength** of the i-th node. The interactions are normally symmetric in the sense that $c_{ij} \neq 0$ if $c_{ji} \neq 0$. The number c_{ii} is called the **self-support** of the i-th node.

4 Invisible states

Let S be the set of states of a finite Markov chain with transition matrix $A = (a_{ij})$. Recall that a subset I of S is **closed** if $a_{ij} = 0$ whenever i is in S and j is not in S. Each state is either transient or is in a unique minimal closed subset. There is a unique stationary distribution on each minimal closed set, and the stationary distributions on S are convex combinations of these unique distributions.

We start with a simple criterion for a state of a chain to have a small probability of being seen.

Theorem 1 *Let $\mathbf{A} = (a_{ij})$ be the transition matrix of a finite Markov chain with a stationary probability distribution $\mathbf{p}$. Let k be a state and I a set of states not containing k. If*

$$\delta = \sum_{i \in I} a_{ki} \text{ and } \varepsilon = \max_{i \in I} \sum_{j \notin I} a_{ij},$$

then $p_k \leq \varepsilon/(\varepsilon + \delta)$.

Proof. If k is a transient state, then $p_k = 0$. If k is a recurrent state, we may assume that $\mathbf{p}$ is the unique stationary distribution supported by the minimal closed set of states containing k. Modify $\mathbf{A}$ by replacing each transition to a state not in I by a transition to k. This doesn't change δ or ε, and can't decrease p_k. The set $I \cup \{k\}$ is now closed, so all other states may be ignored. For $i \in I$, we have $a'_{ik} = \sum_{j \notin I} a_{ij} \leq \varepsilon$. Thus if we combine all the states in I into one state, 2, and consider the Markov chain with transition matrix

$$\begin{pmatrix} 1-\delta & \delta \\ \varepsilon & 1-\varepsilon \end{pmatrix},$$

then $p'_1 \geq p_k$. But

$$p'_1 = \frac{\varepsilon}{\varepsilon + \delta}.$$

■

Let $\mathbf{A} = (a_{ij})$ be the transition matrix of a Markov chain, and δ a positive number. By a **δ-path** from a state i to a state j, we mean a sequence of states $i = k_1, k_2, \ldots, k_m = j$ so that $a_{k_t k_{t+1}} \geq \delta$ for $1 \leq t < m$.

Lemma 2 *Let n be a positive integer, and $r > 1$ a real number. There exists a real number K such that for any Markov chain with n states, and positive number $\delta \leq 1/rn$, if the states can be partitioned into two sets $S_0 \cup S_1$ so that*

- *from each state in S_1, there exists a δ-path to a state in S_0, and*
- *$a_{ij} < r\delta$ whenever $i \in S_1$ and $i \neq j$,*

then the mean time to go from any state in S_1 to some state in S_0 is at most K/δ.

Proof. The only transition probabilities a_{ij} that matter are those for which $i \in S_1$, so we may assume that S_0 consists of a single state 0 with $a_{00} = 1$. First let $\delta = 1/rn$. Clearly

$$a_{i0}^{(kn)} \geq 1 - (1 - \delta^n)^k$$

for each k, so, as $a_{i0}^{(m)}$ increases with m,

$$a_{i0}^{(m)} \geq 1 - (1 - \delta^n)^{\lfloor m/n \rfloor} \geq 1 - bc^m$$

for all m, where $c = (1 - \delta^n)^{1/n}$ and $b = 1/(1 - \delta^n)$. Let $\mathbf{Q}$ be the matrix of the chain restricted to S_1. Then the entries of $\mathbf{Q}^m$ are bounded by bc^m, so the entries of

$$\mathbf{N} = (\mathbf{1} - \mathbf{Q})^{-1} = \mathbf{1} + \mathbf{Q} + \mathbf{Q}^2 + \cdots$$

are bounded by $K = b/(1 - c)$.

The point of looking at the special case $\delta = 1/rn$ is so that K will depend only on n and r, and not on δ or on any other property of the transition matrix $\mathbf{A}$.

Now consider arbitrary $\delta \leq 1/rn$. Let $\lambda = 1/rn\delta$, and let $\mathbf{A}' = \mathbf{1} + \lambda(\mathbf{A} - \mathbf{1})$. Note that $\mathbf{A}'$ is a Markov matrix. The diagonal entries of $\mathbf{A}'$ will be positive because the off-diagonal entries of $\mathbf{A}$ are bounded by $r\delta$, so the off-diagonal entries of $\mathbf{A}'$ will be bounded by $1/n$. Then we are in the $\delta = 1/rn$ case, so the entries of $\mathbf{N}'$ are bounded by K. But

$$\mathbf{N}' = (\mathbf{1} - \mathbf{Q}')^{-1} = (\lambda(\mathbf{1} - \mathbf{Q}))^{-1} = \lambda^{-1}\mathbf{N}$$

so the entries in $\mathbf{N}$ are bounded by λK.

By [5, Theorem 3.2.4] the ij-th entry of $\mathbf{N}$ is the expected number of times the chain is in state j after it starts in state i. So the row sums of $\mathbf{N}$ give the mean time to reach S_0 starting from a state in S_1. These row sums are no greater than $\lambda K n = K/r\delta < K/\delta$. ■

To see the need for some hypothesis in Lemma 2 that bounds the off-diagonal elements, consider the Markov chain with transition matrix

$$\mathbf{A} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1-\delta & \delta \\ \delta & 1-\delta & 0 \end{pmatrix}.$$

The top state is absorbing, and we can get to it from either of the other two states with a δ-path of length at most 2. But

$$\mathbf{N} = (\mathbf{1} - \mathbf{Q})^{-1} = \begin{pmatrix} \delta & -\delta \\ \delta - 1 & 1 \end{pmatrix}^{-1} = \frac{1}{\delta^2} \begin{pmatrix} 1 & \delta \\ 1-\delta & \delta \end{pmatrix}$$

so the mean time to go from the bottom state to the top state is $1/\delta^2$ which can't be bounded by a constant over δ.

Theorem 3 *Let n be a positive integer, and $r > 1$ a real number. Suppose the n states of a Markov chain with transition matrix $\mathbf{A} = (a_{ij})$ contain two disjoint subsets S_0 and S_1 such that*

1. *there is a δ-path in $S_0 \cup S_1$ from each state in S_1 to S_0,*

2. *$a_{ij} < r\delta$ whenever $j \neq i \in S_1$,*

3. *$a_{ij} < \varepsilon$ for $i \in S_0$ and $j \notin S_0$.*

If $\mathbf{p}$ is a stationary distribution for $\mathbf{A}$, then $p_k < K\varepsilon/\delta$ for each k in S_1, where K is a constant depending only on n and r.

Proof. If k is a transient state, then $p_k = 0$. Otherwise we may assume that $\mathbf{p}$ is supported on the minimal closed subset containing k. We may eliminate the states that are not in $S_0 \cup S_1$ by absorbing them into k. The matrix $\mathbf{A}$ is modified by eliminating the rows and columns that correspond to states outside of $S_0 \cup S_1$, and setting

$$a'_{ik} = a_{ik} + \sum_{j \notin S_0 \cup S_1} a_{ij}.$$

The effect is to increase p_k, to decrease n, and to increase ε and r by at most a factor of n. Thus we may assume that $S_0 \cup S_1$ is a partition of the states, and Lemma 2 applies.

Because we are dealing with a Markov chain, each realization consists of the results of a sequence of independent trials, each consisting of transitions starting from the state k and ending the first time k is visited after visiting S_0. Let L be the number of steps in such a trial, and T the number of visits to k during the trial. Schematically, the results of a trial look like

$$k\ k\ \cdots\ k\ i\ k$$

where every visit to k is indicated, i is an element of S_0, and there may be unindicated elements of S_0 to the right of i, but none to the left. As these trials are concatenated, we include only the terminal points of each transition (so we omit the initial k). Because we have to go from S_0 to k, Condition 3 implies $E(L) > 1/\varepsilon$. As T is no greater than the number of steps until you reach S_0, we have $E(T) \leq K/\delta$ by Lemma 2. The number of visits to k during n trials is $T_1 + \cdots + T_n$, and the number of steps is $L_1 + \cdots + L_n$, so $(T_1 + \cdots + T_n)/(L_1 + \cdots + L_n)$ converges to p_k almost surely. Similarly $(T_1+\cdots+T_n)/n$ converges to $E(T)$ and $(L_1+\cdots+L_n)/n$ converges to $E(L)$, so $p_k = E(T)/E(L) < K\varepsilon/\delta$. ■

Returning to the example just before Theorem 3, suppose the transition matrix is

$$\mathbf{A} = \begin{pmatrix} 1-\varepsilon & \varepsilon & 0 \\ 0 & 1-\delta & \delta \\ \delta & 1-\delta & 0 \end{pmatrix}$$

where $\delta^2 \ll \varepsilon \ll \delta$. The stationary distribution is proportional to $(\delta^2, \varepsilon, \delta\varepsilon)$, so the middle state (which is in S_1) is seen the most no matter how small ε/δ is. The problem is the large off-diagonal term $1-\delta$.

5 Indexed probabilities: threshold expressions

We will be dealing with families of transition matrices $\mathbf{A}(\sigma)$ indexed by the positive real numbers, and we set $\mathbf{A}(0^+)$ equal to $\lim_{\sigma\to 0}\mathbf{A}(\sigma)$. We may think of such a family as a matrix of *indexed probabilities* (indexed by σ). By an **indexed probability** we mean a (continuous) function from the positive real numbers to the half-open interval $(0,1]$. In particular, each number in $(0,1]$ is a (constant) indexed probability.

We are interested in small probabilities determined by noise thresholds. It will be convenient to have a way of denoting the probabilities in terms of the thresholds. Let X denote the noise random variable, which depends on the parameter $\sigma > 0$, and is of the type described in Section 2. For r a real number, we denote by $[r](\sigma)$ the probability that $X > r$. More generally,

$$[r_1, r_2, \ldots, r_n](\sigma) = [r_1](\sigma)[r_2](\sigma)\cdots[r_n](\sigma)$$

denotes the joint probability that $X_i > r_i$, where the X_i are independent random variables with the same distribution as X. Thus $[r_1, r_2, \ldots, r_n]$ is an indexed probability—that is, $[r_1, \ldots, r_n]$ is a function. In practice, the r_i will be thresholds for the noise variables at various sites at various times. Note that $[0](\sigma) = 1/2$ and that $[r](0^+) = 0$ for $r > 0$. We call the indexed probability $[r_1, \ldots, r_n]$ an **elementary threshold expression**. The constant function 1 may be thought of as an elementary threshold expression with an empty list of thresholds.

A **threshold expression**, in general, is a linear combination of elementary threshold expressions with positive coefficients. An example is $[2,4] + (1/2)[1,1,3]$. Notice that threshold expressions are closed under addition and multiplication. Indeed, the threshold expressions might best be thought of as polynomials, with positive coefficients, in expressions of the form $[r]$. For normal and logistic noise, we will show that any threshold expression is equivalent to a positive multiple of an elementary one (any polynomial is equivalent to a monomial). We now define the appropriate notion of equivalence.

Let a and b be indexed probabilities. As we are interested only in the values of $a(\sigma)$ and $b(\sigma)$ for σ near zero, we write

$$a \equiv b \text{ if } \lim_{\sigma \to 0} \frac{a(\sigma)}{b(\sigma)} = 1.$$

This means that a and b are essentially interchangeable as far as their behavior in the zero-noise limit is concerned. Note that if $r < 0$, then $[r](0^+) = 1$, so $[r] \equiv 1$, the constant function 1. For this reason, we are not usually concerned with negative thresholds. Also, $[0] \equiv 1/2$, so we could get along without zero thresholds.

Under this equivalence relation, an indexed probability a such that $a(0^+) = 0$ behaves much like an infinitesimal. For example, $1 + a \equiv 1$. Indeed, many different levels of smallness may be defined as follows. If a and b are indexed probabilities, then write

$$a \prec b \text{ if } \lim_{\sigma \to 0} \frac{a(\sigma)}{b(\sigma)} = 0,$$
$$a \preceq b \text{ if } \lim_{\sigma \to 0} \frac{a(\sigma)}{b(\sigma)} \text{ exists.}$$

Note that $a \prec 1$ if and only if $a(0^+) = 0$, and that $a \prec b$ if and only if $a + b \equiv b$. The condition $(*)$ that we imposed on all noise distributions in Section 2 says that if $0 \le r < s$, then $[s] \prec [r]$. In particular, $[r] \prec 1$ whenever $r > 0$.

The preorder $a \preceq b$ gives rise to an equivalence relation $a \approx b$ defined to mean that $a \preceq b$ and $b \preceq a$, or, equivalently, that $a(\sigma)/b(\sigma)$ converges to a positive real number. The idea is that if $a \prec b$, then a and b represent essentially different **time scales** (in the zero-noise limit), while if $a \approx b$, then a and b represent essentially the same time scale. If you have an event that occurs with probability p, then you are apt to see it only if you observe on the order of $1/p$ trials. So if $a \prec b$, then for sufficiently small σ, you can pick a time period in which you will see lots of events that occur with probability $b(\sigma)$ but almost none that occur with probability $a(\sigma)$. Note that $a \approx b$ if and only if $a \equiv tb$ for some positive number t.

A finer partial order, which is consistent with the equivalence $a \equiv b$, is obtained by setting

$$a < b \text{ if } \lim_{\sigma \to 0} \frac{a(\sigma)}{b(\sigma)} < 1.$$

For normal and logistic noise, we will see that this gives a total order on the threshold expressions. If $a < b$ or $a \equiv b$, then a/b is an indexed probability, at least for small values of σ, which is all that matters. We have to perform such divisions when we eliminate invisible states (Corollary 7).

We start with an m-by-m transition matrix $\mathbf{A}(\sigma)$ whose entries are elementary threshold expressions

$$a_{ij}(\sigma) = [r_1, r_2, \ldots, r_m](\sigma).$$

Here $r_k = -f_k(i)s_k(j)$, where $f_k(i)$ is the field at node k when the system is in state i, and $s_k(j)$ is the sign of node k when the system is in state j. We will show that the limiting stationary distribution depends only on the equivalence classes of the indexed probabilities a_{ij}. To do this, it suffices to show that small percentage changes in the entries of a Markov matrix with all positive entries produce small percentage changes in its stationary distribution. First a simple lemma about logarithms of ratios.

Lemma 4 *Let c_i and d_i be positive numbers, $i = 1, \ldots, m$. If $|\log c_i/d_i| \le \delta$ for all i, then*

$$\left|\log \frac{\prod c_i}{\prod d_i}\right| \le m\delta \qquad \textit{and} \qquad \left|\log \frac{\sum c_i}{\sum d_i}\right| \le \delta$$

Proof. To prove the first inequality, just calculate

$$\left|\log \frac{\prod c_i}{\prod d_i}\right| = \left|\sum \log \frac{c_i}{d_i}\right| \le \sum \left|\log \frac{c_i}{d_i}\right| \le m\delta.$$

For the second inequality, the hypothesis says that $c_i = d_i e^{t_i}$ with $-\delta \leq t_i \leq \delta$, so

$$-\delta = \log \frac{\sum d_i e^{-\delta}}{\sum d_i} \leq \log \frac{\sum d_i e^{t_i}}{\sum d_i} \leq \log \frac{\sum d_i e^{\delta}}{\sum d_i} = \delta$$

and the middle term is $\log\left(\sum c_i / \sum d_i\right)$. ■

Theorem 5 *Let A and B be n-by-n Markov matrices with all nonzero entries, and (unique) stationary distributions p and q. If $|\log(a_{ij}/b_{ij})| \leq \delta$ for all i and j, then $|\log(p_i/q_i)| \leq 2(n-1)\delta$ for all i.*

Proof. Consider the matrix $M = A - I$, whose row sums are zero. Let $(-1)^{n-1}u_i$ be the cofactor of the i-th entry in the first column of M, so $(-1)^{n-1}(-1)^{i-1}u_i$ is the determinant of the submatrix of M formed by deleting the first column and the ith row. Then $u = (u_1, \ldots, u_n)$ is a vector in the left null space of M because $\det M = 0$. Each u_i is a sum of monomials of degree $n-1$ in the off-diagonal entries of A (all signs positive—we are indebted to Aaron Meyerowitz for pointing this out). Hence $u = p \sum_{i=1}^{n} u_i$.

Let v be the vector formed from the corresponding sums of monomials in the off-diagonal entries of B. From Lemma 4 (both inequalities) it follows that $|\log(u_i/v_i)| \leq (n-1)\delta$ so

$$|\log(p_i/q_i)| = \left| \log(u_i/v_i) + \log \frac{\sum_{i=1}^{n} v_i}{\sum_{i=1}^{n} u_i} \right| \leq 2(n-1)\delta.$$

■

One consequence of this is that we can eliminate negative thresholds from elementary threshold expressions, and so assume that the entries in the initial matrix $\mathbf{A}$ are either 1 or of the form $[r_1, r_2, \ldots, r_n]$ with each $r_k \geq 0$.

5.1 Normal and logistic noise

If the noise function is normal, and $r_1, r_2, \ldots, r_n$ are positive, then

$$[r_1, r_2, \ldots, r_n](\sigma) \sim \left(\frac{\sigma}{\sqrt{2\pi}}\right)^n e^{-\sum r_i^2/2\sigma^2} \frac{1}{\prod r_i}$$

for small σ. From this we see that

$$
\begin{array}{lll}
[r_1, r_2, \ldots, r_n] \prec [r'_1, r'_2, \ldots, r'_m] & & \text{if } \sum r_i^2 > \sum r_i'^2, \text{ or} \\
 & & \text{if } \sum r_i^2 = \sum r_i'^2 \text{ and } n > m. \\
[r_1, r_2, \ldots, r_n] \equiv \left(\prod \frac{r'_i}{r_i}\right) [r'_1, r'_2, \ldots, r'_n] & & \text{if } \sum r_i^2 = \sum r_i'^2.
\end{array}
$$

If some of the r_i are 0, we simply factor out a power of 1/2 and compare what's left. Thus we can compare these elementary threshold expressions fairly easily. Note that if none of the r_i and r'_i are 0, then $[r_1, r_2, \ldots, r_n] \approx [r'_1, r'_2, \ldots, r'_m]$ exactly when $n = m$ and $\sum r_i^2 = \sum r_i'^2$.

So for a normal noise function, the threshold expression $t[r_1, r_2, \ldots, r_n]$, with positive entries, is completely specified by the triple $\sum r_i^2$, n, and $t/\prod r_i$. Conversely, given any triple (u, v, w) of real numbers, with $w > 0$, and either $u > 0$ or $u = 0$ and $v > 0$, we get an indexed probability

$$\left(\frac{\sigma}{\sqrt{2\pi}}\right)^v e^{-\sum u/2\sigma^2} w$$

The triples (u, v, w) are linearly ordered by setting $(u_1, v_1, w_1) < (u_2, v_2, w_2)$ if

$$u_1 > u_2, \text{ or } u_1 = u_2 \text{ and } v_1 > v_2, \text{ or } u_1 = u_2 \text{ and } v_1 = v_2 \text{ and } w_1 < w_2$$

This ordering corresponds to the ordering $a \prec b$ of indexed probabilities. We multiply and add such triples, and multiply them by positive real numbers, as follows:

- $(a_1, b_1, c_1) \cdot (a_2, b_2, c_2) = (a_1 + a_2, b_1 + b_2, c_1 c_2)$
- $(a_1, b_1, c_1) + (a_2, b_2, c_2) = \begin{cases} (a_1, b_1, c_1) \text{ if } a_1 < a_2, \text{ or } a_1 = a_2 \text{ and } b_1 < b_2, \\ (a_2, b_2, c_2) \text{ if } a_1 > a_2, \text{ or } a_1 = a_2 \text{ and } b_1 > b_2, \\ (a_1, b_1, c_1 + c_2) \text{ if } a_1 = a_2 \text{ and } b_1 = b_2. \end{cases}$
- $r(a, b, c) = (a, b, rc)$.

That the multiplication and scalar multiplication rules give the corresponding operations on indexed probabilities is immediate. The addition rule follows from the observation that if $a \prec b$ are indexed probabilities, then $a + b \equiv b$.

Note that we can also divide any two triples; the result may not correspond to a threshold expression (for example, the first or second element might be negative) but if $(a_1, b_1, c_1) < (a_2, b_2, c_2)$, then the quotient $(a_1 - a_2, b_1 - b_2, c_1/c_2)$ corresponds to an equivalence class of indexed probabilities.

For logistic noise, the situation is even simpler. In this case,

$$[r](\sigma) = (1 + e^{\pi r/\sigma\sqrt{3}})^{-1} \sim e^{-\pi r/\sigma\sqrt{3}},$$

for $r > 0$. So $[r_1, \ldots, r_n] \equiv [r_1 + \cdots + r_n]$, if all the r_i are positive, and $[r] \prec [r']$ exactly when $r > r'$, if $r' \geq 0$.

6 Computing limiting stationary distributions

Let $\mathbf{A}$ be a transition matrix whose entries are indexed probabilities (an indexed transition matrix) as described in the last section. First note that any transient state for the matrix $\mathbf{A}(0^+)$ is invisible. This is because $\mathbf{p}(0^+)$ is a stationary distribution for $\mathbf{A}(0^+)$, by continuity, so must vanish on any transient state of $\mathbf{A}(0^+)$.

We can eliminate invisible states from an indexed transition matrix $\mathbf{A}$ without changing $\mathbf{p}(0^+)$ because if we eliminate a state with a low stationary probability from $\mathbf{A}(\sigma)$, for fixed σ, the stationary probabilities of the remaining states do not change much. Here is how it is done.

Theorem 6 *Let $\mathbf{A} = (a_{ij})$ be the transition matrix of an irreducible Markov process, and $\mathbf{p}$ its stationary distribution. If we eliminate state 1, setting*

$$a'_{ij} = a_{ij} + \frac{a_{i1}a_{1j}}{\sum_{k\neq 1} a_{1k}},$$

for all $i \neq 1$ and $j \neq 1$, then

$$p'_j = \frac{p_j}{1 - p_1}.$$

Proof. Calculate $\mathbf{p}'\mathbf{A}' = \mathbf{p}'$. ■

We use $\sum_{k\neq 1} a_{1k}$ in the denominator, rather than $1 - a_{11}$, so that when implementing the algorithm on an indexed transition matrix, we need only add, multiply, and divide.

Corollary 7 *Let $\mathbf{A} = (a_{ij})$ be an indexed transition matrix. If state 1 is invisible, and we eliminate it, replacing each a_{ij} by*

$$a_{ij} + \frac{a_{i1}a_{1j}}{\sum_{k\neq 1} a_{1k}},$$

then $p_j(0^+)$ does not change for $j \neq 1$. ■

The next theorem is used to identify invisible states. In it we talk about δ-paths for an indexed transition matrix $\mathbf{A} = (a_{ij})$, where δ is an indexed probability. These are defined exactly as for the nonindexed case, with $\delta \leq a_{ij}$ replaced by $\delta \preceq a_{ij}$. The property we will use is that if P is a δ-path for $\mathbf{A}$, then there exists $t > 0$ and $\sigma' > 0$, such that P is a $t\delta(\sigma)$ path for $\mathbf{A}(\sigma)$ for each $\sigma \leq \sigma'$.

Theorem 8 *Let $\mathbf{A} = (a_{ij})$ be an indexed transition matrix, and δ be an indexed probability with $\delta(0^+) = 0$. Suppose there are two disjoint nonempty sets of states S_0 and S_1 such that*

1. *there is a δ-path in $S_0 \cup S_1$ from each state in S_1 to S_0,*
2. *$a_{ij} \preceq \delta$ whenever $j \neq i \in S_1$,*
3. *$a_{ij} \prec \delta$ for $i \in S_0$ and $j \notin S_0$.*

Then the states in S_1 are invisible.

Proof. From Condition 1 there is $t > 0$ and $\sigma_1 > 0$ so that from each state in S_1 there is a path in $S_0 \cup S_1$ to S_0 that is a $t\delta(\sigma)$-path for $\mathbf{A}(\sigma)$ for each $\sigma \leq \sigma_1$.

If $j \neq i \in S_1$, then $a_{ij} \preceq \delta$, so $\lim_{\sigma\to 0} a_{ij}(\sigma)/\delta(\sigma)$ exists. Therefore, there is $\sigma_2 > 0$ and $r > 1$ so that, if $\sigma \leq \sigma_2$, then $a_{ij}(\sigma) < rt\delta(\sigma)$.

If $i \in S_0$ and $j \notin S_0$, then $a_{ij} \prec \delta$, so $a_{ij}(\sigma)/\delta(\sigma) \to 0$. Therefore, for each $\varepsilon > 0$, there is $\sigma_3 > 0$ so that if $\sigma \leq \sigma_3$, then $a_{ij}(\sigma) < \varepsilon t\delta(\sigma)$.

By Theorem 3, if $\sigma \leq \min(\sigma_1, \sigma_2, \sigma_3)$, then $p_k(\sigma) < K\varepsilon t$ for each k in S_1, where K is a constant depending only on r. ■

We extend the idea of a transient state to an indexed transition matrix $\mathbf{A} = (a_{ij})$ as follows. Let δ be the largest off-diagonal entry of $\mathbf{A}$. Then a state i is **transient** for $\mathbf{A}$ if there exists a state j, and a δ-path from i to j, but no δ-path from j to i. This means, roughly, that the state i is transient at the smallest interesting time scale. Let S_1 be all the states that lie on any of the δ-paths from i to j. Let S_0 be all the states you can reach with δ-paths from j. Then Theorem 8 shows that i is invisible.

A state that is transient for $\mathbf{A}(0^+)$ is transient for $\mathbf{A}$. Indeed, if i is a transient state for $\mathbf{A}(0^+)$, then there is $j \neq i$ such that $1 \preceq a_{ij}$.so we can take $\delta = 1$ in the definition of "transient for $\mathbf{A}$" and the condition is exactly that i be transient for $\mathbf{A}(0^+)$.

In addition to identifying and eliminating invisible states, one more technique comes into play when computing limiting stationary distributions. A set S of states may form an irreducible Markov chain on its own at a certain time-scale. Every time you enter S, you stay long enough to achieve a local stationary distribution before you leave. In this case, we can collapse S into a single state, find the probability of seeing this composite state, and multiply by the local stationary distribution to find the probability of seeing the individual states of S. Let's make this more precise.

Let δ be an indexed probability. Call a set S of states **δ-irreducible** if

1. $a_{ij} \prec \delta$ for $i \in S$ and $j \notin S$,
2. If $i, j \in S$ and $i \neq j$, then $a_{ij} \preceq \delta$,
3. If $i, j \in S$ and $i \neq j$, then there is a δ-path from i to j.

What is the local stationary distribution on S? Condition (2) says that, if i and j are distinct elements of S, then $a_{ij}(\sigma)/\delta(\sigma)$ converges to a nonnegative real number t_{ij}. Condition (1) says that $a_{ij}(\sigma)/\delta(\sigma)$ converges to 0 if i is in S and j is not. Condition (3) implies that the matrix indexed by S, with t_{ij} off the diagonal and the diagonal terms chosen so that the row sums are 0, has a one-dimensional left null space.

> To see this, scale the rows so that the diagonal entries are all -1. This doesn't change the right null space, which contains the all 1's column vector v_0, hence doesn't change the dimension of the left null space. The resulting matrix is of the form $-I+B$, where B is the matrix of an irreducible Markov chain. If $Bv = v$, then $B^k v = v$. But B^k converges to a matrix M whose rows are the unique stationary distribution of B. Thus the columns of M are multiples of v_0, and $Mv = v$, so v is a multiple of v_0.

Normalizing a vector in that null space gives the desired stationary distribution.

Let $(q_i)_{i \in S}$ be the local stationary distribution on the δ-irreducible set S. To collapse S into a single (mixed) state 0, we set $a'_{i0} = \sum_{j \in S} a_{ij}$ and $a'_{0j} = \sum_{i \in S} q_i a_{ij}$. We omit the verification that this all works out.

6.1 The algorithm

We can now outline the algorithm for computing limiting stationary distributions for normal noise. This algorithm was implemented using the computational resources provided by the Multi-Disciplinary Research Computing Facility which is funded jointly by the National Science Foundation and Florida Atlantic University.

1. Set up the square matrix $\mathbf{A}$ of triples, indexed by the 2^n states of the network. Each entry is a product of n triples, one for each node, which are of the form $(r^2, 1, 1/r)$ if it requires a noise of r to switch or hold that node to the required sign, of the form $(0, 0, 1)$ if the node will assume the desired sign for any sufficiently small noise, and of the form $(0, 0, 1/2)$ if the node will assume the sign of the noise (that is, the influence field is 0 at that node).

2. Find the largest off-diagonal term δ.

3. Delete the transient states of the matrix, modifying the entries of the matrix according to Corollary 7.

4. Create mixed states for each δ-irreducible subset of the modified matrix. Record the composition of these mixed states together with their local stationary distributions.

5. If more than one state remains, go to step 2.

In actual practice, the matrix $\mathbf{A}$ is indexed by only 2^{n-1} states because of the symmetry of interchanging positive and negative signs on the nodes.

For logistic noise, the program is the same, except that the entries in the matrix $\mathbf{A}$ are pairs of real numbers, (r, s), with $s > 0$, representing the threshold expression $s[r]$. Note that the pair $(0, 2)$ corresponds to the triple $(0, 0, 1)$, and the pair $(0, 1)$ corresponds to the triple $(0, 0, 1/2)$.

7 The single-strength circle

We illustrate how to identify invisible states in a simple case. Consider a group of m people sitting around a circular table, each one influenced only by himself and his two nearest neighbors. The influence matrix, for $m = 6$, is

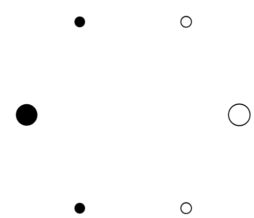
$$\mathbf{C} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{pmatrix}$$

If two adjacent sites have the same sign, then that condition will persist indefinitely in the absence of noise. So a periodic state must consist of alternating signs all around the circle, and this is possible only if m is even.

The stable states are precisely those in which each node is next to a node of the same sign. We shall show that the only visible states are those in which everybody has the same sign (a **homogeneous** state). It is clear that there is a [1]-path from any inhomogeneous stable state to the all-plus state by successively changing to plus those minus sites that are next to plus sites. Neither homogeneous state can be exited with probability greater than [3], and no inhomogeneous stable state can be exited with probability greater than [1]. So each inhomogeneous stable state is invisible by Theorem 8. Moreover any periodic state, if present, can be changed, with noise of 1, to a transient state that evolves to a homogeneous state, so they are also invisible.

8 A two-strength hexagonal network

Next we look at an example where there are visible inhomogeneous states. Consider a group of m people sitting around a circular table, each one influenced only by himself and his two nearest neighbors, but now some are strong and some are weak. We assign **strengths** 1 and $r < 1$ to the strong and weak respectively. Consider specifically the situation where $m = 6$, exactly two are strong, and these are sitting across from each other. Represent the state of such a system by a string of pluses and minuses. Put the two strong nodes in positions 2 and 5, so if the strong nodes are in one state and the weak nodes in the other, then the string is $-+--+-$ or $+-++-+$. Here is the configuration $---+++$



The influence matrix is

$$\mathbf{C} = \begin{pmatrix} r & 1 & 0 & 0 & 0 & r \\ r & 1 & r & 0 & 0 & 0 \\ 0 & 1 & r & r & 0 & 0 \\ 0 & 0 & r & r & 1 & 0 \\ 0 & 0 & 0 & r & 1 & r \\ r & 0 & 0 & 0 & 1 & r \end{pmatrix}$$

The states can be grouped into two-element composite states because of the symmetry of interchanging plus and minus. These are not the mixed states of step 4 of the algorithm; rather they are formed at the start to cut the initial number of states in half, as mentioned in the remark following the algorithm. If $r < 1/2$, then all composite states other than the two that contain $++++++$ and $---+++$ are transient for the matrix $\mathbf{A}(0^+)$. After eliminating the transient states we get the transition matrix

$$\begin{pmatrix} 1 & [2r+1] \\ [2r+1]+2[1,1] & 1 \end{pmatrix}$$

for the two stable composite states. The entry $[2r+1]$ comes from effecting the transition from $++++++$ to $---+++$ by noise of intensity $2r+1$ at the second node (a strong node). This takes you to the transient state $+-++++$ which immediately goes to $---+++$ because $r < 1/2$. The entry $[2r+1]+2[1,1]$ comes from the two different ways of passing from $---+++$ to $++++++$. The first is by noise of intensity $2r+1$ at the second node. The second is noise of intensity 1 at the first or third nodes, followed immediately by noise of intensity 1 at the second node.

Note that the row sums are 1, even though the off-diagonal entries are nonzero, because $[s] \prec 1$ for any $s > 0$. To compute the stationary distribution, we can change the diagonal entries to make the row sums 0, and then compute the left null space. The issue now is how $[2r+1]$ compares to $[1,1]$. That depends on the noise.

For normal noise, they compare as $(2r+1)^2$ and 2, so we must compare $2r+1$ to $\sqrt{2}$. If $r > (\sqrt{2}-1)/2 \approx .20711$, then

$$[2r+1] + 2[1,1] \equiv 2[1,1] \succ [2r+1]$$

so the only visible (composite) state is homogeneous. If $r \leq (\sqrt{2}-1)/2$, then $[2r+1] + 2[1,1] \equiv [2r+1]$ and the two fixed (composite) states are equally likely.

For logistic noise, they compare as $2r+1$ and 2, and $2r+1 < 2$ because $r < 1/2$. So $[1,1] \prec [2r+1]$, whence

$$[2r+1] + 2[1,1] \equiv [2r+1]$$

so the two states are equally likely whenever $r < 1/2$..

If $r = 1/2$, then all states other than $++++++$ and $---+++$ are transient for the matrix $\mathbf{A}(0^+)$, as before. After eliminating transient states, the matrix becomes

$$\begin{pmatrix} 1 & \frac{3}{13}[2r+1] \\ \frac{67}{26}[1,1] & 1 \end{pmatrix}.$$

The difference between this and the $r < 1/2$ case is that, for the latter, a noise of $2r+1$ at a strong node guarantees a transition from $++++++$ to $+-++++$ to $---+++$, but if $r = 1/2$, the transient state $+-++++$ might revert back to $++++++$. Moreover, for the transition from $---+++$ to $++++++$, the transient state $--++++$ might persist awhile, giving more opportunity for a second noise of 1 to switch the negative strong node. The coefficients in this matrix are not so easy to compute—it took some computer runs to get them right. Once computed, however, we see that, for normal noise, only the homogeneous state is visible, while for logistic noise, the stationary distribution on the homogenous and inhomogeneous states is $(67/73, 6/73)$.

Now suppose $r > 1/2$. It will be convenient, for hand computation, to use the symmetries about the horizontal and vertical axes, in addition to interchanging plus and minus, to form composite states. Thus $----++$ is in a composite state containing seven other states. No state where some node has a sign different from both its neighbors is stable. Of these, the state

$-+-+-+$ has period 2 (but is stable as a composite state) and the rest are transient for $\mathbf{A}(0^+)$. For example, $+-+-++$ goes to $++-+++$ goes to $++++++$. The state $++--++$ is stable, as are $----++$ and $+---++$. That's the lot—six stable (composite) states:

$$\begin{array}{c} ++++++ \\ ---+++ \\ -+-+-+ \\ ++--++ \\ ----++ \\ +---++ \end{array}$$

The matrix for these six states is

$$\begin{pmatrix} 1 & \cdot & \cdot & \cdot & 4[1+2r, 2r-1, 2r-1] & \cdot \\ \cdot & 1 & \cdot & \cdot & 4[1] & \cdot \\ 2[2r-1] & \cdot & 1 & \cdot & \cdot & \cdot \\ [2r-1] & \cdot & \cdot & 1 & \cdot & \cdot \\ 2[1] & [2r-1] & \cdot & \cdot & 1 & 2[1] \\ \cdot & \cdot & \cdot & 2[1] & 2[2r-1] & 1 \end{pmatrix}.$$

Some of the entries are difficult to compute by hand, but it is not necessary to compute them all exactly. Those marked by a dot are on a time scale larger than any computed entry in the same row. The entry a_{15}, for example, arises from switching a strong node, which requires a noise of $1+2r$, then holding that strong node in the new state, which takes a noise of $2r-1$, while simultaneously switching an adjacent weak node, which also takes a noise of $2r-1$. The coefficient of 4 comes from the four choices of a strong node together with an adjacent weak node. The coefficient of 4 in a_{25} comes from the fact that we can switch any of the four weak nodes.

States three and four are invisible by Theorem 8 with $\delta = [2r-1]$. Eliminating them we get the matrix

$$\begin{pmatrix} 1 & \cdot & 4[1+2r, 2r-1, 2r-1] & \cdot \\ \cdot & 1 & 4[1] & \cdot \\ 2[1] & [2r-1] & 1 & 2[1] \\ 2[1] & \cdot & 2[2r-1] & 1 \end{pmatrix}.$$

States five and six (now states three and four) are also invisible by Theorem

8 with $\delta = [2r-1]$. Eliminating state six we get the matrix

$$\begin{pmatrix} 1 & \cdot & 4[1+2r, 2r-1, 2r-1] \\ \cdot & 1 & 4[1] \\ 2[1] & [2r-1] & 1 \end{pmatrix}.$$

Eliminating state five we get

$$\begin{pmatrix} 1 & 4[1+2r, 2r-1, 2r-1] \\ \dfrac{8[1,1]}{[2r-1]} & 1 \end{pmatrix}.$$

For both normal and logistic noise $[1+2r, 2r-1, 2r-1] \prec [1,1]/[2r-1]$, whence state two is invisible by Theorem 8 with $\delta = 8[1,1]/[2r-1]$ so the only visible states are the homogeneous ones.

9 A simple line

Consider the case of a line of $n > 1$ weak nodes with two strong nodes on the ends. Here we will see the notion of a δ-irreducible set in action.

$$\CIRCLE \quad \bullet \quad \bullet \quad \circ \quad \circ \quad \circ \quad \bigcirc$$

For $r \leq 1/2$, the stable states are those for which everybody is next to someone of the same sign. For $r > 1/2$, the stable states are those for which each weak node is next to someone of the same sign. We may ignore the unstable states, which are all transient for $\mathbf{A}(0^+)$.

What is the easiest change from a stable state? Denote $+$ and $-$ by p and m, and use upper case for strong nodes.

- For $ppm \to pmm$ the threshold is r.
- For $ppM \to pmM$ the threshold is $2r-1$.
- For $pM \to pP$ the threshold is $1-r$

We can eliminate the set S_1 of states with a weak node next to a strong node of the opposite sign. To see this, note that these are unstable for $r \leq 1/2$. For $r > 1/2$, let S_0 be the other states and $\delta = \max([2r-1], [1-r])$. Then $a_{ij} \leq [1]$ for $i \in S_0$ and $j \in S_1$, and $a_{ij} \leq \delta$ for $i \neq j$, so Theorem 8 applies.

Now let S_0 be the set of states such that each weak node can trace a string of weak nodes of the same sign to a strong node of the same sign, and let S_1 consist of the rest of the states. If $i \in S_0$ and $j \in S_1$, then $a_{ij} \leq \max([3r], [1+r]) \prec [r]$, while if $i \neq j$, then $a_{ij} \leq [r]$. As there is an $[r]$-path from any state to a state in S_0, the states outside of S_0 are invisible by Theorem 8.

A state in S_0 whose strong nodes have the same sign is homogeneous. The set of states in S_0 whose strong nodes have fixed opposite signs forms an $[r]$-irreducible set, and the local action is a random walk of the boundary between the weak nodes of different signs, which can be in any of $n-1$ positions. The visible states are equally likely because the transition matrix (for $n = 6$) is

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & \frac{1}{2} \end{pmatrix}$$

We may thus reduce to two composite states whose constituent states are equally likely: a homogeneous state (2 elements), and an inhomogeneous state ($2(n-1)$ elements). The transition matrix becomes

$$\begin{pmatrix} 1 & 2[1+r] \\ 2[1+r] + [1, 1-r]\frac{2}{n-1} & 1 \end{pmatrix}.$$

The entry $2[1+r] + 2[1, 1-r]/(n-1)$, the probability of going from inhomogeneous to homogeneous, comes about as follows. The term $2[1+r]$ is the probability that a strong node will change sign. The adjacent weak node changes sign at the next step and the resulting state becomes homogeneous in the time scale $[r]$. The other term comes into play when the boundary between weak nodes of different signs is as far left or right as it can be. In those states, a noise of 1 at a weak node, followed immediately by a noise of $1-r$ at the adjacent strong node will change the sign of the strong node. If $n > 2$, the probability of being in one of these situations is $2/(n-1)$. If $n = 2$, then we get a factor of 2 because either of the weak nodes is eligible to change.

If $[1, 1-r] \succ [1+r]$, then the inhomogeneous states will be invisible. This is always true if $r = 1$, that is, when all nodes have the same strength. In the normal case, it happens exactly when $1^2 + (1-r)^2 < (1+r)^2$, that is, when $r > 1/4$. If $r \le 1/4$, then $[1, 1-r] \prec [1+r]$ so the stationary distribution is equally divided between the composite homogeneous state and the composite inhomogeneous state. That is, there are $2n$ visible states: two homogeneous ones each with probability $1/4$ of being seen, and $2(n-1)$ inhomogeneous ones, each with probability $1/4(n-1)$ of being seen.

In the logistic case, $[1, 1-r] \succ [1+r]$ exactly when $2-r < 1+r$, that is, when $r > 1/2$. So if $r > 1/2$, then the inhomogeneous states will be invisible. If $r = 1/2$, then $[1, 1-r] \equiv [1+r]$ and the stationary distribution on the composite homogeneous state and the composite inhomogeneous state is $(n/(2n-1), (n-1)/(2n-1))$, while if $r < 1/2$ it is $(1/2, 1/2)$.

If we take $n = 1$ in this example, then there are essentially two (composite) stable states: one where every node has the same sign, and one where the strong nodes have different signs. The transition matrix is

$$\begin{pmatrix} 1 & 2[1+r] \\ [1-r] & 1 \end{pmatrix}$$

so the inhomogeneous state is invisible for all $r > 0$ (whatever the noise).

10 The two-strength circle

Consider a finite number of nodes arranged in a circle, each having one of two strengths, 1 or $r < 1/2$. We will assume that each weak node is next to another weak node. The stable states are those for which everybody is next to someone of the same sign, and if a weak node is next to a strong node, then they have the same sign.

The thresholds that are no greater than 1 for changing a stable state are

- r for $ppm \to pmm$.
- r for $pPM \to mMM$ (which passes through the unstable state pMM).
- 1 for $PMM \to PPM$.

So the action starts out of the form $ppm \to pmm$ and $pPM \to mMM$. We can find an $[r]$-path from any state to a state where each weak node can trace a chain of weak nodes with the same sign to a strong node of the same sign who is next to another node of the same sign. The probability of exiting these states is on the order of $[3r, r, r] + [1]$. (The $[3r, r, r]$ term comes from changing $pppp$ to $pmpp$, which requires a noise of $3r$, then changing $pmpp$ to $pmmp$, which requires a noise of r to hold the m, simultaneously with a noise of r to change the other p. The $[1]$ term comes from changing $PPMMm$ to $PPPMm$.) So only these states can be visible.

Eliminate all but these possibly visible states. Call two states equivalent if they agree on the strong nodes. Then any equivalence class of states is an $[r]$-irreducible set. Lump them into composite states. Now there is a $[1]$-path from any state to a state where adjacent strong nodes have the same sign, and the probability of exiting this set of states is less than $[2 + r]$. So only these states can be visible.

Consider a fixed assignment of signs to the strong nodes, so that adjacent strong nodes have the same sign. Then, between two groups of strong nodes of opposite signs, we get a random walk as in the example of the simple line. So, in this case, with positive probability, a weak node next to a strong node, which must be the same sign for the state to be visible, is next to a weak node of the opposite sign. That is, the situation between two strong nodes of opposite sign looks like $Ppmm \ldots mmM$ with probability $1/(k-1)$ where k is the number of weak nodes in between.

For what values of r are only the homogeneous states visible? Consider the situation where there is an isolated strong node whose neighboring groups of strong nodes have at least two elements, so they do not switch on the same time scale as the isolated one. The picture might look like

$$MMmmmMmmmmMM.$$

You exit the homogeneous state by switching the isolated strong node with probability $[1+2r]$, and revert with probability $[1+2r]+\theta[1, 1]$, where $\theta > 0$ is a scalar. The $[1, 1]$ is the probability of the transition $pPpm \to pMmm$ by first switching the p and then the P, each of which occurs with probability $[1]$. The θ is the probability that we are in the state $pPpm$ rather than $pPpp$ because of the random walk taking place at the shorter time scale on the weak nodes. In the normal case, we have to compare $(1+2r)^2 = 1+4r+4r^2$ with $1^2 + 1^2 = 2$. They are equal when $r = -\frac{1}{2} + \frac{1}{2}\sqrt{2} = .20711$. What does that say? You won't see a single isolated strong node if $r > .20711$. For logistic noise, the condition is $1 + 2r = 2$, or $r = 1/2$, so for $r < 1/2$, which we have assumed, we are as likely to see the isolated strong node with a different sign as to see the homogeneous state. The analysis in the general case, where there are several isolated strong nodes between two larger groups, must consider random walks on those isolated strong nodes, but we end up comparing the same threshold expressions.

There are two ways to exit a homogeneous state by switching a group of two adjacent strong nodes. One is to switch a strong node, then hold it in place the very next time. The scenario is

$$pPPpp \to pPMpp \to pPMmp$$

This last state is invisible, and, with probability $1/2$, will eventually transit to $pMMmp$ and then to $mMMmp$. So this switch occurs with probability $\frac{1}{2}[2+r, r]$. Or we could switch a weak node first, as above, then switch the adjacent strong node and hold it

$$pPPpp \to pPPmp \to pPMpp \to pPMmp$$

which happens with probability $[1+2r, 2-r, r] \prec [2+r, r]$. So the probability of switching two adjacent nodes in a (locally) homogeneous state is on the order of $[2+r, r]$.

What is the probability of switching two adjacent strong nodes in an inhomogeneous situation? With positive probability the situation is $pPPpm$ so we can switch via

$$pPPpm \to pPPmm \to pPMpm \to pPMmm$$

with probability $[1, 2-r, r]$ or via

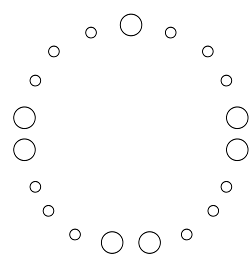
$$pPPpm \to pPMpm \to pPMmm$$

with probability $[2+r, r]$, as in the homogeneous case. Comparing the two probabilities, we see that $[2+r, r] \prec [1, 2-r, r]$ in the normal case exactly when $r > 1/8$, and in the logistic case when $r > 1/2$ (that is, never).

So for logistic noise it appears that the visible states are those for which any two adjacent nodes have the same sign if one of them is strong, and each weak node can trace a chain of weak nodes with the same sign to a strong node of the same sign. Each distribution of signs to the groups of strong nodes is equally likely, and, given such a distribution, we observe random walks of the sign boundary of the weak nodes between two neighboring groups of strong nodes of opposite signs. We had conjectured that this was the case for normal noise, but the behavior there is a lot more complicated.

Let's look at some specific cases with normal noise. Suppose you have one pair of adjacent strong nodes and one isolated strong node. What do you see? If $r > .20711$, you just see the homogeneous states. If $r < .20711$, the isolated strong node is going to change back and forth, so the pair of strong nodes are half the time in the homogeneous situation and half the time not. The pair of strong nodes switch at random, and we have four equally likely states for the two clumps.

Now suppose you have three pairs of strong nodes, and one isolated strong node. Suppose $1/8 < r < .20711$ and the clumps are equally spaced with $k+1$ small nodes between adjacent clumps. Here is the picture with $k = 2$.



There are essentially three states. The isolated node is always switching back and forth. The three pairs can either be the same sign, or the odd pair can be on the end, or the odd pair can be in the middle. Let $q = [1, 2-r, r]$, the probability of a pair switching sign if there is a nearby (almost adjacent) weak node of the opposite sign. Then the transition matrix is

$$\begin{pmatrix} \cdot & \frac{q}{k} & 0 \\ \frac{3q}{2k} & \cdot & \frac{q}{2k} \\ \frac{2q}{k} & \frac{3q}{k} & \cdot \end{pmatrix}$$

that is

$$\frac{q}{k}\begin{pmatrix} \cdot & 1 & 0 \\ 3/2 & \cdot & 1/2 \\ 2 & 3 & \cdot \end{pmatrix}$$

Filling in the dots so that the row sums are zero, the matrix becomes essentially

$$\begin{pmatrix} -2 & 2 & 0 \\ 3 & -4 & 1 \\ 4 & 6 & -10 \end{pmatrix}$$

so its left null space is generated by $(17, 10, 1)$. Thus the invariant distribution on these three composite states is proportional to $(17, 10, 1)$.

If $r \leq 1/8$, then there is no difference between a homogeneous and an inhomogeneous environment, so the 16 states of the four clumps will be equally probable. If $r > .20711$, then only the homogeneous states will be visible.

11 The hexagon with no self-support

Let's look again at the hexagon model of Section 8, but this time with no self-support. There are five composite states, comprising 16 pure states, that we need consider.

$+ + + + ++$	2	The homogeneous states: stable.
$+ - + + ++$	8	Four pairs of period-two states: one odd strong node.
$- + - + -+$	2	A pair of period-two states: alternating.
$+ - + + -+$	2	A pair of period-two states: strong versus weak.
$- - - + ++$	2	The almost homogeneous states: stable.

The other 48 states are

$- + + + ++$	8	One weak node: immediately reverts to homogeneous.
$+ + - - ++$	4	Two adjacent weak nodes: also immediately reverts.
$- + + - ++$	4	Two nonadjacent weak nodes: immediately reverts.
$- - + + ++$	8	Two adjacent nodes, one strong: Goes to $-? - + + +$
$+ - + - ++$	8	Two nonadjacent nodes, one strong, one weak: Goes to $- + - + ++$
$+ + - - -+$	4	Goes to $+? + -?-$, that is $+ + + - --$ or $+ - + - +-$ or $+ + + - +-$
$- + + - -+$	4	Goes to $+? + -?-$
$+ - + - -+$	8	Goes to $- + --?-$

which are all invisible because they are transient (aperiodic).

Suppose $r < 1/3$, that is, $2r < 1-r$. Then, after eliminating the invisible states, the transition matrix is

$$\begin{pmatrix} 1 & 2[2r] & \cdot & \cdot & \cdot \\ \frac{1}{2}[2r] & 1 & \frac{1}{2}[2r] & \frac{1}{2}[2r] & \frac{1}{2}[2r] \\ \cdot & 2[2r] & 1 & \cdot & \cdot \\ \cdot & 2[2r] & \cdot & 1 & \cdot \\ \cdot & 2[2r] & \cdot & \cdot & 1 \end{pmatrix}$$

giving a probability distribution proportional to $1, 4, 1, 1, 1$, so each of the 16 states is equally likely.

Suppose $r > 1/3$, that is $1 - r < 2r$. Then the transition matrix is

$$\begin{pmatrix} 1 & 2[2r] & \cdot & \cdot & \cdot \\ \frac{1}{2}[r] & 1 & \frac{1}{2}[r] & \frac{1}{2}[2r] & \frac{1}{2}[2r] \\ \cdot & 2[2r] & 1 & \cdot & \cdot \\ \cdot & 2[r] & \cdot & 1 & \cdot \\ \cdot & 2[r] & \cdot & \cdot & 1 \end{pmatrix}$$

so states 2, 4 and 5 are invisible. Eliminating them results in the matrix

$$\begin{pmatrix} 1 & [2r] \\ [2r] & 1 \end{pmatrix}$$

so the distribution is fifty-fifty between the homogeneous states $++++++$ and the alternating states $-+-+-+$. This is true even if $r = 1$.

What happens when $r = 1/3$? The matrix is

$$\begin{pmatrix} 1 & 2[2] & \cdot & \cdot & \cdot \\ [2] & 1 & [2] & \frac{1}{2}[2] & \frac{1}{2}[2] \\ \cdot & 2[2] & 1 & \cdot & \cdot \\ \cdot & 4[2] & \cdot & 1 & \cdot \\ \cdot & 4[2] & \cdot & \cdot & 1 \end{pmatrix}$$

and the stationary distribution is proportional to $4, 8, 4, 1, 1$.

12 The single-strength torus

We end with a two-dimensional model: the lattice point on a torus (with the usual self-support, unlike the preceding section). We shall show that if every site on the torus has the same strength, then the only visible states are the homogeneous ones. Note that in this example the limit matrix $\mathbf{A}(0^+)$ is the transition matrix of the dynamical system. We may assume that the transient states have been eliminated (but not the periodic ones).

Let S_1 consist of those stable states that have sites where the local field is ± 1, and S_0 the remaining stable states. Clearly $a_{ij} \preceq [1]$ whenever $j \neq i \in S_1$, and $a_{ij} \prec [1]$ for $i \in S_0$ and $j \notin S_0$. To apply Theorem 8, we need to show that there is a [1]-path from each state in S_1 to S_0.

The algorithm for finding such a path is this. Change pluses to minuses at sites where the local field is 1 until there are no more places where the local field is 1. As you do this, some pluses may spontaneously change to minuses because the field there has turned negative. If, after having done this, the state is not in S_0, then change minuses to pluses at sites where the local field is -1 until the local field is nowhere -1. The key fact is that if the local field is nowhere 1, then the blocks of minuses form rectangles (each plus is next to at least three other pluses—look at the boundary between plus and minus). So when you start changing minuses to pluses, you change the isolated rectangles of minuses (those that don't go around the torus in either direction) to pluses and also those rectangles of width 1, so you end up in S_0.

So suppose the system is in a stable state where the local field is ± 3 or ± 5 at each site. These states all have the form of rows of homogeneous bands of width at least 2, like

$$\begin{matrix} + & + & + & + & + & + \\ + & + & + & + & + & + \\ - & - & - & - & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \end{matrix}$$

Call this state x. We get out of state x with probability $[3, 1, 1]$ by creating a stable two-by-one incursion.

$$\begin{matrix} + & + & + & + & + & + \\ + & + & + & + & + & + \\ - & - & + & + & - & - \\ - & - & - & - & - & - \\ - & - & - & - & - & - \end{matrix}$$

Then we can relax a little and watch a chain on 1-noise which either ends up at the original state or at the state

$$\begin{array}{cccccc} + & + & + & + & + & + \\ + & + & + & + & + & + \\ + & + & + & + & + & + \\ - & - & - & - & - & - \\ - & - & - & - & - & - \end{array}$$

which we call state y. As the intermediate states are invisible, we have $a_{xy} \approx [3, 1, 1]$.

So there is a $[3, 1, 1]$-path from state x to a homogeneous state. On the other hand, to get out of a homogeneous state, it takes $[5, 3, 3, 3, 1, 1, 1]$ to get a two-by-two square, which is the minimal viable configuration. Thus the only visible stable states are the homogeneous ones. There remains the possibility that a periodic state could be visible.

By a theorem of Goles and Olivos, every periodic state has period 2 [3, Theorem 2.1]. With probability at least $[3, 3]$ we can transit from a periodic state to one which has fewer periodic sites (by holding two sites fixed). First eliminate those that have a local field of ± 1 somewhere—these can be exited with probability [1]. The only periodic state with no local field of ± 1 in either phase is the checkerboard (if it exists), but this fact is not needed for the argument.

13 Questions

- We have shown, for the single-strength circle, line, and torus, that the only visible states are homogeneous. Is that true for single-strength networks with any connected geometry whatsoever?

- What is the largest ratio r of weak to strong for which there are visible inhomogeneous states in some two-strength connected network with normal noise? This would be 1 if the answer to the preceding question were "no." For the hexagonal network in Section 8 we saw that the critical value of r was $(\sqrt{2} - 1)/2 \approx .20711$. For the linear networks in Section 9, the critical value of r is $1/4$. For the hexagonal network without self-support in Section 11, the critical value is $1/3$—but it is really a separate question as to what happens when you vary the self-support.

- Can the behavior of two-strength circles with normal noise be completely classified in a reasonable way?

- The algorithm, as it now stands, requires maintaining a large matrix, none of whose entries are zero. This limits the algorithm to fairly small networks, a maximum of 12 nodes in the current implementation. Can this be circumvented?

References

[1] Cox, J. T., and Rick Durrett, Nonlinear voting models, *Random walks, Brownian motion, and interacting particles systems*, 189–201, Birkhäuser 1991

[2] William Feller, *An introduction to probability theory and its application*, Wiley 1970.

[3] Eric Goles and Servet Martínez, *Neural and automata networks*, Kluwer 1990.

[4] Hertz, John, Anders Krogh and Richard G. Palmer, *Introduction to the theory of neural computation*, Addison-Wesley 1991.

[5] John G. Kemeny and J. Laurie Snell, *Finite Markov chains*, D. Van Nostrand 1960.

[6] Maciej Lewenstein, Andrzej Nowak and Bibb Latané, Statistical mechanics of social impact, *Physical Review* **45** (1992) 763–776.

`richman@fau.edu`