Reconstruction of k Impulse Functions from k Measurements of a Linear Dynamical System

(Dated: June 30, 2018)

I. MATHEMATICAL FRAMEWORK

A. Expected Value of Stochastic Discrete Maps

Consider a network represented by directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = 1, \cdots, n$, 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 $\boldsymbol{x}^0$, 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

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

which is the sum of binomial distributions conditioned on random variables that are the states at time t. The expectation of this conditional distribution is given by

$$\begin{aligned}
\mathbb{E}[X_j^{t+1}] &= \mathbb{E}[\mathbb{E}[X_j^{t+1} | \boldsymbol{X}^t]] \\
&= \mathbb{E}[\mathbb{E}[\sum_{i=1}^{n} B(X_i^t, a_{ji})]] \\
&= \mathbb{E}[\sum_{i=1}^{n} \mathbb{E}[B(X_i^t, a_{ji})]] \\
&= \mathbb{E}[\sum_{i=1}^{n} a_{ji} X_i^t] \\
&= \sum_{i=1}^{n} a_{ji} \mathbb{E}[X_i^t] \\
&= \boldsymbol{a}_j \boldsymbol{X}^t.
\end{aligned}$$

In vector form, this expectation becomes

$$\mathbb{E}[\boldsymbol{X}^{t+1}] = A\mathbb{E}[\boldsymbol{X}^t],$$

which we see is a simple discrete linear map.

B. Upper Bound on Variance

As exciting as linearity in these stochastic linear maps are, the distributional variances provide more trouble. First, we establish a useful inequality for our purposes. If $\alpha_1, \cdots, \alpha_n$ are real numbers such that $\alpha_i \geq 0$ where $\sum_{i=1}^{n} = c$, and $X_1, \cdots, X_n$ are random variables. Then

$$\begin{aligned}
Var\left(\sum_{i=1}^{n} \alpha_i X_i\right) &\leq \left(\sum_{i=1}^{n} \alpha_i\right) \sum_{i=1}^{n} \alpha_i Var(X_i) \\
&\leq c \sum_{i=1}^{n} \alpha_i Var(X_i).
\end{aligned}$$

For hierarchical mixture models, we have the following formula

$$\begin{aligned} Var[X_j^{t+1}] &= \mathbb{E}\left[Var\left[X_j^{t+1}|\boldsymbol{X}^t\right]\right] + Var\left[\mathbb{E}\left[X_j^{t+1}|\boldsymbol{X}^t\right]\right] \\ &= \mathbb{E}\left[\sum_{i=1}^n a_{ji}(1-a_{ji})X_i^t\right] + Var\left[\sum_{i=1}^n a_{ji}X_i^t\right] \\ &= \sum_{i=1}^n a_{ji}(1-a_{ji})\mathbb{E}\left[X_i^t\right] + Var\left[\sum_{i=1}^n a_{ji}X_i^t\right] \\ &\leq \sum_{i=1}^n a_{ji}(1-a_{ji})\mathbb{E}\left[X_i^t\right] + c\sum_{i=1}^n a_{ji}Var\left[X_i^t\right], \end{aligned}$$

which, we see is linear in the expected value and variance of X_i^t, and therefore can write

$$Var\left[\boldsymbol{X}^{t+1}\right] \leq (A - A_2)\mathbb{E}\left[\boldsymbol{X}^t\right] + cAVar\left[\boldsymbol{X}^t\right],$$

where A_2 is a matrix with entries $[a_2]_{ij} = [a]_{ij}^2$. We notice this relationship is recurrent, and can propagate all the way back to some initial condition

$$Var\left[\boldsymbol{X}^{t+1}\right] \leq \left(\sum_{i=0}^k c^i\left[A^{k+1} - A^iA_2A^{k-i}\right]\right)\mathbb{E}\left[\boldsymbol{X}^0\right] + c^{k+1}A^kA_2Var\left[\boldsymbol{X}^0\right],$$

and when the branching parameter $c = 1$:

$$Var\left[\boldsymbol{X}^{t+1}\right] \leq \left((k+1)A^{k+1} - \sum_{i=0}^k A^iA_2A^{k-i}\right)\mathbb{E}\left[\boldsymbol{X}^0\right] + A^kA_2Var\left[\boldsymbol{X}^0\right].$$

Another way to represent this inequality is

$$Var[\boldsymbol{X}^{t+1}] \leq \left(\sum_{i=0}^k c^iA^iA^*A^{k-i}\right)\mathbb{E}[\boldsymbol{X}^0] + c^{k+1}A^kA_2Var[\boldsymbol{X}^0].$$

Here we see that $A^* = A - A_2$ acts as a sort of dispersion matrix, where the more A^* projects the eigenvector with largest eigenvalue of A away from itself, the lower the variance will be. One trivial way to decrease the variance is to make the entries $[a]_{ij}$ close to 0 or 1 such that the entries of $[a(1-a)]_{ij}$ become very small. We also see that the variance of the original distribution propagates linearly with the time step, implying that for critical systems with branching parameter = 1, the variance of the original distribution does not factor into the growth of the state variance beyond a constant value.

C. Intuition of Variance Bound and Slope From 1 Eigenvector Case

Suppose our matrix A contained only a single non-zero eigenvalue $\lambda = 1$ composed of some eigenvector $\boldsymbol{v}$ such that $A = \boldsymbol{v}\boldsymbol{v}^T$. Then for $\boldsymbol{v}_2$ containing squared elements of $\boldsymbol{v}$, we have

$$\begin{aligned} Var[\boldsymbol{X}^{t+1}] &\leq \left((k+1)A^{k+1} - \sum_{i=0}^k A^iA_2A^{k-i}\right)\mathbb{E}[\boldsymbol{X}^0] + A^kA_2Var[\boldsymbol{X}^0] \\ &\leq \left((k+1)(\boldsymbol{v}\boldsymbol{v}^T)^{k+1} - \sum_{i=0}^k(\boldsymbol{v}\boldsymbol{v}^T)^i(\boldsymbol{v}_2\boldsymbol{v}_2^T)(\boldsymbol{v}\boldsymbol{v}^T)^{k-i}\right)\mathbb{E}[\boldsymbol{X}^0] + (\boldsymbol{v}\boldsymbol{v}^T)^k\boldsymbol{v}_2\boldsymbol{v}_2^TVar[\boldsymbol{X}^0] \\ &\leq \left((k+1)(\boldsymbol{v}\boldsymbol{v}^T) - (k-1)(\boldsymbol{v}\boldsymbol{v}^T)(\boldsymbol{v}_2\boldsymbol{v}_2^T)(\boldsymbol{v}\boldsymbol{v}^T) - (\boldsymbol{v}_2\boldsymbol{v}_2^T)(\boldsymbol{v}\boldsymbol{v}^T) - (\boldsymbol{v}\boldsymbol{v}^T)(\boldsymbol{v}_2\boldsymbol{v}_2^T)\right)\mathbb{E}[\boldsymbol{X}^0] + (\boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{v}_2\boldsymbol{v}_2^TVar[\boldsymbol{X}^0] \\ &\leq \left((k+1)(\boldsymbol{v}\boldsymbol{v}^T) - (k-1)(\boldsymbol{v}^T\boldsymbol{v}_2)^2(\boldsymbol{v}\boldsymbol{v}^T) - (\boldsymbol{v}_2\boldsymbol{v}_2^T)(\boldsymbol{v}\boldsymbol{v}^T) - (\boldsymbol{v}\boldsymbol{v}^T)(\boldsymbol{v}_2\boldsymbol{v}_2^T)\right)\mathbb{E}[\boldsymbol{X}^0] + (\boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{v}_2\boldsymbol{v}_2^TVar[\boldsymbol{X}^0] \\ &\leq \left(k(1-(\boldsymbol{v}^T\boldsymbol{v}_2)^2) + 1 + (\boldsymbol{v}^T\boldsymbol{v}_2)^2\right)(\boldsymbol{v}\boldsymbol{v}^T)\mathbb{E}[\boldsymbol{X}^0] - (\boldsymbol{v}^T\boldsymbol{v}_2)(\boldsymbol{v}_2\boldsymbol{v}^T + \boldsymbol{v}\boldsymbol{v}_2^T)\mathbb{E}[\boldsymbol{X}^0] + (\boldsymbol{v}\boldsymbol{v}^T)\boldsymbol{v}_2\boldsymbol{v}_2^TVar[\boldsymbol{X}^0]. \end{aligned}$$

While this expression may look very confusing, the key aspect to notice is that the only term that depends on the time step k is

$$1 - (\boldsymbol{v}^T\boldsymbol{v}_2)^2,$$

where $\boldsymbol{v}^T\boldsymbol{v_2}$ is simply a sum of each element of $\boldsymbol{v}$ raised to the third power. Hence, the closer the entries of $\boldsymbol{v}$ are to 1, the smaller this slope will be. To maximize this slope, we just require that every element of $\boldsymbol{v}$ is equal to $\frac{1}{n}$. Hence, for the extreme case where our matrix has rank 1 with one non-zero eigenvalue and corresponding eigenvector, we see that the increase in variance over time is given by a very simple expression.

D. Exact Variance for First Step

Suppose our initial inputs are drawn from independent distributions with mean $\mathbb{E}[\boldsymbol{X}^0]$ and variance $Var[\boldsymbol{X}^0]$. Then the firing rates immediately after have exact variance

$$\begin{aligned} Var[\boldsymbol{X}^1] &= \mathbb{E}[Var[\boldsymbol{X}^1|\boldsymbol{X}^0]] + Var[\mathbb{E}[\boldsymbol{X}^1|\boldsymbol{X}^0]] \\ &= \mathbb{E}[A^*\boldsymbol{X}^0] + Var[A\boldsymbol{X}^0] \\ &= A^*\mathbb{E}[\boldsymbol{X}^0] + A_2 Var[\boldsymbol{X}^0] \\ &= A\mathbb{E}[\boldsymbol{X}^0] + A_2(Var[\boldsymbol{X}^0] - \mathbb{E}[\boldsymbol{X}^0]) \\ &= \mathbb{E}[\boldsymbol{X}^1] + A_2(Var[\boldsymbol{X}^0] - \mathbb{E}[\boldsymbol{X}^0]). \end{aligned}$$

We see that there is a linear aspect in the growth of the variance, where one component of $Var[\boldsymbol{X}^1]$ is $\mathbb{E}[\boldsymbol{X}^1]$. This term increases the variance as quickly as the mean, which is bad. Alternatively, the second term can reduce the variance of $\boldsymbol{X}^1$ if the variance of the initial distribution is lower than the expected value. Suppose very simply that there is no variance in the initial distribution. Then $Var[\boldsymbol{X}^1] = \mathbb{E}[\boldsymbol{X}^1] - A_2\mathbb{E}[\boldsymbol{X}^0]$. From this relation alone, we see that to minimize some measure of variance to expected value, we want our initial condition to excite the largest eigenmode of A_2, not necessarily A.

E. Bayes Theorem and Posteriors

In the quest to understand the question of input reconstruction, suppose we have some initial state $\boldsymbol{X}^0$ drawn from some distribution. The question we are interested in is how to reconstruct this initial state given a measurement later on. Suppose this later measurement occurs at time $T = 1$. Then $X_j^1 \sim \sum_{i=1}^n B(X_i^0, a_{ji})$. If all inputs to the jth neuron are identical, then $X_j^1 \sim \sum_{i=1}^n B(X_i^0, a_j) = B(\boldsymbol{o}_j^T\boldsymbol{X}^0, a_j)$. Hence,

$$P(X_j^1 = x_j^1|\boldsymbol{X}^0 = \boldsymbol{x}_0) = \binom{\boldsymbol{o}^T\boldsymbol{X}^0}{x_j^1} a_j^{x_j^1}(1-a_j)^{\boldsymbol{o}^T\boldsymbol{X}^0 - x_j^1}.$$

Via Bayes Theorem, we can invert this conditional probability

$$P(\boldsymbol{X}^0 = \boldsymbol{x}^0|X_j^1 = x_j^1) = \frac{P(X_j^1 = x_j^1|\boldsymbol{X}^0 = \boldsymbol{x}_0)P(\boldsymbol{X}^0 = \boldsymbol{x}^0)}{P(X_j^1 = x_j^1)}.$$

We note that because what we measure is X_j^1, across all guesses of initial state, $P(X_j^1 = x_j^1)$ remains constant. We also note that if the initial states are drawn independently of each other, $P(\boldsymbol{X}^0 = \boldsymbol{x}^0) = \prod_{i=1}^n P(X_i^0 = x_i^0)$. Hence, we can write

$$P(\boldsymbol{X}^0 = \boldsymbol{x}^0|X_j^1 = x_j^1) \propto \binom{\boldsymbol{o}^T\boldsymbol{X}^0}{x_j^1} a_j^{x_j^1}(1-a_j)^{\boldsymbol{o}^T\boldsymbol{X}^0 - x_j^1}\prod_{i=1}^n P(X_i^0 = x_i^0),$$

which we can write as the log-likelihood

$$\begin{aligned} \log\left(P(\boldsymbol{X}^0 = \boldsymbol{x}^0|X_j^1 = x_j^1)\right) &\propto \log\left(\binom{\boldsymbol{o}^T\boldsymbol{x}^0}{x_j^1}\right) + \log\left(a_j^{x_j^1}(1-a_j)^{\boldsymbol{o}^T\boldsymbol{X}^0 - x_j^1}\right) + \log\left(\prod_{i=1}^n P(X_i^0 = x_i^0)\right) \\ &\propto \log\left(\binom{\boldsymbol{o}^T\boldsymbol{x}^0}{x_j^1}\right) + x_j^1\log(a_j) + (\boldsymbol{o}^T\boldsymbol{x}^0 - x_j^1)\log(1-a_j) + \sum_{i=1}^n \log\left(P(X_i^0 = x_i^0)\right) \\ &\propto \log\left(\binom{\boldsymbol{o}^T\boldsymbol{x}^0}{x_j^1}\right) + x_j^1\log\left(\frac{a_j}{1-a_j}\right) + \boldsymbol{o}^T\boldsymbol{x}^0\log(1-a_j) + \sum_{i=1}^n \log\left(P(X_i^0 = x_i^0)\right), \end{aligned}$$

which in theory we can use to find the most likely initial condition given what we know about the stimulus distribution and measured data by setting $\nabla_{\boldsymbol{x}^0}\log(P(\boldsymbol{X}^0 = \boldsymbol{x}^0|X_j^1 = x_j^1)) = \boldsymbol{0}$.

F. State-Space and Statistical Mechanics

The next approach is to use statistical mechanics to enumerate the tree of probabilities for transitioning from one state to another. Consider a vector $\boldsymbol{X}^t$ of n random variables at time t, drawn from $X_j^{k+1} \sim \sum_{i=1}^{n} B(X_i^k, a_{ji})$. Because the state transition process does not depend on system memory, the probability of transitioning from any state $\boldsymbol{x}_m$ to any other state $\boldsymbol{x}_p$ is completely governed by a non-hierarchical sum of binomial distributions. Hence, we can enumerate all possible discrete states $\boldsymbol{x}_i$, as

$$\boldsymbol{x}_i \in \left(\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}, \cdots, \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}, \cdots, \begin{bmatrix} d \\ d \\ \vdots \\ d \end{bmatrix} \right),$$

and the probability of transitioning from any state $\boldsymbol{x}_m$ and $\boldsymbol{x}_p$ is given simply as the product of sums of binomial distributions.