

Rate-space attractors and low dimensional dynamics interact with spike-synchrony statistics in neural networks

Daniel N. Scott (daniel.scott@brown.edu)

Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, 190 Thayer Street
Providence, Rhode Island, United States

Michael J. Frank (michael.frank@brown.edu)

Carney Institute for Behavioral and Brain Sciences, and Department of
Cognitive, Linguistic, and Psychological Sciences, Brown University, 190 Thayer Street
Providence, Rhode Island, United States

Abstract

Mechanistic models of cognitive phenomena often make use of neural networks, which allow researchers to examine relationships between neurobiology and the computations suspected to underlie cognition. These models typically make use of neural firing rates, as do analyses of in-vivo data, with the dimension of neural dynamics receiving special attention. Treating time-binned spiking activity as a sequence of binary vectors (spike-words) should prove complementary to rate-space analyses, and has been shown to provide links with statistical physics. We investigate the interaction between these two analyses using theory and simulations to show how signatures of rate-dynamics are found in spike-word distributions. We find that a global integration over the eigenvalues of linear dynamics local to attracting subspaces can modify spike-synchrony, and we quantify how this impacts informational and thermodynamic properties of these systems. The research outlined here will have implications for the interpretation of neural data, the use of population codes for tasks such as Bayesian inference, and for various resource rational models attempting to bridge the gap between computation and implementation.

Keywords: rate-space; dynamical systems; spike-words; Ising models; statistical physics;

Introduction

Computational neuroscience has seen exponential growth in the number of simultaneously recorded neurons over approximately 60 years (Stevenson & Kording 2011). The proliferation of micro-electrode arrays, whole-brain calcium imaging (e.g. Kim et al 2017), and other technologies have made spatially and temporally precise large scale neural data relatively common. Temporal resolution at the scale of membrane-fluctuations has recently been achieved in optogenetics (Pitkevich 2018), and an expansion of techniques for analyzing networks of neurons has been ongoing in response to multi-unit data acquisition (Stevenson & Kording 2011). Methods for assessing the dimensions along which neural data vary are especially important in this context, as they allow researchers to ignore unimportant detail and develop intuitive theories of

network function (Cunningham & Yu, 2014; Pang, Lansdell, & Fairhall, 2016).

Rate spaces

A typical first step towards such analyses involves rate-coding a neuron's output by convolving a smoothing kernel with each acquired spike-train. Dimensionality reduction techniques may then be applied on a multidimensional array of neural firing rate data. The array's dimensions will be specified by (1) the neuron being recorded (2) the time within a task for a given recording (3) the repetition (trial) number associated with the recording and (4) any separation by task condition. This applies equally well to data generated by a neural network model. Dimensionality reduction therefore requires determining which covariance matrix to generate from the data for inspection, which in turn depends on a researchers scientific question of interest.

One of the most natural covariance matrices to examine arises from averaging over trials in a condition. This removes one dimension of the four dimensional array, leaving a matrix of firing rates by time for each task condition. Performing PCA on this data (i.e. per condition) will indicate, on average, how neurons' activities covary temporally. If a substantial fraction of the population variance is accounted for by only a few (e.g. tens of) components, single trial population data may then be usefully interpreted by analyzing trajectories in the space these span. This yields low-dimensional dynamics in the most common sense, where, to a good approximation, the state of the whole population evolves over time within a constrained region of the larger, a-priori accessible space.

There are important theoretical reasons to suspect low dimensional dynamics in real neural populations, and they have been widely observed (e.g. Churchland et al 2012). Limb movements, for example, require a radical reduction of dimensionality from controlling populations to effectors. Joint angles, velocities, and accelerations in an arm or leg have tens of degrees of freedom, compared with the hundreds or thousands of dimensions in directly associated neural activity.

Spike-word distributions

One alternative to rate-based analyses involves binning time and encoding spiking activity over a population as a sequence



of binary vectors s , termed spike-words (Schneidman et al 2006). These spike-words then have distributional statistics of interest, which provide an alternative characterization of a network's state, and can be interpreted in statistical mechanical terms with the use of Ising models. Ising models are maximum entropy (i.e. least possibly structured) models defined by the following pair of equations:

$$E = - \sum_{\langle i,j \rangle} J_{ij} s_i s_j - H \sum_i s_i$$

$$P(E) = \frac{1}{Z} \exp(-\beta E)$$

Here E is a function fit to the data by determining J_{ij} and H , whereas s_i and s_j are the components of a spike word (which, again, is a boolean list of which neurons are spiking at a given time). Beta is an analogue of (inverse) temperature, and can be considered constant for the present, as can Z , a normalizing factor. The summation over pairwise data terms is indexed according to a network associated with the data which encodes an effective connectivity hypothesis in a neural network setting. Traditionally, these pairwise terms reflect the influence of proximal lattice sites in a crystalline arrangement of magnetic atoms. In the general case, an arbitrary topology can be used in place of this lattice as a problem-specific way of instantiating node interdependence. This hypothesis is therefore related to pairwise correlations in the data, and for small networks using an all-to-all topology is reasonable, because the complete covariance matrix has $O(n^2)$ terms (given n neurons). The sum over s_i terms is related to the average value of nodes in the connectivity graph, i.e. the spike rates of individual neurons.

Ising models, and statistical physics more generally, are of importance to neuroscientists because they were developed to explain how complex, coordinated behaviors can emerge from systems of many simple, interacting components. Within neuroscience, they have been used to demonstrate that various neural systems appear to be 'nearly critical', which has many implications for their behavior, their information processing, and their responses to perturbation. A review of this literature would require a prohibitive amount of space, but the equations above illustrate that such models essentially rely on first and second order spiking information.

Varying spike-rates produce time-averaged spike-word distributions

Intuitively, spike-word distributions should contain some information about rate-space dynamics, and vice versa. In a two-neuron system, for example, if both neurons always spike synchronously, their rates must always be equal, and dynamics in the rate-space must therefore "live" in the one-dimensional subspace defined by this equality. Is there a corresponding claim we can make for spike-statistics on the basis of rate-dynamics? As will be demonstrated the answer is affirmative, even under the assumption of Poisson spiking.

Consider a neural network defined by the equation

$$\dot{x} = -x/\tau + A\phi(x) + \eta$$

which has been studied extensively (Sompolinsky 1988; Mastrogiuseppe & Ostojic 2018). Here x is a vector of neural firing rates, τ is a time constant, A is a matrix governing cross-neuron interaction (i.e. a connectivity matrix), and η is a noise term, and ϕ is a monotonic nonlinear function. Under certain conditions this model can effectively be reduced to the linear dynamical system:

$$\dot{x} = A'x + \eta' \quad (1)$$

In general, there is also typically some scale at which nonlinear dynamics in a system are locally linear, under mild assumptions, so that studying linear systems provides important insight into nonlinear ones.

Given a system like (1), the connectivity matrix A' can be used to instantiate low-dimensional or attractor dynamics. To gain an understanding of the impacts these dynamics have on spike-word distributions, we can establish attractive points or lines within a two-neuron rate space. As is well known, linear dynamical systems have analytic solutions, so that (1) has a timecourse described by:

$$x(t) = x(0)e^{A't} + \int_0^t e^{A'(t-\tau)} \eta(\tau) d\tau \quad (2)$$

This shows that the rate dynamics are determined by a transient arising from the network's initial conditions and by a convolution of a matrix exponential with the system's input. Matrix exponentials are diagonalized by the same transforms as the matrices being exponentiated, and their eigenvalues are the exponentials of the latter as well. Taken together, this means that the eigenvalues of the connectivity matrix A determine the time constants associated with decay along any dimensions which we care to bias the system against varying along. The eigenvalues of this matrix exhibit a quantitative relationship with the spike-word distribution the system produces.

To every point in a neural network's rate space, there corresponds a spike-word distribution, which may of course be marginal or conditional on other factors. In the low-intensity, independent Poisson case, each point's associated distribution is especially simple, because it is effectively the joint distribution of two independent low-probability coin-flipping processes acting at high temporal resolution. Of course when we observe the system over some period of time, we will only rarely have access to these distributions. Instead, the network will produce spikes according to a given point in rate space only transiently. The observed spike-word distribution will therefore reflect a weighted average of infinitely many individual spike word distributions, with each being weighted by the fraction of time the system spends at the corresponding point in rate-space. That is, the observed spike-word distribution is a path-integral over distributions, with weights determined by some function of the dynamics matrix and input to the system.

Given independent isotropic noise across neurons, the existence of an attractor in the network's rate dynamics will essentially operate as a coloring (that is, the opposite of whitening) transform. Put differently, it will "stretch" the noise according to the eigenvalues of the attractor. In the Bernoulli limit just discussed, the spike-word distribution is then easy to characterize mathematically.

The Bernoulli limit

Consider one particular constant rate configuration, given two neurons, such that neuron 1 fires at x_1 Hz and neuron 2 fires at x_2 Hz. Then assuming independent Poisson spiking, neuron 1 is well described by a Bernoulli process with a probability of spiking in each 1 ms time-bin given by $x_1[\text{Hz}]/1000[\text{ms}] \equiv p_1$ [spikes/ms], and similarly for neuron 2.

Now spike word probabilities are given transparently by the joint distribution

$$\begin{aligned} P(s_1 = 0, s_2 = 0) &= (1 - p_1)(1 - p_2) \\ P(s_1 = 1, s_2 = 0) &= (p_1)(1 - p_2) \\ P(s_1 = 0, s_2 = 1) &= (1 - p_1)(p_2) \\ P(s_1 = 1, s_2 = 1) &= (p_1)(p_2) \end{aligned}$$

where S_1 denotes the binary variable indicating spiking of neuron 1 in a given time-bin, and similarly for neuron 2. Extending this to the case of time dependent probabilities p_i requires computing expectation values, i.e. averages over time. While the details of this integration will be omitted for brevity, they are straightforward and admit various flexible assumptions.

A simple illustrative case involves parameterizing p_1 and p_2 themselves as random variables according to a Gaussian distribution centered at point (\bar{p}_1, \bar{p}_2) , with principal components rotated through an angle θ and variances given by (σ_1^2, σ_2^2) , or equivalently a covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{bmatrix}$$

which is diagonal - without loss of generality - precisely because of our choice of coordinates. Expectation values are easily performed given the associated gaussian integrals, with the resulting spike-word distribution given by:

$$\begin{aligned} C_{\text{indep. var}}(\theta, \sigma_1, \sigma_2) &= \frac{1}{\sqrt{2\pi}} \sin(\theta) \cos(\theta) (\sigma_1 - \sigma_2) \\ P(0, 0) &= (1 - \bar{p}_1)(1 - \bar{p}_2) + C_{\text{indep. var}} \\ P(1, 0) &= (\bar{p}_1)(1 - \bar{p}_2) - C_{\text{indep. var}} \\ P(0, 1) &= (1 - \bar{p}_1)(\bar{p}_2) - C_{\text{indep. var}} \\ P(1, 1) &= (\bar{p}_1)(\bar{p}_2) + C_{\text{indep. var}} \end{aligned}$$

This is a satisfying result for several reasons: First, we see that the same correction term is being applied to each spike word probability, with synchronous firing and synchronous silence being enriched relative to Poissonity by the same factor (and similarly for asynchrony). Next, we see that it is precisely the variances along each dimension in rate space

which dictate the correction terms. Finally, the problem's inherent symmetry between neurons 1 and 2 is reflected in the term $\sin(\theta) \cos(\theta)$ which peaks for $\theta = \pi/4$. For poisson processes with correlation ρ , this can be further extended by a second correction term given

$$C_{\text{corr.}} = \frac{1}{2}(\bar{p}_1 + \bar{p}_2) - \frac{1}{\sqrt{2\pi}} \sigma_2 (\sin \theta + \cos \theta) - \bar{p}_1 \bar{p}_2$$

so that each probability is in fact given by

$$P(s_1, s_2) = P(s_1, s_2)_{\text{static, indep}} + [1 - \rho] C_{\text{indep. var}} + \rho C_{\text{corr.}}$$

which has intuitive properties under inspection asymptotic cases such as the $\rho = 1$ result.

Both of the results indicated above hold equally well for networks comprised of greater numbers of neurons, because pairwise spike-synchrony statistics depend on marginal pairwise distributions of rates. These marginal pairwise distributions can then be examined via the equations given. Third-order spike synchrony statistics are determined by a similar marginalization argument, but these and higher interaction terms grow combinatorially in number with neuron count, and therefore are difficult to estimate from data. The work we will present expands on the mathematics above, relaxing the Poisson-spiking assumption, treating transfer-function related corrections to equation (1), characterizing changes in spike-word distribution entropy, and examining the implications of these for various statistical mechanical quantities.

Network simulation

As noted above, if one injects uncorrelated white noise into a two-dimensional attractor network based on (1), assuming that A has two real, non-positive eigenvalues, the resulting dynamics effectively rotate and color the noise. Variance scaling is induced along the eigendirections according to each one's eigenvalues. Thus, given Gaussian noise, one should observe a Gaussian distribution of firing-rates over sufficiently long time-scales, and the spike-word distribution for the system should thus behave as predicted above.

To assess this prediction, we simulated various linear dynamical systems, and present one such simulation now. An attractor was centered at $x = (25 \text{ Hz}, 25 \text{ Hz})$ in the neural firing rate space with $\theta = \pi/4$ and 10ms auto-correlated Gaussian noise was injected. The noise had a mean of 0 and standard deviation of 0.5, giving a standard deviation along the elongated attractor axis of approximately 12.5 Hz. The specific dynamics matrix used was

$$A = \begin{bmatrix} -12.5 & 7.5 \\ 7.5 & -12.5 \end{bmatrix}$$

which has eigendirections (1,1) and (-1,1) with associated eigenvectors (-5, 30). These firing rates were chosen to be approximately physiologically realistic for cortical neurons, and a simulation-time of 30 minutes was used as a trade-off between longer simulations (which yield better sampling of the

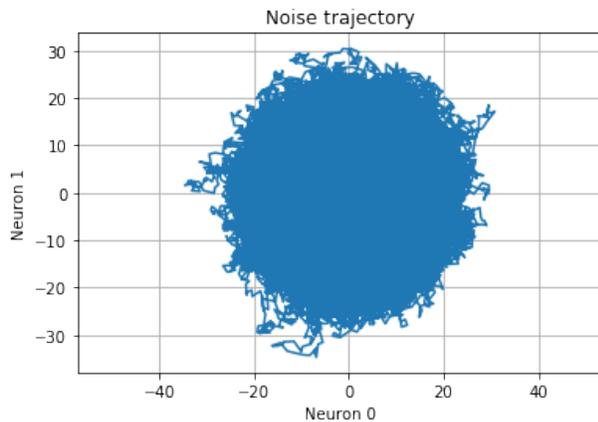


Figure 1: The time course of noise driving our dynamics example. It is a vector comprised of two independent 10ms auto-correlated Gaussian noise processes with mean zero and standard deviation 0.5, sampled every 1ms.

rate space) and shorter ones (which consume fewer computational resources). Given these parameters we consistently find an increase in synchronous spiking corresponding to approximately 10% greater incidence than would be expected under a Poisson distribution generating spikes at the average rates.

Discussion

We have presented a summary of ongoing work undertaken to characterize the relationship between spike-word distributions and rate-space dynamics. We find that attracting subspaces of the rate dynamics do indeed systematically modify spike word distributions. These changes to the spike word distributions can be substantial and they broadly result from the spread of rates along the line of equality in each pairwise subspace, relative to orthogonal spread. This suggests that, in analyses of neural data exhibiting low-dimensional dynamics, the spatial orientations of those dynamics play an important role in determining spike word distributions. In particular, this work suggests neural data should be examined for spread across various dimensions in the pairwise rate spaces in order to bound how much of the observed excesses of pairwise synchrony in spike-words is attributable to such dynamics. This will have implications for our understanding of statistical physics approaches which indicate criticality, and in turn on downstream translations of such work into computational theories of cognitive function.

Acknowledgments

The authors do not wish to acknowledge specific sources of funding at this time.

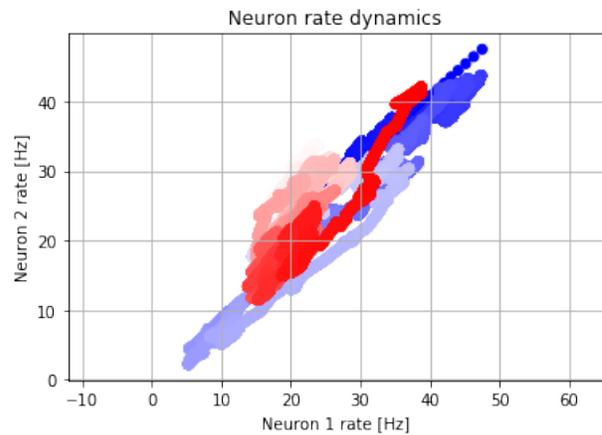


Figure 2: The dynamics of the driven linear dynamical system; Blue denotes earlier simulation times and red denotes later ones.

References

- Churchland, M. M., Cunningham, J. P., Kaufman, M. T., Foster, J. D., Nuyujukian, P., Ryu, S. I., and Shenoy, K. V. (2012). Neural population dynamics during reaching. *Nature*, 487(7405):51–56.
- Cunningham, J. P. and Yu, B. M. (2014). Dimensionality reduction for large-scale neural recordings. *Nature Neuroscience*, 17(11):1500–1509.
- Kim, D. H., Kim, J., Marques, J. C., Grama, A., Hildebrand, D. G. C., Gu, W., Li, J. M., and Robson, D. N. (2017). Pan-neuronal calcium imaging with cellular resolution in freely swimming zebrafish. *Nature Methods*, 14(11):1107–1114.
- Mastrogiuseppe, F. and Ostojic, S. (2018). Linking Connectivity, Dynamics, and Computations in Low-Rank Recurrent Neural Networks. *Neuron*, 99(3):609–623.e29.
- Pang, R., Lansdell, B. J., and Fairhall, A. L. (2016). Dimensionality reduction in neuroscience. *Current Biology*, 26(14):R656–R660.
- Schneidman, E., Berry, M. J., Segev, R., and Bialek, W. (2006). Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012.
- Sompolinsky, H., Crisanti, A., and Sommers, H. J. (1988). Chaos in Random Neural Networks. *Physical Review Letters*, 61(3):259–262.
- Stevenson, I. H. and Kording, K. P. (2011). How advances in neural recording affect data analysis. *Nature Neuroscience*, 14(2):139–142.