

Discovery of salient low-dimensional dynamical structure in neuronal population activity using Hopfield networks

Felix Effenberger, Max-Planck-Institute for Mathematics in the Sciences, Leipzig, Germany

Christopher Hillar, Redwood Center for Theoretical Neuroscience, Berkeley, CA 94720, USA

Summary. We present here a novel method for finding and extracting salient low-dimensional representations of the dynamics of populations of spiking neurons. This is a classical problem in data analysis of parallel spike trains, and quite a number of approaches to detect and classify recurring spatiotemporal patterns (STP) of neural population activity were proposed [GR10]. Yet, most published methods so far assume a noiseless scenario and either focus on synchrony detection and / or seek to classify exactly recurring STP in neuronal activity. Given the usually high variability of population responses to stimuli, the re-occurrence of such exactly repeating STP becomes more and more unlikely with increasing population size, though. Assuming that despite this variability, network activity is not random per se (under the well-supported hypothesis that the population has to code information about stimuli in some form of STP), a much more plausible situation is that some underlying STP appears in several “corrupted” variants differing in a few missing or excess spikes (characterized by a low Hamming distance to a true, underlying STP). The proposed method is robust to this variability in the signal and able to extract the underlying recurring patterns, even for seldomly occurring STP and large population sizes. Modeling furthermore the sequence of occurring STP as a Markov process, we are able to extract low-dimensional representations of neural population activity. We demonstrate the approach on a data set obtained in rat barrel cortex and show that it is able to extract a remarkably low-dimensional yet accurate representation of network activity observed during the experiment. **Background.** We fit Hopfield networks (HN) to windowed, binned spiking activity of a population of cells using minimum probability flow (MPF), a novel probabilistic learning rule for Hopfield networks with many desirable properties [SDBD11]. The training data X are obtained by sliding a window of given length w over a binary matrix of dimension $N \times T$ representing the binned spiking activity of N cells over a time period of T bins, yielding $T - w$ binary vectors of length Nw as training data. After fitting a HN with Nw nodes on the data using MPF, we converge each window of the raw, binned spiking data to its Hopfield memory. We label the sequence of occurring memories by natural numbers in the order of their appearance so that we obtain a memory sequence $S = (s_1, \dots, s_{T-w})$, with $s_i \in \{m_1, \dots, m_k\} = M$, $k \leq T - w$, where M denotes the set of all distinct memories in S . Note that usually $k \ll |S|$, as STP occurring in the raw data that have low Hamming distances are likely to converge to the same memory under the Hopfield dynamics as a result of fitting the network with MPF. For each memory $m_i \in M$ we compute all pairwise one-step Markov transition probabilities to $m_j \in M$ ($1 \leq j \leq k$) using data from S and the entropy over this probability distribution for each m_i , which we call the entropy of the memory m_i and denote by $H(m_i)$. The entropy of a memory is a measure for how predictable the following network state is, according to the observed data. Memories with a more restricted set of following network states have lower entropy, ones with less predictable states have higher entropy. $H(m_i)$ can therefore be seen as a local measure (in time) for how deterministic the network dynamics evolve from that memory. We then construct a directed graph with the elements of M as nodes. Two nodes m_i, m_j are connected by an edge (m_i, m_j) of weight w if their Markov transition probability $w = P(m_j|m_i)$ obtained from S is non-zero. We call this graph the *Markov graph* G_M of S . Paths and cycles (i.e. simple closed paths) in G_M along nodes with low entropy correspond to sequences of memory labels and thus sequences of STP of spiking activity that are prominently and reliably generated by the neuronal population. We applied the proposed method to spiking data of recorded in the rat barrel cortex during repeated whisker stimulation [MCT⁺11] ($N = 16, T = 3.4 \cdot 10^4, w = 10, 1$ ms bins). For this data set we find a ‘central’ node m_α in G_M that has a high degree (sum of in- and out-degrees). This is characteristic for a situation in which the node is the termination (resp. starting) point of prominently occurring STP of network activity. Interestingly, the memory m_α occurs very frequently in the data ($p > 0.9$) and the node has low entropy. This we expect from a network’s low-energy base state that it prominently rests in and repeatedly returns to. Using the information of the stimulus protocol, we indeed found that m_α corresponds to the resting state of the network. Moreover we expect that cycles in G_M starting in m_α can give insight on how the network is driven out of its resting state (by some stimulus) and enters a transient sequence of excited states before falling back to the resting state. Tracing such cycles in G_M (and scoring them by their entropy, obtained as a weighted sum of the entropies of the cycle’s nodes as a measure for how reliably that cycle is “visited” by the network dynamics), we find that the most STP associated with low entropy cycles indeed correspond closely to the average network response to whisker stimulation (that we computed from the raw data using knowledge of the stimulation protocol), but without using any knowledge of the stimulation protocol. We find this a quite remarkable result and propose the method as a novel tool in mining parallel spike trains for possibly low-dimensional underlying network dynamics. An open source software allowing for the wider application of the method is to be released soon. **Note.** Illustrations available online, <http://db.tt/YmOisY86>.

References

- [GR10] S. Grün and S. Rotter. *Analysis of parallel spike trains*. Springer, 2010.
- [MCT⁺11] M Minlebaev, M Colonnese, T Tsintsadze, A Sirota, and R Khazipov. *Science*, 334(6053):226–229, 2011.
- [SDBD11] J. Sohl-Dickstein, P.B. Battaglini, and M.R. DeWeese. *Physical review letters*, 107(22):220601, 2011.