Temporal Infomax on Markov Chains with Input Leads to Finite State Automata

Thomas Wennekers and Nihat Ay [1]

Max-Planck-Institute for Mathematics in the Sciences,

Inselstr. 22–26, D-04103 Leipzig, Germany

Abstract

Information maximization between stationary input and output activity distributions of neural ensembles has been a guiding principle in the study of neural codes. We have recently extended the approach to the optimization of information measures that capture spatial *and* temporal signal properties. Unconstrained Markov chains that optimize these measures have been shown to be almost deterministic. In the present work we consider the optimization of stochastic interaction in constrained Markov chains where part of the units are clamped to prescribed processes. Temporal Infomax in that case leads to finite state automata.

Key words: Markov model; Information maximization; Finite state automata;

[1] Tel +49-341-9959527 Fax +49-341-9959555, Email (wenneker,nay)@mis.mpg.de

1 Introduction

One of the most basic questions in computational neuroscience is that for the nature of neural codes. Experiments suggest a considerable interaction of neurons already on the level of spikes, e.g., expressed by spatio-temporal correlations [1,4,6,7]. A well-known measure that quantifies relations of interacting units is the so-called *mutual information*: The Kullback-Leibler divergence

$$I(p) \;:=\; D(p\,||\,p_1 \otimes \cdots \otimes p_N) \;=\; \sum_{\nu=1}^{N} H(p_\nu) - H(p) \;, \tag{1}$$

where $H(\cdot)$ denotes the Shannon entropy and p_ν the ν'th marginal of p, measures the "distance" of p from the factorized distribution $p_1 \otimes \cdots \otimes p_N$. It is a natural measure for "spatial" interdependence of N stochastic units and a starting point of recent approaches to neural complexity [5,6]. In order to capture intrinsically temporal aspects of dynamic interaction, I in (1) has been extended by Ay [2] to the dynamical setting of Markov chains, where it is referred to as (*stochastic*) *interaction*. The optimization of stochastic interaction in Markov chains has been shown to result in almost deterministic dynamical systems [3]. This work neglegted external input into the considered systems. The present study, therefore, presents optimized Markov chains, where a part of the system is clamped to externally prescribed stochastic processes. Surprisingly, the optimized processes turn out to be finite state automata.

2 Temporal Infomax on Constrained Markov Chains

Consider a set $V = \{1, \ldots, N\}$ of binary units with state sets $\Omega_\nu = \{0, 1\}$, $\nu \in V$. For a subsystem $A \subset V$, $\Omega_A := \{0, 1\}^A$ denotes the set of all configurations restricted to A, and $\bar{P}(\Omega_A)$ is the set of probability distributions on Ω_A. Given two subsets A and B, where B is non-empty, $\bar{K}(\Omega_B \,|\, \Omega_A)$ is the set of all Markov transition kernels from Ω_A to Ω_B. If $A = B$ we use the abbreviation $\bar{K}(\Omega_A)$. For a probability distribution $p \in \bar{P}(\Omega_A)$ and a Markov kernel $K \in \bar{K}(\Omega_A)$, the *conditional entropy* of (p, K) is defined as

$$H(p, K) \;=\; - \sum_{\omega, \omega' \in \Omega_A} p(\omega)\, K(\omega' \,|\, \omega) \, \ln K(\omega' \,|\, \omega) \; . \tag{2}$$

$H(p, K)$ has been used in [2] to generalize (1) to Markov transitions. For that purpose, marginal kernels $K_\nu \in K(\Omega_\nu)$ are defined for strictly positive p as

$$K_\nu(\omega'_\nu \,|\, \omega_\nu) \;:=\; \frac{\sum_{\substack{\sigma, \sigma' \in \Omega_V \\ \sigma_\nu = \omega_\nu, \sigma'_\nu = \omega'_\nu}} p(\sigma)\, K(\sigma' \,|\, \sigma)}{\sum_{\substack{\sigma \in \Omega_V \\ \sigma_\nu = \omega_\nu}} p(\sigma)}, \qquad \omega_\nu, \omega'_\nu \in \Omega_\nu, \;\; \nu \in V \; . \tag{3}$$

Then, the *stochastic interaction measure* of K with respect to p is given by

$$I(p, K) \;:=\; \sum_{\nu \in V} H(p_\nu, K_\nu) \;-\; H(p, K) \; , \tag{4}$$

continuously extended to $\bar{P}(\Omega_V) \times \bar{K}(\Omega_V)$. Equation (4) provides the required generalization of (1) to Markov transitions. $I(p, K)$ is large if the marginal transitions have high entropy, but that of the full transition is low. Thus, supposed the current state $\omega \in \Omega_V$ is known, the next state is predictable with high probability, but, conversely, not much information is gained from knowledge about single units, ω_ν.

In the following we consider Markov chains $X_n = (X_{\nu,n})_{\nu \in V}$, $n = 0, 1, 2, \ldots$, given by an initial distribution $p \in \bar{P}(\Omega_V)$ and a kernel $K \in \bar{K}(\Omega_V)$. We restrict attention to *parallel* Markov chains. A Markov kernel $K \in \bar{K}(\Omega_V)$ is called *parallel* if there exist kernels $K^{(\nu)} \in \bar{K}(\Omega_\nu \,|\, \Omega_V)$, $\nu \in V$, such that

$$K(\omega' \,|\, \omega) \;=\; \prod_{\nu \in V} K^{(\nu)}({\omega'}_\nu \,|\, \omega), \qquad \text{for all} \quad \omega, \omega' \in \Omega_V. \tag{5}$$

Parallel Markov chains are a more natural assumption in neural modeling than general Markov chains, because the action potential generation of a neuron depends only on its own input and internal dynamics. Therefore, each unit ν in (5) is given an individual kernel $K^{(\nu)}$. Spatial correlations in firing patterns may nonetheless result from common input, i.e., the global state ω in (5).

In [3] we studied parallel Markov chains which maximized $I(p, K)$ under no further constraint regarding K. The optimized chains were shown to be almost deterministic and representable in phase space by more or less complex graphs consisting of transient trajectories and nested loops which correspond with repeteting firing patterns. In the present work, in contrast, we show simulations of parallel Markov chains which maximize $I(p, K)$ under the *additional constraint* that the kernels $K^{(\nu)}$ of a subset of units in (5) are fixed during the optimization process. The simulations show that Markov chains with such clamped "input-units" converge to finite state automata.

3 Simulations

The simulations displayed in the following implement the usual Markov dynamics on N binary units together with a random search scheme to optimize the stochastic interaction of the Markov chains: The interaction, I, is computed with respect to an induced stationary probability distribution of a parallel Markov kernel, and starting from initial random values the kernel is iteratively perturbed such that I increases (cf. [3]). In contrast to [3], however, the optimization here is not unconstrained, but $B < N$ kernels $K^{(\nu)}$ in (5) are fixed during the optimization. In especially, these kernels are chosen independent of the state of the internal units. In the simplest case the 'input units' are just B independent Bernoulli processes with fixed rates. We define 'burst-processes' by $a_\nu := K^{(\nu)}(1|0) = 1 - K^{(\nu)}(0|0)$ and $b_\nu := K^{(\nu)}(0|1) = 1 - K^{(\nu)}(1|1)$. If $a_\nu = 1 - b_\nu$ they implement Bernoulli processes with rate a_ν. For other $a_\nu,\ b_\nu \in]0,1[$ they generate burst-like firing patterns with a_ν related to the mean interburst interval and $1 - b_\nu$ to the mean burst duration. Spatial correlations between input units can be introduced by appropriate input kernels.

************ Figure 1 somewhere here ************

Figure 1 shows an optimized system with three units, where unit 3 has been clamped to a burst process with $a_3 = 0.1$ and $b_3 = 0.1$. The unit has a firing probability of $\lambda = a_3/(a_3 + b_3) = 0.5$, and if it has fired the probability that it fires again in the next step is quite high: $K^{(\nu)}(1|1) = 1 - b_3 = 0.9$.

The converged Markov transition kernel is displayed in the upper left of Fig.1 where circle area represents transition probability. Note, that the kernel can be divided into blocks according to the sub-states z, z' of the internal units, i.e., the first $N-B$ bits of the total binary state representation: Given an internal state z and an input a (the B least significant bits of the total state), the internal target state z' in the next step is completely determined, but only the next input a' is random. This means, that the Markov chain can be represented by a deterministic finite state automaton (DFA) with nodes labeled by the internal states and edges labeled by the (random) input patterns.

************ Figure 2 somewhere here ************

************ Figure 3 somewhere here ************

Figures 2 and 3 display examples with three units each and $B = 2$. In Fig. 2 unit 2 is a Bernoulli process with $\lambda = 0.1$ and unit 3 an independent but temporally correlated burst process with $a_3 = b_3 = 0.1$. In Fig. 3 the inputs (2,3) are spatially correlated near Bernoulli processes. The internal unit 1 combines activity from both input units. Both systems converge to comparable DFAs. Nonetheless, in Fig. 2 the externally induced temporal correlations are still expressed in the precise transition probabilites, e.g., three strong transitions to 000 because all units fire relatively seldom, or three strong transitions from states with unit 3 active to 101, because unit 3 fires long bursts but unit 2 is mostly silent. Similarly, spatial correlations are expressed in Fig. 3 by the dominating transitions $000 \to 000$ and $100 \to 100$, which reflect that input units 2 and 3 are simultaneously silent in most steps. Activation of one input unit, however, often induces transitions to states where the same or other input unit fires, since both units excite themselves and each other.

************ Figure 4 somewhere here ************

Finally, if N grows, it turns out that the optimization does not always converge to *deterministic* automata, but internal states become increasingly likely that reveal transitions to several targets for the same input, cf., Fig. 4. Such systems correspond with non-deterministic finite state automatas (NFAs). However, it can be shown mathematically that these NFAs must always be "almost" deterministic revealing a number of targets per node that grows at most linearly in N as compared to the possible 2^N target states [8].

References

[1] M. Abeles, Corticonics: Neural circuits of the cerebral cortex (Cambridge University Press, Cambridge, 1991).

[2] N. Ay, Information Geometry on Complexity and Stochastic Interaction, Submitted (2002).

[3] N. Ay and T. Wennekers, Dynamics of strongly interacting Markov chains, Submitted (2002).

[4] R. Eckhorn, Neural mechanisms of scene segmentation: Recordings from the visual cortex suggest basic circuits for linking field models, IEEE Transactions

on Neural Networks 10 (1999) 464–479.

[5] R. Linsker, From Basic Network Principles to Neural Architecture, Proceedings of the National Academy of Sciences USA 83 (1986) 7508–7512.

[6] F. Rieke, D. Warland, R. Ruyter van Steveninck and W. Bialek, Spikes: Exploring the Neural Code (MIT Press, Cambridge, 1998).

[7] W. Singer and C. M. Gray, Visual feature integration and the temporal correlation hypothesis, Annual Review of Neurosciences 18 (1995) 555-586.

[8] T. Wennekers and N. Ay, Temporal Infomax and Determinism in Markovian Neural Systems, in preparation.

Thomas Wennekers studied physics at the University of Düsseldorf and computer science at the University of Ulm, where he received a Ph.D. in 1998. He does research at the Max-Planck-Institute for Mathematics in the Sciences in Leipzig in the fields of computational neuroscience and brain theory.

Nihat Ay studied Mathematics at the Ruhr-University Bochum and obtained a Ph.D. in Mathematics from the University of Leipzig in 2001. He currently works at the Max-Planck-Institute for Mathematics in the Sciences in Leipzig on information geometry and its applications in complex adaptive systems.

Figure Legends

Fig. 1: Optimized example system with 2 internal units (1,2) and 1 input (3). Upper left: Markov kernel. The total kernel can be segregated into blocks defined by internal states $(z,\ z')$ only. Right: corresponding automaton with internal states (z) as node- and input (a) as edge-symbols. Below: sample trajectory.

Fig. 2: Similar as Fig. 1 but with two independent input units (2,3). Firing of the internal third unit (1) depends on both inputs but strongly reflects unit 3 activity.

Fig. 3: Similar as Fig. 2 but with correlated inputs (2,3) favoring coactivation. The DFA is nonetheless similar to that in Fig. 2. Note also that the internal unit 1 acts as a latch (hysteresis, memory) activated from state 0 to 1 by an input of 1 from unit 3 and inactivated from 1 to 0 by any further additional input of 1.

Fig. 4: Two optimized Markov matrices for $N = 4,\ B = 2$. The example in A is deterministic, but B reveals non-deterministic internal transitions (arrow).

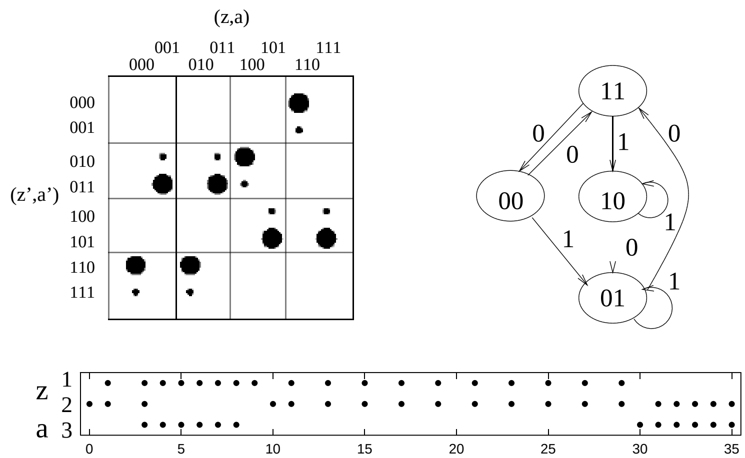

Fig. 1. Wennekers and Ay

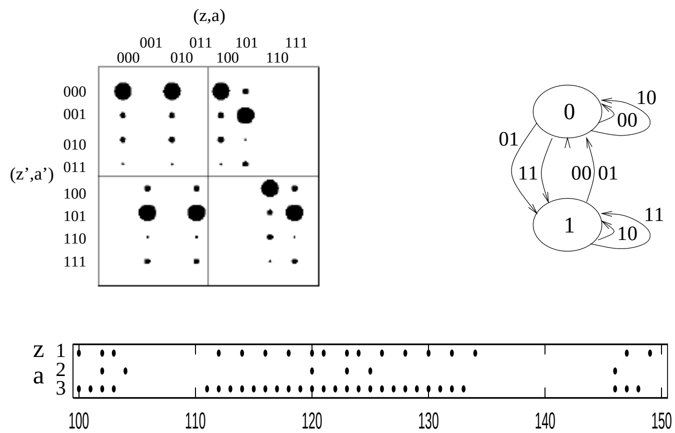

Fig. 2. Wennekers and Ay

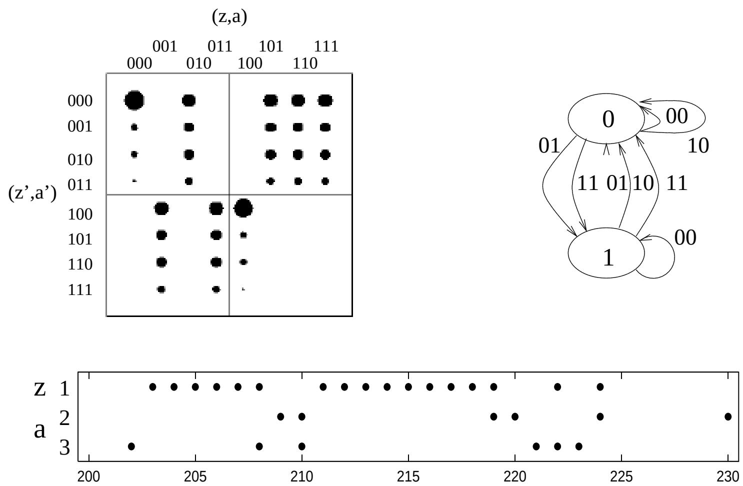

Fig. 3. Wennekers and Ay

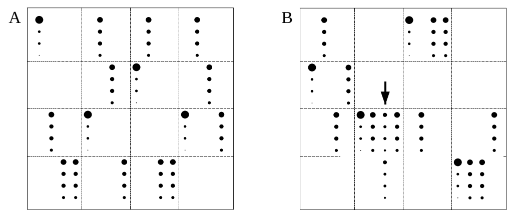

Fig. 4. Wennekers and Ay