View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by INRIA a CCSD electronic archive server

Estimating maximum entropy distributions from periodic orbits in spike trains

Bruno Cessac, Rodrigo Cofre

To cite this version:

Bruno Cessac, Rodrigo Cofre. Estimating maximum entropy distributions from periodic orbits in spike trains. [Research Report] RR-8329, INRIA. 2013. hal-00842776

HAL Id: hal-00842776

https://hal.inria.fr/hal-00842776

Submitted on 9 Jul 2013

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

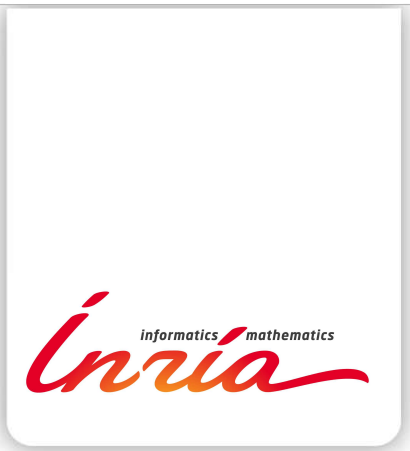
L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating Maximum Entropy distributions from Periodic Orbits in Spike Trains

Bruno Cessac & Rodrigo Cofré

RESEARCH REPORT
N° 8329
July 2013
Project-Team Neuromathcomp

ISSN 0249-6399 ISRN INRIA/RR--8329--FR+ENG



Estimating Maximum Entropy distributions from Periodic Orbits in Spike Trains

Bruno Cessac * & Rodrigo Cofré *

Project-Team Neuromathcomp

Research Report n° 8329 — July 2013 — 36 pages

Abstract: We present a method allowing to compute the shape of a Maximum Entropy potential with spatio-temporal constraints, from the periodic orbits appearing in the spike train.

Key-words: Gibbs distributions, Maximum entropy, Hammersley-Clifford decomposition, Periodic orbits

* NeuroMathComp, INRIA, 2004 Route des Lucioles, 06902 Sophia-Antipolis, France.

RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

estimation des distributions d'entropie maximale à partir des orbites périodiques dans les trains de potentiel d'action

Résumé : Nous présentons une méthode permettant de calculer la forme du potentiel d'entropie maximale en prenant en compte des contraintes spatio-temporelles, à partir des orbites périodiques apparaissant dans le train de potentiel d'action.

Mots-clés : Distributions de Gibbs, Entropie maximale, décomposition de Hammersley-Clifford, orbites périodiques

Contents

1 Introduction

The maximum entropy principle (MaxEnt) has been applied by several authors to characterize statistically the spiking response of neuronal networks, especially in the retina [36, 31, 14, 15, 44]. This approach consists of fixing a set of constraints, determined as the empirical average of quantities "Observables" measured from the spiking activity: for example firing rate of neurons or pairwise correlations. Maximizing the statistical entropy given those constraints provides a unique probability distribution, called a Gibbs distribution, characterizing the data. In particular, fixing firing rates and the probability of pairwise coincidences of spikes leads to a Gibbs distribution having the same form as the Ising model [36]. One of the main interest of this approach is to allow the construction of probabilities fitting the data on the basis of a general principle and a choice of constraints given a priori. This results in an overwhelming reduction of complexity if one compares the (relatively) small number of parameters defining the Gibbs distribution, to the huge dimensionality in the space of spike patterns. This method suffers unfortunately two caveats:

- It assumes stationarity in the data;
- The choice of constraints is ad-hoc.

We focus here on the second aspect.

After [36, 31] several authors have proposed to go "beyond Ising", including additional constraints such as the probability of instantaneous triplets, quadruplets [14]. These additional constraints do not treat memory effects in statistics however. As the consequence, they correspond to statistical models where successive spike times are independent. This is quite questionable as far as neural dynamics is concerned [40, 26]. As a matter of fact, there exists a well-defined theory allowing to handle general spatio-temporal spike events of type: "neuron i_1 is firing at time t_1, neuron i_2 is firing at time t_2" and so on, in the realm of MaxEnt [44]. Obviously, introducing such general spatio-temporal spike events leads to an exponential combinatorial explosion in the number of possible constraints, thus in the number of parameters in the Gibbs distribution, rendering impossible a reliable fit and/or leading to over-fitting. Certainly, not all possible constraints has to be considered: the quest of hidden laws in neural dynamics from spike analysis is largely based on the idea that statistics can be described with a few relevant parameters/constraints. But, how to select them ?

This paper addresses this question on a mathematical ground. Assume that spike statistics has been generated by a hidden stationary Markov process where transition probabilities are known (either because they can be computed exactly in a neural network model, or because some model of fit, like the General Linear Model (GLM) [1, 32], has been proposed, or because they have been estimated from a raster). Using the Hammersley-Clifford decomposition theorem [17] and a result in ergodic theory from Livšic [21], we show that the equilibrium probability of the Markov chain is a Gibbs distribution whose potential can be explicitly computed. This provides us a method to infer the MaxEnt constraints (the shape of the potential) from the transition probabilities. This establishes therefore an analytic relation between structural parameters and MaxEnt parameters. The paper is organized as follows: Section 2 presents the mathematical setting on which our method is grounded. Section 3 presents the method based on periodic orbits decomposition. In Section 4 we present several numerical tests illustrating the method.

This work has been widely inspired by Pollicott and Weiss paper "Free energy as a dynamical invariant (or can you hear the shape of a potential?)" [33] itself inspired by the Kac's seminal paper "Can one hear the shape of a drum?" [18].

2 Setting

In this section we provide the main tools and definitions used in the paper. We present the MaxEnt in the context of spike train statistics and the Hammersley-Clifford decomposition of observables.

2.1 Spike trains

We consider a network of N neurons. We assume that there is a minimal time scale, δ, set to 1 without loss of generality such that a neuron can at most fire a spike within a time window of size δ. This provides a time discretization labeled with an integer time n. To each neuron k and discrete time n one associates a spike variable[1] $\omega_k(n) = 1$ if neuron k fires at time n and $\omega_k(n) = 0$ otherwise. The state of the entire network in time bin n is thus described by a vector $\omega(n) \stackrel{\text{def}}{=} [\omega_k(n)]_{k=1}^N$, called a *spiking pattern*. A *spike block* is a finite ordered list of such vectors, written:

$$\omega_{n_1}^{n_2} = \{\, \omega(n) \,\}_{\{n_1 \leq n \leq n_2\}},$$

where spike times have been prescribed between time n_1 to n_2. The *time-range* (or "range") of block is $n_2 - n_1 + 1$, the number of time steps from n_1 to n_2. The *degree* d of a block is its number of non-zero bits. Here is an example of a spike block with $N = 4$ neurons, range $R = 3$ and degree 7,

$$\begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

A *spike train* or *raster* is a spike block ω_0^T from some initial time 0 to some final time T. Although, experimental spike trains have always a finite duration T, it is useful to us to consider infinite rasters with $T \to +\infty$. To alleviate notations we simply write ω for a spike train. We note $\Omega \equiv \{\, 0, 1 \,\}^{N\,\mathbb{N}}$ the set of spike trains.

The *time shift* $\mathcal{T}$ transforms the raster $\omega = \omega(0)\omega(1)\omega(2)\ldots$ into the raster $\omega' = \mathcal{T}\omega = \omega(1)\omega(2)\ldots$, i.e. it shifts ω left-wise so that such that $\omega'(k) = \omega(k+1)$.

2.1.1 Observables

An *observable* is a function $\mathcal{O}$ which associates a real number $\mathcal{O}(\omega)$ to a spike train. In the realm of statistical physics common examples of observables are the energy or the number of particles. In the context of neural networks examples are the number of neuron firing at a given time n, $\sum_{k=1}^N \omega_k(n)$, or the function $\omega_{k_1}(n_1)\omega_{k_2}(n_2)$ which is 1 if neuron k_1 fires at time n_1 and neuron k_2 fires at time n_2 and is 0 otherwise.

Typically, an observable does not depend on the full raster, but only on a sub-block of it. The *time-range* (or "range") of an observable is the minimal integer $R > 0$ such that, for any raster ω, $\mathcal{O}(\omega) = \mathcal{O}\left(\omega_0^{R-1}\right)$. The range of the observable $\sum_{k=1}^N \omega_k(n)$ is 1; the range of $\omega_{k_1}(n_1)\omega_{k_2}(n_2)$ is $n_2 - n_1 + 1$. From now on, we restrict to observables of range R, fixed and finite. We set $D = R - 1$.

An observable is *time-translation invariant* if, for any time $n > 0$ we have $\mathcal{O}\left(\omega_n^{n+D}\right) \equiv \mathcal{O}\left(\omega_0^D\right)$ whenever $\omega_n^{n+D} = \omega_0^D$. The two examples above are time-translation invariant. The observable $\lambda(n_1)\omega_{k_1}(n_1)\omega_{k_2}(n_2)$, where λ is a real function of time, is not time-translation invariant. Basically, time-translation invariance means that $\mathcal{O}$ does not depend explicitly on time.

[1] We use the notation ω to differentiate our binary variables $\in \{\, 0, 1 \,\}$ to the notation σ or S used for "spins" variables $\in \{\, -1, 1 \,\}$. The choice of a $0, 1$ variable for the spike considerably simplifies the computations and results of the paper.

We focus on time-translation invariant observables from now on.

Prominent examples of time-translation invariant observables with range R are products of the form:

$$m_{p_1,\ldots,p_r}(\omega) \stackrel{\text{def}}{=} \prod_{u=1}^{r} \omega_{k_u}(n_u). \tag{1}$$

where p_u, $u = 1 \ldots r$ are pairs of spike-time events (k_u, n_u), $k_u = 1 \ldots N$ being the neuron index, and $n_u = 0 \ldots D$ being the time index. Such an observable takes therefore values in $\{0, 1\}$ and is 1 if and only if $\omega_{k_u}(n_u) = 1$, $u = 1 \ldots r$ (neuron k_1 fires at time n_1, ..., neuron k_r fires at time n_r). We allow the extension of the definition (1) to the case where the set of pairs $p_1, \ldots, p_r$ is empty and we set $m_\emptyset = 1$. For a number N of neurons and a time range R there are thus $2^{N\,R}$ such possible products.

We show below that any observable of finite range can be represented as a linear combination of products (1). They constitute therefore a canonical basis for observable representation.

Another prominent example of observable is the function called "energy" or *potential* in the realm of the MaxEnt, $\mathcal{H}(\omega) = \sum_{k=1}^{K} \beta_k \mathcal{O}_k(\omega)$, where $\beta_k > -\infty$ is a real number called the parameter conjugated to $\mathcal{O}_k$. Without loss of generality a potential $\mathcal{H}$ of range R can also be written as a linear combination of the $L(N, R) = 2^{NR} - 1$ possible monomials (1):

$$\mathcal{H} = \sum_{l=0}^{L(N,R)} h_l m_l, \tag{2}$$

(where some coefficients h_l in the expansion may vanish). By analogy with spin systems, monomials somewhat constitute spatio-temporal interactions between spikes: the monomial $\prod_{u=1}^{r} \omega_{k_u}(n_u)$ contributes to the total energy $\mathcal{H}(\omega)$ of the raster ω if and only if neuron k_1 fires at time n_1, ..., neuron k_r fires at time n_r in the raster ω. The number of pairs in a monomial (1) defines the degree of an interaction: degree 1 corresponds to "self-interactions", degree 2 to pairwise, and so on.

Note that what are considering spatio-temporal spikes interactions. This allows us to introduce causality in spike statistics estimation, where events arising at consecutive times are not independent.

2.2 The maximum entropy principle

The MaxEnt provides a method to estimate a probability distribution μ from a sample (here a raster). A central assumption is that μ is time-translation invariant (stationarity). We call $\mathcal{M}$ the set of time-translation invariant probability measures on Ω. For an observable $\mathcal{O}$ we denote by $\mu[\mathcal{O}]$ the average of $\mathcal{O}$ with respect to μ. We are not assuming here that successive spikes patterns are independent i.e. $\mu\left[\omega_{n_1}^{n_2}\right] \neq \prod_{n=n_1}^{n_2} \mu[\omega(n)]$. As a consequence, probabilities must be defined on the set Ω of infinite rasters. This can be done thanks to transition probabilities (see section 3.2.2). In this setting, the *entropy rate* (or Kolmogorov-Sinai entropy) of μ is:

$$\mathcal{S}[\mu] = -\limsup_{n\to\infty} \frac{1}{n+1} \sum_{\omega_0^n} \mu[\omega_0^n] \log \mu[\omega_0^n], \tag{3}$$

where the sum holds over all possible blocks ω_0^n.

Given a potential $\mathcal{H}$, the MaxEnt states that there is a unique probability measure $\mu \in \mathcal{M}$ such that:

$$\mathcal{P}[\mathcal{H}] = \sup_{\nu \in \mathcal{M}} (\mathcal{S}[\nu] + \nu[\mathcal{H})]) = \mathcal{S}[\mu] + \mu[\mathcal{H}], \tag{4}$$

where $\nu[\mathcal{H}]$ is the average of $\mathcal{H}$ under ν. This is a variational principle which selects, among all possible probability ν, *a unique* probability μ which realizes the supremum. μ is called *the Gibbs distribution with potential* $\mathcal{H}$.

The quantity $\mathcal{P}[\mathcal{H}]$ is called topological pressure or *free energy*. For a potential of the form (2) this is a convex function of the h_ls and [34, 19]:

$$\frac{\partial \mathcal{P}[\mathcal{H}]}{\partial h_l} = \mu[m_l]. \tag{5}$$

Moreover, the Kullback-Leibler divergence $d_{KL}(\nu, \mu)$ between an invariant probability $\nu \in \mathcal{M}$ and the Gibbs distribution μ with potential $\mathcal{H}$ is given by:

$$d_{KL}(\nu, \mu) = \mathcal{P}[\mathcal{H}] - \nu[\mathcal{H}] - \mathcal{S}[\nu]. \tag{6}$$

2.3 Hammersley-Clifford hierarchy

In this section we show that there is a natural hierarchy on spike blocks, that transposes to monomials, that we call this the Hammersley-Clifford hierarchy in reference to the seminal paper (although unpublished) [17]. See also [2, 23, 30]. The Hammersley-Clifford theorem is widely used in image reconstruction [20], in the realm of spatial interactions between pixels. We use it in a different context, dealing with spatio-temporal interactions between spikes.

2.3.1 Spike blocks representation

To each spike block ω_0^D we associate an integer (*index*) $l \in \{0, \ldots, L(N,R)\}$, with $l = \sum_{k=1}^{N} \sum_{n=0}^{D} 2^{n\,N+k-1}\,\omega_k(n)$. We note $\omega^{(l)}$ the block corresponding to l. As an example, the block $\left[\begin{smallmatrix} 1 & 1 \\ 1 & 0 \end{smallmatrix}\right]$ has an index $l = 7$. The Hammersley-Clifford hierarchy is defined as follows. We define the block inclusion $\sqsubseteq$ by $\omega_0^D \sqsubseteq {\omega'}_0^D$ if $\omega_k(n) = 1 \Rightarrow \omega'_k(n) = 1$ (all bits '1' in ω_0^D are bits '1' in ${\omega'}_0^D$), with the convention that the block of degree 0 is included in all blocks. Note that $\omega_0^D \sqsubseteq \omega_0^D$. We note $\omega_0^D \sqsubset {\omega'}_0^D$ if $\omega_0^D \neq {\omega'}_0^D$. For two blocks $\omega^{(l')}, \omega^{(l)}$ we have $\omega^{(l')} \sqsubseteq \omega^{(l)} \Rightarrow l' \leq l$ but the converse is not true in general.

2.3.2 Monomials representation

Since a monomial is defined by a set of spike events $p_u = (k_u, n_u)$ one can associate to this set a spike block or *mask* where the only bits '1' are located at (k_u, n_u), $u = 1, \ldots, r$. To this mask one can thus associate an integer exactly as in the previous section, $l = \sum_{(k_u,n_u)} 2^{n_u\,N+k_u-1}$. Thus, we may label the interaction by this integer index and write m_l instead of $m_{p_1,\ldots,p_r}$. The monomial $m_\emptyset$ is written m_0. The degree of a monomial is the degree of the corresponding mask. Here is the mask corresponding to the monomial $\omega_1(2)\omega_2(2)$ for $N = 3, R = 2$: $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$ (instantaneous pairwise interaction between neurons 1 and 2).

A mask is a spike block, so the relation $\sqsubseteq$ can be used on monomials as well. For two integers l, l', $m'_l(\omega^{(l)}) = 1$ if and only if $\omega^{(l')} \sqsubseteq \omega^{(l)}$: the monomial corresponding to the mask $\omega^{(l')}$ is equal to 1 whenever the spike block $\omega^{(l)}$ has 1s at each 1s position in the mask $\omega^{(l)}$. In particular, $m_l\left(\omega^{(l)}\right) = 1$.

2.3.3 Decomposition of observables and potentials

Time-translation invariant observables of range R are real functions of blocks $\omega_0^D = \omega^{(l)}$, $l = 0 \dots L(N,R)$. They can therefore be represented by row vectors (linear forms) in $\mathbb{R}^{L(N,R)}$, with entries $O_l \stackrel{\text{def}}{=} \mathcal{O}\left(\omega^{(l)}\right)$. Thus, each monomial m_l can be represented by a row vector M_l with entries $M_{l',l} = m_l(\omega^{(l')})$. This defines an upper triangular matrix M with entries :

$$M_{l,l'} = \mathbb{1}_{l \sqsubseteq l'} \stackrel{\text{def}}{=} \begin{cases} 1, & \text{if} \quad \omega^{(l)} \sqsubseteq \omega^{(l')}; \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

M is invertible with inverse [2] $M^{-1}_{l,l'} = (-1)^{d(l')-d(l)} \mathbb{1}_{l \sqsubseteq l'}$. In this way, M defines a coordinate transformation in $\mathbb{R}^{L(N,R)}$ from the canonical basis, to a new basis where basis vectors are monomials. In the canonical basis, an observable $\mathcal{O}$ is written as a row vector with entries $O_l = \mathcal{O}\left(\omega^{(l)}\right)$. In the monomial basis it takes the form $\sum_{l=0}^{L(N,R)} o_l \, m_l$. This decomposition is unique.

We have [3]:

$$O_{l'} = \sum_{l=0}^{L(N,R)} o_l M_{l,l'} = \sum_{l \sqsubseteq l'} o_l. \tag{8}$$

and the inverse formula

$$o_{l'} = \sum_{l=0}^{L(N,R)} O_l M^{-1}_{l,l'} = \sum_{l \sqsubseteq l'} (-1)^{d(l')-d(l)} O_l. \tag{9}$$

This decomposition holds for any observable of range R. From now on, for such a function, say g, we shall use capital letters for the representation in the canonical basis (i.e. $g(\omega^{(l)}) = G_l$) and small letters g_l for the decomposition in the monomial basis. In particular, a potential $\mathcal{H}$ decomposes as (2), as announced.

We call the decomposition (8) the Hammersley-Clifford decomposition [17] and (9) the Mousouris inversion formula [23].

---

[2]
$$\sum_{j=1}^{L(N,R)} M_{ij} M^{-1}_{jk} = \sum_{j=1}^{L(N,R)} \mathbb{1}_{i \sqsubseteq j} (-1)^{d(k)-d(j)} \mathbb{1}_{j \sqsubseteq k} = \sum_{m=0}^{d(k)} (-1)^m \sum_{j,\, d(k)-d(j)=m} \mathbb{1}_{i \sqsubseteq j} \mathbb{1}_{j \sqsubseteq k}$$
$$= \sum_{m=0}^{d(k)} (-1)^m \# \{ j;\, i \sqsubseteq j \sqsubseteq k;\; d(k) - d(j) = m \}.$$
If $i = k$, $i = j = k$ and $\sum_{j=1}^{L(N,R)} M_{ij} M^{-1}_{jk} = 1$.
If $i \neq k$, $\# \{ j;\, i \sqsubseteq j \sqsubseteq k;\; d(k) - d(j) = m \} = C^{d(k)-m}_{d(k)}$ so that $\sum_{m=0}^{d(k)} (-1)^m C^{d(k)-m}_{d(k)} = (1-1)^{d(k)} = 0$.

[3]
$$O_{l'} = \mathcal{O}\left(\omega^{(l')}\right) = \sum_{l=0}^{L(N,R)} o_l \, m_l(\omega^{(l')}) = \sum_{l=0}^{L(N,R)} o_l \, M_{l,l'}$$

3 Method

3.1 Determining the shape of the potential from a raster

We address now the following problem. Assume that we are given a raster ω generated by an unknown Gibbs probability[4] $\mu^{(ex)} \in \mathcal{M}$ with a hidden potential $\mathcal{H}^{(ex)} = \sum_{l=0}^{L} h_l^{(ex)} m_l$. How to determine $\mathcal{H}^{(ex)}$ from the observation of ω ? This question has two parts:

(i) **Determining the shape of $\mathcal{H}^{(ex)}$**. This means, determining the set of monomials having a non zero coefficient $h_l^{(ex)}$. In the context of statistical physics / thermodynamics this amounts to selecting an *ensemble* where specific forms of energy depending on the problem are considered and the characterization of the potential / Gibbs distribution is made from first principles in mechanics and thermodynamics. When dealing with neuronal networks, such principles are not (yet ?) available, and one has either to guess the potential's shape by testing different types of interactions mostly based on statistical physics analogy [36, 14, 15, 41, 42]. Or, one has to consider the most general potential form (2) including all types of interactions, and find a strategy to eliminate as many terms as possible. This is the point of view adopted in this paper.

(ii) The shape of the potential being given, find the value of the non vanishing $h_l^{(ex)}$s. This issue has been widely discussed in [44, 8, 24] where the strategy was to (numerically) minimize the KL divergence $d_{KL}\left(\mu^{(ex)}, \mu\right)$. Here we adopt a distinct point of view providing exact results.

We answer these questions using the Hammersley-Clifford decomposition of the potentials and the equivalence of potentials (cohomology).

3.2 Equivalent potentials

3.2.1 Definition

We say that two potentials are *equivalent* if they correspond to the same Gibbs distribution. These potentials can be characterized in terms of *cohomology* [29, 10, 35] defined as follows.

Two potentials $\mathcal{H}^{(1)}, \mathcal{H}^{(2)}$, of range[5] $R = D + 1$ are *cohomologous* if there exists a function $f : \{0,1\}^{ND} \to \mathbb{R}$, of range D, such that:

$$\mathcal{H}^{(2)}\left(\omega_0^D\right) = \mathcal{H}^{(1)}\left(\omega_0^D\right) - f\left(\omega_0^{D-1}\right) + f\left(\omega_1^D\right) + \Delta, \tag{10}$$

where $\Delta = \mathcal{P}\left[\mathcal{H}^{(2)}\right] - \mathcal{P}\left[\mathcal{H}^{(1)}\right]$ is the difference between the free energies of $\mathcal{H}^{(2)}$ and $\mathcal{H}^{(1)}$. From now on, we set:

$$\mathcal{G}(\omega_0^D) = f\left(\omega_0^{D-1}\right) - f\left(\omega_1^D\right) - \Delta, \tag{11}$$

Two potentials are equivalent (correspond to the same Gibbs distribution) *if and only if* they are cohomologous.

It is convenient to use the vector representation for $H^{(1)}, H^{(2)}, F$ so that (10) reads:

$$H^{(2)}_{\omega_0^D} = H^{(1)}_{\omega_0^D} + F_{\omega_1^D} - F_{\omega_0^{D-1}} + \Delta = H^{(1)}_{\omega_0^D} - G_{\omega_0^D}. \tag{12}$$

[4] Note that, dealing with finite rasters, $\mu^{(ex)}$ can only be approached by the so-called empirical measure on the raster. This issue is postponed to section 3.4.7.

[5] The definition of cohomology is given in the more general context of infinite range potentials. Here we stick at range-R potentials.

for each entry ω_0^D.

Since the choice of f is arbitrary there are a priori infinitely many potentials equivalent to a given one.

3.2.2 Normalized potential

A prominent example of cohomology associates a potential $\mathcal{H}$ to a unique *normalized* potential, namely the log of a transition probability:

$$\phi\left(\omega_0^D\right) = \log P\left[\omega(D) \,\middle|\, \omega_0^{D-1}\right]. \tag{13}$$

So, in this case $\mathcal{G}$ acts as a normalization function. For potentials of range $R = 1$ ($D = 0$), we have $\phi = \mathcal{H} - \log Z$ where Z is the partition function. Therefore, $\mathcal{G} = \log Z$, whereas ϕ is a function of the spike pattern $\omega(0)$ only: $\phi\left(\omega(0)\right) = \log P\left[\omega(0)\right]$. Here, therefore, $P\left[\omega(0)\right] = \frac{e^{\mathcal{H}(\omega(0))}}{Z}$.

For potential with range larger that 1 the normalization does not reduce to subtracting a constant but also involves a function $\mathcal{G}$ which can be computed in terms of largest eigenvalue and related right eigenvector of a transfer matrix [44, 24, 8].

Transition probabilities defined via (13) are always positive provided $\mathcal{H} > -\infty$. They define therefore a Markov chain with memory depth D having a *unique* invariant measure: the Gibbs distribution μ. Note that the free energy density of a normalized potential is always 0.

3.2.3 There are infinitely many equivalent potentials

The hidden measure $\mu^{(ex)}$ which has generated the observed raster is the invariant probability of a Markov chain determined by transition probabilities $P\left[\omega(D) \,\middle|\, \omega_0^{D-1}\right]$ hence, by a normalized potential (13). These probabilities can be determined either empirically (section 4.1) or analytically in some examples of neural network models (see section 4.2). They can also be fitted by canonical models such as LN or GLM [1, 28]. Knowing ϕ, there exist however *infinitely many equivalent potentials* of the form (2) depending on the choice of the cohomology function f. The situation is summarized in the following diagram. Given a family of cohomologous potentials $\{\mathcal{H}^{(1)}, \mathcal{H}^{(2)}, \ldots, \mathcal{H}^{(n)}, \ldots\}$ there is one only normalized potential ϕ corresponding to all of them. But, on the opposite, given a normalized potential ϕ, there are infinitely many potentials $\mathcal{H}$ equivalent to it, via the cohomology.

$$\begin{array}{ccc}
\mathcal{H}^{(1)} & \Rightarrow & \phi = \mathcal{H}^{(1)} - \mathcal{G}^{(1)} \\
 & & \Updownarrow f \\
\mathcal{H}^{(2)} & \Rightarrow & \phi = \mathcal{H}^{(2)} - \mathcal{G}^{(2)} \\
\vdots & & \\
\mathcal{H}^{(n)} & \Rightarrow & \phi = \mathcal{H}^{(n)} - \mathcal{G}^{(n)}
\end{array}$$

So the issue is: We can determine ϕ (or approximate it from data), but then we have infinitely many candidates for $\mathcal{H}^{(ex)}$.

Additionally, the normalization function $\mathcal{G}$ obeys the Hammersley-Clifford decomposition theorem: $\mathcal{G} = \sum_{l=0}^{L(N,R)} g_l\, m_l$, as well as the normalized potential $\phi = \sum_{l=0}^{L(N,R)} \phi_l\, m_l$. So, the cohomology relation (10) reads:

$$\phi_l = h_l - g_l, \quad l = 0, \ldots, L(N,R). \tag{14}$$

This has a dramatic consequence. Even if the hidden potential $\mathcal{H}^{(ex)}$ as a small number of terms, the normalized potential ϕ *has extra terms brought by the normalization function* $\mathcal{G}$. This means that working on ϕ we have in general many extra Hammersley-Clifford terms which are somewhat irrelevant. How to eliminate them ?

3.3 Equivalent interactions

Let us consider a potential of the form $\mathcal{H}^{(1)} = \mathcal{H}^{(0)} + h_{l_1} m_{l_1} + h_{l_2} m_{l_1} \circ \mathcal{T}$ where the monomial m_{l_1}'s mask has no 1's in the last column D. Hence, $m_{l_1}\left(\omega_0^D\right) \equiv m_{l_1}\left(\omega_0^{D-1}\right)$ and $m_{l_1} \circ \mathcal{T}\left(\omega_0^D\right) \equiv m_{l_1}\left(\omega_1^D\right)$. This potential is equivalent [6] to $\mathcal{H}^{(2)} = \mathcal{H}^{(0)} + [\, h_{l_1} + h_{l_2} \,] m_{l_1}$ via the cohomology function $f = -h_{l_2} m_1$ (with $\Delta = 0$, so that these potentials have the same free energy density). In this example, the cohomology function f allows us to eliminate the term $m_{l_1} \circ \mathcal{T}$ in $\mathcal{H}^{(1)}$ leading to a simplified potential $\mathcal{H}^{(2)}$. This result generalizes. If a potential contains terms of the form $m_{l_1} \circ \mathcal{T}^k$, $k = 0, \ldots, D$, where $m_{l_1}\left(\omega_0^D\right) = m_{l_1}\left(\omega_0^{D-k}\right)$, k of these terms can be removed by cohomology, only keeping one, still having an equivalent potential.

This has the following interpretation. The average of $\mathcal{H}$ appearing in (4) is given by $\mu[\,\mathcal{H}\,] = \sum_{l=0}^{L(N,R)} h_l \mu[\,m_l\,]$. Thus, the average value of each m_l constitutes a priori a constraint in the variational problem (4). However, the monomials m_{l_1} and $m_{l_1} \circ \mathcal{T}$ have the same average, whatever the Gibbs distribution, thanks to the time translation invariance of Gibbs distributions. Their averages constitute therefore redundant constraints that cannot be determined independently, and which make the variational problem (4) under-determined: there are less independent constraints than parameters. An easy example is given by the monomials $\omega_k(0)$ and $\omega_k(1)$ whose average is the firing rate of neuron k. This quantity is independent of time from the time-translation invariance hypothesis, thus $\mu[\,\omega_k(0)\,] = \mu[\,\omega_k(1)\,]$ whatever $\mu \in \mathcal{M}$.

As a consequence, we say that two monomials m_{l_1}, m_{l_2} correspond to *equivalent interactions* if $m_{l_2} = m_{l_1} \circ \mathcal{T}^k$, for some $0 < k \leq D$ and if the mask of one of them have the last column full of zeros. If we have n equivalent interactions, one can remove $n-1$ of them in the potential. We call the set of non equivalent interactions *canonical interactions*. We call *canonical form* the remaining potential. By construction, it contains interactions with a mask having at least one '1' in the last column (time D).

3.3.1 Canonical interactions cannot be eliminated by cohomology

Assume now that we are given two potentials $\mathcal{H}^{(1)}, \mathcal{H}^{(2)}$ in the canonical form, where $\mathcal{H}^{(1)}$ has a zero coefficient for the canonical interaction m_l whereas $\mathcal{H}^{(2)} = \mathcal{H}^{(1)} + h_l\, m_l$, $h_l \neq 0$. Let us show that these two potentials are not equivalent. From (12), they are equivalent if one can find a $L(N,D)$-dimensional vector F such that, $\forall \omega_0^D$:

$$F_{\omega_1^D} - F_{\omega_0^{D-1}} + \Delta + h_l \mathbb{1}_{\omega^{(l)} \sqsubseteq \omega_0^D} = 0.$$

The block only composed by '1's contains all other blocks, and it is translation invariant so that the terms involving F cancel in the equation above. We have therefore $\Delta + h_l = 0$. The block only composed by '0's is also translation invariant and, if $l > 0$ we obtain $\Delta = 0$, so that $h_l = 0$, in contradiction with the hypothesis.

[6]
$$\begin{aligned}
\mathcal{H}^{(2)}\left(\omega_0^D\right) &= \mathcal{H}^{(0)}\left(\omega_0^D\right) + [\, h_{l_1} + h_{l_2} \,] m_{l_1}\left(\omega_0^D\right) \\
&= \mathcal{H}^{(1)}\left(\omega_0^D\right) - h_{l_1} m_{l_1}\left(\omega_0^D\right) - h_{l_2} m_{l_1} \circ \mathcal{T}\left(\omega_0^D\right) + [\, h_{l_1} + h_{l_2} \,] m_{l_1}\left(\omega_0^D\right) \\
&= \mathcal{H}^{(1)}\left(\omega_0^D\right) + h_{l_2} m_{l_1}\left(\omega_0^{D-1}\right) - h_{l_2} m_{l_1}\left(\omega_1^D\right) = \mathcal{H}^{(1)}\left(\omega_0^D\right) + f \circ \mathcal{T}\left(\omega_1^D\right) - f\left(\omega_0^{D-1}\right).
\end{aligned}$$

We arrive therefore at the following important conclusions:

(i) Two canonical potentials of the form (2), where the sum holds on canonical interactions, are equivalent if and only if they have the same coefficients (except the constant h_0). We set $h_0 = 0$ from now on.

(ii) A generic canonical potential has a number of coefficients growing like 2^{NR}. The number of equivalent interactions can be computed as the number of blocks with last column full of '0's, therefore considering N neurons and range R there are (only) $2^{N(R-1)}$ interactions that can be eliminated. A generic potential contains therefore a priori $2^{NR} - 2^{N(R-1)}$ canonical terms.

3.4 Computing coefficients

We now describe a procedure allowing to compute the h_ls when the normalized potential is known. It is based on periodic orbits sampling of the phase space.

3.4.1 Periodic orbits invariants

A raster ω can be viewed as a sequence of range-R blocks $\omega^{(l_1)} \equiv \omega_0^D$, $\omega^{(l_2)} \equiv \omega_1^{D+1}$, an so on where $\omega^{(l_k)}$ is mapped to $\omega^{(l_{k+1})}$ by the time shift $\mathcal{T}$. Since the set of range-R blocks is finite an infinite raster contains typically infinitely many repetitions of each block. More generally, periodic repetitions of blocks sequences occur in a recurrent way. The recurrence of such periodic patterns are especially useful to characterize the probability $\mu^{(ex)}$ and the potential $\mathcal{H}^{(ex)}$.

A raster ω is a *periodic orbit* of period τ if $\omega^{(l_{k\tau+n})} = \omega^{(l_n)}$, $k \geq 0$, $0 \leq n \leq \tau$. Now, it is clear from (10) that if $\mathcal{H}^{(1)}$ and $\mathcal{H}^{(2)}$ are equivalent then:

$$\sum_{n=1}^{\tau} \mathcal{H}^{(2)}\left(\omega^{(l_n)}\right) = \sum_{n=1}^{\tau} \mathcal{H}^{(1)}\left(\omega^{(l_n)}\right) + \tau\Delta, \tag{15}$$

because the sum of terms involving f along the periodic orbit cancel each other. This reflects an invariance property of the value that equivalent potentials take on periodic orbits. Reciprocally, it can be shown that $\mathcal{H}^{(2)}$ and $\mathcal{H}^{(1)}$ are equivalent if and only if (15) holds for *all* possible periodic orbits in Ω (thus infinitely many with a period ranging from 1 to $+\infty$) [33].

Although powerful, this last result is therefore of little practical use. However, the consideration of specific periodic orbits, combined with Hammersley-Clifford decomposition leads to particularly useful results.

3.4.2 Hammersley-Clifford decomposition on specific periodic orbits

We define the modulo-R periodic shift σ which act as follows:

$$\omega_0^D = \omega(0)\,\omega(1)\,\ldots\omega(D) \rightarrow \sigma\omega_0^D = \omega(1)\,\omega(2)\,\ldots\omega(D)\,\omega(0)$$

With a slight abuse of notation we note σl the index of the block $\sigma\omega^{(l)}$. By iterating σ on all possible blocks in $\{0,1\}^{N\,R}$ one generates all periodic orbits having a period $\tau \leq R$ where τ divides R. Clearly, this set is quite small compared to the (infinite) set of all possible periodic orbits in Ω. Nevertheless, it gives useful and tractable information on $\mathcal{H}^{(ex)}$. We call $\mathcal{C}$ the set of these specific periodic orbits. It only depends on N, R but not on the potentials. An element c of $\mathcal{C}$ can be represented by the index $l \equiv l(c)$ of the first block. We call $\mathcal{C}^*$ the set $\mathcal{C}$ minus the pattern corresponding to $l = 0$ (block with no spike).

For two equivalent potentials $\mathcal{H}^{(1)}, \mathcal{H}^{(2)}$ as in (10) we have, for each periodic orbit $c \in \mathcal{C}$:

$$\sum_{n=1}^{R} \mathcal{H}^{(2)}\left(\omega^{(\sigma^n l)}\right) = \sum_{n=1}^{R} \mathcal{H}^{(1)}\left(\omega^{(\sigma^n l)}\right) + R\Delta, \tag{16}$$

This equality is necessary but not a sufficient condition for $\mathcal{H}^{(2)}$ and $\mathcal{H}^{(1)}$ to be equivalent.

We can now use the Hammersley-Clifford decomposition (8) to obtain:

$$\sum_{n=1}^{R} \sum_{l'_n \sqsubseteq \sigma^n l} h^{(2)}_{l'_n} = \sum_{n=1}^{R} \sum_{l'_n \sqsubseteq \sigma^n l} h^{(1)}_{l'_n} + R\Delta. \tag{17}$$

We have $\sum_{n=1}^{R} \sum_{l'_n \sqsubseteq \sigma^n l} h_{l'_n} = \sum_{l' \sqsubseteq l} \sum_{n=1}^{R} h_{\sigma^n l'}$. Indeed, in the left-hand side one sums over the blocks $\sigma^n l$, $n = 1, \ldots, R$ in the periodic orbit, then sums over all sub-blocks l'_n of $\sigma^n l$. This corresponds to a list of sub-blocks which can be rearranged in a list of R-periodic sequences of sub-blocks of l: each periodic list correspond to a sub-block of l_1 and its R shifts under σ. Note that this commutation of sums property only holds for elements of $\mathcal{C}$.

We have finally:

$$\sum_{l' \sqsubseteq l} \sum_{n=1}^{R} h^{(2)}_{\sigma^n l'} = \sum_{l' \sqsubseteq l} \sum_{n=1}^{R} h^{(1)}_{\sigma^n l'} + R\Delta \tag{18}$$

for all $l \equiv l(c), c \in \mathcal{C}$.

3.4.3 Invariants

This defines a hierarchy of relations between blocks / interactions with increasing degree d. For the block of degree $d = 0$ we obtain $h^{(2)}_0 - h^{(1)}_0 = \Delta$. The difference in the constant terms $h^{(2)}_0 - h^{(1)}_0$ is the difference of free energy densities.

For blocks of degree $d = 1$ we have $\sum_{n=1}^{R} h^{(2)}_{\sigma^n l} + R h^{(2)}_0 = \sum_{n=1}^{R} h^{(1)}_{\sigma^n l} + + R h^{(1)}_0 + R\Delta$, so that $\sum_{n=1}^{R} h^{(2)}_{\sigma^n l} = \sum_{n=1}^{R} h^{(1)}_{\sigma^n l}$. By recursion this relation holds for all blocks of degree $d > 0$. We write this result in a more compact form introducing the notation

$$S_h(c) = \sum_{n=1}^{R} h_{\sigma^n l}, \quad l \equiv l(c), \tag{19}$$

Thus, for all periodic orbit $c \in \mathcal{C}^*$,

$$S_{h^{(2)}}(c) = S_{h^{(1)}}(c). \tag{20}$$

This relation reveals thus an interesting symmetry property between interactions. The blocks appearing in an element $c \in \mathcal{C}$ corresponds to a set of interactions mapped on each other by the periodic time shifts. An example for $R = 2$ is e.g. $\omega_{i_1}(1)\omega_{i_2}(2)$ and $\omega_{i_1}(2)\omega_{i_2}(1)$. Equation (20) establishes therefore that the sum of these interactions coefficients is an invariant for equivalent potentials.

3.4.4 An ansatz to eliminate some h_ls

These relations hold in particular between the normalized potential ϕ and any potential $\mathcal{H}$ equivalent to ϕ. We have:

$$\mathcal{P}[\mathcal{H}] = h_0 - \phi_0, \tag{21}$$

whereas from (20):

$$S_h(c) = S_\phi(c), \quad c \in \mathcal{C}^*. \tag{22}$$

These relations can be used to reduce the number of coefficients in the sought potential $\mathcal{H}^{(ex)}$. Indeed, for a periodic orbit $c \in \mathcal{C}$ denote K_c the set of canonical interactions (section 3.3). Thus, c contains $R - K_c$ equivalent interactions whose coefficient can be set to zero in $S_{h^*}(c)$. Thus, if $K_c = 1$ there remains only one coefficient, $h_l^{(ex)} = S_\phi(c)$, $l \equiv l(c)$. More generally, the relation (22) constrains the sum of canonical coefficients.

If $\mathcal{H}^{(ex)}$ has vanishing h_ls, then some sums $S_{h^{(ex)}}(c)$ may vanish. From the invariance property (16) this implies $S_\phi(c) = 0$, a property which can be observed from data. As a corollary, when observing that $S_\phi(c) = 0$ we may conjecture that *the hidden potential $\mathcal{H}^{(ex)}$ is such that all canonical coefficients in the orbit c are vanishing*. This non rigorous argument is based on the following reasoning.

If $S_\phi(c) = 0$, any potential $\mathcal{H}$ equivalent to ϕ, satisfies the condition $S_h(c) = 0$. This however does not mean that all canonical h_{l_k}'s in the periodic orbit defining $S_h(c)$ vanish. The fact that a sum is equal to zero does not imply that all coefficients in the sum are zero ! However, when seeking a hidden potential $\mathcal{H}^{(ex)}$, a reasonable assumption is that if $S_h(c) = 0$ then *generically*[7] all h_l's in the orbit vanish. Certainly, one can construct potentials violating this assumption, by tuning sharply the coefficients. For example, the following potential $\mathcal{H}^{(ex)} = \omega_1(0)\omega_2(1) - \omega_1(1)\omega_2(0)$, with $N = 2, R = 2$ has a vanishing sum $S_h(c)$ for: $c = \left\{ \left[\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix} \right]; \left[\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix} \right] \right\}$. But such potentials are neither generic (one cannot obtain them e.g. by drawing the h_l's at random) nor robust (if the sum $S_h(c)$ vanishes accidentally, a small variation of one coefficient makes the sum non zero). On the opposite, the condition $h_{l_k} = 0$ for all l_k in the orbit "physically" corresponds to *an absence of interaction* between spikes. An absence of interactions does not arise accidentally but reflects deep causal effects in the dynamics generating spikes.

Therefore, setting $h_{l_k} = 0$ for all l_k in an orbit such that $S_\phi(c) = 0$ provides an efficient method to reducing the number of terms in the guess potential, with a phenomenological interpretation.

3.4.5 A general algorithm to compute the h_ls

We now give a general method to compute all h_ls. This method is not based on the assumption made is the previous section, so it can be used to compute all the h_l's, even those belonging to the orbits in which $S_\phi(c) = 0$. However, it implies heavy computations if there are many terms in the potential. Presumably, if $S_\phi(c) = 0$ for some c, using the ansatz of the previous section will speed up the computation.

Applied to ϕ, $\mathcal{H}^{(ex)}$, eq. (15) gives, for a periodic orbit of period τ:

$$\sum_{n=1}^{\tau} \phi\left(\omega^{(l_n)}\right) = \sum_{n=1}^{\tau} \mathcal{H}^{(ex)}\left(\omega^{(l_n)}\right) - \tau \mathcal{P}\left[\mathcal{H}^{(ex)}\right]. \tag{23}$$

It does not contain the cohomology function and provides new equations that can be used to fully determine the interactions coefficients.

The idea is to proceed iteratively. One computes first the coefficients of degree 1 interactions, then degree 2 and so on. Assume that we want to compute the coefficient of a canonical interaction with mask $\omega^{(l)}$ of degree d. We claim that it is possible to construct a periodic orbit with

[7] A potential with L parameters can be viewed as a point in a (compact) subset $\mathcal{S} \in \mathbb{R}^L$. A subset $s \in \mathcal{S}$ is generic in a metric sense if it has a positive Lebesgue probability. It is generic in a topological sense if it is dense in $\mathcal{S}$.

period $2d$ such that (23) holds, with only one unknown, the coefficient $h_l^{(ex)}$. Such an orbit can be constructed as follows (there are other possibilities).

- Step 1. Shift periodically $\omega^{(l)}$ to the left until the left-most spiking pattern has at least one 1. All the generated masks correspond to the same canonical interactions so their coefficient is zero and do not contribute to (23).

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

- Step 2. Continue to periodically left shift but, before shifting, remove the 1 with the lower neuron index, on the left most spike pattern. Tag the 1's that has been removed. Do this until the total number of left shifts including step 1 and 2 is R. The masks obtained this way have a degree $< d$ so their coefficient is known.

$$\begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

- Step 3. Same as step 1. All these masks correspond to the same canonical constraint so they do not contribute to (23).

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix}$$

- Step 4. This is the inverse step as step 2. One restore the $'1'$ that has been removed on the left most spike pattern and left shift. In this way one finally regenerate $\omega^{(l)}$. All the mask generated this way (expect $\omega^{(l)}$) have a degree $< d$ so their coefficient is known.

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

As claimed we have generated a periodic orbit of period $2R$ where (23) has only one unknown, $h_l^{(ex)}$. Proceeding iteratively one can compute all remaining $h_l^{(ex)}$'s.

3.4.6 Where to stop ?

When getting to larger and larger interactions degree the computation might become quite expensive. However, monomials constrain the entropy $\mathcal{S}[\mu]$ and free energy density $\mathcal{P}[\mathcal{H}]$ as follows. For the normalized potential ϕ (eq. (13)) we have $\mathcal{P}[\phi] = 0$, so that (4) gives:

$$\mathcal{S}[\mu] = -\sum_{l=0}^{L} \phi_l \mu[m_l]. \tag{24}$$

Each monomial m_l contributes to the entropy density with a weight $\phi_l \mu[m_l]$.

Therefore, from (4), any potential $\mathcal{H}$ equivalent to ϕ obeys:

$$\mathcal{P}[\mathcal{H}] = \sum_{l=0}^{L} g_l \mu[m_l], \tag{25}$$

with $g_l = h_l - \phi_l$. Each monomial m_l contributes to the free energy density with a weight $g_l \mu[m_l]$. Note that, from (21), $\mathcal{P}[\mathcal{H}] = g_0$, therefore $\sum_{l=1}^{L} g_l \mu[m_l] = 0$.

Now, we have:

$$\mu[m_l] = \sum_{l \sqsubseteq l'} \mu\left[\omega^{(l')}\right]. \tag{26}$$

Since [8] $0 < \mu\left[\, m_l \,\right] < 1$:

$$\omega^{(l_1)} \sqsubseteq \omega^{(l_2)} \Rightarrow \mu\left[\, m_{l_1} \,\right] \geq \mu\left[\, m_{l_2} \,\right] \tag{27}$$

The average value of m_l's are decreasing when the degree of the interaction increases.

Thus, the contribution of high degree interactions to entropy and free energy becomes rapidly negligible. Thus, one can iterate the computation described in the previous section with increasing degree until the entropy contribution of the last step is below a certain threshold. A possible criterion of convergence is:

$$\left| \frac{\Delta \mathcal{S}^{(d+1)}}{\mathcal{S}^{(d)}} \right| < \epsilon, \tag{28}$$

for some $\epsilon > 0$.

Using this criterion, one replaces the exact potential $\mathcal{H}^{(ex)}$ by a truncated potential $\mathcal{H}^{(co)} = \sum_{l=0}^{l_{co}} h_l\, m_l$, where "co" stands for "cut-off". How much do we loose with this approximation? From (6) we have[9]:

$$d_{KL}(\mu^{(ex)}, \mu^{(co)}) = \sum_{l=l_{co}+1}^{L(N,R)} h_l^{(ex)}\, \mu^{(ex)}\left[\, m_l \,\right]. \tag{29}$$

If we assume that the h_ls stay bounded as l grows, this quantity is bounded by the sum of monomial averages $\sum_{l=l_{co}+1}^{L(N,R)} \mu^{(ex)}\left[\, m_l \,\right]$ which tends to 0 as l_{co} grows.

3.4.7 Finite size sampling

We now focus on the realistic case where the raster has a finite duration T. The previous results assume that the conditional probabilities $P\left[\, \omega(D) \,\middle|\, \omega_0^{D-1} \,\right]$ are known and are positive. However, when dealing with a finite raster of length T one only obtains an estimate:

$$P^{(T)}\left[\, \omega(D) \,\middle|\, \omega_0^{D-1} \,\right] = \frac{\pi^{(T)}\left[\, \omega_0^D \,\right]}{\pi^{(T)}\left[\, \omega_0^{D-1} \,\right]}, \tag{30}$$

of these probabilities, where $\pi^{(T)}\left[\, \omega_0^D \,\right]$ is the number of occurrence of ω_0^D in the raster, divided by the total number of blocks of range R, $T - R + 1$, observed in the raster. Thus, $P^{(T)}\left[\, \omega(D) \,\middle|\, \omega_0^{D-1} \,\right]$ is a random variable, with (typically Gaussian) fluctuations tending to $P\left[\, \omega(D) \,\middle|\, \omega_0^{D-1} \,\right]$ as $T \to +\infty$.

In a raster of length T one observes at most $T - R + 1$ distinct blocks with $T << L(N,R)$. As a consequence, for many blocks, the probabilities $P^{(T)}\left[\, \omega(D) \,\middle|\, \omega_0^{D-1} \,\right]$ cannot be determined, even approximately. Indeed, to compute (30) one needs to define some $\epsilon > \frac{1}{T-R+1} > 0$ such that

[8] This is a consequence of the Perron-Frobenius theorem and of the assumption that $\mathcal{H} > -\infty$.

[9]
$$d_{KL}(\mu^{(ex)}, \mu^{(co)}) = \mathcal{P}\left[\, \mathcal{H}^{(co)} \,\right] - \mu^{(ex)}\left[\, \mathcal{H}^{(co)} \,\right] - \mathcal{S}[\mu^{(ex)}],$$

where from (4),

$$\mathcal{P}\left[\, (\mathcal{H}^{(ex)} \,\right] = \mu^{(ex)}\left[\, \mathcal{H}^{(ex)} \,\right] + \mathcal{S}[\mu^{(ex)}],$$

so that:

$$\begin{aligned} d_{KL}(\mu^{(ex)}, \mu^{(co)}) &= \mathcal{P}\left[\, \mathcal{H}^{(co)} \,\right] - \mu^{(ex)}\left[\, \mathcal{H}^{(co)} \,\right] - \mathcal{P}\left[\, \mathcal{H}^{(ex)} \,\right] - \mu^{(ex)}\left[\, \mathcal{H}^{(ex)}) \,\right] \\ &= \sum_{l=l_{co}+1}^{L(N,R)} h_l^{(ex)} \mu^{(ex)}\left[\, m_l \,\right] + \mathcal{P}\left[\, \mathcal{H}^{(co)} \,\right] - \mathcal{P}\left[\, \mathcal{H}^{(ex)} \,\right]. \end{aligned}$$

We have $\mathcal{P}\left[\, \mathcal{H}^{(co)} \,\right] - \mathcal{P}\left[\, \mathcal{H}^{(ex)} \,\right] = h_0^{(co)} - h_0^{(ex)}$. Since we are considering two potentials having the same Hammersley-Clifford expansion up to some order, $h_0^{(co)} = h_0^{(ex)}$.

$\pi^{(T)}\left[\omega_0^D\right] > \epsilon$ *and* $\pi^{(T)}\left[\omega_0^{D-1}\right] > \epsilon$. Since $\pi^{(T)}\left[\omega_0^{D-1}\right] = \sum_{\omega(D)} \pi^{(T)}\left[\omega_0^D\right] \geq \pi^{(T)}\left[\omega_0^D\right]$ it is sufficient to have $\pi^{(T)}\left[\omega_0^D\right] > \epsilon$, i.e., the number of occurrences of ω_0^D in the experimental raster is large enough.

This condition can be violated in two ways. First, $\pi^{(T)}\left[\omega_0^{D-1}\right] < \epsilon$ so that $\pi^{(T)}\left[\omega_0^D\right] < \epsilon$. Then, it is not possible to estimate reliably $P^{(T)}\left[\omega^{(l)}(D) \,\middle|\, \omega_0^{D-1}\right]$. Second, $\pi^{(T)}\left[\omega_0^D\right] < \epsilon$ and $\pi^{(T)}\left[\omega_0^{D-1}\right] > \epsilon$. In this case the empirical estimation of $P^{(T)}\left[\omega^{(l)}(D) \,\middle|\, \omega_0^{D-1}\right]$ is interpreted as being 0. A vanishing transition probability corresponds to a forbidden transition so that $\Phi(\omega_0^D) = -\infty$. This has the physical interpretation of a hard core potential. The existence of forbidden transitions can result either in the non existence of an invariant Gibbs measure, or its non-uniqueness (first order phase transitions [16]). Were the transition probabilities to be exactly known, would the vanishing of transition probabilities have the serious impact on spike train statistics estimation. Note that this problem is not intrinsic to our approach, but to any approach attempting to fit a Markov chain with too small samples. In this situation, there is no practical way to decide whether a transition probability is indeed vanishing or if its has a too small value to be accessed from the statistical sample. In this paper we make the simplifying assumption that this case corresponds also to blocks whose transition probabilities are undetermined.

We divide therefore the set of blocks $\Omega_{N,R} = \{0,1\}^{NR}$ into two subsets. We call Γ the set of blocks such that $\pi^{(T)}\left[\omega_0^D\right] > \epsilon$. All blocks in Γ have a well defined (estimated) transition probability. The set $\Omega_{N,R}\backslash\Gamma$ contains blocks either having an undetermined transition probability or corresponding to (empirically) forbidden transitions. We consider here that this set only contains blocks with undetermined transition probabilities.

When sampling the periodic orbit structure and the associated conditional probabilities we have, up to now, assumed that all blocks in each periodic orbit have a well defined conditional probability. This is not the case if the raster is finite. Here, the most practical issue is to consider that an undetermined conditional probability on the orbit c is set to 1, so that its logarithm does not contribute to the weight $S_c(\phi)$. If all masks in c didn't appear in the raster $S_c(\phi) = 0$ all h_l's in c are set to zero.

The average value of a monomial decreases with its degree, so that the error made in estimating the empirical conditional probabilities increases with degree. This error propagates to the estimation of the corresponding Hammersley-Clifford coefficient. Conversely, the most reliable terms are those with lowest degree. Fortunately, these terms contribute the most to the entropy. These aspects are numerically illustrated in the next section.

4 Two examples

In this section we illustrate our method in two cases:

(i) **Finite-size effects.** The Gibbs distribution $\mu^{(ex)}$ is known and used to generate a finite raster of length T. Then, we use our method to recover the potential $\mathcal{H}^{(ex)}$ from the raster. This example illustrates the finite-size effects discussed in section 3.4.7.

(ii) **Exact case.** The transition probabilities are exactly known, hence the normalized potential, and we use the method to construct the canonical potential. The chosen example is a discrete-time Leaky Integrate and Fire model where the transition probabilities are known. The goal here is to compare the so-called functional interactions introduced in the realm of MaxEnt, to the real interactions (the synaptic weights).

4.1 Finite-size sampling of a known Gibbs distribution

Given a Gibbs distribution $\mu^{(ex)}$ with spatio-temporal parametric potential :

$$\mathcal{H}^{(ex)} = \sum_{l=0}^{L} h_l^{(ex)} m_l, \tag{31}$$

given a raster ω with length T generated by $\mu^{(ex)}$, the goal is to estimate the parametric form of $\mathcal{H}^{(ex)}$ from that raster.

4.1.1 Periodic orbits and entropy

Since the estimation is based on periodic orbits expansion, it is important to have some characterization of the set of periodic orbits which allows to reconstruct a given potential. Indeed, some orbits will have a zero weight $S_\Phi(l)$, leading to cancel the corresponding h_ls; some others have a small weight that can thresholded at zero, provided a threshold is suitably defined; finally some have a large contribution. The number of periodic orbits of period τ in the support of a measure μ grows, roughly, like $e^{\tau \mathcal{S}[\mu]}$ [27]. Thus, the higher $\mathcal{S}[\mu]$, the larger the number of periodic orbits in a raster of length T, the closer are we from equiprobability and the smaller the probability to observe specific blocks. From this handwaving argument, one expects that the algorithm will perform more or less badly, depending on the entropy.

To investigate this aspect, we considered two different classes of potentials.

1. Monomials and coefficients are drawn at random. There are NR monomials of range $k = 1, \ldots, R$. Coefficients are drawn with a Gaussian distribution with mean 0 and variance $\frac{1}{NR}$ to ensure a correct scaling of the coefficients dispersion as NR increases. This produces typically dense rasters (fig. 1). The entropy of such potentials is extensive and growths rapidly with N. On average, it is close to the maximum entropy that a system with N neurons can have, $N \log 2$, corresponding to a Bernoulli system where spikes are drawn at random with equiprobability. However, due to the pairwise, triplets, and so on coefficients, the distribution corresponding to a given $\mathcal{H}$ is not Bernoulli (see fig. 5 and explanations below). The value of $\mathcal{H}$, averaged over samples, is equal to 0. This explains[10] why, in fig. 4, the pressure, averaged over samples, is equal to the entropy. We call this family of potentials "dense", since they produce dense rasters.

2. Monomials and coefficients are still drawn at random, but with a different distribution as in the dense case. There are NR monomials of range $k = 1, \ldots, R$. Thus, there are N rate coefficients. They are very negative[11] whereas other coefficients are drawn with a Gaussian distribution with mean 0.8 and variance 1. This produces sparse rasters with strong multiple correlations (fig. 2). Such rasters resemble much more to retina spike trains (fig. 3). The entropy of such potentials is growing slowly with N; the pressure is roughly constant. We call this family of potentials "sparse".

These two classes of potential display correlations which are significantly different from a Bernoulli model. To check this, we made the following test. We generate 100 potentials of a given type. For each potential, we generate a raster of length T. Then, we compute the empirical

[10] We have $\mathcal{P}[\mathcal{H}] = \mathcal{S}[\mu] + \mu[\mathcal{H}]$. In the dense case, the average of $\mu[\mathcal{H}]$, taken out of several potentials samples is close to 0 explaining why $\mathcal{P}[\mathcal{H}] \sim \mathcal{S}[\mu]$, when averaged over several samples, in the dense case.

[11] The rate coefficient $h_i = \log\left(\frac{r_i}{1-r_i}\right)$ is the chosen so that the rate $r_i \in [0 : 0.01]$ with a uniform probability distribution.

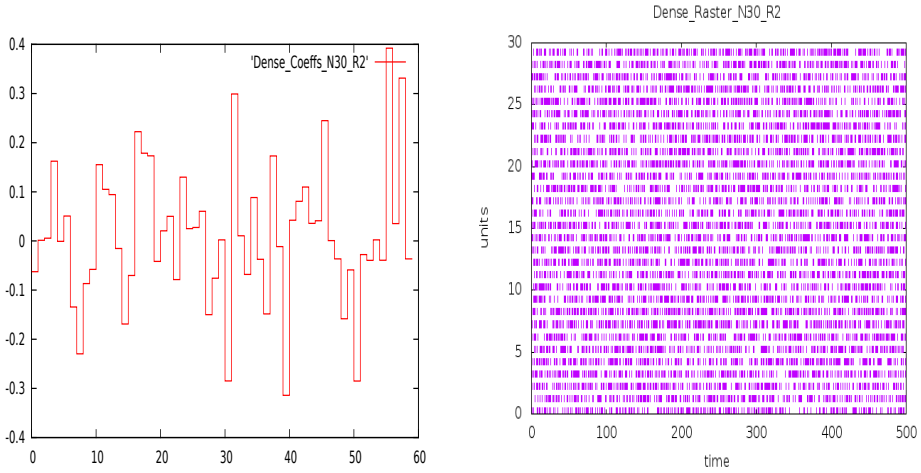


Figure 1: Left. Distribution of coefficients in a pairwise potential ($N = 30, R = 2$) producing a dense raster (Right).

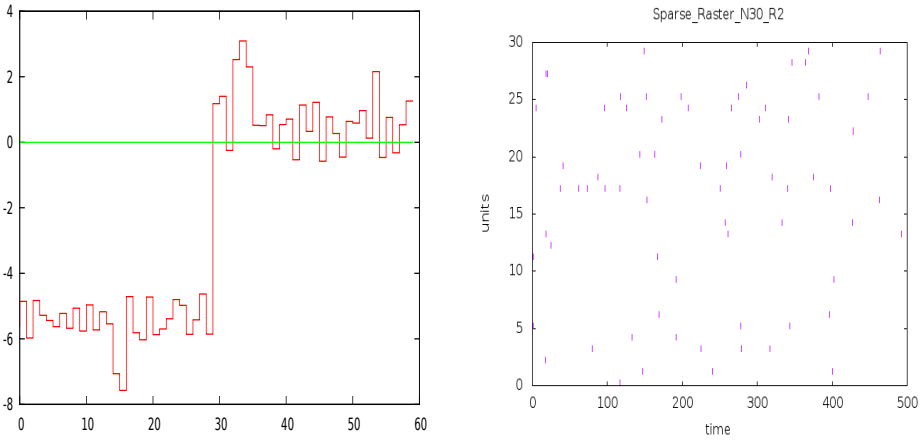


Figure 2: Left. Distribution of coefficients in a pairwise potential ($N = 30, R = 2$) producing a sparse raster (Right).

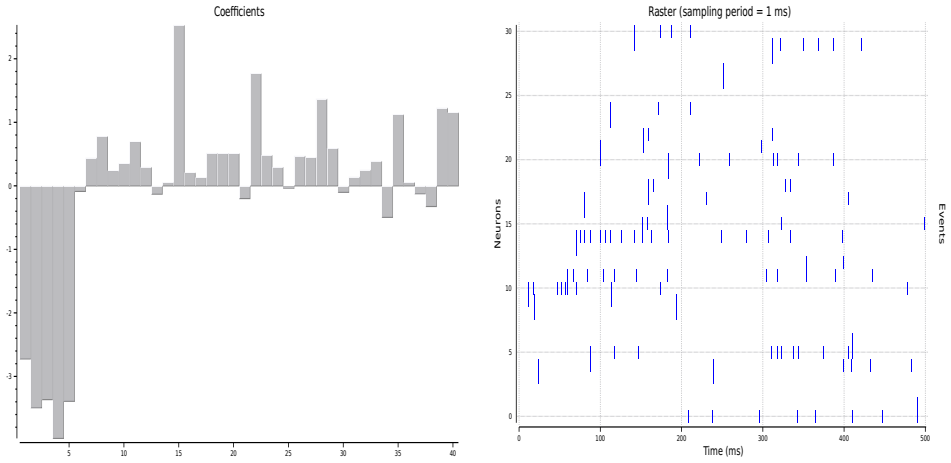


Figure 3: (Left.) Distribution of coefficients in a pairwise potential ($R = 2$) whose coefficients have been estimated in a raster from retinal spiking activity(Right). Raster: courtesy of M. J. Berry and O. Marre).

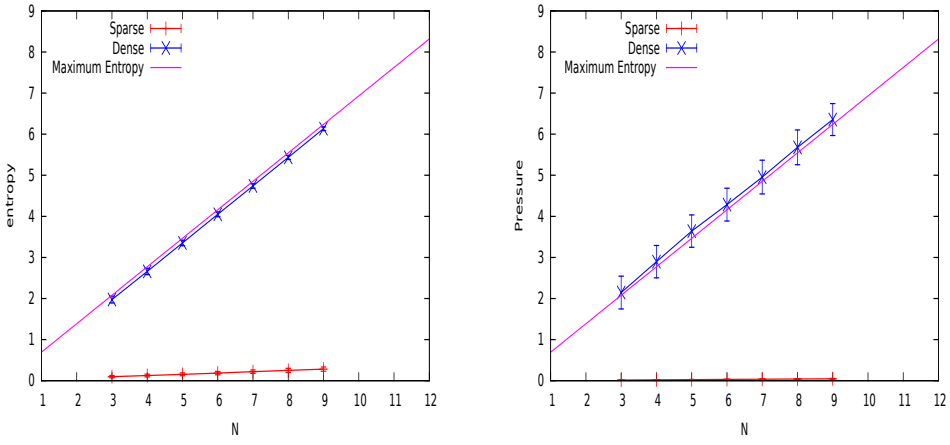


Figure 4: (Left) Average entropy as a function of N in the sparse and dense case. (Right) Average pressure as a function of N in the sparse and dense case. Averages and error bars has been taken out of 240 samples.

average $\pi_\omega^{(T)}[\,m_l\,]$ of each monomial m_l, with degree > 1, appearing in the potential. We compare this value with the average value $\mu_B[\,m_l\,]$ that this monomial would have if the raster were generated by a Bernoulli process. That is, if $m_l = \prod_r \omega_{k_r}(n_r)$ then $\mu_B[\,m_l\,] = \prod_r \mu_B[\,\omega_{k_r}(n_r)\,]$. If the raster were generated by a Bernoulli process then $\pi_\omega^{(T)}[\,m_l\,]$ would be a Gaussian random variable with mean $\mu_B[\,m_l\,]$ and mean-square deviation $\sigma_B[\,m_l\,] = \frac{1}{\sqrt{T}}\sqrt{\mu_B[\,m_l\,]\,(1-\mu_B[\,m_l\,])}$. As a consequence, the probability that $\left|\pi_\omega^{(T)}[\,m_l\,] - \mu_B[\,m_l\,]\right| > 5\sigma_B[\,m_l\,]$ would be smaller than 3×10^{-7}. Here, we cannot estimate exactly $\mu_B[\,m_l\,]$ for an arbitrary potential, since this requires to compute numerically the pressure, a task that becomes cumbersome as NR grows. Instead, we replace $\mu_B[\,\omega_{k_r}(n_r)\,]$ by the empirical probability $\pi_\omega^{(T)}[\,\omega_{k_r}(n_r)\,]$, so that $\mu_B[\,m_l\,]$ is approximated by $\pi_B[\,m_l\,] = \prod_r \pi_\omega^{(T)}[\,\omega_{k_r}(n_r)\,]$. We then compute, over the 100 potential samples the probability of the event:

$$D \stackrel{\text{def}}{=} \bigcup_l \left\{ \left|\pi_\omega^{(T)}[\,m_l\,] - \pi_B[\,m_l\,]\right| > \frac{5}{\sqrt{T}}\sqrt{\pi_B[\,m_l\,]\,(1-\pi_B[\,m_l\,])} \right\}. \tag{32}$$

Even if the raster is Bernoulli, this approximated probability deviates from the exact value 3×10^{-7}, due to the finite size fluctuations in $\pi_\omega^{(T)}[\,\omega_{k_r}(n_r)\,]$. Therefore, we compute this probability in the case of a Bernoulli raster of length T. Then, under the same conditions we compute this probability in the sparse and dense case, as a function of N, for $T = 10^6$. Results are shown in fig. 5.

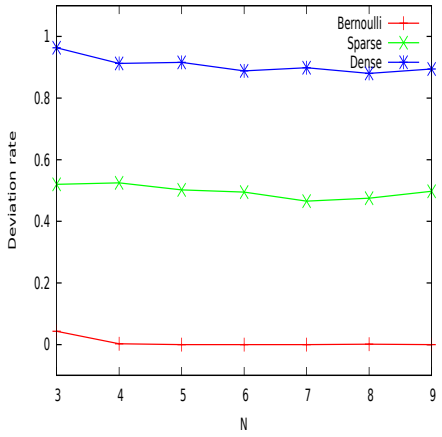


Figure 5: Probability of deviation from Bernoulli.

4.1.2 Estimating the topological pressure from the marked spectrum

We have estimated the pressure from the periodic orbit marked spectrum, when the parametric form of $\mathcal{H}$ is known. The algorithm is :

1. Compute the set of periodic orbits appearing in the raster ω and their relative weight: the conditional probabilities are estimated by empirical probabilities.

2. Compute the sum $\sum_{l=1}^{M} S_{\Phi(l)}$ where M is the total number of computed periodic orbits. Since $\sum_{l=1}^{M} S_{\mathcal{H}(l)}$ is known the quantity:

$$\hat{\mathcal{P}} \stackrel{\text{def}}{=} \frac{1}{M} \sum_{l=1}^{M} \left[S_{\mathcal{H}(l)} - S_{\Phi(l)} \right], \tag{33}$$

provides an estimate of $\mathcal{P}[\mathcal{H}]$. The empirical average over all periodic orbits reduces the error in empirical conditional probabilities estimation. However, when sampling the raster, some periodic orbit may appear with a very small (and unreliable) probability. This is prone to generate a strong noise in the pressure estimation. As a consequence, we introduced a threshold allowing to cut off orbits that didn't appear more than a certain fraction of the raster size. This is tuned by a variable called "threshold" θ.

We have plotted the average pressure computed from periodic orbits, for different values of θ, and compared it with the exact pressure (Perron-Frobenius) . For N ranging from 3 to 9, $R = 2$, we have drawn 240 potentials with random coefficients (sparse and dense case). For each potential, we have computed the exact pressure, and the estimated pressure (33) for a raster of length T where only the periodic orbits with blocks appearing more than $\theta \times T$ where stored. In Fig. 6 we have plotted the average exact and estimated pressures, with error bars, for different θ values, as a function of N.

The role of the threshold is clearly seen on these figures. In the dense case, the threshold has little impact for small N. When N increases, however, the estimated pressure departs from the exact one, and eventually ceases to be defined. This is because, as N growths, the entropy growths and the phase space to sample is larger and larger. As a consequence, a raster of size T contains relatively few blocks compared to the number of blocks in the support of the measure. Thus, the conditional probabilities computed from the raster have large errors inducing large errors in the pressure computation. This explains the observed deviation as N growths. Increasing the threshold increases the accuracy of empirical conditional probabilities, and reduces fluctuations. But then, less and less periodic orbits are selected until there is no periodic orbit any more. At this point the experimental curves ceases to be defined. Increasing T improves the situation as shown in fig. 6 left.

In the sparse case (Fig. 6 right), the estimated pressure has large error bars, decreasing as the threshold increases, while the average value tends to the exact one. Note that, in the tests we made, the estimated value of the pressure is completely wrong for a zero threshold, namely, it is largely outside the error bars tolerance (not shown). Also note that we were obliged to consider rasters of size $T = 10^6$ to have correct results.

These two figures clearly show that the estimation of the pressure is harder in the case of sparse potentials.

4.1.3 Estimating the shape of a potential from a raster

The Hammersley-Clifford decomposition allows an exact reconstruction of the potential from transition probabilities. We have implemented and tested. This method cannot be used for $NR > 20$ since it has to sample the set of all blocks.

Here, we reconstruct the potential from a raster of size T where transition probabilities are estimated from empirical averages. Two parameters are important: the raster size T and the threshold θ. As we saw in the previous sections θ can play a crucial role, since blocks with unreliable probabilities lead to severe errors in the free energy estimation. This effect is expected to be even more prominent in the potential parameters estimation since the Hammersley-Clifford hierarchy propagates errors.

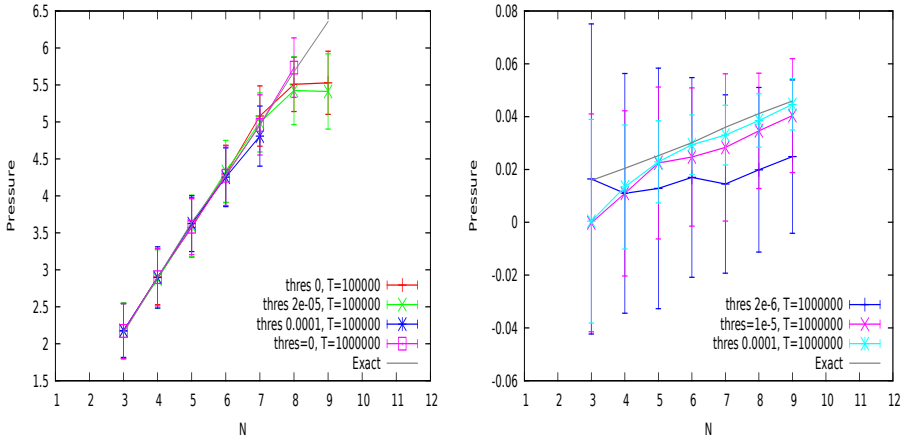


Figure 6: Average pressure estimated over periodic orbits as a function of N. (Right) Dense. (Left) Sparse. Averages and error bars has been taken out of 240 samples. In the sparse case, a threshold $\theta = 0$ give unreliable results so we didn't plotted it.

To verify these points we have drawn 10 random potentials (in the dense and sparse case). For each potential, we draw 24 rasters of size T and estimate the parameters for each raster. This provides error bars. Then, we compare to the exact value of the parameters. The results are drawn fig 7 and 8.

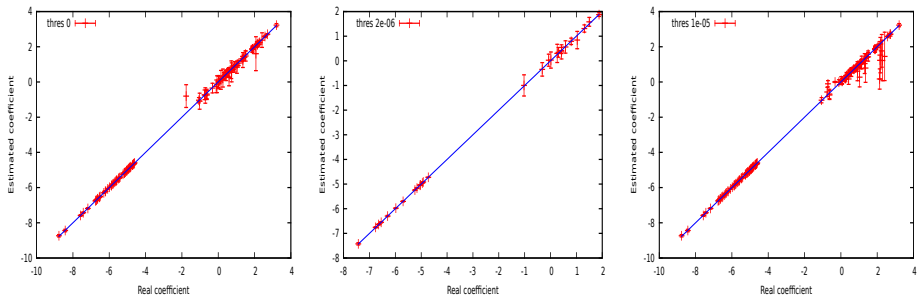


Figure 7: Comparison between estimated and reconstructed coefficients in the sparse case, for $T = 1000000$, $N = 7$, $R = 2$. From left to right: $\theta = 0; 2.10^{-6}; 10^{-5}$, corresponding respectively to keep blocks that has appeared at least $0; 2; 10$ times in the grammar. The estimation of coefficients was done over 10 potentials and the coefficients of all these potentials have been plotted. For each potential, the method is run 24 times, providing averages and error bars. The same potentials are used in each figure.

4.2 Exact recovery: The discrete time Leaky Integrate and Fire model

In this section we test our result in a stochastic discrete-time leaky Integrate-and-Fire model with noise and stimulus, first introduced by G. Beslon, O. Mazet and H. Soula (BMS) in [39] and ana-

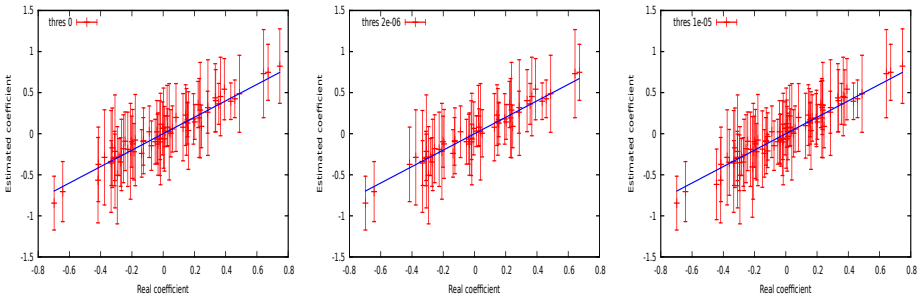


Figure 8: Same as fig. 7 in the dense case.

lyzed rigorously in [4, 6]. For this model the conditional probabilities can be explicitly computed, and the normalized potential can be constructed. We compute the interaction terms explicitly as a function of synaptic weights and stimulus for degree 0 and 1. For pairwise interactions we compute the instantaneous and 1-step time shifted case. We provide numerical simulations to illustrate our results. We also compare the synaptic weights and effective interactions graphs and discuss their differences.

This model is a discretization of the usual leaky Integrate-and-Fire model. Its dynamics reads:

$$V(t+1) = F(V(t)) + \sigma_B B(t), \tag{34}$$

where $V(t) = (V_i(t))_{i=1}^N$ is the vector of neuron's membrane potential at time t; $F(V)$ is a vector-valued function with entries:

$$F_i(V) = \gamma V_i(1 - Z[V_i]) + \sum_{j=1}^{N} W_{ij} Z[V_j] + I_i, \quad i = 1 \dots N$$

where $\gamma \in [0, 1[$, is the (discrete-time) "leak rate[12]"; Z is a function characterizing the neuron's firing: for a firing threshold $\theta > 0$, $Z(x) = 1$ whenever $x \geq \theta$ and $Z(x) = 0$ otherwise; I_i is an external current. In the most general version of this model, I_i depends on time. Here, we focus on the case where I_i is constant, ensuring the stationarity of dynamics and the existence/uniqueness of a Gibbs distribution[13]. Finally, in (34), $\sigma_B > 0$ is a variable controlling the noise intensity, where the vector $B(t) = (B_i(t))_{i=1}^N$ is an additive noise. It has Gaussian independent and identically distributed entries with zero mean and variance 1.

4.2.1 The normalized potential

The spike transition probabilities can be analytically computed in the model (34), and the normalized potential ϕ (13) as well. This potential has infinite range $R \to +\infty$. However, thanks to the leak term $\gamma < 1$, which ensures the existence and uniqueness of a Gibbs distribution in this model, one can approximate the exact infinite range by a finite range potential, where R has to be larger that the characteristic time scale of the leak $\frac{1}{|\log \gamma|}$. The approximate potential

[12] Thus, it corresponds to $\gamma = 1 - \frac{dt}{RC}$ in the continuous-time LIF model.

[13] In the sense of the variational principle (4) which is the common way to define Gibbs distributions in statistical physics. Note however that a more general definition of Gibbs distributions exist, which encompasses the non stationary case. The model (34) has a Gibbs distribution in this general case [6].

is given by:

$$\phi(\omega_0^D) = \sum_{k=1}^{N} \Big[\omega_k(D) \log \pi \left(X_k \left(\omega_0^{D-1} \right) \right) + \left(1 - \omega_k(D) \right) \log \left(1 - \pi(X_k \left(\omega_0^{D-1} \right) \right) \Big], \qquad (35)$$

where the function π is given by:

$$\pi(x) = \frac{1}{\sqrt{2\pi}} \int_x^{+\infty} e^{\frac{-u^2}{2}} du.$$

All functions appearing below depend on the spike block ω_0^{D-1} and make explicit the dependence of the network state (membrane potentials) on the spike history of the network.

The term:

$$X_k \left(\omega_0^{D-1} \right) = \frac{\theta - \mathcal{V}_k^{(det)} \left(\omega_0^{D-1} \right)}{\sigma_k(\omega_0^{D-1})}, \qquad (36)$$

contains the network spike history dependence of the neuron k at time D. More precisely, the term $\mathcal{V}_k^{(det)} \left(\omega_0^{D-1} \right)$ contains the deterministic part of the membrane potential of neuron k at time D, given the network spike history ω_0^{D-1}, whereas $\sigma_k(\omega_0^{D-1})$ characterizes the variance of the integrated noise in the neuron k's membrane potential (see [5] for details). We have:

$$\mathcal{V}_k^{(det)} \left(\omega_0^{D-1} \right) = \sum_{j=1}^{N} W_{kj} \, \eta_{kj} \left(\omega_0^{D-1} \right) + I_k \frac{1 - \gamma^{D - \tau_k \left(\omega_0^{D-1} \right)}}{1 - \gamma}.$$

The first term is the network contribution to the neuron k's membrane potential, where:

$$\eta_{kj} \left(\omega_0^{D-1} \right) = \sum_{l=\tau_k \left(\omega_0^{D-1} \right)}^{D-1} \gamma^{D-1-l} \omega_j(l) \qquad (37)$$

is the sum of spikes emitted by j in the past, with a weight γ^{D-1-l} corresponding to the leak decay of the spike influence as time goes on. The notation $\tau_k \left(\omega_0^{D-1} \right)$ means the last time before $D-1$ where neuron k has fired, with the convention that this time is 0 if neuron k didn't fire between 0 and $D-1$ in the block ω_0^{D-1}. In the definition of $\eta_{kj} \left(\omega_0^{D-1} \right)$ we sum from the initial time $\tau_k \left(\omega_0^{D-1} \right)$: this is because the membrane potential of neuron k is reset whenever k fires, hence loosing the memory of its past. Finally, in (36), we have:

$$\sigma_k^2(\omega_0^{D-1}) = \sigma_B^2 \frac{1 - \gamma^{2\left(D - \tau_k \left(\omega_0^{D-1} \right) \right)}}{1 - \gamma^2}.$$

The form (35) is very close to the Generalized Linear Model used for retina spike trains analysis [1] taking into account that time is discrete in our model. Equation (35) provides an example where the validity of the GLM can be mathematically proved and where the different terms can be interpreted in terms of the underlying network structure [3].

4.2.2 Explicit calculation of the canonical Hammersley-Clifford interactions

The goal now is to derive from (35) a non-normalized canonical potential $\mathcal{H}$ of the form (2) whose spike interactions terms h_l's are functions of the network parameters: the synaptic weight matrix $\mathcal{W}$ and the external stimulus I, $h_l \equiv h_l(\mathcal{W}, I)$.

Equation (17) gives a relation between the normalized potential and a cohomologous non-normalized potential. From this equation, after considering the elimination of equivalent interactions is it possible to compute explicitly the values of the interaction terms h_l's as a function of network parameters.

We have, from (21), $\Delta = \mathcal{P}[\mathcal{H}] - \mathcal{P}[\phi] = h_0(\mathcal{W}, I) - \phi_0(\mathcal{W}, I)$ with $\mathcal{P}[\phi] = 0$ (normalized potential). We may take $h_0(\mathcal{W}, I) = 0$. Additionally, $\phi_0(\mathcal{W}, I) = \phi(\omega^{(0)}) = \sum_{k=1}^{N} \log(1 - \pi(X_k(0))$, where $X_k(0)$ is the value of (36) when setting all spikes from 0 to D equal to 0. Thus, $\eta_{kj}(0) = 0, \sigma_k^2(0) = 1$ and $X_k(0) = \theta - I_k \frac{1-\gamma^D}{1-\gamma}$. As a consequence, $\phi_0 \equiv \phi_0(I) = \sum_{k=1}^{N} \log\left[1 - \pi\left(\frac{\theta - I_k \frac{1-\gamma^D}{1-\gamma}}{\sigma_B \sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}}\right)\right]$, which does not depend on $\mathcal{W}$.

Firing rates: Without loss of generality (since we are summing on periodic shifts) we take $\omega^{(l)}$ to be the block with a spike located at time $n = 0$. From time $n = 0$ to time $n = D - 1$ we have $\tau_k\left(\omega^{(\sigma^n l)}\right) = n\delta_{ik}$ whereas $\tau_k\left(\omega^{(\sigma^D l)}\right) = 0, \forall k$. From (37), for $0 \leq n < D$, $\eta_{kj}\left(\omega^{(\sigma^n l)}\right) = \gamma^{D-1-n}\delta_{ij}, \forall k$, whereas $\eta_{kj}\left(\omega^{(\sigma^D l)}\right) = 0, \forall k$. Likewise, for $0 \leq n < D$, $\sigma_k(\omega^{(\sigma^n l)}) = \sigma_B \sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}$ for $k \neq i$ and $\sigma_i(\omega^{(\sigma^n l)}) = \sigma_B \sqrt{\frac{1-\gamma^{2(D-n)}}{1-\gamma^2}}$, whereas $\sigma_k(\omega^{(\sigma^n l)}) = \sigma_B \sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}$ for all k's when $n = D$. We have thus:

$$X_k\left(\omega^{(\sigma^n l)}\right) = \begin{cases} \frac{\theta - W_{ki}\gamma^{D-1-n} - I_k \frac{1-\gamma^D}{1-\gamma}}{\sigma_B \sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}}, \; 0 \leq n < D, \; k \neq i; \\ \frac{\theta - W_{kk}\gamma^{D-1-n} - I_k \frac{1-\gamma^{D-n}}{1-\gamma}}{\sigma_B \sqrt{\frac{1-\gamma^{2(D-n)}}{1-\gamma^2}}}, \; 0 \leq n < D, \; k = i; \\ \frac{\theta - I_k \frac{1-\gamma^D}{1-\gamma}}{\sigma_B \sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}}, \; \forall k, \; n = D. \end{cases} \tag{38}$$

From (16):

$$\sum_{n=1}^{R} \mathcal{H}\left(\omega^{(\sigma^n l)}\right) = \sum_{n=1}^{R} \phi\left(\omega^{(\sigma^n l)}\right) + R\Delta.$$

Using the Hammersley-Clifford decomposition (8) for $\mathcal{H}$, and since $\omega^{(l)}$ has degree 1:

$$\sum_{n=1}^{R} \sum_{l'_n \sqsubseteq \sigma^n l} h_{l'_n} = Rh_0 + \sum_{n=1}^{R} h_{\sigma^n l} = \sum_{n=1}^{R} \phi\left(\omega^{(\sigma^n l)}\right) + R(h_0 - \phi_0).$$

The coefficients $h_{\sigma^n l}$ correspond to the same canonical constraint so we can fix all of them to 0 but one (section 3.3). We have thus:

$$h_l = \sum_{n=1}^{R} \phi\left(\omega^{(\sigma^n l)}\right) - R\phi_0. \tag{39}$$

The block $\omega^{(l)}$ has one spike corresponding for instance to neuron i. To make this dependence explicit we note $h_l \equiv \mathtt{h}_i$: this Hammersley-Clifford coefficient has the statistical physics interpretation of a local field acting on neuron i.

Finally, combining equations (35) , (38) and (39) we obtain:

$$\begin{array}{l} \mathtt{h}_i = \sum_{n=1}^{D-1} \sum_{k=1}^{N} \log\left[1 - \pi\left(X_k\left(\omega^{(\sigma^n l)}\right)\right)\right] + \\ \sum_{k \neq i} \log\left[1 - \pi\left(X_k\left(\omega^{(\sigma^D l)}\right)\right)\right] + \log\left[\pi\left(X_i\left(\omega^{(\sigma^D l)}\right)\right)\right] - R\phi_0. \end{array} \tag{40}$$

which is an explicit function of synaptic weights and stimuli. As a conclusion:

- The "local field" of neuron depends non linearly on *all* stimuli (not only I_i).
- It depends non linearly on the incoming synaptic weights connected to i. This dependence is weak since the synaptic weight W_{ki} is multiplied by a factor γ^D.

Pairwise interactions (instantaneous). As an example of degree 2 interaction let us compute the instantaneous pairwise terms ("Ising" interaction). The instantaneous interaction between neuron $j \to i$, denoted from now on $\mathtt{J}_{ij}$ to match statistical physics uses, correspond to a block $\omega^{(l)}$ with 2 '1's on the last column (time D). Again, the choice of the last time to locate the 1's is arbitrary.

The computation of $\phi\left(\omega^{(\sigma^n l)}\right)$ is very similar to the case of degree 1. Consider $\omega^{(l)}$ to be the block with two instantaneous spikes corresponding to neurons i and j located at time $n=0$. For $0 \leq n < D$ the last firing time is $\tau_k\left(\omega^{(\sigma^n l)}\right) = n\delta_{ik}\delta_{jk}$ whereas $\tau_k\left(\omega^{(\sigma^D l)}\right) = 0, \forall k$. From (37), for $0 \leq n < D$, $\eta_{kj'}\left(\omega^{(\sigma^n l)}\right) = \gamma^{D-1-n}(\delta_{ij'} + \delta_{jj'})$, $\forall k$, whereas $\eta_{kj}\left(\omega^{(\sigma^D l)}\right) = 0, \forall k$. Likewise, for $0 \leq n < D$, $\sigma_k(\omega^{(\sigma^n l)}) = \sigma_B\sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}$ for $k \neq i, j$ and $\sigma_{i,j}(\omega^{(\sigma^n l)}) = \sigma_B\sqrt{\frac{1-\gamma^{2(D-n)}}{1-\gamma^2}}$, whereas $\sigma_k(\omega^{(\sigma^n l)}) = \sigma_B\sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}$ for all k's when $n = D$.

We have thus:

$$X_k\left(\omega^{(\sigma^n l)}\right) = \begin{cases} \frac{\theta - (W_{ki}+W_{kj})\,\gamma^{D-1-n} - I_k\frac{1-\gamma^D}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}},\ 0 \leq n < D,\ k \neq i,j; \\ \frac{\theta - (W_{kk}+W_{kj})\,\gamma^{D-1-n} - I_k\frac{1-\gamma^{D-n}}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2(D-n)}}{1-\gamma^2}}},\ 0 \leq n < D,\ k = i; \\ \frac{\theta - (W_{kk}+W_{ki})\,\gamma^{D-1-n} - I_k\frac{1-\gamma^{D-n}}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2(D-n)}}{1-\gamma^2}}},\ 0 \leq n < D,\ k = j; \\ \frac{\theta - I_k\frac{1-\gamma^D}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}},\ \forall k,\ n = D. \end{cases} \tag{41}$$

We have, from (16) and the Hammersley-Clifford decomposition (8) for $\mathcal{H}$:

$$\mathtt{J}_{ij} = \sum_{n=1}^{R} \phi\left(\omega^{(\sigma^n l)}\right) + R\Delta - \sum_{n=1}^{R}\sum_{l'_n \sqsubset \sigma^n l} h_{l'_n}. \tag{42}$$

For blocks $l'_n \sqsubset \sigma^n l$ of degree 1 the spike is either on neuron i or neuron j. The contribution of these blocks is $\mathtt{h}_i + \mathtt{h}_j$. In the blocks $l'_n \sqsubset \sigma^n l$ there is also the block $\omega^{(0)}$, whose contribution is $R\phi_0$. Therefore, we have:

$$\mathtt{J}_{ij} = \sum_{n=1}^{R} \phi\left(\omega^{(\sigma^n l)}\right) - \mathtt{h}_i - \mathtt{h}_j - R\phi_0. \tag{43}$$

Replacing (41) in (35) to use in equation (43), one finally obtain $\mathtt{J}_{ij}$ as a explicit function of synaptic weights and stimulus.

Remarks:

- The "instantaneous pairwise" interaction depends not only on W_{ij}, but in all synaptic weights of neurons connected with i or j.
- It also depends in the stimulus of all neurons in the network.

Spatio-Temporal pairwise Interactions (1-time step). We compute the non-instantaneous 1-time step spike pairwise terms. The interaction between neuron $j \to i$ 1-time step shifted, denoted from now on $\mathtt{J}^1_{ij}$, correspond to a block $\omega^{(l)}$ with a '1' at neuron j in time $n-1$ and other 1 at neuron i in time n . Following as in previous examples, we obtain in this case:

$$X_k\left(\omega^{(\sigma^n l)}\right) = \begin{cases} \frac{\theta-(W_{ki}\gamma^{D-1-n}+W_{kj}\gamma^{D-2-n})-I_k\frac{1-\gamma^D}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}}, \; 0\leq n<D,\; k\neq i,j; \\ \frac{\theta-(W_{kk}\gamma^{D-1-n}+W_{kj}\gamma^{D-2-n})-I_k\frac{1-\gamma^{D-n}}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2(D-n)}}{1-\gamma^2}}}, \; 0\leq n<D,\; k=i; \\ \frac{\theta-(W_{ki}\gamma^{D-1-n}+W_{kk}\gamma^{D-2-n})-I_k\frac{1-\gamma^{D-n}}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2(D-(n+1))}}{1-\gamma^2}}}, \; 0\leq n<D,\; k=j; \\ \frac{\theta-W_{ij}-I_k\frac{1-\gamma^D}{1-\gamma}}{\sigma_B\sqrt{\frac{1-\gamma^{2D}}{1-\gamma^2}}}, \; \forall k,\; n=D. \end{cases} \tag{44}$$

As in the previous example we obtain $\mathtt{J}^1_{ij}$ as a explicit function of synaptic weights and stimulus. Note that in $\mathtt{J}^1_{ij}$ the synaptic weights have different influences in $X_k\left(\omega^{(\sigma^n l)}\right)$ due to the fact that they appear in different times in the raster.

4.2.3 When do effective interaction vanish ?

An interesting case to analyze is under which conditions on the synaptic weights and stimulus this coefficients vanishes. Consider e.g. $N=2$ and $R=2$ in a BMS model. The firing rate of the neuron 1 corresponds to the canonical interaction associated to the block which has only the neuron number 1 firing at time 1: $\left[\begin{smallmatrix} 0 & 1 \\ 0 & 0 \end{smallmatrix}\right]$, following the spike block representation, presented in (2.3) this block has associated the monomial h_4, which can be computed explicitly.

$$\begin{aligned} h_4 &= \log(1-\pi(\frac{\theta-(W_{11}+I_1)}{\sigma_B^2}))+\log(1-\pi(\frac{\theta-(W_{21}+I_2)}{\sigma_B^2}))) \\ &+ \log(\pi(\frac{\theta-I_1}{\sigma_B^2}))+\log(1-\pi(\frac{\theta-I_2}{\sigma_B^2})))-2(\log(1-\pi(\frac{\theta-I_1}{\sigma_B^2}))+\log(1-\pi(\frac{\theta-I_2}{\sigma_B^2}))) \end{aligned}$$

Which vanishes for all the parameters of the network satisfying this equation with $h_4=0$. Take for instance: $\theta=1, \sigma_B=1, W_{11}=2.1238, I_2+W_{21}=0.1409, I_1=0.158, I_2=1$

Therefore, in general, this parameters do not vanish. More generally, all h_l's in this model are *generically* non zero.

4.2.4 A numerical investigation

Here, we compute the parametric potential

$$\mathcal{H} = \sum_{l=0}^{L(N,R)} h_l m_l, \tag{45}$$

with range R, equivalent to the normalized BMS potential. The goal is to compare the effective graph of interactions provided by the h_ls to the real synaptic interactions.

For this, we generate first a sparse random graph of synaptic interactions. Each node receives K arrows among the N possible. Each arrow is weighted by a random Gaussian synaptic weight with mean zero and variance $\frac{J^2}{N}$. In the simulations, we took $K = 2$ and $J^2 = 3$. The input is constant equal to 0.7 (with a firing threshold equal to 1). An example of a resulting synaptic graph are given in fig. 9

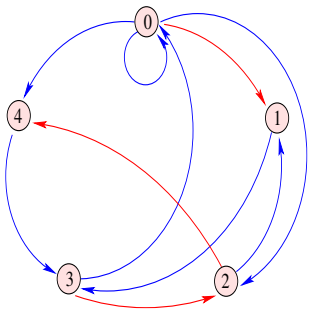


Figure 9: (Left). Synaptic graph for $N = 5$.

Then, we compute the potential (2) equivalent to the truncation of BMS potential to a range R. For low $N\,R$ ($N\,R \leq 20$) this is done by sampling the whole set of blocks (2^{NR}) in order to have an exact transformation and using the exact transformation for ϕ to $\mathcal{H}$. With $N = 5, R = 4$ we have thus already $2^{20} = 1048576$ blocks. This is already quite huge and requires large memory and CPU time. If we were to study e.g. the case $N = 8$ for $R = 4$ we would be out of the limit of what is numerically computable ($2^{32} = 4294967296$).

We have thus made the subsequent tests in the case $N = 5$, $\gamma = 0.2$, $\sigma_B = 0.2$. In fig. 10 we have represented the KL divergence (6) between the empirical measure of a raster produced by (34) and the Gibbs measure with potential (35) as a function of T for $R = 2, 3, 4$. Seemingly, $R = 3$ (2^{15} blocks) provides already a good convergence. We compute $\mathcal{H}$ with our method:

- We compute the conditional probabilities $P\left[\omega(0) \,\middle|\, \omega_0^{D-1}\right]$ associated to (2) and compare them to the exact BMS conditional probabilities of blocks ω_0^D: for the latter it corresponds to compute the conditional probability of infinite blocks where the patterns $\omega(t)$ corresponding to $t > D$ are all set to 0. (fig. 11, left).
- We compute the probabilities of blocks ω_0^k, $k = 1, \ldots$ predicted by (2) to the empirical probabilities of the same blocks, obtained in raster generated with the BMS model. (fig. 11, right).

Remark. We would like here to point out that the exact determination of the potential raises, beyond the number of blocks, additional problems. The computation, based on Hammersley-Clifford hierarchy, sums up terms which are logarithm of conditional probabilities. Due to noise, every conditional probability is positive so that logarithms are finite. Nevertheless, it can be that some of these probabilities is very small, leading to large negative terms in the Hammersley-Clifford hierarchy. This provides a $\mathcal{H}$ potential with large terms. Now, the exact computation of the conditional and joint probabilities corresponding to $\mathcal{H}$ is done via Perron-Frobenius theorem which requires taking the exponential of $\mathcal{H}$. This can lead to severe numerical instabilities, as we checked. Using e.g. MonteCarlo methods is more robust, but less accurate.

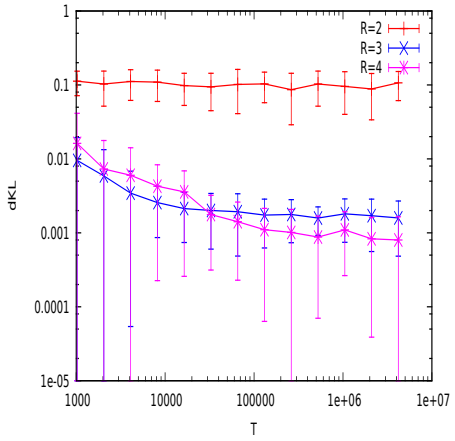


Figure 10: Plot of the Kullback-Leibler divergence (6) as a function of T for $N = 5$, $R = 2, 3, 4$.

4.2.5 Graphs of interactions

Finally, we compute the graph of "Ising" effective interactions J_{ij} associated with $\mathcal{H}$ as well as the one step pairwise interactions $J_{ij}(1)$ (Fig. 12).

5 Conclusion

In this work, we have introduced a method allowing to recover the potential that has generated a raster distributed according to a Gibbs distribution having this potential.

The method is exact when the raster is infinite, or, equivalently, when transition probabilities are known exactly. This situation arises in several known examples in the literature: heuristic form of conditional intensities (LN or GLM), or neural networks models. Concerning neural network models, although we have focused in this paper on a discrete time leaky Integrate-and-Fire model, this analysis extends as well to continuous time conductance-based Integrate-and-Fire models with chemical and electric synapses [7, 11].

The method works as well for finite size rasters, although, in this case, fluctuations on conditional probabilities estimations dramatically affect the reliability of coefficients, depending on their rank in the Hammersley-Clifford hierarchy. This effect is however not intrinsic to our method. Instead, it simply reflects a well known effect in statistics. Adding more and more parameters to get a better fit ultimately leads to a breakdown of the estimation for the whole set of parameters. In the realm of the maximum entropy principle, this can be rephrased as follows. To estimate a set of parameters h_l one computes the empirical average of the conjugated monomials m_l. These averages have fluctuations δm_l. This induces fluctuations of the coefficient h_l. These fluctuations, when they are small, are related by $\delta m = \chi \delta h$, where $\delta m, \delta h$ are the vectors of $\delta m_l, \delta h_l$ and χ is the second derivative of free energy, that can be written as a sum of correlations functions. Although χ is a positive matrix, some eigenvalues are close to zero. The larger the order of the potential the more small eigenvalues there are, the larger are the fluctuations on the h_ls. Presumably, our method has therefore to be used in combination to filtering methods based on the raster observation to eliminate initially, the degree of freedom

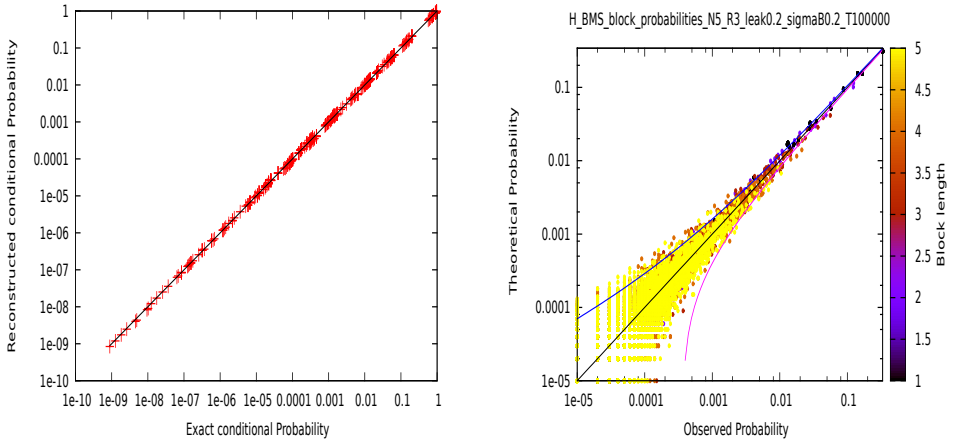


Figure 11: (Left). Exact conditional probabilities for blocks of range R and BMS potential, vs exact conditional probabilities associated with the potential (2). (Right) Empirical probabilities of blocks ω_0^k, $k = 1, \dots 5$, obtained in a BMS raster of size $T = 100000$ vs the probabilities of the same blocks predicted by (2). This figure corresponds to $N = 5, R = 3, \gamma = 0.2, \sigma_B = 0.2$ and to the synaptic graph of Fig. 9).

(monomials) leading to such large fluctuations (and determining the maximal degree and the potential range). Such methods will be presented in a separate work.

This work leaves several open questions and further directions for future research.

- **Functional connectivity and stimulus dependence.** Our method provides a detailed explicit description of how the synaptic weights and stimulus in a leaky Integrate-and-Fire neural network shape the form of the maximum entropy potential. Previous studies have investigated how specific aspects of the network structure or stimulus affects spike correlations [45, 43]. Our results show that no straightforward correspondence exists between structural parameters (stimulus, synaptic weights) and Maximum Entropy parameters. The latter are functions of the former, but the relationship is non-linear. Moreover, while there are of order N^2 structural parameters, the number of Maximum Entropy parameters increases exponentially fast with N, the number of neurons. Thus, there is considerable amount of redundant information in the h_l's.

- **Typical form of potential in real data.** As we showed, there is in general no reason why a (canonical) coefficient h_l ought to vanish. Therefore, one expects that a *generic* MaxEnt potential contains *all* canonical terms. As an example, in the discrete time leaky Integrate-and-Fire model that we have presented, and even focusing on a finite range approximation of the potential, there are generically $2^{NR} - 2^{N(R-1)}$ canonical terms. This rules out any hope to apply the Maximum Entropy Principle to characterize the statistics of Gibbs distributions generically arising in neural networks models. However, this statement contradicts the indisputable success of applying MaxEnt to fit real spike trains in the retina ([36, 31, 14, 15, 44]) and in the cortex [22]. This may be due to two reasons.

  - Real neural networks are *not generic*. The structural parameters of a real neural network are not drawn at random from some a priori distribution: they are the result

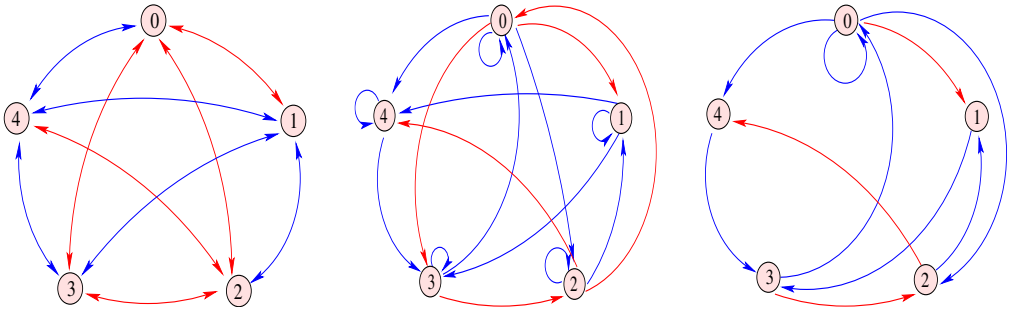


Figure 12: Effective graph for $N = 5, R = 3, \gamma = 0.1, \sigma_B = 0.3$. (Left) Ising Graph. (Middle) $J_{ij}(1)$ graph. (Right) Synaptic graph. As expected these graphs differ. Ising and $J_{ij}(1)$ graph are complete and non-causal.

of a long term genetic evolution and mechanisms such as synaptic or intrinsic plasticity. Especially, the effect of synaptic plasticity on Gibbs statistics can be studied mathematically (see next section).

- Binning is currently used in spike train data. It has the effect of removing time-correlations and it clearly simplifies the shape of the potential. The effect of binning will be studied in a separate paper.

- **Effects of synaptic plasticity.** Plasticity mechanisms induce changes on synaptic weights, depending (in a first approximation) to moments in the spikes distribution (rates, pairwise). These changes induce in turn variations in the spike statistics. These variations can be studied in the realm of MaxEnt, assuming that synaptic changes are slow enough so that spike dynamics can be considered as stationary (adiabatic approximation) [9]. It has been shown in [12, 37, 38, 25] that Hebbian and intrinsic plasticity drive a neural network at "the edge of chaos", where it becomes highly sensitive to learned stimuli. The same type of effect has been reported in [39] when a Integrate and Fire model is submitted to Spike-Time Dependent Synaptic Plasticity. These examples show that plasticity clearly generates non generic dynamical systems. What is the typical Gibbs distribution of spiking neural networks models evolving under such plasticity rules ? This will be studied in a forthcoming paper.

- **Non stationary data.** As mentioned in the introduction, the MaxEnt principle heavily relies on the highly questionable assumption of stationarity. Although Gibbs distributions can also be defined for non stationary processes [13], in the context of neural networks [3], they do not obey the MaxEnt principle. Although stationarity can be defended in the case of movies presented to a retina [36], in many experiments the spike response to flashed stimuli is considered. In this case, there is no way to defend that spike train statistics is stationary. So, is there any use in studying MaxEnt Gibbs distributions ? An interesting hypothesis would be to assume, in the case of flashed images, that spontaneous dynamics is stationary, and has a MaxEnt Gibbs distributions. Then, the response to a stimuli could be analyzed in the realm of (linear) response theory. From this, analytic form of response kernels ("receptive fields") can be inferred. But the correct estimation of the response relies on a correct characterization of the spontaneous spike activity. The method presented here could be a way to have a correct description of the Gibbs distribution corresponding to spontaneous activity.

The spike train statistics depend not only on the stimulus, but also on a constantly changing underlying neural structure. In particular, for the retina we believe our work is a step forward toward the understanding of multi-electrode arrays data and how the statistics obtained from the spiking activity link with the retinal properties and stimulus. We hope that new methods in physiology and data analysis can help to gain new insights about these relationship.

Acknowledgments This work was supported by the French ministry of Research and University of Nice (EDSTIC), INRIA, ERC-NERVI number 227747, KEOPS ANR-CONICYT and European Union Project # FP7-269921 (BrainScales), Renvision # 600847 and Mathemacs # FP7-ICT-2011.9.7.

References

[1] Yashar Ahmadian, Jonathan W. Pillow, and Liam Paninski. Efficient Markov Chain Monte Carlo Methods for Decoding Neural Spike Trains. *Neural Computation*, 23(1):46–96, January 2011.

[2] J. Besag. Spatial interaction and the statistical analysis of lattice systems (with discussion). *Journal of Royal Statistical Society*, 2:192–236, 1974.

[3] B. Cessac and R. Cofré. Spike train statistics and gibbs distributions. *J. Physiol. Paris*, 2013. In Press.

[4] Bruno Cessac. A discrete time neural network model with spiking neurons. rigorous results on the spontaneous dynamics. *J. Math. Biol.*, 56(3):311–345, 2008.

[5] Bruno Cessac. A view of neural networks as dynamical systems. *International Journal of Bifurcations and Chaos*, 20(6):1585–1629, 2010.

[6] Bruno Cessac. A discrete time neural network model with spiking neurons ii. dynamics with noise. *J. Math. Biol.*, 62:863–900, 2011.

[7] Bruno Cessac. Statistics of spike trains in conductance-based neural networks: Rigorous results. *Journal of Mathematical Neuroscience*, 1(8), 2011.

[8] Bruno Cessac and Adrian Palacios. Spike train statistics from empirical facts to theory: the case of the retina. In Frédéric Cazals and Pierre Kornprobst, editors, *Modeling in Computational Biology and Biomedicine: A Multidisciplinary Endeavor*, Lectures Notes in Mathematical and Computational Biology (LNMCB). Springer-Verlag, 2012.

[9] Bruno Cessac, H. Rostro-Gonzalez, Juan-Carlos Vasquez, and Thierry Viéville. How gibbs distribution may naturally arise from synaptic adaptation mechanisms: a model based argumentation. *J. Stat. Phys*, 136(3):565–602, August 2009.

[10] J.R. Chazottes and G. Keller. *Mathematics of Complexity and Dynamical Systems*, chapter Pressure and Equilibrium States in Ergodic Theory, pages 1422–1437. Springer, 2011.

[11] Rodrigo Cofré and Bruno Cessac. Dynamics and spike trains statistics in conductance-based integrate-and-fire neural networks with chemical and electric synapses. *Chaos, Solitons and Fractals*, 50(8):13–31, 2013.

[12] E. Daucé, M. Quoy, Bruno Cessac, B. Doyon, and M. Samuelides. Self-organization and dynamics reduction in recurrent networks: stimulus presentation and learning. *Neural Networks*, 11:521–33, 1998.

[13] Roberto Fernandez and Grégory Maillard. Chains with complete connections : General theory, uniqueness, loss of memory and mixing properties. *J. Stat. Phys.*, 118(3-4):555–588, 2005.

[14] Elad Ganmor, Ronen Segev, and Elad Schneidman. The architecture of functional interaction networks in the retina. *The journal of neuroscience*, 31(8):3044–3054, 2011.

[15] Elad Ganmor, Ronen Segev, and Elad Schneidman. Sparse low-order interaction network underlies a highly correlated and learnable neural population code. *PNAS*, 108(23):9679–9684, 2011.

[16] Hans-Otto Georgii. *Gibbs measures and phase transitions.* De Gruyter Studies in Mathematics:9. Berlin; New York, 1988.

[17] J. M. Hammersley and P. Clifford. Markov fields on finite graphs and lattices. *unpublished*, 1971.

[18] Mark Kac. Can one hear the shape of a drum? *The American Mathematical Monthly*, 73(4):1–23, April 1966.

[19] G. Keller. *Equilibrium States in Ergodic Theory.* Cambridge University Press, 1998.

[20] R. Kindermann and J.L. Snell. *Markov Random Fields and Their Applications.* American Mathematical Society, 1980.

[21] A. Livšic. Cohomology properties of dynamical systems. *Math. USSR- Izvestia*, (6):1278–1301, 1972.

[22] O. Marre, S. El Boustani, Y. Frégnac, and A. Destexhe. Prediction of spatiotemporal patterns of neural activity from pairwise correlations. *Physical review letters*, 102(13), April 2009.

[23] John Moussouris. Gibbs and markov random systems with constraints. *J. Stat. Phys.*, 10(1):11–33, 1974.

[24] H. Nasser, O. Marre, and B. Cessac. Spatio-temporal spike trains analysis for large scale networks using maximum entropy principle and monte-carlo method. *Journal Of Statistical Mechanics*, 2013.

[25] Jeremie Naude, Bruno Cessac, Hugues Berry, and Bruno Delord. Effects of cellular homeostatic intrinsic plasticity on dynamical and computational properties of biological recurrent neural networks. *J. of Neuroscience, to appear*, 2013.

[26] Ifije E. Ohiorhenuan, Ferenc Mechler, Keith P. Purpura, Anita M. Schmid, Qin Hu, and Jonathan D. Victor. Sparse coding and high-order correlations in fine-scale cortical networks. *Nature*, 466(7306):617–621, 2010.

[27] D. Ornstein and B. Weiss. Entropy and data compression schemes. *IEEE Trans. Inform. Theory*, 39:78–83, 1993.

[28] Srdjan Ostojic and Nicolas Brunel. From spiking neuron models to linear-nonlinear models. *PLoS Comput Biol*, 1(7), 2011.

[29] W. Parry and M. Pollicott. *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, volume 187–188. Asterisque, 1990.

[30] P.Clifford. *Disorder in Physical Systems: A Volume in Honour of John M. Hammersley,*, pages 19–32. Oxford University Press, 1990.

[31] J W Pillow, J Shlens, L Paninski, A Sher, A M Litke, E J Chichilnisky, and E P Simoncelli. Spatio-temporal correlations and visual signaling in a complete neuronal population. *Nature*, 454(7206):995–999, Aug 2008.

[32] Jonathan W. Pillow, Yashar Ahmadian, and Liam Paninski. Model-based decoding, information estimation, and change-point detection techniques for multineuron spike trains. *Neural Comput.*, 23(1):1–45, 2011.

[33] Mark Pollicott and Howard Weiss. Free energy as a dynamical invariant (or can you hear the shape of a potential?). *Communications in Mathematical Physics*, 240:457–482, 2003.

[34] D. Ruelle. *Statistical Mechanics: Rigorous results*. Benjamin, New York, 1969.

[35] O. Sarig. Thermodynamic formalism for countable markov shifts. http://www.wisdom.weizmann.ac.il/ sarigo/TDFnotes.pdf, 2010.

[36] E. Schneidman, M.J. Berry, R. Segev, and W. Bialek. Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature*, 440(7087):1007–1012, 2006.

[37] B. Siri, H. Berry, Bruno Cessac, B. Delord, and M. Quoy. Effects of hebbian learning on the dynamics and structure of random networks with inhibitory and excitatory neurons. *Journal of Physiology, Paris*, 101(1-3):138–150, 2007. e-print: arXiv:0706.2602.

[38] B. Siri, H. Berry, Bruno Cessac, B. Delord, and M. Quoy. A mathematical analysis of the effects of hebbian learning rules on the dynamics and structure of discrete-time random recurrent neural networks. *Neural Comp.*, 20(12):12, dec 2008. e-print: arXiv:0705.3690v1.

[39] H. Soula, G. Beslon, and O. Mazet. Spontaneous dynamics of asymmetric random recurrent spiking neural networks. *Neural Computation*, 18(1), 2006.

[40] Aonan Tang, David Jackson, Jon Hobbs, Wei Chen, Jodi L. Smith, Hema Patel, Anita Prieto, Dumitru Petrusca, Matthew I. Grivich, Alexander Sher, Pawel Hottowy, Wladyslaw Dabrowski, Alan M. Litke, and John M. Beggs. A maximum entropy model applied to spatial and temporal correlations from cortical networks *In Vitro*. *The Journal of Neuroscience*, 28(2):505–518, January 2008.

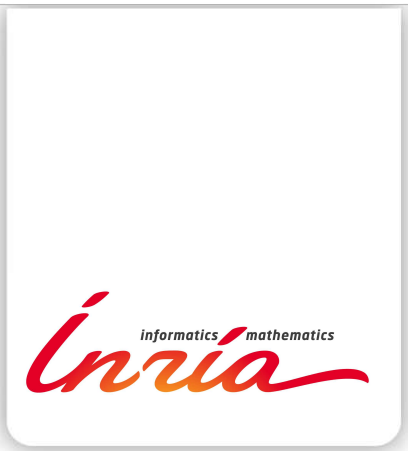
[41] G. Tkačik, E. Schneidman, M.J. Berry II, and W Bialek. Ising models for networks of real neurons. *arXiv q-bio/0611072*, 2006.

[42] G. Tkačik, Elad Schneidman, Michael J. Berry II, and William Bialek. Spin glass models for a network of real neurons. *arXiv: 0912.5409v1*, 2009.

[43] S. Cardanobile V. Pernice, B. Staude and S. Rotter. How structure determines correlations in neuronal networks. *PLoS Computational Biology*, 7(5):P03008, 2013.

[44] Juan Carlos Vasquez, Olivier Marre, Adrian G Palacios, Michael J Berry, and Bruno Cessac. Gibbs distribution analysis of temporal correlation structure on multicell spike trains from retina ganglion cells. *J. Physiol. Paris*, 106(3-4):120–127, May 2012.

[45] K. Josić Y. Hu, J.Trousdale and E. Shea-Brown. Motif statistics and spike correlations in neuronal networks. *Journal of Statistical Mechanics*, page P03012, 2013.



RESEARCH CENTRE
SOPHIA ANTIPOLIS – MÉDITERRANÉE

2004 route des Lucioles - BP 93
06902 Sophia Antipolis Cedex

Publisher
Inria
Domaine de Voluceau - Rocquencourt
BP 105 - 78153 Le Chesnay Cedex
inria.fr

ISSN 0249-6399