From inverse problems to learning: a Statistical Mechanics approach

Carlo Baldassi¹, Federica Gerace², Luca Saglietti² and Riccardo Zecchina¹

¹ Bocconi Institute for Data Science and Analytics, Bocconi University, Milan, Italy
² Dipartimento di Scienza Applicata e Tecnologia, Politecnico di Torino, Italy
E-mail: riccardo.zecchina@unibocconi.it

Abstract. We present a brief introduction to the statistical mechanics approaches for the study of inverse problems in data science. We then provide concrete new results on inferring couplings from sampled configurations in systems characterized by an extensive number of stable attractors in the low temperature regime. We also show how these result are connected to the problem of learning with realistic weak signals in computational neuroscience. Our techniques and algorithms rely on advanced mean-field methods developed in the context of disordered systems.

1. Introduction

The enormous amount of data which is rapidly accumulating across scientific disciplines has given to inverse problems a renewed fundamental role. For instance, the problem of inferring effective or direct interactions in systems composed of many heterogeneous degrees of freedom is at the root of many fields, ranging from natural sciences to social sciences. Quite often, the systematic experimental measurements of the interactions are difficult to perform, and to estimate them one needs to resort to indirect statistical inference methods. These need to be computationally efficient in order to process experimental datasets which are typically huge. During the last decade, statistical inference challenges have stimulated lots of research efforts in the statistical physics community [1] in the attempt to adapt rigorous theories and algorithms to the problem of going from observables to parameters, somehow the inverse of what statistical physics has mostly been developed for.

Spectacular examples of inverse problems can be found in the biological sciences where experimental data on neuronal populations [2, 3], proteins [4], biomolecules and genetic interactions [5], have largely surpassed the understanding of the underlying complex chemical mechanisms. In physics too there exist many examples of inverse statistical problems, e.g. the design of many-body systems with particular desired properties (e.g. [6], see refs. [1]) or to learn from data some effective parametrizations of wave-functions (e.g. [7]). The technological progress is rapidly opening up many more applications.

Inverse problems and statistical inference are closely linked to machine learning. Here we will focus on unsupervised learning, i.e. extracting information from unlabeled data [8]. The vast majority of the data are indeed unlabeled, and unsupervised learning is possibly the most
challenging branch of machine learning, which has a direct mapping onto maximum entropy ideas [9] and statistical mechanics methods.

Adopting the physics notation, the unsupervised learning problem can be stated as follows. We assume that a set of $M$ i.i.d. sampled configurations $D = \{\sigma^{(\mu)}\}_{\mu=1,\ldots,M}$ drawn from an unknown distribution $P(\sigma) = \frac{1}{Z} e^{-H(\sigma)}$ are given. The configurations components $\sigma^{(\mu)}$ can be either continuous or discrete. We want to identify a probability distribution over the energy functions $H(\sigma)$ which is compatible with the data. The likelihood of the model given the data is defined as the posterior distribution $P(H|D) \propto \exp(-M (\langle H \rangle + \log Z)) P(H)$, where $\langle H \rangle = \frac{1}{M} \sum_{\mu=1}^{M} H(\sigma^{(\mu)})$ is the average over the data and $P(H)$ is the prior about $H$. The number of samples $M$ plays a crucial role, quite similar to inverse temperature: when $M$ is large, $P(H|S)$ concentrates on the maximums with respect to $H$ of the log-likelihood $\mathcal{L} = \langle H \rangle + \log Z$. Inferring the most probable $H$ from the data amounts at identifying the maximum of $\mathcal{L}$. This is a formidable optimization problem, due to the difficulty of estimating $Z$ on a candidate solution for $H$.

In what follows, for the sake of simplicity, we assume that the dynamical variables $\{\sigma\}$ are binary, i.e. Ising spins. The simplest non-trivial scenario assumes that we can have access to random configurations which allow us to measure non vanishing pairwise correlations between spins. In absence of correlations, the problem trivializes. In a general setting, the couplings characterizing the $H(\sigma)$ we want to infer will not all be positive. Moreover they will be embedded in some high-dimensional space. A concrete working example is provided by a system of $N$ spins $\{\sigma_i \equiv s_i = \pm 1\}$ which interact according to a generic (unknown) spin glass like Hamiltonian, i.e. spins are coupled by pairwise couplings $J_{ij}$ and are subject to external magnetic fields $h_i$. The unknown distribution from which the random configurations (the data) are extracted is a Boltzmann distribution

$$P(\{s_i\}) = \frac{1}{Z} e^{\sum_{i<j} J_{ij} s_i s_j + \sum_i h_i s_i} \quad (1)$$

where we have included the inverse temperature factor into the couplings and fields. The problem we want to solve is the determination of the most probable values for the couplings $J_{ij}$ and local fields $h_i$, given a set of $M$ observed spin configurations. Depending on the particular nature of the system at hand, the functional form of the Hamiltonian may be need to include higher order terms. Or even the choice of the form of the probability distribution may not be appropriate (e.g. if the systems are out of equilibrium). Still, the basic ideas are already captured by the above Ising like choice. It is in fact well known that the distribution (1) is not only the equilibrium distribution of the Ising model but, more importantly in our context, it is also the form of the distribution which maximizes the (Gibbs) entropy

$$S[P] = - \sum_{\{s_i\}} P(\{s_i\}) \ln P(\{s_i\}) \quad (2)$$

with $P(\{s_i\})$ satisfying the constraints of normalization and of having first and second moments (magnetizations and correlations) that match some given experimentally measured values. The maximum entropy $P(\{s_i\})$ is thus the probability distribution which matches the constraints given by some observables, without making any further assumption on the data, i.e. being the least biased distribution [10]. In settings where higher moments are not captured by (1) one must resort to maximum entropy distributions which match such moments and have multi-body interactions. Finding the parameters of (1) is a delicate computational problem: any small change in the couplings affects correlations between many spin variables and vice versa. Determining correlations for a given value of the parameters is a difficult computational problem which needs to be repeated multiple times in the search for the optimal couplings. Several techniques can be
used, ranging from Monte Carlo chains, to variational methods to mean-field techniques. Inverse statistical problems are obviously not new in statistical mechanics as well as in statistics and applied probability. Language barriers quite often have hampered the communication between disciplines.

In what follows we will provide some preliminary results in which we discuss how the above ideas can be used to infer couplings in systems characterized by multiple attractors. As a by-product we will show how the overall technique can lead to some interesting learning mechanisms which might be of interest in modeling neural systems. We will perform sampling by a mean-field technique developed in the context of disordered systems[11]. Still other approaches, e.g. Monte Carlo Markov chains, could be used.

2. Attractor networks

The celebrated Hopfield model [12] suggests a way of modeling associative memory as a process through which a set of $M$ memories $\{\xi^\mu\}_{\mu=1}^M$ are stored within a recurrent neural network as the minima of an Ising-like Hamiltonian. The network is modeled as consisting of $N$ binary stochastic units representing the neurons. At each time $t$, a unit state $s_i^t \in \{-1, 1\}$ (representing whether the neuron is active or quiescent) is determined from the other units according to the probability distribution

$$P(s_i^t) \propto \exp\left(\beta s_i^t \left( \sum_{j \neq i} J_{ij} s_j^{t-1} + h_i^{\text{ext}} \right) \right) \tag{3}$$

where $\beta$ controls the level of stochasticity, the quantities $J_{ij}$ model the synaptic connection strengths and $h_i^{\text{ext}}$ some external input. As $\beta \to \infty$, the neurons become deterministic and their state is simply given by the sign of the total local field $h_i^{\text{tot}} = \sum_{j \neq i} J_{ij} s_j^{t-1} + h_i^{\text{ext}}$ acting upon them. The above equation implements a parallel dynamics, in which time is discretized and all units are updated simultaneously; it is straightforward to implement other update protocols (e.g. random sequential); in such case, and if the synaptic connections are assumed to be symmetric, $J_{ij} = J_{ji}$, the probability distribution of the whole network matches Eq. (1) with $h = h_i^{\text{ext}}$. This insight allows one to use the tools from statistical mechanics to study the behavior or the network for given $J$ and $h_i^{\text{ext}}$.

In its original formulation, the Hopfield model assumed that the synaptic couplings are shaped according to the Hebb’s learning rule, reinforcing the connections inducing positive correlations among neurons and depressing the couplings leading to negative ones. Then, if the network is presented, during a training phase, with a number $M$ of binary input patterns $\xi^\mu$ (by setting in succession $h_i^{\text{ext}} = \lambda \xi_i^\mu$ for all $\mu$, with $\lambda \gg 1$ such that the external fields completely clamp the state of the network in Eq. (3) to $s^t = \xi^\mu$) the resulting couplings will have the symmetric form $J_{ij} \propto \sum_{\mu} \xi_i^\mu \xi_j^\mu$. It can then be shown that, for sufficiently high $\beta$ and sufficiently low $M$, there exist a regime in which this model behaves as an attractor network, meaning that initializing the state of the network in the proximity of any of the patterns, i.e. $s^0 = \xi^\mu + \nu$ with $\nu$ some binary noise that inverts some fraction of the units, the dynamics of the network, Eq. (3), spontaneously converges towards the basin of attraction of the memory to be stored, $s^t \sim \xi^\mu$ for $t \gg 1$. The amount of noise $\nu$ for which this is possible determines the size of the basin of attraction of the patterns. In this way, the input patterns $\{\xi^\mu\}$ are said to be stored as (associative) memories in the network, since they are encoded in the $J$ matrix and can be recovered from partial or corrupted inputs.

The Hopfield model with the Hebb’s learning rule has notoriously a number of problematic features [13], among which the most prominent are: a rather low capacity (the maximum $M$ scales as $0.14N$, while it is known that a network could reach $2N$ if $\beta \to \infty$ and if the symmetry constraint on $J$ is dropped); the existence of attractors of the dynamics that were not in $\{\xi^\mu\}$
(called “spurious” attractors); the phenomenon of *catastrophic forgetting* (when \( M \) exceeds 0.14N, the network enters a spin-glass phase in which all the \( \xi^\mu \) attractors are suddenly forgotten and an exponential number of spurious attractors are present instead); the inability to deal in general with correlated input patterns (although Hebb’s rule can be amended in some simpler cases, e.g. for biased patterns); the requirement of very strong external fields during the training phase (expressed as \( \lambda \gg 1 \) above), which is somewhat biologically implausible, while the learning rule requires that each unit should somehow be aware of both its current state (clamped to \( \xi^\mu \)) and the state that it would have in absence of its own external field, which is even more biologically problematic.

The choice of using Hebb’s rule was motivated by its simplicity and by biological considerations; moving away from this choice allows to amend many of the above-mentioned issues (see e.g. Refs. [14, 15]). Here, we extend the learning protocol proposed in Ref. [16], which specifically addressed the problem of learning with small external fields, to fully-connected networks.

As shown in more detail in the next section, the approach frames the learning problem as an inverse problem, in which the memories \( \{ \xi^\mu \} \) represent the given data \( D \), and we seek to find a coupling matrix \( J \) such that the memories would be typical samples from the distribution of Eq. (1), i.e. representatives of the basins of attraction of the final model.

### 3. Description of the learning protocol

Generalizing the work of Ref. [16], we study the problem of learning stable attractors by employing finite signals, in a fully-connected recurrent neural network. The inverse problem, i.e. the unsupervised learning task, is defined as follows: each pattern \( \xi^\mu \), \( \mu = 1, \ldots, \alpha N \), is presented to the network in the form of an external field of intensity \( \lambda \), oriented along the direction of the pattern itself, namely \( h^{\text{ext}} = \lambda \xi^\mu \). The relative balance between the intensities of the recurrent and the external contributions to the local field \( h_i = \sum_{j \neq i} J_{ij} s_j + h^{\text{ext}} \), can produce a dynamics which goes from a clamped regime, where the state of the network is set by the external field in a deterministic fashion, to a noise dominated regime, where the dynamics of the network becomes fully stochastic. The goal of the learning rule is that of adapting the coupling matrix \( J \) (i.e., the synaptic couplings), so that the driving effect of the external field is effectively absorbed in the recurrent interaction of the network. In this way, the freely evolving dynamics of the network is required to maximally resemble the driven dynamics.

This prescription naturally translates into a simple learning rule, inspired by the close connection between learning and inference in inverse problems:

\[
J_{ij}^{t+1} = J_{ij}^t + \eta \left( \langle s_i s_j \rangle_{\lambda_1} - \langle s_i s_j \rangle_{\lambda_2=0} \right)
\]  

(4)

where the parameter \( \eta \) plays the role of a learning rate.

The analogy with the Inverse Ising problem is immediately evident: the difference between the empirical correlations and the ones estimated from the partition function of the inferred model, is here replaced by the correlation difference between two models with the same synaptic couplings \( J \), but subject to different external field intensities, \( \lambda_1 \) and \( \lambda_2 \). However, we should note that, contrary to what happens in inference problems, the two correlations will never perfectly match.

As in Ref. [16], we note that in the limit of high temperatures and very strong biasing signals, i.e. \( \lambda_1 \to \infty \), we can recover the classic Hebb rule, since \( \langle s_i s_j \rangle_0 \simeq 0 \) and \( \Delta J_{ij} \propto \langle s_i s_j \rangle_\infty = \xi^\mu_i \xi^\mu_j \). At lower temperatures, in the same clamping limit, the learning protocol translates into the so-called Contrastive Divergence algorithm Ref. [17].

The computational problem associated with this learning strategy is that of the correlations estimation: in the following sections we propose a mean-field method developed in the Spin Glass theory framework.
Input: parameters: \( N \), \( p \), \( \eta \), \( T_L \), \( t_L \), \( \lambda_2^{\text{rate}} \)
Randomly initialize \( \xi^\mu \), \( \mu = 1, \ldots, p \) with \( \xi^\mu_l = \{-1, 1\} \);
Initialize \( J_s \) to zero;
Choose a field intensity \( \lambda_1 \);
Set a field intensity \( \lambda_2^{\text{init}} = \lambda_1 / 2 \);

\[
\text{for } T = 1 \text{ to } T_L \text{ do}
\]
\[
\quad \text{while } \lambda_2 > 0 \text{ do}
\]
\[
\quad \quad \lambda_2 \leftarrow \lambda_2 - \lambda_2^{\text{init}} / \lambda_2^{\text{rate}};
\]
\[
\quad \quad \text{for } \mu \text{ in random permutation of } p \text{ patterns do}
\]
\[
\quad \quad \quad \text{Initialize magnetization close to the chosen pattern } \mu;
\]
\[
\quad \quad \quad \text{Compute } <s_i s_j>_{\lambda_1};
\]
\[
\quad \quad \quad \text{for } t = 1 \text{ to } t_L \text{ do}
\]
\[
\quad \quad \quad \quad \text{ Compute } <s_i s_j>_{\lambda_2};
\]
\[
\quad \quad \quad \quad \lambda_3 \leftarrow J_s + \eta(<s_i s_j>_{\lambda_1} - <s_i s_j>_{\lambda_2});
\]
\[
\quad \quad \end{algorithm}
\]
\[
\end{algorithm}
\]

Algorithm 1: Learning

4. Correlations recording
In the previous section we have seen how the analogy between inference and learning leads to a new learning protocol, according to which the parameters of the network, namely the synaptic couplings \( J_{ij} \), have to be modulated in order to allow the storage of a set of \( p \) memories. The resulting update rule relies on the matching between the pair-wise correlations, estimated in presence of an external field of intensity \( \lambda_1 \), and the ones recorded in presence of a weaker external field \( \lambda_2 \), progressively vanishing.

The central issue is then how to compute the pair-wise correlations. In some contexts, more biologically inspired, the evolution in time of the network is simulated through the Glauber dynamics. The correlations among neuronal state variables are then estimated by means of the Monte Carlo method. However, it is known that this procedure becomes unfeasible for too large network sizes: the time needed for reaching a stationary state consistently increases when a bigger number of network components comes into play. We thus need to resort to more efficient methods.

In this work, we take advantage of the Belief Propagation algorithm (BP). This mean-field cavity method provides an estimation of the marginals associated to the variables that characterize a given physical system. Despite the fact that this estimation is exact only in tree-like graphical models, BP can be still successfully employed on loopy-graphs, when they can be locally approximated as trees.

By computing the (cavity) marginals, other interesting physical observables, such as the pair-wise correlations, can then be estimated. The BP algorithm accomplishes this task by defining a set of messages (also called cavity beliefs) running across the edges of the graph and corresponding to the marginals of a given variable in absence of a specific node of the graph (i.e. in presence of a “cavity”). The marginals can then be obtained as the product of the messages over all the cavity sites. BP provides a set of self-consistent equations that need to be solved iteratively, in order to find a closed expression for the cavity beliefs.

In the model of a fully-connected neural network, the BP equations of the messages are known, and given by:
The cavity message $\pi_{i\to j}(s_i)$ then determines the marginal of the $s_i$ variable in absence of the node $j$ in the network. In the thermodynamic limit, we can get the asymptotic expression:

$$\pi_{i\to j}(s_i) \simeq \prod_{k \neq j} \exp \left( \beta \frac{J_{ki}}{\sqrt{N}} s_i m_{k\to i} \right)$$

(6)

where we have defined $m_{i\to j}$ as the cavity magnetization:

$$m_{i\to j} = \sum_{s_i} s_i \pi_{i\to j}(s_i) = \tanh \left( \beta \left( \lambda \xi_i^\mu + \sum_{k \neq j} \frac{J_{ki}}{\sqrt{N}} m_{k\to i} \right) \right)$$

(7)

This self-consistent system of equations can be solved by iteration. Once a fixed point is reached, the pair-wise connected correlations can be recovered, on the basis of the cavity magnetization, as:

$$\langle s_i s_j \rangle = \tanh (\beta J_{ij}) \frac{1 - m_{j\to i}^2 (1 - m_{i\to j}^2)}{(1 + \tanh (\beta J_{ij}) m_{i\to j} m_{j\to i})^2} + m_i m_j$$

(8)

### 4.1. The TAP Approach

The BP algorithm is a very powerful tool. As we have seen, once the cavity messages are known, we are able to derive the main interesting observables characterizing a physical system, such as the pair-wise correlations.

However, one of the drawbacks of this method is that BP requires the storage of a number $O(N^2)$ of cavity messages, thus considerably affecting the computational performances. In order to overcome this difficulty, we can resort to the Thouless-Anderson-Palmer (TAP) approach. It can be derived through the high temperature expansion, or, equivalently, the limit of small synaptic couplings, of the mean field equations, on which the BP algorithm is based.

By performing the expansion, we get the TAP self-consistent equations, to be solved iteratively:

$$m_{i\to j}^{t+1} \simeq \tanh \left[ \beta \left( \lambda \xi_i^\mu + \sum_{k \neq i} m_k^t J_{ki} - \left( 1 - (m_k^t)^2 \right) m_{i\to j}^{t-1} J_{ki}^2 \right) \right]$$

(9)

where the time indices have been assigned so as to ensure the convergence of the iterative scheme [18]. We can notice that, instead of involving the cavity magnetizations, the TAP equations refer directly to the magnetizations, thus reducing the number of equations to be solved to $O(N)$.

In this framework, the pair-wise correlations can be computed as:

$$\langle s_i s_j \rangle = \frac{\beta J_{ij} + m_i m_j - (1 - m_i^2) \beta J_{ij} m_j^2 - (1 - m_j^2) \beta J_{ij} m_i^2}{1 + \beta J_{ij} m_i m_j}$$

(10)
Figure 1. Memory retrieval as a function of the external field intensity $\lambda_1$, in a network with $N = 201$ neurons. The purple and green curves show, respectively, the case of symmetric synapses, where the network learns by matching the equal time correlations, and the case of asymmetric synapses, where the time delayed correlations are instead employed. In both cases the correlations are obtained in the TAP approximation at a fixed inverse temperature $\beta = 2$. When running the simulations at a fixed number of learning cycles $T_L$, we can observe that learning with weaker fields is not only possible, but allows the network to reach higher storage loads. The points are obtained by averaging over 20 samples.

5. Asymmetric Synaptic Couplings
The Hebb’s learning rule is characterized by the unrealistic feature of symmetric synaptic couplings, namely $J_{ij} = J_{ji}$: there is no reason to assume that the signal sent by neuron $i$ towards neuron $j$ has to be of the same strength of the one sent by neuron $j$ towards neuron $i$. In fact, kinetically persistence states can be still observed in asymmetric networks.

In this section we thus consider a recurrent and fully connected neural network characterized by asymmetric synaptic couplings. The main difficulty in dealing with such a kinetic model is represented by the impossibility of defining a Hamiltonian. Therefore, in principle, we are not allowed to naively resort to the mean field approach. Here we follow the solution proposed in Ref. [19], where the authors show how the TAP equations can still be derived from a Kullback-Leibler divergence minimization even in the context of asymmetric synaptic couplings. We extend their results to the case of a parallel neuronal dynamics. The fixed point equations for the marginal probabilities of the neuronal states read:

$$m_i = \tanh \left( h_i + \sum_j J_{ij} m_j - m_i \sum_j (1 - m_j^2) J_{ij}^2 \right)$$  \hspace{1cm} (11)
The TAP approach can be thus exploited in the asymmetric context for the computation of the pair-wise correlations, as in [19]. However, in this specific case, we need to slightly modify the update rule of Eq. (4), in favor of the time-delayed correlations $\langle s_i^{t+1} s_j^t \rangle$:

$$
\langle s_i^{t+1} s_j^t \rangle = (1 - m_j^2) (1 - m_i^2) \left( J_{ij} + \sum_l J_{il} J_{jl} (1 - m_l^2) + 2m_i m_j J_{ij}^2 \right) + m_i m_j \quad (12)
$$

Indeed, by definition, the equal time correlations considered so far are not the proper quantity for learning in the asymmetric scenario.

Here, we present the results we obtained by applying the new learning protocol to a recurrent neural network, both in the symmetric and in the asymmetric cases. In the retrieval phase, we say that a pattern $\xi^\mu$ has been stored if the overlap of the stationary state, reached at convergence of the TAP equations, is consistently higher than all the other overlaps with all the other patterns $\xi^\nu$, with $\nu \neq \mu$.

In order to guarantee convergence of the TAP equations, instead of directly comparing the $\lambda_1$ intensity with $\lambda_2 = 0$, we slowly decrease the magnitude of the second external field, starting from an initial value of $\lambda_1/2$, as detailed in Algorithm 1.

Fig. 1 shows the number of patterns that have been retrieved by the network, as a function of the intensity of the reference external field $\lambda_1$. Notice that we work at storage loads that are higher than the critical capacity estimated for the Hebb’s rule. We can notice that, the network is able to learn even in the presence of weak external field.

6. Learning correlated patterns

As pointed out above, one of the major drawbacks of the classic Hebb rule is in the inability of embedding correlated memories. We test the learning protocol of Eq. (4) in this setting, introducing an inter-pattern correlation inspired to the hierarchical scheme as discussed e.g. in Ref [20]. In this scheme, the patterns are organized in $L$ categories, each defined by an independently generated random prototype. Each category contains $\alpha N/L$ patterns, obtained as independent perturbations of the corresponding prototype, with a fraction $1 - x$ of randomly flipped components. By increasing $x$, one can thus enhance the correlations in the dataset. The purple curve in Fig. 2 shows the maximum storage load $\alpha_c$ at which perfect memorization can be achieved, in the case of $N = 201$ and $L = 5$, as a function of the correlation $x$. The reference field intensity $\lambda_1$ was set to 0.3. As the learning problem becomes harder, the performance of the networks remains stable before suffering from a sudden drop, for values of $x > 0.5$.

In order to overcome this problem, we slightly modified the learning prescription, by updating the synapses as:

$$
J_{ij}^{t+1} = J_{ij}^t + \eta \left( \langle s_i s_j \rangle_{\lambda_1} - \langle s_i s_j \rangle_{\lambda_2 = -\lambda_1} \right) \quad (13)
$$

The main intuition behind this adjustment is the following: we are not only requiring the free dynamics of the network to resemble the biased dynamics, but we are also inducing some robustness into the memorization process. In fact, in this case, the network tries to match the initial correlations with the one measured in the presence of an antagonist external field, whose effect needs to be overcompensated by the recurrent couplings of the network. The green curve in Fig. 2 shows the results with the new update rule, in the same settings as above: by implementing this protocol, the network is indeed able to learn even in presence of stronger correlations.
Figure 2. Storage of correlated patterns, with a comparison between rules Eq. (4) and Eq. (13), in a symmetric network with $N = 201$ neurons. The purple curve shows the learning behavior when the weaker external field $\lambda_2$ tends to zero: after the pattern correlation exceeds a given threshold ($x = 0.5$), the ability of the network to store stable attractors sharply deteriorates. The green curve, instead, shows the case where $\lambda_2$ tends to a weak negative value ($-0.3$): the performance in the highly correlated regime is clearly improved. All the points are obtained by averaging over 20 samples.

Conclusions
We have shown how statistical physics methods can be used to infer couplings in systems characterized by multiple attractors, given relatively few samples. These results are of direct interest for computational neuroscience in that they lead to explicit learning protocols for weak realistic signals.

References
[1] Nguyen H C, Zecchina R and Berg J 2017 Advances in Physics 66 197–261
[2] Schneidman E, Berry M J, Segev R and Bialek W 2006 Nature 440 1007–1012
[3] Cocco S, Leibler S and Monasson R 2009 Proceedings of the National Academy of Sciences 106 14058–14062
[4] Morcos F, Pagnani A, Lunt B, Bertolino A, Marks D S, Sander C, Zecchina R, Onuchic J N, Hwa T and Weigt M 2011 Proceedings of the National Academy of Sciences 108 E1293–E1301
[5] Lezon T R, Banavar J R, Cieplak M, Maritan A and Fedoroff N V 2006 Proceedings of the National Academy of Sciences 103 19033–19038
[6] Miskin M Z, Khaira G, de Pablo J J and Jaeger H M 2016 Proceedings of the National Academy of Sciences 113 34–39
[7] Carleo G and Troyer M 2017 Science 355 602–606
[8] MacKay D J 2003 Information theory, inference and learning algorithms (Cambridge university press)
[9] Jaynes E T 1957 Physical review 106 620
[10] Jaynes E T 1986 Maximum-Entropy and Bayesian Methods in Applied Statistics (Cambridge university press)
[11] Mezard M and Montanari A 2009 Information, physics, and computation (Oxford University Press)
[12] Hopfield J J 1982 Proc Natl Acad Sci U S A 79 2554–2558 ISSN 0027-8424 6953413[pmid] URL http://www.ncbi.nlm.nih.gov/pmc/articles/PMC346238/
[13] Amit D J 1989 Modeling Brain Function: the World of Attractor Neural Networks (New York, NY, USA: Cambridge University Press) ISBN 0-521-36100-1
[14] Amit D J, Gutfreund H and Sompolinsky H 1987 Physical Review A 35 2293
[15] Alemi A, Baldassi C, Brunel N and Zecchina R 2015 PLOS Computational Biology 11 1–23 URL http://dx.doi.org/10.1371/journal.pcbi.1004439
[16] Braunstein A, Ramezanpour A, Zecchina R and Zhang P 2011 Physical Review E 83 056114 URL https://doi.org/10.1103/PhysRevE.83.056114
[17] Ackley D H, Hinton G E and Sejnowski T J 1985 Cognitive science 9 147–169
[18] Bolthausen E 2014 Communications in Mathematical Physics 325 333–366
[19] Kappen H J and Spaniers J J 2000 Phys. Rev. E 61(5) 5658–5663 URL http://link.aps.org/doi/10.1103/PhysRevE.61.5658
[20] Mezard M 2017 Phys. Rev. E 95(2) 022117 URL http://link.aps.org/doi/10.1103/PhysRevE.95.022117