

Beyond Inverse Ising Model: Structure of the Analytical Solution

Iacopo Mastromatteo

Received: 8 October 2012 / Accepted: 29 January 2013 / Published online: 13 February 2013
© Springer Science+Business Media New York 2013

Abstract I consider the problem of deriving couplings of a statistical model from measured correlations, a task which generalizes the well-known inverse Ising problem. After reminding that such problem can be mapped on the one of expressing the entropy of a system as a function of its corresponding observables, I show the conditions under which this can be done without resorting to iterative algorithms. I find that inverse problems are *local* (the inverse Fisher information is sparse) whenever the corresponding models have a factorized form, and the entropy can be split in a sum of small cluster contributions. I illustrate these ideas through two examples (the Ising model on a tree and the one-dimensional periodic chain with arbitrary order interaction) and support the results with numerical simulations. The extension of these methods to more general scenarios is finally discussed.

Keywords Statistical learning · Inverse Ising problem · Maximum entropy models · Complex systems

1 Introduction

The estimation of the direct interactions among the microscopic constituents of an extended system is a problem that has recently received considerable interest from the literature of several communities (biology [28, 33], genetics [4, 5], neuroscience [6, 25, 27], economy [15, 16, 20]). This is mainly due to the availability of large datasets across various fields, which have introduced the possibility to directly fit from data the interaction structure (e.g., the wiring pattern of neurons, the regulatory network of genes, the cross-influence of traders) of a system which is operating according to an unknown, or a partially known, mechanism. Despite this abundance of data, the problem of estimating the most appropriate model in order to describe a system is far from being solved: even if the maximum entropy principle (MEP) is invoked in order deal with the issue of model selection [12, 13], one is typically left with a difficult inference problem. In particular fitting a model under the MEP requires

I. Mastromatteo (✉)
Capital Fund Management, Rue de l'Université 23, 75007, Paris, France
e-mail: Iacopo.Mastromatteo@cfm.fr

matching a set of observables estimated from data with the ones predicted by the model (the so-called *inverse problem*), a task which is known to be generically hard [31]. While the *Boltzmann learning* algorithm can be used to obtain accurate solutions for the inverse problem only by investing a large amount of computational power [1], mean-field techniques [14, 22–24, 26, 29], message-passing methods [3, 11, 17, 19] and pseudo-likelihood approximations [2, 32] have proved to be very effective in order to fit approximately the interactions of the system in a short amount of time.

In this work I show that for a specific class of models the inverse problem can be solved exactly and efficiently. In particular, if the probability distribution of the system has a factorized form, the entropy of the system can be written as a sum of cluster contributions which are easy to manipulate analytically. This is consistent with the results of [7, 8], who show that efficient, fast inference procedures can be constructed if the entropy of the system can be approximated by the sum of a suitable set of small cluster contributions. For these systems one can see that the robustness of the inverse problem is linked with the factorization property of the probability distribution (equivalently, the additivity of the entropy). This corroborates the general belief that inverse problems are more stable with respect to the direct ones, as the response of couplings due to a change in the observables is expected to be much more localized than the response of the observables due to a shift of the couplings. This property has been observed in the context of the inverse Ising problem [8], but it is not clear to which classes of problems it might be generalized to [30]. Additionally, finite sampling effects can in principle spoil this feature, hence it is relevant to understand how the structure of the Fisher information matrix behaves when sampling noise is added (see for example formulas (18) and (27)). In the following section I introduce a very generic setup for the inverse problem, in order to stress the generality of the results. In Sect. 2 its exact solution is presented for two specific cases. In Sect. 3 the results are scrutinized against numerical simulations involving synthetic data, and in Sect. 4 the method introduced is extended to more general cases. In Sect. 5 some conclusions are drawn.

I consider the problem of estimating a set of M couplings $g = (g_1, \dots, g_M)$ describing a probability density for a set of N binary spins $\underline{s} = (s_1, \dots, s_N) \in \{-1, 1\}^N$ of the form

$$p(\underline{s}) = \frac{1}{Z(g)} \exp\left(\sum_{\mu=1}^M g_{\mu} \phi_{\mu}(\underline{s})\right), \quad (1)$$

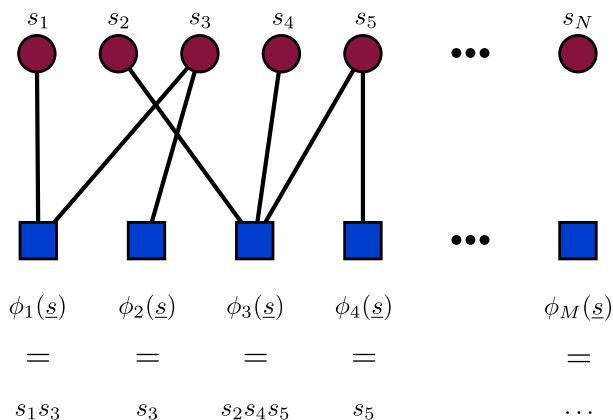
where $\phi = (\phi_1, \dots, \phi_M)$ is a set of known functions of s (usually referred as *potentials*, or *sufficient statistics*), and the *partition function* $Z(g)$ enforces the normalization of $p(\underline{s})$. In the following discussion we will assume all of the $\phi_{\mu}(\underline{s})$ to be monomials, i.e., to be functions of the form

$$\phi_{\mu}(\underline{s}) = \prod_{i \in \Gamma_{\mu}} s_i, \quad (2)$$

where Γ_{μ} is a given subset (cluster) of the *vertex* set $V = \{1, \dots, N\}$ associated with the operator μ . Within this general formulation, the Ising (or *graphical*) model can for example be constructed by using single-body clusters $\{i\}_{i=1}^N$ and two-body clusters $\{i, j\}_{i < j=1}^N$, together with their associated couplings usually denoted with $\{h_i\}_{i=1}^N$ and $\{J_{ij}\}_{i < j=1}^N$. Figure 1 shows factor graph associated with the probability density (1) in a generic case. In order to estimate such model, a set of T i.i.d. observations $\hat{s} = \{\underline{s}^{(t)}\}_{t=1}^T$ of the system is provided, so that the log-likelihood function for the model can be written as

$$\log L(\hat{s}|g) = T \left(-\log Z(g) + \sum_{\mu=1}^M g_{\mu} \bar{\phi}_{\mu} \right) \quad (3)$$

Fig. 1 Example of a typical factor graph associated with the probability density (1)



and depends on the data through the empirical averages $\bar{\phi}_\mu = \frac{1}{T} \sum_{t=1}^T \phi_\mu(\underline{s}^{(t)})$. Maximizing such function in order to find the coupling set g^* best describing the data \hat{s} in absence of prior leads to the set of M conditions

$$\langle \phi_\mu \rangle_g = \bar{\phi}_\mu, \quad (4)$$

where the quantities $\langle \phi_\mu \rangle_g = \partial \log Z(g) / \partial g_\mu$ describe the averages of the functions $\phi_\mu(\underline{s})$ under the model parametrized by g . Thus, in order to find the best model in order to describe data \hat{s} one has to match empirical averages $\bar{\phi}_\mu$ with ensemble averages $\langle \phi_\mu \rangle_{g^*}$. An equivalent characterization of the problem of finding the optimal g^* is provided by computing the Legendre transform

$$-S(\bar{\phi}) = \max_g \left(\sum_{\mu=1}^M g_\mu \bar{\phi}_\mu - \log Z(g) \right), \quad (5)$$

which can be identified as the Shannon entropy of the optimal model expressed as a function of the empirical averages $\bar{\phi}$. Such quantity is the cumulant generating function for the maximum likelihood estimator of the couplings $g^* = g^*(\bar{\phi})$. From this perspective, the entropy $S(\bar{\phi})$ plays a role similar to the one of the free-energy $F(g) = -\log Z(g)$ in a statistical mechanics problem (i.e., estimating the ensemble averages $\langle \phi_\mu \rangle_g$ given the couplings g). In fact, $F(g)$ satisfies the relations

$$\langle \phi_\mu \rangle_g = -\frac{\partial F(g)}{\partial g_\mu}, \quad (6)$$

$$\chi_{\mu,v}(g) = -\frac{\partial^2 F(g)}{\partial g_\mu \partial g_v} \quad (7)$$

being $\chi_{\mu,v}(g)$ the covariance matrix $\langle \phi_\mu \phi_v \rangle_g - \langle \phi_\mu \rangle_g \langle \phi_v \rangle_g$. Such matrix $\chi_{\mu,v}$ is a central object in the field of information theory, where it is customarily referred as *Fisher information* [9, 18]. For $S(\bar{\phi})$ it holds

$$g_\mu^*(\bar{\phi}) = -\frac{\partial S(\bar{\phi})}{\partial \bar{\phi}_\mu}, \quad (8)$$

$$\chi_{\mu,v}^{-1}(\bar{\phi}) = -\frac{\partial^2 S(\bar{\phi})}{\partial \bar{\phi}_\mu \partial \bar{\phi}_v}. \quad (9)$$

Just as the quantity $\chi_{\mu,v}(g)$ can be used to express how fast in T the empirical averages $\bar{\phi}$ converge to the ensemble values $\langle \phi \rangle$, its matrix inverse expresses the rate of convergence of the inferred couplings $g^*(\bar{\phi})$ to the asymptotic values $g^*(\langle \phi \rangle)$. In particular the covariance matrix for the estimator $g^*(\bar{\phi})$ is given by

$$\frac{\chi_{\mu,v}^{-1}}{T} = \left((g_{\mu}^*(\bar{\phi}) - g_{\mu}^*(\langle \phi \rangle)) (g_v^*(\bar{\phi}) - g_v^*(\langle \phi \rangle)) \right)_T - \left(g_{\mu}^*(\bar{\phi}) - g_{\mu}^*(\langle \phi \rangle) \right)_T \left(g_v^*(\bar{\phi}) - g_v^*(\langle \phi \rangle) \right)_T, \quad (10)$$

where $\langle \cdot \rangle_T = \sum_{\underline{s}} \prod_{i=1}^T p(\underline{s}^{(i)})$. Hence the matrix $\chi^{-1}(\bar{\phi})$ expresses the stability of the inferred couplings with respect to the statistical error due to finite sampling.

I will show in the following section how Eqs. (8) and (9) can be used to determine analytically $g^*(\bar{\phi})$. In particular, if it is possible to compactly express the entropy of a model as a function of the empirical averages $\bar{\phi}$, the maximum likelihood estimator $g^*(\bar{\phi})$ can be found easily.

2 Two Factorizable Models

In this section I present two models for which the inverse problem can be exactly solved. The solution strategy can be straightforwardly extended to more general cases.

In order to discuss this approach to the inverse problem, it is nevertheless necessary to introduce some extra notation. In particular, we will call *cluster* any subset $\Gamma \subseteq V$, and denote the set of spins contained in Γ with $\underline{s}_{\Gamma} = (s_i)_{i \in \Gamma}$. Given a cluster Γ , its associated *marginal* p_{Γ} is then defined as the probability distribution on \underline{s}_{Γ}

$$p_{\Gamma}(\underline{s}_{\Gamma}) = \sum_{s_i | i \notin \Gamma} p(\underline{s}). \quad (11)$$

Finally, the *cluster entropy* S_{Γ} is defined as

$$S_{\Gamma}(p_{\Gamma}) = - \sum_{\underline{s}_{\Gamma}} p_{\Gamma}(\underline{s}_{\Gamma}) \log p_{\Gamma}(\underline{s}_{\Gamma}). \quad (12)$$

It is easy to prove that marginals and cluster entropies can be exactly expressed as functions of the averages of all the possible monomials contained in the cluster Γ . More precisely, given the set of $2^{|\Gamma|}$ monomials $\{\hat{\phi}_{\Gamma'}\}_{\Gamma' \subseteq \Gamma}$ associated with the clusters $\Gamma' \subseteq \Gamma$, it is easy to show that

$$p_{\Gamma}(\underline{s}_{\Gamma}) = \frac{1}{2^{|\Gamma|}} \sum_{\Gamma' \subseteq \Gamma} \hat{\phi}_{\Gamma'}(\underline{s}_{\Gamma'}) \langle \hat{\phi}_{\Gamma'} \rangle_g. \quad (13)$$

This formula characterizes the marginals $p_{\Gamma}(\underline{s}_{\Gamma})$ as *local* objects, as they are uniquely determined by the averages of potentials contained inside Γ .

2.1 Tree-Like Graphs

Consider a model of the form

$$p(\underline{s}) = \frac{1}{Z(h, J)} \exp \left(\sum_{i \in V} h_i s_i + \sum_{(i,j) \in E} J_{ij} s_i s_j \right), \quad (14)$$

where $E = \{(i, j) \in V \times V \mid i < j\}$ is a given set of *edges*. Suppose additionally that the graph associated with the edge set E is acyclic, so that the model (14) corresponds to an Ising model on a tree (more precisely, a forest). In that case, it can be shown by induction (see [18, 31]) that the total probability density $p(\underline{s})$ can be factorized according to

$$p(\underline{s}) = \prod_{(i,j) \in E} p_{[i,j]}(s_i, s_j) \prod_{i \in V} [p_{\{i\}}(s_i)]^{1-|\partial i|}, \quad (15)$$

where $\partial i = \{(j, k) \in E \mid i \in (j, k)\}$. Then, the entropy of the model is additive and can be written as

$$S(p) = \sum_{(i,j) \in E} S_{[i,j]}(m_i, m_j, c_{ij}) + \sum_{i \in V} (1 - |\partial i|) S_{\{i\}}(m_i), \quad (16)$$

where one has defined the *magnetizations* $m_i = \frac{1}{T} \sum_{t=1}^T s_i^{(t)}$ and the *correlations* $c_{ij} = \frac{1}{T} \sum_{t=1}^T s_i^{(t)} s_j^{(t)}$. After the use of formula (13), one finds by differentiation that the solution of the inverse Ising model on an acyclic graph is given by

$$\begin{aligned} h_i^* &= \frac{1}{4} \sum_{j \in \partial i} \sum_{s_i, s_j} s_i \log \left[\frac{1}{4} (1 + m_i s_i + m_j s_j + c_{ij} s_i s_j) \right] \\ &\quad + \frac{1}{2} (1 - |\partial i|) \sum_{s_i} s_i \log \left[\frac{1}{2} (1 + m_i s_i) \right], \\ J_{ij}^* &= \frac{1}{4} \sum_{s_i, s_j} s_i s_j \log \left[\frac{1}{4} (1 + m_i s_i + m_j s_j + c_{ij} s_i s_j) \right], \end{aligned} \quad (17)$$

and the stability of the solution is determined by

$$\begin{aligned} \chi_{[i,j],\{k,l\}}^{-1} &= \frac{1}{16} \sum_{s_i, s_j} \frac{\delta_{i,k} \delta_{j,l} + \delta_{i,l} \delta_{j,k}}{\bar{p}_{[i,j]}(s_i, s_j)}, \\ \chi_{[i,j],\{k\}}^{-1} &= \frac{1}{16} \sum_{s_i, s_j} \frac{\delta_{i,k} s_j + \delta_{j,k} s_i}{\bar{p}_{[i,j]}(s_i, s_j)}, \\ \chi_{[i],\{j\}}^{-1} &= \frac{1}{16} \sum_{k \in \partial i} \sum_{s_i, s_k} \frac{\delta_{i,j} + s_i s_k \delta_{k,j}}{\bar{p}_{[i,k]}(s_i, s_k)} + \frac{1}{4} (1 - |\partial i|) \sum_{s_i} \frac{\delta_{i,j}}{\bar{p}_{\{i\}}(s_i)}. \end{aligned} \quad (18)$$

The quantities

$$\bar{p}_{\{i\}}(s_i) = \frac{1}{2} (1 + m_i s_i), \quad (19)$$

$$\bar{p}_{[i,j]}(s_i, s_j) = \frac{1}{4} (1 + m_i s_i + m_j s_j + c_{ij} s_i s_j) \quad (20)$$

appearing in the previous expression describe the empirical frequencies associated with the clusters $\{i\}$ and $\{i, j\}$.

Notice that the problem of finding any of the mean magnetizations $\langle s_i \rangle_{(h,J)}$ or a correlations $\langle s_i s_j \rangle_{(h,J)}$ usually requires the use of iterative algorithms such as belief propagation, which converge in a number of steps linear with the number of vertices $|V|$ [18]. For the inverse problem no iterative algorithm is required, as the solution can be found by simply evaluating formula (17). This is a consequence of the fact that—as observed in [7, 8]—inverse problems tend to be more local than their direct counterparts. Equivalently, due to sparsity of χ^{-1} , by shifting an empirical average (either m_i or J_{ij}) just neighboring couplings are changed, whereas for the direct problem the change of a coupling (say, h_i), has an

effect on a finite number of vertices (being the size of the perturbation roughly determined by the correlation length of the system). These results are consistent with the ones of [21, 22, 34], where the solution of the inverse Ising problem under the Bethe approximation is derived. In particular, Eq. (17) can be recovered by specializing the results in [21, 22, 34] to the case of a tree, for which the Bethe approximation is exact.

This solution requires indeed the knowledge of the full topology specified by E , so that this approach cannot be used to determine the graph structure from data. Rather, it allows to exactly recover the coupling strength given an acyclic set of edges E .

2.2 One-Dimensional Models

Within this approach it is also possible to find the exact solution of the inverse problem for a one-dimensional periodic chain with interactions of arbitrary order. Even if a heuristic solution for this model was first proposed in [10], its detailed derivation is provided in the following. The model is defined by a probability distribution $p(\underline{s})$ on a set of binary spins $\underline{s} \in \{-1, 1\}^N$ specified by a family of potentials acting on the first R spins $\phi(s_1, \dots, s_R) = (\phi_1(s_1, \dots, s_R), \dots, \phi_M(s_1, \dots, s_R))$ subject to the periodic boundary conditions $s_i = s_{i+N}$. These potentials are meant to model spin-spin interactions of range up to R . A periodicity parameter $\rho \leq R$ is also introduced, in order to enforce the invariance of the system under translations by multiples of ρ . This is formally achieved by using *translation operators* $T = \{T_n\}_{n=0}^{N/\rho-1}$, defined through their action on the ϕ

$$T_n \phi_\mu(s_1, \dots, s_R) = \phi_\mu(s_{1+n\rho}, \dots, s_{R+n\rho}). \quad (21)$$

Finally one can define a *one-dimensional periodic chain* as the probability distribution

$$p(\underline{s}) = \frac{1}{Z(g)} \exp \left(\sum_{\mu=1}^M g_\mu \sum_{n=0}^{N/\rho-1} T_n \phi_\mu(\underline{s}) \right). \quad (22)$$

Suppose that the family $\phi = (\phi_1, \dots, \phi_M)$ contains all and only the $2^R(1 - 2^{-\rho})$ monomials describing the unit cell (i.e., any potential of range smaller than R along the chain can be uniquely obtained by translating one of the ϕ_μ by mean of a T_n operator). Then one can exploit the factorization

$$p(\underline{s}) = \prod_n \frac{p_{\Gamma_n}(\underline{s}_{\Gamma_n})}{p_{\gamma_n}(\underline{s}_{\gamma_n})}, \quad (23)$$

where one has defined the sets $\Gamma_n = \{n\rho + 1, \dots, n\rho + R\}$ and $\gamma_n = \{(n+1)\rho + 1, \dots, n\rho + R\}$. This leads to the additivity formula for the entropy

$$S(p) = \sum_n [S_{\Gamma_n}(p_{\Gamma_n}) - S_{\gamma_n}(p_{\gamma_n})] \quad (24)$$

proved in the [Appendix](#). Translational symmetry and formula (13) lead finally to

$$\begin{aligned} S(T\bar{\phi}) &= \frac{N}{\rho} \left\{ \sum_{\underline{s} \in I_0} \left[\frac{1}{2^R} \sum_{\Gamma' \subseteq I_0} \sum_{\mu=1}^M c_{\mu, \Gamma'} \bar{\phi}_\mu \hat{\phi}_{\Gamma'}(\underline{s}_{\Gamma'}) \right] \right. \\ &\quad \times \log \left[\frac{1}{2^R} \sum_{\Gamma' \subseteq I_0} \sum_{\mu=1}^M c_{\mu, \Gamma'} \bar{\phi}_\mu \hat{\phi}_{\Gamma'}(\underline{s}_{\Gamma'}) \right] \end{aligned}$$

$$\begin{aligned}
& - \sum_{\underline{s}_{\gamma_0}} \left[\frac{1}{2^{R-\rho}} \sum_{\gamma' \subseteq \gamma_0} \sum_{\mu=1}^M c_{\mu, \gamma'} \bar{\phi}_{\mu} \hat{\phi}_{\gamma'}(\underline{s}_{\gamma'}) \right] \\
& \times \log \left[\frac{1}{2^{R-\rho}} \sum_{\gamma' \subseteq \gamma_0} \sum_{\mu=1}^M c_{\mu, \gamma'} \bar{\phi}_{\mu} \hat{\phi}_{\gamma'}(\underline{s}_{\gamma'}) \right] \Bigg\}, \quad (25)
\end{aligned}$$

where the $\hat{\phi}_{\Gamma}(\underline{s}_{\Gamma})$ are monomials used to expand the marginals as in (13), while $c_{\mu, \Gamma} = 1$ if it exists an n such that $T_n \phi_{\mu} = \hat{\phi}_{\Gamma}$ and $c_{\mu, \Gamma} = 0$ otherwise. The couplings are given by

$$\begin{aligned}
g_{\mu}^* &= \sum_{\underline{s}_{\Gamma_0}} \left[\frac{1}{2^R} \sum_{\Gamma' \subseteq \Gamma_0} c_{\mu, \Gamma'} \hat{\phi}_{\Gamma'}(\underline{s}_{\Gamma'}) \right] \log \left[\frac{1}{2^R} \sum_{\Gamma' \subseteq \Gamma_0} \sum_{v=1}^M c_{v, \Gamma'} \bar{\phi}_v \hat{\phi}_{\Gamma'}(\underline{s}_{\Gamma'}) \right] \\
& - \sum_{\underline{s}_{\gamma_0}} \left[\frac{1}{2^{R-\rho}} \sum_{\gamma' \subseteq \gamma_0} c_{\mu, \gamma'} \hat{\phi}_{\gamma'}(\underline{s}_{\gamma'}) \right] \log \left[\frac{1}{2^{R-\rho}} \sum_{\gamma' \subseteq \gamma_0} \sum_{v=1}^M c_{v, \gamma'} \bar{\phi}_v \hat{\phi}_{\gamma'}(\underline{s}_{\gamma'}) \right] \quad (26)
\end{aligned}$$

while the inverse susceptibilities can be written as

$$\begin{aligned}
\chi_{\mu, v}^{-1} &= \frac{\rho}{N} \left\{ \sum_{\underline{s}_{\Gamma_0}} \frac{[\frac{1}{2^R} \sum_{\Gamma' \subseteq \Gamma_0} c_{\mu, \Gamma'} \hat{\phi}_{\Gamma'}(\underline{s}_{\Gamma'})][\frac{1}{2^R} \sum_{\Gamma' \subseteq \Gamma_0} c_{v, \Gamma'} \hat{\phi}_{\Gamma'}(\underline{s}_{\Gamma'})]}{\bar{p}_{\Gamma_0}(T\bar{\phi})} \right. \\
& \left. - \sum_{\underline{s}_{\gamma_0}} \frac{[\frac{1}{2^{R-\rho}} \sum_{\gamma' \subseteq \gamma_0} c_{\mu, \gamma'} \hat{\phi}_{\gamma'}(\underline{s}_{\gamma'})][\frac{1}{2^{R-\rho}} \sum_{\gamma' \subseteq \gamma_0} c_{v, \gamma'} \hat{\phi}_{\gamma'}(\underline{s}_{\gamma'})]}{\bar{p}_{\gamma_0}(T\bar{\phi})} \right\}. \quad (27)
\end{aligned}$$

The structure of this solution is analogous to the one of (17), in the sense that once that the $S(p)$ is written as a sum of cluster entropies $S_{\Gamma_n}(p_{\Gamma_n})$ and $S_{\gamma_n}(p_{\gamma_n})$, the expression for the couplings is obtained by summing their separate contributions to the estimator $g_{\mu}^*(T\phi)$. The main difference in this case is that one can exploit translational symmetry in order to equate the cluster contributions due to different unit cells. Thus, the number of sampled unit cells is $\sqrt{TN/\rho}$, rather than \sqrt{T} . This rules the convergence of the couplings to their asymptotic values: if the system is large translational symmetry allows to obtain reliable estimations of the couplings even with a small number of observations.

3 Applications

In order to validate the results shown above, some examples involving synthetic datasets are discussed. In the case of a tree-like model, a set of $T = 10^6$ configurations of a systems of size $N = 50$ has been simulated. The configurations have been obtained by Monte Carlo sampling from a distribution $p(\underline{s})$ specified by a tree-like graph E , while the couplings J and h have been uniformly drawn from the interval $[0, 1]$. The node degrees have been extracted from a distribution of fixed mean $\langle k_i \rangle = \langle |\partial i| \rangle = 2(N-1)/N$. The variances of the inferred couplings for specific instances of the problem have been shown in Fig. 2, where they are compared against their expected scaling. The plot indicates that formula (17) correctly predicts the value of the inferred couplings and of their fluctuations. I have also considered the problem in which every coupling is multiplied by an *inverse temperature* parameter β , in order to control the overall intensity of the fluctuations. The presence of strong noise corresponds to the regime of large β , while the one in which noise is suppressed is associated with the small β region. In this example the couplings J and h have been uniformly drawn in $[0, 1]$ and multiplied by a β parameter in the range $[0.01, 10]$. Subsequently, $T = 10^5$

Fig. 2 Variance of the inferred couplings h (red line) and J (blue line) against the number of samples T for a pairwise tree. The dashed lines indicate the values of the error calculated by using formula (18) (Color figure online)

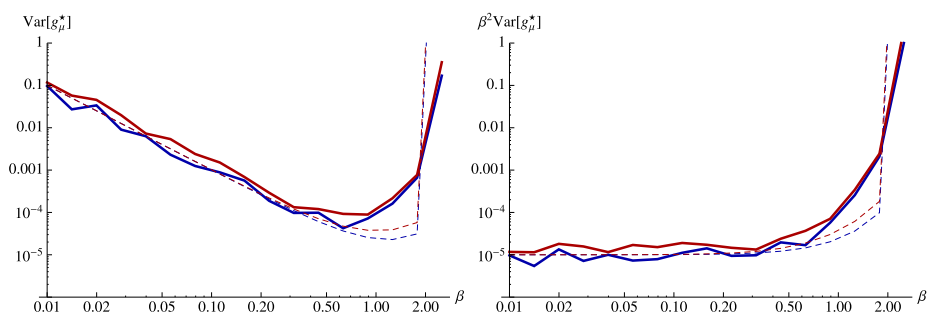
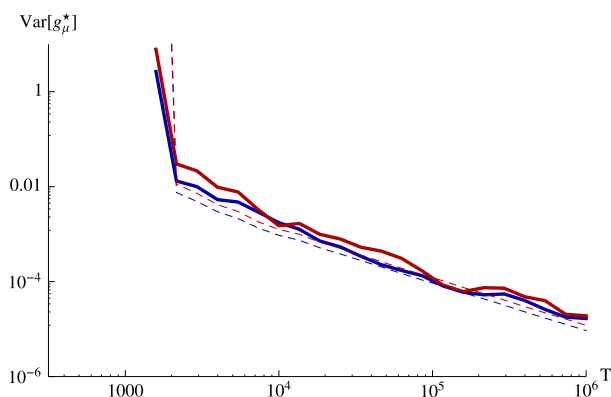


Fig. 3 Variance of the inferred couplings h^* (red line) and J^* (blue line) against the inverse temperature β for a pairwise tree, obtained by using $T = 10^5$ Monte Carlo samples. Both the variances of h^* and J^* (left panel) and the ones of the products $h^*\beta$ and $J^*\beta$ (right panel) are plotted. The dashed lines indicate the estimations of the error obtained by using expression (18) (Color figure online)

samples have been extracted via Monte Carlo. In Fig. 3 the variance of the inferred couplings is plotted against the inverse temperature β . One can notice from the plots that, as the inverse problem depends just upon the products βJ and βh , it is not possible to distinguish between an overall strength of the fluctuations at the temperature parameter β . Hence, while the lowest inference error for the products βJ and βh is obtained for $h_i = J_{ij} = 0$, the lowest error for the values J and h (given a fixed, known value for β) is obtained by finding a compromise between maximum signal (favoring high couplings) and minimum noise (favoring high temperature, or equivalently low β). One also finds that the quality of the reconstruction degrades progressively in the large β regime by increasing the inverse temperature, until eventually some of the two-state marginals are not sufficiently sampled, in which case the inferred couplings diverge. The same type of study has been carried on for the model (22). As an example, I show the results obtained for a system of size $N = 50$, characterized by interactions of range $R = 4$ and periodicity parameter $\rho = 2$. A number of configurations $T = 10^6$ have been simulated via Monte Carlo for a system with random interactions (g_μ uniformly extracted in $[0, 0.01]$). The results are shown in Fig. 4, where the variance of the inferred couplings is plotted against the expected scaling. Also for the one dimensional case I analyzed the effect of an inverse temperature parameter $\beta \in [0.01, 1]$ modulating an interaction strength g_μ randomly drawn from the interval $[0, 1]$. Figure 5 shows the errors in the reconstructed couplings as a function of the β parameter. Notice that

Fig. 4 Variance of the inferred coupling vector g^* (solid line) plotted against the number of sampled unit cells NT/ρ , obtained by Monte Carlo sampling of a model describing a complete one-dimensional periodic chain of size $N = 50$, range $R = 4$ and periodicity $\rho = 2$. The dashed line shows the error predicted by Eq. (27)

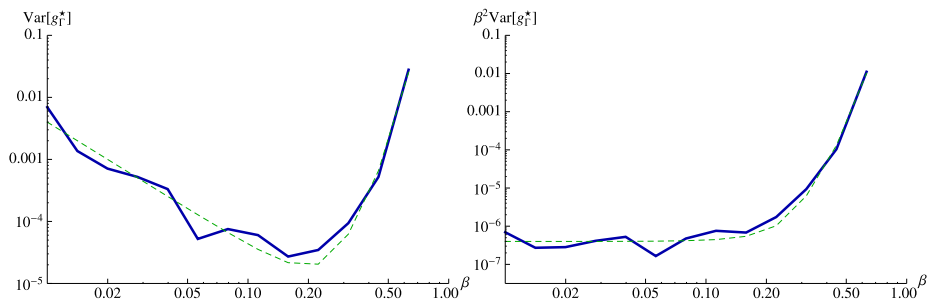
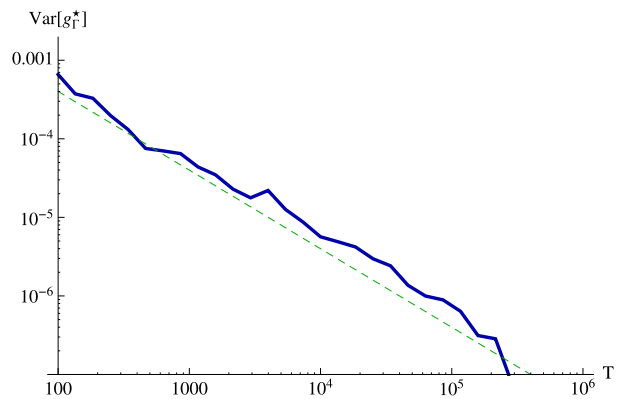


Fig. 5 Variance of the inferred couplings g^* (solid line) against the inverse temperature β for a one dimensional periodic chain. $T = 10^5$ configurations have been sampled via Monte Carlo so to construct the empirical averages $\hat{\phi}$. Both the results for the inferred couplings g^* (left panel), and the one for the products βg^* (right panel) are shown. The dashed line indicates the error obtained by using Eq. (27)

the error on βg^* is minimum for a flat distribution ($\beta = 0$), while the one on g^* is minimized by an intermediate value of β . The quality of the reconstruction is determined by how uniformly is the R -spin marginal sampled. If local fluctuations are not sampled well-enough, the error on the inferred couplings is large as predicted by Eq. (27). As observed above, it is not necessary to probe global fluctuations of the system in order to accurately reconstruct the couplings.

4 General Structure of Exact Solutions

What has been shown for models (14) and (22) can be easily extended to more general scenarios. In particular any factorization property for the probability distribution $p(\underline{s})$ analogous to (15) and (23) breaks the entropy in a set of cluster contributions of the form

$$S(p) = \sum_{\Gamma \in \mathcal{G}} c_{\Gamma} S_{\Gamma}(p_{\Gamma}), \quad (28)$$

where \mathcal{G} is a suitably chosen collection of clusters associated with a set of $\{c_{\Gamma}\}_{\Gamma \in \mathcal{G}}$ coefficients. Equation (13) implies that cluster entropies $S_{\Gamma}(p_{\Gamma})$ are local in the empirical observables associated with Γ . Hence, if the model is large enough (i.e., if ϕ includes all

the monomials located within the clusters $\Gamma \in \mathcal{G}$), the cluster entropies can be expressed as functions of the observables $\bar{\phi}$ conjugated to the couplings g . In this case the following facts hold:

- Due to linearity of the derivative any inferred coupling $g_\mu^*(\bar{\phi})$ receives contributions just by clusters containing its conjugated potential $\phi_\mu(\underline{s})$.
- The matrix χ^{-1} is sparse, so that by shifting any of the empirical observables $\bar{\phi}_\mu$ the only clusters which are affected are the ones including variables contained inside the monomial $\phi_\mu(\underline{s})$.

The expressions (18) and (27) for the stability matrix χ^{-1} have then a clear interpretation in term of cluster contributions. Specifically, the calculation of the Fisher information requires a summation over all the possible micro-states s_Γ for a set of relevant clusters Γ (a cluster is relevant to the term $\chi_{\mu,\nu}^{-1}$ if $S_\Gamma(p_\Gamma)$ depends both on $\bar{\phi}_\mu$ and $\bar{\phi}_\nu$). Each of the terms in the sum is weighted according to the inverse empirical marginal $\bar{p}_\Gamma(\bar{\phi}_\Gamma)$ estimated from data. If the clusters necessary to describe $p(\underline{s})$ are uniformly sampled (i.e., the empirical frequencies \bar{p}_Γ are approximately flat), the error on the inferred couplings is small. Conversely, as soon as one or more configurations \underline{s}_Γ are not sampled, the asymptotic error diverges. This is consistent with the observation that if a $p_\Gamma(\underline{s}_\Gamma)$ assigns zero mass to some configurations, infinite couplings are required to describe the marginals (in that case either some regularization can be used to enforce finiteness of the solution, or infinite solutions may be accepted as describing a hard constraint on the set of accessible configurations). As a further comment, notice that a reliable estimation of the couplings g doesn't require sampling global fluctuations in the configuration space: it is simply necessary to sample fluctuations which are local with respect to clusters in \mathcal{G} in order to accurately estimate the couplings g .

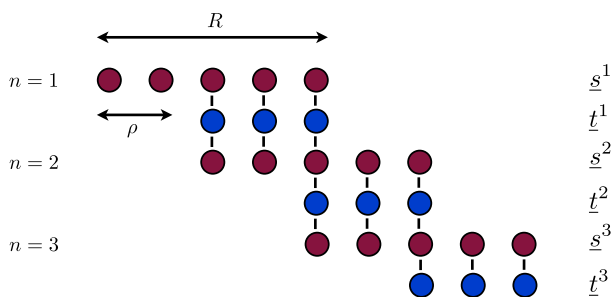
These ideas can be useful even in the case in which the entropy doesn't strictly break in the sum of small cluster terms. In particular it has been observed in [7, 8] that for several models it is possible to account for a large fraction of the entropy $S(p)$ by using just a few local terms (corresponding to the remark that in many cases the matrix χ^{-1} is approximately sparse). In this case fast, although approximate, inference schemes could be obtained by using these techniques.

Notice that the price to pay in order to analytically solve this inverse problem is the introduction of a potentially high number of parameters: a coefficient has to be fitted for each of the monomials included in any of the clusters $\Gamma \in \mathcal{G}$. In particular, the number of parameters to be considered is bound by $M \leq |\mathcal{G}| \times 2^{\max_{\Gamma \in \mathcal{G}} |\Gamma|}$. Thus, even if a number of clusters polynomial in N is usually sufficient to properly describe the entropy of a system, their size enters exponentially in the number of fitted parameters.

5 Conclusions

The task of determining a probability distribution whose associated momenta match a given set of values is a known hard problem, with relevant implications in the field of high-dimensional inference. In this work I have shown that its exact solution can be found whenever the probability density for the system has a factorized structure. The sparsity of the Fisher information matrix $\chi^{-1}(\bar{\phi})$ and the additivity of the entropy can be shown to rely upon this independence property. These result is completely general (in particular, it is independent of the interaction structure of the system). In order to provide illustrative examples, these ideas have been used to solve two specific inverse problems: the inverse Ising model

Fig. 6 Two dimensional auxiliary model $p^\lambda(s)$ associated with the original distribution $p(s)$ describing a one-dimensional periodic chain



on a tree and the one-dimensional periodic chain with arbitrary order interactions. The interactions of both systems can be accurately reconstructed if the clusters describing the system are uniformly sampled, while the inferred couplings diverge as soon as some relevant micro states are unobserved. These findings have been confirmed by numerical simulations, which show an excellent agreement with the analytical predictions. Techniques for the approximate solution of generic inverse problems could in principle be derived by developing the methods presented in this work.

Acknowledgements I acknowledge M. Marsili, G. Gori and S. Cocco for very useful discussions.

Appendix: Factorization Property for the One-Dimensional Chain

I show in the following that for a one-dimensional periodic chain defined as in (22), the factorization property

$$p(\underline{s}) = \prod_{n=0}^{N/\rho-1} \frac{P_{\Gamma_n}(\underline{s}_{\Gamma_n})}{P_{\gamma_n}(\underline{s}_{\gamma_n})} \quad (29)$$

holds, where $\Gamma_n = \{n\rho + 1, \dots, n\rho + R\}$ and $\gamma_n = \{(n+1)\rho + 1, \dots, n\rho + R\}$. To obtain this result, one can consider the auxiliary two-dimensional model defined by the log-probability

$$\begin{aligned} \log p^\lambda(\underline{s}, \underline{t}) = & -\log Z(g) + \sum_{n=0}^{N/\rho-1} \sum_{\mu=1}^M g_\mu \phi_\mu(s_{1+n\rho}^n, \dots, s_{R+n\rho}^n) \\ & + \lambda \sum_{n=0}^{N/\rho-1} \sum_{i=(n+1)\rho+1}^{n\rho+R} [(t_i^n - s_i^n)^2 + (t_i^n - s_i^{n+1})^2], \end{aligned} \quad (30)$$

in which the configuration space contains the original degrees of freedom are $s_i^n \in \{-1, 1\}$ (with $n = 0, \dots, N/\rho - 1$ and $i = 1 + n\rho, \dots, R + n\rho$) and the auxiliary ones $t_i^n \in \{-1, 1\}$ (with $n = 0, \dots, N/\rho - 1$ and $i = 1 + (n+1)\rho, \dots, R + n\rho$). The relation between the original model and the auxiliary one is sketched in Fig. 6. In particular the coupling λ controls the strength of the bond in the auxiliary dimension (labeled by n), so that the limit $\lambda \rightarrow \infty$ describes the original chain with the obvious identification $s_i^n \rightarrow s_i$ and $t_i^n \rightarrow s_i$. By defining the row variables $\underline{s}^n = \{s_i^n\}_{i=1+n\rho}^{i=R+n\rho}$ and $\underline{t}^n = \{t_i^n\}_{i=1+(n+1)\rho}^{i=R+n\rho}$, the log-probability for the two dimensional model can be written as

$$\log p^\lambda(\underline{s}, \underline{t}) = -\log Z^\lambda(g) - \sum_{n=0}^{N/\rho-1} [\mathcal{H}_n^\lambda(\underline{s}^n) + \mathcal{H}_{n,n}^\lambda(\underline{s}^n, \underline{t}^n) + \mathcal{H}_{n,n+1}^\lambda(\underline{t}^n, \underline{s}^{n+1})]. \quad (31)$$

Hence the distribution over the degrees of freedom \underline{s}^n and \underline{t}^n defines a *tree*, because only successive row of variables interact.¹ The marginals associated with the clusters Γ_n and γ_n can be used in order to express the probability $p^\lambda(\underline{s}, \underline{t})$ as

$$p^\lambda(\underline{s}, \underline{t}) = \frac{\prod_n p_{\Gamma_n \cup \gamma_n}^\lambda(\underline{s}^n, \underline{t}^n) p_{\gamma_n \cup \Gamma_{n+1}}^\lambda(\underline{t}^n, \underline{s}^{n+1})}{\prod_n p_{\Gamma_n}^\lambda(\underline{s}^n) p_{\gamma_n}^\lambda(\underline{t}^n)}, \quad (32)$$

where for the two-dimensional model Γ_n and γ_n are analogously defined. By taking the $\lambda \rightarrow \infty$ limit, the identification

$$p_{\Gamma_n \cup \gamma_n}^\lambda(\underline{s}^n, \underline{t}^n) \rightarrow p_{\Gamma_n}(s_{n\rho+1}, \dots, s_{n\rho+R}), \quad (33)$$

$$p_{\Gamma_n}^\lambda(\underline{s}^n) \rightarrow p_{\Gamma_n}(s_{n\rho+1}, \dots, s_{n\rho+R}), \quad (34)$$

$$p_{\gamma_n}^\lambda(\underline{t}^n) \rightarrow p_{\gamma_n}(s_{(n+1)\rho+1}, \dots, s_{n\rho+R}) \quad (35)$$

allows to recover the factorization property (23).

References

1. Ackley, D., Hinton, G., Sejnowski, T.: A learning algorithm for Boltzmann machines. *Cogn. Sci.* **9**, 147 (1985)
2. Aurell, E., Ekeberg, M.: Inverse Ising inference using all the data. *Phys. Rev. Lett.* **108**(9), 090201 (2012)
3. Aurell, E., Ollion, C., Roudi, Y.: Dynamics and performance of susceptibility propagation on synthetic data. *Eur. Phys. J. B* **77**(4), 587–595 (2010)
4. Bailly-Bechet, M., Braunstein, A., Pagnani, A., Weigt, M., Zecchina, R.: Inference of sparse combinatorial-control networks from gene-expression data: a message passing approach. *BMC Bioinform.* **11**(1), 355 (2010)
5. Braunstein, A., Pagnani, A., Weigt, M., Zecchina, R.: Inference algorithms for gene networks: a statistical mechanics analysis. *J. Stat. Mech.* **2008**(12), P12001 (2008)
6. Cocco, S., Leibler, S., Monasson, R.: Neuronal couplings between retinal ganglion cells inferred by efficient inverse statistical physics methods. *Proc. Natl. Acad. Sci. USA* **106**, 14058 (2009)
7. Cocco, S., Monasson, R.: Adaptive cluster expansion for inferring Boltzmann machines with noisy data. *Phys. Rev. Lett.* **106**(9), 090601 (2011)
8. Cocco, S., Monasson, R.: Adaptive cluster expansion for the inverse Ising problem: convergence, algorithm and tests. *J. Stat. Phys.* **147**(2), 252–314 (2012)
9. Cover, T., Thomas, J.: *Elements of Information Theory*, vol. 6. Wiley, New York (1991)
10. Gori, G., Trombettoni, A.: The inverse Ising problem for one-dimensional chains with arbitrary finite-range couplings. *J. Stat. Mech.* **2011**, P10021 (2011)
11. Higuchi, S., Mézard, M.: 2009, Susceptibility propagation for constraint satisfaction problems. ArXiv preprint. <http://arxiv.org/abs/0903.1621v1>
12. Jaynes, E.: Information theory and statistical mechanics. *Phys. Rev.* **106**(4), 620–630 (1957)
13. Jaynes, E.: Information theory and statistical mechanics. II. *Phys. Rev.* **108**(2), 171 (1957)
14. Kappen, H., Rodriguez, F.: Efficient learning in Boltzmann machines using linear response theory. *Neural Comput.* **10**, 1137–1156 (1998)
15. de Lachapelle, D., Challet, D.: Turnover, account value and diversification of real traders: evidence of collective portfolio optimizing behavior. *New J. Phys.* **12**, 075039 (2010)
16. Lillo, F., Moro, E., Vaglica, G., Mantegna, N.: Specialization and herding behavior of trading firms in a financial market. *New J. Phys.* **10**, 043019 (2008)
17. Marinari, E., Van Kerrebroeck, V.: Intrinsic limitations of the susceptibility propagation inverse inference for the mean field Ising spin glass. *J. Stat. Mech.* **2010**, P02008 (2010)
18. Mézard, M., Montanari, A.: *Information, Physics and Computation*. Oxford University Press, London (2009)

¹Periodic boundary conditions enforce the presence of a single loop of length N , so that the model is not exactly a tree. Nevertheless, for N large enough and for g sufficiently distant from critical points of the model, if any, the presence of such loop can be neglected.

19. Mézard, M., Mora, T.: Constraint satisfaction problems and neural networks: a statistical physics perspective. *J. Physiol. Paris* **103**, 107–113 (2009)
20. Moro, E., Vicente, J., Moyano, L., Gerig, A., Farmer, J.D., Vaglica, G., Lillo, F., Mantegna, N.: Market impact and trading profile of hidden orders in stock markets. *Phys. Rev. E* **80**, 066102 (2009)
21. Nguyen, H., Berg, J.: Bethe–Peierls approximation and the inverse Ising problem. *J. Stat. Mech. Theory Exp.* **2012**(03), P03,004 (2012)
22. Ricci-Tersenghi, F.: The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods. *J. Stat. Mech.* **2012**(08), P08015 (2012)
23. Roudi, Y., Aurell, E., Hertz, J.: Statistical physics of pairwise probability models. *Front. Comput. Neurosci.* **3**(22), 1–15 (2009)
24. Roudi, Y., Hertz, J.: Mean field theory for nonequilibrium network reconstruction. *Phys. Rev. Lett.* **106**(4), 048702 (2011)
25. Schneidman, E., Berry, M. II, Segev, R., Bialek, W.: Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* **440**, 1007–1012 (2006)
26. Sessak, V., Monasson, R.: Small-correlation expansions for the inverse Ising problem. *J. Phys. A, Math. Theor.* **42**, 055,001 (2009)
27. Shlens, J., Field, G., Gauthier, J., Grivich, M., Petrusca, D., Sher, A., Litke, A., Chichilnisky, E.: The structure of multi-neuron firing patterns in primate retina. *J. Neurosci.* **26**(32), 8254–8266 (2006)
28. Socolich, M., Lockless, S., Russ, W., Lee, H., Gardner, K., Ranganathan, R.: Evolutionary information for specifying a protein fold. *Nature* **437**, 512–518 (2005)
29. Tanaka, T.: Mean field theory of Boltzmann machine learning. *Phys. Rev. E* **58**, 2302 (1998)
30. Tarantola, A.: *Inverse Problem Theory and Methods for Model Parameter Estimation*. SIAM, Philadelphia (2005)
31. Wainwright, M., Jordan, M.: Graphical models, exponential families, and variational inference. *Found. Trends Mach. Learn.* **1**(1–2), 1–305 (2008)
32. Wainwright, M., Ravikumar, P., Lafferty, J.: High-dimensional graphical model selection using ℓ_1 -regularized logistic regression. *Adv. Neural Inf.* **19**, 1465–1472 (2006)
33. Weigt, M., White, R., Szurmant, H., Hoch, J., Hwa, T.: Identification of direct residue contacts in protein–protein interaction by message passing. *Proc. Natl. Acad. Sci. USA* **106**, 67 (2009)
34. Welling, M., Teh, Y.: Approximate inference in Boltzmann machines. *Artif. Intell.* **143**(1), 19–50 (2003)