The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods

J. Stat. Mech. (2012) P08015

(http://iopscience.iop.org/1742-5468/2012/08/P08015)

Journal of Statistical Mechanics: Theory and Experiment
An IOP and SISSA journal

The Bethe approximation for solving the inverse Ising problem: a comparison with other inference methods

Federico Ricci-Tersenghi

Dipartimento di Fisica, INFN—Sezione di Roma 1, and CNR-IPCF, UOS di Roma, Università La Sapienza, Piazzale Aldo Moro 5, I-00185 Roma, Italy
E-mail: Federico.Ricci@roma1.infn.it

Received 30 March 2012
Accepted 6 August 2012
Published 24 August 2012

Online at stacks.iop.org/JSTAT/2012/P08015
doi:10.1088/1742-5468/2012/08/P08015

Abstract. The inverse Ising problem consists in inferring the coupling constants of an Ising model given the correlation matrix. The fastest methods for solving this problem are based on mean-field approximations, but which one performs better in the general case is still not completely clear. In the first part of this work, I summarize the formulas for several mean-field approximations and I derive new analytical expressions for the Bethe approximation, which allow one to solve the inverse Ising problem without running the susceptibility propagation algorithm (thus avoiding the lack of convergence). In the second part, I compare the accuracy of different mean-field approximations on several models (diluted ferromagnets and spin glasses) defined on random graphs and regular lattices, showing which one is in general more effective. A simple improvement over these approximations is proposed. Also a fundamental limitation is found in using methods based on TAP and Bethe approximations in the presence of an external field.

Keywords: spin glasses (theory), statistical inference

1742-5468/12/P08015+23$33.00

Contents

Mean-field approximations (MFA) are very important tools in statistical mechanics, since they provide an approximated description of a physical system in terms of just a few parameters (e.g. local magnetizations). Among the MFA the one based on the Bethe approximation (BA) is particularly effective. In recent years the BA—originally derived for the ferromagnetic model on regular lattices [1]—has been extended, under the name of the cavity method, to models having arbitrary couplings and topologies [2]. Although the BA is exact only for tree-like topologies, its application to models defined on random graphs has proved very successful: see e.g. the cases of low density parity check codes, spin glasses and constraint satisfaction problems, all nicely reviewed in [3].

The inverse Ising problem, originally known as Boltzmann machine learning, consists in inferring coupling constants for an Ising model (both pairwise interactions and external fields) given the vector of magnetizations and the matrix of pairwise correlations. In recent years the inverse Ising problem has received a lot of attention, especially in connection with inference in biological problems [4]–[7].

The inverse Ising problems can be viewed as the dual problems with respect to the 'direct' problems of estimating magnetizations and correlation given the Hamiltonian. So, under any MFA, the inverse problem can be solved by inverting (if possible) the analytic expressions that give the magnetizations and the correlations as a function of interactions and fields. Although MFA usually do not directly provide the correlations between distant variables, these correlations can be computed by using the linear response theorem [8].

J. Stat. Mech. (2012) P08015

Within the BA, a very fast and efficient way of estimating correlations between any pair of variables is given by the susceptibility propagation (SuscProp) algorithm recently introduced in [9]. SuscProp is an iterative algorithm for solving a set of self-consistency equations; when it converges it is very fast, but sometimes it may not converge. Indeed the application of the BA to the inverse Ising problem has been limited up to now by the range of convergence of SuscProp [10, 11].

In this work I present an analytical expression for the fixed point of SuscProp, thus avoiding any problem related to its lack of convergence. Actually such an expression already appeared in [12], but was unknown to the statistical mechanics community (including the author): otherwise there would be no need for the SuscProp algorithm introduced in [9].

In the first part of this work I derive new analytical expressions for solving the inverse Ising problem under the BA. These analytical expressions allow for a fair comparison among different MFA, over a wide range of temperatures, both for the problem of estimating two-point correlations given the couplings (the direct problem) and for the problem of estimating couplings given correlations and magnetizations (the inverse problem).

In the present work I consider MFA obtained from the so-called Plefka expansion [13]–[15] and from a small correlations expansion [16]. For the inverse Ising problem I compare methods that take as input only the correlation matrix. More complex, but usually slower, inference methods exist that require many samples of equilibrium configurations [17]–[20] and not only the correlation matrix.

Improvements on these MFA would be very welcome. It is well known that MFA ignore loops, so correcting these MFA by adding the loop contributions would be the right direction to follow. However at present all the methods which have been developed for considering the loops explicitly [21]–[25] do not provide analytical expressions for the correlations, which are simple enough to be inverted. For this reason, I have not considered these improved algorithms in the comparison of MFA for solving the inverse Ising problem.

Nonetheless, I am proposing a simple improvement of inference methods (both for the direct and the inverse problems) based on the idea that the loops may modify similarly self-correlations and correlations between variables that are close by.

1. The model and the mean-field approximations

In order to keep the presentation simple, I prefer to deal only with binary variables (Ising spins) $s_i = \pm 1$ and Hamiltonians containing interactions of up to two bodies, i.e. external fields and pairwise couplings. Thus, the most general model that I want to study is defined by the following joint probability distribution over N Ising variables:

$$P(s_1, \ldots, s_N) = \frac{1}{Z(\boldsymbol{J}, \boldsymbol{h})} \exp\left[\sum_{i \neq j} J_{ij} s_i s_j + \sum_i h_i s_i\right], \tag{1}$$

where the partition function $Z(\boldsymbol{J}, \boldsymbol{h})$ is a normalizing constant that depends on all the couplings $\boldsymbol{J} = \{J_{i,j}\}$ and the external fields $\boldsymbol{h} = \{h_i\}$. Please note that the temperature parameter has been absorbed in the definition of external fields and couplings. All the

J. Stat. Mech. (2012) P08015

required information about the model is encoded in the free energy

$$F(\boldsymbol{J},\boldsymbol{h}) = \ln Z(\boldsymbol{J},\boldsymbol{h}). \tag{2}$$

In the rest of this section, I summarize the most common MFA to the free energy: I am particularly interested in deriving the self-consistency equations for the magnetizations that are used in section 2 for obtaining two-point correlations.

The simplest MFA, also known as the naive MF (nMF) approach, approximates the model in terms of local magnetizations $m_i = \langle s_i \rangle$, where the angular brackets represent the average w.r.t. the measure in equation (1). The corresponding approximation to the free energy is

$$F_{\mathrm{nMF}} = \sum_i \left[H\left(\frac{1+m_i}{2}\right) + H\left(\frac{1-m_i}{2}\right)\right] + \sum_i h_i m_i + \sum_{i\neq j} J_{ij} m_i m_j, \tag{3}$$

where $H(x) \equiv -x\ln(x)$ and the m_i must be fixed according to the self-consistency equations

$$\frac{\partial F_{\mathrm{nMF}}}{\partial m_i} = \sum_j J_{ij} m_j + h_i - \operatorname{atanh}(m_i) = 0 \quad \Rightarrow \quad m_i = \tanh\left[h_i + \sum_j J_{ij} m_j\right]. \tag{4}$$

A better MFA can be obtained by considering also the Onsager reaction term [26], leading to the following TAP approximated free energy and self-consistency equations:

$$F_{\mathrm{TAP}} = \sum_i \left[H\left(\frac{1+m_i}{2}\right) + H\left(\frac{1-m_i}{2}\right)\right] + \sum_i h_i m_i + \sum_{i\neq j} \left(J_{ij} m_i m_j + \tfrac{1}{2} J_{ij}^2 (1-m_i^2)(1-m_j^2)\right), \tag{5}$$

$$m_i = \tanh\left[h_i + \sum_j J_{ij}(m_j - J_{ij}(1-m_j^2)m_i)\right]. \tag{6}$$

In the TAP approximation, when computing the marginal probability of spin s_i (i.e. its magnetization m_i), the reaction term modifies the marginal probabilities of the neighboring spins, $m_j \to (m_j - J_{i,j}(1-m_j^2)m_i)$, in order to try to remove the effect of the spin s_i under study. It has been recognized [13, 14] that F_{nMF} and F_{TAP} are only the first two terms of the expansion of $F(\boldsymbol{J},\boldsymbol{h})$ in small couplings $\boldsymbol{J}$ at fixed magnetizations $\boldsymbol{m} = \{m_i\}$. This expansion contains [14] both loop terms, like $J_{ij}J_{j\ell}J_{\ell i}$, and terms with higher powers of a single coupling, i.e. J_{ij}^k: the latter terms, that correspond to considering recursively the reaction to the reaction between spins s_i and s_j, can be resummed and lead to the BA.

The BA gives a description of the model in terms of magnetizations m_i and connected correlations $c_{ij} = \langle s_i s_j\rangle - m_i m_j$ between neighboring spins (i.e. spins connected by a non-zero coupling J_{ij}). The BA can be derived in two equivalent ways. The first way consists

J. Stat. Mech. (2012) P08015

in finding values of $\boldsymbol{m}$ and $\boldsymbol{c}$ minimizing the following free energy:

$$F_{\rm BA} = \sum_{i\neq j}\left[H\left(\frac{(1+m_i)(1+m_j)+c_{ij}}{4}\right)+H\left(\frac{(1-m_i)(1-m_j)+c_{ij}}{4}\right)\right.$$
$$\left. + \; H\left(\frac{(1+m_i)(1-m_j)-c_{ij}}{4}\right)+H\left(\frac{(1-m_i)(1+m_j)-c_{ij}}{4}\right)\right]$$
$$+ \sum_i (1-d_i)\left[H\left(\frac{1+m_i}{2}\right)+H\left(\frac{1-m_i}{2}\right)\right]+\sum_i h_i m_i$$
$$+ \sum_{i\neq j} J_{ij}(c_{ij}+m_i m_j), \qquad (7)$$

where d_i is the degree of spin s_i, i.e. the number of its neighboring spins. In equation (7) the last two terms correspond to the average value of the energy at given magnetizations and neighboring correlations, while the first two terms correspond to the entropy of the Bethe approximation to the joint probability distribution of the N spin variables,

$$P(s_1,\ldots,s_N) \overset{BA}{\simeq} \prod_{(ij)}\frac{p_{ij}(s_i,s_j)}{p_i(s_i)p_j(s_j)}\prod_i p_i(s_i), \qquad (8)$$

where the first product runs over all pairs of neighboring spins and the two-spin and single-spin marginal probabilities are given respectively by $p_{ij}(s_i,s_j) = [(1+m_i s_i)(1+m_j s_j)+c_{ij}s_i s_j]/4$ and $p_i(s_i) = (1+m_i s_i)/2$. The conditions $\partial F_{\rm BA}/\partial c_{ij} = 0$ can be solved analytically and lead to

$$J_{ij} = \frac{1}{4}\ln\left(\frac{((1+m_i)(1+m_j)+c_{ij})((1-m_i)(1-m_j)+c_{ij})}{((1+m_i)(1-m_j)-c_{ij})((1-m_i)(1+m_j)-c_{ij})}\right), \qquad (9)$$

$$c_{ij}(m_i,m_j,t_{ij}) = \frac{1}{2t_{ij}}\left(1+t_{ij}^2-\sqrt{(1-t_{ij}^2)^2-4t_{ij}(m_i-t_{ij}m_j)(m_j-t_{ij}m_i)}\right)-m_i m_j. \qquad (10)$$

where $t_{ij} = \tanh(J_{ij})$. Please note that equation (9) is identical to equation (26) in [16] and this is a further confirmation that resumming all two-spin terms in the Plefka expansion leads to the BA. Moreover equation (9) has been used in the literature [7, 27] as the independent-pair (IP) approximation for inferring couplings from magnetizations and correlations: such an approximation infers the coupling J_{ij} by assuming that spins s_i and s_j form an isolated pair with magnetizations m_i and m_j and correlation c_{ij}. Unfortunately under this IP approximation, computing the external fields is not immediate and moreover even the estimates of the couplings are rather poor (see section 5).

By making the substitution $c_{ij} \to c_{ij}(m_i,m_j,t_{ij})$ in $F_{\rm BA}$ one can obtain the Bethe free energy only in terms of magnetizations, from which the self-consistency equations for the magnetizations can be derived. However this derivation requires a rather complicated algebra and I prefer to obtain the same equations in a much simpler alternative way.

In the so-called cavity method [2], local magnetizations m_i and neighboring correlations c_{ij} are expressed in terms of some auxiliary variables, the cavity magnetizations $m_i^{(j)}$ (i.e. the mean value of s_i in the absence of a neighboring spin s_j):

$$m_i = \frac{m_i^{(j)}+t_{ij}\,m_j^{(i)}}{1+m_i^{(j)}\,t_{ij}\,m_j^{(i)}}, \qquad (11)$$

J. Stat. Mech. (2012) P08015

$$m_j = \frac{t_{ij}\, m_i^{(j)} + m_j^{(i)}}{1 + m_i^{(j)}\, t_{ij}\, m_j^{(i)}}, \tag{12}$$

$$c_{ij} = \frac{t_{ij} + m_i^{(j)}\, m_j^{(i)}}{1 + m_i^{(j)}\, t_{ij}\, m_j^{(i)}} - m_i m_j. \tag{13}$$

Cavity magnetizations must satisfy the self-consistency equations

$$m_i^{(j)} = \tanh\left[h_i + \sum_{k(\neq j)} \operatorname{atanh}(t_{ik}\, m_k^{(i)})\right]. \tag{14}$$

These equations are often solved by an iterative algorithm known as belief propagation (BP) [28]: in the case of convergence, the fixed point of BP gives directly the Bethe free energy that admits an expression in terms of cavity magnetizations only [2].

In order to obtain a closed set of self-consistency equations in the magnetizations $\boldsymbol{m}$, I solve equations (11) and (12) for the cavity magnetizations and find

$$m_i^{(j)} = f(m_i, m_j, t_{ij}) \qquad m_j^{(i)} = f(m_j, m_i, t_{ij}), \tag{15}$$

where

$$f(m_1, m_2, t) = \frac{1 - t^2 - \sqrt{(1-t^2)^2 - 4t(m_1 - m_2 t)(m_2 - m_1 t)}}{2t(m_2 - m_1 t)}. \tag{16}$$

The sign in front of the square root has been chosen such that $f(0, 0, t) = 0$, as it should. A consistency check can be made by substituting expressions (15) in equation (13) to obtain again the result in equation (10). Finally, combining equations (11) and (14), it is possible to obtain the self-consistency equation for the magnetizations under the BA:

$$m_i = \tanh\left[h_i + \sum_j \operatorname{atanh}(t_{ij} f(m_j, m_i, t_{ij}))\right]. \tag{17}$$

It is only fair to comment that the use of this formula for finding Bethe magnetizations is not a good idea: indeed an iterative solution of equation (17) is typically more unstable than BP for solving equation (14). My interest in this formula is that it involves only physical magnetizations (not cavity ones) and can be used to obtain correlations (see section 2) and to solve quickly the inverse Ising problem (see section 5).

A series expansion of the exponent in equation (17) for small couplings gives

$$h_i + \sum_j \operatorname{atanh}(t_{ij} f(m_j, m_i, t_{ij})) \simeq h_i + \sum_j (J_{ij} m_j - J_{ij}^2 (1 - m_j^2) m_i + \cdots), \tag{18}$$

and one recognizes that the first two terms of the expansion are the naive MF approximation and the Onsager reaction term. This expansion should make it clearer that the BA is a way of considering recursively all the reactions between a pair of neighboring variables.

J. Stat. Mech. (2012) P08015

2. Computing correlations using the linear response

A preliminary step to solving the inverse Ising problem by using any MFA is to derive an analytical expression for the pairwise correlations as a function of the coupling constants. Actually, the MFA discussed in section 1 do not provide information about the correlation between distant variables: indeed, naive MF and TAP approximations give $c_{ij} = 0$ for any pair of variables, and the BA only provides an expression for the correlation between neighboring spins (see equation (10)), which is trivially $c_{ij} = t_{ij}$ in the case of null magnetizations.

Nonetheless, a closed set of equations for the connected correlations[1], $C_{ij} \equiv \langle s_i s_j \rangle - \langle s_i \rangle \langle s_j \rangle$ for any pair i, j, can be derived from the magnetizations self-consistency equations, equations (4), (6) and (17), through the linear response [8, 12]:

$$C_{ij} = \frac{\partial m_i}{\partial h_j}, \qquad (C^{-1})_{ij} = \frac{\partial h_i}{\partial m_j}. \tag{19}$$

The inverse correlation matrices C^{-1} for the three MFA discussed above are given by the following expressions:

$$\text{naive MF} \quad (C^{-1}_{\text{nMF}})_{ij} = \frac{\delta_{ij}}{1 - m_i^2} - J_{ij}, \tag{20}$$

$$\text{TAP} \quad (C^{-1}_{\text{TAP}})_{ij} = \left[\frac{1}{1 - m_i^2} + \sum_k J_{ik}^2 (1 - m_k^2) \right] \delta_{ij} - \left(J_{ij} + 2 J_{ij}^2 m_i m_j \right), \tag{21}$$

$$\text{Bethe} \quad (C^{-1}_{\text{BA}})_{ij} = \left[\frac{1}{1 - m_i^2} - \sum_k \frac{t_{ik} f_2(m_k, m_i, t_{ik})}{1 - t_{ik}^2 f(m_k, m_i, t_{ik})^2} \right] \delta_{ij} - \frac{t_{ij} f_1(m_j, m_i, t_{ij})}{1 - t_{ij}^2 f(m_j, m_i, t_{ij})^2}, \tag{22}$$

where $f_1(m_1, m_2, t) \equiv \partial f(m_1, m_2, t)/\partial m_1$ and $f_2(m_1, m_2, t) \equiv \partial f(m_1, m_2, t)/\partial m_2$. From these expressions one can obtain directly any correlation by simply computing the inverse of a matrix.

Please note that equation (22) gives exactly the same solution as is found by the SuscProp iterative algorithm [9], which is currently considered among the best inference algorithms. The main advantage of equation (22) is that it always provides the correlation matrix, even in those cases where SuscProp does not converge to the fixed point. Moreover inverting a matrix takes roughly the same time as a single iteration of SuscProp, and so using equation (22) is much faster than running SuscProp, even when the latter converges.

Nevertheless, it is only fair to note that the use of equation (22) does not solve all the problems related to the lack of convergence of SuscProp. Indeed, during the many tests that I have run, I noticed that often the lack of convergence of SuscProp does correspond to the BA fixed point becoming unphysical: in these cases, by inverting the correlation matrix provided by equation (22), one gets an unphysical correlation matrix (e.g. a correlation matrix with negative diagonal elements!) In this sense the lack of convergence of SuscProp gives a warning that the 'blind' use of equation (22) does not provide. So, a general suggestion when using the above formulas, providing an analytical expression for the

[1] Please do not confuse the correlation C_{ij} with the parameter c_{ij} appearing in the BA: the two coincide only when the BA is exact.

J. Stat. Mech. (2012) P08015

correlation matrices under a MFA, is to check explicitly the physical consistency of the outcome.

One may comment that equation (22) contains the magnetizations, and the iterative computation of these (i.e. the BP algorithm) suffers the same convergence problems as SuscProp: this is easy to prove, given that the homogeneous SuscProp equations are nothing but the iterative equations for evolving under BP a small perturbation in the magnetization, and so BP is unstable if SuscProp does not converge. However there are provably convergent algorithms for the computation of magnetizations under the BA [29, 30]: the use of these algorithms in conjunction with equation (22) allows a direct computation of correlations under the BA. Moreover there are situations where magnetizations are known *a priori* and equation (22) can be applied directly: e.g. when symmetries in the probability measure force magnetizations to be zero, or in the inverse Ising problem, where magnetizations are given as an input to the problem. In the rest of the paper I deal mainly with these two cases.

2.1. Estimating correlations in the case of null magnetizations

A preliminary ranking of MFA can be done on the basis of how good their estimates of correlations are, given the couplings. Indeed I expect that the better this estimate is, the better the solution to the inverse problem will be.

For simplicity I concentrate on models with no external fields and the couplings are multiplied by a parameter β (the inverse temperature) such that the difficulty of the inference problem increases with β.

In the case of null magnetizations, the expressions for the inverse correlation matrices simplify a lot:

$$\text{naive MF} \qquad (C^{-1}_{\text{nMF}})_{ij} = \delta_{ij} - J_{ij}, \tag{23}$$

$$\text{TAP} \qquad (C^{-1}_{\text{TAP}})_{ij} = \left[1 + \sum_k J^2_{ik}\right] \delta_{ij} - J_{ij}, \tag{24}$$

$$\text{Bethe} \qquad (C^{-1}_{\text{BA}})_{ij} = \left[1 + \sum_k \frac{t^2_{ik}}{1 - t^2_{ik}}\right] \delta_{ij} - \frac{t_{ij}}{1 - t^2_{ij}}, \tag{25}$$

since $f_1(0, 0, t) = 1/(1 - t^2)$ and $f_2(0, 0, t) = -t/(1 - t^2)$.

Given that for $m_i = 0$, the expressions for the correlation matrices are much simpler, I report also those that can be obtained from the Plefka expansion at the third and fourth order:

$$\text{Third order} \quad (C^{-1}_3)_{ij} = \left[1 + \sum_k J^2_{ik} + 2 \sum_{k,\ell} J_{ik} J_{k\ell} J_{\ell i}\right] \delta_{ij} - \left(J_{ij} + \tfrac{2}{3} J^3_{ij}\right), \tag{26}$$

$$\text{Fourth order} \quad (C^{-1}_4)_{ij} = \left[1 + \sum_k J^2_{ik} + 2 \sum_{k,\ell} J_{ik} J_{k\ell} J_{\ell i} + \tfrac{1}{3} \sum_k J^4_{ik} + 2 \sum_{k,\ell,m} J_{ik} J_{k\ell} J_{\ell m} J_{mi}\right] \delta_{ij}$$
$$- \left(J_{ij} + \tfrac{2}{3} J^3_{ij} + 2 J^2_{ij} \sum_k J_{ik} J_{kj}\right). \tag{27}$$

J. Stat. Mech. (2012) P08015

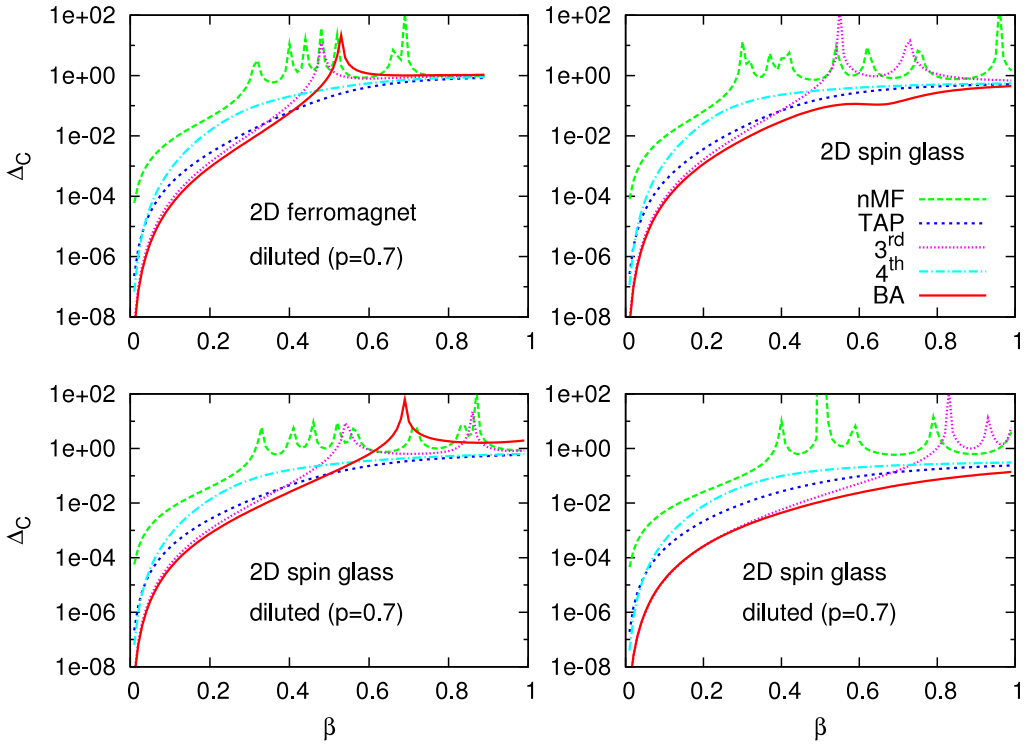

Figure 1. Error incurred by five mean-field approximations in estimating the correlation matrix, given the couplings. Shown are typical samples of size $N = 5^2$ (the qualitative behavior does not change for larger sizes).

The purpose is to understand whether and to what extent the estimate of the correlation matrix improves upon adding terms in the Plefka expansion.

I have tested the accuracy of formulas in equations (23)–(27) for ferromagnetic ($J_{ij} = 1$) and spin glass ($J_{ij} = \pm 1$) models defined on fully connected (FC) topologies, on a 2D square lattice and on a 3D cubic lattice. In diluted versions of these models a fraction $(1 - p)$ of couplings has been set to zero. In models defined on FC graphs the couplings have been normalized so as to have a critical inverse temperature $\beta_c = 1$ in the thermodynamic limit.

The discrepancy between true correlations C and those inferred C' is defined as

$$\Delta_C \equiv \sqrt{\frac{1}{N^2}\sum_{i,j}(C_{ij} - C'_{ij})^2}. \tag{28}$$

In figures 1 and 2 I report the typical behavior of the error Δ_C between exact and estimated correlation matrices for five different MFA. Figure 1 shows results for models defined on a 2D square lattice, while figure 2 refers to FC and 3D topologies. In order to compare the MFA estimates with the exact correlation matrices I am studying small systems here, but the qualitative behavior does not change for larger sizes.

Although the quantitative behavior of Δ_C depends on the specific sample, some general statements can be made:

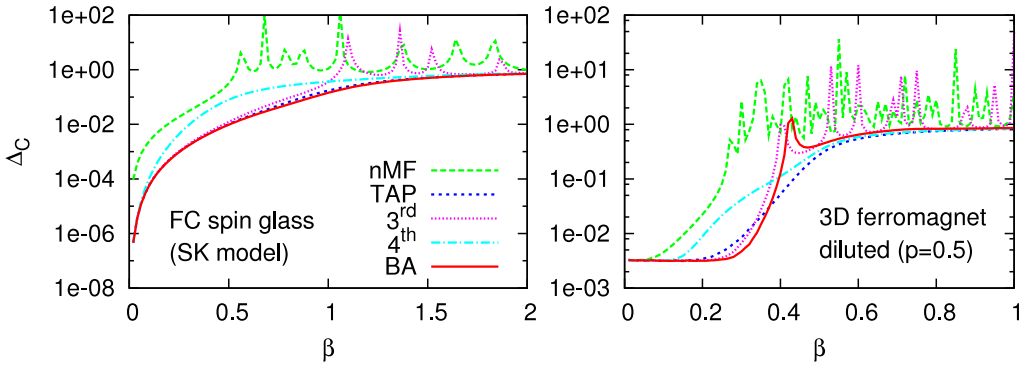

Figure 2. Same as figure 1, but for typical samples of the fully connected spin glass (SK model) of size $N = 20$ and of the 3D diluted ferromagnet of size $N = 3^5$.

- naive MF is typically the worst MFA and shows many spurious singularities (roughly one for each peak in Δ_C);
- TAP and fourth-order approximations typically show no (or very rare) singularities;
- the best estimate is typically provided by BA and TAP, with BA being the best unless it has a singularity (in this case TAP becomes the best at lower temperatures, and higher β).

These results suggest that increasing the number of terms in the Plefka expansion does not always improve the estimate of the correlation matrix (as one could have naively expected). On the basis of these preliminary results I believe that it is appropriate to consider only TAP and BA for the inverse Ising problem, together with other inference methods (see section 4).

In the left panel of figure 2 the results obtained from TAP and BA are almost perfectly superimposed (indeed the former is not clearly visible). This is expected since the TAP approximation is exact for the SK model at high temperatures ($\beta < \beta_c = 1$) in the large N limit: so the BA cannot improve it, except in $1/N$ corrections. Indeed a careful analysis shows a tiny improvement of BA over TAP around the critical temperature, where $1/N$ corrections are stronger.

Please note that in the right panel of figure 2 the high temperature (small β) behavior of Δ_C is very different to that in previous plots: indeed for $\beta \to 0$, Δ_C goes to a constant, instead of decreasing with a power law in β (as in figure 1 and in the left panel of figure 2). This is due to the fact that the comparison has not been made with the exact correlation matrix, but with correlations measured from a Monte Carlo (MC) simulation. Actually, in this case, I have used the Wolff algorithm and the correlation matrix has been computed from 10^5 independent measures. The difference between the error due to the MFA and the error due to MC noisy data can be better appreciated in figure 3: in the high temperature region the error does not decrease below a limiting value given roughly by the inverse of the square root of the number of measures.

3. Improving inference algorithms

Expressions in equations (23)–(27) are intrinsically approximated, and turn out to be correct only in some particular cases. Naive MF and TAP approximations (as well as

J. Stat. Mech. (2012) P08015

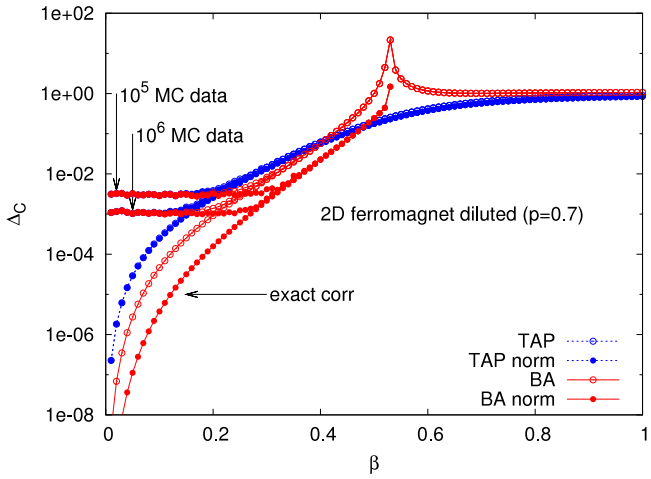

Figure 3. Same as figure 1, but for a typical sample of the 2D diluted ferromagnet of size $N = 5^2$. The error Δ_C has been computed with respect to the exact correlation matrix and with respect to the one measured in MC simulations. Full points show the error obtained with the normalization trick.

third-order and fourth-order approximations), being the first orders in a small coupling expansion, are exact only in the limit of very weak couplings (either high temperature or fully connected models in the large N limit). The BA, in contrast, is exact also for coupling intensities O(1), but only if the interacting network is a tree; on random graph models (which are locally tree-like) the BA turns out to be correct as long as the model has only one state (modulo the known symmetries). On any other model those expressions are approximated and it is worth trying to improve on this.

Let me first note that any of the above MFA returns in general a value for the self-correlation differing from the exact one, i.e. $C_{ii} \neq 1$ (for simplicity I consider the case of null magnetizations, but the argument is general). This fact can be easily explained, noticing that all the above MFA assume that correlations along loops are vanishingly small (at least in the large N limit). However, on any loopy graph, e.g. a regular lattice one, correlations along loops are important and may significantly alter the mean-field estimates. A general solution to this problem is still not available, although a lot of work is in progress on including loop contributions to MFA [21]–[25].

What I am proposing here is a simple heuristic improvement. Once the correlation matrix C_{ij} is computed by one of the approximations described in section 2, a properly *normalized* correlation matrix can be defined:

$$\widehat{C}_{ij} \equiv \frac{C_{ij}}{\sqrt{C_{ii}C_{jj}}}. \tag{29}$$

By definition $\widehat{C}_{ii} = 1$, and also an off-diagonal element may approximate better the true correlations. The reason for this is that the loops neglected in MFA actually modify in a similar way both self-correlations C_{ii} and off-diagonal correlations C_{ij}, and the heuristic normalization in equation (29) is assuming that the modifying factor only depends on the loop structure around sites i and j (which is certainly wrong for distant sites, but may be a reasonable approximation for sites that are close by).

J. Stat. Mech. (2012) P08015

In figure 3 full points show that the error Δ_C in the BA decreases by roughly one order of magnitude if normalized correlations are used. In contrast, the TAP result is not very sensitive to this normalization: the reason is that the estimates of the self-correlations in TAP remain quite close to the right value, especially if compared to BA estimates that diverge at the singularity (marked by a peak in figure 3). On the right of such a peak the error obtained by the normalized BA is not reported because equation (29) cannot be used, since several BA estimates for self-correlations are negative. This is the problem of the BA fixed point becoming (strongly) unphysical, already discussed in section 2: indeed by running SuscProp on this sample one would observe convergence only for β smaller than the peak location. I would like to stress again that checking the physical consistency of a solution based on a MFA is very important: for the sample shown in figure 3, even without knowing the exact correlations, one should switch from the BA to the TAP when the former reaches the singularity (that manifests e.g. in SuscProp not converging or in self-correlations diverging)[2].

Moreover there are cases (e.g. homogeneous FC models) where the spurious singularity induced by the MFA in a system of finite size is such that C_{ii} and C_{ij} diverge with the same law at the spurious critical point, while the normalized correlation $\widehat{C}_{ij}$ stays finite (and much closer to the true one). For example for the FC ferromagnetic model the normalized correlation $\widehat{C}_{\mathrm{MFA}}$ estimates the true correlation with an error roughly half that of C_{MFA} for any of the five MFA considered here.

4. Methods for the inverse Ising problem

I consider four different approximations for solving the inverse Ising problem. The simplest one is the independent-pair (IP) approximation, already discussed in section 1 and recalled here for convenience:

$$J_{ij}^{\mathrm{IP}} = \frac{1}{4}\ln\left(\frac{((1+m_i)(1+m_j)+C_{ij})((1-m_i)(1-m_j)+C_{ij})}{((1+m_i)(1-m_j)-C_{ij})((1-m_i)(1+m_j)-C_{ij})}\right). \tag{30}$$

Among the MFA which can be derived from the Plefka expansion, I consider only TAP and BA, because those are the ones that perform better for the direct problem of estimating correlations (see section 2). The corresponding expressions for the inferred couplings can be obtained by solving the equation

$$2m_i m_j J_{ij}^2 + J_{ij} + (C^{-1})_{ij} = 0 \qquad \forall (i \neq j) \tag{31}$$

for TAP and the equation

$$(C^{-1})_{ij} = \frac{-t_{ij} f_1(m_j, m_i, t_{ij})}{1 - t_{ij}^2 f(m_j, m_i, t_{ij})^2} = \frac{-t_{ij}}{\sqrt{(1-t_{ij}^2)^2 - 4t_{ij}(m_i - t_{ij}m_j)(m_j - t_{ij}m_i)}} \qquad \forall (i \neq j) \tag{32}$$

[2] Actually for a ferromagnet one knows how to break the up–down symmetry and let BP converge even at low temperatures: once BP returns non-zero magnetizations m_i, the correlation matrix can be computed by means of equation (22). However in the general case, BP does not converge in the presence of long range correlations, i.e. after the singularity, and one must resort to other MFA.

J. Stat. Mech. (2012) P08015

for the BA, thus leading to

$$J_{ij}^{\mathrm{TAP}} = \frac{\sqrt{1-8m_i m_j (C^{-1})_{ij}}-1}{4m_i m_j}, \tag{33}$$

$$J_{ij}^{\mathrm{BA}} = -\mathrm{atanh}\left[\frac{1}{2(C^{-1})_{ij}}\sqrt{1+4(1-m_i^2)(1-m_j^2)(C^{-1})_{ij}^2}-m_i m_j - \frac{1}{2(C^{-1})_{ij}}\sqrt{\left(\sqrt{1+4(1-m_i^2)(1-m_j^2)(C^{-1})_{ij}^2}-2m_i m_j (C^{-1})_{ij}\right)^2-4(C^{-1})_{ij}^2}\right]. \tag{34}$$

The fourth approximation that I am considering has been obtained from a small correlation expansion by Sessak and Monasson [16] and has been further simplified in [27] to the following expression

$$J_{ij}^{\mathrm{SM}} = -(C^{-1})_{ij} + J_{ij}^{\mathrm{IP}} - \frac{C_{ij}}{(1-m_i^2)(1-m_j^2)-(C_{ij})^2}. \tag{35}$$

For each approximation, I measure the error in inferred couplings J'_{ij} with respect to the true ones J_{ij} by using the following expression:

$$\Delta_J = \sqrt{\frac{\sum_{i<j}(J'_{ij}-J_{ij})^2}{\sum_{i<j}J_{ij}^2}}. \tag{36}$$

I study both the diluted ferromagnetic model with a fraction p of non-zero couplings ($J_{ij}=\beta$) and undiluted spin glass models ($J_{ij}=\pm\beta$ with probability 1/2). I also consider several topologies: 2D square lattices, 3D cubic lattices, random regular graphs with fixed degree $c=4$ and fully connected (FC) graphs. In the latter case the couplings are normalized in order to have a phase transition at $\beta_{\mathrm{c}}=1$ in the thermodynamic limit. I restrict the study to models of small sizes, with N ranging between 20 and 100, because these are the sizes for problems of biological interest. Moreover, as discussed below, the number M of independent measurements of the correlation matrix that make inferred coupling reasonably good grows linearly with N, and so for larger systems the number of measurements needed becomes too large. The data shown in section 5 have been obtained with $M=10^6$ independent measures of the correlation matrix (unless stated otherwise) and going to much larger values seems to me rather unrealistic as compared with practical applications.

4.1. The normalization trick for the inverse Ising problem

The trick of normalizing the correlation matrix to improve inference (see section 3) can be extended to the inverse Ising problem. In practice, it corresponds to solving *all* the equations relating the inverse correlation matrix C^{-1} to the couplings J_{ij}, including also those for the diagonal elements which are usually ignored.

Let me illustrate the new method for the simple case of the TAP approximation with null magnetizations. In this case, solving the inverse Ising problem only on the off-diagonal elements is equivalent to solving the equations

$$(C^{-1})_{ij} = -J_{ij} \equiv D_{ij} \qquad \forall (i \neq j),$$

J. Stat. Mech. (2012) P08015

but the diagonal equations are in general unsatisfied:

$$(C^{-1})_{ii} \neq 1 + \sum_k J_{ik}^2 \equiv D_{ii},$$

where D is the inverse correlation matrix estimated by using TAP once the couplings J_{ij} are given. Please note that diagonal elements D_{ii} are fully determined once the off-diagonal elements are known.

Following equation (29) I would like to normalize the estimated matrix D, and produce a normalized inverse correlation matrix $\widehat{D}$, matching better the true inverse correlation matrix C^{-1}. In other words I would like to solve the equations

$$(C^{-1})_{ij} = \widehat{D}_{ij} = D_{ij}\lambda_i\lambda_j \qquad \forall\ i, j, \tag{37}$$

where the N variables λ_i are exactly those necessary for solving the N new diagonal equations. Physically speaking, λ_i^2 should be the self-correlation C_{ii} produced by using the MFA, when using the right couplings J_{ij}. When λ_i becomes very different from 1, then the MFA is working very badly; however I expect situations where some of the errors produced by the MFA can be compensated for by using this normalization trick.

Unfortunately the solution to the new equations, those involving both J_{ij} and λ_i, does not have an analytical expression and has to be solved numerically. I have adopted an iterative solution, which is very fast (when it converges). In practice, I start with all $\lambda_i = 1$ and then, iteratively, first I solve the off-diagonal equations, thus getting an estimate for the couplings, and then I solve the diagonal equations to obtain a new estimate for the λ values. I repeat this iterative procedure, updating the values of λ with a damping factor, until the variations of the λ values are below a threshold (typically 10^{-8}).

From the many tests that I have run, I noticed that this normalization trick is more relevant for unfrustrated models (such as the diluted ferromagnets studied below) or models containing regions very weakly frustrated (those usually leading to Griffith singularities). Most probably in these weakly frustrated regions correlations get self-reinforced by the loops (ignored in the MFA) and thus the normalization trick may improve the coupling estimates. In contrast, for strongly disordered models, like spin glasses, the effect of the normalization trick depends a lot on the disordered sample and it does not seem to give a clear improvement on average.

The use of the normalization trick for improving inference in the inverse Ising problem may resemble the use of the diagonal-weight trick introduced in [8], but it is actually very different. In the diagonal-weight trick, the self-couplings J_{ii} are allowed to take non-zero values in order to solve all the equations $(C^{-1})_{ij} = D_{ij}$, while in the normalization trick the self-couplings J_{ii} remain null. It has been shown [15] that the first-order approximation (nMF) with the diagonal-weight trick improves over the second-order approximation (TAP) in estimating magnetizations. However the estimates for the couplings do not improve at all, because the off-diagonal equations are left unchanged by the diagonal-weight trick. In contrast, the normalization trick used here does change the estimates for the couplings. Moreover, being based on the very general requirement that self-correlations must take the right value, it can be applied to any approximation.

J. Stat. Mech. (2012) P08015

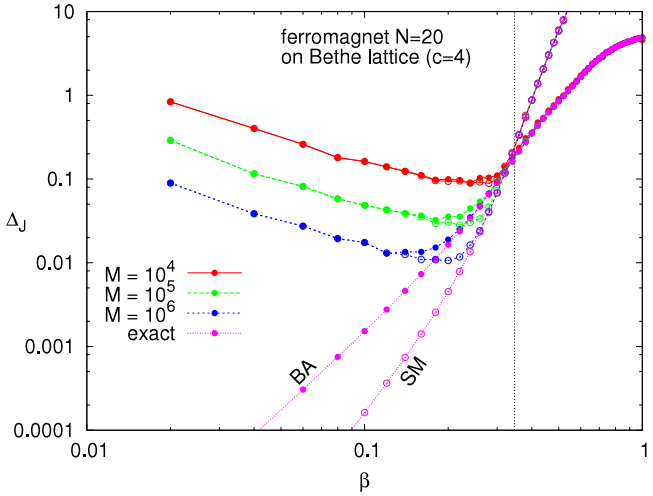

Figure 4. Dependence of the error Δ_J in inferred couplings on the number M of independent measures for the correlation matrix. The model is a ferromagnet of $N = 20$ variables on a regular random graph of fixed degree $c = 4$, whose critical temperature is marked by the vertical dotted line.

5. Numerical results on the inverse Ising problem

Let me start by making some general statements that summarize the numerical results shown in this section. Among the four approximations studied (IP, TAP, SM and BA) it seems in general true that:

- IP always provides the worst estimate, especially at high temperatures (low β);
- BA always outperforms TAP;
- as regards SM and BA, the former is better in the high temperature (low β) phase, while the latter is better at lower temperatures (higher β values);
- at low β, the error Δ_J in couplings inference is completely dominated by the uncertainty in the correlation matrix and it does not depend on the inference method; for this reason the range where SM is the best method becomes tiny, especially for noisy data;
- at high β, the errors Δ_J produced by TAP and SM diverge, the one for IP stays limited, but very high, and only BA may have a reasonable error;
- for diluted ferromagnetic models the normalization trick works well and thus BA with the normalization trick is the best method overall;
- in the presence of an external field, i.e. when magnetizations are different from zero, TAP and BA stop working at high enough β values (i.e. these methods do not admit a solution); given that at high β neither IP nor SM provides acceptable inferred couplings, I conclude that in such a situation the inverse Ising problem needs to be solved by other methods not explored in the present work.

5.1. Ferromagnetic models

Let me start by discussing the high temperature (low β) regime. In figure 4 there are reported the errors Δ_J in inferring the couplings of a ferromagnetic model of $N = 20$

J. Stat. Mech. (2012) P08015

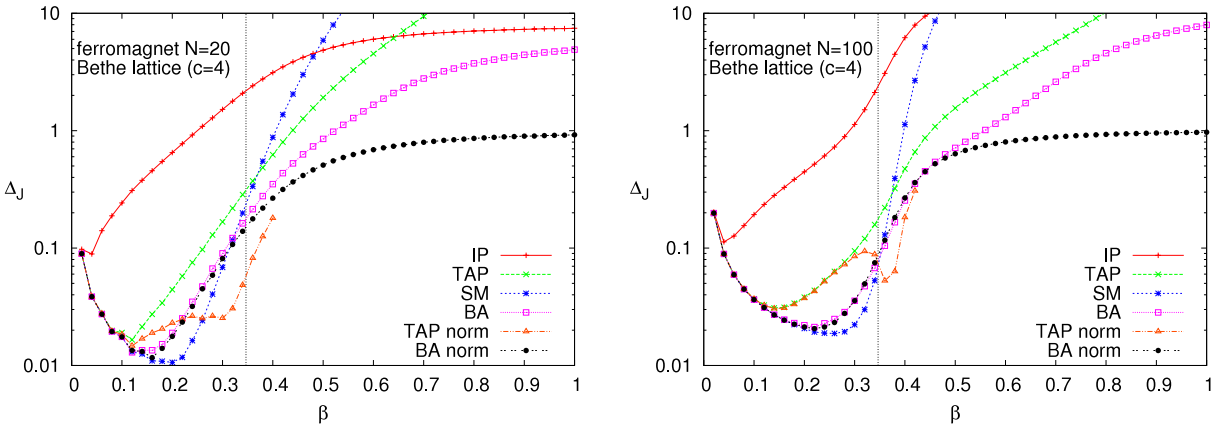

Figure 5. Errors in the couplings inferred from several approximations. The model is a ferromagnet on a random regular graph of fixed degree 4. Shown are two typical samples of sizes $N = 20$ (left) and $N = 100$ (right). The vertical dotted lines mark the locus of the ferromagnetic phase transition in the thermodynamic limit.

variables on a regular random graph of fixed degree $c = 4$. For simplicity I plot only the results obtained with SM and BA. Data in the upper curves have been computed using a correlation matrix averaged over M independent measurements, while data in the lower curves have been computed from the exact correlation matrix. It is clear that the low β regime is completely dominated by the uncertainty in the correlation matrix (as already noticed in [10]), and the error in this regime is independent of the inference method used. For this reason the good performances of SM in this regime are actually washed out and, even for $M = 10^6$ measurements, the improvement of SM over BA is very limited (see figure 4). Moreover such an improvement tends to become smaller on increasing the system size, because in the low β regime the error goes like

$$\Delta_J \propto \frac{1}{\beta}\sqrt{\frac{N}{M}}. \tag{38}$$

I think that comparing inference methods by using the exact correlation matrix is rather unrealistic, given that in any practical application the correlations are always known with some uncertainty. So in presenting below the numerical results, I always consider the case with $M = 10^6$ independent measurements for the magnetizations and the correlations.

In figure 5 I am showing the error in the couplings inferred from several approximations for a ferromagnet on a random regular graph with fixed degree $c = 4$. The two panels correspond to sizes $N = 20$ (left) and $N = 100$ (right) and show that the qualitative behavior is mostly size independent. Also the dependence on the specific sample (i.e., on the random graph) is rather weak. The data in figure 5 support many of the statements written out above: (i) IP is a very bad approximation even in the low β regime; (ii) BA always outperforms TAP; (iii) SM is better than BA only in the low β regime, but here the error is dominated by the uncertainty in the correlation and increases with the system size, so the improvement of SM over BA is tiny; (iv) errors in TAP and SM diverge for large β, while those in IP and BA remain finite, although very large; (v) the normalization

J. Stat. Mech. (2012) P08015

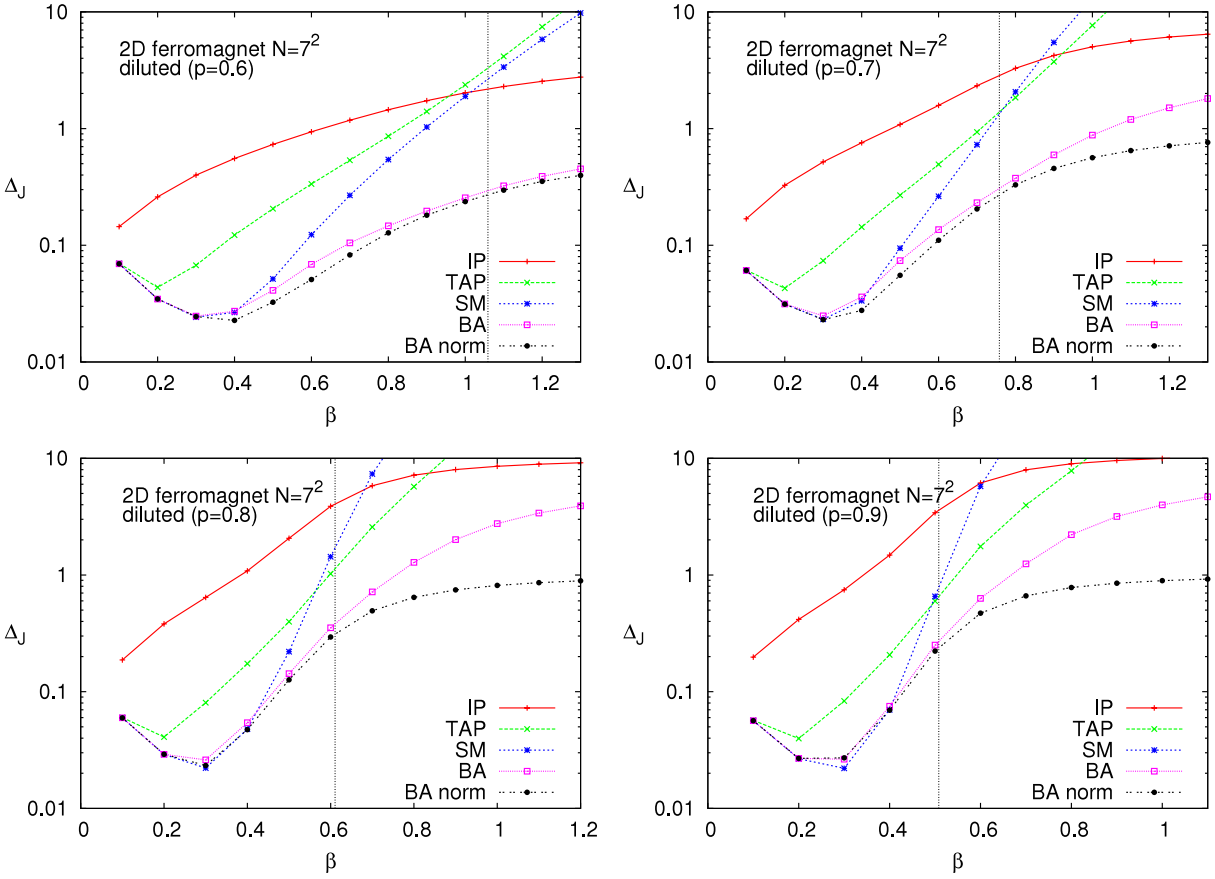

Figure 6. Errors in the couplings inferred from several approximations. The model is a diluted ferromagnet on a 2D square lattice of size $N = 7^2$. Shown are typical samples for four different dilutions: 0.6, 0.7, 0.8 and 0.9. The vertical dotted lines mark the loci of the ferromagnetic phase transitions in the thermodynamic limit.

trick works nicely and gives actually the best result over a wide range of temperatures. The data for TAP with the normalization trick (labeled 'TAP norm') are interrupted because at large β the iterative procedure that I am using for finding the parameters $\{\lambda_i\}$ stops converging.

The same qualitative conclusions can be reached by studying a diluted ferromagnet on a 2D square lattice for several different dilutions (see figure 6). In particular the relative quality of the approximations seems to be independent of the dilution, and the BA with the normalization trick outperforms the other inference methods. However I notice that, while the error of BA (with and without the normalization trick) at the critical temperature is roughly independent of the dilution, the errors made by TAP and SM tend to increase when the dilution is stronger and the system becomes more heterogeneous.

All the conclusions reached for the 2D case apply perfectly also to the case of a diluted ferromagnetic model on a 3D cubic lattice (see figure 7). So it is very reasonable to conclude that the statements made at the beginning of this section apply to any diluted ferromagnet independently of the specific topology.

J. Stat. Mech. (2012) P08015

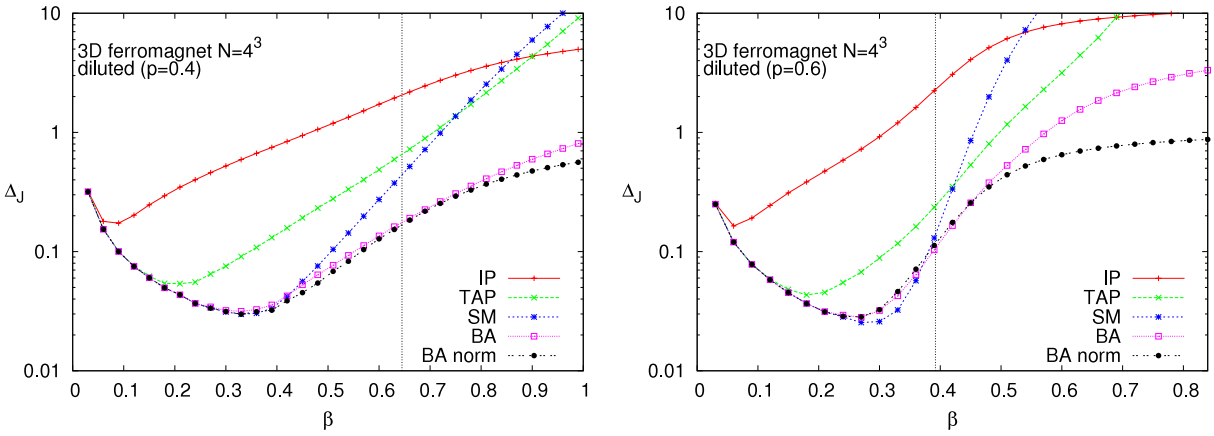

Figure 7. Errors in the couplings inferred from several approximations. The model is a diluted ferromagnet on a 3D cubic lattice of size $N = 4^3$. Shown are two typical samples with different dilutions. The vertical dotted lines mark the loci of the ferromagnetic phase transitions in the thermodynamic limit.

5.2. Spin glass models

Also for spin glass models, the conclusions reached above apply very well: the relative quality of the various approximations is roughly unchanged. The only major difference is that the normalization trick no longer provides a clear improvement: its performances strongly depend on the disorder sample. In figure 8 I report the error on the inferred coupling for spin glass models on a random regular graph with fixed degree $c = 4$, on a 3D cubic lattice and on a 2D square lattice. Again, as regards suggesting a general purpose inference method, the favored choice is clearly the BA.

Please note that it would be impossible to obtain the results shown in figure 8 by running SuscProp, because of the limited range of convergence of such an algorithm. Indeed for spin glass models on a random graph, SuscProp stops converging around the critical temperature and for spin glass models on a regular lattice it stops converging even before, well into the high temperature phase: for example for the spin glass model on a 2D square lattice it converges up to $\beta_{\rm BP} \simeq 0.66$ [25]. In this sense, the use of the new formula in equation (34) is really innovative.

5.3. Spin glass models with an external field

Let me finally come to the most surprising case: frustrated models in the presence of an external field. As shown in figure 9, once more the relative levels of accuracy of the four approximations tested are the same, but there is a major difference with respect to case of zero field (which has been reported in the upper left panel of figure 9 for the reader's convenience). At high enough β values, both TAP and BA cease to have a solution to equations (31) and (32) or equivalently the expressions under the square roots in equations (33) and (34) become negative. This fact has never, to my knowledge, been noticed in the past, although the TAP approximation for inferring couplings is greatly used.

In the appendix I sketch the analytical solution for the simplest model showing this phenomenon, namely a system of three spins connected by antiferromagnetic couplings in the presence of an external field: such an analytical solution should convince the reader

J. Stat. Mech. (2012) P08015

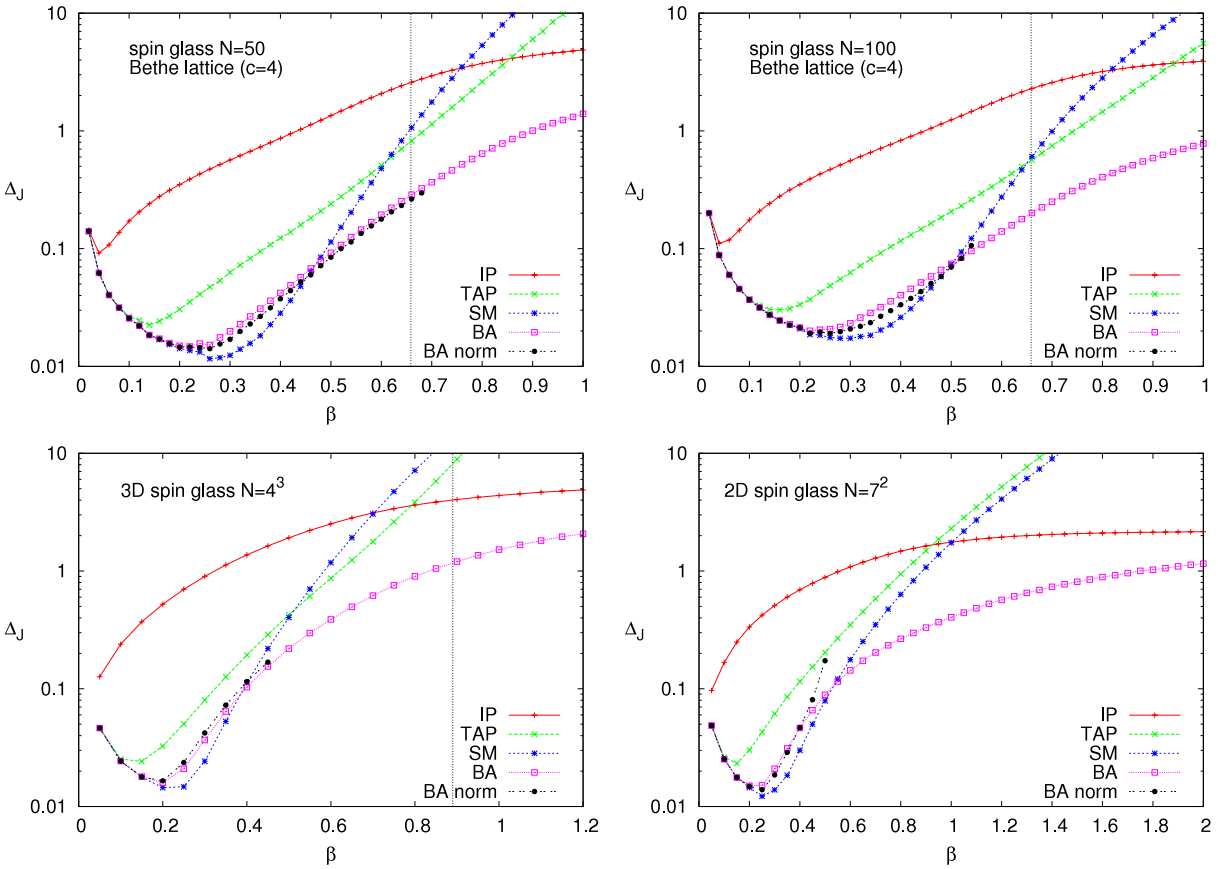

Figure 8. Errors in the couplings inferred from several approximations for a typical sample of spin glass models ($J_{ij} = \pm\beta$) on different topologies: random regular graphs with fixed degree 4 (upper panels), and 3D cubic and 2D square lattices (lower panels). The vertical dotted lines mark the loci of the spin glass phase transitions in the thermodynamic limit.

that the phenomenon is not due to any numerical inaccuracy related to the complexity of the models studied here, but it can be mathematically proved in a very simple model.

The absence of solutions in TAP and BA is evident in figure 9 where the corresponding curves are interrupted at a β value that becomes smaller for larger fields. Beyond the point where BA stops providing inferred couplings, one should resort to other inference methods. Unfortunately at that point both SM and IP already give quite large errors, that keep growing fast. So in practice none of the methods studied in the present work are valid for inferring couplings in a frustrated model in the presence of an external field at low enough temperatures.

6. Conclusions

The purpose of the present work is to make a detailed comparison among several approximations for solving the inverse Ising problem, i.e., estimating couplings and fields from magnetizations and correlations.

After having explained how to derive the mean-field approximations based on the Plefka expansion (the naive mean-field one, TAP, Bethe approximations, etc), I ranked

J. Stat. Mech. (2012) P08015

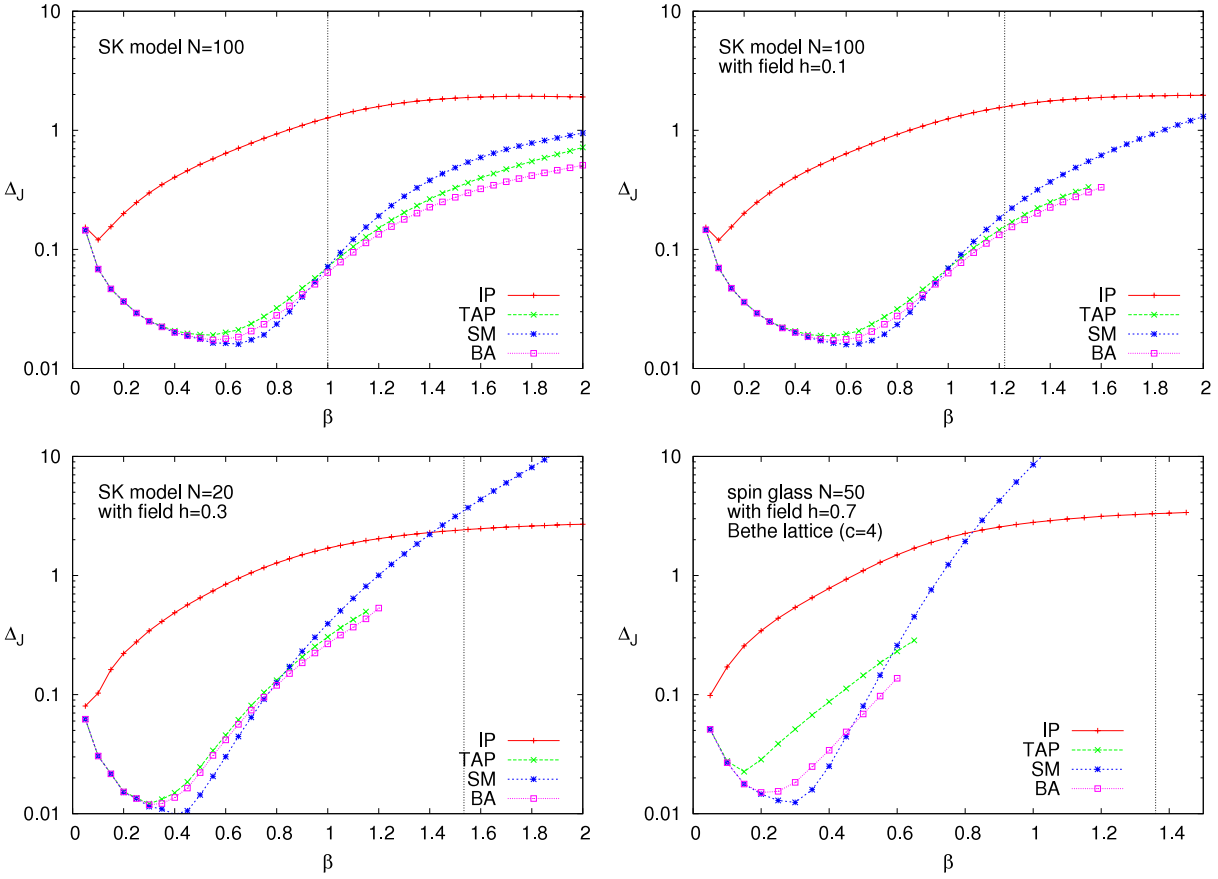

Figure 9. Errors in the couplings inferred from several approximations: the SK model with $h = 0$ (upper left, just for comparison), the SK model with $h = 0.1$ (upper right), the SK model with $h = 0.3$ (lower left) and the spin glass on a random regular graph with fixed degree $c = 4$ and $h = 0.7$ (lower right). The vertical dotted lines mark the loci of the spin glass phase transitions in the thermodynamic limit. In the presence of an external field, TAP and BA cease to have a solution for high enough β values.

these approximations on the basis of how good they are at solving the direct problem (i.e., computing the correlations given the couplings). TAP and Bethe turned out to be in general the best approximations available.

Secondly I derived the new analytical formula (34) for inferring couplings from magnetizations and correlations under the Bethe approximation. This formula allows one to infer couplings without running the susceptibility propagation algorithm, thus avoiding all the serious problems related to the lack of convergence of such an algorithm.

After having summarized the formulas giving the inferred couplings for the four approximations tested (the independent-pair one, TAP, the Bethe one and the small correlation expansion of [16]), I introduced a trick that, normalizing the correlation matrix, improves the TAP and the Bethe approximations in the case of models being unfrustrated or weakly frustrated.

Finally I presented the results of the comparison among the four approximations for inferring couplings in diluted ferromagnetic models and spin glass models. I have

J. Stat. Mech. (2012) P08015

used several different topologies: fully connected graphs, regular random graphs, 3D cubic lattices and 2D square lattices.

At the beginning of section 5 a list of general statements about the performances of these approximations in solving the inverse Ising problem is given. The bottom-line suggestion is to use the Bethe approximation, equation (34), eventually with the normalization trick if the model is weakly frustrated or unfrustrated.

In the case of frustrated models with an external field (that is with non-zero magnetizations) I have found an important limitation for the TAP and Bethe approximations: at low enough temperatures these approximations stop having a solution and can no longer be used for solving the inverse Ising problem. This is a fundamental limitation, that arises also for very simple systems (see the Appendix) and that was not noticed before.

Moreover, when the Bethe approximation stops inferring couplings, the other methods already have a rather large error. So, in my opinion, it is still an open problem to find an approximation that, using only the correlation matrix, is able to solve the inverse Ising problem in a frustrated model with a field at low enough temperatures.

Acknowledgments

I acknowledge interesting conversations with A Decelle, S Franz, A Pelizzola, I Perez-Castillo and J Raymond, and financial support by the Italian Research Minister through the FIRB Project No. RBFR086NN1 on 'Inference and Optimization in Complex Systems: From the Thermodynamics of Spin Glasses to Message Passing Algorithms'.

After the completion of the present manuscript I learnt of a manuscript by H C Nguyen and J Berg [31] showing a formula similar to equation (34).

Appendix. Limits of TAP and BA inference methods for a frustrated model in a field

In this appendix I show explicitly that the formulas derived with TAP and BA for solving the inverse Ising problem do not always admit a solution for the case of frustrated models. In order to simplify the computation I focus on the simplest model showing this problem, namely a system of three spins interacting with antiferromagnetic couplings of intensity $J < 0$, in the presence of an external field of intensity h, whose probability distribution is

$$P(s_1, s_2, s_3) \propto \exp[J(s_1 s_2 + s_2 s_3 + s_3 s_1) + h(s_1 + s_2 + s_3)].$$

Thanks to the symmetries in the above measure, each spin has the same local magnetization $m(J, h)$ and each pair of spins has the same correlation $c(J, h)$.

When using the TAP approximation for the inverse problem one has to solve the following equation for each coupling J_{ij}:

$$2m_i m_j J_{ij}^2 + J_{ij} + (C^{-1})_{ij} = 0,$$

and the above equation admits a solution only if its discriminant is non-negative:

$$\Delta^{\mathrm{TAP}} \equiv 1 - 8m_i m_j (C^{-1})_{ij} \geq 0. \tag{A.1}$$

In the present case, the discriminant is the same for each coupling and it is a function of the two parameters J and h, that I report schematically in figure A.1. The full curve shown

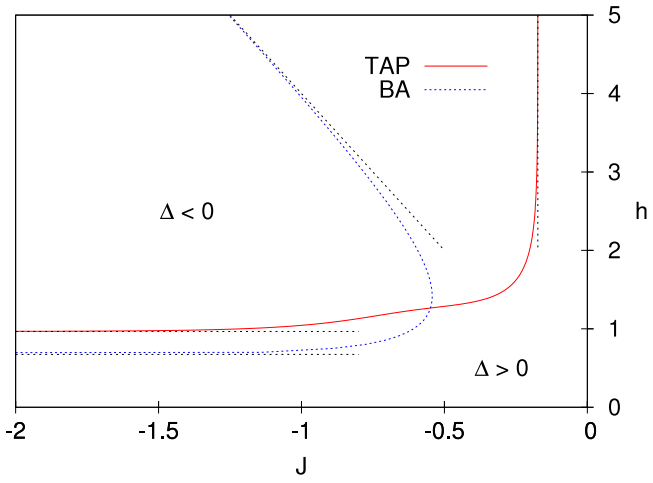

Figure A.1. Limit of validity of the TAP inference method for a system of three spins interacting with an antiferromagnetic coupling J in an external field h. In the region where the discriminant Δ is negative, the TAP inference method does not work.

in figure A.1 corresponds to $\Delta^{\mathrm{TAP}}(J,h)=0$ and has two asymptotes at $h^*=0.966\,759\ldots$ and $J^*=-\ln(2)/4$. It is clear that for any non-zero field h and any antiferromagnetic coupling J the inference method based on the TAP approximation will fail at sufficiently small temperatures (i.e., large absolute values of h and J).

The same phenomenon arises also for the inference method based on the BA. In this case the discriminant that may become negative is

$$\Delta^{\mathrm{BA}}=\left(\sqrt{1+4(1-m_i^2)(1-m_j^2)(C^{-1})_{ij}^2}-2m_im_j(C^{-1})_{ij}\right)^2-4(C^{-1})_{ij}^2. \tag{A.2}$$

In figure A.1 the dashed line corresponds to $\Delta^{\mathrm{BA}}(J,h)=0$ and has two asymptotes at $h^*=0.673\,689$ and along the line $h=-4J$ (meaning that for this simple system of three spins the BA can work even at very low temperatures if the external field is large enough).

References

[1] Bethe H A, 1935 *Proc. R. Soc.* **150** 552
[2] Mézard M and Parisi G, 2001 *Eur. Phys. J.* B **20** 217
Mézard M and Parisi G, 2003 *J. Stat. Phys.* **111** 1
[3] Mézard M and Montanari A, 2009 *Information, Physics, and Computation* (Oxford: Oxford University Press)
[4] Weigt M, White R A, Szurmant H, Hoch J A and Hwa T, 2009 *Proc. Nat. Acad. Sci.* **106** 67
[5] Cocco S, Leibler S and Monasson R, 2009 *Proc. Nat. Acad. Sci.* **106** 14058
[6] Bailly-Bechet M, Braunstein A, Pagnani A, Weigt M and Zecchina R, 2010 *BMC Bioinformat.* **11** 355
[7] Roudi Y, Aurell E and Hertz J A, 2009 *Front. Comput. Neurosci.* **3** 22
[8] Kappen H J and Rodriguez F B, 1998 *Neural Comput.* **20** 1137
[9] Mézard M and Mora T, 2009 *J. Physiol. Paris* **103** 107
[10] Marinari E and Van Kerrebroeck V, 2010 *J. Stat. Mech.* P02008
[11] Aurell E, Ollion C and Roudi Y, 2010 *Eur. Phys. J.* B **77** 587
[12] Welling M and Teh Y W, 2003 *Artif. Intell.* **143** 19
[13] Plefka T, 1982 *J. Phys. A: Math. Gen.* **15** 1971
[14] Georges A and Yedidia J S, 1991 *J. Phys. A: Math. Gen.* **24** 2173

J. Stat. Mech. (2012) P08015

[15] Tanaka T, 1998 *Phys. Rev.* E **58** 2302
[16] Sessak V and Monasson R, 2009 *J. Phys. A: Math. Gen.* **42** 0055001
[17] Ravikumar P, Wainwright M J and Lafferty J D, 2010 *Ann. Stat.* **38** 1287
[18] Bento J and Montanari A, 2009 arXiv:0910.5761v1
[19] Aurell E and Ekeberg M, 2011 arXiv:1107.3536v2
[20] Cocco S and Monasson R, 2011 *Phys. Rev. Lett.* **106** 090601
[21] Yedidia J S, Freeman W T and Weiss Y, 2005 *IEEE Trans. Inf. Theory* **51** 2282
[22] Montanari A and Rizzo T, 2005 *J. Stat. Mech.* P10011
[23] Mooij J M and Kappen H J, 2007 *J. Mach. Learn. Res.* **8** 1113
[24] Gomez V, Kappen H J and Chertkov M, 2010 *J. Mach. Learn. Res.* **11** 1273
[25] Domínguez E, Lage-Castellanos A, Mulet R, Ricci-Tersenghi F and Rizzo T, 2011 *J. Stat. Mech.* P12007
[26] Thouless D J, Anderson P W and Palmer R G, 1977 *Phil. Mag.* **35** 593
[27] Roudi Y, Tyrcha J and Hertz J, 2009 *Phys. Rev.* E **79** 051915
[28] Pearl J, 1988 *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference* (San Francisco, CA: Morgan Kaufmann)
[29] Yuille A L, 2002 *Neural Comput.* **14** 1691
[30] Heskes T, Albers C and Kappen H, 2003 *Proc. Conf. on Uncertainty in Artificial Intelligence* ed C Meek and U Kjaerulff (San Francisco, CA: Morgan Kaufmann) p 313
[31] Nguyen H C and Berg J, 2012 *J. Stat. Mech.* P03004

J. Stat. Mech. (2012) P08015