Mean-Field Theory for the Inverse Ising Problem at Low Temperatures

H. Chau Nguyen* and Johannes Berg†

Institute for Theoretical Physics, University of Cologne, Zülpicher Straße 77, 50937 Köln, Germany

(Received 8 May 2012; published 1 August 2012)

The large amounts of data from molecular biology and neuroscience have lead to a renewed interest in the inverse Ising problem: how to reconstruct parameters of the Ising model (couplings between spins and external fields) from a number of spin configurations sampled from the Boltzmann measure. To invert the relationship between model parameters and observables (magnetizations and correlations), mean-field approximations are often used, allowing the determination of model parameters from data. However, all known mean-field methods fail at low temperatures with the emergence of multiple thermodynamic states. Here, we show how clustering spin configurations can approximate these thermodynamic states and how mean-field methods applied to thermodynamic states allow an efficient reconstruction of Ising models also at low temperatures.

DOI: 10.1103/PhysRevLett.109.050602 PACS numbers: 05.10.−a, 02.30.Zz, 05.50.+q, 89.75.−k

Taking a set of spin configurations sampled from the equilibrium distribution of an Ising model, can the underlying couplings between spins be reconstructed from a large number of such samples? This inverse Ising problem is a paradigmatic inverse problem with applications in neural biology [1,2], protein structure determination [3], and gene expression analysis [4]. Typically a large number of spins (representing the states of neurons, genetic loci, or genes) is involved, as well as a large number of interactions between them.

Such large system sizes makes the inverse Ising model intrinsically difficult: solving the inverse problem involves first solving the Ising model, in some manner, for a given set of couplings and external fields. Then one can ask how couplings between spins and external fields need to be adjusted in order to match the inferred model with the observed statistics of the samples. An early and fundamental approach to the inverse Ising model, Boltzmann machine learning [5], follows this prescription quite literally. Proceeding iteratively, couplings and fields are updated in proportion to the differences of magnetizations and two-point correlations resulting from the current model parameters and the corresponding values observed in data. To compute the magnetizations and two-point correlations, each iteration involves a numerical simulation of the Ising model, so this approach is limited to small systems.

Instead, mean-field theory is the basis of many approaches to the inverse Ising problem used in practice [6,7]. Under the mean-field approximation, the Ising model can be solved easily for the magnetizations and correlations between spins. The mean-field solution is then inverted (see below) to yield the parameters of the model (couplings and external fields) as a function of the empirical observables (magnetizations and correlations). Yet, as temperature is decreased and correlations between spins grow and become more discernible, the reconstruction given by mean-field theory becomes less accurate not, as one might expect, more accurate. This effect has been called "an embarrassment to statistical physics" [8]. Mean-field reconstruction of the Ising model even breaks down entirely near the transition to a low-temperature phase [9]: in the low-temperature phase there is no correlation between reconstructed and underlying couplings. This low-temperature failure equally affects all refinements related to mean-field theory like the Thouless-Anderson-Palmer (TAP) approach [6,7,9], susceptibility propagation [9,10], the Sessak-Monasson expansion [11], and Bethe reconstruction [12].

The breakdown of mean-field reconstructions can have different roots: the emergence of multiple thermodynamic states at a phase transition, an increasing correlation length at lower temperatures, or the freezing of the spins into a reduced set of configurations at low temperatures requiring more samples to measure the correlations between spins. To address this issue, we first consider a very simple case where mean-field theory is exact: the Curie-Weiss model. The zero-field Hamiltonian of N binary spins s_i is $\mathcal{H}_J(\mathbf{s}) = -J/N\sum_{i<j} s_i s_j$ with $J = 1$. This corresponds to equal couplings $J_{ij}^0 = J/N$ between all pairs of spins, a fact that is of course not known when reconstructing the couplings. M samples of spin configurations are taken from the equilibrium measure $\exp\{-\beta\mathcal{H}_J(\mathbf{s})\}/Z$, where β is the inverse temperature and Z is the partition function. In a real-life reconstruction, these configurations would come from experimental measurements of neural activity, gene expression levels, etc. One then can calculate the observed magnetizations $\bar{m}_i = \frac{1}{M}\sum_\mu s_i^\mu$ and connected correlations $\bar{c}_{ik} = \frac{1}{M}\sum_\mu s_i^\mu s_k^\mu - \bar{m}_i\bar{m}_k$, with $\mu = 1, \ldots, M$ denoting the sampled configurations.

The mean-field prediction for the magnetizations of the Curie-Weiss model is given by the solution of the self-consistent equation

$$m_i = \tanh\Big(\sum_{j\neq i} J_{ij} m_j + h_i\Big), \tag{1}$$

where the couplings are rescaled with temperature $J_{ij} = \beta J^0_{ij}$. The connected correlations follow from Eq. (1) by considering the linear response

$$c_{ik} = \frac{\partial m_i}{\partial h_k} = (1 - m_i^2)\Big(\sum_{j\neq i} J_{ij}\frac{\partial m_j}{\partial h_k} + \delta_{ik}\Big)$$
$$= (1 - m_i^2)\Big(\sum_{j\neq i} J_{ij} c_{jk} + \delta_{ik}\Big). \tag{2}$$

Inserting the observed magnetizations and correlations into Eq. (2) gives [6]

$$\sum_{j\neq i} J_{ij}\bar{c}_{jk} = -\delta_{ik} + \bar{c}_{ik}/(1 - \bar{m}_i^2), \tag{3}$$

which can be solved directly for the couplings $J_{ij} = -(\bar{c}^{-1})_{ij}$ $(i \neq j)$ and the fields $h_i = \operatorname{arctanh}\bar{m}_i - \sum_{j\neq i} J_{ij}\bar{m}_j$ using Eq. (1).

Figure 1(a) shows how well this reconstruction performs at different inverse temperatures β and different number of samples M. For $\beta < \beta_c = 1$, the reconstruction error goes to zero with the number of samples as $M^{-1/2}$: since for the Curie-Weiss model the self-consistent equation (1)

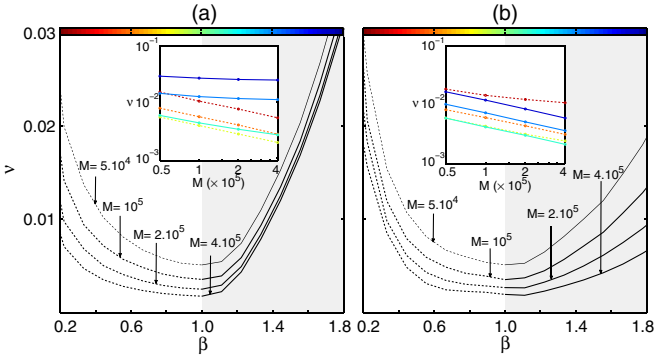

FIG. 1 (color online). Reconstructing couplings of the Curie-Weiss ferromagnet. The root-mean-squared deviation between the reconstructed couplings and underlying couplings, $\nu = \sqrt{\frac{2}{N(N-1)}\sum_{i<j}(J_{ij}/\beta - J/N)^2}$, is plotted against the inverse temperature β for different numbers of configurations ($J = 1$). The system size is $N = 100$. The insets show this deviation on a logarithmic scale versus the number of samples M at different inverse temperatures indicated by the colors of the curves ($\beta = 0.3, 0.58, 0.86, 1.14, 1.42, 1.7$). (a) Reconstruction based on a single thermodynamic state breaks down in the low-temperature phase $\beta > 1$, and the deviation between reconstructed and underlying couplings does not vanish with an increasing number of samples M (see the top curves in the inset). (b) Reconstruction based on two thermodynamic states is asymptotically exact. The deviation between reconstructed and underlying couplings vanishes as $M^{-1/2}$ at low temperatures.

is exact, the reconstruction is limited only by fluctuations of the measured correlations resulting from the finite number of samples and by the finite system size.

Yet for $\beta > \beta_c$, the difference between the underlying couplings and the reconstructed couplings does not vanish with increasing number of samples. While the self-consistent equation (1) is still correct, the identification of its solutions with the observed magnetizations $\bar{m}_i$ is mistaken. For the ferromagnetic phase at $\beta > \beta_c$, there are two solutions of the self-consistent equation, denoted $m_i^{\pm} = \pm m$. The observed magnetizations are averages over these two thermodynamic states, and they have nothing to do with either of the two solutions of Eq. (1). The same holds for the connected correlations c_{ij}^{+} and c_{ij}^{-} in the two states and the observed correlations $\bar{c}_{ij}$. Any method explicitly or implicitly connecting the magnetization in low-temperature states with the average magnetization over samples will thus fail at low temperatures. Note that this does not affect Boltzmann machine learning, where the magnetization is averaged over all states.

A simple cure suggests itself: since each sample stems from one of the two thermodynamic states, we divide the M configurations into those configurations with positive total magnetization $\sum_i s_i^{\mu}$ and those with negative total magnetization. Then the magnetizations in the two thermodynamic states can be calculated separately, giving $\bar{m}_i^{+} = \frac{1}{M_+}\sum_{\mu\in+} s_i^{\mu}$ and similarly for $\bar{m}_i^{-}$ and the connected correlations. Identifying these magnetizations with the solutions of the self-consistent equation (1), we obtain in place of Eq. (3) two sets of equations:

$$\sum_{j\neq i} J_{ij}\bar{c}_{jk}^{+} = -\delta_{ik} + \bar{c}_{ik}^{+}/[1 - (\bar{m}_i^{+})^2], \tag{4}$$

$$\sum_{j\neq i} J_{ij}\bar{c}_{jk}^{-} = -\delta_{ik} + \bar{c}_{ik}^{-}/[1 - (\bar{m}_i^{-})^2]. \tag{5}$$

Reconstructing the couplings using a single state only, by solving say Eq. (4), the observed positive magnetization can be accounted for equally well by positive external fields (even though the samples were generated by a model with zero field) or, alternatively, by ferromagnetic couplings between the spins. One finds that solving Eq. (4) leads to an underestimate of the couplings, and positive external fields calculated by Eq. (1) follow. Correspondingly, basing the reconstruction only on data from the down state by solving Eq. (5) also leads to an underestimate of the couplings and large negative fields. This effect has already been noted in the context of the inverse Hopfield problem [13]. We thus demand that the reconstructed fields obtained from either state are equal to each other,

$$\sum_{j\neq i} J_{ij}(\bar{m}_j^{+} - \bar{m}_j^{-}) = \operatorname{arctanh}\bar{m}_i^{+} - \operatorname{arctanh}\bar{m}_i^{-}, \tag{6}$$

and claim that jointly solving Eqs. (4)–(6) gives the correct mean-field reconstruction at low temperatures.

Already Eqs. (4) and (5) are two linear equations per coupling variable, so in general, there is no solution to these equations. However, we expect that the underlying couplings used to generate the M configurations actually solve these equations, at least up to fluctuations due to the finite number of configurations sampled and the finite size effect. For an overdetermined linear equation of the form $\mathbf{A}\cdot\mathbf{x}=\mathbf{b}$ with vectors of different dimensions $\mathbf{x}$ and $\mathbf{b}$ and a nonsquare matrix $\mathbf{A}$, the Moore-Penrose pseudoinverse $\mathbf{A}^+$ [14,15] gives a least-squares solution $\mathbf{x}=\mathbf{A}^+\cdot\mathbf{b}$ such that the Euclidean norm $\|\mathbf{A}\cdot\mathbf{x}-\mathbf{b}\|_2$ is minimized. In this sense, the Moore-Penrose pseudoinverse allows us to solve Eqs. (4)–(6) as well as possible. The linear equations (4)–(6) can be written as a single matrix equation $\mathbf{J}\cdot\mathbf{A}=\mathbf{B}$, where $\mathbf{A}$ is the $N\times(2N+1)$ matrix $(\bar{\mathbf{c}}^+,\bar{\mathbf{c}}^-,\bar{\mathbf{m}}^+-\bar{\mathbf{m}}^-)$ and $\mathbf{B}$ is the $N\times(2N+1)$ matrix $(\bar{\mathbf{b}}^+,\bar{\mathbf{b}}^-,\tilde{\mathbf{m}}^+-\tilde{\mathbf{m}}^-)$, with $\bar{b}^+_{ij}=-\delta_{ij}+\bar{c}^+_{ij}/[1-(\bar{m}^+_i)^2]$ and analogously for $\bar{b}^-_{ij}$, and $\tilde{m}^+_i=\operatorname{arctanh}\bar{m}^+_i$ and analogously for $\tilde{m}^-_i$. The Moore-Penrose inverse is calculated using singular value decomposition [16] and right multiplied with $\mathbf{B}$ to obtain the the optimal solution $\mathbf{J}$. In general, this matrix will not be symmetric, and we use $(J_{ij}+J_{ji})/2$, $i\neq j$ for the reconstructed couplings. The external fields can be computed for each state from $h^+_i=\operatorname{arctanh}\bar{m}^+_i-\sum_{j\neq i}J_{ij}\bar{m}^+_j$ and analogously for h^-_i. Their average over the two states is used for the reconstructed fields.

Figure 1(b) shows how the reconstruction error now vanishes as $M^{-1/2}$ also in the ferromagnetic phase, albeit with a prefactor which grows as the temperature decreases. So while the mean-field reconstruction from many samples is still successful at low temperatures, more configurations are needed to obtain a certain reconstruction error: at very low temperatures, most spins will be in the same state (either up or down); the connected correlations are small as a result and require many samples for their accurate determination. The quality of the reconstruction depends on configurations being correctly assigned to the thermodynamic states. Artificially introdcuing mistakes in this assignment, we find the reconstruction error increases linearly with the fraction of mistakes in the assignment of configurations to states.

In practice, couplings between spins will not all be equal to each other as they are in the Curie-Weiss model. Ferromagnetic as well as antiferromagnetic couplings may be present in magnetic alloys, neurons have excitatory and inhibitory interactions, regulatory interactions between genes can either enhance or suppress the expression of a target gene. The Curie-Weiss ferromagnet is not a good model for all those cases where the couplings are of different signs and magnitudes. In fact, in models where all spins interact with each other via couplings that can be positive or negative [17], the low-temperature regime may be characterized not by two but by many thermodynamic states [18,19]. These so-called glassy states cannot be identified simply by the total magnetization of each sample, as is the case for the ferromagnet. Nevertheless, configurations μ, μ' from the same thermodynamic state are typically close to each other, having a large overlap $(1/N)\sum_i s^\mu_i s^{\mu'}_i$. Glassy thermodynamic states thus appear as clusters in the space of configurations [20,21].

We use the k-means clustering algorithm [22] to find these clusters in the sampled spin configurations. Starting with a set of randomly chosen and normalized cluster centers, each configuration is assigned to the cluster center it has the largest overlap with. Then the cluster centers are moved to lie in the direction of the center of mass of all configurations assigned to that cluster, and the procedure is repeated until convergence. We also tried out different algorithms from the family of hierarchical clustering methods but found no significant difference in the reconstruction performance. Then, magnetizations and connected correlations are computed for each cluster separately. Equations (4)–(6) can be written for k thermodynamic states. The mean-field equation for each state and the condition that the external fields are equal in all states can be written again in the form of a matrix equation $\mathbf{J}\cdot\mathbf{A}=\mathbf{B}$. $\mathbf{A}$ is the $N\times(kN+k)$ matrix $(\bar{\mathbf{c}}^1,\ldots,\bar{\mathbf{c}}^k,\bar{\mathbf{m}}^1-\langle\bar{\mathbf{m}}\rangle,\ldots,\bar{\mathbf{m}}^k-\langle\bar{\mathbf{m}}\rangle)$, where $\langle\cdot\rangle$ denotes the average over clusters, $\langle\bar{\mathbf{m}}\rangle=(1/k)\sum_{a=1}^k\bar{\mathbf{m}}^{\mathbf{a}}$, and analogously for $\mathbf{B}$. The pseudoinverse of $\mathbf{A}$ can be computed in $\mathcal{O}(kN^3)$ steps [16], so up to a factor of k coming from the number of clusters this is just as fast as the high-temperature mean-field reconstruction based on Gaussian elimination to invert the correlation matrix.

We test this approach using couplings drawn independently from a Gaussian distribution of zero mean and variance $1/N$ (the Sherrington-Kirkpatrick model [17]). Figure 2(a) shows the reconstruction at low temperatures improving with the number of clusters k and configuration samples M. We note that at high temperatures, magnetizations are $0\pm\mathcal{O}(N^{-1/2})$, so for small system sizes clustering erroneously identifies distinct clusters with small magnetization. Thus, at high temperatures the low-temperature reconstruction based on many clusters does not work as well as the standard approach based on a single cluster.

A further improvement is possible. For disordered systems, the self-consistent equation (1) is not exact. An additional term is required, the so-called Onsager reaction term describing the effect a spin has on itself via the response of its neighboring spins. The TAP equation [18],

$$m_i=\tanh\Big[\sum_{j\neq i}J_{ij}m_j-m_i\sum_{j\neq i}J^2_{ij}(1-m^2_j)+h_i\Big], \qquad (7)$$

turns out to be exact for fully connected models. For each state a we now obtain instead of Eq. (4)

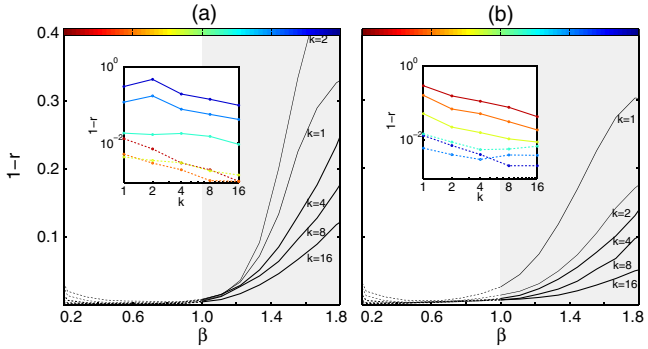

FIG. 2 (color online). Reconstructing couplings of the Sherrington-Kirkpatrick model. The Pearson correlation coefficient r quantifies the correlation between reconstructed couplings and underlying couplings, $r = \frac{\frac{1}{N(N-1)}\sum_{i\neq j}(J_{ij}-\langle J\rangle)(J^0_{ij}-\langle J^0\rangle)}{\sqrt{\mathrm{var}(J)\mathrm{var}(J^0)}}$, where $\langle J\rangle$, $\mathrm{var}(J)$ are the mean and variance of the reconstructed couplings across bonds and similarly for the underlying couplings. $r = 1$ or $1 - r = 0$ corresponds to perfect reconstruction. The main plots show $1 - r$ against the inverse temperature β for different numbers of clusters k. The insets show how $1 - r$ depends on the number of clusters k at different inverse temperatures indicated by the colors of the curves ($\beta = 0.3$, 0.58, 0.86, 1.14, 1.42, 1.7). The numbers of samples M are scaled with the numbers of clusters $M = k \times 5 \times 10^4$ to ensure a constant average number of states per cluster. The system size is $N = 100$. (a) Reconstruction based on mean-field approximation. (b) Reconstruction based on the TAP approximation with gradient descent.

$$\sum_{j\neq i} J_{ij}\bar{c}^a_{jk} = -\delta_{ik} + \bar{c}^a_{ik}/[1-(\bar{m}^a_i)^2] + \bar{c}^a_{ik}\sum_{j\neq i} J^2_{ij}[1-(\bar{m}^a_j)^2] - 2\bar{m}^a_i\sum_{j\neq i} J^2_{ij}\bar{m}^a_j\bar{c}^a_{jk}. \quad (8)$$

These equations are no longer linear in the couplings J_{ij} and cannot be solved by the pseudoinverse. A simple gradient descent method still allows us to solve these equations in $\mathcal{O}(kN^3)$ steps per iteration. We define a quadratic cost function S for the couplings $\mathbf{J}$ by squaring the difference between the lhs and rhs of Eq. (8) and summing over all spin pairs i, k and states a. Differences in the external fields $h^a_i = \mathrm{arctanh}\bar{m}^a_i - \sum_{j\neq i} J_{ij}\bar{m}^a_j + \bar{m}^a_i\sum_{j\neq i} J^2_{ij}[1-(\bar{m}^a_j)^2]$ across thermodynamic states are penalized by an additional term $\sum_{i,a}(h^a_i - \langle h_i\rangle)^2$. The iterative prescription with rate η, $J_{ij} \leftarrow J_{ij} - \eta\partial S/\partial J_{ij}$, converges to a point near the solution of the TAP equation with small differences in the external fields across states (the deviations resulting from the finite number of samples and finite system size). Figure 2(b) shows how the reconstruction error asymptotically tends to zero with growing k and M.

Mean-field theories exist beyond the Curie-Weiss or the Sherrington-Kirkpatrick model discussed here [23]. We have shown that the use of mean-field methods to solve the inverse Ising problem at low temperatures hinges on our ability to reconstruct the thermodynamic states from the sampled data. With this proviso, the entire range of mean-field methods can be now be used, for instance, for treelike couplings [12] or couplings with local loops [24].

We placed our focus on mean-field approaches, since they result in computationally efficient reconstructions independently of the underlying model (for instance a full connectivity matrix J_{ij} versus a sparse matrix). Reconstructions based on pseudolikelihood [25] can fail at low temperatures as well [26], although [27] finds a good reconstruction for several models also at low temperatures, albeit at a large computational cost. The adaptive cluster expansion recently introduced by Cocco and Monasson [28] is not affected by the transition to a low-temperature phase but becomes computationally unwieldy for highly connected models due to the large number of clusters to be considered.

We thank Erik Aurell, Filippos Klironomos, and Nico Riedel for discussions. Funding by the DFG under SFB 680 and BCGS is acknowledged.

*cnguyen@thp.uni-koeln.de
†berg@thp.uni-koeln.de

[1] E. Schneidman, M. J. Berry, II, R. Segev, and W. Bialek, Nature (London) **440**, 1007 (2006).
[2] S. Cocco, S. Leibler, and R. Monasson, Proc. Natl. Acad. Sci. U.S.A. **106**, 14058 (2009).
[3] M. Weigt, R. A. White, H. Szurmant, J. A. Hoch, and T. Hwa, Proc. Natl. Acad. Sci. U.S.A. **106**, 67 (2009).
[4] M. Bailly-Bechet, A. Braunstein, A. Pagnani, M. Weigt, and R. Zecchina, BMC Bioinf. **11**, 355 (2010).
[5] D. Ackley, G. Hinton, and T. Sejnowski, Cogn. Sci. **9**, 147 (1985).
[6] H. J. Kappen and F. B. Rodríguez, Neural Comput. **10**, 1137 (1998).
[7] Y. Roudi, E. Aurell, and J. A. Hertz, Front. Comput. Neurosci. **3**, 22 (2009).
[8] E. Aurell (private communication).
[9] M. Mézard and T. Mora, J. Physiol. (Paris) **103**, 107 (2009).
[10] M. Welling and Y. W. Teh, Neural Comput. **16**, 197 (2004).
[11] V. Sessak and R. Monasson, J. Phys. A **42**, 055001 (2009).
[12] H. C. Nguyen and J. Berg, J. Stat. Mech. (2012) P03004; F. Ricci-Tersenghi, arXiv:1112.4814.
[13] A. Braunstein, A. Ramezanpour, R. Zecchina, and P. Zhang, Phys. Rev. E **83**, 056114 (2011).
[14] E. H. Moore, Bull. Am. Math. Soc. **26**, 394 (1920).
[15] R. Penrose, Math. Proc. Cambridge Philos. Soc. **51**, 406 (1955).
[16] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in C++* (Cambridge University Press, Cambridge, 2002).
[17] D. Sherrington and S. Kirkpatrick, Phys. Rev. Lett. **35**, 1792 (1975).

[18] D. J. Thouless, P. W. Anderson, and R. G. Palmer, Philos. Mag. **35**, 593 (1977).
[19] M. Mézard, G. Parisi, and M. A. Virasoro, *Spin Glass Theory and Beyond* (World Scientific Publishing, Singapore, 1987).
[20] G. Hed, A. K. Hartmann, D. Stauffer, and E. Domany, Phys. Rev. Lett. **86**, 3148 (2001).
[21] F. Krazakala, A. Montanari, F. Ricci-Tersenghi, G. Semerjian, and L. Zdeborová, Proc. Natl. Acad. Sci. U.S.A. **104**, 10318 (2007).
[22] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006).
[23] *Advanced Mean-Field Methods: Theory and Practice*, edited by M. Opper and D. Saad (The MIT Press, Cambridge, 2001).
[24] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[25] P. Ravikumar, M. J. Wainwright, and J. D. Lafferty, Ann. Stat. **38**, 1287 (2010).
[26] J. Bento and A. Montanari, in *Advances in Neural Information Processing Systems 22* (Curran Associates, Red Hook, NY, 2009).
[27] E. Aurell and M. Ekeberg, Phys. Rev. Lett. **108**, 090201 (2012).
[28] S. Cocco and R. Monasson, Phys. Rev. Lett. **106**, 090601 (2011).