Mean-field method with correlations determined by linear response

Jack Raymond

Dipartimento di Fisica, Università La Sapienza, Piazzale Aldo Moro 5, I-00185 Rome, Italy

Federico Ricci-Tersenghi

Dipartimento di Fisica, Università La Sapienza, Piazzale Aldo Moro 5, I-00185 Rome, Italy
and INFN–Sezione di Roma 1 and CNR–IPCF, UOS di Roma, I-00185 Rome, Italy

We introduce a mean-field approximation based on the reconciliation of maximum entropy and linear response for correlations in the cluster variation method. Within a general formalism that includes previous mean-field methods, we derive formulas improving on, e.g., the Bethe approximation and the Sessak-Monasson result at high temperature. Applying the method to direct and inverse Ising problems, we find improvements over standard implementations.

I. INTRODUCTION: CLUSTER VARIATIONAL AND REGION BASED METHODS

The cluster variational method (CVM) is a unifying framework for many approximation methods on graphical models, with variational parameters in correspondence with the marginal probability distributions one is interested in Refs. [1–3]. Fast and provably convergent methods are known for the minimization of the CVM free energy [4], and systematic expansion methods about minima have been shown [5,6]. Generalizations and convex approximations to CVM have allowed for the development of fast and secure inference methods [7,8]. A common practice is to use linear response (LR) to improve the correlation estimates [9–13]. Another important application of LR has been in inverse problems, the methods have been applied, e.g., to infer protein folding structures and information processing in the retina [14,15].

We develop an extension of the standard method for fixing parameters in CVM that allows the marginal probabilities to be made consistent with the LR estimates. Our model improves over standard implementations on arbitrary graphs for high temperature. From the Bethe approximation, we recover the Sessak-Monasson expression for correlation estimation [16] from a variational framework, and with an alternative CVM approximation we improve upon the formula. We apply the method to homogeneous lattice models and demonstrate improvements with respect to the standard implementation. We also apply the method to the inverse problem of estimating couplings given correlations, demonstrating results superior to the best mean-field methods for a range of temperatures. For brevity we focus only on binary variables (spins), pairwise interactions, and three standard region selection rules, but the principle we outline is flexible with respect to these criteria. The framework offers many avenues for improvement, e.g., many of the extensions outlined in introductory comments can be directly incorporated.

A paradigmatic problem in physics is the determination of the thermodynamics and marginal probabilities of a system with N spins $\{\sigma_i = \pm 1\}$ with a Hamiltonian determined by external fields H_i and symmetric pair couplings ($J_{ii} = 0$

and $J_{ij} = J_{ji}$)

$$\mathcal{H}(\sigma) = -\sum_i H_i \sigma_i - \sum_{i<j} J_{ij} \sigma_i \sigma_j. \tag{1}$$

The free energy

$$\beta F(H,J) = -\log \mathrm{Tr}\left[\exp\left(-\beta \mathcal{H}(\sigma)\right)\right] \tag{2}$$

is in most cases computational intractable for moderate system sizes and the analytic solution is unknown in the large N limit. $\mathrm{Tr}[\cdot]$ denotes a summation over all spin variables in the expression. The cluster variational method (CVM) offers insight into approximations [1,2]: the free-energy functional

$$F_{\mathrm{CVM}}(b,J,H) = E(b,J,H) - \frac{1}{\beta} S(b) \tag{3}$$

is the sum of energetic part

$$E(b,J,H) = -\sum_{i<j} J_{ij} \mathrm{Tr}\left[b_{ij}\sigma_i\sigma_j\right] - \sum_i H_i \mathrm{Tr}\left[b_i\sigma_i\right] \tag{4}$$

and entropic part

$$S(b) = -\sum_R c_R \mathrm{Tr}\left[b_R \log b_R\right], \tag{5}$$

where R are subsets of variables, c_R are integer counting or Möbius numbers, σ_R denotes the set of variables $\{\sigma_i : i \in R\}$, and $b_R(\sigma_R)$ are beliefs over the variables in R (arguments omitted for brevity). Selecting a set of regions that forms a junction tree (e.g., selecting all possible regions is sufficient) the approximation is exact, Eq. (2) becomes equivalent to (3), when the beliefs are equal to the marginal probability distributions.

A complete parametrization of the beliefs in terms of connected correlation parameters C is

$$b_R(\sigma_R) = \frac{1}{2^{|R|}} \left[1 + \sum_{p \in P(R)} \prod_{s \in p} C_s \prod_{i \in s} \sigma_i \right], \tag{6}$$

where $P(R)$ are nonempty partitions over R and s are the elements (subsets) in each partition. Explicit examples for small subsets are given in Appendix A, in the CVM framework any correlation parameter C_s is zero unless $s \in R$ for some

region R in the approximation. The beliefs are by this choice normalized and share parameters so as to be consistent on all marginals. We can interpret b_R as locally consistent probabilities provided $0 \leqslant b_R(\sigma_R) \leqslant 1$.

The beliefs are, in a standard implementation, fixed by minimizing the free-energy subject to local consistency requirements (maximum entropy). The beliefs determined in this way equal the marginal probabilities only for the special case of a junction tree. There are two possible sources of error in the approximate case: (A) region selection, (B) marginal distributions.

A junction tree is composed of large regions in many cases, making calculation of (5) impractical even for known marginals. A compromise is to select only a subset of small regions. For graphs of special topology, or where out-of-region correlations are weak, these approximations are good, or occasionally asymptotically exact, and b_R are close to the true marginal probabilities.

A hierarchy of mean-field approximations are recovered from CVM: The naive mean-field (NMF) approximation is achieved by selecting single-variable regions, since (i,j) are not contained in any regions $C_{ij} = 0$ and $b_{ij} = b_i b_j$ for purposes of evaluating the energy (4), and (5) simplifies to

$$S_N(C) = -\sum_i \text{Tr}\,[b_i \log b_i]. \qquad (7)$$

Note that in this expression, and henceforth, we write the variational dependence as C rather than b, and all beliefs are functions of C through (6). The Bethe approximation includes NMF regions and adds one pair region for each nonzero coupling ($J_{ij} \neq 0$). The entropy approximation is $S_B = S_N + \Delta S_B$, where the correction to the entropy is

$$\Delta S_B(C) = -\sum_{ij:J_{ij} \neq 0} \text{Tr}\left[b_{ij} \log\left(\frac{b_{ij}}{b_i b_j}\right)\right]. \qquad (8)$$

Assuming small pair correlations we could consider an expansion of (8) to quadratic order in the pair correlation parameters; from such an approach one can derive the TAP equations. One possibility beyond Bethe (that we call the P_X approximation) includes all plaquette regions up to maximum size X. A plaquette P is a closed loop of coupled variables, without chords—an ordered set of $|P|$ variables $(i_1, \ldots, i_{|P|})$ such that $J_{i_{x-1}, i_x} \neq 0$ (allowing $i_0 = i_{|P|}$). The entropy is $S_{P_X} = S_B + \Delta S_{P_X}$, with

$$\Delta S_{P_X}(C) = -\sum_P \text{Tr}\left[b_P \log\left(\frac{b_P \prod_{x=1}^{|P|} b_{i_x}}{\prod_{x=1}^{|P|} b_{i_{x-1}, i_x}}\right)\right], \qquad (9)$$

assuming plaquettes overlap on at most one edge.

A choice is made to approximate the entropy by a particular region selection, and the parameters (constrained beliefs, or equivalently correlation parameters) are fixed by minimizing the free energy by

$$\left.\frac{\partial F_{\text{CVM}}}{\partial C_s}\right|_{C=C^*} = 0, \qquad (10)$$

where the Hessian should also be positive definite. Linear response (LR) about this minimum then approximates the

connected correlation on subset s as

$$\chi_s = \left.\frac{\partial^{|s|} F_{\text{CVM}}(C)}{\prod_{i \in s} \partial H_i}\right|_{C=C^*}. \qquad (11)$$

Except for cases where the free energy is exact the parameters C_s^* and LR estimates χ_s disagree, on subsets s of size $|s| > 1$. By a simple argument [17] it is expected that parameters fixed by the saddle-point criteria (10) will be poorer estimates to the true correlations than those estimated by LR about the saddle point (11). CVM is fundamentally a variational method, inducing an approximation to the probability measure $P(\sigma) = P_{\text{exact}}(\sigma) + \epsilon\,\delta P(\sigma)$. Since $O(\epsilon^d)$ errors arise in quantity determined by d^{th} derivatives of the free energy, the correlation parameters determined by first derivative conditions (10) are of lower fidelity than the LR estimates obtained by higher-order derivatives (11). However, by interpreting χ as the true correlations we implicitly accept the inaccuracy of b_R as the correct marginal probabilities at the saddle point.

In our method we require consistency between the two correlation estimates over a set Ω

$$C_s^* = \chi_s : s \in \Omega. \qquad (12)$$

The elements of Ω are subsets in at least two indices, and these subsets must all be contained in some region forming the CVM approximation.

If the approximation is NMF, consisting only of vertex regions, the set Ω must be empty and so we do not change the standard approximation. For purposes of this article we consider two simple modifications for the Bethe and Plaquette approximations: For the direct problem the set Ω is all edge regions with nonzero J_{ij} ($\Omega = \{(i,j) : J_{ij} \neq 0\}$); for the inverse problem the set Ω is of maximum size (all possible subsets of regions in the approximation without repetitions, $\Omega = \cup_R \{s : s \subset R, |s| \geqslant 2\}$). For the Bethe approximation both definitions of Ω are equivalent, and for the Plaquette approximation we do not constrain three-point (and higher-order) connected correlations in the direct problem, whereas we do for the inverse problem. For the inverse problem we certainly want χ_s to equal C_s for all s, and we want to use the data to fix both correlations to the same value. For the direct problem implementation of the plaquette method we do not include the constraint on three-point correlations for several reasons as follows: brevity in explanation, technical convenience (it proves much simpler to fix three-point correlations by maximum entropy than by the linear response identity), and the intuition that three-point correlations will be less significant than two-point correlations.

To implement correlation constraints we invoke a modified entropy approximation (5), with slack parameters λ

$$S_\lambda(C) = S(C) - \sum_{s \in \Omega} \lambda_s C_s. \qquad (13)$$

Choosing λ to satisfy (10) leaves $C_\Omega = \{C_s : s \in \Omega\}$ unconstrained by the saddle-point equations, such that we can fix it self-consistently from (11) and (12). λ may be large for poor region selection and $\lambda = 0$ when CVM is exact.

In Appendix I we reformulate the free energy as a CVM approximation with modifications of the fields and couplings and additional variational parameters, this provides some

additional insight into the complexity relative to a standard CVM implementation of the modification (13).

II. A GENERAL FRAMEWORK FOR LINEAR RESPONSE

In our method a minimum of the free energy is first determined. The saddle-point equations, $\partial F_{\text{CVM}}/\partial C_i$, are

$$0 = -\beta H_i - \sum_j \beta J_{ij} C_j + \text{atanh}(C_i) + L_i(C),$$

$$L_i(C) = \sum_{R(\neq i):i\in R} c_R \text{Tr}\left[\frac{\sigma_i}{2} b_{R\backslash i} \log\left(\frac{b_R}{b_i}\right)\right], \quad (14)$$

where $b_{R\backslash i}$ is the belief over the region R excluding i. Equation (14) includes the NMF result and the correction $L_i(C)$. The saddle-point equations, $\partial F_{\text{CVM}}/\partial C_s$ for larger sets $|s| > 1$, are

$$\beta J_s - \lambda_s = \sum_{R:s\in R} c_R \text{Tr}\left[\prod_{i\in s}\left(\frac{\sigma_i}{2}\right) b_{R\backslash s} \log(b_R)\right], \quad (15)$$

with $J_s = 0$ for sets $|s| > 2$ and $\lambda_s = 0$ if $s \notin \Omega$. Equations (14) and (15) fix a subset of parameters C, λ: in our method we fix λ and $C \setminus C_\Omega$, whereas the standard approach fixes all of C (given $\lambda_s = 0, \forall s$). More details of the derivation of the saddle-point equations (14) and (15) are in Appendix B.

Going forward, we assume that this is a well-defined minimum in the variational parameters (C): differentiable and not on the boundary of the feasible parameter space. In response to a variation in the fields $H_z \to H_z + \delta H_z$ the parameters are perturbed $C_s \to C_s^* + \delta C_s$, and a quadratic order free energy describes the fluctuation

$$F_{\text{CVM}}(C^*) + \sum_i (C_i^* + \delta C_i)\delta H_i + \sum_{s,s'} \delta C_s Q_{s,s'} \delta C_{s'}/2. \quad (16)$$

In this expansion we treat $\partial\lambda/\partial H$ as zero, i.e., λ is not a variational parameter. The saddle-point equations are then

$$\sum_s Q_{i,s}\delta C_s = \delta H_i; \quad \forall i \quad (17)$$

and

$$\sum_s Q_{s',s}\delta C_s = 0; \quad \forall s' : |s'| > 1. \quad (18)$$

Solving this system of linear equations in δC_i, and identifying $\delta C_i = \beta \sum_j \chi_{ij}\delta H_j$ as linear responses, we can write a system of N^2 equations,

$$[\chi^{-1}]_{i,j} = -\beta J_{ij} + \Phi_{i,j}(C^*). \quad (19)$$

The detailed relationships between the Hessian Q, Φ and some other linear response identities are described in Appendix C.

For our choice of Ω in the direct problem, Eqs. (12), (14), (15), and (19) form a closed set for the determination of C and λ given J and H. Requiring consistency with respect to different choices of Ω, e.g., three-point correlations, is possible but requires higher-order derivates.

For the inverse problem we do constrain all higher-order (more than two-point) correlations in the case of the Plaquette approximation, but note that, for any CVM approximation, Eqs. (12) and (14) and the off-diagonal component of (19)

already form a closed set of equations for the determination of H and J given C. We need not know either λ or the structure of the equation determining the higher-order responses in order to complete the inference of J and H.

III. COMPARISON OF METHODS

Equation (19) applies for any λ value. Standard ($\lambda = 0$) LR results are reproduced by solving (14), (15), and (19) without requiring (12) [10,18].

In our approach, beginning from the Bethe approximation, we determine the on-diagonal elements in (19) to be

$$\Phi_{ii}^B = \frac{1}{1 - C_i^2}\left[1 + \sum_{j:ij\in\Omega} \frac{C_{ij}^2}{(1 - C_i^2)(1 - C_j^2) - C_{ij}^2}\right]. \quad (20)$$

The entropic off-diagonal matrix components, for edge regions $\{ij : J_{ij} \neq 0\}$, are

$$\Phi_{ij}^B = J^{\text{IP}}(C_{ij}, C_i, C_j) - \frac{C_{ij}}{(1 - C_i^2)(1 - C_j^2) - C_{ij}^2}, \quad (21)$$

where other off-diagonal components are zero. J^{IP} is the independent pair approximation

$$J^{\text{IP}}(C_{ij}, C_i, C_j) = \text{Tr}\left[\frac{\sigma_i\sigma_j}{4} \log b_{ij}\right]. \quad (22)$$

More details are provided in Appendix D. Using Φ_{ij}^B in (19) the Sessak-Monasson [16] result for small correlation parameters is recovered. For the plaquette method we have corrections $\Phi^P = \Phi^B + \Delta\Phi^P$. In the simplest case with triangular plaquettes, and spin-symmetric probabilities ($C_i = 0, C_{ijk} = 0$), we have a correction to the nonzero off-diagonal elements given by

$$\Delta\Phi_{ij}^{P_3} = \sum_{k(\neq i,j)} c_{ijk}\left\{\frac{1}{4}\sum_{a=\pm 1} a \log\left[1 - \frac{(C_{jk} + aC_{ik})^2}{(1 + aC_{ij})^2}\right]\right.$$
$$\left. + \frac{(C_{ik} - C_{jk}C_{ij})(C_{jk} - C_{ik}C_{ij})}{(1 - C_{ij}^2)(1 - C_{jk}^2 - C_{ik}^2 + 2C_{jk}C_{ik}C_{ij} - C_{ij}^2)}\right\}, \quad (23)$$

where $c_{ijk} = 1(0)$ for included (excluded) plaquettes, as described in Appendix E.

A. High-temperature expansions

For a fully connected model we can consider the leading order errors in the high-temperature regime by an expansion in β, as described in Appendix G. For the symmetric case, an approximation inclusive of all edge regions yields a correction in Φ_{ij},

$$\Phi_{ij}^B - \Phi_{ij}^{\text{exact}} = -\beta^5 \sum_{k(\neq i,j)} 2J_{ij}J_{ik}^2 J_{jk}^2 + O(\beta^6), \quad (24)$$

to be compared to the $O(\beta^4)$ error in standard Bethe. For the Plaquette approximation, including all triplet regions, the error is improved to $O(\beta^6)$. In the special case of zero external field

(a symmetric solution) it is

$$
\begin{aligned}
\Phi_{ij}^{P_3} - \Phi_{ij}^{\text{exact}} = \beta^7 \sum_{k<l(\neq i,j)} & 2J_{kl}[J_{ij}J_{kl}(J_{ik}J_{jl} + J_{il}J_{jk})^2 \\
& + 2J_{ik}J_{jk}J_{il}J_{jl}(J_{ij}J_{kl} + J_{ik}J_{jl} + J_{il}J_{jk})] \\
& + O(\beta^8),
\end{aligned}
\tag{25}
$$

to be compared to $O(\beta^5)$ errors for the standard LR implementation. The errors in (24) and (25) are evaluated given the exact value of C, which is pertinent to an idealized inverse problem application. If instead we consider the direct problem, we are more interested in errors on the statistics C; these errors depends on the on-diagonal component error of Φ, which is unimproved in our method. In Ref. [19] an analysis and remedy is proposed that involves including *on-diagonal* constraints, as discussed in Appendix H.

B. Direct problem

The direct problem of determining magnetizations $\{C_i\}$ and correlations (e.g., $\{C_{ij}\}$) given H,J requires the simultaneous solution to (12), (14), (15), and (19). A possible iterative scheme for the NMF, Bethe, and Plaquette approximations is

$$
C_i^{t+1} \leftarrow \tanh\left[\beta\left(H_i + \sum_j J_{ij}C_j^t\right) - L_i(C^t)\right], \tag{26}
$$

$$
C_{ij}^{t+1} \leftarrow \chi_{ij}^t = \{[-\beta J + \Phi(C^t)]^{-1}\}_{ij}, \tag{27}
$$

$$
b_P^t \leftarrow \text{argmin}\left\{\text{Tr}\left[b_P^t \log b_P^t\right] \big| \{C_i^t\}, C_\Omega^t\right\}. \tag{28}
$$

All approximations (NMF, Bethe, and P_X) require (26), only the Bethe and P_X approximations ($\lambda \neq 0$) require (27), and only the P_X approximations require (28), where b_P^t is the belief parameterized by the correlations $\{C_s^t : s \in P\}$. Thus, (28) assigns the maximum entropy estimate to all connected correlations on a plaquette region P not fixed by (26) and (27). Equations (28) is an easily solved local convex optimization, subject to linear constraints determined by C_Ω^t and $\{C_i^t\}$. At sufficiently high temperature the scheme is convergent and the solution stable. However, at lower temperatures the process may be unconvergent, and so strong damping and/or special update ordering is required. We describe in more detail solving the equations for the special case of homogeneous solutions on a lattice in Appendix F.

We find our method to be promising for models with many short loops. The homogeneous triangular lattice model (HTL), with $H_i = 0$, $J_{ij} = 1$ for nearest neighbors and 0 otherwise, is a well-understood canonical model that has a ferromagnetic transition point at $\beta_c = 0.275$ [20] and for $\beta < 0$ is fully frustrated with no long-range order but a Kosterlitz-Thouless transition [21,22]. For finite lattice implementations we choose periodic boundary conditions (periodicity L). Figure 1 shows the corresponding region based approximations in the direct problem.

Figure 2 shows nearest-neighbor correlation estimates obtained in the thermodynamic limit by Fourier techniques, as described in Appendix F 2. We show the exact result by the black line, standard LR methods in red (label χ), standard methods minimizing F in the variational parameters in green
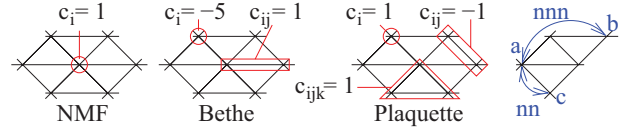

FIG. 1. (Color online) Regions and counting numbers for a triangular lattice. $\{a,c\}$ are nearest neighbors (nn), and $\{a,b\}$ are next-nearest neighbors (nnn). We abbreviate $\{\chi_{\text{nn}}, C_{\text{nn}}, \lambda_{\text{nn}}\}$ and χ_{nnn} for the corresponding homogeneous quantities.

(label C), and our method in blue. All methods perform well at high temperature (small $|\beta|$), and magnetized solutions are accurate for $\beta \gg \beta_c$. Standard methods undergo spurious continuous transitions for $\beta \lesssim \beta_c$, and the NMF and Plaquette approximations also undergo a transition in the frustrated regime $\beta < 0$. LR estimates diverge at these spurious critical points. The standard ($\lambda = 0$) P_3 method performs well in the estimate of C_{nn} (nearest-neighbor correlation), for the unmagnetized solutions, but only in the stable range $\beta \in (-1.01, 0.255)$, while our P_3 method performs well in the entire frustrated region and up to the true critical temperature $\beta < \beta_c$ (see the inset of Fig. 2). However, the unmagnetized solution does not exhibit continuous phase transitions for our methods for $\beta \sim \beta_c$, as it should. At low temperature, convergence problems hinder the construction of solutions (iteration of (27) fails due to large gradients, as shown in Fig. 5), and the unmagnetized P_3 solution is constructed only for $\beta < 0.3$. Certainly the P_3 unmagnetized solution is unfeasible for $\beta > 0.35$ (already significantly below the critical temperature), since the Hessian becomes singular for any $C_{\text{nn}} < 1$, the unmagnetized Bethe solution is stable to much larger β as shown.

Lattice models with finite L do not, strictly speaking, exhibit a phase transition, but many phenomena are well described by models with this feature. However, data collected in real applications often do not show any phase-transition phenomena [14,15]; a more general test of inference methods is the quality of the marginals predicted.
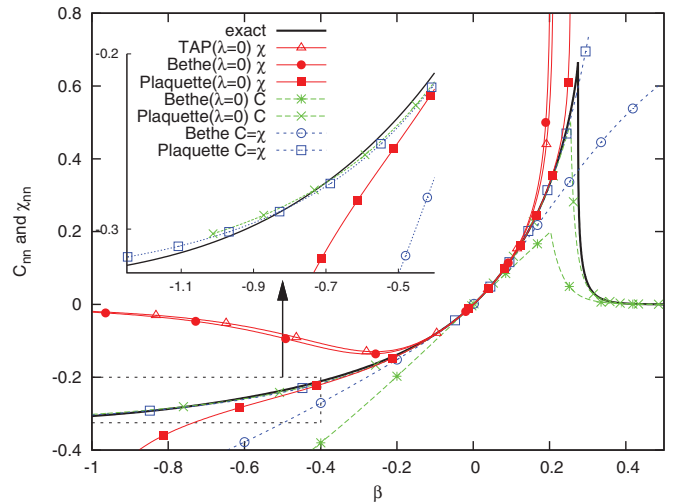

FIG. 2. (Color online) Nearest-neighbors correlation estimates for the asymptotic ($L \to \infty$) HTL. Magnetized branches are shown only for Bethe and Plaquette ($\lambda = 0, \beta > 0$) correlation parameters.
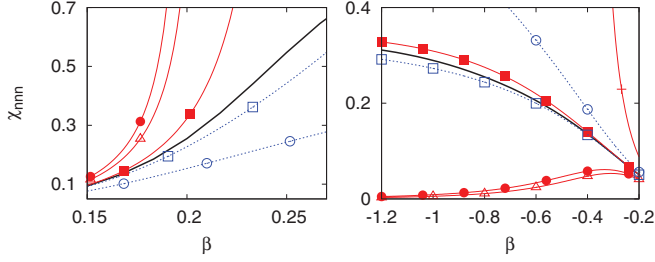
FIG. 3. (Color online) Next-nearest-neighbor correlations for the HTL $L = 5$, unmagnetized branches. Curves as Fig. 2.

Correlations that extend beyond the approximation regions are not among the CVM parameters, and LR is required to determine pair correlations at distances larger than 1. For $\beta \in (-\infty, \beta_c]$, our method improves on standard implementations for many significant terms in χ and χ^{-1}. Figure 3 shows the next-nearest correlations calculated on a finite model $L = 5$: Our method estimates are superior to their counterparts for most β. The values calculated for $L = 5$ are close to those for $L \to \infty$ for $\beta < \beta_c$, although in the case of $L = 5$ the tripartite lattice symmetry is broken so for $\beta < 0$ there is extra frustration, and the unmagnetized solutions are more stable. The NMF unmagnetized solution is unstable for $\beta < -0.382$, but other unmagnetized solutions are stable for $-1.2 < \beta$. An interesting feature of this method is that it overcompensates the error of the standard method, so a combination of the two can lead to even better results.

We can see that the LR (χ_{nn}) and max-entropy (C_{nn}) estimates are, for the Plaquette approximations, much closer, and relatively accurate, compared to those for Bethe and NMF. Correspondingly, the values of λ_{nn} are much smaller in absolute value in the Plaquette approximation as shown in Fig. 4. For the asymptotic curves ($L \to \infty$) we observe a monotonic trend in λ_{nn}, and for the finite model ($L = 5$) the effect of the boundary causes the curve to become nonmonotonic. However, note that for $L = 5$ the effective potential ($\beta - \lambda_{nn}$, discussed Appendix I) remains monotonic as a function of β for all
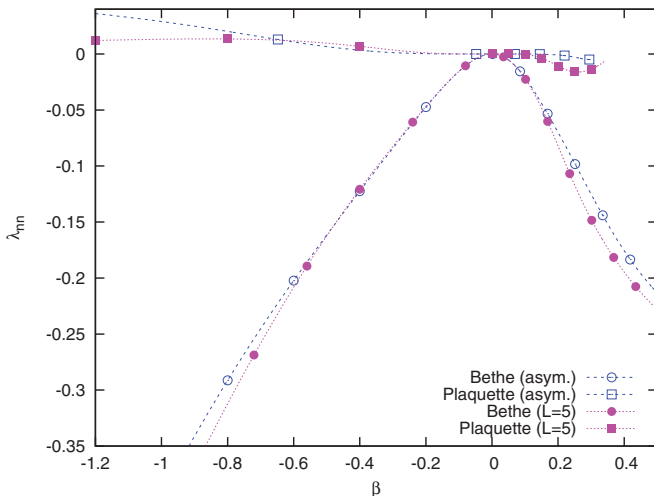

FIG. 4. (Color online) Values of λ_{nn} that achieve the curves of Figs. 2 and 3 in the Bethe and Plaquette approximations with $L = 5$ and $L \to \infty$ (asymptotic).

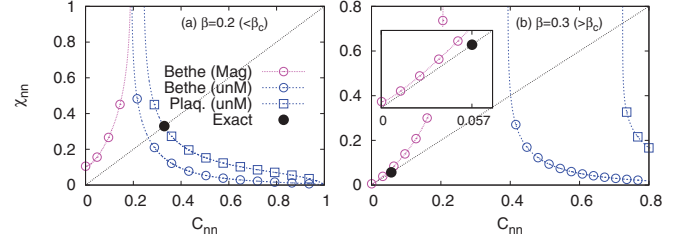FIG. 5. (Color online) The response as a function of the correlation parameter for the asymptotic ($L \to \infty$) HTL with (left) $\beta < \beta_c$ and (right) $\beta > \beta_c$. Inset (right): A magnified version of the magnetized branch. Bethe (magnetized and unmagnetized) and P_3 (unmagnetized) approximation results are shown. Our method requires the intersection point ($\chi_{nn} = C_{nn}$), to be compared with the exact solution.

solutions presented. In the ferromagnetic region ($\beta > 0$) the value of λ_{nn} is negative for both approximations, which has the effect to reduce the ferromagnetic susceptibility of the unmagnetized solution.

In all the results discussed so far we have used an annealing technique to obtain the curves, so the curves presented are those that are obtained continuously from the unique high-temperature solution to the equations. These are all unmagnetized solutions. In the zero field lattice models tested, we found that for large β either a magnetized solution appeared discontinuously or that no magnetized solution exists.

As shown in Fig. 5 we plot the curves for the linear response correlation estimate χ_{nn} against the parameter value in the asymptotic ($L \to \infty$) HTL; a similar pattern of curves applies for a variety of translationally invariant lattice models we solved. We show the magnetized branch only for the Bethe approximation, since the P_3 approximation cannot be solved asymptotically in the magnetized case by the Fourier method outlined in Appendix F 2. Our method dictates solutions according to (12), i.e., the point(s) $\chi_{nn} = C_{nn}$, which can be achieved by a choice of λ_{nn}.

For $\beta < \beta_c$ the true solution for a lattice model is unmagnetized. Our approximations also determine unique unmagnetized solutions that slightly underestimate the true value; a typical case is Fig. 5(a). For frustrated lattice models, such as the HTL with $\beta < 0$, the unmagnetized solution is also found to be unique, and the curve for $-\chi_{nn}$ versus $-C_{nn}$ is similar to the unmagnetized branch of Fig. 5 (with the magnetized branch absent).

For $\beta > \beta_c$ the true solution is magnetized and the connected correlation decreases from a peak at β_c. Our method typically exhibits an unmagnetized solution for $\beta \gtrsim \beta_c$, with large connected correlation; a typical case is 5(b). This solution can persist to very large $\beta \gg \beta_c$, as shown in Fig. 2. Alongside this, we typically find either a magnetized solution, or magnetized pseudo-solution ($\chi_{nn} \approx C_{nn}$), for which the marginals are better estimated. A typical pseudosolution behavior is shown Fig. 5[(b),inset]. We found the pseudosolution behavior to be typical of standard 2D lattice models. By contrast, moving to the 3D model, we found for $\beta \gtrsim \beta_c$ the coexistence of stable magnetized and unmagnetized solutions—the figure is similar to 5(b) except that the magnetized curve crosses at two points to give one locally stable magnetized solution alongside
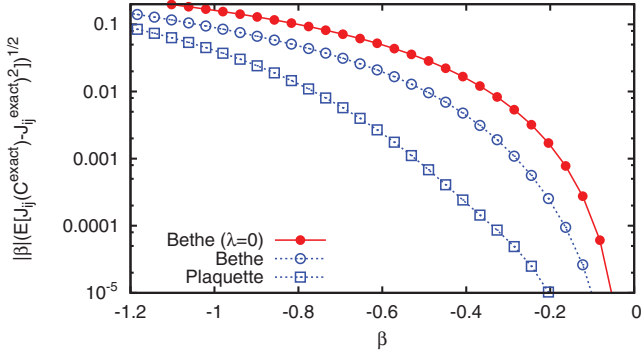
FIG. 6. (Color online) THL: error in inference of J from exact statistics.

the locally stable unmagnetized solution. A discontinuous transition from the paramagnetic to the ferromagnetic solution is apparent in these cases.

The curves of Fig. 5 also dictate the dynamics of the iterative procedure (27) with other terms [(26) and (28)] at their fixed point values. Due to the large gradient at the unmagnetized fixed point strong damping is required in the proposed iterative method to find the unmagnetized solutions for large β (large $|\beta|$ in the case of frustrated regimes). By contrast, the magnetized solution (or pseudosolution, if we allow $\chi_{nn} \sim C_{nn}$) has benign dynamics.

C. Inverse problem

A simpler application of our method is for the inverse problem: Given sample statistics, determine J and H [18,23]. With ignorance of the distribution of couplings (and topology), we must have unbiased region selection: all edges for Bethe and all (triangular) Plaquettes for P_3. In our method we take C and χ equal to the correlation statistics and solve, first, (19) in the off-diagonal elements for J_{ij} and then (14) for H. Equation (15) can be used to determine λ, which would be a measure of model fidelity. In standard mean-field methods the same assumptions are made on region selection, but only χ and $\{C_i\}$ are determined from the statistics; all other C obey the saddle-point equations (15) with $\lambda = 0$ (thus making equations solvable for Bethe and TAP [18]).

Figure 6 demonstrates the results for estimation of matrix J in the HTL $L = 5$ based on exact data. The improved scaling at small $|\beta|$ is as anticipated in Eqs. (24) and (25). However, even at low temperature, reconstruction is significantly improved by our methods. Although Φ_{nn} determines the error, note that the approximation differs from that used in the direct problem: The 2D triangular structure is discovered, unlike in the direct problem, where it is assumed in the region selection.

Figure 7 demonstrates results for an instance of a 7×7 diluted square lattice Ising model in zero field. Each coupling is assigned according to the probability distribution $P(J) = 0.7\delta_{J,1} + 0.3\delta_{J,0}$. The reconstruction assumes $H_i = 0$, but no knowledge of J. We generated the pair-correlation matrix from independent Monte Carlo measures. Sampling errors limit all methods for small β. When β is large enough, the error of the method exceeds the sampling one. A β interval exists in
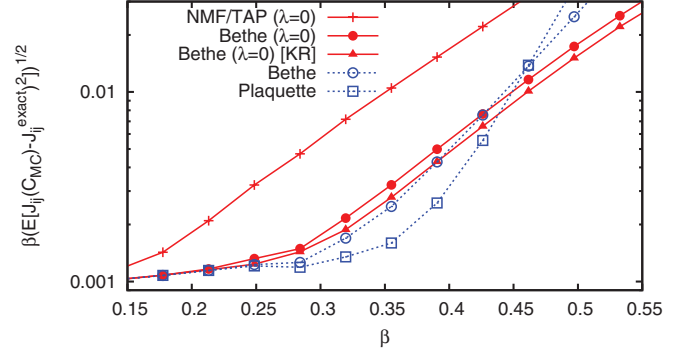

FIG. 7. (Color online) Error in inferring couplings J for a diluted 2D square ferromagnet, from statistics of 10^6 independent samples. [KR] employs the Kappen-Rodriguez normalization [10].

which our methods improve over standard ones. The triangular Plaquette approximation improves over Bethe, despite the absense of triangles in the model (the shortest loop is of length 4). For larger β the model undergoes a rapid growth in correlation length, edge and/or triangular regions are relatively small and account for only a small fraction of the significant correlations, all mean-field methods are prone to significant errors.

Since at the Bethe level our method coincides with the Sessak-Monasson expression it is not a surprise that we outperform other mean-field methods at high temperature due to the improved scaling (24). The plaquette approximation is by no means guaranteed to outperform the Sessak-Monasson expression outside the weak-coupling regime where a superior (25) scaling applies, but this advantage persists at intermediate temperatures for the two models presented. Realization of the high temperature scaling is only feasible if data are of very high quality; in practical applications this is unlikely, as sampling will be subject to errors and performance may not be significantly improved over NMF. The more interesting comparison of methods is for intermediate and low temperatures.

We studied, for example, the 2D square lattice Edward-Anderson model with zero field and $J = \pm 1$ coupling distribution across a range of system sizes $L = 4$ to $L = 32$ at intermediate temperatures $\beta \in [0.3, 0.8]$ and a small number of samples for each system size [24,25]. We found that the $\lambda \neq 0$ outperform standard implementations of Bethe and NMF in all samples; however, for nearest couplings we found that the Bethe ($\lambda \neq 0$) method provided a marginally superior estimate (i.e., estimates of J_{ij} were closer to 1 in absolute value for nearest neighbors i,j), while the $P_3 (\lambda \neq 0)$ approximation provided a marginally superior estimate for the absent couplings (i.e., estimates were closer to 0 for J_{ij}, where ij are not nearest neighbors). We will present a more detailed analysis of the inverse problem across a range of problems in a forthcoming paper.

IV. DISCUSSION AND CONCLUSION

We propose a minimal modification to the mean-field free-energy functional in order to make max-entropy estimates of correlations consistent with LR ones; in other words,

the Hessian consistent with the location of the free-energy minimum. To do this, we introduce a new set of parameters (λ) and constraints (12) and argue that this may move belief estimates closer to the true marginals and, hence, the free-energy estimate is improved. An alternative argument is that physical quantities, as measured by different combinations of derivatives of the variational free energy, should be consistent at a good evaluation point, as discussed in Appendix H.

For the direct problem, the value of λ is unique and can be found by expansion in the weak-coupling regime, but even for simple models we find that a solution meeting criteria (12) may not exist at low temperature. This is in contrast to standard implementations of CVM where minima always exist [7]. The absence of a continuous transition to a magnetized solution in some ferromagnetic lattice models might be viewed as a pathology, although the standard mean-field exponent 1/2 is itself a significant underestimation of the abruptness of these transitions. We have proposed a strict implementation of the condition (12), but by relaxing this condition slightly we might find better solutions such as the magnetized pseudosolution for the HTL. By contrast, for the inverse problem there always exists a unique set of values J,H solving (19), (14), and (12), which determines unique values λ through the saddle-point equations (15); this in an improvement relative to the standard implementation of CVM methods for the inverse problems where solutions may not exist for certain models [18].

In our framework we are not able to produce a general argument that guarantees that our variational approach produces either an upper or lower bound to the free energy (standard CVM has the same problem), and, more importantly, we cannot provide guarantees of feasibility or uniqueness of λ. Despite this, it seems that our method is very effective in both the inverse and direct problems, indicating that the constraints we are introducing are beneficial as extensions of variatonal frameworks, and we maintain the property of CVM that when the region selection is correct the exact free energy is found by minimization ($\lambda = 0$ is a solution). The framework is inclusive of previous NMF and Bethe approaches for the inverse problem, while providing a new variational basis for the Sessak-Monasson expression.

Another important set of issues are algorithmic, for example, regarding when feasible values of λ exist how can they be found and how can the free energy even be minimized given fixed λ. The constraints we introduce to the CVM are linear in the connected correlation parameters but nonlinear in the belief paramaterization as discussed in Appendix I. This restricts the class of methods available for (local or global) minimization, and we have presented a simple iterative algorithmic method which is sufficient for high temperature but leaves room for improvement. We have been able to develop message passing equations for our framework, which will be presented alongside the analysis of the direct problem in a broader class of models in a forthcoming work. Another important topic not covered in this paper is on-diagonal constraints [19,26], discussed in Appendix H. These are important for improving estimation in the direct problem, and the estimation of the magnetizations in the inverse problem and can be straightforwardly incorporated in our framework.

APPENDIX A: BELIEFS AS FUNCTIONS OF CORRELATIONS

For the derivations restricted to the NMF, Bethe, and P_3 approximations, we indicate the explicit forms for the beliefs in terms of the connected correlation parameters (6),

$$b_i = \frac{1 + C_i \sigma_i}{2}, \tag{A1}$$

$$b_{(i_1,i_2)} = b_{i_1 i_2} = \prod_{x=1}^{2} b_{i_x} + C_{i_1 i_2} \prod_{x=1}^{2} \left[\frac{\sigma_{i_x}}{2} \right], \tag{A2}$$

$$b_{(i_1,i_2,i_3)} = b_{i_1 i_2 i_3} = \prod_{x=1}^{3} b_{i_x} + C_{i_1 i_2 i_3} \prod_{x=1}^{3} \left[\frac{\sigma_{i_x}}{2} \right]$$
$$+ \sum_{x=1}^{3} C_{i_1 i_2 i_3 \backslash i_x} \left(C_{i_x} + \sigma_{i_x} \right) \prod_{y=1}^{3} \left[\frac{\sigma_{i_y}}{2} \right]. \tag{A3}$$

Note that unlike the representation of the beliefs in terms of the full correlations, the representation in terms of connected correlations is nonlinear. Linearity and convexity are not preserved properties in the new basis; in particular, the entropy term $\text{Tr}\,[b_R \log b_R]$ is not a convex function of C in general. However, since the parameters are variational, and the correlation parameters span the same space of beliefs, the free-energy estimate is unaltered.

In the CVM framework parameters are nonzero only if they are subsets of some region in the approximation. Hence, for the NMF approximation, (A2) and (A3) reduce to products of the marginal probabilities (A1).

APPENDIX B: DERIVATION OF SADDLE-POINT EQUATIONS: (14) AND (15)

The Hamiltonian of the main text (1) is a special case of

$$H = - \sum_s J_s \prod_{i \in s} \sigma_i, \tag{B1}$$

where in the main text $J_i = H_i$ (the external fields) and $J_s = 0$ for sets larger than 2 (couplings are pairwise).

The free-energy components in a generic CVM approximation become

$$E(J,H,b) = -\beta \sum_s J_s \text{Tr} \left[b_s \prod_{i \in s} \sigma_i \right] \tag{B2}$$

and

$$S(b) = \sum_R c_R \text{Tr}\,[b_R \log b_R] + \sum_{s \in \Omega} \lambda_s C_s. \tag{B3}$$

R are the regions forming the CVM approximation, s are subsets of variables, and Ω is the set for which we require $\chi_s = C_s$.

The saddle points and Hessian are determined by derivatives with respect to the variational parameter. A useful identity given our parametrization is

$$\frac{\partial b_R}{\partial C_s} = \prod_{i \in s} \left[\frac{\sigma_i}{2}\right] b_{R \setminus s}, \qquad (B4)$$

where $R \setminus s$ is the set complement. The belief over the empty set is defined to be $b_{s \setminus s} = 1$ and we define $b_{R \setminus s} = 0$ for cases where s is not entirely inside R ($s \setminus R$ is not empty).

The saddle-point equation for any correlation parameter $\partial \beta F_{CVM} / \partial C_t$ is, thus,

$$0 = -\beta \sum_{s:t \subset s} J_s \mathrm{Tr} \left[\prod_{i \in t} \left[\frac{\sigma_i}{2}\right] b_{s \setminus t} \prod_{i \in s} \sigma_i \right]$$
$$+ \sum_{R:t \subset R} c_R \mathrm{Tr} \left[\prod_{i \in t} \left[\frac{\sigma_i}{2}\right] b_{R \setminus t} \log b_R \right] + \sum_{s \in \Omega} \delta_{t,s} \lambda_s. \qquad (B5)$$

For t in at least two indices this is identical to (15) recognizing that the first term reduces to $-\beta J_{ij}$ for pairwise couplings. We arrive at (14) from (B5) employing the counting numbers identity $\sum_{R:i \in R} c_i = 1$ true for any i and $\mathrm{Tr} [\sigma_i b_{i \cup R} \log b_i] = \mathrm{Tr} [\sigma_i b_i \log b_i]$ for any set R. In this way

$$\sum_R c_R \mathrm{Tr} \left[\frac{\sigma_i}{2} b_{R \setminus i} \log b_R \right]$$
$$= \mathrm{Tr} \left[\frac{\sigma_i}{2} \log b_i \right] + \sum_{R \setminus i} c_R \mathrm{Tr} \left[\frac{\sigma_i}{2} b_{R \setminus i} \log \left(\frac{b_R}{b_i}\right) \right], \qquad (B6)$$

thereby separating the naive mean-field term for $t = i$ from corrections in (14).

APPENDIX C: DERIVATION OF HESSIAN AND Φ (16) AND (19) AND AN ALTERNATIVE RESPONSE IDENTITY

Following the derivation of the saddle-point equations, we construct the Hessian by a derivative of (B5) with respect to C_u,

$$Q_{t,u} = -\beta \sum_{s:t,u \in s} J_s \mathrm{Tr} \left[\prod_{i \in t} \left[\frac{\sigma_i}{2}\right] \prod_{i \in u} \left[\frac{\sigma_i}{2}\right] b_{(s \setminus t) \setminus u} \prod_{i \in s} \sigma_i \right]$$
$$+ \sum_{R:t,u \in R} c_R \mathrm{Tr} \left[\prod_{i \in t} \left[\frac{\sigma_i}{2}\right] \prod_{i \in u} \left[\frac{\sigma_i}{2}\right] b_{(R \setminus t) \setminus u} \log b_R \right]$$
$$+ \sum_{R:t,u \in R} c_R \mathrm{Tr} \left[\prod_{i \in t} \left[\frac{\sigma_i}{2}\right] \prod_{i \in u} \left[\frac{\sigma_i}{2}\right] \frac{b_{R \setminus t} b_{R \setminus u}}{b_R} \right]. \qquad (C1)$$

The Hessian should be positive definite; for a consistent method, this can be checked. We separate the matrix into blocks (in the case of NMF we have only the block $Q^{(1)}$ and we can take $\Phi = Q$ in that special case),

$$Q = \begin{pmatrix} Q^{(1)} & Q^{(2+,1)} \\ [Q^{(2+,1)}]^T & Q^{(2+)} \end{pmatrix}, \qquad (C2)$$

where $Q^{(1)}$ is the submatrix where both s and t are single indices, $Q^{(2+)}$ is the submatrix formed where neither s nor t

are single indices, and, naturally, $Q^{(2+,1)}$ is the case where u is a single index set and t is a multi-index set. We can define by analogy vector forms $C_1 = \{C_i\}$ and $C_{2+} = C \setminus C_1$, with associated fluctuations δC_1 and δC_{2+} about the minima of the free energy. The saddle-point equations of the main text derived from the quadratic expansion about the minima (16) can be written as vector equations,

$$Q^{(1)} \delta C_1 + [Q^{(2+,1)}]^T \delta C_{2+} = \beta \delta H, \qquad (C3)$$

$$Q^{(2+,1)} \delta C_1 + Q^{(2+)} \delta C_{2+} = 0. \qquad (C4)$$

The second equation can be solved in δC_{2+}, leaving one equation for δC_1,

$$Q^{(1)} \delta C_1 - [Q^{(2+,1)}]^T [Q^{(2+)}]^{-1} Q^{(2+,1)} \delta C_1 = \beta \delta H. \qquad (C5)$$

Since we are expanding about a proper minimum, Q is invertible, and, hence, so is the submatrix $Q^{(2+)}$. We can identify δC with linear responses; it is a sum of perturbations due to each independent field fluctuation

$$\delta C_s = \beta \sum_z \chi_{s,z} \delta H_z. \qquad (C6)$$

We denote by χ^{NN} the $N \times N$ matrix with components $d \delta C_i / d \delta H_j$, and the off-diagonal elements are sufficient to enforce the correlation constraints $C_\Omega = \{C_{(i,j)}\}$. Equation (C5) becomes

$$\{[Q^{(1)} - [Q^{(2+,1)}]^T [Q^{(2+)}]^{-1} Q^{(2+,1)}] \chi^{NN}\} \delta H = \delta H. \qquad (C7)$$

The result holds for all possible perturbations of the field, so (C7) implies

$$\{Q^{(1)} - [Q^{(2+,1)}]^T [Q^{(2+)}]^{-1} Q^{(2+,1)}\} \chi^{NN} = I, \qquad (C8)$$

where I is the identity matrix. Finally, we separate the coefficient of χ^{NN} into the energetic part and entropic part (Φ). Restricting attention to the pairwise model, the energetic term is $-\beta J$ and arises in $Q^{(1)}$, and, thus, the identification among the Hessian Q, Φ, and the coupling matrix is

$$\Phi - \beta J = \{Q^{(1)} - [Q^{(2+,1)}]^T [Q^{(2+)}]^{-1} Q^{(2+,1)}\}. \qquad (C9)$$

Since we are dealing with a minima, the covariance matrix χ is invertible, and we arrive at (19).

In the direct problem we do require knowledge of the structure of higher-order response equations. This requires expanding the free energy to third order or higher, which is quite complicated. Aside from maximum entropy, an alternative mechanism for fixing correlations is the use of alternative linear response identities. For example, the three-point connected correlation can also be defined,

$$\chi_{s,z} = \frac{\partial^2 F_{CVM}}{\partial H_z \partial \lambda_s} = \frac{\partial C_s}{\partial H_z}. \qquad (C10)$$

This quantity can be determined from (C4) and (C6) as

$$\chi_{s,z} = \{[Q^{(2+)}]^{-1} Q^{(2+,1)} \chi^{NN}\}_{s,z}. \qquad (C11)$$

Thus, for a three-point correlation, we might consider $C_{i_1 i_2 i_3} = (\chi_{i_1 i_2, i_3} + \chi_{i_1 i_3, i_2} + \chi_{i_2 i_3, i_1})/3$ in place of the standard constraint $C_{i_1 i_2 i_3} = \chi_{i_1, i_2, i_3}$.

APPENDIX D: DERIVATION OF SESSAK-MONASSON EXPRESSION: (21) AND REF. [16]

The form is found from the Bethe approximation to the entropy. For the Bethe approximations $Q^{(2+)}$ is a diagonal matrix so the inverse is simple, and $Q^{(2+,1)}$ is also sparse,

$$Q_{ij,ij}^{(2+)} = \mathrm{Tr}\left[\frac{1}{16b_{ij}}\right]; \quad Q_{ij,k}^{(2+,1)} = \mathrm{Tr}\left[\frac{\delta_{k,j}b_i\sigma_i + \delta_{k,i}b_j\sigma_j}{8b_{ij}}\right].$$

For a pairwise model,

$$Q_{i,j}^{(1)} = -\beta J_{ij} + \delta_{i,j}\left[\mathrm{Tr}\left[\frac{1}{4b_i}\right] + \sum_{k(\neq i)}c_{ik}\mathrm{Tr}\left[\frac{b_k^2}{4b_{ik}}\right]\right]$$
$$+ (1-\delta_{i,j})c_{ij}\mathrm{Tr}\left[\frac{\sigma_i\sigma_j}{4}\left\{\frac{b_ib_j}{b_{ij}} + \log b_{ij}\right\}\right], \quad \text{(D1)}$$

where $c_{ij} = 1(0)$ for included (excluded) regions. Exploiting the sparseness of matrices (C9) becomes

$$\Phi_{ij}^B = c_{ij}\mathrm{Tr}\left[\frac{\sigma_i\sigma_j}{4}\left(\frac{b_ib_j}{b_{ij}} + \log b_{ij}\right)\right] + c_{ij}\frac{Q_{ij,i}^{(2+,1)}Q_{ij,j}^{(2+,1)}}{Q_{ij,ij}^{(2+)}}. \quad \text{(D2)}$$

Thus, for included regions $c_{ij} = 1$, we recover the Sessak-Monasson expression

$$[C^{-1}]_{ij} = -J_{ij} + J^{\mathrm{IP}}(C) - \frac{C_{ij}}{(1-C_i^2)(1-C_j^2)-C_{ij}^2} \quad \text{(D3)}$$

by a combination of (12), (19), and (21); note that we need not know λ to estimate J given the correlations.

APPENDIX E: DERIVATION OF $\Delta\Phi^{P_3}$ IN THE SPIN-SYMMETRIC CASE: (23)

If we consider the special case of triangles without symmetry breaking, then $C_s = 0$ for any set of odd parity. We can rearrange the matrices $Q^{(2+)}$ and $Q^{(2+,1)}$, ordering components $Q_{s,t}$ according to the parity, $|s|$ and $|t|$ (where notation $|\cdot|$ indicates the number of elements in the set). Due to symmetry, $Q_{t,u}$, where $|t|$ and $|u|$ are of different parity, are null

$$Q^{(2+)} = \begin{pmatrix} Q^{(\mathrm{even})} & 0 \\ 0 & Q^{(\mathrm{odd})} \end{pmatrix}, \quad \text{(E1)}$$
$$Q^{(2+,1)} = (0 \quad Q^{(\mathrm{odd},1)}). \quad \text{(E2)}$$

As such, the expression (C9) in a symmetric case becomes

$$\Phi = [Q^{(1)} + \beta J] - [Q^{(\mathrm{odd},1)}]^T[Q^{(\mathrm{odd})}]^{-1}Q^{(\mathrm{odd},1)}. \quad \text{(E3)}$$

For the plaquette approximation $Q^{(\mathrm{odd})}$ is block diagonal, and for the case P_3 the matrix is diagonal; $Q^{(\mathrm{odd},1)}$ is also sparse. The relevant matrix components are

$$Q_{ijk,ijk}^{P_3} = \mathrm{Tr}\left[\frac{1}{2^6 b_{ijk}}\right],$$
$$Q_{ijk,l}^{P_3} = \mathrm{Tr}\left[\frac{\delta_{l,i}b_{jk}\sigma_j\sigma_k + \delta_{l,j}b_{ik}\sigma_i\sigma_k + \delta_{l,k}b_{ij}\sigma_i\sigma_j}{2^4 b_{ijk}}\right].$$
$$\text{(E4)}$$

The matrix $Q^{(1)}$ for a pairwise model has components

$$Q_{i,j}^{P_3} = \delta_{i,j}\left[\sum_{R:i,j\in R}c_R\frac{b_{R\setminus i}^2}{4b_R}\right] + (1-\delta_{i,j})\left[-\beta J_{i,j}\right.$$
$$\left. + \sum_{R:i,j\in R}c_R\frac{\sigma_i\sigma_j}{4}\left(b_{R\setminus i,j}\log b_R + \frac{b_{R\setminus i}b_{R\setminus j}}{b_R}\right)\right].$$

Again we exploit the sparseness of these matrices to write for the off-diagonal component

$$\Phi_{i,j}^{P_3} = Q_{i,j}^{P_3} - \sum_{k(\neq i,j)}c_{ijk}\frac{Q_{i,ijk}^{P_3}Q_{j,ijk}^{P_3}}{Q_{ijk,ijk}^{P_3}}, \quad \text{(E5)}$$

where for each included plaquette $c_{ijk} = 1$ (and zero otherwise). Taking $\Phi^{P_3} - \Phi^B$ (21) and representing the beliefs in their symmetric forms ($C_i = 0, C_{ijk} = 0$), we arrive at (23).

APPENDIX F: HOMOGENEOUS LATTICE MODEL SOLUTIONS

On a translationally invariant lattice we can exploit redundancy of the parameters to find the homogeneous solution. We can label the variables by their geometric coordinates, and we take the standard dot-product on the Euclidean vector space, $x \cdot y = \sum_i x_i y_i$.

A simple Hamiltonian for variables embedded on a hypercubic lattice $\mathcal{L}$ of dimension d is

$$\mathcal{H} = \sum_{i\in\mathcal{L}}\sum_{\xi\in\Xi}\sigma_i\sigma_{i+\xi}/2, \quad \text{(F1)}$$

where Ξ is the set of vectors describing the relative position of coupled variables. Couplings are taken to be 1 to remove clutter in notation, β controls the strength of interaction, and we allow $\beta < 0$ to describe the antiferromagnetic model.

For planar lattice models the free energy can be calculated exactly, including in the thermodynamic limit [20,21]. The energy $E = \sum_{i,j}\langle\sigma_i\sigma_j\rangle$ is related to the mean nearest-neighbor correlation $\hat{c}$ by $E = 3\hat{c}N$, forming the basis for the comparison of Fig. 2.

In the case of the triangular lattice the nearest neighbors are defined $\Xi = \{\pm(0,1),\pm(1,0),\pm(1,1)\}$ [see Fig. 8(a)].

1. Homogeneous symmetric solution

To avoid clutter we present the simplest case of the spin-symmetric solution. In this simple case, the beliefs are determined by a single parameter $C_{i,i+\xi} = \hat{c}, \forall \xi \in \Xi$, which describes the solution to (F1) in the absence of spontaneous symmetry breaking,

$$b_i^* = \frac{1}{2}, \quad \text{(F2)}$$
$$b_{ij}^* = \frac{(1+\hat{c}\sigma_i\sigma_j)}{4}, \quad \text{(F3)}$$
$$b_{ijk}^* = \frac{1+\hat{c}(\sigma_i\sigma_j + \sigma_i\sigma_k + \sigma_j\sigma_k)}{8}. \quad \text{(F4)}$$

Once Φ is computed $\hat{c}$ is determined as the fixed point of (27).
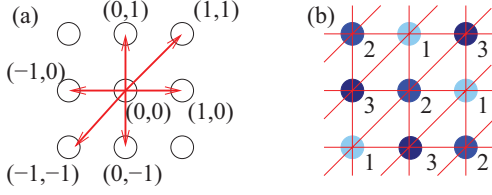
FIG. 8. (Color online) The HTL lattice model embedded on a square lattice. (a) The origin and the relative position of (coupled) nearest neighbors. The elements of the set Ξ that define the displacement of neighbors are shown in red for the HTL. (b) The interactions are tripartite: each sublattice labeled 1 to 3 in a regular pattern as shown forms a set of variables that do not self-interact (except possibly at the boundary). The susceptibility of the lattice for $\beta < 0$ is largest with respect to fluctuations breaking this symmetry.

For this solution, it is also relatively simple to identify symmetry-breaking instabilities with respect to homogeneous perturbations of $\hat{c}$ and with respect to perturbations breaking the tripartite symmetry of the HTL [see Fig. 8(b)]. For finite L, where the tripartite symmetry is broken by the boundary conditions (e.g., $L = 5$), the full Hessian must be constructed and analyzed to determine stability.

By translational invariance Φ is a function only of the displacement between its components $\Phi_{i,j} = \phi_{i-j}$. For a symmetric model the NMF solution is found from $\phi_i^N = \delta_{i,0}$. For the Bethe approximation, with edge regions $\{(i, i + \xi) : \xi \in \Xi\}$ (removing sets that differ only in ordering), the homogeneous solution is $\phi_i^B = \phi_i^N + \Delta\phi_i^B$, where

$$\Delta\phi_i^B = \delta_{i,0}\left[\sum_{\xi \in \Xi}\frac{\hat{c}^2}{1 - \hat{c}^2}\right] + \sum_{\xi \in \Xi}\delta_{i,\xi}\left[\text{atanh}(\hat{c}) - \frac{\hat{c}}{1 - \hat{c}^2}\right]. \quad \text{(F5)}$$

The P_3 approximation, which has plaquettes $\{(i, i + \xi_1, i + \xi_2) : \xi_1 + \xi_2 \in \Xi\}$ (without repetitions), is determined by

$$\phi^{P_3} = \phi^B + \Delta\phi^{P_3} \text{ with}$$

$$\Delta\phi_i^{P_3} = \delta_{i,0}\left[-\sum_{\xi_1,\xi_2,\xi_3 \in \Xi}\delta_{\xi_1+\xi_2+\xi_3,0}\frac{2\hat{c}^3}{(1 + 2\hat{c})(1 - \hat{c}^2)}\right]$$

$$+ \sum_{\xi_1,\xi_2 \in \Xi}\delta_{\xi_1+\xi_2+i,0}\left\{\frac{(\hat{c} - \hat{c}^2)^2}{(1 - \hat{c}^2)(1 - 3\hat{c}^2 + 2\hat{c}^3)}\right.$$

$$\left.+ \frac{1}{4}\log\left[1 - \frac{(4\hat{c}^2)}{(1 + \hat{c})^2}\right]\right\}. \quad \text{(F6)}$$

To determine the solutions for large $|\beta|$ in practice (27) is applied making use of damping, annealing, and symmetry.

2. The asymptotic solution

Using these approximations, we can solve the linear system of equations for lattice models

$$(-\beta J + \Phi(\hat{c}))\chi = I. \quad \text{(F7)}$$

For a translationally invariant lattice embedded in a hypercubic lattice of dimension d, we can exploit the Fourier representation of $\chi_{i,j}$

$$\chi_{i,j} = \chi_{i-j} = \int d\mu \, \tilde{\chi}(\mu)\exp[-i2\pi(i - j) \cdot \mu]. \quad \text{(F8)}$$

with the integral over the unit hypercube centered on the origin. The inverse transform is

$$\tilde{\chi}(\mu) = \sum_z \chi_z \exp(i2\pi\mu \cdot z). \quad \text{(F9)}$$

A general solution in the thermodynamic limit ($L \to \infty$) is found for (F1) as

$$\tilde{\chi}(\mu) = \frac{1}{\phi_0 + \sum_{\xi \in \Xi}[\phi_\xi - \beta]\exp(2\pi i\mu \cdot \xi)}. \quad \text{(F10)}$$

From which any element of the matrix $\chi_{a,b}$ can be constructed by the inverse discrete Fourier transform. In the standard framework ($\lambda_{nn} = 0$), we solve for $\hat{c}$ by minimization, and calculate χ once. For our method we begin with an estimate of $\hat{c}^0$ and update according to (27), and for the special case of the triangular homogeneous lattice, (F10) yields

$$\hat{c}^{t+1} = \frac{1}{3}\int_{-1/2}^{1/2}d\mu_1\int_{-1/2}^{1/2}d\mu_2\frac{\cos(2\pi\mu_1) + \cos(2\pi\mu_2) + \cos[2\pi(\mu_1 + \mu_2)]}{\phi_0(\hat{c}^t) + 2[\phi_1(\hat{c}^t) - \beta]\{\cos(2\pi\mu_1) + \cos(2\pi\mu_2) + \cos[2\pi(\mu_2 + \mu_2)]\}}. \quad \text{(F11)}$$

where due to homogeneity $\phi_\xi = \phi_1$ and

$$\phi_i = \delta_{i,0}\phi_0 + \sum_{\xi \in \Xi}\delta_{i,\xi}\phi_1. \quad \text{(F12)}$$

APPENDIX G: HIGH-TEMPERATURE EXPANSION

Loops cause the failure of the Bethe approximation, and the shortest loops contribute the leading order errors in β. To understand errors up to $O(\beta^X)$, it is sufficient to consider a diagrammatic expansion, where diagrams of size greater than X do not contribute. The free energy can be explicitly con-

structed, as such the exact connected correlations and entropy, $\Phi_{i,j}$ and L_i. This allows a comparison to our approximation which again can be determined by a diagrammatic expansion. The diagrams that contribute to the leading-order corrections in Bethe are apparent on a single triangle (fully connected graph of three variables), while for the P_3 approximation it is sufficient to consider a tetrahedron (fully connected graph of four variables). The two smallest graphs are not solved exactly by the respective approximations.

Regarding the inverse problem, in our method χ and C coincide with the data, whereas in the standard method χ and C_1 coincide with the data and C_{2+} is determined by

maximum entropy. The coupling J_{ij} is determined by (19), hence, the error arises only in the term $\Phi_{i<j}(C)$. The fields H_i are determined by (14), so the error arises in $\sum_j J_{ij} C_j + L_i$. We call C_{2+}^* the value determined by maximum entropy in the standard approach, and as a perturbation about the data (C) we can solve the linearized saddle-point equation to determine the leading-order error,

$$0 = \frac{\partial F}{\partial C_{2+}}(C) + [Q^{(2+,2+)}(C)](C_{2+}^* - C_{2+}). \quad (G1)$$

The sampling error in the data can dominate the method error in practical scenarios for both methods, and for high-fidelity data and sufficiently large β it is the errors outlined that are most significant. Our analysis assumes that the sampling error is negligible compared to the error in the entropy approximation.

In the direct problem the error on C depends in a coupled manner on the errors in all Φ, L, and the maximum entropy procedure for $C \setminus C_\Omega$. Thus, the increased accuracy of $\Phi_{i<j}$ is not necessarily realized in increased accuracy of C_{ij}, for example. A minor modification, including the on-diagonal constraints discussed in Appendix H, can ameliorate this error [19].

We demonstrate the leading-order diagrams for the saddle-point term L_i and the matrix Φ for the weak-coupling limit (J small, β finite), using notation $\doteq$ to indicate the asymptotic nature. For the weak-coupling limit we abbreviate $t_i = \tanh(\beta H_i)$ and $T_i = 1 - t_i^2$.

1. Bethe errors in the fully connected model

The error in the inference of H is limited by

$$L_i^B - L_i^{\text{exact}} \doteq 2\beta^3 t_i \sum_{j<k(\neq i)} T_j T_k J_{ij} J_{ik} J_{jk}, \quad (G2)$$

errors are $O(J^3, \beta^4)$ in the weak-coupling and high-temperature expansions respectively. The error in inference of J is limited by, for $i \neq j$,

$$\Phi_{ij}^B - \Phi_{ij}^{\text{exact}} \doteq -2\beta^4 \sum_{k(\neq i,j)} T_k (2J_{kij} + 2J_{jki} + J_{ijk}) J_{ik} J_{jk}, \quad (G3)$$

where $J_{ijk} = J_{ik} J_{jk} t_i t_j T_k$ and errors are $O(J^4, \beta^5)$. The high-temperature result is dominated by a different set of diagrams, as shown (24). The on-diagonal component error in Φ is determined by the diagrams

$$\Phi_{ii}^B - \Phi_{ii}^{\text{exact}} \doteq 2\beta^3 \sum_{j<k(\neq i)} T_j T_k J_{ij} J_{ik} J_{jk}, \quad (G4)$$

and, respectively, $O(J^3, \beta^3)$. This error is insignificant for the inverse problem and significant only for the direct problem.

If we fix correlation parameters according to maximum entropy, the standard method, new errors are introduced,

$$C_{ij}^* - C_{ij} \doteq -\beta^2 T_i T_j \sum_{k \neq (i,j)} T_k J_{i,k} J_{j,k}, \quad (G5)$$

which are $O(J^2, \beta^2)$. The error on $\Phi_{i<j}$ for exact data (G3) is worsened after considering this additional error source to $O(J^4, \beta^4)$; thus, in the high-temperature limit we gain one order of magnitude in the inference of J_{ij}. In the weak-coupling

limit fewer diagrams contribute to the error but the order remains the same (unless $H = 0$). The diagrams contributing to the leading-order errors in Φ_{ii} (G4) and L_i (G2) are the same after considering this error source but contribute with opposite sign. Overestimations become underestimations and vice versa, a pattern we see realized in both direct and inverse problem applications, even for models with stronger coupling.

2. Plaquette and P_3 errors on the fully connected model

Defining $J_{ijkl} = J_{ij} J_{jk} J_{kl} J_{il} T_j T_k T_l$ we can write

$$L_i^{P_3} - L_i^{\text{exact}} \doteq -2\beta^4 t_i \sum_{j<k<l} [J_{ijkl} + J_{ikjl} + J_{ijlk}], \quad (G6)$$

the error is $O(J^4, \beta^5)$. For the off-diagonal component,

$$\Delta\Phi_{ij}^{P_3} \doteq -2\beta^4 J_{ij} t_i t_j \sum_{k<l(\neq i,j)} J_{kl} [J_{ik} J_{jl} + J_{il} J_{jk}] T_k T_l, \quad (G7)$$

the errors are $O(J^4, \beta^6)$. For the special case of a strictly pairwise Hamiltonian $H = 0$ we recover (25) for our method, and the leading-order diagram is $O(\beta^7)$. For the on-diagonal component,

$$\Delta\Phi_{ii}^{P_3} \doteq -2\beta^4 \sum_{j<k<l(\neq i)} [J_{ijkl} + J_{ikjl} + J_{ijlk}], \quad (G8)$$

the errors are $O(J^4, \beta^4)$.

If we fix the correlation parameters according to maximum entropy, the standard method, then new errors are introduced,

$$C_{ij}^* - C_{ij} \doteq \beta^3 T_i T_j \sum_{k<l(\neq i,j)} T_k T_l J_{kl} (J_{il} J_{jk} + J_{ik} J_{jl}), \quad (G9)$$

$$C_{ijk}^* - C_{ijk} \doteq 2\beta^3 T_i T_j T_k \sum_{l(\neq i,j,k)} t_l T_l J_{il} J_{jl} J_{kl}. \quad (G10)$$

As for the Bethe approximation, the forms (G6) and (G8) undergo a sign change with this modification. In the case of high temperature, the off-diagonal component error in Φ is worsened relative to the case of exact correlations C (G7) to $O(J^4, \beta^5)$. In the high-temperature limit we thus improve the estimation of J_{ij} by an order of magnitude, and in the special case of zero external field the improvement is two orders of magnitude.

APPENDIX H: DIAGONAL AND ON-DIAGONAL CONSTRAINTS

In this paper we have argued that since the CVM approximation parameters should be properly interpreted as marginal probabilities it is sensible to consider the most accurate beliefs possible, which may be linear response estimates rather than maximum entropy estimates. We then have shown how the the free energy should be modified to make the linear response self-consistent.

An alternative argument could be that for any variational framework two different derivatives (responses), or functions of derivatives, that determine the same quantity ought to be in agreement for a good approximation, those at lowest order being most important. Consider, for example, the two possible estimates of the pair correlation $E[\sigma_i \sigma_j]$, which can be estimated either by a first derivative with respect to J_{ij} or

by a nonlinear function of the responses

$$\frac{\partial F_{\text{CVM}}}{\partial J_{ij}} = \frac{\partial^2 F_{\text{CVM}}}{\partial H_i \partial H_j} + \left[\frac{\partial F_{\text{CVM}}}{\partial H_i}\right]\left[\frac{\partial F_{\text{CVM}}}{\partial H_j}\right]. \quad \text{(H1)}$$

This constraint is equivalent to (12) restricted to the case of pairs. Another important relation at second order in the derivatives would be the self-response

$$1 - \left[\frac{\partial F_{\text{CVM}}}{\partial H_i}\right]^2 = \frac{\partial^2 F_{\text{CVM}}}{\partial H_i^2}, \quad \text{(H2)}$$

which is an identity not considered in this paper but has been proposed as a constraint simultaneously in two recent papers [19,26]. The power of the on-diagonal constraint is readily apparent, and at the NMF level approximation one already recovers the adaptive TAP equations which are well tested [11,12,27].

It is natural to call the first class of constraints [(H1) and (12)] off-diagonal, and the second class (H2) on-diagonal, since they relate to the simplest possible off- and on-diagonal identities for the response matrix χ at second order. We have shown in this paper that the off-diagonal constraints can yield performance gains in isolation, and combining both constraints also can be effective [19].

APPENDIX I: CVM FREE ENERGY WITH EFFECTIVE FIELDS AND COUPLINGS

The constraint whereby pair connected correlation parameters are fixed leads to a simple modification of the entropy in (13). For the standard parameterization using constrained beliefs ($\{b_R\}$) in place of connected correlations (C), the additional term in (13) becomes a nonlinear function of the parameters. If Ω contains only pairs of variables (pair constraints) the same entropy can be written concisely in terms of the beliefs

$$S_\lambda(b) = S(b) - \sum_{(i,j)\in\Omega}\sum_{R:(i,j)\in R} c_R(\text{Tr}\,[b_R\sigma_i\sigma_j]$$
$$- \text{Tr}\,[b_R\sigma_i]\text{Tr}\,[b_R\sigma_j]). \quad \text{(I1)}$$

It is clear that this new term is not linear, convex, or concave in the beliefs, which makes certain standard and robust methods of minimization for the CVM method defunct [4,7]. Recall that c_R are the counting numbers for region R in the approximation, and the sum restricted to inclusion of any subset (e.g., $\{R : i \in R\}$ or $\{R : i, j \in R\}$) sums to 1.

If we define a new set of variational parameters, with corresponding constraints

$$M_i = \sum_{R:i\in R} c_R\text{Tr}\,[b_R\sigma_i], \quad \text{(I2)}$$

we can redefine the entropy as

$$S_\lambda(b,M) = S(b) - \sum_{(i,j)\in\Omega}\sum_{R:(i,j)\in R} c_R(\text{Tr}\,[b_R\sigma_i\sigma_j]$$
$$- \text{Tr}\,[b_R\sigma_i]M_j - \text{Tr}\,[b_R\sigma_j]M_i + M_iM_j). \quad \text{(I3)}$$

At which point we notice a convenient factorization for the full free energy

$$F_{\text{CVM}}(J,H,b,M) = \sum_i \tilde{H}_i\text{Tr}\,[b_i\sigma_i]$$
$$+ \sum_{ij} \tilde{J}_{ij}\text{Tr}\,[b_{ij}\sigma_i\sigma_j] - \frac{1}{\beta}S(b), \quad \text{(I4)}$$

which is the same as a standard CVM free energy but with effective fields and couplings

$$\tilde{H}_i = H_i + \frac{1}{\beta}\sum_R c_R\sum_{j\in R} M_j\lambda_{ij}, \quad \text{(I5)}$$

$$\tilde{J}_{ij} = J_{ij} - \frac{1}{\beta}\lambda_{ij}, \quad \text{(I6)}$$

and additional variational parameters $\{M_i\}$ in one-to-one correspondence with the additional constraint (I2). In the Bethe case, the effective field simplifies to

$$\tilde{H}_i^B = H_i + \sum_{j\in\partial_i} \lambda_{ij}M_j, \quad \text{(I7)}$$

where ∂_i are the neighbors of i.

If M and λ were fixed external parameters, our method would be equivalent to a reassignment of J and H. However, since M is variational there is a new reaction term in the saddle-point equation for M, and a new linear response term (even for paramagnetic solutions, where $M = 0$), each linear in λ. Parameters λ are fixed according to this modified linear response criteria. This can be counterintuitive. For example, by decreasing λ we increase the effective coupling, but the intuition that the susceptibility towards a ferromagnetic solution would increase is incorrect—in fact, the susceptibility of the solution decreases.

By converting the free energy to the form (I4) it becomes clearer how to apply standard methods to construct message passing framework alternatives to the iterative approach we outline and what may cause instabilities of these frameworks. In particular, we note that for fixed M and λ we can use any standard message passing or susceptibility propagation procedure [7]; the subtlety is then in the selection of update rules for λ and M (which can be chosen to require only local information).

[1] G. An, J. Stat. Phys. **52**, 727 (1988).

[2] A. Pelizzola, J. Phys. A **38**, R309 (2005).

[3] M. Wainwright and M. Jordan, Found. Trends Mach. Learn. **1**, 1 (2007).

[4] A. L. Yuille, Neural Comput. **14**, 2002 (1691).

[5] H.-J. Zhou and C. Wang, J. Stat. Phys. **148**, 513 (2012).

[6] M. Chertkov and V. Y. Chernyak, Phys. Rev. E **73**, 065102 (2006).

[7] J. Yedidia, W. Freeman, and Y. Weiss, Inf. Theory: IEEE Trans. **51**, 2282 (2005).

[8] Y. Weiss, C. Yanover, and T. Meltzer, Technical Report, Hebrew University (2006).

[9] M. Welling and Y. Teh, Neural Comput. **16**, 197 (2004).

[10] H. Kappen and F. Rodriguez, Neural Comput. **20**, 1137 (1998).

[11] M. Opper and O. Winther, Phys. Rev. Lett. **86**, 3695 (2001).

[12] M. Opper and O. Winther, Phys. Rev. E **64**, 056131 (2001).

[13] A. Montanari and T. Rizzo, J. Stat. Mech. (2005) P10011.

[14] F. Morcosa, A. Pagnani, B. Lunta, A. Bertolinoc, D. Marks, C. Sander, R. Zecchina, J. Onuchica, T. Hwaa, and M. Weigt, Proc. Natl. Acad. Sci. USA **108**, E1293 (2011).

[15] S. Cocco, S. Leibler, and R. Monasson, Proc. Natl. Acad. Sci. USA **106**, 14058 (2009).

[16] V. Sessak and R. Monasson, J. Phys. A **42**, 055001 (2009).

[17] G. Parisi, *Statistical Field Theory* (Addison-Wesley, Boston, MA, 1987).

[18] F. Ricci-Tersenghi, J. Stat. Mech. (2012) P08015.

[19] J. Raymond and F. Ricci-Tersenghi, in *IEEE ICC'13—Workshop on Networking across Disciplines: Communication Networks, Complex Systems and Statistical Physics (NETSTAT)* (IEEE, Budapest, Hungary, 2013).

[20] R. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic Press, New York, 1982).

[21] G. H. Wannier, Phys. Rev. **79**, 357 (1950).

[22] J. Stephenson, J. Math. Phys. **11**, 413 (1970).

[23] E. Aurell and M. Ekeberg, Phys. Rev. Lett. **108**, 090201 (2012).

[24] S. Edwards and P. Anderson, J. Phys. F **5**, 965 (1975).

[25] H.-J. Zhou, personal communication (2012).

[26] M. Yasuda and K. Tanaka, Phys. Rev. E **87**, 012134 (2013).

[27] H. Haiping and Y. Kabashima, arXiv:1303.2810.