POLITECNICO DI TORINO
Repository ISTITUZIONALE

Loop Corrections in Spin Models through Density Consistency

Original
Loop Corrections in Spin Models through Density Consistency / Braunstein, A., Catania, G., Dall'Asta, L.. - In: PHYSICAL REVIEW LETTERS. - ISSN 0031-9007. - 123:2(2019). [10.1103/PhysRevLett.123.020604]

Availability:
This version is available at: 11583/2741852 since: 2019-07-14T09:50:10Z

Publisher:
APS

Published
DOI:10.1103/PhysRevLett.123.020604

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
default_article_editorial [DA NON USARE]

-

(Article begins on next page)

06 November 2026

Loop Corrections in Spin Models through Density Consistency

Alfredo Braunstein,[1,2,3,*] Giovanni Catania,[1] and Luca Dall'Asta[1,3]
[1]Politecnico di Torino, Corso Duca Degli Abruzzi 24, 10129, Torino, Italy
[2]Italian Institute for Genomic Medicine, Via Nizza 52, 10126, Torino, Italy
[3]Collegio Carlo Alberto, Via Real Collegio 30, 10024, Moncalieri, Italy
and INFN Sezione di Torino, Via P. Giuria 1, 10125, Torino, Italy

(Received 5 December 2018; revised manuscript received 3 May 2019; published 12 July 2019)

Computing marginal distributions of discrete or semidiscrete Markov random fields (MRFs) is a fundamental, generally intractable problem with a vast number of applications in virtually all fields of science. We present a new family of computational schemes to approximately calculate the marginals of discrete MRFs. This method shares some desirable properties with belief propagation, in particular, providing exact marginals on acyclic graphs, but it differs with the latter in that it includes some loop corrections; i.e., it takes into account correlations coming from all cycles in the factor graph. It is also similar to the adaptive Thouless-Anderson-Palmer method, but it differs with the latter in that the consistency is not on the first two moments of the distribution but rather on the value of its density on a subset of values. The results on finite-dimensional Isinglike models show a significant improvement with respect to the Bethe-Peierls (tree) approximation in all cases and with respect to the plaquette cluster variational method approximation in many cases. In particular, for the critical inverse temperature β_c of the homogeneous hypercubic lattice, the expansion of $(d\beta_c)^{-1}$ around $d = \infty$ of the proposed scheme is exact up to d^{-4} order, whereas the latter two are exact only up to d^{-2} order.

DOI: 10.1103/PhysRevLett.123.020604

Introduction.—Markov random fields (MRFs) are undirected probabilistic graphical models in which random variables satisfy a conditional independence property so that the joint probability measure can be expressed in a factorized form, with each factor involving a possibly different subset of variables [1]. Computing marginal distributions of discrete or semidiscrete Markov random fields is a fundamental step in most approximate inference methods and high-dimensional estimation problems [2], such as the evaluation of equilibrium observables in statistical mechanics models. The exact calculation of marginal distributions is, however, intractable in general, and it is common to resort to stochastic sampling algorithms, such as the Monte Carlo Markov chain model, to obtain unbiased estimates of the relevant quantities. On the other hand, it is also useful to derive approximations of the true probability distribution, for which marginal quantities can be deterministically computed. An important family of approximations is the one of mean-field (MF) schemes. The simplest is naive mean-field (NMF) scheme, which neglects all correlations between random variables. An improved MF approximation [3] called Thouless-Anderson-Palmer (TAP) equations, works well for models with weak dependences, but it is usually unsuitable for MRFs on diluted models. Here, a considerable improvement is provided by the Bethe-Peierls approximation or belief propagation (BP), which is exact for probabilistic models defined on graphs without loops [4]. It is a fact that BP has been successfully employed even in loopy probabilistic models both in physics and in applications (see, e.g., Berrou *et al.* [5]), yet the lack of analytical control on the effect of loops calls for novel approaches that could systematically improve with respect to BP. A traditional way to account for the effect of short loops is by means of cluster variational methods (CVMs) that exactly treat correlations generated between variables within a finite region $\mathcal{R}$ [6–8]. The main limitation of the CVM resides in its algorithmic complexity that grows exponentially with the size of the region $\mathcal{R}$. A completely different path to systematically improve BP is represented by loop series expansions [9–12] in which BP is obtained as a saddle point in a corresponding effective field theory. Loop corrections to BP equations can be alternatively introduced in terms of local equations for correlation functions, as first suggested for pairwise models [13] and later extended to arbitrary factor graphs [14–16]. This method consists of considering deformed local marginal probabilities on a "cavity graph" obtained by removing a factor node (i.e., interaction) and imposing a consistency condition on single-node marginals. On trees, BP equations are recovered, whereas on loopy graphs, the obtained set of equations is strongly underdetermined and requires additional constraints. Linear-response relations were exploited for this purpose in Ref. [13], even though other moment closure methods are possible [14].

A different approach to approximate inference exploits the properties of multivariate Gaussian distributions that

 © 2019 American Physical Society

have the advantage of retaining information on correlations albeit allowing for explicit calculations. In particular, expectation propagation (EP, which can be thought as an adaptive variant of TAP) is a very successful algorithmic technique in which a tractable approximate distribution is obtained as the outcome of an iterative process in which the parameters of a multivariate Gaussian are optimized by means of local moment matching conditions [17,18]. EP has been applied to problems involving discrete random variables by employing atomic measures [19]. In the present work, we put forward a new family of computational schemes to calculate approximate marginals of discrete MRFs. We exploit the flexibility of multivariate Gaussian approximation methods, but, unlike EP and inspired by the beliefs marginalization condition in BP, we impose that marginals over factors are locally consistent on each variable, a condition that we call density consistency. When the underlying graph is a tree, the set of equations produced is equivalent to BP. As for Ref. [13], the density consistency condition leaves an underdetermined system of equations on loopy graphs that can be solved once they are supplemented with a further set of closure conditions. As a result of employing Gaussian distributions, higher-order correlation functions between neighbors of a given variable are, at least partially, taken into account.

The model.—Consider a factorized distribution of binary variables $x_1, \ldots, x_n \in X = \{-1, 1\}$ for arbitrary positive factors $\psi_a \colon X_a \to \mathbb{R}_+$, each depending on a subvector $\boldsymbol{x}_a = \{x_i\}_{i \in \partial a} \in X_a$,

$$p(\boldsymbol{x}) = \frac{1}{z} \prod_{a \in A} \psi_a(\boldsymbol{x}_a). \tag{1}$$

The bipartite graph $G = (V, E)$ with $V = I \cup A$ the disjoint union of variable indices $I = \{1, \ldots, n\}$ and factor indices and $E = \{(ia) : i \in \partial a\}$ is called the factor graph of the factorization (1), and as we will see, some of its topological features are crucial to devise good approximations. Particular important cases of Eq. (1) include, e.g., Ising spin models, many neural network models, and the uniform distribution of the solutions of k-SAT formulas. Computing marginal distributions from Eq. (1) is, in general, NP hard (i.e., computationally intractable).

Density consistency.—Following Gaussian expectation propagation (otherwise called adaptive TAP or expectation consistency) [17,18], we will approximate an intractable $p(\boldsymbol{x})$ by a normal distribution $g(\boldsymbol{x})$. To do so, we will replace each $\psi_a(\boldsymbol{x}_a)$ by an appropriately defined multivariate normal distribution $\phi_a(\boldsymbol{x}_a) = \mathcal{N}(\boldsymbol{x}_a; \boldsymbol{\mu}^a, \boldsymbol{\Sigma}^a)$. The parameters $\boldsymbol{\mu}^a$, $\boldsymbol{\Sigma}^a$ will be selected as follows. First define

$$g(\boldsymbol{x}) = \frac{1}{z_g} \prod_a \phi_a(\boldsymbol{x}_a) = \frac{1}{z'} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x}-\boldsymbol{\mu})},$$

and $g^{(a)}(\boldsymbol{x}) = \frac{1}{z_a} g(\boldsymbol{x}) \sum_{\hat{\boldsymbol{x}}_a \in X_a} \delta(\boldsymbol{x}_a - \hat{\boldsymbol{x}}_a) [\psi_a(\hat{\boldsymbol{x}}_a)/\phi_a(\hat{\boldsymbol{x}}_a)]$, $g^{(i)}(\boldsymbol{x}) = \frac{1}{z_i} g(\boldsymbol{x}) \sum_{\hat{x}_i \in X} \delta(x_i - \hat{x}_i)$ as auxiliary distributions. Matching between the marginals $g^{(i)}(x_i)$ and $g^{(a)}(x_i)$ results in

$$\frac{\mu_i}{\Sigma_{ii}} = \operatorname{atanh} \langle x_i \rangle_{g^{(a)}}, \tag{2}$$

$\forall\, i \in \partial a, a \in A$, giving $\sum_a |\partial a|$ equations. As we will see, Eq. (2) is chosen because it ensures exactness on acyclic graphs. We call density consistency (DC) any scheme that enforces Eq. (2). We propose to complement Eq. (2) with matching of the first moments and Pearson correlation coefficients $\operatorname{corr}_Q(x,y) = (\langle xy \rangle_Q - \langle x \rangle_Q \langle y \rangle_Q)(\langle x^2 \rangle_Q - \langle x \rangle_Q^2)^{-\frac{1}{2}} (\langle y^2 \rangle_Q - \langle y \rangle_Q^2)^{-\frac{1}{2}}$ (although other closures are possible; see the Supplemental Material [20])

$$\mu_i = \langle x_i \rangle_{g^{(a)}}, \qquad \frac{\Sigma_{ij}}{\sqrt{\Sigma_{ii}\Sigma_{jj}}} = \rho \operatorname{corr}_{g^{(a)}}(x_i, x_j) \tag{3}$$

for $i \neq j$ where $\rho \in [0, 1]$ is an interpolating parameter that is fixed to 1 for the time being. Relations (2) and (3) give a system of $\sum_a |\partial a|(|\partial a| + 3)/2$ equations and unknowns that can be solved iteratively to provide an approximation for the first $\langle x_i \rangle_p$ and second moments $\langle x_i x_j \rangle_p$ of the original distribution. In a parallel update scheme (in which all factor parameters are updated simultaneously), the computational cost of each iteration is $O(N^3)$ dominated by the calculation of Σ.

On acyclic factor graphs, the method converges in a finite number of iterations and is exact; i.e., on a fixed point, $\langle x_i \rangle_g$ equals the magnetization $\langle x_i \rangle_p$. Therefore, as both the DC scheme and BP are exact on acyclic graphs, their estimation of marginals must coincide. However, a deeper connection can be pointed out. If a DC scheme applies zero covariances [e.g., by setting $\rho = 0$ in Eq. (3)], on a DC fixed point on any factor graph, the quantities $m_{ai} = \tanh(\mu_i^a / \Sigma_{ii}^a)$ satisfy the BP equations. Moreover, DC dynamically follows a BP update. In particular, when the equations converge, the magnetizations $m_i = \tanh(\mu_i / \Sigma_{ii})$ are equal to the corresponding belief magnetizations (proof in the Supplemental Material [20]).

Interestingly, the DC scheme can be thought of as a Gaussian pairwise EP scheme with a modified consistency condition. The latter can be obtained by keeping Eq. (3) and replacing $\operatorname{atanh}(x)$ on the rhs of Eq. (2) with the qualitatively similar $x/(1 - x^2)$. This of course renders the method inexact on acyclic graphs and turns out to give generally a much worse approximation in many cases (see the Supplemental Material [20]). In addition, as it also happens with the EP method, the Gaussian densities in factors ψ_a can be moved freely between factors (sharing the same variables) without altering the approximation (details in the Supplemental Material [20]).

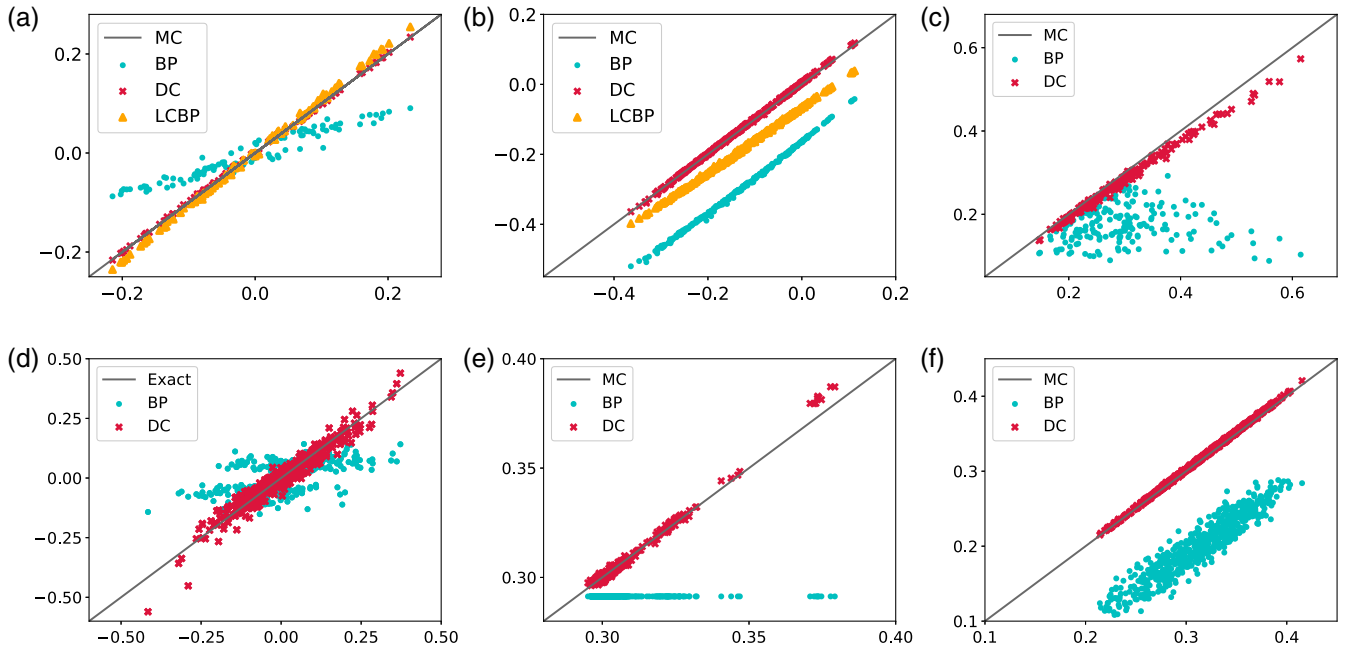

FIG. 1. Comparison of the DC, BP, and LCBP on single instances of disordered systems. (a) Magnetizations of the antiferromagnetic Ising model on a triangular lattice with $N = 100$, $|E| = 6N$, $J = -1$, $\beta = 0.52$, and random binary fields of $|h_i| = 0.2$. (b) Magnetization of the ferromagnetic Ising model on a random regular (RR) graph, $N = 300$, degree 4, $\beta = 0.35$, $J = 1$, and random binary fields of $|h_i| = 0.3$. (c) Correlations of the (heterogeneous) Ising model on a Barabasi-Albert graph, $N = 100$, $n_0 = k = 2$ without external fields (the solution is found by using $\rho^* = 0.95$, and it is divergent for $\rho > \rho^*$). (d) Magnetizations of a random 4-SAT instance at $\alpha = (M/N) = 4$ at $\beta \to \infty$. (e) Correlations of (heterogeneous) the ferromagnetic Ising model on a RR graph, $N = 300$, degree 4, and $\beta = 0.3$. (f) Correlations on a 3D hypercubic toroidal lattice ferromagnetic (heterogeneous) Ising model, $N = 6^3$ and $\beta = 0.21$, and no external fields. In heterogeneous ferromagnetic models, couplings are drawn from a uniform distribution in (0.5,1.5).

Numerical results.—We tested the method on the Ising model in many different scenarios of heterogeneous systems, with a selection of results given in Fig. 1. The true values for the magnetization and correlations were computed approximately with long Monte Carlo runs ($1 \times 10^6 N$–$2 \times 10^6 N$ Monte Carlo Gibbs-sampling steps) for Ising models and with the exact (exponential) trace for up to $N = 28$ in the case of k-SAT. All simulations have been performed with a damping parameter of around 0.95 to improve convergence. The DC method provides a substantial correction to BP magnetizations and correlations in almost all cases (details in the Supplemental Material [20]); it also improves single-node marginal estimates with respect to loop corrected belief propagation (LCBP) [14] in several cases. LCBP simulations were performed using the code provided in Ref. [21]. We underline that despite the computational cost per iteration of LCBP on bounded-degree graphs being $O(N^2)$, the prefactor strongly depends on the degree distribution (with even exponential scaling in some cases); also, the number of iterations required to converge is normally much larger than the one of DC. For instance, for antiferromagnetic models [like the one shown in Fig. 1(a)], LCBP does not seem to converge at smaller temperatures.

Homogeneous Ising model.—Consider a homogeneous ferromagnetic Ising model with coupling constant J and external field h^{ext} on a d-dimensional lattice with periodic (toroidal) boundary condition: because of the translational invariance, all Gaussian factors ϕ_a are identical, and the covariance matrix admits an analytic diagonalization. Therefore, it is possible to estimate equilibrium observables through an analytical DC scheme also in the thermodynamic limit. After some calculations (see the Supplemental Material [20]), at a given temperature the DC solution is found by solving the following system of three fixed point equations $\sigma_0 = (m/\text{atanh}m)$, $\sigma_1 = \rho[(c-m^2)/(1-m^2)]\sigma_0$, and $y = m(\gamma_0 + \gamma_1)$ in the Gaussian parameters y, γ_0, γ_1, where $m = \langle x_i \rangle_{g^{(a)}}$, $c = \langle x_i x_j \rangle_{g^{(a)}}$ are the moments computed under the "tilted" distribution $g^{(a)}$, and σ_0, σ_1, γ_0, γ_1 equal, respectively, Σ_{ii}, Σ_{ij}, $(2d)^{-1}(\Sigma^{-1})_{ii}$, $(\Sigma^{-1})_{ij}$ for i, j two first lattice neighbors. Defining $R_d(r) = \frac{1}{2}\int_0^\infty dt\, e^{-dt}\mathcal{I}_0^d(rt)$, where $\mathcal{I}_0$ is the modified Bessel function of the first kind of order 0, and after some straightforward algebraic manipulations, we finally obtain the following equations (here, $h^{\text{ext}} = 0$ for simplicity) for variables β, m, $r = \gamma_1\gamma_0^{-1}$,

$$\beta = \text{ath}\left[\frac{k_r}{\rho} + m^2\left(1 - \frac{k_r}{\rho}\right)\right] - g_r\frac{h}{m} - \text{ath}[\text{th}^2(f_r h)], \quad (4)$$

$$m = \text{th}\left\{f_r h + \text{ath}\left[\text{th}\left(\beta + \frac{g_r h}{m}\right)\text{th}(f_r h)\right]\right\}, \quad (5)$$

TABLE I. Critical values obtained with different approximation schemes of the inverse temperature β marking the onset of spontaneous magnetization in the homogeneous Ising model on infinite d-dimensional hypercubic lattices. The values of β_{BP}, β_{PCVM}, and β_{LCB}, respectively, refer to the Bethe-Peierls approximation, plaquette cluster variational method [23], and loop corrected Bethe [13] approximation, while β_p and β_m, respectively, correspond to the maximum β of the paramagnetic DC solution and the minimum β of the magnetized DC solution, and β_c indicates the currently best-known approximation up to numerical accuracy (Ref. [23] for $d \leq 5$, Ref. [24] for $d = 6$). Results in bold indicate the closest value to the last column.

d	β_{BP}	β_{PCVM}	β_{LCB}	β_m	β_p	β_c
2	0.346 57	**0.412 258**	$(\cdots)$	0.388 448	0.376 93	0.440 687
3	0.202 73	0.216 932	0.238 520	0.218 908	**0.222 223**	0.221 654(6)
4	0.143 84	0.148 033	0.151 650	0.149 835	**0.149 862**	0.149 66(3)
5	0.111 57	0.113 362	0.114 356	**0.113 946**	**0.113 946**	0.113 88(3)
6	0.091 16	0.092 088	0.092 446	**0.092 304**	**0.092 304**	(0.092 253 0)

where $k_r = [(1-2dR_d(r))/r2dR_d(r)]$, $g_r = [k_r/(1-k_r^2)+rR_d(r)]$, $f_r = [1/(1+k_r)-(r+1)R_d(r)]$, $h = \mathrm{ath}m$, while the functions th and ath denote the hyperbolic tangent and its inverse, respectively. Substituting Eq. (4) into Eq. (5), we get a single equation for m, r, allowing for a parametric solution $m(r)$, $\beta(r)$.

The maximum value of β for which a paramagnetic solution exists can be analytically derived by substituting $m = 0$ and taking $\sup_{-1<r\leq 0}\beta(r)$ from Eq. (4). For $d \geq 3$ [22], the maximum is realized at $r = -1$, obtaining

$$\beta_p = \mathrm{ath}\left(1 - \frac{1}{x_d}\right) - \frac{x_d(x_d-1)}{2x_d-1} + \frac{x_d}{2d}, \tag{6}$$

where $x_d = 2dR_d(-1)$. Values of β_p for various dimensions d are reported in Table I. The paramagnetic solution is stable in the full range $0 \leq \beta < \beta_p$ for $d \geq 3$.

Expanding Eq. (6) in powers of d^{-1}, we get $1/(2d\beta_p) = 1 - \frac{1}{2}d^{-1} - \frac{1}{3}d^{-2} - (13/24)d^{-3} - (979/720)d^{-4} - (2039/480)d^{-5} + O(d^{-6})$, which is exact up to the d^{-4} order (the correct coefficient of d^{-5} is $-(2009/480)$) [24]. For comparison, NMF is exact up to the d^0 order, BP is exact up to the d^{-1} order, while the loop corrected Bethe (LCB) [13] approximation and plaquette CVM (PCVM) [23] are exact up to the d^{-2} order.

The minimum value of β for which a magnetized solution exists can also be computed by seeking a point with $(d\beta/dr) = 0$ with the complication that m is defined implicitly by Eqs. (4) and (5) (details in the Supplemental Material [20]). The resulting equation has a single solution that has been numerically computed and shown in Table I as β_m. It turns out to be smaller but always very close to β_p and coincident up to numerical precision for $d \geq 5$. Note that for inverse temperatures in the (albeit small) range $\beta_m < \beta < \beta_p$, the DC approximation has both magnetized and paramagnetic stable solutions, suggesting a phase coexistence that should be absent in the real system [25].

Discussion.—We proposed a general approximation scheme for distributions of discrete variables that show interesting properties, including being exact on acyclic factor graphs and providing a form of loop corrections on graphs with cycles.

In the same spirit as PCVM and LCBP, the DC approximation can be thought of as a method to correct the cavity independence (or absence of cycles) assumption in the Bethe-Peierls approximation. Whereas PCVM deals only with local (short) cycles, it is true that LCBP and DC both attempt to correct for arbitrarily long cycles in the interaction graph. However, they do so through crucially different approaches. LCBP works by computing several BP fixed points (one for each cavity distribution in which one node and all the factors connected to it are removed) and then imposing consistency over single-node beliefs among them. Therefore, for each cavity distribution, it computes fixed points by still assuming a tree factorization, i.e., by neglecting the correlations coming from other cycles in the graph. Therefore it computes a higher-order approximation by relying on a lower order one, computed on a modified or simplified interaction graph. In this sense, it can be considered as a first-order correction to BP, and indeed it improves BP estimates of single-node marginals, as shown in Fig. 1. With this perspective, DC can be considered as a new approximation in which all two-point cavity correlations are taken into account (of course, in an approximate way, through a Gaussian distribution) in a single self-consistent set of equations in which correlations arise simultaneously from all cycles in the graph.

The method can be solved analytically for homogeneous systems such as finite-dimensional hypercubic lattices with periodic conditions. Analytical predictions from the model show a number of interesting features that are not shared by other mean-field approaches: The method provides finite-size corrections which are in close agreement with numerical simulations; the paramagnetic solution exists only for $\beta < \beta_p$ (in PCVM and BP, the paramagnetic solution exists for all $\beta \geq 0$, although it stops being stable at a finite value of β); it can capture some types of heterogeneity where the Bethe-Peierls approximation cannot (such as in RR graphs). Numerical simulations are in good agreement with

Monte Carlo estimates on different models, including the random-field Ising model with various topologies and random k-SAT. On lattices, the method could in principle be rendered more accurate by taking into account small loops explicitly. The DC scheme can be extended for models with q-state variables by replacing each of them with q binary variables. Again, in this setup it is possible to get a similar set of closure equations that are exact on acyclic graphs and recover BP fixed points on any graph when neglecting cavity correlations. This will be the subject of future research.

We warmly thank F. Ricci-Tersenghi, A. Lage-Castellanos, A. P. Muntoni, I. Biazzo, A. Fendrik, and L. Romanelli for interesting discussions. A. B. and G. C. thank Universidad Nacional de General Sarmiento for hospitality. The authors acknowledge funding from EU Horizon 2020 Marie Sklodowska-Curie Grant Agreement No. 734439 (INFERNET) and PRIN (Progetti di Rilevante Interesse Nazionale) Project No. 2015592CTH.

*Corresponding author.
alfredo.braunstein@polito.it

[1] S. L. Lauritzen, *Graphical Models* (Clarendon Press, Oxford, 1996), Vol. 17.
[2] D. J. MacKay, *Advanced Mean Field Methods: Theory and Practice* (MIT Press, Cambridge, 2001).
[3] M. Opper, D. Saad *et al.*, *Advanced Mean Field Methods: Theory and Practice* (2001), Vol. 7.
[4] M. Mezard and A. Montanari, *Information, Physics, and Computation* (Oxford University Press, New York, 2009).
[5] C. Berrou, A. Glavieux, and P. Thitimajshima, in *Proceedings of ICC '93—IEEE International Conference on Communications* (1993), Vol. 2, pp. 1064–1070.
[6] R. Kikuchi, Phys. Rev. **81**, 988 (1951).
[7] J. S. Yedidia, W. T. Freeman, and Y. Weiss, Exploring artificial intelligence in the new millennium **8**, 236 (2003).
[8] A. Pelizzola, J. Phys. A **38**, R309 (2005).
[9] M. Chertkov and V. Y. Chernyak, J. Stat. Mech. (2006) P06009.
[10] V. Gómez, J. M. Mooij, and H. J. Kappen, J. Mach. Learn. Res. **8**, 1987 (2007).
[11] G. Parisi and F. Slanina, J. Stat. Mech. (2006) L02003.
[12] A. Altieri, M. C. Angelini, C. Lucibello, G. Parisi, F. Ricci-Tersenghi, and T. Rizzo, J. Stat. Mech. (2017) 113303.
[13] A. Montanari and T. Rizzo, J. Stat. Mech. (2005) P10011.
[14] J. Mooij, B. Wemmenhove, B. Kappen, and T. Rizzo, in *Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics*, edited by M. Meila and X. Shen Proceedings of Machine Learning Research (2007), pp. 331–338.
[15] T. Rizzo, B. Wemmenhove, and H. J. Kappen, Phys. Rev. E **76**, 011102 (2007).
[16] M. Ohzeki, J. Phys. Conf. Ser. **473**, 012005 (2013).
[17] T. P. Minka, in *Proceedings of the Seventeenth Conference on Uncertainty in Artificial Intelligence, UAI'01* (Morgan Kaufmann Publishers Inc., San Francisco, 2001), pp. 362–369.
[18] M. Opper and O. Winther, Phys. Rev. E **64**, 056131 (2001).
[19] M. Opper and O. Winther, J. Mach. Learn. Res. **6**, 2177 (2005).
[20] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.123.020604 for more details.
[21] J. M. Mooij, J. Mach. Learn. Res. **11**, 2169 (2010).
[22] $R_2(-1)$ is divergent, and the maximum for $d = 2$ is located in $-1 < r^* < 0$. The solution for $d = 2$ is thus qualitatively different, see Ref. [20].
[23] E. Domínguez, A. Lage-Castellanos, R. Mulet, and F. Ricci-Tersenghi, Phys. Rev. E **95**, 043308 (2017).
[24] M. E. Fisher and D. S. Gaunt, Phys. Rev. **133**, A224 (1964).
[25] J. L. Lebowitz, J. Stat. Phys. **16**, 463 (1977).