Radboud Repository

Radboud University Nijmegen

PDF hosted at the Radboud Repository of the Radboud University Nijmegen

The following full text is an author's version which may differ from the publisher's version.

For additional information about this publication click this link.
http://hdl.handle.net/2066/32808

Please be advised that this information was generated on 2022-08-25 and may be subject to change.

On the properties of the Bethe approximation and Loopy Belief Propagation on binary networks

JM Mooij and HJ Kappen

Dept. of Biophysics, Inst. for Neuroscience, Radboud Univ. Nijmegen, 6525 EZ Nijmegen, the Netherlands

E-mail: j.mooij@science.ru.nl, b.kappen@science.ru.nl

Abstract. We analyze the local stability of the high-temperature fixed point of the Loopy Belief Propagation (LBP) algorithm and how this relates to the properties of the Bethe free energy which LBP tries to minimize. We focus on the case of binary networks with pairwise interactions. In particular, we state sufficient conditions for convergence of LBP to a unique fixed point and show that these are sharp for purely ferromagnetic interactions. In contrast, in the purely anti-ferromagnetic case, the undamped parallel LBP algorithm is suboptimal in the sense that the stability of the fixed point breaks down much earlier than for damped or sequential LBP; we observe that the onset of instability for the latter algorithms is related to the properties of the Bethe free energy. For spin-glass interactions, damping LBP only helps slightly. We estimate analytically the temperature at which the high-temperature LBP fixed point becomes unstable for random graphs with arbitrary degree distributions and random interactions.

PACS numbers: 75.10.Nr, 02.50.Tt, 89.75.-k

Submitted to: *Journal of Statistical Mechanics: theory and experiment (JSTAT)*

1. Introduction

Techniques that were originally developed in the statistical physics of lattice models are nowadays increasingly often and successfully applied in diverse application areas such as information theory, coding theory, combinatorial optimization and machine learning. A prominent example is the Bethe-Peierls approximation [1, 2], an extension of the ordinary Mean Field method that takes into account correlations between nearest-neighbour sites. A more general and powerful approximation scheme, which is also currently being used as a general inference tool in applications in the aforementioned areas, is the Cluster Variation Method (CVM) [3, 4], also called Kikuchi approximation. The CVM treats arbitrarily large clusters of sites exactly; the Bethe approximation can be seen as the simplest nontrivial case (the pair approximation) of the Cluster Variation Method.

The problems arising in the aforementioned application domains can often be reformulated as inference problems on graphical models, i.e. as the calculation of marginal probabilities of some probability distribution. Typically, this probability distribution is proportional to a product of many factors, each factor depending on only a few variables; this structure can be expressed in terms of a graph, hence the name *graphical model*. An illustrative example can be found in image restoration [5], where the 2D classical Ising model can be used to model features of monochromatic images. The pixels in the image correspond to the Ising spins, the local external fields correspond to observed, noisy pixels and the probability distribution over different images corresponds to the equilibrium Boltzmann distribution of the Ising model. The underlying graph is in this example the 2D rectangular lattice, and the interactions between the nearest-neighbours correspond to factors in the probability distribution. By taking the interactions to be of the ferromagnetic type, one can obtain a smoothing filter.

In statistical physics, one is predominantly interested in the thermodynamic limit of infinitely large systems, and furthermore, in the case of disordered systems, one usually averages over a whole ensemble of such systems. In contrast, in the applications in computer science the primary interest lies in the properties of individual, finite systems—in the example above, one would be interested in individual images. Given the probability distribution, the task is then to calculate marginal probabilities, which in principle amounts to performing a summation or integral. Unfortunately, the required computational time is generally exponential in the number of variables, and the calculation quickly becomes infeasible for real-world applications.

Therefore, one is often forced to use approximative methods, such as Monte Carlo methods or "deterministic approximations". A prominent example of the latter category is the successful Belief Propagation algorithm [6], which was originally developed as a fast algorithm to calculate probabilities on graphical models without loops (i.e. on trees), for which the results are exact. The same algorithm can also be applied on graphs containing loops, in which case the results are approximative, and it is then often called

Loopy Belief Propagation (LBP) to emphasize the fact that the graph may contain loops. The results can be surprisingly good, even for small graphs with many short loops, e.g. in the case of decoding error-correcting codes [7, 8]. An important discovery was that the LBP algorithm in fact tries to minimize the Bethe free energy (more precisely, fixed points of the LBP algorithm correspond to stationary points of the Bethe free energy) [9]. This discovery has lead to renewed interest in the Bethe approximation and related methods and to cross-fertilization between disciplines, a rather spectacular example of which is the Survey Propagation (SP) algorithm, which is now the state of the art solution method for some difficult combinatorial optimization problems [10]. Other examples are the generalizations of LBP obtained by replacing the Bethe free energy by the more complicated Kikuchi free energy, which has resulted in algorithms that are much faster than the NIM algorithm developed originally by Kikuchi [4].

This article is organised as follows. We start in section 2 with a brief review of the Bethe approximation and the Loopy Belief Propagation algorithm, trying to combine the two different points of view, namely the statistical physicist's perspective and the one found in machine learning and computer science. A notorious problem plaguing applications of LBP is the fact that it does not always converge to a fixed point. With the aim of better understanding these convergence issues, in section 3 we discuss the local stability of LBP fixed points, state "global" conditions for convergence towards a unique fixed point, and discuss the stability of the high-temperature Bethe free energy minimum. In section 4, we qualitatively discuss how these properties are related and connect them with phase transitions in the thermodynamic limit. In section 5, we quantify the results of the previous section by estimating the phase transition temperatures for random graphs with random interactions.

This article is written primarily for statistical physicists, but we tried to make it also understandable for readers with a background in computer science, which may explain some seemingly redundant remarks.

2. The Bethe approximation and the LBP algorithm

2.1. The graphical model

Let $G = (V, B)$ be an undirected labelled graph without self-connections, defined by a set of vertices $V = \{1, \ldots, N\}$ and a set of edges $B \subseteq \{(i, j) \,|\, 1 \leq i < j \leq N\}$. The *adjacency matrix* M corresponding to G is defined as follows: $M_{ij} = 1$ if $(ij) \in B$ or $(ji) \in B$ and 0 otherwise. Denote by N_i the set of neighbours of vertex i, and the degree (connectivity) of vertex i by $d_i := |N_i| = \sum_{j \in V} M_{ij}$.

To each vertex $i \in V$ we associate a random variable s_i (called a "spin"), taking values in $\{-1, +1\}$. We put weights J_{ij} on the edges (ij): let J be a symmetric $N \times N$ matrix that is compatible with the adjacency matrix M, i.e. $J_{ij} = 0$ if $M_{ij} = 0$. Let $\theta \in \mathbf{R}^N$ be local "fields" (local "evidence") acting on the vertices. We will study the

Boltzmann distribution corresponding to the Hamiltonian

$$H = -\sum_{(i,j)\in B} J_{ij} s_i s_j - \sum_i \theta_i s_i = -\frac{1}{2}\sum_{i,j} J_{ij} M_{ij} s_i s_j - \sum_i \theta_i s_i, \tag{1}$$

i.e. the probability of the configuration $s = (s_1, \ldots, s_N) \in \{-1, +1\}^N$ is given by:

$$P(s) = \frac{1}{Z} \exp\left(\beta \sum_{(i,j)\in B} J_{ij} s_i s_j + \beta \sum_i \theta_i s_i\right) \tag{2}$$

with $\beta > 0$ the inverse temperature and Z a normalization constant. The problem that we would like to solve is calculating the first and second moments $\langle s_i \rangle$ and $\langle s_i s_j \rangle$ under this distribution. In general, this is an NP-complete problem, so in practice we often have to settle for approximations of these quantities.

The general model class that we have described above has been the subject of numerous investigations in statistical physics. There one often takes a lattice as the underlying graph G, or studies an ensemble of random graphs (including the fully-connected SK model as a limiting case). The weights J_{ij} and the local fields θ_i are often taken to be i.i.d. according to some probability distribution (a special case is where this probability distribution is a delta function—this corresponds to uniform, deterministic interactions). In these cases one can take the thermodynamic limit $N \to \infty$, which is the subject of investigation of the major part of statistical physics studies (except for the studies of "finite size effects"). Depending on these weight distributions and on the graph structure, macroscopic order parameters can be identified that distinguish between different phases, e.g. the ferromagnetic phase for large positive weights or a spin-glass phase for weights that are distributed around zero.

The probability distribution (2) is a special case of the class of probability distributions over N discrete random variables $\{X_i\}_{i=1}^N$, with X_i taking values in some finite set $\mathcal{X}_i$, that factorize as a product of factors ψ (often called "potentials" in computer science literature—not to be confused with the potentials in statistical physics, which are the logarithms of the factors) in the following way:

$$P(X = x) = \frac{1}{Z} \prod_{(ij)\in B} \psi_{ij}(x_i, x_j) \prod_{i\in V} \psi_i(x_i) \tag{3}$$

with Z the normalization constant. These probability distributions are known in machine learning as *undirected graphical models* (in this case consisting of N nodes with pairwise potentials) or as *Markov Random Fields*. In fact, it is easy to see that (2) is equivalent to (3) when all variables are binary (and the factors are positive); in this case, (2) can obviously be written in the form of (3), but the converse also holds. Applications include decoding of error-correcting codes [7], artificial vision [11] and medical diagnosis [12]. In contrast with statistical physics studies, the number of variables is usually finite and one is interested in a single instance instead of the properties of an ensemble of instances.

In the following three subsections, we describe the LBP algorithm and the Bethe approximation for the graphical model (3), and what is known about the relation between the two.

2.2. Bethe approximation

The calculation of properties such as marginals $P(s_i)$ of the probability distribution (2) is an NP-complete problem. Only in cases with much symmetry (e.g. when all weights J_{ij} are equal and the field is uniform, i.e. $\theta_i = \theta$, and the graph has a high permutation symmetry, such as e.g. translation symmetry in case of a 2D rectangular lattice), or if N is small, or if the graph contains no cycles, it is possible to calculate marginals exactly. In other cases, one has to use approximate methods, such as Monte Carlo methods or "deterministic" approximation methods, the simplest of which is the well-known Mean Field method. An extension of the Mean Field method that treats pairs of neighbouring spins exactly is the Bethe approximation, also known as the Bethe-Peierls approximation [1, 2].

The Bethe approximation consists of minimizing the *Bethe free energy*, which for the factorizing probability distribution (3) is defined as the following functional [9]:

$$F_{Bethe}(\{b_i, b_{ij}\}) = \sum_{(ij)\in B} \sum_{x_i,x_j} b_{ij}(x_i, x_j) \log \frac{b_{ij}(x_i, x_j)}{\psi_{ij}(x_i, x_j)\psi_i(x_i)\psi_j(x_j)} \tag{4}$$

$$- \sum_i (d_i - 1) \sum_{x_i} b_i(x_i) \log \frac{b_i(x_i)}{\psi_i(x_i)}. \tag{5}$$

Its arguments, called *beliefs*, are single-node marginals $b_i(x_i)$ and pairwise marginals $b_{ij}(x_i, x_j)$. The *Bethe approximation* is obtained by minimizing the Bethe free energy with respect to the beliefs under the following normalization and consistency constraints

$$\sum_{x_i} b_i(x_i) = 1 \qquad \text{for all } i \in V, \tag{6}$$

$$\sum_{x_i} b_{ij}(x_i, x_j) = b_j(x_j) \qquad \text{for all } (ij) \in B. \tag{7}$$

The values of these variables at the minimum of F_{Bethe} are then taken as approximations for the marginal distributions $P(x_i)$ and $P(x_i, x_j)$. The beliefs are the exact marginals when the underlying graph G contains no cycles [13]. The rationale for *minimizing* the Bethe free energy is that the Bethe free energy is an approximate Gibbs free energy with an exact energy term, but in which the entropy term is approximated by only the single-node and pairwise entropies. Minimizing the exact Gibbs free energy would recover the exact marginal distributions $P(x_i)$ and $P(x_i, x_j)$, but is infeasible; minimizing its approximation, the Bethe free energy, gives approximations b_i and b_{ij} to the exact marginal distributions [14].

2.3. LBP algorithm

A popular and efficient algorithm for obtaining the Bethe approximation is Loopy Belief Propagation (LBP), also known under the names Sum-Product Algorithm [15] or simply Belief Propagation [6]. The adjective "Loopy" is used to emphasize the fact that the graph may contain cycles, i.e. that the beliefs are only approximations of the exact marginals.

The LBP algorithm consists of the iterative updating of a set of *messages* $\{\mu_{ij} : (ij) \in B \vee (ji) \in B\}$. The new message μ_{ij}^{new} that vertex i sends to its neighbour j is given in terms of all incoming messages by the following update rule [9]:‡

$$\mu_{ij}^{\text{new}}(x_j) \propto \sum_{x_i} \psi_{ij}(x_i, x_j)\psi_i(x_i) \prod_{k \in N_i \setminus j} \mu_{ki}(x_i), \tag{8}$$

where one usually normalizes messages such that $\sum_{x_j} \mu_{ij}^{\text{new}}(x_j) = 1$. The update schedule can be chosen to be parallel ("flooding schedule"), sequential ("serial schedule") or random; the update schedule influences convergence properties.

When the messages μ_{ij} have converged to some fixed point μ_{ij}^{∞}, the approximate marginal distributions (beliefs) $\{b_i\}_{i \in V}$ and $\{b_{ij}\}_{(ij) \in B}$ are calculated by

$$b_i(x_i) \propto \psi_i(x_i) \prod_{k \in N_i} \mu_{ki}^{\infty}(x_i), \tag{9}$$

$$b_{ij}(x_i, x_j) \propto \psi_{ij}(x_i, x_j)\psi_i(x_i)\psi_j(x_j) \left(\prod_{k \in N_i \setminus j} \mu_{ki}^{\infty}(x_i) \right) \left(\prod_{k \in N_j \setminus i} \mu_{kj}^{\infty}(x_j) \right). \tag{10}$$

Note that these beliefs satisfy the normalization and consistency constraints (6) and (7).

Unfortunately, LBP does not always converge. It can get trapped in limit cycles, or it can wander around chaotically, depending on the problem instance. This non-robust behaviour hampers application of LBP as a "black box" inference algorithm. Furthermore, there is some empirical evidence that if LBP does not converge, the quality of the Bethe approximation (which can also be obtained by using double-loop algorithms [16] that are guaranteed to converge, but are slower than LBP) is low. The analysis that we will perform in subsequent sections should be seen as first steps in obtaining a better understanding of these issues.

2.4. The connection between LBP and the Bethe approximation

Using Lagrange multipliers, one can prove [9] that the beliefs $b(\mu^{\infty})$ corresponding to a LBP fixed point μ^{∞} are a stationary point of the Bethe free energy under the constraints (6) and (7). Conversely, a set of messages μ for which the corresponding beliefs $b(\mu)$ are a stationary point of the constrained Bethe free energy, are a fixed point of LBP. In other words: stationary points of the Bethe free energy correspond one-to-one to fixed points of LBP.

‡ Here and in the following, if X is a set, we write $X \setminus i$ as a shorthand notation for $X \setminus \{i\}$.

It takes considerably more effort to prove that *(locally) stable* LBP fixed points are *(local) minima* of the constrained Bethe free energy [17]. The converse does not necessarily hold (as was already observed by Heskes [17]), i.e. a minimum of the Bethe free energy need not be a stable fixed point of LBP. In that case, LBP cannot be used to obtain the Bethe approximation. We will see examples of this in section 4.

3. Stability analysis for binary variables

From now on, we consider the special case (2) for which all variables are binary. In this section, we derive conditions for the local stability of fixed points of parallel LBP, in the undamped and damped cases. We state sufficient conditions for the uniqueness of the fixed point and "global" convergence properties of parallel, undamped LBP. Finally, we discuss the properties of Bethe energy minima for binary variables. In section 4 we will study the relations between those properties. We will start with reformulating LBP for the case of binary variables.

3.1. LBP for binary variables

In the case of binary variables, we can parameterize each message μ_{ij} by a single real number. A canonical choice is to transform to the variables ν_{ij} defined by

$$\nu_{ij} := \tanh^{-1}\Big(\mu_{ij}(s_j = 1) - \mu_{ij}(s_j = -1)\Big). \tag{11}$$

The LBP update equations (8) can be written in terms of these new messages as:

$$\tanh(\nu_{ij}^{\text{new}}) = \tanh(\beta J_{ij})\tanh(\beta h_{i\backslash j}), \tag{12}$$

where we defined the "cavity field" $h_{i\backslash j}$ by

$$\beta h_{i\backslash j} := \beta\theta_i + \sum_{k\in N_i\backslash j} \nu_{ki}. \tag{13}$$

Our usage of the term "cavity field" corresponds to that in [18] and is motivated by the fact that $h_{i\backslash j}$ is the effective field that acts on spin i in the absence of spin j (under the assumption that the spins $k \in N_i$ are independent in the absence of spin j).

The single-node beliefs $b_i(s_i)$ can be parameterized by their means ("magnetizations")

$$m_i := \langle s_i\rangle_{b_i} = \sum_{s_i} s_i b_i(s_i), \tag{14}$$

and the pairwise beliefs $b_{ij}(s_i, s_j)$ can be parameterized by m_i, m_j and the second order moment ("correlation")

$$\chi_{ij} := \langle s_i s_j\rangle_{b_{ij}} = \sum_{s_i,s_j} s_i s_j b_{ij}(s_i, s_j). \tag{15}$$

The beliefs (9) and (10) at a fixed point ν^∞ can then simply be written as:

$$m_i = \tanh(\beta h_{i\backslash j}^\infty + \nu_{ji}^\infty), \tag{16}$$

$$\chi_{ij} = \tanh\left(\beta J_{ij} + \tanh^{-1}\left(\tanh(\beta h_{i\backslash j}^\infty)\tanh(\beta h_{j\backslash i}^\infty)\right)\right). \tag{17}$$

3.2. Local stability of undamped, parallel LBP fixed points

For the parallel update scheme, we can consider the update mapping $F : \nu \mapsto \nu^{\text{new}}$ written out in components in (12). Its derivative ("Jacobian") is given by:

$$F'(\nu) = \frac{\partial \nu_{ij}^{\text{new}}}{\partial \nu_{kl}} = \frac{1 - \tanh^2(\beta h_{i\backslash j})}{1 - \tanh^2(\beta J_{ij}) \tanh^2(\beta h_{i\backslash j})} \tanh(\beta J_{ij}) \, \mathbf{1}_{N_i \backslash j}(k) \, \delta_{i,l} \tag{18}$$

where $\mathbf{1}$ is the indicator function (i.e. $\mathbf{1}_X(x) = 1$ if $x \in X$ and 0 otherwise) and δ the Kronecker delta function.

Let ν be a fixed point of parallel LBP. We call ν *locally stable* if starting close enough to the fixed point, LBP will converge to it. A fixed point ν is locally stable if all eigenvalues of the Jacobian $F'(\nu)$ lie inside the unit circle in the complex plane [19]:

$$\nu \text{ is locally stable } \iff \sigma(F'(\nu)) \subseteq \{\lambda \in \mathbf{C} : |\lambda| < 1\}, \tag{19}$$

where $\sigma(F')$ denotes the *spectrum* (set of eigenvalues) of the matrix F'. If at least one eigenvalue lies outside the unit circle, the fixed point is unstable.

3.3. Local stability conditions for damped, parallel LBP

The LBP equations can in certain cases lead to oscillatory behaviour, which may be remedied by damping the update equations. This can be done by replacing the update map $F : \nu \mapsto \nu$ by the convex combination $F_\epsilon := (1-\epsilon)F + \epsilon I$ of F and the identity I, for damping strength $0 \leq \epsilon < 1$. Fixed points of F are also fixed points of F_ϵ and vice versa. The spectrum of the local stability matrix of the damped LBP update mapping becomes:

$$\sigma(F'_\epsilon(\nu)) = (1-\epsilon)\sigma(F'(\nu)) + \epsilon.$$

In words, all eigenvalues of the local stability matrix *without damping* are simply interpolated with the value 1 for damped LBP. It follows that the condition for (local) stability of a fixed point ν under *arbitrarily large* damping is given by

$$\nu \text{ is stable under } F_\epsilon \text{ for some damping } \epsilon \quad \iff \sigma(F'(\nu)) \subseteq \{\lambda \in \mathbf{C} : \Re\lambda < 1\}, \tag{20}$$

i.e. all eigenvalues of $F'(\nu)$ should have real part smaller than 1.

Note that conditions (19) and (20) do not depend on the chosen parameterization of the messages. In other words, the local stability of the LBP fixed points does not depend on whether one uses μ_{ij} messages or ν_{ij} messages, or some other parameterization, i.e. the choice made in (11) has no influence on the results, but it does simplify the calculations.

3.4. Uniqueness of LBP fixed points and convergence

The foregoing conditions are local and by themselves are not strong enough for drawing conclusions about global behaviour, i.e. whether or not LBP will converge for any initial set of messages.

In [20] we have derived sufficient conditions for the uniqueness of the LBP fixed point and convergence of undamped, parallel LBP to the unique fixed point, irrespective of the initial messages. For the binary case, our result can be stated as follows:§

Theorem 1 *If the spectral radius‖ of the square matrix*

$$B_{ij,kl} := \tanh(\beta\,|J_{ij}|)\delta_{i,l}\mathbf{1}_{N_i\setminus j}(k) \tag{21}$$

is strictly smaller than 1, undamped parallel LBP converges to a unique fixed point, irrespective of the initial messages.

Proof. See [20]. □

Note that the matrix B, and hence the sufficient condition, depends neither on the fields θ_i, nor on the sign of the weights J_{ij}.

These conditions are sufficient, but by no means necessary, as we will see in the next section. However, for ferromagnetic interactions without local fields, they are sharp, as we will prove later on. First we discuss some properties of the Bethe free energy that we will need in section 4.

3.5. Properties of the Bethe free energy for binary variables

For the case of binary variables, the Bethe free energy (4) can be parameterized in terms of the means $m_i = \langle s_i \rangle_{b_i}$ and correlations $\chi_{ij} = \langle s_i s_j \rangle_{b_{ij}}$; it becomes:

$$\begin{aligned} F_{Be}(m,\chi) := & -\beta \sum_{(ij)\in B} \beta J_{ij}\chi_{ij} - \beta \sum_i \theta_i m_i \\ & + \sum_{i=1}^{N}(1-d_i) \sum_{s_i=\pm 1} \eta\left(\frac{1+m_i s_i}{2}\right) \\ & + \sum_{(ij)\in B} \sum_{s_i,s_j=\pm 1} \eta\left(\frac{1+m_i s_i + m_j s_j + s_i s_j \chi_{ij}}{4}\right) \end{aligned} \tag{22}$$

where $\eta(x) := x \log x$. The normalization and consistency constraints (6) and (7) are satisfied automatically; however now we need to enforce positivity constraints

$$\begin{aligned} & -1 \le m_i \le 1 \\ & -1 \le \chi_{ij} \le 1 \\ & 1 + m_i\sigma + m_j\sigma' + \chi_{ij}\sigma\sigma' \ge 0 \quad \text{for all } \sigma,\sigma' = \pm 1 \end{aligned}$$

which guarantee that the beliefs $\{b_i\}_{i\in V}$ and $\{b_{ij}\}_{(ij)\in B}$ are positive. The stationary points of the Bethe free energy (22) are the points where the derivative of (22) vanishes;

§ An equivalent result but formulated in terms of an algorithm was derived independently in [21].
‖ The spectral radius $\rho(B)$ of a matrix B is defined as $\rho(B) := \sup|\sigma(B)|$, i.e. it is the largest absolute value of the eigenvalues of B.

this yields the following equations:

$$0 = \frac{\partial F_{Be}}{\partial m_i} = -\beta\theta_i + (1-d_i)\tanh^{-1} m_i + \\ + \sum_{j\in N_i} \tfrac{1}{4}\log\frac{(1+m_i+m_j+\chi_{ij})(1+m_i-m_j-\chi_{ij})}{(1-m_i+m_j-\chi_{ij})(1-m_i-m_j+\chi_{ij})}. \tag{23}$$

$$0 = \frac{\partial F_{Be}}{\partial \chi_{ij}} = -\beta J_{ij} + \tfrac{1}{4}\log\frac{(1+m_i+m_j+\chi_{ij})(1-m_i-m_j+\chi_{ij})}{(1+m_i-m_j-\chi_{ij})(1-m_i+m_j-\chi_{ij})}. \tag{24}$$

The last equation has a unique solution χ_{ij} as a function of m_i and m_j [22].

From now on we consider the special case of vanishing local fields (i.e. $\theta_i = 0$) in the interest of simplicity. Note that in this case, the LBP update equations (12) have a trivial fixed point, namely $\nu_{ij} = 0$. The corresponding beliefs have $m_i = 0$ and $\chi_{ij} = \tanh(\beta J_{ij})$, as follows directly from (16); of course, this also follows from (23) and (24). We call this fixed point the *paramagnetic fixed point* (or the *high-temperature fixed point* to emphasize that it exists for high enough temperature, i.e. for β small enough).

Whether the paramagnetic stationary point of the Bethe free energy is indeed a minimum depends on whether the Hessian of F_{Be} is positive-definite. The Hessian at the paramagnetic stationary point is given by:

$$\frac{\partial^2 F_{Be}}{\partial m_j \partial m_i} = \delta_{ij}\left(1 + \sum_{k\in N_i}\frac{\chi_{ik}^2}{1-\chi_{ik}^2}\right) + M_{ij}\frac{-\chi_{ij}}{1-\chi_{ij}^2} =: U_{ij}, \tag{25}$$

$$\frac{\partial^2 F_{Be}}{\partial m_k \partial \chi_{ij}} = 0,$$

$$\frac{\partial^2 F_{Be}}{\partial \chi_{kl} \partial \chi_{ij}} = \delta_{(ij),(kl)}\frac{1}{1-\chi_{ij}^2}.$$

The Hessian is of block-diagonal form; the χ-block is always positive-definite, hence the Hessian is positive-definite if and only if the the m-block (U_{ij}) is positive-definite. This depends on the weights J_{ij} and on the graph structure; for β small enough (i.e. high temperature), this is indeed the case. A consequence of the positive-definiteness of the Hessian of the Bethe free energy is that the approximate covariance matrix, given by U^{-1}, is also positive-definite.

4. Phase transitions

In this section we discuss various phase transitions that may occur, depending on the distribution of the weights J_{ij}. We take the local fields θ_i to be zero. Our usage of the term "phase transition" is somewhat inaccurate, since we actually mean the finite-N manifestations of the phase transition in the Bethe approximation and in the dynamical behaviour of the LBP algorithm, instead of the common usage of the word, which refers to the $N \to \infty$ behaviour of the exact probability distribution. We conjecture though, that at least for the ferromagnetic and spin-glass phase transitions, these different notions coincide in the $N \to \infty$ limit.

4.1. Ferromagnetic interactions

Consider the case of purely ferromagnetic interactions, by which we mean that all interactions J_{ij} are positive. In that case, the local LBP stability matrix $F'(0)$ at the trivial fixed point, given by

$$F'(0) = \tanh(\beta J_{ij})\mathbf{1}_{N_i \setminus j}(k)\delta_{i,l} \tag{26}$$

is equal to the matrix B in Theorem 1. For high temperature (i.e. small β), the paramagnetic fixed point is locally stable, as is evident from (26). Theorem 1 guarantees that this is the only LBP fixed point and that parallel undamped LBP will converge to it. When we gradually lower the temperature (i.e. increase β), at a sudden point the paramagnetic LBP fixed point generally becomes unstable. This seems to hold for all graphs that have more than one cycle. By a generalization of Perron's theorem (Theorem 3 in the Appendix), the eigenvalue of the matrix $F'(0)$ (which has positive entries) with the largest absolute value is actually positive. This property of the spectrum can be clearly seen in figure 1.I(a), where most eigenvalues are distributed in a roughly circular form, except for one outlier on the positive real axis. Thus the onset of instability of the paramagnetic LBP fixed point coincides with this outlier crossing the complex unit circle; the paramagnetic fixed point bifurcates and two new stable fixed points arise, describing the two ferromagnetic states. Since $B = F'(0)$, we conclude that the sufficient condition in Theorem 1 for convergence to a unique fixed point is sharp in this case.

At high temperature, the corresponding stationary point of the Bethe free energy is a minimum. However, as illustrated in figure 1.II(a), at a certain critical temperature the Hessian is no longer positive-definite. In the Appendix, we prove the following theorem:

Theorem 2 *For $J_{ij} \geq 0$ and $\theta_i = 0$, the critical temperature at which the paramagnetic Bethe free energy minimum disappears is equal to the critical temperature at which the paramagnetic LBP fixed point becomes unstable.*

Proof. See Appendix. □

Beyond the transition temperature, LBP converges to either of the two new fixed points describing the two ferromagnetic phases. As can be seen in figure 1.I(c), the number of LBP iterations needed for convergence has a peak precisely at the critical temperature; far from the phase transition, LBP converges rapidly to a stable fixed point.

4.2. Anti-ferromagnetic interactions

For purely anti-ferromagnetic interactions, i.e. all $J_{ij} < 0$, the situation is different. Again, for high temperature, the paramagnetic fixed point is the unique fixed point, is locally stable and has the complete message space as an attractor. Since the local stability matrix $F'(0)$ is exactly the same as in the ferromagnetic case, except for

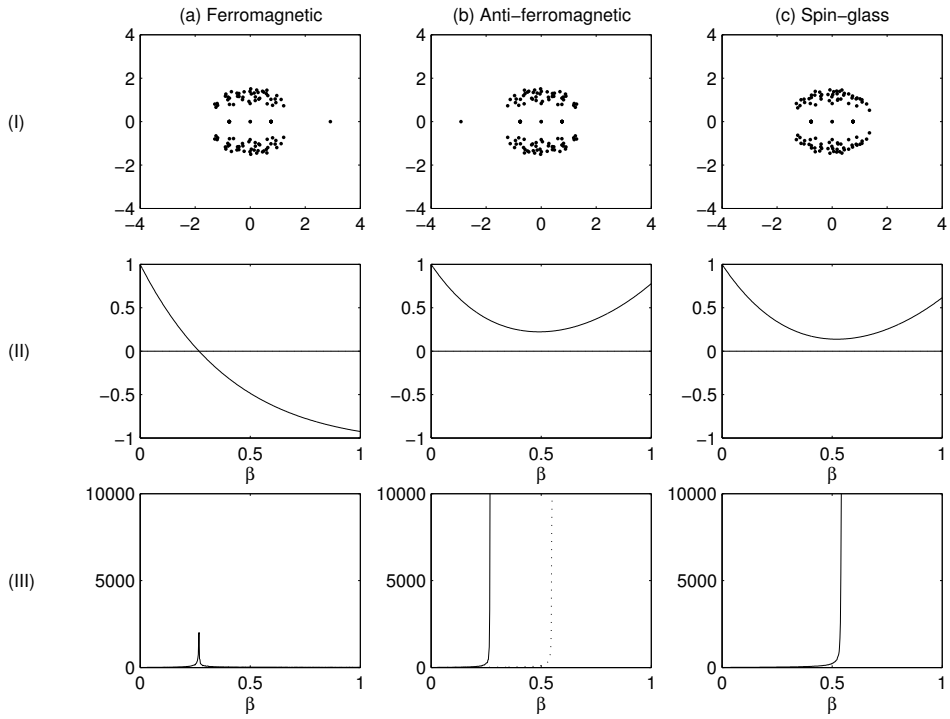

Figure 1. From top to bottom: (I) spectrum of the local LBP stability matrix F' at the trivial fixed point $\nu = 0$, for $\beta = 1$; (II) minimal eigenvalue of $U_{ij} = \partial^2 F_{Be}/\partial m_i \partial m_j$ at the paramagnetic solution, as a function of inverse temperature β; (III) number of undamped, parallel LBP iterations needed for convergence as a function of inverse temperature β (dotted line in antiferromagnetic case shows the number of iterations for a sequential update scheme). From left to right: (a) ferromagnetic interactions $J = M$ (b) antiferromagnetic interactions $J = -M$; (c) spin-glass interactions $J = \pm M$ with equal probability for positive or negative interaction.

The underlying graph G is a random graph with Poissonian degree distribution, $N = 50$ and average degree $d = 4$; the local fields are zero.

the minus sign (as can be seen in figure 1.I(b)), the local stability of the trivial fixed point is invariant under a sign change $J \mapsto -J$. Hence the paramagnetic fixed point becomes locally unstable for undamped LBP exactly at the same temperature as in the ferromagnetic case, for fixed weight strengths $|J_{ij}|$. However, the spectral radius of $F'(0)$ is now determined by a negative eigenvalue. Hence in this case damping helps to some extent. Empirically, we find that also changing the update scheme from parallel to sequential helps, as illustrated by the dotted line in figure 1.III(b). Note that the temperature where sequential LBP stops converging roughly coincides with the minimum of the smallest eigenvalue of U (compare figure 1.II(b) and 1.III(b)). This observation seems to be generic, i.e. not just a coincidence for the particular instance in figure 1. We have no theoretical explanation for this at the moment, but it might be possible to get such an explanation by relating U with $F'(0)$, using a technique similar to the one applied in the proof of Theorem 2 given in the Appendix.

4.3. Spin-glass interactions

Now consider spin-glass interactions, i.e. all J_{ij} are distributed around 0 such that $\langle J_{ij} \rangle \approx 0$. This case is illustrated in figure 1(c). Here the eigenvalues of the local stability matrix are distributed in a roughly circular form, without an outlier with a large absolute value. Note the suprising similarity between the spectra in the different cases; we have no explanation for this similarity, nor for the roughly circular form of the distribution of the majority of the eigenvalues.

Although the paramagnetic Bethe free energy minimum generally does not disappear when lowering the temperature, LBP does not converge anymore once the trivial fixed point becomes unstable, despite the possible existence of other, stable, fixed points. Neither damping nor changing the update scheme seems to help in this case. Empirically we find that the temperature at which the trivial LBP fixed point becomes locally unstable roughly coincides with the temperature at which the lowest eigenvalue of U attains its minimal value [23]. Again, we have no theoretical explanation for this observation.

5. Estimates of the phase transition temperatures

In this section we estimate the critical temperatures corresponding to the onset of instability of the LBP paramagnetic fixed point (which we discussed qualitatively in the previous section) for a random graph with random interactions. The method is closely related to the cavity method at the replica-symmetric level (see e.g. [24, 18, 25]). A similar analysis of the stability of the LBP paramagnetic fixed point has been done by Kabashima [26]; however, the results reported in that work are limited to the case of infinite connectivity (i.e. the limit $N \to \infty$, $d \to \infty$). In this case, the results turn out to be identical to the condition of replica symmetry breaking derived by Almeida and Thouless (the "AT line") [27]. The analysis we present below essentially extends the analysis of [26] to the larger class of arbitrary degree distribution random graphs, which includes Erdős-Rényi graphs (with Poissonian degree distribution, as well as fixed degree random graphs) and power-law graphs (which have power-law degree distributions), amongst others.

5.1. Random graphs with arbitrary degree distributions

We consider arbitrary degree distribution random graphs [28]. This class of random graphs has a prescribed expected degree distribution $P(d)$; apart from that they are completely random. Given an expected degree distribution $P(d)$ and the number of nodes N, a particular sample of the corresponding ensemble of random graphs can be constructed as follows: for each node i, independently draw an expected degree δ_i from the degree distribution $P(d)$. Then, for each pair of nodes (i, j), independently connect them with probability $\delta_i \delta_j / \sum_i \delta_i$; the expected degree of node i is then indeed

$\langle d_i \rangle = \delta_i$. We define the *average degree* $\langle d \rangle := \sum_d P(d)d$ and the second moment $\langle d^2 \rangle := \sum_d P(d)d^2$.

We consider the case of vanishing local fields (i.e. $\theta_i = 0$) and draw the weights J_{ij} independently from some probability distribution $P(J)$. We also assume that the weights are independent of the graph structure.

5.2. Estimating the PA-FE transition temperature

Assume $P(d)$ to be given and N to be large. Assume that x is an eigenvector with eigenvalue 1 of $A := F'(0)$, the Jacobian of the parallel LBP update at the paramagnetic fixed point $\nu = 0$. Using (18):

$$x_{ij} = \sum_{kl} A_{ij,kl} x_{kl} = \tanh(\beta J_{ij}) \sum_{k \in N_i \setminus j} x_{ki}. \tag{27}$$

Consider an arbitrary spin i; conditional on the degree d_i of that spin, we can calculate the expected value of x_{ij} as follows:

$$\begin{aligned} E\left(x_{ij} \,|\, d_i\right) &= E\left(\tanh(\beta J_{ij}) \textstyle\sum_{k \in N_i \setminus j} x_{ki} \,|\, d_i\right) && (28a) \\ &= E\left(\tanh(\beta J_{ij})\right) E\left(\textstyle\sum_{k \in N_i \setminus j} x_{ki} \,|\, d_i\right) && (28b) \\ &= \langle \tanh \beta J \rangle \left(d_i - 1\right) \sum_{d_k} P(d_k \,|\, d_i, k \in N_i) E\left(x_{ki} \,|\, d_i, d_k\right) && (28c) \\ &\approx \langle \tanh \beta J \rangle \left(d_i - 1\right) \sum_{d_k} P(d_k \,|\, d_i, k \in N_i) E\left(x_{ki} \,|\, d_i\right) && (28d) \end{aligned}$$

using, subsequently: (a) equation (27); (b) the independence of the weights from the graph structure; (c) conditioning on the degree d_k of spin k and the equivalence of the various $k \in N_i \setminus j$; and finally, (d) neglecting the correlation between x_{ki} and d_k, given d_i. We have no formal argument for the validity of this approximation, but the result accurately describes the outcomes of numerical experiments.

For arbitrary degree distribution random graphs, the probability of d_k given the degree d_i and the fact that k is a neighbour of i is given by (see [28]):

$$P(d_k \,|\, d_i, k \in N_i) = \frac{d_k P(d_k)}{\langle d \rangle}.$$

Hence we obtain the relation

$$E\left(x_{ij} \,|\, d_i\right) = \langle \tanh \beta J \rangle \left(d_i - 1\right) \sum_{d_k} \frac{d_k P(d_k)}{\langle d \rangle} E\left(x_{ki} \,|\, d_k\right) \tag{29}$$

A self-consistent nontrivial solution of these equations is $E\left(x_{ij} \,|\, d_i\right) \propto (d_i - 1)$, provided that

$$1 = \langle \tanh \beta J \rangle \left(\frac{\langle d^2 \rangle}{\langle d \rangle} - 1 \right). \tag{30}$$

which gives us the critical temperature at which the paramagnetic–ferromagnetic phase transition occurs, or in other words, where the paramagnetic LBP fixed point undergoes

a pitchfork bifurcation. This result is identical to the one obtained by the replica method in the replica-symmetric setting [29] and to the one found by applying the cavity method [25], as expected. Figure 2 illustrates the estimate; note that the accuracy is quite high already for low N ($N = 50$ in this case), for higher N it becomes even better.

Extending the analysis to the case of non-vanishing local fields does not appear to be straightforward, since in that case the value of the fixed point ν is not known. However, since the elements of A are upper bounds for the elements of $F'(\nu)$, we can at least qualitatively conclude that in the case of non-vanishing local fields, the transition temperature will be lower.

5.3. The antiferromagnetic case

This is similar to the ferromagnetic case, however the eigenvalue is now -1 instead of $+1$. This yields the following equation for the transition temperature:

$$1 = \langle \tanh(-\beta J) \rangle \left(\frac{\langle d^2 \rangle}{\langle d \rangle} - 1 \right). \tag{31}$$

Again the prediction turns out to be quite accurate (see figure 2), as was to be expected.

5.4. Estimating the PA-SG transition temperature

For the paramagnetic–spin-glass phase transition, we can perform a similar calculation, now assuming that x is an eigenvector with eigenvalue λ on the complex unit circle:

$$\begin{aligned} E\left(|x_{ij}|^2 \mid d_i\right) &= E\left(|\tanh(\beta J_{ij})|^2 \left| \textstyle\sum_{k \in N_i \setminus j} x_{ki} \right|^2 \mid d_i \right) \\ &= \left\langle \tanh^2(\beta J) \right\rangle E\left(\left| \textstyle\sum_{k \in N_i \setminus j} x_{ki} \right|^2 \mid d_i \right) \\ &\approx \left\langle \tanh^2(\beta J) \right\rangle E\left(\textstyle\sum_{k \in N_i \setminus j} |x_{ki}|^2 \mid d_i \right) \\ &\approx \left\langle \tanh^2(\beta J) \right\rangle (d_i - 1) \sum_{d_k} P(d_k \mid d_i, k \in N_i) E\left(|x_{ki}|^2 \mid d_i \right), \end{aligned}$$

where, in addition to the assumptions in the PA-FE case, we assumed that the correlations between the various x_{ki}'s can be neglected. Again, we can only motivate this assumption in that it appears to give correct results.

Using relation (29), we find a nontrivial self-consistent solution $E\left(|x_{ij}|^2 \mid d_i\right) \propto (d_i - 1)$, if the following equation holds:

$$1 = \left\langle \tanh^2(\beta J) \right\rangle \left(\frac{\langle d^2 \rangle}{\langle d \rangle} - 1 \right). \tag{32}$$

This result is again identical to the one obtained by the cavity method [25], as expected. As illustrated in figure 2 (the dashed line), the accuracy is somewhat less than that of the ferromagnetic transition, but is nevertheless quite good, even for $N = 50$.

For completeness we would like to state that the numerical results reported in [23], in which we numerically studied the behaviour of the lowest eigenvalue of U, are

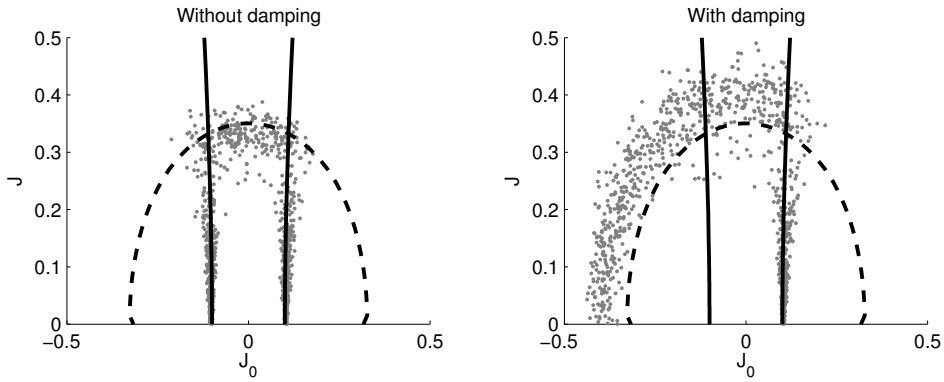

Figure 2. Onset of instability of the paramagnetic LBP fixed point, for random graphs with $N = 50$ and a Poissonian degree distribution with $d = 10$. The weights J_{ij} are independently drawn from a Gaussian distribution with mean J_0 and variance J^2. The solid thick lines show the expected value for the (anti)ferromagnetic transitions (30) and (31), the dashed thick line for the spin-glass transition (32). The dots show for individual instances at which temperature the paramagnetic fixed point becomes unstable, for undamped LBP (left) and for damped LBP (right). The lines in the right graph (the damped case) are for reference only, they should not be interpreted as theoretical predictions, except for the ferromagnetic transition (the solid line on the right-hand side).

accurately described by the predictions (30) and (32), which supports the hypothesis that these notions coincide in the $N \to \infty$ limit.

6. Conclusions

We have derived conditions for the local stability of parallel LBP fixed points, both in the undamped and damped case for binary networks with pairwise interactions. We have shown how these relate to the sufficient conditions for uniqueness of the LBP fixed point and convergence to this fixed point. In particular, we have shown that these sufficient conditions are sharp in the ferromagnetic case, exactly describing the pitchfork bifurcation of the paramagnetic fixed point into two ferromagnetic fixed points. For undamped LBP, the local stability of the paramagnetic fixed point (for vanishing local fields) is invariant under a sign change of the interactions. For anti-ferromagnetic interactions, parallel undamped LBP stops converging at the PA-FE transition temperature. Damping or using a sequential update scheme remedy this defect. However, although the paramagnetic minimum of the Bethe free energy does not disappear, the trivial fixed point becomes locally unstable even for damped LBP at roughly the PA-SG transition temperature. Finally, for interactions that are dominantly of the spin-glass type, using damping only marginally extends the domain of convergence of LBP.

We estimated the PA-FE transition temperature and the PA-SG transition temperature for arbitrary degree distribution random graphs. The results are in good

agreement with numerical simulations. How this relates to the AT line is an open question and beyond the scope of this work.

We believe that the case that we have considered in detail in this work, namely vanishing local fields $\theta_i = 0$, is actually the worst-case scenario: numerically it turns out that adding local fields helps LBP to converge more quickly. We have no proof for this conjecture at the moment; the local fields make an analytical analysis more difficult and we have not yet been able to extend the analysis to this more general setting. We leave the generalization to non-zero local fields as possible future work.

Acknowledgments

The research reported here is part of the Interactive Collaborative Information Systems (ICIS) project, supported by the Dutch Ministry of Economic Affairs, grant BSIK03024. We thank Bastian Wemmenhove and the anonymous reviewer for valuable comments on a draft of this paper.

Appendix: Proof of Theorem 2

For a square matrix B, we write $B \geq 0$ iff all entries of B are non-negative. $\sigma(B)$ is the set of all eigenvalues of B, $\rho(B)$ is the spectral radius of B, i.e. $\rho(B) := \max|\sigma(B)|$. We will use the following generalization of Perron's theorem:

Theorem 3 *If $B \geq 0$, then the spectral radius $\rho(B) \in \sigma(B)$ and there exists an associated eigenvector $x \geq 0$ such that $Bx = \rho(B)x$.*

Proof. See [30, p. 670]. □

Applying this theorem to the matrix B defined in (21), we deduce the existence of an eigenvector $x \geq 0$ with $Bx = \rho(B)x$. Writing $C_{ij} := \tanh(\beta\,|J_{ij}|)$ and $\lambda := \rho(B)$, we derive:

$$\begin{aligned} x_{ij} &= \lambda^{-1} C_{ij} \left(\sum_{k \in N_i} x_{ki} - x_{ji} \right) \\ &= \lambda^{-1} C_{ij} \left(\sum_{k \in N_i} x_{ki} - \lambda^{-1} C_{ji} \left(\sum_{k \in N_j} x_{kj} - x_{ij} \right) \right). \end{aligned}$$

Defining $X_i := \sum_{k \in N_i} x_{ki}$, we obtain by summing over $i \in N_j$:

$$X_j = \sum_{i \in N_j} \lambda \frac{C_{ij}}{\lambda^2 - C_{ij}C_{ji}} X_i - \sum_{i \in N_j} \frac{C_{ij}C_{ji}}{\lambda^2 - C_{ij}C_{ji}} X_j,$$

i.e. X is an eigenvector with eigenvalue 1 of the matrix

$$M_{ij} \frac{\rho(B)\tanh(\beta\,|J_{ij}|)}{\rho(B)^2 - \tanh^2(\beta\,|J_{ij}|)} - \delta_{ij} \sum_{k \in N_i} \frac{\tanh^2(\beta\,|J_{ik}|)}{\rho(B)^2 - \tanh^2(\beta\,|J_{ik}|)} \tag{A.1}$$

Now, if all J_{ij} are positive, and if $\rho(B) = 1$, this matrix is exactly $I - U$, where U_{ij} is defined in (25). Hence, since in this case $B = F'(0)$, the critical temperature at which the paramagnetic LBP fixed point becomes unstable coincides with the matrix $I - U$ having an eigenvalue 1, or in other words U having eigenvalue 0. Thus the onset of instability of the paramagnetic LBP fixed point in this case exactly coincides with the disappearance of the paramagnetic Bethe free energy minimum.

References

[1] Bethe H 1935 *Proc. R. Soc. A* **150** 552
[2] Peierls R E 1936 *Proc. Cambridge Philos. Soc.* **32** 477
[3] Kikuchi R 1951 *Phys. Rev.* **81** 988
[4] Pelizzola A Aug 2005 *J. Phys. A: Math. Gen.* **38** R309
[5] Tanaka K 2002 *J. Phys. A: Math. Gen.* **35** R81
[6] Pearl J 1988 *Probabilistic Reasoning in Intelligent systems: Networks of Plausible Inference* (San Francisco, CA: Morgan Kaufmann)
[7] McEliece R J, MacKay D J C and Cheng J F Feb 1998 *IEEE J. Select. Areas Commun.* **16** 140
[8] Nishimori H 2001 *Statistical Physics of Spin Glasses and Information Processing - an Introduction* (Oxford: Oxford Press)
[9] Yedidia J S, Freeman W T and Weiss Y 2001 in *Advances in Neural Information Processing Systems 13 (NIPS*2000)*, eds. L K Saul, Y Weiss and L Bottou (Cambridge, MA: MIT Press) pp. 689–695
[10] Braunstein A and Zecchina R 2004 *J. Stat. Mech.: Theor. Exp.* P06007
[11] Freeman W, Pasztor E and Carmichael O 2000 *Int. J. Comput. Vis.* **40** 25
[12] Kappen H 2002 in *Modelling biomedical signals* (Singapore: World Scientific) pp. 3–27
[13] Baxter R 1982 *Exactly Solved Models in Statistical Mechanics* (New York: Academic)
[14] Welling M and Teh Y W 2003 *Artificial Intelligence* **143** 19
[15] Kschischang F R, Frey B J and Loeliger H A Feb 2001 *IEEE Trans. Inform. Theory* **47** 498
[16] Heskes T, Albers C and Kappen H J 2003 in *Proc. of the 19th Annual Conf. on Uncertainty in Artificial Intelligence (UAI-03)* (San Francisco, CA: Morgan Kaufmann Publishers) pp. 313–320
[17] Heskes T Nov 2004 *Neural Computation* **16** 2379
[18] Mézard M and Parisi G 2001 *Eur. Phys. Jour. B* **20** 217
[19] Kuznetsov Y A 1988 *Elements of Applied Bifurcation Theory* vol. 112 of *Applied Mathematical Sciences* (New York: Springer) 2nd ed.
[20] Mooij J and Kappen H 2005 in *Proc. of the 21st Annual Conf. on Uncertainty in Artificial Intelligence (UAI-05)* (Corvallis, Oregon: AUAI Press) pp. 396–403
[21] Ihler A T, Fisher J W and Willsky A S May 2005 *Journal of Machine Learning Research*
[22] Max M W and Teh Y W 2001 in *Proc. of the 17th Annual Conf. on Uncertainty in Artificial Intelligence (UAI-01)* (San Francisco, CA: Morgan Kaufmann Publishers) pp. 554–561
[23] Mooij J and Kappen H 2005 in *Advances in Neural Information Processing Systems 17 (NIPS*2004)*, eds. L K Saul, Y Weiss and L Bottou (Cambridge, MA: MIT Press)
[24] Mezard M, Parisi G and Virasoro M A 1987 *Spin glass theory and beyond* (Singapore: World Scientific)
[25] Wemmenhove B, Nikoletopoulos T and Hatchett J P L 2004 *preprint* **cond-mat/0405563**
[26] Kabashima Y 2003 *J. Phys. Soc. Japan* **72** 1645
[27] de Almeida J R L and Thouless D J 1978 *J. Phys. A* **11** 983
[28] Newman M E J, Strogatz S H and Watts D J Aug 2001 *Phys. Rev. E* **64** 026118
[29] Leone M, Vázquez A, Vespignani A and Zecchina R 2002 *Eur. Phys. Jour. B* **28** 191
[30] Meyer C D 2000 *Matrix Analysis and Applied Linear Algebra* (Philadelphia: SIAM)