# On Loopy Belief Propagation – Local Stability Analysis for Non-Vanishing Fields

**Christian Knoll**
Graz University of Technology
christian.knoll.c@ieee.org

**Franz Pernkopf**
Graz University of Technology
pernkopf@tugraz.at

## Abstract

In this work we obtain all fixed points of belief propagation and perform a local stability analysis. We consider pairwise interactions of binary random variables and investigate the influence of non-vanishing fields and finite-size graphs on the performance of belief propagation; local stability is heavily influenced by these properties. We show why non-vanishing fields help to achieve convergence and increase the accuracy of belief propagation. We further explain the close connections between the underlying graph structure, the existence of multiple solutions, and the capability of belief propagation (with damping) to converge. Finally we provide insights into why finite-size graphs behave better than infinite-size graphs.

## 1 INTRODUCTION

Belief propagation (BP) is a method that exploits the structure of probabilistic graphical models and performs local operations to approximate the marginal distribution. BP has a long success story with applications including computer vision, speech processing, and medical diagnosis systems (Koller and Friedman, 2009; Pernkopf et al., 2014; Jordan, 2004)

BP may not converge on graphs with loops – only for restricted classes of graphs sufficient conditions are known that guarantee convergence and uniqueness of the solution (Ihler et al., 2005; Heskes, 2004; Weiss, 2000). If BP does not converge or has a bad approximation quality we may be confronted with: (i) multiple solutions where the performance depends on the initialization, or (ii) potentially good approximate solutions that are unstable.

To overcome some of these issues a lot of effort was put into finding provably convergent algorithms. Inspired by the connection between BP and the Bethe free energy (Yedidia et al., 2005) it is natural to minimize the Bethe free energy directly (Welling and Teh, 2001), or find convex surrogates to minimize (Meshi et al., 2009). Alternatively, if multiple fixed points exist, one can try to obtain one stationary point of the Bethe free energy directly (Yuille and Rangarajan, 2003), or get at least some approximation of the stationary points (Shin, 2012). For graphs where fixed points close to the exact solution exist, but BP fails to converge, (Weller and Jebara, 2014) provide a method to obtain an approximation of the best possible fixed point.

All these methods do still raise particular questions, like: under which conditions can we expect BP to perform well?– or, given that multiple fixed points exist, which of these fixed points are stable and how does stability relate to accuracy?

We consider BP as a discrete-time map, and – drawing from the theory of dynamical systems – analyze the local stability of all fixed points. This is inspired by the thorough analysis for Ising models with vanishing local fields (Mooij and Kappen, 2005); for non-vanishing fields, however, local stability analysis was left open. The main difficulty of local stability analysis of Ising models with arbitrary parameters on finite-size graphs is the need for *all* fixed points. To find multiple fixed points one could resort to survey propagation (Braunstein et al., 2005) or its efficient approximation scheme (Srinivasa et al., 2016), but these methods fail to obtain unstable fixed points. Unfortunately, unstable fixed points cannot be obtained in general, which, in general, makes local stability analysis much more intricate.

In this work, we resort to the numerical polynomial-homotopy-continuation (NPHC) method (Knoll et al., 2016), where the update equations of BP are reformulated as a nonlinear system of equations. The NPHC method solves this system of equations and obtains all fixed points of BP – even for Ising graphs with arbitrary

parameters. Consequently, we extend the work of (Mooij and Kappen, 2005) and find answers to the following questions: (i) Why does the presence of non-vanishing fields enhance the convergence properties? (ii) Which influence do non-vanishing fields have on the approximation accuracy? (iii) Under which circumstances does damping help to enforce convergence of BP? (iv) How does the number of fixed points and their stability depend on the number of variables for finite-size graphs?

The paper is structured as follows: Section 2 provides a brief background on probabilistic graphical models, belief propagation, and its connection to statistical physics. In Section 3 we present tools from dynamical systems to perform the stability analysis. We discuss empirical observations and extend the stability analysis to graphs with non-vanishing local fields in Section 4 and provide a more formal analysis in Section 5. Finally, we conclude the paper in Section 6.

# 2 BACKGROUND

Let us consider an undirected graph $G = (\mathbf{X}, \mathbf{E})$ defined by a set of nodes $\mathbf{X} = \{X_1, \dots, X_N\}$ and a set of undirected edges $\mathbf{E}$. Two nodes $X_i$ and $X_j$ are joined by an edge if $e_{ij} \in \mathbf{E}$. Let the set of neighbors of $X_i$ be defined by $\partial(X_i) = \{X_j \in \mathbf{X} : e_{ij} \in \mathbf{E}\}$ and denote the degree of $X_i$ by $d_i = |\partial(X_i)|$. Consider the adjacency matrix $\boldsymbol{A}$ with rows and columns indexed by the set of nodes; $\boldsymbol{A}$ is a symmetric 0-1-matrix with $a_{ij} = 1$ if and only if $e_{ij} \in \mathbf{E}$. We further define the average degree of $G$ as $\langle d \rangle_G = \frac{|\mathbf{E}|}{N}$. A graph is called bipartite if the set of nodes can be decomposed into two disjoint subsets $\mathbf{Y}$ and $\mathbf{Z}$ so that $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$ and where the neighbors $\partial(Y_i) \in \mathbf{Z}$ for all $Y_i \in \mathbf{Y}$ and vice versa. It follows immediately that bipartite graphs are 2-colorable.

Consider a finite set of $N$ binary random variables $\mathbf{X} = \{X_1, \dots, X_N\}$ with $X_i$ taking values in the finite set $\mathbb{S}_i$. We assume a one-to-one correspondence between the variables and the nodes. The joint probability factorizes as a product of potentials, i.e., $P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \prod_{l=1}^{L} \Phi_{C_l}(C_l)$, where the potentials $\Phi_{C_l}(C_l)$ are specified over the maximal cliques $C_l$ of $G$ (Pearl, 1988, p.105). If we restrict all potentials to consist of two variables at most: then, the joint distribution is factorized according to the product

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \prod_{(X_i, X_j) \,:\, e_{ij} \in \mathbf{E}} \Phi_{X_i, X_j}(x_i, x_j) \prod_{i=1}^{N} \Phi_{X_i}(x_i), \quad (1)$$

where $Z \in \mathbb{R}$ guarantees proper normalization. We refer to $\Phi_{X_i, X_j}(x_i, x_j)$ and $\Phi_{X_i}(x_i)$ as pairwise and local potentials. The marginal probabilities are obtained by summing over all variables except $X_i$, i.e., $P(X_i = x_i) = \sum_{\sim X_i} P(\mathbf{X})$.

## 2.1 BELIEF PROPAGATION (BP)

BP[1] approximates the marginals by recursively updating local messages between random variables. The message $\mu_{ij}^{n+1}(x_j)$ from $X_i$ to $X_j$ is given by the recursive update rule

$$\mu_{ij}^{n+1}(x_j) = \alpha_{ij}^{n} \sum_{x_i \in \mathbb{S}_i} \Phi_{X_i, X_j}(x_i, x_j) \Phi_{X_i}(x_i) \prod_{X_k \in \partial(X_{i \setminus j})} \mu_{ki}^{n}(x_i), \quad (2)$$

where $n$ denotes the iteration, and the shorthand notation $\partial(X_{i \setminus j}) = \{\partial(X_i) \setminus X_j\}$ is used. To compute the message, BP collects all messages sent to $X_i$, except for $X_j$ and multiplies this product with the local potential and the pairwise potential. The sum over all states of $X_i$ is sent to $X_j$. The messages are often normalized by $\alpha_{ij}^{n}$ so that $\sum_{x_j \in \mathbb{S}_i} \mu_{ij}^{n+1}(x_j) = 1$. If BP is converged the approximate marginals are obtained by

$$\hat{P}(X_i = x_i) = \Phi_{X_i}(x_i) \prod_{X_k \in \partial(X_i)} \mu_{ki}(X_i = x_i). \quad (3)$$

BP is guaranteed to converge and the messages correspond to the correct marginals on tree-structured graphs

## 2.2 ISING MODELS

We focus on one specific model: the Ising model, in which every variable $X_i$ has an associated spin taking values from $\mathbb{S} = \{+1, -1\}$. Let us define couplings $J_{ij} \in \mathbb{R}$ assigned to each edge $e_{ij} \in \mathbf{E}$ and a local field $\theta_i \in \mathbb{R}$ acting on each variable $X_i \in \mathbf{X}$. Let the Ising potentials be defined as $\Phi_{X_i}(x_i) = \exp(\theta_i x_i)$ and as $\Phi_{X_i, X_j}(x_i, x_j) = \exp(J_{ij} x_i x_j)$. The Boltzmann distribution

$$P(\mathbf{X} = \mathbf{x}) = \frac{1}{Z} \exp\left(\beta \sum J_{ij} x_i x_j + \beta \sum \theta_i x_i\right) \quad (4)$$

gives the probability of a certain configuration $\mathbf{x}$. Instead of changing the inverse temperature $\beta > 0$ we assume that $\beta = 1$ for the remainder of this work and alter the configuration by changing $J_{ij}$ and $\theta_i$. Whenever $J_{ij}$ and $\theta_i$ are constant over all edges and variables we drop the subscripts, and – because of its physical interpretation – refer to $\theta$ as external field.

Let us distinguish three different interactions on Ising models. First, in the ferromagnetic case all interactions

[1] If BP is applied on graphs with loops it is sometimes named loopy belief propagation; independent of the graph we just stick to the term belief propagation.

$J_{ij}$ are positive; secondly, for antiferromagnetic interactions all $J_{ij}$ are negative. Finally, spin-glass interactions contain both: positive and negative $J_{ij}$. An Ising model is frustrated whenever local constraints exist that cannot be satisfied simultaneously, i.e., a cycle exists, along which the product of all $J_{ij}$ is negative (Mezard and Montanari, 2009).

BP has a close connection to some concepts from statistical mechanics; in particular the correspondence between fixed points of BP and stationary points of the Bethe free energy motivated a large body of research. An excellent treatment of these connection can be found in (Yedidia et al., 2005; Mezard and Montanari, 2009).

Note that in statistical physics the behavior of graphs is analyzed in the thermodynamical limit (i.e. for $N \to \infty$ many variables) and over ensembles of graphs. In this work, however, we consider finite-size graphs with distinct potentials. Note that we use the term phase transition for the finite-size manifestation of the phase transition in the thermodynamical limit. These two notions are similar but do not coincidence in general (cf. Theorem 3).

### 2.3 BP FOR BINARY VARIABLES

For binary variables we parametrize the normalized messages as a single message (cf. Mooij and Kappen (2005)) defined by

$$\nu_{ij}^n = \operatorname{atanh}(\mu_{ij}^n(X_j = 1) - \mu_{ij}^n(X_j = -1)). \quad (5)$$

This transformation reduces the number of variables and eases some calculations but does not change the properties of BP. The update rule in (2) can be rewritten for Ising models to:

$$\tanh(\nu_{ij}^{n+1}) = \tanh(J_{ij}) \tanh(h_{ij}), \quad (6)$$

where we define the effective field $h_{ij}$ that acts on $X_i$ while neglecting the incoming message from $X_j$

$$h_{ij} = \theta_i + \sum_{X_k \in \partial(X_{i \setminus j})} \nu_{ki}. \quad (7)$$

BP corresponds to the cavity method (Mezard et al., 1987) that computes the effect of removing one node and the according edges. This connection becomes obvious by looking at (7) (cf. (Opper and Winther, 2001)).

We can further parametrize the marginals $P(X_i = x_i)$ in (3) by the expected value: $m_i = P(X_i = 1) - P(X_i = -1)$. The mean magnetization of the system $\langle m \rangle = \frac{1}{N} \sum_{i=1}^N m_i$ is often considered to describe the response of the system to an external field (Mezard and Montanari, 2009). Note that the difference between the mean magnetization of an approximate solution $\langle \hat{m} \rangle$ and the exact solution $\langle m \rangle$ is the same as the error averaged over all marginals, i.e., $|\langle m \rangle - \langle \hat{m} \rangle| = \frac{2}{N} |\sum_{i=1}^N P(x_i) - \hat{P}(x_i)|$.

### 2.4 FIXED POINTS OF BP

At least one fixed point always exists for BP (Yedidia et al., 2000); if the graph has loops, however, multiple fixed points may exist. It further depends on the potentials whether BP converges or not. We obtain all fixed points with the NPHC method and analyze convergence properties.

Let $\underline{\nu}$ be the set of all messages; then BP acts as a map on the messages in discrete time such that $\underline{\nu}^{n+1} = \mathcal{BP}(\underline{\nu}^n)$. If successive messages remain unchanged under this map, then BP did converge to some fixed point

$$\underline{\nu}^* = \mathcal{BP}(\underline{\nu}^*). \quad (8)$$

If, however, the messages oscillate, one can either try to achieve convergence by changing the update rule (Elidan et al., 2006; Sutton and McCallum, 2007; Knoll et al., 2015), or by replacing the messages with a weighted average of the last messages (Murphy et al., 1999). We denote the update map of BP with damping as

$$\mathcal{BP}_D(\underline{\nu}^n) = (1 - \epsilon)\mathcal{BP}(\underline{\nu}^n) + \epsilon \underline{\nu}^n, \quad (9)$$

where $\epsilon \in [0, 1)$. Note that fixed points of $\mathcal{BP}_D(\underline{\nu}^*)$ – if BP with damping did converge – must be fixed points of BP without damping as well, because (8) holds in the equilibrium and $\mathcal{BP}_D(\underline{\nu}^*) = (1 - \epsilon)\mathcal{BP}(\underline{\nu}^*) + \epsilon \underline{\nu}^* = \mathcal{BP}(\underline{\nu}^*)$. Although the fixed points remain unchanged under any form of damping, the local stability of these fixed points may change (cf. Section 4).

## 3 STABILITY OF FIXED POINTS

We start by considering BP as a discrete-time map and proceed according to the usual procedure in dynamical systems. For an in-depth treatment of dynamical systems we refer the reader to (Teschl, 2003; Scheinerman, 2000).

A fixed point $\underline{\nu}^*$ is locally stable if a neighborhood $U(\underline{\nu}^*)$ exists such that messages inside this neighborhood converge to $\underline{\nu}^*$ under the considered map (Teschl, 2003, pp.170). Local stability of BP cannot be analyzed by looking at the Bethe free energy as: stable fixed points must be local minima, but local minima must not be stable (Heskes et al., 2003).

In general stability can be investigated by the method of Ljapunov (Scheinerman, 2000, pp.93), though it is sufficient for all graphs considered in this work to restrict

our analysis to linearization; we therefore obtain *all* fixed points by solving $\mathcal{BP}\left(\underline{\nu}^*\right) = \underline{\nu}^*$ and subsequently analyze the Jacobian matrix.

### 3.1 FINDING FIXED POINTS

As a first step, irrespective of the stability, we have to obtain all fixed points of the map. The trivial solution $\nu_{ij}^* = 0$ for all $e_{ij} \in \mathbf{E}$ is known in the special case of vanishing external field (Mooij and Kappen, 2005). The extension to non-vanishing local fields is not straightforward, because in general the unstable fixed point is not known. There are indeed several methods that aim to capture and combine multiple fixed points of BP (e.g., (Srinivasa et al., 2016)), but none of which does also account for unstable fixed points.

In order to find all fixed point solutions we reformulate the update rule of BP (2) as a system of polynomial equations. To solve this nonlinear system of equations we resort to the NPHC method that overcomes all problems of both iterative (e.g., Newton's method) and symbolic methods (e.g., Gröbner basis). The NPHC method solves a similar, but trivial system of equations and further refines these solutions until all (complex) solutions to the system of interest are obtained. A detailed treatment is beyond the scope of this paper but we refer the interested reader to (Sommese and Wampler, 2005; Chen and Li, 2015) and the references therein. Recently the NPHC method was applied to analyze energy landscapes in general (Mehta et al., 2015; Ballard et al., 2017) and to find *all* solutions of BP (Knoll et al., 2016).

### 3.2 LINEARIZATION

After obtaining all fixed points the obvious question remains, whether these fixed points are stable or unstable. Therefore, we approximate $\mathcal{BP}\left(\cdot\right)$ by a linear function in every fixed point and analyze the behavior of the linearized system. This is done by taking the partial derivatives of all messages, i.e., by analyzing the Jacobian matrix $\boldsymbol{F}'(\underline{\nu}^*)$ with the elements defined as

$$\boldsymbol{F}'(\underline{\nu}^*)_{mn} = \frac{\partial \nu_{ij}^*}{\partial \nu_{kl}^*}, \tag{10}$$

where – given some ordering – $\nu_{ij}^*$ and $\nu_{kl}^*$ are the $m^{\text{th}}$ and the $n^{\text{th}}$ message, respectively.

For Ising models parametrized as in Section 2.3 the Jacobian is given as follows: without loss of generality[2] we consider only messages where $e_{ij} \in \mathbf{E}$ and $e_{kl} \in \mathbf{E}$ such that

$$\boldsymbol{F}'(\underline{\nu}^*)_{mn} = \begin{cases} \frac{\tanh(J_{ij})\left(1-\tanh^2(h_{ij})\right)}{1-\tanh^2(J_{ij})\tanh^2(h_{ij})} & \text{if } i = l \text{ and } k \in \partial(X_{i\setminus j}) \\ 0 & \text{else.} \end{cases} \tag{11}$$

Let the spectrum of the Jacobian matrix be the set of all eigenvalues $\Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right) = \{\lambda_1, \cdots, \lambda_n\}$ with spectral radius $\rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right) = \max\{|\lambda_1|, \cdots, |\lambda_n|\}$.

A fixed point $\underline{\nu}^*$ is locally stable if all eigenvalues have absolute value strictly smaller than one

$$\rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right) < 1, \tag{12}$$

i.e., all eigenvalues lie inside the unit circle; BP converges to locally stable fixed points if initialized sufficiently close enough. A fixed point is unstable with respect to BP if at least one eigenvalue exists outside the unit circle such that $\rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right) > 1$. For $\rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right) = 1$ stability of the nonlinear system cannot be inferred by just looking at the linear system (Teschl, 2003, Hartman Grobman Theorem).

If we use damping as in (9) the eigenvalue spectrum of $\mathcal{BP}_D\left(\cdot\right)$ becomes

$$\Gamma\left(\boldsymbol{F}'_D(\underline{\nu}^*)\right) = \Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right) \cdot (1-\epsilon) + \epsilon. \tag{13}$$

It follows, that all eigenvalues are reduced by a factor $(1-\epsilon)$ and experience a shift by $\epsilon$ along the real axis, i.e., $\Re\left(\Gamma\left(\boldsymbol{F}'_D(\underline{\nu}^*)\right)\right) = \epsilon + \Re\left((1-\epsilon)\Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right)\right)$. A fixed point $\underline{\nu}^*$ is locally stable under $\mathcal{BP}_D\left(\cdot\right)$ if

$$\Re\left(\Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right)\right) < 1. \tag{14}$$

## 4 EMPIRICAL STABILITY ANALYSIS OF BP ON ISING MODELS

In this section we present specific instances of the Ising model, compute all fixed points, and analyze the local stability. Then, we compare the results to known results of both: infinite graphs, and finite-size graphs with vanishing local fields. We restrict our analysis to ferromagnetic and antiferromagnetic interactions (cf. Section 2.2), because this allows to change the behavior of BP with just a single parameter.

We briefly present the results of (Mooij and Kappen, 2005) for vanishing fields in Section 4.2. Then we extend the analysis to graphs with non-vanishing fields, discuss some empirical observations, and interpret the implications. We further obtain the exact solution by the junction tree algorithm (Lauritzen and Spiegelhalter, 1988) shown as red line in Figure 2. Fixed points of BP are depicted in blue (stable under $\mathcal{BP}\left(\cdot\right)$), black (stable under $\mathcal{BP}_D\left(\cdot\right)$), and green (unstable). A more formal analysis is presented in Section 5.

[2] If all possible edges are considered, all rows and columns that correspond to a message $\nu_{vw}$, where $e_{vw} \notin \mathbf{E}$ include only zero-values and do not change the eigenvalues.

### 4.1 GRAPHS

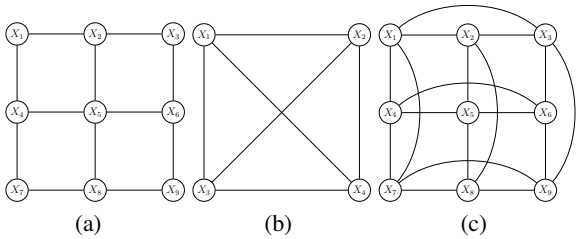



Figure 1: Considered Ising graphs: (a) grid-graph with 9 RVs; (b) complete graph; (c) grid-graph with periodic boundary conditions.

We consider different realizations of the Ising model on complete graphs, grid-graphs, and grid-graphs with periodic boundary conditions (Figure 1).

For the complete graph each pair of nodes is connected by an edge; it follows by definition that it is a regular graph, i.e., all variables have equal degree $d_i = N-1$ for all $X_i$. Note that we can associate an infinite, connected cycle-free, graph – a Bethe lattice – to the complete graph, which has the same fixed point solutions (Taga and Mase, 2004). The grid-graph has all edges aligned along the two-dimensional square lattice. Finite-size realizations contain variables with varying degree. Further we consider grid-graphs with periodic boundary conditions, where nodes on the boundary are joined by edges so that all variables have equal degree.

### 4.2 VANISHING LOCAL FIELD

It is assumed that the case of vanishing local fields is the worst case scenario; we will confirm this assumption empirically and analytically. For $\theta = 0$ a trivial solution exists – namely a distribution that is uniform over all states, i.e., for all $X_i \in \mathbf{X}$ the magnetization is $m_i = 0$.

For purely ferromagnetic interactions $J_{ij} > 0$, BP converges to the trivial (paramagnetic) fixed point, which is unique and stable. As coupling-strength increases, the largest eigenvalue $\lambda_{max}$ increases as well (see Theorem 3) and as $\lambda_{max}$ crosses the unit circle the paramagnetic fixed point remains unchanged but becomes unstable. At the same time two additional fixed points, however, do appear – these symmetry-broken fixed points are both stable (Figure 2a and 2b).

For purely antiferromagnetic interactions $J_{ij} < 0$ all entries of the Jacobian swap in sign $\boldsymbol{F}'_-(\underline{\nu}^*) = -\boldsymbol{F}'_+(\underline{\nu}^*)$, where $-$ and $+$ indicate the antiferromagnetic and the ferromagnetic case, respectively. It follows that $\Gamma\left(\boldsymbol{F}'_-(\underline{\nu}^*)\right) = -\Gamma\left(\boldsymbol{F}'_+(\underline{\nu}^*)\right)$. Consequently the local stability of the fixed point is invariant under a sign-change of $J$ and instability occurs precisely for the same coupling strength as before (Figure 2a and 2b). There is one important difference though – the dominant eigenvalue is negative now, such that damping helps to achieve convergence.

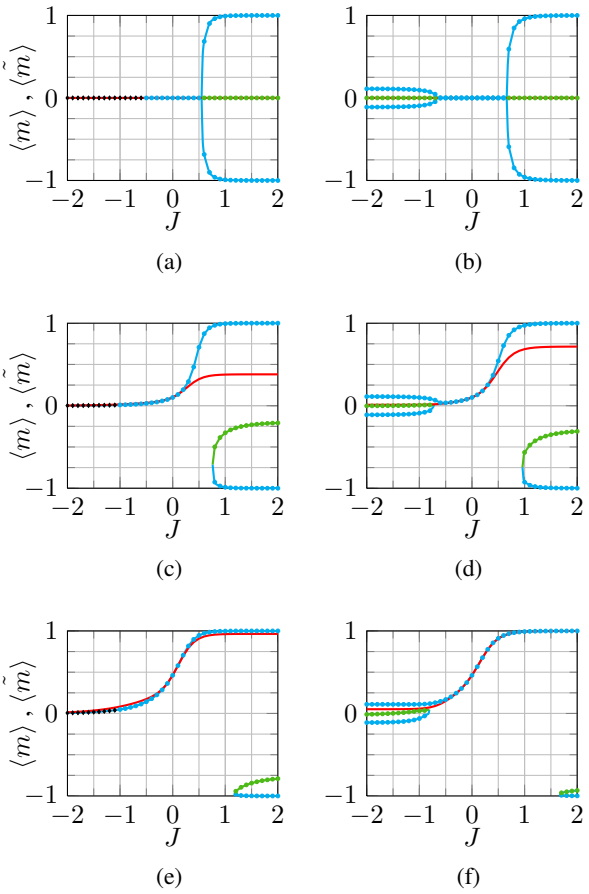
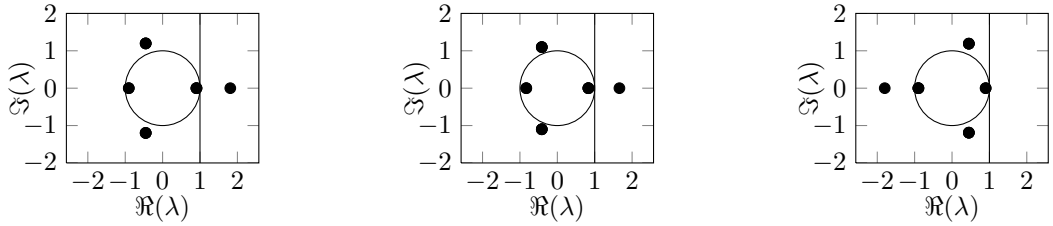



Figure 2: Mean magnetization for couplings $J \in [-2, 2]$ (exact in red, and approximate in blue ($\rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right) < 1$), black ($\rho\left(\boldsymbol{F}'_D(\underline{\nu}^*)\right) < 1$), and green ($\rho\left(\boldsymbol{F}'_D(\underline{\nu}^*)\right) > 1$)); left column shows results for the complete graph with $N = 4$; right column shows results for the grid-graph with $N = 9$. (a) and (b): $\theta = 0$; (c) and (d): $\theta = 0.1$; (e) and (f): $\theta = 0.5$. Increasing $\theta$ enhances convergence properties and accuracy of BP.

### 4.3 NON-VANISHING LOCAL FIELD

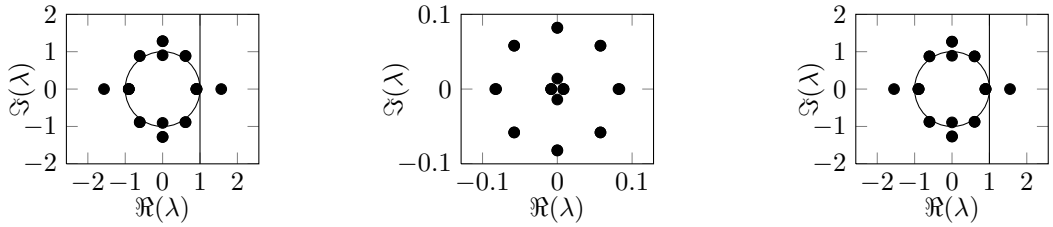
For reasons of simplicity we restrict our analysis to graphs where $\theta_i = \theta \neq 0$ for all variables $X_i$. Because the Ising model is symmetric with respect to the local fields it is sufficient to consider only non-negative local fields. Moreover, the same qualitative results hold if we allow for $\theta_i \geq 0$ in general.




(a) (b) (c)

Figure 3: Eigenvalue spectra of the unstable fixed point of the complete graph with $N = 4$. If all eigenvalues are inside the unit-circle BP converges without damping. If no eigenvalues lie on the right-hand side of the vertical line BP converges with damping. (a) $J = 1.5$ and $\theta = 0$: notice that $\Re(\lambda_{max}) > 1$, i.e., damping does not help; (b) $J = 1.5$ and $\theta = 0.5$: notice how the external field reduces the magnitude of the eigenvalues (cf. Theorem 2); (c) $J = -1.5$ and $\theta = 0.5$: the fixed point is unstable but can be stabilized with an damping because $\Re(\lambda_i) < 1$.

.




(a) (b) (c)

Figure 4: Eigenvalue spectra of the grid-graph with $N = 9$; all eigenvalues are symmetric because the graph is bipartite (cf. Theorem 4). (a) $J = 1.5$, and $\theta = 0$: the shown fixed point is unstable and damping does not help; (b) $J = 1.5$ and $\theta = 0.5$: only a unique stable fixed point exists (cf. Figure 2f); note the qualitative difference in the spectrum; (c) $J = -1.5$ and $\theta = 0.5$: because of the symmetric spectrum, damping does not help in the antiferromagnetic case.

#### 4.3.1 FERROMAGNETIC INTERACTIONS

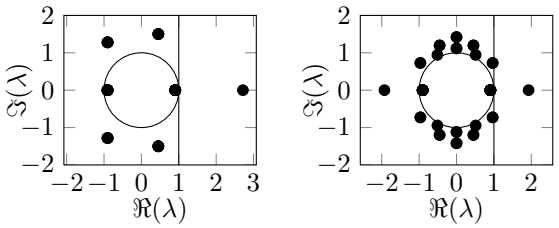
For small values of $J > 0$ a unique fixed point exists to which BP converges and for which all eigenvalues lie inside the unit circle. If we gradually increase the coupling-strength to the point where instability occurs for vanishing external field, we observe that the spectral radius $\rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right) < 1$ and the fixed point is stable (cf. Theorem 2 for more details). If we further increase $\theta$, the spectral radius decreases and a unique stable fixed point exists for even larger values of $J$ (compare Figure 2c and 2d with Figure 2e and 2f).

Now, if we increase $J$ – although the fixed point remains stable – two additional solutions emerge beyond some critical point $J_{C+}$ (Figure 2c and 2d). As opposed to vanishing local fields, the unique fixed point is continuously deformed and remains stable though. The second stable fixed point corresponds to some self-preserving state of magnetization[3] and is accompanied by another unstable fixed point (Figure 3b).

The graph structure is supposed to influence the spectral radius as well. This becomes obvious if we enlarge the size of a graph while keeping its local structure unchanged, e.g., by increasing the grid-graph from $N_1 = 9$ to $N_2 = 16$. A comparison of Figure 4 with Figure 5b reveals an increased spectral radius, i.e., $\rho\left(\boldsymbol{F}'_{N_1}(\underline{\nu}^*)\right) < \rho\left(\boldsymbol{F}'_{N_2}(\underline{\nu}^*)\right)$ (cf. Theorem 3). It is interesting that the largest eigenvalue – and its symmetric counterpart for bipartite graphs – are the only eigenvalues that experience a relevant increase in their real part. The real part of all other eigenvalues $\tilde{\lambda}_i \in \Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right) \setminus \{\lambda_\rho : |\Re(\lambda_\rho)| = \rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right)\}$ is bounded by $|\Re(\tilde{\lambda}_i)| < 1$.

In terms of accuracy BP performs better for non-vanishing fields; compared to vanishing fields a stable fixed point exists that lies closer to the exact solution. Loosely speaking increased local fields effectively reduce the influence of the couplings. This does not only lead to better convergence properties of BP (cf. Corollary 3.1), but reduces the approximation-error as well.

[3] Self-preserving states are stable fixed points where $\langle\tilde{m}\rangle$ points in the opposite direction as the external field.




(a) (b)

Figure 5: Eigenvalue spectra for $J = 1.5$ and $\theta = 0.5$ of the (a) grid-graph with $N = 9$ and periodic boundary conditions; (b) grid-graph with $N = 16$.

#### 4.3.2 ANTIFERROMAGNETIC INTERACTIONS

For small values of $J < 0$ a unique fixed point exists for which all eigenvalues lie inside the unit circle. If we further decrease $J$, a change of behavior can be observed beyond some critical value $J_{C-}$. Interestingly, $|J_{C-}| \leq |J_{C+}|$ with equality if and only if $\theta = 0$. In Figure 3b and 3c we can also see that $\rho\left(\boldsymbol{F}'_{-}(\underline{\nu}^*)\right) > \rho\left(\boldsymbol{F}'_{+}(\underline{\nu}^*)\right)$, i.e., the invariance of local stability under sign-change of $J$ does not hold in general.

Let $J < J_{C-}$, then it depends on the graph structure whether a unique fixed point exists and if damping is useful. It turns out that these two properties are closely connected: First, consider the complete graph with $N = 4$ or the grid-graph with periodic boundary conditions; these graphs have unique fixed points but frustrations exist because of cycles with odd-length. Consequently $\Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right)$ is non-symmetric (cf. Theorem 4). Besides the sign-change, the spectral radius increases as well so that $\rho\left(\boldsymbol{F}'_{-}(\underline{\nu}^*)\right) > \rho\left(\boldsymbol{F}'_{+}(\underline{\nu}^*)\right)$; the dominant eigenvalue, however, has negative sign and $\Re\left(\Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right)\right) < 1$ so that an appropriate damping term exists that enforces BP to converge.

Second, consider both grid-graphs with $N_1 = 9$ and $N_2 = 16$; these graphs are bipartite and have a symmetric spectrum $\Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right)$ (see Figure 4 and 5b, and Theorem 4). Because of the symmetric spectrum BP behaves similar as in the ferromagnetic case: i.e., no damping term exists that would stabilize the unstable fixed point; however, two additional stable fixed points exist.

Multiple fixed points for graphs with purely antiferromagnetic interactions only exist if the underlying graph is bipartite. Such graphs can be decomposed into two disjoint subsets $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$ such that the magnetization follows a "checkerboard-distribution" where for all $X_i \in \mathbf{Y}$ the magnetization is $m_i > 0$ and for all $X_j \in \mathbf{Z}$ the magnetization is $m_j < 0$, or exactly the other way round. This explains the existence of two stable fixed points. Note that the according mean magnetization $\langle \hat{m} \rangle$ behaves as follows: (i) if $N$ is even both stable fixed points are symmetric and have the same mean magnetization; (ii) if $N$ is odd both stable fixed points have different mean magnetization, because the subsets differ in size, i.e., $|\mathbf{Y}| \neq |\mathbf{Z}|$. The difference in $\langle \hat{m} \rangle$ reduces as the number of variables increases. In the limit of $N = \infty$, the mean magnetization of all solutions collapses onto the same value.

To conclude our observations: either a unique fixed point exists, which may be unstable but can be stabilized by damping (black); or a fixed point exists which is unstable under any form of damping (green) but is accompanied by two stable fixed points (blue) (cf. Figure 2). Therefore, non-vanishing fields increase the accuracy of BP (Figure 2) and lead to better convergence properties (Theorem 2) in all experiments.

## 5 THEORETICAL ANALYSIS

Here we present some more formal arguments and explain the observations made in Section 4. We start with the Perron-Frobenius Theorem and specify some implications for the particular form of $\boldsymbol{F}'(\underline{\nu}^*)$. Then we provide properties of the eigenvalue spectrum $\Gamma\left(\boldsymbol{F}'(\underline{\nu}^*)\right)$ and explain the influence of finite-size graphs in non-vanishing local fields. We consider only connected graphs[4] where all $X_i$ have minimum degree $\min d_i \geq 2$. Note that any $X_i$ with $d_i = 1$ can be absorbed in larger potentials.

**Lemma 1.** *The Jacobian matrix $\boldsymbol{F}'(\underline{\nu}^*)$ of a connected graph $G$ is irreducible if* $\min d_i \geq 2$.

*Proof.* The adjacency matrix of a connected undirected graph is irreducible. Let us construct the Jacobian $\boldsymbol{F}'(\underline{\nu}^*)$ as in (10); note that we can partition $\boldsymbol{F}'(\underline{\nu}^*)$ into $N \times N$ block matrices $[\boldsymbol{F}'(\underline{\nu}^*)]_{ik}$ of size $d_i \times d_k$ each. These blocks contain non-zero values if and only if $e_{ij}, e_{kl} \in \mathbf{E}$ for $X_j, X_k \in \mathbf{X}$; i.e., if $a_{ik} = 1$. It follows that $\boldsymbol{F}'(\underline{\nu}^*)$ is irreducible as well. □

**Theorem 1** (Perron Frobenius Theorem)**.** *A real non-negative square matrix $\boldsymbol{B}$ has its spectral radius bounded as follows:*

$$\min_i \sum_j b_{ij} \leq \rho\left(\boldsymbol{B}\right) \leq \max_i \sum_j b_{ij}. \qquad (15)$$

*If $\boldsymbol{B}$ is irreducible the largest eigenvalue is a positive real number $\lambda_{max} = \rho\left(\boldsymbol{B}\right)$.*

[4] A graph G is *connected* if there is a path for all $X_i, X_j \in \mathbf{X}$, i.e., $X_j$ can be reached from $X_i$ along a series of edges.

**Corollary 1.1** (Implications for regular graphs)**.** *A regular graph has $d_i = d$ for all $X_i \in \mathbf{X}$. If the couplings and the fields are constant $\sum_n \boldsymbol{F}'(\underline{\nu}^*)_{mn} = c$ is constant for all rows. By the Perron Frobenius Theorem and by Lemma 1 it follows that $\lambda_{max} = c$.*

Consider an infinite-size grid-graph and some constant coupling-strength $J > 0$. Under these assumptions it is fairly straightforward to provide theoretical insights that explain why the existence of non-vanishing fields enhances the convergence properties. We generalize the results subsequently, as the same argument can be applied to finite-size graphs with different $J_{ij}$.

**Theorem 2.** *Let $G_\infty$ be an infinite-size graph with $d_i = d$ and purely ferromagnetic interactions $J > 0$. Then, the existence of a non-vanishing external field $\theta \neq 0$ stabilizes BP.*

*Proof.* First assume a vanishing external field, i.e., $\theta = \theta_0 = 0$, in which case the trivial fixed point is $\nu^*_{ij} = 0$ and (11) reduces to

$$\boldsymbol{F}'_{\theta_0}(\underline{\nu}^*)_{mn} = \begin{cases} \tanh(J) & \text{if } i = l \text{ and } k \in \partial(X_{i\setminus j}) \\ 0 & \text{else.} \end{cases} \tag{16}$$

Because all variables $X_i \in \mathbf{X}(G_\infty)$ have equal degree $\sum_n \boldsymbol{F}'(\underline{\nu}^*)_{mn} = c$ for all $m$. It follows from (11) and the Perron-Frobenius Theorem that $\rho\left(\boldsymbol{F}'_{\theta_0}(\underline{\nu}^*)\right) = \tanh(J) \cdot (d_i - 1)$. Without loss of generality we exploit symmetry properties and assume an external field $\theta_\delta > 0$. By (2) and (6) it becomes obvious that any fixed point message $\nu^*_{ij} \neq 0$ and $h_{ij} \neq 0$. Note that, qualitatively it does not matter whether $h_{ij}$ is positive or negative as we only consider $\tanh^2(h_{ij})$.

Because $0 \leq \tanh(J) < 1$ it follows that $\tanh(J) > \tanh^2(J)$ and consequently, as

$$\boldsymbol{F}'_{\theta_\delta}(\underline{\nu}^*)_{mn} = \begin{cases} \frac{\tanh(J)\left(1-\tanh^2(h_{ij})\right)}{1-\tanh^2(J)\tanh^2(h_{ij})} & \text{if } i = l \text{ and } k \in \partial(X_{i\setminus j}) \\ 0 & \text{else,} \end{cases} \tag{17}$$

all non-zero entries of $\boldsymbol{F}'_{\theta_0}(\underline{\nu}^*)_{mn}$ are element-wise larger than $\boldsymbol{F}'_{\theta_\delta}(\underline{\nu}^*)_{mn}$. Let us define a field-dependent scaling term $\kappa_\theta \in (0, 1)$, then $\boldsymbol{F}'_{\theta_\delta}(\underline{\nu}^*) = \kappa_\theta \boldsymbol{F}'_{\theta_0}(\underline{\nu}^*)$.

Loosely speaking the existence of some non-vanishing field reduces all entries of the Jacobian matrix and consequently reduces the spectral radius. □

Now choose a critical value $J_C$ at the onset of instability[5] so that $\rho\left(\boldsymbol{F}'_{\theta_0}(\underline{\nu}^*)\right) = \lim_{\epsilon \to 0}(1 + \epsilon)$. If the external field dampens the spectral radius, so that $\rho\left(\boldsymbol{F}'(\underline{\nu}^*)\right) < 1$ the unstable fixed point vanishes, and only a unique stable fixed point remains. Interestingly, this observation holds in all experiments, i.e., in the ferromagnetic case an unstable fixed point exists if and only if multiple fixed points are present.

Suppose we either have purely ferromagnetic or purely antiferromagnetic interactions, then all entries of the Jacobian matrix have the same sign as the couplings and are bounded.

**Lemma 2.** *All entries of the Jacobian are bounded by $\boldsymbol{F}'(\underline{\nu}^*)_{mn} \in [0, 1)$ if $J_{ij} > 0$ and by $\boldsymbol{F}'(\underline{\nu}^*)_{mn} \in (-1, 0]$ if $J_{ij} < 0$.*

*Proof.* If all couplings have the same sign, the qualitative result of (17) holds as well. For purely antiferromagnetic interactions we just have to swap signs. □

**Theorem 3.** *The finite-size manifestation of phase transitions occur beyond the theoretical phase transitions. Let us define a parameter-set $(J_C, \theta_C)$ for which $\lambda_{max}$ crosses the unit-circle. Then, for a graph $G_N$ with $N$ nodes, the values of the critical parameters $(J_C, \theta_C)$ decrease as $N$ increases.*

*Consider two finite-size graphs with identical structure*[6] *with different size $N_1 < N_2$, then $\rho\left(\boldsymbol{F}'_{N_1}(\underline{\nu}^*)\right) \leq \rho\left(\boldsymbol{F}'_{N_2}(\underline{\nu}^*)\right) \leq \rho_\infty\left(\boldsymbol{F}'(\underline{\nu}^*)\right)$.*

*Proof.* Assume $J > 0$ and w.l.g. $\theta = 0$ (for the influence of non-vanishing external field see Theorem 2). Then on $G_\infty$ – or for any other grid-graph with periodic boundary conditions – the degree $d_i = d$ is constant for all $X_i \in \mathbf{X}$. It follows from Corollary 1.1 that the largest eigenvalue is given by $\lambda_{max} = (d_i - 1) \cdot \tanh(J)$[7]. Suppose we increase the couplings to $J^{new} > J$ and denote the change of its parameters as $\kappa = \frac{\tanh(J^{new})}{\tanh(J)}$. It is obvious that $\lambda^{new}_{max} = \kappa \cdot \lambda_{max}$.

Suppose we have some finite-size graph where $d_i$ is not constant but depends on $X_i$. If the couplings increase as before each row of $\boldsymbol{F}'(\underline{\nu}^*)$ experiences a different amount of scaling. In all generality this is described by

$$\begin{bmatrix} c_1 & & [0] \\ & \ddots & \\ [0] & & c_K \end{bmatrix} \cdot \boldsymbol{F}'(\underline{\nu}^*) \tag{18}$$

[5] If $\lambda_{max} = 1$ we cannot infer from the linearized to the nonlinear map (Teschl, 2003). We introduce an $\epsilon$-small term to avoid this subtlety.

[6] E.g., a grid-graph with $N = 9$ or $N = 16$ variables.

[7] For vanishing fields $J_{ij}$ determines the Jacobian matrix.

where $c_k$ depends on $d_i$ and $K = \sum_{i=1}^{N} d_i$. Let us reformulate (18) in all detail to

$$\boldsymbol{W} = \sum_{m=2}^{\max(d_i)} \boldsymbol{C}_m \boldsymbol{F}'(\underline{\nu}^*), \tag{19}$$

where $\boldsymbol{C}_m$ is a diagonal matrix with values $\boldsymbol{C}_{m;kk} = \kappa_m$ if the $k^{\text{th}}$ line of $\boldsymbol{F}'(\underline{\nu}^*)$ corresponds to a variable with $d_k \geq m$ and 0 otherwise. Since the largest eigenvalue is a positive real number it follows that the largest eigenvalue is the sum of the individual eigenvalues, so that

$$\rho\left(\boldsymbol{W}\right) = \sum_{m=2}^{\max(d_i)} \rho\left(\boldsymbol{C}_m \boldsymbol{F}'(\underline{\nu}^*)\right). \tag{20}$$

We still have to show that

$$\rho\left(\boldsymbol{F}'_{N_1}(\underline{\nu}^*)\right) \overset{(a)}{\leq} \rho\left(\boldsymbol{F}'_{N_2}(\underline{\nu}^*)\right) \overset{(b)}{\leq} \rho\left(\boldsymbol{F}'_{\infty}(\underline{\nu}^*)\right), \tag{21}$$

where $(b)$ follows from the existence of an associated eigenvector $x$ such that $\rho\left(\boldsymbol{W}\right) x \leq (d_i - 1) \cdot \tanh(J)$ with equality if and only if all $\boldsymbol{C}_m$ have only non-zero values on the main diagonal, i.e., all variables have equal degree. As $G_{N_2}$ has a larger proportion of nodes with $\max d_i$ than $G_{N_1}$, the average degree is also higher: $\langle d \rangle_{G_{N_1}} < \langle d \rangle_{G_{N_2}}$. Then $(a)$ follows from (20).

□

By combination of Theorem 2 and Theorem 3 we get:

**Corollary 3.1.** *The existence of a non-vanishing external field stabilzes BP on finite-size graphs.*

Next we extend the above observations to varying couplings and fields. We assume that, all couplings $J_{ij}$ have the same sign, and all fields $\theta_i$ have the same sign, not necessarily the same as the couplings. Then the scaling coefficients $c_k$ in (18) depend not only on $d_i$, but on $J_{ij}$ and $\theta_i$ as well. Still, $\boldsymbol{F}'(\underline{\nu}^*)$ can only contain either positive entries or negative entries; it follows that:

**Corollary 3.2.** *For infinite-size grid-graphs with either purely ferromagnetic interactions $J_{ij} > 0$ or purely antiferromagnetic interactions $J_{ij} < 0$, the existence of some non-vanishing fields $\theta_i \neq 0$ stabilizes BP.*

**Theorem 4.** *The eigenvalue-spectrum of the Jacobian $\boldsymbol{F}'(\underline{\nu}^*)$ is symmetric if and only if the underlying graph is bipartite.*

*Proof.* The adjacency matrix of any bipartite graph can be rearranged and written in block form

$$\boldsymbol{A} = \begin{bmatrix} [0] & \boldsymbol{M} \\ \boldsymbol{M}^T & [0] \end{bmatrix}, \tag{22}$$

so that the eigenvalue-spectrum is symmetric (Brouwer and Haemers, 2011). By the same arguments as in Lemma 1 it follows that $\boldsymbol{F}'(\underline{\nu}^*)$ has the same structure and a symmetric spectrum as well. □

**Corollary 4.1.** *Assume we only have purely antiferromagnetic interactions. Then a graph is bipartite if and only if no frustrations occur.*

*Proof.* A cycle is frustrated if and only if the product of all $J_{ij}$ along the corresponding edges is negative (Mezard and Montanari, 2009, p.45). Frustrations can therefore only occur in graphs with cycles of odd length, which implies that the graph cannot be bipartite (Korte and Vygen, 2005, Prop. 2.27). □

# 6 CONCLUSION

In this paper we applied the NPHC method to belief propagation (BP) and obtained all fixed points – stable and unstable ones. Subsequently we performed a local stability analysis for various Ising models with non-vanishing external field.

On the basis of our empirical observations and our theoretical analysis we answered several questions: (i) graphs with vanishing fields are indeed a worst-case scenario. Non-vanishing fields reduce the magnitude of all eigenvalues and help to achieve convergence. (ii) Additionally, we showed that existence of non-vanishing fields increases the accuracy of the best stable fixed point. (iii) Damping is assumed to work best in the antiferromagnetic case; we argue that damping can only be used to stabilize a fixed point if it is unique. (iv) We showed the influence of the graph structure regarding stability and uniqueness of BP: if the graph increases in size, the eigenvalues tend to increase, and as a result the performance of BP degrades; furthermore, because of the symmetric spectrum, damping does not help on bipartite graphs. Nonetheless, for all bipartite graphs considered a stable fixed point does always exist and BP consequently converges.

Only two types of unstable fixed points could be observed: either the fixed point is unique (which has to be a minimum of the corresponding Bethe free energy) and stable under optimal damping. Or, a fixed point exists that is unstable under any form of damping; then, at least two additional fixed points are present that are both stable.

## References


Ballard, A. J., Das, R., Martiniani, S., Mehta, D., Sagun, L., Stevenson, J. D., and Wales, D. J. (2017). Perspective: Energy landscapes for machine learning. *arXiv preprint arXiv:1703.07915*.

Braunstein, A., Mézard, M., and Zecchina, R. (2005). Survey propagation: an algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226.

Brouwer, A. E. and Haemers, W. H. (2011). *Spectra of Graphs*. Springer Science & Business Media.

Chen, T. and Li, T.-Y. (2015). Homotopy continuation method for solving systems of nonlinear and polynomial equations. *Commun. Inf. Syst.*, 15(2):119–307.

Elidan, G., McGraw, I., and Koller, D. (2006). Residual belief propagation: Informed scheduling for asynchronous message passing. In *Proceedings of UAI*, pages 165–173.

Heskes, T. (2004). On the uniqueness of loopy belief propagation fixed points. *Neural Computation*, 16(11):2379–2413.

Heskes, T. et al. (2003). Stable fixed points of loopy belief propagation are minima of the Bethe free energy. In *NIPS*, volume 15, pages 359–366.

Ihler, A., Fisher, J., and Willsky, A. (2005). Loopy belief propagation: convergence and effects of message errors. In *Journal of Machine Learning Research*, pages 905–936.

Jordan, M. (2004). Graphical models. *Statistical Science*, pages 140–155.

Knoll, C., Pernkopf, F., Mehta, D., and Chen, T. (2016). Fixed points of belief propagation–an analysis via polynomial homotopy continuation. *arXiv preprint arXiv:1605.06451*.

Knoll, C., Rath, M., Tschiatschek, S., and Pernkopf, F. (2015). Message scheduling methods for belief propagation. In *Machine Learning and Knowledge Discovery in Databases*, pages 295–310. Springer.

Koller, D. and Friedman, N. (2009). *Probabilistic Graphical Models: Principles and Techniques*. MIT press.

Korte, B. and Vygen, J. (2005). *Combinatorial Optimization*, volume 3. Springer.

Lauritzen, S. and Spiegelhalter, D. (1988). Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society*, pages 157–224.

Mehta, D., Daleo, N. S., Dörfler, F., and Hauenstein, J. D. (2015). Algebraic geometrization of the Kuramoto model: equilibria and stability analysis. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 25(5):053103.

Meshi, O., Jaimovich, A., Globerson, A., and Friedman, N. (2009). Convexifying the Bethe free energy. In *Proceedings of UAI*, pages 402–410.

Mezard, M. and Montanari, A. (2009). *Information, Physics, and Computation*. Oxford Univ. Press.

Mezard, M., Parisi, G., and Virasoro, M. (1987). *Spin Glass Theory and Beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Co Inc.

Mooij, J. and Kappen, H. (2005). On the properties of the Bethe approximation and loopy belief propagation on binary networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2005(11):P11012.

Murphy, K., Weiss, Y., and Jordan, M. (1999). Loopy belief propagation for approximate inference: an empirical study. In *Proceedings of UAI*, pages 467–475.

Opper, M. and Winther, O. (2001). Tractable approximations for probabilistic models: the adaptive Thouless-Anderson-Palmer mean field approach. *Physical Review Letters*, 86(17):3695.

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers.

Pernkopf, F., Peharz, R., and Tschiatschek, S. (2014). *Introduction to Probabilistic Graphical Models*. Academic Press Library in Signal Processing.

Scheinerman, E. R. (2000). *Invitation to Dynamical Systems*. Prentice Hall.

Shin, J. (2012). Complexity of Bethe approximation. In *Proceedings of AISTATS*, pages 1037–1045.

Sommese, A. and Wampler, C. (2005). *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*, volume 99. World Scientific.

Srinivasa, C., Ravanbakhsh, S., and Frey, B. (2016). Survey propagation beyond constraint satisfaction problems. In *Proceedings of AISTATS*, pages 286–295.

Sutton, C. and McCallum, A. (2007). Improved dynamic schedules for belief propagation. In *Proceedings of UAI*, pages 376–383.

Taga, N. and Mase, S. (2004). *On the convergence of belief propagation algorithm for stochastic networks with loops*.

Teschl, G. (2003). *Ordinary Differential Equations and Dynamical Systems*.

Weiss, Y. (2000). Correctness of local probability propagation in graphical models with loops. *Neural Comp.*, 12(1).

Weller, A. and Jebara, T. (2014). Approximating the Bethe partition function. pages 858–867.

Welling, M. and Teh, Y. W. (2001). Belief optimization for binary networks: A stable alternative to loopy belief propagation. In *Proceedings of UAI*, pages 554–561. Morgan Kaufmann Publishers Inc.

Yedidia, J. S., Freeman, W. T., and Weiss, Y. (2005). Constructing free-energy approximations and generalized belief propagation algorithms. *IEEE Transactions on Information Theory*, 51(7):2282–2312.

Yedidia, J. S., Freeman, W. T., Weiss, Y., et al. (2000). Generalized belief propagation. In *NIPS*, volume 13, pages 689–695.

Yuille, A. and Rangarajan, A. (2003). The concave-convex procedure. *Neural Comp.*, 15(4):915–936.