DIAGONAL STATIONARY POINTS OF
THE BETHE FUNCTIONAL

GRZEGORZ SIUDEM
Faculty of Physics, Warsaw University of Technology
Faculty of Mathematics and Information Science
Warsaw University of Technology
Koszykowa 75, PL-00-662 Warsaw, Poland

GRZEGORZ ŚWIATEK
Faculty of Mathematics and Information Science
Warsaw University of Technology
Koszykowa 75, PL-00-662 Warsaw, Poland

(Communicated by Rafael de la Llave)

Abstract. We investigate stationary points of the Bethe functional for the
Ising model on a 2-dimensional lattice. Such stationary points are also fixed
points of message passing algorithms. In the absence of an external field,
by symmetry reasons one expects the fixed points to have constant means at
all sites. This is shown not to be the case. There is a critical value of the
coupling parameter which is equal to the phase transition parameter on the
computation tree, see [13], above which fixed points appear with means that
are variable though constant on diagonals of the lattice and hence the term
“diagonal stationary points”. A rigorous analytic proof of their existence is
presented. Furthermore, computer-obtained examples of diagonal stationary
points which are local maxima of the Bethe functional and hence stable equi-
libria for message passing are shown. The smallest such example was found on
the 100 × 100 lattice.

1. Introduction. Inference from graphical models is a general term for the wide
range of problems which arises from the AI, machine vision, statistical physics and
many others fields. This multitude of aspects is a consequence of the fact that many
theoretical and real-world problems can be reduced to the problem of inferring the
probability distribution from the appropriate graphical model (e.g. Bayesian net-
works, random Markov fields, Tanner graphs, etc. see [17]). There is always the
brute force approach - summing over all nodes leads to the desired distribution,
but this is actually only a theoretical solution — the number of summands grows
exponentially with the size of the graphical model. A computable approach to
the problem is embodied by a class of heuristic algorithms, generally known as be-
 lief propagation or message passing, which are widely used in statistical inference,
combinatorial optimization, image processing and other applications [17]. These
algorithms iterate some abstract beliefs, and if they converge this allows a fast

2010 Mathematics Subject Classification. Primary: 37N40, 90C26; Secondary: 37A60.
Key words and phrases. Bethe approximation, Ising model, message passing algorithms.
Both authors supported in part by Narodowe Centrum Nauki - grant 2015/17/B/ST1/00091.
* Corresponding author: g.swiatek@mini.pw.edu.pl.
approximation of the desired probability distribution. The interest in those tech-
niques derives from the work of Pearl \[9\] and their importance was emphasized by
the Turing prize awarded to Pearl in 2011 for fundamental contributions to artifi-
cial intelligence through the development of a calculus for probabilistic and causal
reasoning. As we will see belief propagation algorithms are strictly connected with
Bethe approximation.

Bethe approximation is a technique which originated in statistical physics, see \[2\].
To explain the idea behind this approach let us consider an Ising model with the
energy function \( U(\sigma) \) where \( \sigma = (\sigma_i)_{i \in V} \) is a random configura-
tion of spins and \( U \) is assumed to a sum of terms which depend on spins at single sites or pairs of adjacent
sites. As a general convention, we will denote by \( \sigma = (\sigma_i) \) random variables indexed
by the sites and taking values \( \pm 1 \) and by \( s = (s_i) \) particular configurations of such
values. Given a probability distribution \( P \) on the set of all configurations we define
the entropy
\[
S^P = - \sum_s P(s) \ln P(s),
\]
and the average energy
\[
\langle U(\sigma) \rangle^P = \sum_s P(s) U(s).
\]
These two terms allows one to define the free energy functional
\[
F^P = \langle U(\sigma) \rangle^P + S^P.
\]
As is well-known the free energy functional has a unique maximum which is given
by the exponential distribution. Even though the exponential distribution is given
explicitly, the formula is not usable in practice since it involves summing up over all
configurations whose number is exponential in terms of the size of the model. As
noted, this is a common problem with inference from graphical models. To address
this, Bethe approximation is defined on a much smaller space of pseudo-marginals
\( (b_i(s_i)), (b_{ij}(s_i, s_j)) \) where \( i, j \) run over all pairs of neighboring sites. When a prob-
ability distribution \( P \) on the set of configurations is given, one can define actual
marginals \( b_i(s_i) := P(\sigma_i = s_i) \) and \( b_{ij}(s_i, s_j) = P(\sigma_i = s_i, \sigma_j = s_j) \), however
pseudo-marginals are only required to satisfy natural marginalization conditions
(see Eq. \[5\]) and may not derive from any probability distribution on the set of
configurations. Given a system of pseudo-marginals \( b \) the entropy part of the Bethe
functional is
\[
S^b = \sum_{i,j} \sum_{s_i, s_j = \pm 1} b_{ij}(s_i, s_j) \ln \frac{b_{ij}(s_i, s_j)}{b_i(s_i)b_j(s_j)}
\]
with the first summation extended over all pairs of closest neighbors \( i, j \). Then the
Bethe functional is
\[
H(b) = \langle U(\sigma) \rangle^b + S^b.
\]
Note that the expected value of \( U(\sigma) \) is defined given a pseudo-marginal due to
the assumption that \( U \) is the a sum of terms which depend on spins at most at
two neighboring sites, while the entropy term has been modified in a way that may
appear arbitrary. It turns out that the Bethe functional equals the free energy
functional for Ising models on trees, see \[14\] (4.13). The equality means that if
a pseudo-marginal is an actual marginal derived from a probability distribution
\( P \) on the set of all configurations, then both functionals are equal. The Bethe
functional is defined on the space of all pseudo-marginals whose dimension grows
polynomially with the size of the model and hence is computable. The idea is that
knowing the global maximizer of the Bethe functional provides an approximation of the marginals for the exponential distribution. The precise formulation of the concepts introduced here is worked out in Section 2.

The basic connection between Bethe approximation and belief propagation algorithms is that fixed points of belief propagation algorithms are stationary points of the Bethe functional, see [3, 17]. This is natural since the classical belief propagation algorithm can be seen as a discrete approximation of the gradient flow of the Bethe functional, see [16]. Local maxima of the Bethe functional can be effectively found by those algorithms, while on the other hand the analysis of belief propagation must largely rely on the understanding of the Bethe functional [11]. Since the global maximum of the Bethe functional is the object of interest, clearly the analysis of the set of local maxima is relevant to evaluating and improving performance of belief propagation methods.

It appears that the theory of the Bethe approximation has lagged behind its algorithmic development. In the literature one can find local results concerning fixed points of belief propagation and their stability [5, 6, 7, 10]. Studies of the global aspect of the problem, such as the structure of the fixed point set have been almost exclusively the domain of experimentation.

From the dynamical point of view, one can consider the gradient flow of the Bethe functional. This makes the system formally very simple, but complication arises because the high dimension of the phase space. From the algorithmic point of view, the desirable situation is convergence to the global maximum and the danger lies in the existence of local maxima.

1.1. The goals of this paper. We deal with an Ising model on a finite two-dimensional square lattice with periodic boundary conditions i.e. on the toral lattice. Formally, the sites are elements of the set \( \{0, \ldots, n - 1\} \times \{0, \ldots, n - 1\} \) wrapped doubly periodically so that neighbors are calculated mod \( n \). We will denote such a graph by \( T_n \). The interaction between neighbors is given by \( J \sigma_i \sigma_j \), for a fixed \( J > 0 \), where \( \sigma_i \) are variables of a random Markov field which take values \( \pm 1 \).

One can also add the sum over all sites of the terms \( h_i \sigma_i \) in which case the vector \( (h_i) \) is the external field.

This paper grew out of a simple question posed during our research on the previous work [12]: if the external field possesses a certain symmetry, does the same symmetry hold for stationary points of the Bethe functional? After some initial investigation the problem was narrowed to the situation in which the external field is simply 0 and the question becomes whether Bethe stationary points necessarily have constant means for all sites of the lattice.

Surprisingly, the answer to even such a simple question appeared unknown and even more surprisingly turned out to be negative. The goal of this paper is to describe a class of non-constant stationary points, prove their existence by rigorous analytic methods and study some of their properties by a combination of analytic and numerical methods.

1.2. Critical coupling and phase transitions. A difficulty in studying the set of stationary points of the Bethe functional lies the fact that except for special cases the functional is not convex (strictly speaking, using the sign convention of this paper: its negative is not convex). One paper which deals with this problem is [18]. For a given graph, its approach is to construct the computation tree, or a universal covering in the language of topology, which is generally infinite. For Markov fields on
infinite graphs one expects a phenomenon known as \textit{phase transition} which occurs when $J$ exceeds a critical value $J_c$. The main result of [13] is that the absence of a phase transition on the computation tree implies the uniqueness of the Bethe stationary point and convergence of the belief propagation algorithm. A similar construction for the Bethe approximation had been known in statistical mechanics, see [1], Chap. 4. Rather than “computational tree”, the term used there is “Cayley tree” and in the case of zero external field the critical value $J_c = \frac{1}{2} \ln 2$ is obtained. This is equivalent to $\tanh J_c = \frac{\exp(J_c) - \exp(-J_c)}{\exp(J_c) + \exp(-J_c)} = \frac{1}{3}$.

In this context, let us state our main result

\textbf{Theorem 1.1.} In the case of zero external field, the following alternative holds. If $\tanh(J) \leq 1/3$, then there is only one stationary point of the Bethe functional with all singleton components $b_i$ equal to $1/2$. On the other hand, for every $J : \tanh(J) > 1/3$, there exists $n(J)$ such that for every $k \geq n(J)$ there is a stationary point of the Bethe functional $H_{T_n}^\mathbf{b} (\{(b_i), (b_{ij})\}, J)$ with values of singleton components $b_i$ both greater and smaller than $1/2$, depending on $i$.

Stationary points of Theorem 1.1 will be referred to as \textit{diagonal points} since as will become clear from the proof the values of $b_i$ are constant on diagonals of the lattice.

We conclude first that the result of [13] is sharp in this case (Ising model with zero external field): for $J \leq J_c$ there is no phase transition on the computation tree and for $J > J_c$ there are examples with many fixed stationary points. But Theorem 1.1 goes further than that in stating not only that multiple stationary points exist, but that they are different from the obvious ones: one with fixed positive means and the other its negative image.

1.3. Stability. Both from the point of view of the dynamical systems theory and applications, the important issue is the stability of diagonal fixed points. For the Bethe gradient flows stability of a fixed point is clearly equivalent to its being a strict local maximum and the same holds for discrete versions of belief propagation, see [5]. Here, we have only been able to obtain a numerically-supported result.

\textbf{Finding 1.} For the value of $J = 0.5$ and $n = 100$ there exists a pseudo-marginal on $T_n$ with singleton components fixed on diagonals, but otherwise taking values both greater and less than $1/2$, which is a strict local maximum of the Bethe functional with zero external field.

Finding 1 is less robust than Theorem 1.1. First, it requires the use of a computer with floating-point arithmetic and we have made no attempt to rigorously verify either the correctness of the program, or the numerical error. Secondly, the example is only for a particular choice of parameters and we do not know whether $J$ can be lowered to a value arbitrarily close to $J_c$.

1.4. Consequences and directions for further research. From the mathematical point of view, our results show a complex nature of the bifurcation in the set of stationary points of the Bethe functional which occurs when $J$ crosses the critical threshold. A key question here is whether stable equilibria can be obtained for $J$ arbitrarily close to $J_c$.

These results and particularly Finding 1 are bad news for algorithmic uses of belief propagation. Even in the arguably simplest setting spurious local maxima appear. Although in case of no external field it is clear what the global maximum
BETHE FUNCTIONAL’S STATIONARY POINTS

is, under the actual circumstances one can easily see how finding it could become a computationally hopeless task. On the other hand, it offers a non-trivial dynamical system to research.

Let’s take a quick glimpse into the possibilities. Owing to the diagonal symmetry, if an example of the type we consider exists on a toral lattice $T_n$, it can also be realized to $T_{kn}$ for any $k$ natural or on the infinite lattice $\mathbb{Z}^2$ by periodic repetition on blocks of $n$ consecutive diagonals. Hence, if indeed stable equilibria for zero external field appear for any $J > J_c$ and $T_n$ sufficiently big, there would be infinitely many on the infinite lattice for any $J > J_c$, all essentially different, i.e. not related by shifts. This would be analogous to the Newhouse phenomenon, see [8], and while an infinite lattice is needed, it would still be striking given the complete homogeneity of the phase space in the absence of an external field.

1.5. Plan of the work. In Section 2 we present a setting for our work and develop methods which will be used in the sequel. That part of the work is done in a general setting of Ising models on arbitrary graphs and allows for an external field. The most important result is Proposition 2 which provides a necessary and sufficient condition for Bethe stationary points without reference to message passing, but relying instead on a harmonicity-type condition expressed in terms of the neighbor average function.

In Section 3 the discussion is specialized to the case of two-dimensional toral lattice and no external field. They lead to the basic formula [8] and the statement of Theorem 3.3 about a connection between the neighbor average function and arithmetic mean.

In Section 4 we build on those results to construct the diagonal stationary points from Theorem 1.1. This is followed by numerical studies which visualize these points and illustrate difficulties inherent in studying their stability. This section ends with the description of numerical evidence in favor of Finding 1.

The last part of the work is an appendix in which we group calculations which are important for the verification of our work, but somewhat complicated and detached from the main line of the argument.

With the exception of Sections 4.2 and 4.3 which are based on numerical work, rigorous analytic proofs are provided. We tried as much as possible to use the terminology and conventions of [14] in order to make the work accessible to application-oriented readers. A dynamist will have no difficulty recognizing that “exponential families” are “Gibbs distributions”. Perhaps more confusingly, in combinatorial optimization one maximizes things instead of minimizing as is the convention in thermodynamic formalism, hence reversed signs in several formulas.

2. Ising models in general setting. We work with an undirected graph $G$ with no self-connections. Its nodes will be denoted with small Latin subscripts such as $i$ and the set of all vertices will be denoted with $V$. The edge set $E$ is identified with a subset of unordered pairs of vertices $\{i, j\}$. For each $i$ we consider its neighborhood $\mathcal{N}(i) = \{j \in V : \{i, j\} \in E\}$. With each site we associate a variable $\sigma_i$ which takes values $\pm 1$. Given a vector $s := (s_i)_{i \in V}$, $s_i = \pm 1$, called a configuration and canonical parameters $f = (f_i)_{i \in V \cup \{0\}} \subseteq \mathbb{R}^{\left| V \right| + 1}$ we define the potential

$$U(f, s) = \sum_{i \in V} f_i s_i + f_0 \sum_{\{i, j\} \in E} s_i s_j,$$

we also denote $s_0 = \sum_{\{i, j\} \in E} s_i s_j$. 


The family of exponential distributions [14] parametrized by the canonical parameters $f$ on the set of all configurations is given by

$$G_f(s) = \frac{\exp[U(f, s)]}{Z(f)},$$

where

$$Z(f) = \sum_s \exp[U(f, s)],$$

is the normalizing factor known as the partition function. Given a probability distribution $P$ on the set of configurations the expected values will be written as $\langle \cdot \rangle^P$. Recall that we denote by $\sigma_i$ random variables and by the $(s_i)$ a certain configuration.

We will also use covariance:

$$\text{Cov}_P(\sigma_i, \sigma_j) = \langle \sigma_i \sigma_j \rangle^P - \langle \sigma_i \rangle^P \langle \sigma_j \rangle^P.$$

The following identities are well known for all $i, j \in V \cup \{0\}$.

$$\langle \sigma_i \rangle^{G_f} = \frac{\partial \log Z(f)}{\partial f_i},$$

and

$$\text{Cov}^{G_f}(\sigma_i, \sigma_j) = \frac{\partial^2 \log Z(f)}{\partial f_i \partial f_j}.$$

These formulas offer a convenient way for expressing the marginals of the exponential distribution, which are a particular case of pseudo-marginals and can be used to express the Bethe functional.

We rely on the following:

**Proposition 1.** Let $V' \subset V$ and take any vector $m' = (m_j)_{j \in V'}, m_j \in (-1, 1)$. Fix canonical parameters $\hat{f} = (f_i)_{i \in (V \setminus V') \cup \{0\}}$. There is a unique choice of canonical parameters $f' = (f_j)$ for $j \in V'$ such that when $f$ is the joint vector $f' \cup \hat{f}$, then the equality $\langle \sigma_j \rangle^{G_f} = m_j$ holds for all $j \in V'$.

Proposition essentially follows from Theorem 3.3 [14], but the difference is that we fixed some of the canonical parameters, which requires an additional argument. The formal proof is given in Appendix A.

**Proposition 1** implicitly defines the function

$$\Psi_{V, V'} : (-1, 1)^{V'} \times \mathbb{R}^{(V \setminus V') \cup \{0\}} \to \mathbb{R}^{V \cup \{0\}},$$

given by

$$\Psi_{V, V'}((m_j)_{j \in V'}, (\hat{f})_{i \in (V \setminus V') \cup \{0\}}) = (f_i)_{i \in V \cup \{0\}},$$

where $f$ is characterized uniquely by the conditions

$$\forall j \in V', \langle \sigma_j \rangle^{G_f} = m_j, \forall i \in V \cup \{0\} f_i = \hat{f}_i.$$
2.1. Pseudo-marginals and the Bethe variational Problem. We now proceed to define formally pseudo-marginals, which have been informally discussed in the introduction. By a pseudo-marginal we mean a system of positive singleton quantities \((b_i(s_i))_{i\in V}\) and their pairwise counterparts \((b_{ij}(s_i, s_j))_{(i,j)\in E}\), symmetric meaning \(b_{ij}(s_i, s_j) = b_{ji}(s_j, s_i)\) and subject to the local consistency conditions for all \(i \in V\) and \(s_i = \pm 1\):

\[
\sum_{s_i} b_i(s_i) = 1,
\]

\[
\forall j \in \mathcal{N}(i), \sum_{s_j} b_{ij}(s_i, s_j) = b_i(s_i).
\]

The space of all pseudo-marginals will be denoted with \(\mathcal{B}\). One can easily see the that \(\mathcal{B}\) can be parametrized by variables \(b_i := b_i(1)\) for \(i \in V\) and \(b_{ij} = b_{ij}(1, 1)\) for \(\{i, j\} \in E\) subject to inequalities \(0 < b_{ij} < \min(b_i, b_j)\). Then \(b_i(s_i) = \frac{1-b_{ij}}{s_i b_i + s_j b_i}\) and \(b_{ij}(s_i, s_j)\) has a similarly simple expression through \(b_{ij}, b_i, b_j\) (for exact formulas cf. \([16]\)).

Given a pseudo-marginal and a set of canonical parameters \((f_i)_{i \in V \cup \{0\}}\) we can define the Bethe functional on \(b \in \mathcal{B}\):

\[
H_G(b, (f_i)) = \sum_{i \in V} \sum_{s_i} f_i s_i b_i(s_i) + f_0 \sum_{\{i,j\} \in E} \sum_{s_i, s_j} s_i s_j b_{ij}(s_i, s_j) + \]

\[
+ \sum_{i \in V} \left[(\mathcal{N}(i) - 1) b_i(s_i) \ln b_i(s_i) \right] - \sum_{\{i,j\} \in E} \sum_{s_i, s_j} b_{ij}(s_i, s_j) \ln b_{ij}(s_i, s_j)
\]

where the summation with respect to \(s_i, s_j\) is always meant to range over the values \(\pm 1\). In the future, we will often use the same letter for a graph and its vertex set, so \(H_G\) could also be written as \(H_V\).

Discussed above the Bethe Variational Problem consists in finding the stationary pseudo-marginals \(\{(b_i), (b_{ij})\}\) under a given ensemble of canonical parameters \((f_i)\).

2.1.1. Pairwise maximality.

**Definition 2.1.** A pseudo-marginal is called pairwise maximal, given a set of canonical parameters \((f_i)\), provided that its pairwise components \((b_{ij})\) maximize the Bethe functional under its fixed singleton components \((b_i)\).

Finding the maximizing value of \(b_{ij}\) given \(b_i, b_j\) is tantamount to maximizing the Bethe functional in the case of a subgraph \(W\) which consists of \(i, j\) and the edge between them. Since \(b_i, b_j\) are fixed, the first and third sums in Eq. \(6\) become irrelevant and hence the problem is solved by taking the marginal of \(G_{W}\) where \(f^W = \Psi_{W,W}(2b_i - 1, 2b_j - 1, (f_0))\). The concept of pairwise maximality was examined in \([16]\), from where an algebraic formula for \(f^W\) can be derived. We will not need it and will simply introduce a function \(B^{\text{max}}\) defined by \(b_{ij} = B^{\text{max}}(b_i, b_j)\) for all edges and a pairwise maximal pseudo-marginal.

2.1.2. The Bethe Variational Problem on a subgraph.

If \(W\) is a subgraph \(V\), we will write \(\partial W\) for the set of the vertices of \(W\) which are connected to a vertex not in \(W\),

**Lemma 2.2.** Suppose that \(W\) is a subgraph of \(V\) with the edge set \(E_W\), \(b = \{(b_i), (b_{ij})\}\) is a pseudo-marginal on \(V\) and \((f_i)\) a vector of canonical parameters. Then:
Definition 2.3. For every neighbor average function and the criterion of stationarity.

2.1.3. Sometimes for all \( i \in W \setminus \partial W \) and \( \{i, j\} \in E_W \)

\[
\frac{\partial H_V(b, (f_i))}{\partial b_i} = \frac{\partial H_W(b, (f_i))}{\partial b_i} \\
\frac{\partial H_V(b, (f_i))}{\partial b_{ij}} = \frac{\partial H_W(b, (f_i))}{\partial b_{ij}}
\]

there exists a unique set of canonical parameters \( (f_i^W) \) on \( W \), equal to \( f \) for \( i \notin \partial W \) such that

\[
\frac{\partial H_V(b, (f_i))}{\partial b_i} = \frac{\partial H_W(b, (f_i^W))}{\partial b_i}
\]

for all \( i \in W \),

- if \( W \) is a tree and \( b \) is a stationary point for \( H_V(b, (f_i)) \), then its restriction to \( W \) is stationary for \( H_W(b, f^W) \), where

\[
f^W = \Psi_{W, \partial W} ((2b_i - 1), \{i \in \partial W \setminus \emptyset \})
\]

Proof. The first claim follows by the inspection of formula (6), since all terms which depend on the variables \( b_i, b_{ij} \) occur in the same form in both \( H_V \) and \( H_W \).

For \( b_i \) such that \( i \in \partial W \) this is no longer true since the orders of vertices which occur in the third term of (6) are different and \( b_i \) also enter into \( b_{ik}(s_i, s_k) \) for \( k \notin W \) and such terms are missing in \( H_W \). Note that the first term in \( H_W \) can be rewritten as

\[
\sum_{i \in W} \sum_{s_i} f_i s_i b_i(s_i) = \sum_{i \in W} f_i \sum_{s_i} \left[ s_i \left( \frac{1 - s_i}{2} + s_i b_i \right) \right] = \sum_{i \in W} f_i (2b_i - 1)
\]

and these are the only terms which depend on \( f_i \). Hence, \( f_i \) for \( i \in \partial W \) can be adjusted in a unique way to bring about the equality of partial derivatives which is referred to in the second claim.

The third claim when \( W = V \) reduces to the well-known fact that for trees the unique stationary point of the Bethe functional is the marginal of the corresponding exponential distribution, see Theorem 4.2(b) in [14]. In the general situation, the third claim follows for this and the second claim for some \( f_i^W \). Since the mean values are \( (2b_i - 1) \) for \( i \in \partial W \) while at the remaining vertices \( f_i^W = f_i \), \( f^W \) must be in the form given by Proposition 1.

2.1.3. Neighbor average function and the criterion of stationarity.

Definition 2.3. For every \( i \in V \) consider the star graph \( C_i \) which consists of \( i \), its neighbors from \( N(i) \) and exactly the edges in the form \( \{i, j\}, j \in N(i) \).

Definition 2.4. Define the neighbor average function \( A : (-1, 1)^{|N(i)|} \times \mathbb{R} \to (-1, 1) \to \mathbb{R} \) by

\[
A_{f_0}((m_j)_{j \in N(i)}, f_i) = \langle \sigma_i \rangle^G_{f_0}
\]

where

\[
\mathcal{F} = \Psi_{N(i), i, N(i)}((m_j)_{j \in N(i)}((f_0, f_i))
\]

Sometimes \( f_0 \) will be skipped when its value is fixed and not important in the context.

The neighbor average function provides a local condition for stationarity among pairwise maximal pseudo-marginals, as shown by the following properties.
Lemma 2.5. For any star subgraph $C_i$, any choice of canonical parameters
\[(f_j)_{j \in \{0,i\} \cup N(i)}\]
as well as a vector \((m_j)_{j \in N(i)}: \forall j m_j \in (-1,1)\), the marginal of the exponential
distribution \(G_{\Psi_{N(i)\cup\{i\}},N(i),\{(m_j),(f_0,f_i)\}}\) is the unique conditional stationary point and
a maximizer of \(H_{C_i}(b,(f_j))\) in the set of pseudo-marginals on star graph \(C_i\) whose
singleton means on \(N(i)\) agree with \((m_j)\).

Proof. Since \(C_i\) is a star graph and thus a tree, by the fact already used in the proof
of the third claim of Lemma 2.2 the marginal of the exponential distribution is the
unique maximizer of \(H_{C_i}(b,f_G)\) where
\[f_G = \Psi_{N(i)\cup\{i\},N(i),((m_j),(f_0,f_i))} .\]
The difference is
\[H_{C_i}(b,(f_j)) - H_{C_i}(b,f_G) = \sum_{j \in N(i)} (f_j - f_j^G) b_j\]
which is independent of \(b\) under the condition \(2b_j - 1 = \beta_j\).

We get a Corollary:

Corollary 1. Among all pairwise maximal pseudo-marginals (cf. Definition 2.1)
on \(V\) with singleton means \((m_j)\) fixed except for \(j = i\), the unique maximizer of the
Bethe functional is given by the condition
\[m_i = A_{f_0}((m_j)_{j \in N(i)}, f_i) .\]

This follows from Lemma 2.5 since \(H_V\) is the sum of \(H_{C_i}\) and terms which are
independent of \(b_i\) and \((b_{ij})_{j \in N(i)}\).

There is a simple characterization of stationary points of the Bethe functional in
terms of the neighbor average function.

Proposition 2. A pseudo-marginal given by \(b = \{(b_i),(b_{ij})\}\) is a stationary point of
the Bethe functional \(H_V(b,(f_i))\) if and only if it is pairwise maximal (cf. Definition
2.1) and for every \(i\) the equality
\[2b_i - 1 = A((2b_j - 1)_{j \in N(i)}, f_i)\]
holds.

So, stationary points of the Bethe functional are effectively characterized only
through their singleton components.

Proof. First assume that \(b\) is stationary. Then the third claim of Lemma 2.2 can
be used to all subgraphs which are trees. For subgraph which consists of a pair
of vertices and an edge between them, this reduces to the condition of pairwise
maximality. For a star graph \(C_i\) the vector of canonical parameters \(f_C\), mentioned
in the third claim of Lemma 2.2 is the same as \(F\) in Definition 2.3. Hence the
marginals are the same which implies the equality of the claim of Proposition 2.

To prove the opposite implication, suppose that \(b\) is not stationary. Then there
is an arc of pseudo-marginals \(b(t)\), with \(b(0) = b\) such that
\[\frac{d}{dt} H_V(b(t),(f_i))|_{t=0} > 0 .\]
Replace \( b(t) \) with an arc of pairwise maximal pseudo-marginals (for proof of the differentiability of \( b^{\text{max}} \) cf. [16]), namely
\[
b^{\text{max}}(t) = \{ (b_i(t)), B^{\text{max}}(b_i(t), b_j(t)) \}.
\]
In other words, we keep the singleton components from the original \( b(t) \) and adjust the pairwise components to gain pairwise maximality. By definition,
\[
H_V(b^{\text{max}}(t), (f_i)) \geq H_V(b(t), (f_i)),
\]
for all \( t \) and hence
\[
\frac{d}{dt} H_V(b^{\text{max}}(t), (f_i))|_{t=0} > 0.
\]
What we have proved is that \( b \) is also non-stationary on the submanifold of pairwise maximal pseudo-marginals. In other words, if we define
\[
H_V^{\text{max}}((b_i), (f_i)) := H_V(\{ (b_i), B^{\text{max}}(b_i, b_i) \}, (f_i)),
\]
then
\[
\exists \ i \in V \frac{\partial}{\partial b_i} H_V^{\text{max}}((b_i), (f_i)) \neq 0
\]
at \( b_i \) given by \( b \).

For that site \( i \) consider an arc of pairwise maximal pseudo-marginals which varies only \( b_i(t) \) and leaves all other \( b_j(t) \neq i \) fixed. Then the only \( b_j(t) \) that vary are those which correspond to the edges of \( V \) which come out of \( i \). From the first claim of Lemma 2.2 we conclude that \( b \) restricted to the star graph \( C_i \) is not a stationary point of \( H_{C_i} \).

On the other hand, we are assuming the equality
\[
2b_i - 1 = A ((2b_j - 1)_{j \in N(i)}, f_i),
\]
which means that it implies the same singleton means as \( G_F \) from Definition 2.3. The pairwise components \( (b_j)_{j \in N(i)} \) are also maximizing since \( b \) was pairwise maximal. Thus, \( b \) restricted to the star graph \( C_i \) is the actual marginal of \( G_F \). Then it is stationary for \( H_{C_i} \) by the Lemma 2.5.

This contradiction ends the proof.

2.2. Monotonicity properties. Let us begin with the following observation of the nature of the function \( \Psi_{N(i) \cup \{ i \}, N(i)} \).

**Lemma 2.6.** Consider the star graph \( C_i \) (cf. Definition 2.3) for any \( i \in V \). Choose \( k \in N(i) \) and for \(-1 < m < 1\) define the vector-valued function
\[
F(m) := \Psi_{N(i) \cup \{ i \}, C_i} ((m_i, m_k = m, m_j)_{j \in N(i) \setminus \{ k \}}, f_0).
\]
Then, for \( m_i, (m_j)_{j \in N(i) \setminus \{ k \}} \) fixed the functions \( F(m)_{j \in N(i) \setminus \{ k \}} \) are all constant.

**Proof.** Pick a \( j \in N(i) \setminus \{ k \} \). Using conditional expectations
\[
m_j = (\sigma_j)^{G_F(m)} = \sum_{s_i} (\sigma_j | \sigma_i = s_i)^{G_F(m)} G_{F(m)}(\sigma_i = s_i) = \sum_{s_i} (\sigma_j | \sigma_i = s_i)^{G_F(m)} \frac{1 + m_j s_i}{2}.
\]

Furthermore, it is known that conditional expectations of exponential distributions on a subgraph \( W \) are again exponential with the conditions playing the role
of boundary conditions on \( \partial W \). In this case, \( W \) is just the edge from \( i \) to \( j \) and the exponential distribution is readily computed to give
\[
(\sigma_j | \sigma_i = s_i)^{G_{\mathcal{F}(m)}} = \tanh(f_0 s_i + \mathcal{F}(m)_j) .
\]
Combining these formulas we see that with \( m_j, m_i, f_0 \) fixed the only variable is \( \mathcal{F}(m)_j \) which must therefore be fixed as well.

Let \( G_f \) be any exponential distribution with the vector \( f \) of canonical parameters. As we have already observed, for any \( i \in V \)
\[
\frac{\partial (\sigma_i)^{G_f}}{\partial f_j} = \text{Cov}^{G_f}(\sigma_i, \sigma_j) .
\]
This is always non-negative when \( f_0 > 0 \) as a consequence of the Fortuin-Kasteleyn-Ginibre (FKG) inequality, see [4]. As a corollary, we get the following Fact which we refer to as the monotonicity of expectations:

**Remark 1.** If \( f_0 \geq 0 \), then for any \( i, j \in V \) \( (\sigma_i)^{G_f} \) is a non-decreasing function of any \( f_j \). It is increasing if \( i = j \).

The increase follows since
\[
\frac{\partial}{\partial f_i} (\sigma_i)^{G_f} = \text{Cov}^{G_f}(\sigma_i, \sigma_i) > 0 .
\]

Fact [1] leads to a proposition concerning the monotonicity of the neighbor average functions.

**Proposition 3.** If \( f_0 \geq 0 \) then the function \( A((m_j)_{j \in N(i)}, f_i) \) is increasing with respect to \( f_i \) and non-decreasing with respect to all other variables.

**Proof.** Let us first prove monotonicity with respect to \( f_i \). Clearly, the limit of \( A((m_j)_{j \in N(i)}, f_i) \) is 1 when \( f_i \) tends to \( \infty \) and \(-1 \) when it tends to \(-\infty \). So assuming the contrary means that for different values \( f_i^a \) and \( f_i^b \) and fixed \((m_j)\)
\[
A((m_j), f_i^a) = A((m_j), f_i^b) .
\]
This contradicts the uniqueness in Proposition [1].

Now fix \( k \in N_j \) and choose \( m_k^b > m_k^a \) regarding \( m = m_k \) as variable, while fixing the remaining \((m_j)\). Define \( \mathcal{F}(m) \) as in Lemma [2.6]. This yields canonical parameters \( \mathcal{F}(m)_k \) and \( \mathcal{F}(m)_i \) while all other components of \( \mathcal{F}(m)_k \) are fixed. Again, as in the proof of Lemma [2.6] the exponential distribution \( G_{\mathcal{F}(m)} \) becomes trivial after conditioning on \( \sigma_i \), which leads to
\[
m = \sum_{s_i} \frac{1 + m_i s_i}{2} \tanh(f_0 s_i + F(m)_k) .
\]
Hence, \( \mathcal{F}(m)_k \) is an increasing function of \( m \). From Fact [1] \( \mathcal{F}(m)_i \) must be a non-increasing function of \( m \) in order to keep \( m_i \) fixed. So, \( \mathcal{F}(m_k^b)_i \leq \mathcal{F}(m_k^a)_i \). We get
\[
m_i = A((m_k^b, m_{j \in N(i) \setminus \{k\}}), \mathcal{F}(m_k^a)_i) = A((m_k^b, m_{j \in N(i) \setminus \{k\}}), \mathcal{F}(m_k^a)_i) \geq A((m_k^b, m_{j \in N(i) \setminus \{k\}}), \mathcal{F}(m_k^a)_i) ,
\]
where the final inequality follows from the monotonicity of \( A \) with respect to the last variable, which has been proven at the beginning. This ends the proof of Proposition [3].
3. Special case of the Ising model. In this section we will consider the Ising model on a torus without external field.

We will chiefly be interested in the lattice
\[ T_n := (\mathbb{Z} \times \mathbb{Z})/(n\mathbb{Z} \times n\mathbb{Z}), \]
made into a graph by assuming connections between the closest neighbors. This will serve as graph \( V \) from the general setting. Furthermore, the potential takes the simple form
\[ U(J, (s_i)) = J \sum_{\{i,j\} \in E} s_is_j. \]
In relation to the general setting, the canonical parameter \( f_0 = J \) and all others are 0.

3.1. Exponential distribution of the cross subgraph.

3.1.1. The partition function. We will now establish certain properties of the neighbor average function by direct calculations. Star graphs (cf. Definition 2.3) \( C_i \) are isomorphic for all \( i \) and we will refer to such a graph as a cross. We will label its vertices \( c \) for the central one and 1 through 4 for the others.

Consider the partition function of the Gibbs distribution on the cross
\[ Z(J, f_c, f_1, f_2, f_3, f_4) = \sum_{(s_i)} \exp \left[ U(J, (s_i)) + \sum_{i=c,1,\ldots,4} f_i s_i \right], \]
where \( i = c, 1, 2, 3, 4 \). Note the presence of the additional linear term in the energy in spite of our hypothesis of no external field.

Lemma 3.1.
\[ Z(J, f_c, f_1, f_2, f_3, f_4) = 16 \left[ \exp(f_c) \prod_{i=1}^4 \cosh(J + f_i) + \exp(-f_c) \prod_{i=1}^4 \cosh(J - f_i) \right]. \]

Proof. The proof is by inspection. The factor 16 corresponds to removing \( 1/2 \) from the hyperbolic cosines. Then, after multiplying out, on the right-hand side one gets the sum of 32 exponentials and each of them is \( \exp(U(J, (s_i))) \) for a particular configuration \((s_i)_{i=c,1}^4\). To observe this it helps to notice that the terms resulting from the first product correspond exactly to those configurations with \( s_c = +1 \).

3.1.2. Symmetric form of the means on the cross. To simplify further calculations, we assume that \( m_1 = m_2 \) and \( m_3 = m_4 \). From Proposition 1 we get
\[ (J, 0, f_1, f_2, f_3, f_4) = \Psi_{N^c(i) \cup (i), N^i(j)}((m_j)_{j=1}^4, (J, 0)). \]
Note that \( f_c \) is always 0 since we will be looking for stationary points with 0 canonical parameters.

Notice that \( f_1 = f_2 \) and \( f_3 = f_4 \). If \( f_1 \neq f_2 \), then we consider the canonical parameters \( f'_1 = f_2, f'_2 = f_1, f'_3 = f_3, f'_4 = f_4 \) and denote this set of canonical parameters by \( f' \). By symmetry,
\[ \langle \sigma_1 \rangle_{G'}^{f'} = m_2, \langle \sigma_2 \rangle_{G'}^{f'} = m_1, \langle \sigma_3 \rangle_{G'}^{f'} = m_3, \langle \sigma_4 \rangle_{G'}^{f'} = m_4. \]
But since $m_1 = m_2$, this is the same as the original set of $\langle \sigma_i \rangle^{G_f}$ and $f_2 = f'_1 = f_1$ from the uniqueness in Proposition 1. We will write $m = m_1$, $M = m_3$ and $f = f_1$, $F = f_3$.

Also, adopt notations $c_\pm = \cosh(J \pm f)$ and $C_\pm = \cosh(J \pm F)$. In the future we will also use $s_\pm$, $S_\pm$, $t_\pm$, $T_\pm$ replacing $\cosh$ with $\sinh$ and $\tanh$, respectively. Then,

$$Z(J, f, F) := Z(J, 0, f, F, F) = 16 \left[ c_+^2 C_+^2 + C_-^2 c_-^2 \right].$$

3.1.3. Expected values. As it is well known (cf. Eq. (2)) expected values with respect to the exponential distribution are obtained by

$$\langle \sigma_i \rangle^{G_f} = \frac{\partial \log Z(J, f_c, f_1, f_2, f_3, f_4)}{\partial f_i} \bigg|_{(f_i) = (0, f, f, F, F)},$$

for $i = c, 1, \cdots, 4$. Using our notations this leads to

$$m = \langle \sigma_1 \rangle^{G_f} = \frac{c_+ s_+ C_+^2 - c_- s_- C_-^2}{c_+^2 C_+^2 + c_-^2 C_-^2},$$

$$M = \langle \sigma_2 \rangle^{G_f} = \frac{c_+^2 C_+ S_+ - c_-^2 C_- S_-}{c_+^2 C_+^2 + c_-^2 C_-^2},$$

$$A_J(m, m, M, M) = \langle \sigma_0 \rangle^{G_f} = \frac{c_+^2 C_+^2 - c_-^2 C_-^2}{c_+^2 C_+^2 + c_-^2 C_-^2}. \tag{7}$$

Here, $A_J(m, m, M, M, 0)$ is the neighbor average function (cf. Definition 2.4) in which we have made the dependence on $J$ explicit with the use of the subscript.

3.1.4. The second difference. We will now endeavor to compute the second difference $\Delta_2(J, f, F) = 2A_J(m, m, M, M, 0) - m - M$. We will denote $\gamma = f + F$, $\delta = F - f$. Observe that

$$\Delta_2(J, f, F)Z(J, f, F)/16 =$$

$$= c_+ C_+^2 (c_+ - s_+) + c_+^2 C_+ (C_+ - S_+) - c_- C_-^2 (c_- - s_-) - c_+^2 C_+^2 (C_+ - S_+) - c_-^2 C_-^2 (C_+ - S_-)$$

$$= c_+ C_+ e^{-J-f} + c_+^2 C_+ e^{-J-F} - c_- C_- e^{-J+f} - c_-^2 C_- e^{-J+F} =$$

$$= \frac{1}{2} C_+ c_+ \left[ e^{\delta} + e^{-2J-\gamma} + e^{-\delta} + e^{-2J+\gamma} \right] -$$

$$- \frac{1}{2} C_+ c_- \left[ e^{\delta} + e^{-2J+\gamma} + e^{-\delta} + e^{-2J-\gamma} \right] =$$

$$= C_+ c_+ \left[ \cosh(\delta) + e^{-2J-\gamma} \right] + C_+ c_- \left[ \cosh(\delta) + e^{-2J+\gamma} \right].$$

Now recall the formula

$$\cosh a \cosh b = \frac{1}{2} \left[ \cosh(a + b) + \cosh(a - b) \right].$$
When we use this to transform $C_-c_-$ and $C_+c_+$, the result is
\[
\Delta_2(J, f, F)Z(J, f, F)/16 = \frac{1}{2} \left[ \left[ \cosh(2J + \gamma) + \cosh(\delta) \right] \left[ \cosh(\delta) + e^{-2J-\gamma} \right] + \right. \\
- \left[ \cosh(2J - \gamma) + \cosh(\delta) \right] \cdot \left[ \cosh(\delta) + e^{-2J+\gamma} \right] \right] = \\
\frac{1}{2} \cosh(\delta) \left[ \cosh(2J + \gamma) - \cosh(2J - \gamma) + e^{-2J-\gamma} - e^{-2J+\gamma} \right] + \\
\frac{1}{2} \left[ \cosh(2J + \gamma)e^{-2J-\gamma} - \cosh(2J - \gamma)e^{-2J+\gamma} \right].
\]

Taking into account the identity
\[
\cosh(a + b) - \cosh(a - b) = 2 \sinh a \sinh b,
\]
we arrive at
\[
\Delta_2(J, f, F)Z(J, f, F)/16 = \\
\cosh(\delta) \left[ \sinh(2J) \sinh(\gamma) - \exp(-2J) \sinh(\gamma) \right] + \\
+ \frac{1}{4} \left[ \exp(-4J - 2\gamma) - \exp(-4J + 2\gamma) \right] = \\
\frac{1}{2} \sinh(\gamma) \left[ \cosh(\delta) \left[ \exp(2J) - 3 \exp(-2J) \right] - 2 \exp(-4J) \cosh(\gamma) \right],
\]
where in getting to the final term we have also used the identity
\[
\sinh(2\gamma) = 2 \cosh(\gamma) \sinh(\gamma).
\]

3.2. Consequences of formula (8). We proceed to derive properties of the neighbor average function.

**Lemma 3.2.** For each real value of $J$, the equation $A_f(m, m, m, m, 0) = m$ has at most one positive solution denoted by $\tilde{m}(J)$. The corresponding value of the canonical parameter at $i = 1, 2, 3, 4$ denoted by $\tilde{f}(J)$ satisfies the condition
\[
cosh[2\tilde{f}(J)] = \exp(6J) - 3 \exp(2J)
\]
and the solution exists if and only if $\tanh(J) > \frac{1}{3}$. When $\tanh(J) \leq \frac{1}{3}$, $A_f(m, m, m, m, 0) < m$ for all positive $m$.

**Proof.** The necessary condition stated by the Lemma is obtained from formula (8) by setting $\delta = 0, \gamma = 2\tilde{f}(J), \Delta_2(J, \tilde{f}, \tilde{f}) = 0$. The uniqueness of $\tilde{m}(J)$ then follows from Proposition 1. Denoting $y := \exp(2J)$ we see that the solution exists if and only if $y^3 - 3y - 2 > 0$. This is only possible when $y > 1$. On the interval $(1, +\infty)$ the left-hand side is increasing with $y$ and by inspection one verifies that for $y = 2$ the equality holds. Hence, the necessary and sufficient condition for the existence of a solution is $\exp(2J) > 2$ which in turn is equivalent to $\tanh(J) > \frac{1}{3}$. If this condition is not satisfied, then the expression in the curly bracket of formula (8) is negative for $\delta = 0$ and any $\gamma > 0$, which leads to the final claim of the Lemma. □
3.2.1. Proof of Theorem 1.1 for the case of \( \tanh(J) \leq 1/3 \). Suppose that there is a stationary point with non-zero mean somewhere and without loss of generality let \( m_i = 2b_i - 1 > 0 \). Then \( m_i = A_J((m_j)_{j \in \mathcal{N}(i)}, 0) \) and by Proposition 3 \( m_i \leq A_J((m_j = m_i)_{j \in \mathcal{N}(i)}, 0) \). If \( F \) is the corresponding value of the canonical parameters given by \( \Psi_{\mathcal{N}(i) \cup (i)}((m_j = m_i)_{j \in \mathcal{N}(i)}, (J, 0)) = (J, 0, F, F, F) \), then \( \Delta_2(J, F, F) = 2A_J((m_j = m_i)_{j \in \mathcal{N}(i)}, 0) - 2m_i \geq 0 \) contrary to the final claim of Lemma 3.2.

3.2.2. Neighbor average function and the arithmetic mean.

**Theorem 3.3.** Suppose that \( \tanh(J) > 1/3 \), \( m, M \in (-1, 1) \) are chosen so that 
\[
0 \leq \min(m, M) \leq \max(m, M) \leq \tilde{m}(J).
\]
Then \( A_J(m, m, M, M, 0) \geq \frac{m + M}{2} \) with the equality only possible if \( m = M \) and both are either 0 or \( \tilde{m}(J) \).

The claim of Theorem 3.3 is equivalent to showing that the second difference \( \Delta_2(J, f, F) \) is non-negative or zero, respectively. That is given explicitly by formula (8). The only difficulty is that the formula is given in terms of the canonical \( f, F \) while the hypothesis of the Proposition involves the means \( m, M \). The proof of Theorem 3.3 contains therefore some amount of detail not important for the main line of the paper and can be found in the Appendix.

4. Diagonal stationary points. In this section we prove the following proposition.

**Proposition 4.** If \( \tanh(J) > 1/3 \), there exists \( n(J) \) such that for every \( k \geq n(J) \) one can construct on \( T_{2k} \) a pseudo-marginal \( \{(b_i), (b_{ij})\} \) for which \( (b_i) \) are fixed on diagonals, but vary by taking both positive and negative values on different diagonals, and for every \( i \) the equality
\[
m_i = A_J((m_j)_{j \in \mathcal{N}}, 0)
\]
holds and \( m_i = 2b_i - 1 \) are the singleton means.

Observe that Theorem 1.1 then follows in the light of Proposition 4.

4.1. Construction. We construct a pseudo-marginal which is pairwise maximal and hence defined by its singleton means. The largest positive mean \( t_0 \) is taken on some diagonal, then a smaller but still positive mean \( t_1 \) on both adjacent diagonals, then even smaller \( t_2 \) on two diagonals one more step away from the initial and so on until for some \( k \) we have \( t_k = 0 \). The construction is illustrated on Fig. 1. The neighbor average condition from Proposition 4 is equivalent to
\[
\begin{align*}
t_0 &= A_J(t_1, t_1, t_1, t_1, 0) \\
t_\ell &= A_J(t_{\ell-1}, t_{\ell-1}, t_{\ell+1}, t_{\ell+1}, 0) \\
t_k &= 0
\end{align*}
\]
for \( 0 < \ell < k \). Recall that the final zero argument of \( A \) corresponds to the fact that there is no external field. If these conditions can be satisfied, then the means are further defined by symmetry. First, we set \( t_{k+\ell} = -t_{k-\ell} \) for \( 0 < \ell < k \). This results in \( t_{k+1} = -t_{k-1} \) which means that (9) also holds for \( \ell = k \). Also, \( t_{2k} = -t_0 \) and (9) hold for \( k < \ell < 2k \) by symmetry. Finally, \( t_{2k+\ell} = -t_\ell \) for \( 0 \leq \ell < 2k \) which leads to \( t_{4k-1} = -t_{4k-1} = t_1 \) consistent with our original setting.

The neighbor average conditions are then automatically satisfied and so Proposition 4 will follow.
The possibility of satisfying (9). It remains to show that conditions (9) can be satisfied for any $J > J_c$ with some $k$. Pick $J$ so that $\tanh(J) > 1/3$ and specify $t_1 \in (0, \tilde{m}(J))$ and set

$$t_0 = A_J(t_1, t_1, t_1, 0).$$

From Theorem 3.3, $t_0 > t_1$.

Further construction depends on the following Lemma.

**Lemma 4.1.** Suppose that $0 \leq m < M \leq \tilde{m}(J)$. Then there is a unique real number $\phi(M, m) < m$ such that $m = A_J(\phi(M, m), \phi(M, m), M, M, 0)$.

**Proof.** Uniqueness follows from the monotonicity of $A$ (cf. Proposition 3). To get existence, note that $0 = A_J(-M, -M, M, M, 0) \leq m$ but

$$A_J(m, m, M, M, 0) \geq \frac{M + m}{2} > m,$$

by Theorem 3.3. Then the equality $A_J(x, x, M, M, 0) = m$ is achieved for some $-M \leq x < m$ by the intermediate value theorem. $\square$

Hence, one can define $t_{\ell+1} = \phi(t_{\ell-1}, t_{\ell})$ for $\ell \geq 1$ as long as $t_{\ell} \geq 0$. This gives a decreasing sequence in which the second condition of (9) is satisfied. $t_{\ell}$'s constructed in this way are continuous functions of $t_1$ and since $t_1 = \tilde{m}(J)$ would lead to a constant sequence, there must be

$$\lim_{t_1 \to \tilde{m}(J)} t_{\ell}(t_1) = \tilde{m}(J).$$

Now suppose that $t_k < 0$ for some $k$ which would prevent a further construction of the sequence. Then we can start increasing $t_1$. Since $t_k$ depends on $t_1$ in a continuous fashion and its limit as $t_1 \to \tilde{m}(J)$ is $\tilde{m}(J)$, then for some $t_1$ we have $t_k(t_1) = 0$. It could also happen that at some increased value of $t_1$, $t_{k-1}(t_1) = 0$ preventing us from even constructing $t_k(t_1)$, but then we would have solved (9) with $k - 1$ instead of $k$.

So it remains to show that the construction of numbers $t_\ell$ cannot always give positive values. This follows from Theorem 3.3 since

$$t_\ell \geq \frac{t_{\ell-1} + t_{\ell+1}}{2}, \quad (10)$$

for as long as all these numbers are non-negative. Then $t_\ell - t_{\ell+1} \geq t_{\ell-1} - t_{\ell}$ for $\ell \geq 1$. It means that the sequence $t_\ell$ decreases faster than the arithmetic progression with initial terms $t_0$ and $t_1$.

**4.1.2. Proof of Proposition 4.** Once for a given $J$ a value of $k$ was chosen as described above, we can use it as $(J)$ in Proposition 3. We will show by induction that for a given $J$ if (9) has a solution with some $k_0$, it also has a solution with any $k > k_0$. As in the the preceding reasoning allow $t_1$ to tend to $\tilde{m}(J)$. Without loss of generality $t_k(t_1)$ is then a continuous function of $t_1$ which takes only positive values and tends to $\tilde{m}(J)$ as $t_1 \to \tilde{m}(J)$. Then $t_{k+1}(t_1)$ can be defined and also tends to $\tilde{m}(J)$ with $t_1$, so it will be 0 at some point.
4.2. Numerical simulations for diagonal equilibria. For $J = 0.5$ and $k = 4$ we obtain from Eq. (6) the following $4k = 16$ values of singleton means
\[
\begin{align*}
t_0 &= 0.82136, & t_1 &= 0.78063, & t_2 &= 0.63953, & t_3 &= 0.37019, \\
t_4 &= 0., & t_5 &= -0.37019, & t_6 &= -0.63953, & t_7 &= -0.78063, \\
t_8 &= -0.82136, & t_9 &= -0.78063, & t_{10} &= -0.63953, & t_{11} &= -0.37019, \\
t_{12} &= 0., & t_{13} &= 0.37019, & t_{14} &= 0.63953, & t_{15} &= 0.78063.
\end{align*}
\] (11)

A visualization of this diagonal stationary point of the Bethe functional is presented on Fig. 1.

Eq. (11) shows a vector of means, fixed on diagonals of $T_{16}$, which according to Propositions 2 and 4 is a stationary point of the Bethe functional. One can also check this numerically. For this purpose let $B_0$ be the original vector of means ($t_\ell$) (see Eq. (11)), and $B_\eta$ the following vector
\[
B_\eta = (t_\ell + \eta X_\ell)_{\ell = 1, \ldots, 16},
\] (12)

where $t_\ell$ are given with Eq. (11), $X_\ell$ are positive numbers drawn independently from the uniform distribution on the interval [0, 1] and $\eta$ is a real parameter. For thus constructed $B_\eta$ one can calculate the Bethe functional from Eq. (6) according to the following rules
\begin{itemize}
  \item $b_\ell = \frac{1 + t_\ell}{2} \text{ from Proposition 2, of course } b_\ell(-1) = 1 - b_\ell(1)$,
  \item $b_{\ell,\ell+1}$ are then determined from the requirement of pairwise maximality.
\end{itemize}

Calculated values of the Bethe functional for the $B_\eta$ as a function of $\eta$ are presented in Fig. 2. Fig. 2 may misleadingly suggest that the diagonal stationary point is a local maximum. By the method of Section C one can show that is not the case. This shows the danger of trying to determine the type of a stationary point by random experimentation — not one of our perturbations detected the non-definiteness of its Hessian. Only after computations described in Section C one
gets the perturbation $P = (X_\ell)$ which directly shows that the stationary point is a saddle:

$$P = \{0.07888, 0.09712, 0.14878, 0.21143, 0.24065, 0.21143, 0.14878, 0.09712, 0.07888, 0.09712, 0.14878, 0.21143, 0.24065, 0.21143, 0.14878, 0.09712\}.$$  

Values of the negative Bethe functional calculated in the direction of $P$ are presented in Fig. 3. So, while it is much more likely to find a stable direction, as illustrated by Fig. 2, Fig. 3 shows that at least one unstable direction given by $P$ also exists.

4.3. **Existence of a local maximum of diagonal type.** Let us consider the evidence of Finding 1. To begin with, we provide the listing of the means of a
diagonal stationary point (cf. Eq. 13) which was found by methods of Section 4.1 for \( J = 0.5 \) and \( k = 25 \). Unlike an example from Section 4.2 (given by Eq. 11), this stationary point is stable.

The stability of the stationary point given by Eq. (13) was verified using two numerically-based methods. Firstly, we checked that the negative Hessian matrix of the Bethe functional is positive definite by the method of Section 4.2. The resulting value of \( \chi \) was

\[
\chi = 9.9956 \cdot 10^{-9} > 0,
\]

which implies stability of the stationary point given by Eq. (13).

\[
t_0 = 0.92858, \quad t_1 = 0.92858, \quad t_2 = 0.92858, \quad t_3 = 0.92858, \\
t_4 = 0.92858, \quad t_5 = 0.928584, \quad t_6 = 0.928584, \quad t_7 = 0.928584, \\
t_8 = 0.928584, \quad t_9 = 0.928584, \quad t_{10} = 0.928583, \quad t_{11} = 0.928583, \\
t_{12} = 0.92858, \quad t_{13} = 0.928571, \quad t_{14} = 0.928544, \quad t_{15} = 0.928461, \\
t_{16} = 0.928203, \quad t_{17} = 0.927403, \quad t_{18} = 0.924926, \quad t_{19} = 0.917331, \\
t_{20} = 0.894648, \quad t_{21} = 0.831454, \quad t_{22} = 0.679936, \quad t_{23} = 0.395056, \\
t_{24} = 0., \quad t_{25} = -0.395056, \quad t_{26} = -0.679936, \quad t_{27} = -0.831454, \\
t_{28} = -0.894648, \quad t_{29} = -0.917331, \quad t_{30} = -0.924926, \quad t_{31} = -0.927403, \\
t_{32} = -0.928203, \quad t_{33} = -0.928461, \quad t_{34} = -0.928544, \quad t_{35} = -0.928571, \\
t_{36} = -0.92858, \quad t_{37} = -0.928583, \quad t_{38} = -0.928583, \quad t_{39} = -0.928584, \\
t_{40} = -0.928584, \quad t_{41} = -0.928584, \quad t_{42} = -0.928584, \quad t_{43} = -0.928584, \\
t_{44} = -0.928584, \quad t_{45} = -0.928584, \quad t_{46} = -0.928584, \quad t_{47} = -0.928584, \\
t_{48} = -0.928584, \quad t_{49} = -0.928584, \quad t_{50} = -0.928584, \quad t_{51} = -0.928584, \\
t_{52} = -0.928584, \quad t_{53} = -0.928584, \quad t_{54} = -0.928584, \quad t_{55} = -0.928584, \\
t_{56} = -0.928584, \quad t_{57} = -0.928584, \quad t_{58} = -0.928584, \quad t_{59} = -0.928584, \\
t_{60} = -0.928583, \quad t_{61} = -0.928583, \quad t_{62} = -0.92858, \quad t_{63} = -0.928571, \\
t_{64} = -0.928544, \quad t_{65} = -0.928461, \quad t_{66} = -0.928203, \quad t_{67} = -0.927403, \\
t_{68} = -0.924926, \quad t_{69} = -0.917331, \quad t_{70} = -0.894648, \quad t_{71} = -0.831454, \\
t_{72} = -0.679936, \quad t_{73} = -0.395056, \quad t_{74} = 0., \quad t_{75} = 0.395056, \\
t_{76} = 0.679936, \quad t_{77} = 0.831454, \quad t_{78} = 0.894648, \quad t_{79} = 0.917331, \\
t_{80} = 0.924926, \quad t_{81} = 0.927403, \quad t_{82} = 0.928203, \quad t_{83} = 0.928461, \\
t_{84} = 0.928544, \quad t_{85} = 0.928571, \quad t_{86} = 0.92858, \quad t_{87} = 0.928583, \\
t_{88} = 0.928583, \quad t_{89} = 0.928584, \quad t_{90} = 0.928584, \quad t_{91} = 0.928584, \\
t_{92} = 0.928584, \quad t_{93} = 0.928584, \quad t_{94} = 0.928584, \quad t_{95} = 0.928584, \\
t_{96} = 0.928584, \quad t_{97} = 0.928584, \quad t_{98} = 0.928584, \quad t_{99} = 0.928584.
\]

The second method consists in checking that after a small perturbation (in our case it was a positive “bump” \( \varepsilon = 10^{-5} \) added to all terms of Eq. (13) the pseudo-marginal returns to its non-perturbed form under iterated neighborhood average function. This test, which is explained on Fig. 5 in an algorithmic form, is very sensitive, because it breaks the symmetry of sign flipping. In the absence of stability this should lead to convergence to the fixed-mean positive solution \( \hat{m}(J) \). That happens for saddle point given by Eq. (11).

Notice that by Corollary 1 replacing the mean at the center of any cross by the value of its neighborhood average function increases the value of the Bethe functional, so local maxima attract their neighborhoods under the iteration of this procedure. On the other hand, for a saddle convergence could only occur if the initial point was chosen on the stable invariant manifold under the scheme, which has probability 0.
Acknowledgments. We thank the referees for many useful comments. We also thank Jurek Jaromczyk for inspiration and discussions concerning the BP algorithm and Tomek Miller for proofreading the manuscript. Both authors are partially supported by NCN grant UMO-2015/17/B/ST1/00091.

Appendix A. Proof of Proposition 1. Recall the concept of the correlation polytope

\[ M = \{(\langle \sigma_i \rangle)_{i \in V \cup \{0\}}\}, \]

where \( P \) ranges over all probability distributions on the set of configurations.

As a consequence of Theorem 3.3 of \cite{14} we get the claim of Proposition 1 in the situation when \( V' = V \cup \{0\} \) and provided that \( m' \) belongs to the interior of the correlation polytope \( M \).

Consider the set

\[ W = \{ \hat{f} \in \mathbb{R}^{\left| V' \right| + 1} : \exists f' \in \mathbb{R}^{\left| V' \right|} \forall j \in V' \langle \sigma_j \rangle^{G_{j\cup j}} = m_j \}. \]

First observe that \( W \) is non-empty. This follows from the Theorem quoted above once we notice that the vector \( m' \) can be extended by

\[ \hat{m} = (m_i)_{i \in (V \setminus V') \cup \{0\}}, \]

so that \( m' \cup \hat{m} \in M^o \), where \( M^o \) means the interior of the correlation polytope. This can be realized by a Bernoulli distribution.

Test's algorithm:

1) for \( \alpha = 0 \) to 99 do
2) \( t_\alpha = t_\alpha + \varepsilon; \)
3) end
4) do
5) for \( \alpha = 0 \) to 99 do
6) \( t_\alpha = \text{A}(t_{\alpha+1 \mod 100}, t_{\alpha+1 \mod 100}, t_{\alpha-1 \mod 100}, t_{\alpha-1 \mod 100}, 0); \)
7) end
8) until the scheme has converged

Figure 5. The stability test algorithm.
The condition \( \langle \sigma_j \rangle^{G_{t/\omega}} = m_j \) for \( j \in V' \) can be seen as a system of equations with variables \( f' \) and \( \hat{f} \). By the implicit function theorem it can be locally solved for \( \hat{f} \) in terms of \( f' \) provided that the matrix
\[
\frac{\partial \langle \sigma_j \rangle^{G_{t/\omega}}}{\partial f_\ell} = \text{Cov} \langle \sigma_j, \sigma_\ell \rangle,
\]
is non-singular, where \( j, \ell \in V' \). That is the case for the covariance matrix in the Ising model. Thus, \( W \) is open and one can also conclude that the map \( f' \mapsto m' \) is a local diffeomorphism.

Suppose that \( \hat{f}_\infty \in \partial W \) and let \( \hat{f}_n \) be a sequence from \( W \) which converges to \( \hat{f}_\infty \), while \( f'_n \) are the corresponding vectors supported on \( V' \) for which the condition which defines \( W \) holds. Observe that for every \( j \in V' \) the sequence \( (f_{j,n})_{n \in \mathbb{N}} \) is bounded. Indeed, the canonical parameters \( f_{0,n} \) form a convergent sequence and are bounded, which means that the coupling between different nodes in the Ising model remains bounded. In that case we can see from formula \( 1 \) that \( f_j, n \) tending to \( +\infty \), or \( -\infty \), would result in \( \langle \sigma_j \rangle^{G_{f,n}} \) tending to \( 1, -1 \), resp., contrary to the fact that this quantity remains fixed at the value of \( m_j \). Without loss of generality, then, \( f_{j,n} \to f_j, \infty \) for all \( j \in V' \). Writing \( f'_\infty = (f_{j,\infty})_{j \in V'} \) and \( f_\infty = f'_\infty \cup \hat{f}_\infty \), by continuity we conclude that
\[
\langle \sigma_j \rangle^{G_{f,\infty}} = \lim_{n \to \infty} \langle \sigma_j \rangle^{G_{f,n}} = m_j,
\]
hence \( \hat{f}_\infty \in W \).

Since \( W \) is non-empty, open and closed it must be all of \( \mathbb{R}^{|V'|-|V'|+1} \). This proves the existence part of Proposition \( 1 \).

To see uniqueness, fix \( f \) and recall the observation that \( f' \mapsto m' \) is a local diffeomorphism. It is also proper, that is the pre-image of any compact set in \((-1, 1)^{|V'|} \) is compact. This again follows from the observation that with \( f_0 \) fixed, \( f_j \) tending to \( \pm \infty \) implies \( \langle \sigma_j \rangle^{G_{f}} \) tends to \( \pm 1 \). By topology, a local homeomorphism which is proper and onto a simply-connected space is a global homeomorphism.

**Appendix B. Proof of Theorem 3.3.** Let us recall that we are dealing with the exponential distribution on the cross \( C_i \) with canonical parameters \( f_0 = 0, f_1 = f_2 = f, f_3 = f_4 = F \) and means \( m_1 = m_2 = m, m_3 = m_4 = M \). Also, notations from the fragment \( 3.1.4 \) will be used, in particular \( \delta = F - f, \gamma = f + F \).

**Lemma B.1.** If \( m + M \geq 0 \), then \( f + F \geq 0 \) and \( m + M = 0 \) iff \( f + F = 0 \).

**Proof.** From the symmetric form of the exponential distribution \( f = -F \) implies \( m = -M \) and as \( f \) changes from \(-\infty \) to \( \infty \), all values of \( m \in (-1, 1) \) are taken. This implies the equivalence of \( m + M = 0 \) and \( f + F = 0 \) in view of Proposition \( 1 \).

Now, to get a contradiction, suppose that the first claims fails, which means that \( f < -F \leq 0 \leq F \). Consider the pair of canonical parameters \( (f', F') := (-F, -f) \). This pair is greater than \( (f, F) \). Writing \( m', M' \) for the means with respect to this new choice of canonical parameters, in view of the monotonicity of expectations (cf. Fact \( 3 \)) this implies \( m' > m, M' \geq M \). But from the symmetry of the exponential distribution, \( m' = -M, M' = -m \). Hence, \( -(M + m) > m + M \) which contradicts the hypothesis \( m + M \geq 0 \).

By Lemma B.1, in formula \( 8 \), \( \gamma \geq 0 \) and so the inequality in Theorem 3.3 is equivalent to
\[
\cosh(\delta) [\exp(2J) - 3 \exp(-2J)] - 2 \exp(-4J) \cosh(\gamma) \geq 0. \tag{14}
\]
One case in which inequality (14) is easy to see is when \( \min(f,F) < 0 \). Then \( \gamma < \delta \) and the inequality follows from \( \exp(6J) - 3\exp(2J) - 2 > 0 \) which corresponds to \( \tanh(J) > \frac{1}{3} \). In the sequel, we will assume that \( \min(f,F) \geq 0 \).

Recall the parameter \( \tilde{m}(J) \) defined when \( \tanh(J) > \frac{1}{3} \) by the statement of Lemma B.2.

**Lemma B.2.** If \( 0 < m \leq \tilde{m}(J) \) and \( m = M \), then the inequality (14) holds with the equality only when \( m = \tilde{m}(J) \).

**Proof.** In such a case we have \( \delta = 0 \) and \( \cosh(\gamma) \leq \cosh(2\tilde{m}(J)) \). Condition (14) now follows from Lemma B.2. \( \square \)

The proof of inequality (14) in the general setting of Theorem 3.3 is based on the next Proposition.

First consider the notation \( \gamma(M) := 2f(M) \) where

\[
(J,0,f(M),f(M),f(M)) = \Psi_{N(i) \cup \{j\}}((M,M,M,M),(J,0)) .
\]

**Proposition 5.** Suppose that \( J > 0 \) and \( \max(m,M) \leq \tilde{m}(J) \). Assume further \( F \geq f \geq 0 \). Then \( \gamma \leq \gamma(M) \).

We postpone the proof of Proposition 5 and finish the proof of Theorem 3.3. Given \( m,M \) which satisfy its hypothesis, we fix \( m \leq M \) for definiteness. Inequality (14) holds when \( m := M \). For a smaller value of \( m \), the term \( \delta \) becomes non-zero so that \( \cosh(\delta) > 1 \), but \( \gamma \) does not increase according to Proposition 5, so that a sharp inequality follows.

Concerning the conditions for equality in Theorem 3.3, one has to have either \( \sinh(\gamma) = 0 \) in formula (8), or the equality in (14). By Lemma B.1, the first possibility occurs only when \( f = -F \). This entails \( m = -M \) by symmetry and since \( m \geq 0 \), the only possibility is \( m = M = 0 \). As we have just argued from Proposition 3, the equality in (14) may only occur when \( m = M = \tilde{m}(J) \) by Lemma B.2.

**B.0.1. Proof of Proposition 3.** We return to the general setting of the exponential distribution on the cross, with canonical parameters \( f = (J,0,f_i), i = 1,\ldots,4 \), and means \( m_i = \langle \sigma_i \rangle_{G_f} \). We will be interested in derivatives \( \frac{\partial \langle \sigma_i \rangle_{G_f}}{\partial f_j} \).

**Lemma B.3.** For every \( 1 \leq i,j \leq 4 \) and any \( f \),

\[
\frac{\partial \langle \sigma_i \rangle_{G_f}}{\partial f_j} = \frac{\partial \langle \sigma_j \rangle_{G_f}}{\partial f_i} .
\]

**Proof.** Recall that

\[
\frac{\partial \langle \sigma_i \rangle_{G_f}}{\partial f_j} = \text{Cov}_{G_f}(\sigma_i,\sigma_j) ,
\]

and the covariance matrix is symmetric. \( \square \)

We will now work in the setting where \( \varphi = (J,0,f,f,F) \) and \( \langle \sigma_1 \rangle_{G_{\varphi}} = \langle \sigma_2 \rangle_{G_{\varphi}} = m(\varphi) \) while \( \langle \sigma_3 \rangle_{G_{\varphi}} = \langle \sigma_4 \rangle_{G_{\varphi}} = M(\varphi) \).

Let \( E_q \) denote the event \( \sigma_3 + \sigma_4 = q \). \( E_q \) are not empty for \( q \in \{-2,0,2\} \). Then \( p_q(\varphi) \) denotes the probability of \( E_q \) with respect to the exponential distribution \( G_{\varphi} \).

**Lemma B.4.** If \( F \geq f \geq 0 \) and \( q = -2,0 \), then

\[
\frac{\partial p_q(\varphi)}{\partial F} < 0 ,
\]

and \( E_{G_{\varphi}}(\sigma_1|E_0) \geq 0 \).
Proof. To see the first claim, note that for all configurations in $E_q$ the value of

$$U(\varphi, \sigma) = J \sum_{i=1}^{4} \sigma_i \sigma_i + f(\sigma_1 + \sigma_2) + F(\sigma_3 + \sigma_4),$$

depends on $F$ through the term $F(\sigma_3 + \sigma_4) = Fq$. Hence, if $q \leq 0$, then $p_q(\varphi)$ has the form $g(f, F) / Z(J, 0, f, f, F, F)$ where $\partial g(f, F) / \partial F \leq 0$. On the other hand,

$$\partial Z / \partial F = Z \partial \log Z / \partial F = Z \langle \sigma_3 \rangle^G \varphi + \langle \sigma_4 \rangle^G \varphi \geq 0,$$

by the monotonicity of expectations, since all canonical parameters are non-negative.

As for the second claim, the conditional exponential distribution is given by the exponential distribution on the sub-graph which consists of nodes $c, 1, 2$ with canonical parameters $0, f, f$, respectively, and again the expected value of $\sigma_0$ is non-negative since $f \geq 0$.

Lemma B.5. If $f \geq 0$, then

$$\frac{\partial m(\varphi)}{\partial F} \leq \frac{\partial M(\varphi)}{\partial F}.$$

Proof. We start by observing that

$$M(\varphi) = \frac{1}{2} \left[ (\langle \sigma_3 \rangle^G \varphi + \langle \sigma_4 \rangle^G \varphi) = p_2(\varphi) - p_{-2}(\varphi) \right],$$

hence

$$\frac{\partial M(\varphi)}{\partial F} = \frac{\partial p_2(f)}{\partial F} - \frac{\partial p_{-2}(f)}{\partial F}.$$

Similarly,

$$m(\varphi) = \sum_{q \in \{-2, 0, 2\}} E^{G^*_q}(\sigma_1 | E_q) p_q(\varphi).$$

Now the conditional expectations with respect to $E_q$ no longer depend on $F$, so

$$\frac{\partial m(\varphi)}{\partial F} = \sum_{q \in \{-2, 0, 2\}} E^{G^*_q}(\sigma_1 | E_q) \frac{\partial p_q(\varphi)}{\partial F}.$$

Subtracting, we get

$$\frac{\partial M(\varphi)}{\partial F} \frac{\partial m(\varphi)}{\partial F} =$$

$$= \left[ 1 - E^{G^*_q}(\sigma_1 | E_2) \right] \frac{\partial p_2(\varphi)}{\partial F} - E^{G^*_q}(\sigma_1 | E_0) \frac{\partial p_0(\varphi)}{\partial F} -$$

$$- \left[ 1 + E^{G^*_q}(\sigma_1 | E_{-2}) \right] \frac{\partial p_{-2}(\varphi)}{\partial F}.$$

From Lemma B.4 and the monotonicity of expectations all three terms are seen to be non-negative. \qed

Lemma B.6.

$$\frac{\partial M(f)}{\partial f} = \frac{\partial m(f)}{\partial F}.$$

Proof. In terms of the partial derivatives with respect to the canonical parameters $f_i$, we can write

$$\frac{\partial M(\varphi)}{\partial f} = \frac{\partial \langle \sigma_3 \rangle^G \varphi}{\partial f_1} + \frac{\partial \langle \sigma_4 \rangle^G \varphi}{\partial f_2}.$$
From Lemma B.3, this is equal to
\[
\frac{\partial \langle \sigma_1 \rangle_{G \phi}}{\partial f_3} + \frac{\partial \langle \sigma_2 \rangle_{G \phi}}{\partial f_4} = \frac{\partial m(\phi)}{\partial F}.
\]

To prove Proposition 5, we consider a one-parameter family
\[
\varphi(f) := (J, 0, f, f, F(f), F(f)),
\]
of canonical parameters parametrized by \(f\). \(f\) varies from its initial value to \(f(M)\) (cf. Eq. (15)) while \(F(f)\) is determined by the condition \(M(\varphi(f)) = M\) for all \(f\).

From the implicit function theorem, this translates into
\[
\frac{\partial M(\varphi)}{\partial f} + \frac{\partial M(\varphi)}{\partial F} \frac{dF}{df} = 0.
\]

In view of Lemmas B.6 and B.5 \(\frac{dF}{df} < -1\). Hence, the sum \(f + M(f)\) decreases to its final value \(\gamma(M)\), which proves Proposition 5.

Appendix C. Testing of diagonal stationary points for stability. Recall the toral lattice \(T_n := (\mathbb{Z} \times \mathbb{Z})/(\mathbb{Z} \times \mathbb{Z})\). The vertices can be naturally labeled by pairs \(p, q \in \mathbb{Z}_n\). In order to explore the diagonal symmetry we will prefer the coordinates \((\alpha, \beta) := (p + q, q)\). So, the four neighbors of \((\alpha, \beta)\) are \((\alpha + 1, \beta)\), \((\alpha + 1, \beta + 1)\), \((\alpha - 1, \beta)\), \((\alpha - 1, \beta - 1)\). Further, we consider a stationary pseudo-marginal \(b\). Let \(t_{\alpha, \beta}\) denote the singleton means of \(b\). We assume diagonal symmetry: \(t_{\alpha, \beta} =: t_\alpha\) for all \(\beta\). We can introduce further
\[
V(\alpha) = \text{Cov}(\sigma_{\alpha, \beta}, \sigma_{\alpha, \beta}) = 1 - t_{\alpha}^2,
\]
where the probability distribution corresponds to an edge from \((\alpha, \beta)\) to its neighbor and is uniquely given by (necessarily pairwise maximal) singleton and pairwise components of \(b\). For the variance in Eq. (16) the choice of the neighbor does not matter and for covariances it is clear which neighbor is meant.

We further have the correlation coefficient \(C\), which is always positive by the monotonicity of expectations and given by
\[
C^2(\alpha, \alpha + 1) = C^2(\alpha + 1, \alpha) = \frac{\text{Cov}^2(\sigma_{\alpha, \beta}, \sigma_{\alpha + 1, \beta})}{V(\alpha)V(\beta)} = \frac{\sinh^2(2J)}{[\cosh(2J) + 2f][\cosh(2J) + 2F]},
\]
where \(f\) and \(F\) are such that the probability distribution induced on the edge \{\((\alpha, \beta), (\alpha + 1, \beta)\)\} is equal to the exponential distribution with canonical parameters \((J, f, F)\).

Recall that for pairwise maximal pseudo-marginals the Bethe functional is a function of singleton components \((b_i)\) only. Since in turn singleton pseudo-marginal components are trivially connected to the means by \(b_i = \frac{1 + m_i}{2}\), we can introduce
\[
H((t_\alpha)) := H_V \left(\left(\frac{1 + t_\alpha}{2}\right), B^{\text{max}} \left(\frac{1 + t_\alpha}{2}, \frac{1 + t_{\alpha + 1}}{2}\right), (0)\right),
\]
where \(B^{\text{max}}\) is the function which derives the maximizing value of the pairwise component of a pseudo-marginal from its singleton components and the final \((0)\) represents the fact that we use zero canonical parameters. The computation of the
Hessian matrix $D^2H$ is then an elementary though somewhat tedious task, which has been done in \cite{15}, Theorem 7. In our notation the result can be written as

$$
- (u_{\alpha,\beta})^T D^2H((t_{\alpha}) (u_{\alpha,\beta}) = 
\sum_{\alpha} \sum_{\beta} \left\{ \left[ 2 \left( \frac{1}{1 - C^2(\alpha, \alpha + 1)} + \frac{1}{1 - C^2(\alpha - 1, \alpha)} \right) - 3 \right] u_{\alpha,\beta}^2 / V(\alpha) - 
\frac{\sqrt{C(\alpha, \alpha + 1)}}{\sqrt{V(\alpha)V(\alpha + 1)(1 - C^2(\alpha, \alpha + 1))}} u_{\alpha,\beta}(u_{\alpha+1,\beta} + u_{\alpha+1,\beta+1}) - 
\frac{\sqrt{C(\alpha, \alpha - 1)}}{\sqrt{V(\alpha)V(\alpha - 1)(1 - C^2(\alpha, \alpha - 1))}} u_{\alpha,\beta}(u_{\alpha-1,\beta} + u_{\alpha-1,\beta-1}) \right\}. 
$$

Thus,

$$
\chi((u_{\alpha,\beta})) := -(u_{\alpha,\beta})^T D^2H((t_{\alpha}) (u_{\alpha,\beta}) = 
\sum_{\alpha} \sum_{\beta} \left\{ v(\alpha) u_{\alpha,\beta}^2 - c(\alpha, \alpha + 1) u_{\alpha,\beta}(u_{\alpha+1,\beta} + u_{\alpha+1,\beta+1}) - c(\alpha, \alpha - 1) u_{\alpha,\beta}(u_{\alpha-1,\beta} + u_{\alpha-1,\beta-1}) \right\},
$$

where

$$
v(\alpha) := V^{-1}(\alpha) \left[ 2 \left( \frac{1}{1 - C^2(\alpha, \alpha + 1)} + \frac{1}{1 - C^2(\alpha - 1, \alpha)} \right) - 3 \right],
$$

$$
c(\alpha, \alpha + 1) = c(\alpha + 1, \alpha) = \frac{\sqrt{C(\alpha, \alpha + 1)}}{\sqrt{V(\alpha)V(\alpha + 1)(1 - C^2(\alpha, \alpha + 1))}}.
$$

It is important to observe that $v(\alpha)$ and $c(\alpha, \alpha + 1)$ are always positive.

C.1. Minimization problem for $\chi$. Our goal is to decide whether $-D^2H(t_{\alpha})$ is positive definite. This will be done by minimizing $\chi((u_{\alpha,\beta}))$ under the constraint

$$
\sum_{\alpha} v(\alpha) \sum_{\beta} u_{\alpha,\beta}^2 = n.
$$

**Lemma C.1.** There exists a minimizer $(u_{\alpha,\beta})$ for which for all $\alpha, \beta, \beta'$ one gets

$$
u_{\alpha,\beta} = u_{\alpha,\beta'} > 0.
$$

**Proof.** Start by observing that without loss of generality all $u_{\alpha,\beta}$ are non-negative. Indeed, flipping the signs of $u_{\alpha,\beta}$ where needed to make them positive will not violate the constraint by make all mixed terms in \cite{18} negative.

Lagrange multipliers give

$$
\lambda v(\alpha) u_{\alpha,\beta} = c(\alpha, \alpha + 1)(u_{\alpha+1,\beta} + u_{\alpha+1,\beta+1}) + c(\alpha, \alpha - 1)(u_{\alpha-1,\beta} + u_{\alpha-1,\beta-1}). \quad (19)
$$

From here we first infer that all $u_{\alpha,\beta}$ are strictly positive. To this end choose $\alpha_0$ so that the diagonal set $\{u_{\alpha,\beta}\}_{\beta \geq \alpha_0}$ contains the largest possible number of zero elements. Since $c(\alpha, \alpha + 1)$ are all positive and $u_{\alpha,\beta}$ non-negative, condition \cite{19} implies that $u_{\alpha,\beta}, u_{\alpha,\beta+1}$ are both 0 (ditto for the $\alpha - 1$ diagonal, but we shall not need it). Thus, the $\alpha + 1$ diagonal set contains more 0 elements, unless all or none $u_{\alpha_0,\beta}$ were 0. All zeros are not possible since it would violate the constraint, so none were 0.

Notice that we now also know that $\lambda > 0$. 

**The End.**
Continuing along the same line or reasoning, now define
\[
M(\alpha) = \max\{u_{\alpha,\beta} : \beta \in \mathbb{Z}_n\}, \\
\text{and} \\
m(\alpha) = \min\{u_{\alpha,\beta} : \beta \in \mathbb{Z}_n\}.
\]
Then pick \(\alpha_0\) so that \(M(\alpha_0)/m(\alpha_0)\) is the largest possible and secondarily, if several diagonals maximize this ratio, choose among them so that \(M(\alpha_0)\) is taken for the largest number of values of \(\beta\).

Conditions \(\text{(19)}\) imply that for any \(\beta, \beta'\)
\[
\frac{u_{\alpha_0,\beta'}}{u_{\alpha_0,\beta}} \leq \frac{\min(u_{\alpha_0+1,\beta'}, u_{\alpha_0+1,\beta'+1})}{\max(u_{\alpha_0+1,\beta}, u_{\alpha_0+1,\beta+1})}.
\]

So, for this ratio to be equal to \(M(\alpha_0)/m(\alpha_0)\), we must have \(M(\alpha_0 + 1)/m(\alpha_0 + 1) = M(\alpha_0)/m(\alpha_0)\) and \(u_{\alpha_0+1,\beta'+1} = u_{\alpha_0+1,\beta'} = M(\alpha_0 + 1)\). So the \(\alpha_0 + 1\) shows the same value of the ratio \(M(\alpha)/m(\alpha)\) and \(M_{\alpha_0+1}\) is achieved for a larger number of values of \(\beta\), unless \(M_{\alpha_0} = u_{\alpha_0,\beta}\) already for all \(\beta\), which concludes the proof of the Lemma.

C.2. Simplified version of the minimization problem. If we write \(u_\alpha\) for the common value of \(u_{\alpha,\beta}\), the problem of determining the positive definiteness of the Hessian is reduced to finding the minimum of
\[
n^{-1}\chi((u_\alpha)) = \sum_\alpha [v(\alpha)u_{\alpha}^2 - 2u_\alpha(c(\alpha, \alpha + 1)u_{\alpha+1} + c(\alpha, \alpha - 1)u_{\alpha-1})],
\]
under the constraint \(\|u_\alpha\| = 1\) where
\[
\|u_\alpha\|^2 = \sum_\alpha v(\alpha)u_{\alpha}^2.
\]
Additionally, \(u_\alpha\) are positive.

This leads to the Lagrange multiplier setting of
\[
\lambda v(\alpha)u_\alpha = c(\alpha, \alpha + 1)u_{\alpha+1} + c(\alpha, \alpha - 1)u_{\alpha-1}.
\]

(20)

Let \(A\) denote the linear transformation
\[
(u'_\alpha) = A(u_\alpha) = v(\alpha)^{-1} [c(\alpha, \alpha + 1)u_{\alpha+1} + c(\alpha, \alpha - 1)u_{\alpha-1}].
\]

\(A\) has all non-negative coefficients and in fact \(A^n\) has them all strictly positive. By the Perron–Frobenius theorem \(A\) has a unique eigenvector with positive entries which can be obtained by iterating \(A\) on any initial non-negative vector. This eigenvector provides a solution to the minimization problem modulo the normalization.

This establishes the following iterative procedure.

• Choose an initial approximation \((u_\alpha)\) which satisfies the constraint and is positive, for example \(u_\alpha = 1/\sqrt{v(\alpha)}\).
• Given \((u_\alpha)\), set the next approximation
\[
(u'_\alpha) = \|A(u_\alpha)\|^{-1}A(u_\alpha).
\]

This will converge exponentially fast to a minimizing vector \((u_\alpha)\). The sign of \(\chi((u_\alpha))\) is then positive if and only if \(-D^2H((t_i))\) is positive definite, 0 iff it is positive semi-definite but not definite and negative if it is not positive semi-definite.
REFERENCES

[1] R. J. Baxter, Exactly solved models in statistical mechanics. Integrable Systems in Statistical Mechanics, 1 (1985), 5–63.

[2] H. A. Bethe, Statistical theory of superlattices. Selected Works of Hans A Bethe, 18 (1997), 245–270.

[3] S. Dorogovtsev, A. Goltsev and J. Mendes, Critical phenomena in complex networks. Rev. Mod. Phys., 80 (2008), 1275–1335.

[4] C. Fortuin, P. Kasteleyn and J. Ginibre, Correlation inequalities on some partially ordered sets. Commun. Math. Phys., 22 (1971), 89–103.

[5] T. Heskes, Stable fixed points of loopy belief propagation are local minima of the Bethe free energy, In S. Thrun S. Becker and K. Obermayer, editors, Advances in Neural Information Processing Systems 15, MIT Press, Cambridge, MA, (2003), 343–350.

[6] J. M. Mooij and H. J. Kappen, On the properties of the Bethe approximation and loopy belief propagation on binary networks. J. Stat. Mech. Theor. Exp., 11 (2005), P11012.

[7] J. M. Mooij and H. J. Kappen, Sufficient conditions for convergence of the sum-product algorithm. IEEE Transactions on Information Theory, 53 (2007), 4422–4437.

[8] S. Newhouse, Diffeomorphisms with infinitely many sinks. Topology, 13 (1974), 9–18.

[9] J. Pearl, Reverend Bayes on inference engines: A distributed hierarchical approach. Proceedings of the Second National Conference on Artificial Intelligence, (1982), 133–136.

[10] T. G. Roosta, M. J. Wainwright nd S. S. Sastry, Convergence analysis of reweighted sum-product algorithms. IEEE Transactions on Signal Processing, 56 (2008), 4293–4305.

[11] J. Shin, The complexity of approximating a bethe equilibrium. IEEE Transactions on Information Theory, 60 (2014), 3959–3969.

[12] G. Siudem and G. Świątek, Dynamics of the belief propagation for the ising model. Acta Physica Polonica A, 127 (2015), 3A145–3A149.

[13] S. Tatikonda and M. Jordan, Loopy belief propagation and Gibbs measures, in Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufman, San Francisco, (2002), 493–500.

[14] M. J. Wainwright and M. I. Jordan, Graphical models, exponential families, and variational inference. Foundations and Trends in Machine Learning, 1 (2008), 1–305.

[15] A. Weller and T. Jebara, Bethe bounds and approximating the global optimum. Journal of Machine Learning Research W& CP, 31 (2013), 618–631.

[16] M. Welling and Y.-W. Teh, Belief Optimization for Binary Networks: A Stable Alternative to Loopy Belief Propagation, in Proc. 17th Conference on Uncertainty in Artificial Intelligence (UAI), 2001.

[17] J. Yedidia, W. Freeman and Y. Weiss, Constructing free-energy approximations and generalized belief propagation algorithms. IEEE Trans. on Information Theory, 51 (2005), 2282–2312.

Received March 2016; revised December 2016.

E-mail address: siudem@if.pw.edu.pl
E-mail address: g.swiatek@mini.pw.edu.pl