Applications of Gibbs Measure Theory to Loopy Belief Propagation Algorithm

Nobuyuki Taga[1] and Shigeru Mase[1]

Tokyo Institute of Technology, Ookayama 2-12-1-w8-28,
Meguro-ku, Tokyo 152-8552, Japan
{Nobuyuki.Taga, mase}@is.titech.ac.jp

Abstract. In this paper, we pursue application of Gibbs measure theory to LBP in two ways. First, we show this theory can be applied directly to LBP for factor graphs, where one can use higher-order potentials. Consequently, we show beliefs are just marginal probabilities for a certain Gibbs measure on a computation tree. We also give a convergence criterion using this tree. Second, to see the usefulness of this approach, we apply a well-known general condition and a special one, which are developed in Gibbs measure theory, to LBP. We compare these two criteria and another criterion derived by the best present result. Consequently, we show that the special condition is better than the others and also show the general condition is better than the best present result when the influence of one-body potentials is sufficiently large. These results surely encourage the use of Gibbs measure theory in this area.

1 Introduction

Inference problems using graphical models are important in various application fields. The belief propagation (BP) algorithm is an efficient method for computing marginal probabilities of probabilistic networks without loops. BP can be formally applied also to networks with loops (LBP). However, if networks have loops, the algorithm may not converge and beliefs may not equal to exact marginal probabilities. Nevertheless, applications of LBP algorithm have been reported to be remarkably useful such as in the coding theory [1, 4, 6].

In analysis of LBP, Tatikonda and Jordan [8] applied Gibbs measure theory using the concept of computation trees, which was first introduced by Weiss [9]. They also gave a sufficient convergence criterion based on Simon's condition of Gibbs measure theory. Nevertheless, to use this theory seems not to be so popular.

In this paper, we pursue Gibbs measure approach. This paper is composed of two parts. First, we show that this theory can directly be applied to general potentials case. The concept of computation tree is important to apply Gibbs measure theory to LBP. However, it is discussed only for pair potential case and it is still unclear how to construct it where higher-order potentials exist. We give a construction of computation trees according to the LBP for factor graphs. Second, we show the effectiveness of Gibbs measure approach. Tatikonda and

Jordan derived a criterion based on Simon's condition of Gibbs measures theory in their paper. However, Ihler et al. [3] and Mooij and Kappen [5] independently proposed stronger conditions than it. In this paper, we apply a well-known condition called Dobrushin's condition and compare the convergence criteria derived from this and Ihler's and Mooij's approaches for Ising models on complete graphs. This model has only pair potentials. However, it is remarkable since the complete characterization of the phase transition region of the associated Gibbs measure is known. We also use the convergence criterion derived from this characterization to compare.

In Sect. 2, we review the LBP algorithm on factor graphs and derive its computation tree. We also give some results using Gibbs measure theory with the computation tree. In Sect. 3, we compare three LBP convergence criteria for Ising models. We give a conclusion in Sect. 4. In Appendix, we show how to check Dobrushin's condition for Ising models on complete graphs.

2 LBP Algorithm on Factor Graphs and Its Computation Trees

To analyze LBP algorithm, Tatikonda and Jordan [8] utilize *Gibbs measure theory*. This theory deals with so-called Gibbs measures, defined on a set of infinite nodes. Its main concern is to investigate the phase transition phenomenon. They used the concept of computation tree, which was first introduced by Weiss [9], to connect LBP algorithm with Gibbs measure theory.

In their paper, they discussed mainly about the BP algorithm for pair potentials. On the other hand, if there exists higher-order potentials, the concept of computation trees is never clear. It should be noted that conversion of higher-order potential case into pair potential case may have the validity of the application of Gibbs measure theory become uncertain since the assumption of the positivity of the probability function is not necessarily preserved. For example, see [9]. In this section, we look at the BP algorithm for factor graphs, with which one can use probability functions with general potentials, and investigate how to construct its computation trees. We also give some results applying Gibbs measure theory with the computation tree.

2.1 BP Algorithm on Factor Graphs

Let consider a network on the node set $\{1, 2, \ldots, n\}$ and an associated set of random variables $X = \{X_1, X_2, \ldots, X_n\}$. Assume that each state space E_i of X_i is discrete and finite. We consider a target probability function for X which is factorized as follows:

$$p(x) = \frac{1}{Z} \prod\nolimits_{A \in \mathbb{A}} f_A(x_A) \ ,$$

where $f_A : E^A \mapsto (0, \infty)$ denotes a positive and non-constant finite function on $E^A = \prod_{i \in A} E_i$ and $\mathbb{A}$ is *factor set* which is a collection of non-empty subsets of

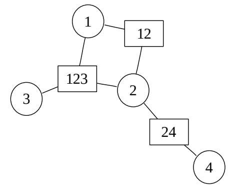

Fig. 1. A factor graph for $p(x_1, x_2, x_3, x_4) \propto f_{\{1,2\}}(x_1, x_2) f_{\{1,2,3\}}(x_1, x_2, x_3) f_{\{2,4\}}(x_2, x_4)$

$\{1, 2, \ldots, n\}$. Throughout this paper, Z stands for normalizing constants and are not always the same. A *factor graph* is an undirected graph associated with X and $\mathbb{A}$. A factor graph has two kinds of nodes; *variable nodes* and *factor nodes*. A variable node i and a factor node A are associated with the random variable X_i and the function f_A respectively. An edge is drawn between a variable node i and a factor node A if $i \in A$. We assume that no node is isolated. As an example, the factor graph corresponding to the probability function

$$p(x_1, \ldots, x_4) \propto f_A(x_1, x_2) f_B(x_1, x_2, x_3) f_C(x_2, x_4) ,$$

where $A = \{1, 2\}$, $B = \{1, 2, 3\}$ and $C = \{2, 4\}$ is shown in Fig. 1. Variable (resp. factor) nodes are represented by circles (resp. squares). The neighbors of a variable node i is $\{A \in \mathbb{A} : i \in A\}$ and is denoted by ∂i, and those of a factor node A is $\{i : i \in A\}$, i.e., A itself. Two kinds of *messages* are used in BP algorithm for factor graphs. They are defined reciprocally as follows:

$$n^{(t+1)}_{i \to A}(x_i) \equiv \frac{1}{Z} \prod_{C \sim \partial i \setminus \{A\}} m^{(t)}_{C \to i}(x_i) \tag{1}$$

$$m^{(t+1)}_{A \to i}(x_i) \equiv \frac{1}{Z} \sum_{x_{A \setminus \{i\}}} f_A(x_A) \prod_{j \in A} n^{(t)}_{j \to A}(x_j) \tag{2}$$

for each step $t = 0, 1, 2, \ldots$ and $x_i \in E_i$. We assume in this paper $n^{(0)}_{i \to A}(\cdot) = m^{(0)}_{A \to i}(\cdot) \equiv 1$ for convenience. Actually any initializations which are positive are possible. If messages $n^{(t)}_{i \to A}(\cdot)$ and $m^{(t)}_{A \to i}(\cdot)$ converge, the limits are denoted by $n_{i \to A}(\cdot)$ and $m_{A \to i}(\cdot)$. For these limit messages, the belief for each variable node i is defined by the normalized product:

$$b_i(x_i) = \frac{1}{Z} \prod_{A \in \partial i} m_{A \to i}(x_i), \quad x_i \in E_i . \tag{3}$$

Beliefs for a set of variable nodes can also be defined. In particular, the belief for variables associated with a factor node A (briefly, belief for a factor node A) is defined as follows:

$$b_A(x_A) = \frac{1}{Z} f_A(x_A) \prod_{i \in A} n_{i \to A}(x_i), \; x_A \in E^A . \tag{4}$$

If the factor graph has no loops, the beliefs will be exact marginal probabilities. If the factor graph has loops, the BP algorithm is still applicable and is reported to give often a good approximation. However, whether it converges or not becomes uncertain.

2.2 Computation Trees for BP Algorithm on Factor Graphs

In this section, we construct the computation trees for the BP on factor graphs corresponding to the BP update rules (1) and (2). In the construction of the computation tree of the BP for pair potentials, each node added to a computation tree as a message is updated is associated with a variable to be summed in the message update relation. It is similar for the factor graph case.

In order to construct the computation tree for the factor graph case, we first eliminate $n_{i\to A}(x_i)$ messages and rewrite the message update relations only by $m_{A\to i}(x_i)$ messages. Substituting $n_{i\to A}(x_i)$ messages for $m_{A\to i}(x_i)$ messages in (2), we have

$$m_{A\to i}^{(t+1)}(x_i) \propto \sum_{x_{A\setminus\{i\}}} f_A(x_A) \prod_{j\in A} \prod_{C\sim\partial j\setminus\{A\}} m_{C\to j}^{(t-1)}(x_j) \ .$$

According to this relation, we can construct the computation tree for the BP on factor graphs which can be summarized as the following proposition.

Proposition 1. *Let G be a factor graph and V_G and F_G be the set of variable nodes and factor nodes respectively. The computation tree T_k for a belief $b(x_k)$, $k \in V_G$, is constructed as follows:*

- *let $N_i = 0$, $i \neq k$, and $N_k = 1$. For convenience, let $T_k^{(0)} = \{k^{(0)}\}$ where $k^{(0)}$ is a copy of k.*
- *Let $\{A_1, A_2, \ldots\} = \partial k$, $C_i = A_i\setminus\{k\}$. The m-th node of $T_k^{(1)}$ is composed of $T_k^{(0)}$ and C_m in the following way. Let $C_m = \{i, j, \ldots\}$ and $N_i \leftarrow N_i+1$,$N_j \leftarrow N_j+1$,.... Add $A' = \{i^{(N_i)}, j^{(N_j)}, \ldots\}$ as a multi-state variable node to $T_k^{(0)}$ with the corresponding edge $(k^{(1)}, A')$ where $i^{(N_i)}, j^{(N_j)}, \ldots$ are the copies of $i, j, \ldots$ respectively. Let $S_{\{k^{(1)}\}A'} = \{k\}$.*
- *If the t-th computation tree $T_k^{(t)}$ is defined, the next computation tree $T_k^{(t+1)}$ is defined to be $T_k^{(t)}$ augmented by new nodes and edges repeating the following steps:*
  - *For each edge (A, A') of $T_k^{(t)}$ with $A \notin T_k^{(t-1)}$, let $A = \{i^{(N_i)}, j^{(N_j)}, \ldots\}$ and $S_A = \{s\}$.*
  - *For each element $i^{(N_i)} \in A$, let $\partial i = \{C_1, C_2, \ldots\}$ where $C_k \cup \{i\}, k = 1, 2, \ldots$ are elements of F_G except $\{s\} \cup A$. Add the k-th node and edge associated with $i^{(N_i)}$ as follows. Let $C_k = \{h, j, \ldots\}$ and $N_h \leftarrow N_h + 1, N_j \leftarrow N_j+1, \ldots$. Add the new copies $\{h^{(N_h)}, j^{(N_j)}, \ldots\}$ as a multi-state variable node A'' and the corresponding edge (A, A'') and let $S_{AA''} = \{i\}$*

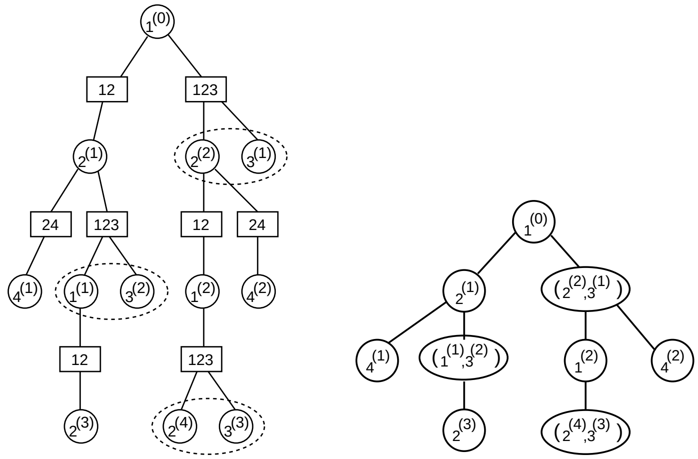

Fig. 2. Construction of the computation tree for the factor graph in Fig. 1 up to a few updates. The computation tree with factor nodes (left) and the computation tree (right). (i, j) indicates that the variable of the corresponding node has the compound state space $x_{(i,j)} \equiv (x_i, x_j) \in E_i \times E_j$

The potential functions for the corresponding Gibbs measure are defined such that $\phi_{AC} = -\log f_{(S_{AC})\cup C'}$ where C' is the set of index of C which are stripped of superscripts. For example, in Fig. 2,

$$\phi_{\{2^{(1)}\}\{1^{(1)},3^{(2)}\}}(x_1, x_2, x_3) = -\log f_{\{1,2,3\}}(x_1, x_2, x_3) \ ,$$
$$\phi_{\{2^{(2)},3^{(1)}\}\{4^{(2)}\}}(x_2, x_3, x_4) = -\log f_{\{2,4\}}(x_2, x_4) \ .$$

As is seen from the BP relation (5), one node added to the computation tree may be multi-state (i.e., a product of certain states) which is associated with $A\backslash\{i\}$ for some $i \in A$. That is, the state space is $E^{A\backslash\{i\}}$. In the following, we sometimes use Greek letters like α for expressing nodes on computation trees for factor graphs.

To ease the construction of the computation trees for factor graphs, it is helpful to draw the computation tree with factor nodes at first, which is similar to that of the BP for pair potentials where factor nodes are temporarily regarded as variable nodes. We give an example in Fig. 2

When a probability function has only two variable functions the computation tree for factor graph case is equivalent to the one for pair potential case discussed in [8]. In that case, in particular, $n_{i\to a}$ messages are equivalent to *boundary laws* of corresponding Gibbs measure, see [2].

It should be noted that unlike the pair potential case, the topology of the computation tree for a factor graph may depend on the choice of the root node. The topology of computation tree is related to the convergence property of message updates since it is sometimes related to the absence condition of phase transition. The reason for the dependency comes from the fact that, in the m message update, the set of variables to be summed depend on the direction of the m message to be updated if functions which have more than two variables are used.

Even if state spaces of variables in the computation tree may be different from those of the original graph, results in Tatikonda and Jordan [8] are also valid. That is, each belief $b_i(x_i)$ of (3) is the marginal probability of a Gibbs measure $\mu(X_{i^{(1)}} = x_i)$ on the corresponding computation tree and the absence of phase transition guarantees the convergence of LBP. In addition, we can show that beliefs $b_A(x_A)$ of (4) are also certain marginal probabilities of the Gibbs measure on the computation tree for factor graphs. We give the outline of the proof. For relevant concepts and references, see [7].

Corollary 1. *Let $T_k^{(t)}$ be the computation tree for a root node k after t message-update steps, and $\{Q_{ij}(x_i, x_j)\}_{i,j,E_i \times E_j}$ be the associated transfer matrices. $\ell_{ij}^{T_k^{(t)}} \in [0,\infty)^{E_i}$ denotes the boundary law for each adjacent sites i, $j \in T_k^{(t)}$ and the state space E_i, then*

$$m_{ik}^{(t)}(x_k) \propto \sum_{x_i \in E_i} \ell_{ik}^{T_k^{(t)}}(x_i) Q_{ik}(x_i, x_k)$$

for all neighboring node i of k and $x_k \in E_k$. If no phase transition occurs, there exists an unique boundary law $\ell_{ij}(x_i)$ such that $\ell_{ij}^{T_k^{(t)}}(x_i) \to \ell_{ij}(x_i)$ as $t \to \infty$. Therefore, using the limit boundary law,

$$m_{ik}^{(t)}(x_k) \to m_{ik}(x_k) \equiv \frac{1}{Z} \sum_{x_i} \ell_{ik}(x_i) Q_{ik}(x_i, x_k)$$

as $t \to \infty$.

Proof. See [7].

Proposition 2. *For each $i \in A \in \mathbb{A}$, there exists a node β adjacent to the root node $\{i^{(0)}\}$ in the computation tree for i such that $\beta = \{j : j$ is a copy of $A \backslash \{i\}\}$. Let $\alpha = \{\{i^{(0)}\}, \beta\}$ and $\partial\alpha = \partial i^{(0)} \cup \partial\beta \backslash \{i^{(0)}, \beta\}$. Then, if no phase transition occurs, the belief $b_A(x_A)$ defined by (4) is a marginal probability of the unique Gibbs measure on the computation tree.*

Proof. If no phase transition occurs there exists an unique Gibbs measure μ on the computation tree and

$$\begin{aligned}
\mu(x_\alpha) &= \sum_{x_{\partial\alpha}} \mu(x_{\alpha \cup \partial\alpha}) \propto \sum_{x_{\partial\alpha}} f_A(x_\alpha) \prod_{\kappa \in \partial\alpha} \ell_{\kappa\kappa'}(x_\kappa) f_{\kappa\kappa'}(x_\kappa, x_{\kappa'}) \\
&= f_A(x_\alpha) \prod_{\kappa \in \partial\alpha} \sum_{x_\kappa} \ell_{\kappa\kappa'}(x_\kappa) f_{\kappa\kappa'}(x_\kappa, x_{\kappa'}) \\
&\propto f_A(x_\alpha) \prod_{\kappa \in \partial\alpha} m_{\kappa\kappa'}(x_{\kappa'}) \ ,
\end{aligned} \tag{5}$$

where (5) comes from Gibbs measure theory and (5) comes from Corollary 1. Here κ' is $\{i^{(0)}\}$ or β, which is adjacent to κ. After some tiresome check of the correspondence between messages on the computation tree and original m and n messages on the factor graph, it can be shown that

$$\mu(x_A) \propto f_A(x_A) \prod_{i \in A} n_{i \to A}(x_i) \ .$$

Thus the proposition is complete. □

The above convergence criterion is based on the phase transition property of the associated Gibbs measure on the computation tree. In the factor graph case, since the topology of computation trees may depend on the choice of a root node, even the application of *Simon's condition* is not straightforward. We show a procedure how to check Simon's condition for the factor graph case.

Proposition 3. *For the LBP algorithm on factor graphs, the convergence condition based on Simon's condition can be checked as follows:*

STEP 0 : Let G be the index set of random variables and let $M = m = 0$. Go to STEP 1.
STEP 1 : If G is empty, break and return M. Otherwise, fix a factor i in G, let $\mathbb{A}_i = \{A \in \mathbb{A} : i \in A\}$ and go to STEP 2.
STEP 2 : If $\mathbb{A}_i$ is empty, $G \leftarrow G \setminus \{i\}$ and go to STEP 1. Otherwise, fix a factor A in $\mathbb{A}$, $m \leftarrow m + \delta(f_A)$ where $\delta(f_A)$ is the oscillation of the function f_A and go to STEP 3.
STEP 3 : If $A \setminus \{i\}$ is empty, $\mathbb{A}_i \leftarrow \mathbb{A}_i \setminus \{A\}$, $M \leftarrow \max\{M, m\}$, $m \leftarrow 0$ and go to STEP 2. Otherwise, fix a factor b in $A \setminus \{i\}$, let $\mathbb{A}_b = \{A \in \mathbb{A} : b \in A\}$ and go to STEP 4.
STEP 4 : If $\mathbb{A}_b \setminus \{A\}$ is empty, $A \leftarrow A \setminus \{b\}$ and go to STEP 3. Otherwise, fix a factor c in $\mathbb{A}_b \setminus \{A\}$, $m \leftarrow m + \delta(f_c)$, $\mathbb{A}_b \leftarrow \mathbb{A}_b \setminus \{c\}$ and go to STEP 4.

If the $M < 2$, the LBP algorithm converges. □

3 Comparison of Three Convergence Criteria

In this section, we compare LBP convergence criteria due to two approaches, including the one derived from Dobrushin's condition, in order to see their effectiveness. As shown in the previous chapter, Gibbs measure approach is still useful for LBP with general potentials through factor graphs. On the other hand, it is generally difficult to characterize the phase transition region (i.e., a certain parameter region) precisely. One remarkable exception is *Ising models on Cayley trees*; Its complete characterization are known. It is noted that a Cayley tree is a certain computation tree of a complete graph. For this reason, we use Ising models on a complete graph as a target probabilistic network for comparing two approaches, nevertheless Ising models have only pair potentials so that it does not have to be expressed by factor graphs any longer.

For given parameters J and $h \in \mathbb{R}$, the Ising model is defined by

$$p(x) \propto \exp\Big(h \sum_i x_i + J \sum_{i \sim j} x_i x_j\Big)$$

where $x_i \in \{-1, 1\}$ for all $i \in G$ and $\sim$ means the neighbor relation. For a complete graph, the corresponding computation tree is called a Cayley tree or *Bethe lattice*. Let $d+2$ be the number of vertex of the complete graph. We assume $d \geq 2$. In fact, for Ising models, the convergence conditions derived from Ihler's and Mooij's approaches are same such as

$$d \tanh |J| < 1 \ . \tag{6}$$

This is the best criterion at present. Their approaches do not rely on Gibbs measures and may be valid even when phase transition occurs. The criterion based on Dobrushin's condition becomes

$$\frac{(d+1)\sinh(2|J|)}{g(h, J) + \cosh(2J)} < 1 \ , \tag{7}$$

where

$$g(x, y) = \min_{z_i \in \{-1,1\}, i=1,\ldots,d} \cosh 2\Big(x + y \sum_{i=1}^{d} z_i\Big) \ .$$

Dobrushin's condition is not so simple to verify. We give a derivation of this criterion at length in Appendix.

For Ising models on Cayley trees, there is the following complete condition for the lack or existence of phase transitions is known, see [2] for details. Let $J(d) = \operatorname{arccoth}(d) = \frac{1}{2} \log \frac{d+1}{d-1}$, and

$$h(J, d) \equiv \begin{cases} 0 & \text{if } |J| \leq J(d) \ , \\ d \operatorname{arctanh}\Big(\frac{dw-1}{d/w-1}\Big)^{1/2} - \operatorname{arctanh}\Big(\frac{d-1/w}{d-w}\Big)^{1/2} & \text{if } J > J(d) \ , \\ d \operatorname{arctanh}\left(\frac{dw-1}{d/w-1}\right)^{1/2} + \operatorname{arctanh}\left(\frac{d-1/w}{d-w}\right)^{1/2} & \text{if } J < -J(d) \ , \end{cases}$$

where $w = \tanh |J|$. The phase transition region (the set of parameters where the phase transition occurs) consists of the *ferromagnetic (antiferromagnetic) phase transition region* F (AF) defined by

$$\begin{aligned} (F) \ d > 1, \ J > J(d), \ |h| \leq h(J, d) \ , \\ (AF) \ d > 1, \ J < -J(d), \ |h| < h(J, d) \ . \end{aligned}$$

Leftmost figure in Fig. 3 shows F and AF regions for $d = 2$. The region AF is open. The region F includes its boundary except for the singular point $(h, J) = (0, J(d))$. The region other than F and AF is the LBP convergence region.

In Fig. 3, we also give other two LBP convergence regions derived from Dobrushin's and Ihler's and Mooij's. Also we show these regions together.

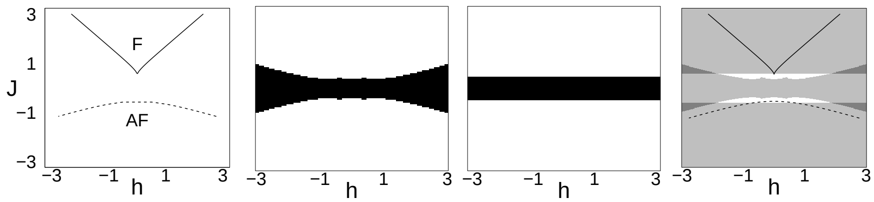

Fig. 3. Convergence regions derived from the complete characterization and Dobrushin's, and Ihler's (Mooij's) conditions (left to right). The region other than F and AF is that of the complete characterization, and black regions are those of Dobrushin's and Ihler's. The rightmost figure shows these regions together with the boundary curves of F and AF. Difference sets of Dobrushin's and Ihler's (Mooij's) regions are represented in white and dark gray

It should be noted that since Dobrushin's condition is a sufficient condition of absence of phase transition, the region derived form Dobrushin's condition is naturally included in that of the complete characterization. Therefore, we look at other relationships here.

Ihler's (Mooij's) region is completely included in that of the complete characterization. This is proved by the fact $d \tanh J(d) = 1$. As a result, the condition obtained from the complete characterization is stronger than that of Ihler's (Mooij's). On the other hand, neither Dobrushin's nor Ihler's (Mooij's) region is a subset of the other. Nevertheless, when $|h|$ is sufficiently large, Dobrushin's region always includes that of Ihler's (Mooij's), that is, Dobrushin's condition is stronger than Ihler's when the influence of one-body potentials is sufficiently large in this case.

4 Conclusion

In this paper, we show two applications of Gibbs measure theory to LBP algorithm. We first show this theory can be applied directly to probability functions with general potentials through factor graphs. Second, we show the usefulness of the application of Gibbs measure theory in the sense that an elaborate application of this theory gives a better result than the best present result in a special case. These two results are not so prominent but sure to encourage the use of Gibbs measure theory in this area.

References

1. Frey, B. J.: Graphical Models for Pattern Classification, Data Compression and Channel Coding, MIT press, Cambridge (1998).

2. Georgii, H. -O.: Gibbs Measures and Phase Transitions, Walter de Gruyter, Berlin · New York (1988).
3. Ihler, A. T., Fisher, J. W., Willsky, A. S.: Loopy Belief Propagation: Convergence and Effects of Message Errors. Journal of Machine Learning Research **6**. Cambridge, MA: MIT press (2005) 905-936.
4. McEliece, R. J., MacKay, D. J. C., Cheng, J. F.: Turbo Decoding as an Instance of Pearl's "Belief Propagation" Algorithm, IEEE Journal on Selected Areas in Communication, **16(2)** Springer Verlag, New York, Berlin, Heidelberg (1998) 140-152.
5. Mooij, J. M., Kappen, H. J.: Sufficient conditions for convergence of Loopy Belief Propagation, Proc. of 21st Conf. on Unc. in Art. Int., Edinburgh (2005) 396-403.
6. Murphy, K. P., Weiss, Y., Jordan, M. I.: Loopy belief propagation for approximate inference: an empirical study, Proc. of the 15th Conf. on Unc. in Art. Int., Morgan Kaufmann, San Francisco (1999) 467-475.
7. Taga, N., Mase, S.: On the Convergence of Loopy Belief Propagation Algorithm for Different Update Rules. IEICE transactions on Fundamentals of Electronics, Communications and Computer Sciences, Vol.E89-(2). Tokyo (2006) 575-582.
8. Tatikonda, S. C., Jordan, M. I.: Loopy Belief Propagation and Gibbs Measures, Proc. of the 18th Conf. on Unc. in Art. Int., Morgan Kaufmann, San Francisco (2002) 493-500.
9. Weiss, Y.: Correctness of Local Probability Propagation in Graphical Models with Loops, Neur. Comp., vol.12, MIT press. Cambridge (2000) 1-41.

Appendix

In this appendix, we show how to check Dobrushin's condition for Ising models on the Cayley tree of degree d or the complete graph of $d+2$ vertices. Let S be the vertex set of the Cayley tree and E_i be $\{-1,1\}$ for $i \in S$. Let $\Omega = \prod_{i \in S} E_i$ be the configuration space, $(\Omega, \mathcal{F})$ be the measurable space with the Borel set $\mathcal{F}$ of Ω and γ be a *specification* on $(\Omega, \mathcal{F})$. γ satisfies Dobrushin's condition if

$$c(\gamma) \equiv \sup_{i \in S} \sum_{j \in S} C_{ij}(\gamma) < 1 ,$$

where

$$C_{ij}(\gamma) = \sup_{\zeta,\eta \in \Omega,\ \zeta_{S\setminus\{i\}} = \eta_{S\setminus\{i\}}} \| \gamma_i^0(\cdot \mid \zeta) - \gamma_i^0(\cdot \mid \eta) \| \tag{8}$$

with the norm $\|f(\cdot)\| = \max_{A \in \mathcal{F}} \left| \sum_{\zeta \in A} f(\zeta) \right|$ for a real function f on Ω. In the Ising potential case, we have

$$\gamma_i^0(x_i \mid \zeta) = \frac{1}{Z(\zeta)} \exp\left[x_i \Big(h + J \sum_{j \in \partial i} \zeta_j \Big) \right] ,$$

where $Z(\zeta) = 2\cosh(h + J \sum_{j \in \partial i} \zeta_j)$ and ∂i denotes the set of neighbors of $i \in S$. Note that $C_{ij}(\cdot) = 0$ when $j \notin \partial i$ (for specifications with nearest neighbor potentials). Then, taking the supremum in the right side of eq. (8), we can restrict ourselves to consider configurations ζ, η such that for some $j \in \partial i$, $\eta_j = -\zeta_j$ and $\eta_k = \zeta_k$ for $k \neq j$. Now $\gamma_i^0(\emptyset \mid \zeta) = 0$ and $\gamma_i^0(E_i \mid \zeta) = 1$ for any configuration ζ, hence we have

$$\gamma_i^0(A \mid \zeta) - \gamma_i^0(A \mid \eta) = \begin{cases} \sinh(2J)/Z'_{ij}(\zeta) & \text{if } A = \{1\} \ , \\ -\sinh(2J)/Z'_{ij}(\zeta) & \text{if } A = \{-1\} \ , \\ 0 & \text{otherwise} \ , \end{cases}$$

where $Z'_{ij}(\zeta) = \cosh 2(h + J \sum_{k \in \partial i \setminus \{j\}} \zeta_k) + \cosh(2J)$ for configurations ζ, η such that for some $j \in \partial i$, $\eta_j = -\zeta_j$ and $\eta_k = \zeta_k$ for $k \neq j$. Therefore

$$\|\gamma_i^0(\cdot \mid \zeta) - \gamma_i^0(\cdot \mid \eta)\| = \frac{1}{Z'_{ij}(\zeta)} \sinh(2|J|) \ ,$$

For the Cayley tree of degree d, the number of neighbors for each vertex is $d+1$. Therefore we have

$$C_{ij}(\gamma) = \max_{\zeta \in \Omega} \frac{1}{Z'_{ij}(\zeta)} \sinh(2|J|) = \frac{\sinh(2|J|)}{g(h,J) + \cosh(2J)}$$

where

$$g(x,y) = \min_{z_i \in \{-1,1\},\ i=1,\ldots,d} \cosh 2\left(x + y \sum_{i=1}^{d} z_i\right) \ .$$

That is, $C_{ij}(\gamma)$ is independent of i, j and it follows

$$c(\gamma) = \frac{(d+1)\sinh(2|J|)}{g(h,J) + \cosh(2J)} \ .$$

Using this, we can check Dobrushin's condition easily.