On the convergence of loopy belief propagation algorithm under several update rules

Nobuyuki Taga
Tokyo Institute of Technology,
O-okayama 2-12-1, Meguro-ku, Tokyo 152-8552, Japan,
E-mail: Nobuyuki.Taga@is.titech.ac.jp

Shigeru Mase
Tokyo Institute of Technology,
O-okayama 2-12-1, Meguro-ku, Tokyo 152-8552, Japan,
E-mail mase@is.titech.ac.jp

***Abstract*— The belief propagation (BP) algorithm is a tool with which one can calculate beliefs, marginal probabilities, of stochastic networks without loops (e.g., Bayesian networks) in a time proportional to the number of nodes. For networks with loops, it may not converge and, even if it converges, beliefs may not be equal to exact marginal probabilities although its application is known to give remarkably good results such as in the coding theory. Tatikonda and Jordan show a theoretical result on the convergence of the algorithm for stochastic networks with loops in terms of the theory of Markov random fields on trees and give a sufficient condition of the convergence of the algorithm. In this paper, we show a new aspect of convergence property of BP algorithm. We discuss the "impatient" update rule as well as the "lazy" update rule discussed in Tatikonda and Jordan. Then we show for simple models that the impatient update rule converges faster than the lazy one and if the algorithm converge, the beliefs obtained by impatient update rule give good approximations of exact marginal probabilities.**

I. Introduction

The belief propagation (BP) algorithm is a tool with which one can calculate beliefs, marginal probabilities, of stochastic networks without loops (e.g., Bayesian networks) in a time proportional to the number of nodes. It has the origin in the probabilistic expert system theory proposed by Pearl *et al.*, see [7]. Similar algorithms appear in several applications such as Viterbi algorithm in hidden Markov models, iterative algorithms for Gallager codes and turbocodes, Kalman filter and the transfer-matrix approach in physics.

Lauritzen [1] formulated it as an algorithm on derived networks called junction trees. On the other hand, Yedidia [12] showed that it can be formulated as the algorithm of minimizing Bethe free energy of Gibbs distributions with pairwise potentials on trees.

As such, it can be formally applicable to networks with loops. However, if networks have loops, the algorithm may not converge and, even if it converges, beliefs may not equal to exact marginal probabilities. Nevertheless, applications of the BP algorithm for networks with loops have been reported to be remarkably useful such as in the coding theory [2] [5][6]. We call BP algorithm for networks with loops $loopy\ belief\ propagation$ algorithm.

Weiss [10] discussed the BP algorithm on networks with a single loop and Weiss and Freeman [11] discussed the BP algorithm on Gaussian networks with loops. A basic idea of Weiss is the fact that the calculation of BP algorithm for networks with loops is equivalent to the one on corresponding infinite trees called unwrapped networks.

Tatikonda and Jordan [9] pursued his idea and formulated the convergence problem as the one of Markov random fields on unwrapped networks. Surprisingly enough, an essentially equivalent (finer in a sense) problem is already discussed in the theory of Markov random fields on trees in order to study their phase transition property, see [3]. They showed a relation between the convergence of BP algorithm and phase transition phenomena on unwrapped networks in a proceeding paper.

In this paper, we show that the convergence of BP algorithm for networks with loops depends on update rules and discuss the "impatient" update rules, which is used in practice, as well as the "lazy" one discussed in Weiss or Tatikonda and Jordan. We give numerical experiments to check the difference between the convergence of both algorithms for simple models. It is shown that the convergence with the impatient update rule is faster than the one with the lazy update rule. We also show some other results concerning the property of BP algorithm with impatient update rule.

Models used in numerical experiments are known to show phase transitions and a fairly complete condition on the existence and the absence of phase transitions is known. For parameters corresponding to the absence of phase transitions, the BP algorithm converges as theoretical result shows for both update rules. In the ferromagnetic phase transition region, the BP algorithm still converges. However, in the anti-ferromagnetic phase transition region, while the BP algorithm with the lazy update rule did not converge for all parameters, the one with the impatient update rule did converge around the boundary. It is also shown that, if messages converge, the beliefs give good approximations of marginal probabilities. The present result suggests that the BP algorithm is valid in broader region , as suggested in the literature.

We give a review of BP algorithm and introduce the theory of Markov random fields in Section II and III. In Section IV, we state the relation between unwrapped networks and update rules. In Section V, results of numerical experiments are shown. In Section VI, we give a conclusion and a remark.

II. BP algorithm and unwrapped networks

There are several existing formulations of the BP algorithm. The one used in this paper is as follows. Let X be a connected and undirected finite network. A random variable X_i is associated with each $i \in X$ and y_i is its observation. The state space E_i of X_i is finite. Some y_i may be missing. We consider a probability function on X of the form

$$p(x \mid y) = \frac{1}{Z} \prod_{i \sim j} \phi_{ij}(x_i, x_j) \prod_{i \in X} \phi_i(x_i, y_i), \tag{1}$$

where $\sim$ denotes the neighborhood relationship, and the first product extends over all neighboring nodes (i, j). We call X a $stochastic\ network$ with the joint distribution p. Throughout this paper, Z stands for normalizing constants and are not always the same. Usually, the existence of a data y_i restricts the state space E_i to $\{y_i\}$ effectively. We will adopt this convention and, further, suppress the dependencies of ϕ_i's on $\{y_i\}$. Therefore, it takes the form

$$p(x) = \frac{1}{Z} \prod_{i \sim j} \phi_{ij}(x_i, x_j) \prod_{i \in X} \phi_i(x_i). \tag{2}$$

It is the basic assumption of this paper that $\phi_{ij}(\cdot, \cdot)$ and $\phi_i(\cdot)$ are all positive.

For each neighboring node (i, j) and each state $x_j \in E_j$, we consider the *message* $m_{ij}^{(n)}(x_j)$, $n = 1, 2, \ldots$. These messages obey the following update rule called the *Belief Propagation* (BP):

$$m_{ij}^{(n+1)}(x_j) = \frac{1}{Z} \sum_{x_i \in E_i} \phi_{ij}(x_i, x_j) \phi_i(x_i) \prod_{k \in \partial i \setminus \{j\}} m_{ki}^{(n)}(x_i), \tag{3}$$

where ∂i denotes the set of all neighboring nodes of node i. We call this update relation "lazy", which is easier to analyze. In practice, one use the update rule which always utilizes the "most recently updated messages", that is, some of $m_{ki}^{(n)}(x_i)$ are $m_{ki}^{(n+1)}$ in the right side of 3. We call this update relation is "impatient" .

In the following, $|A|$ for a set A means its cardinality. All messages are initialized as $m_{ij}^{(0)}(x_j) \equiv 1/|E_j|$. If messages $m_{ij}^{(n)}(x_j)$ converges, its limit is denoted by $m_{ij}(x_j)$. They satisfy the relation:

$$m_{ij}(x_j) = \frac{1}{Z} \sum_{x_i \in E_i} \phi_{ij}(x_i, x_j) \phi_i(x_i) \prod_{k \in \partial i \setminus \{j\}} m_{ki}(x_i). \tag{4}$$

For these messages, the *belief* for each node i is the normalized product

$$b_i(x_i) = \frac{1}{Z} \phi_i(x_i) \prod_{k \in \partial i} m_{ki}(x_i), \quad x_i \in E_i. \tag{5}$$

If the stochastic network has no loops, i.e., tree-like, it is known that all the messages $\{m_{ij}^{(n)}(x_j)\}$ converge after a finite number of BP updates and that the belief $b_i(\cdot)$ is equal to the marginal probability $\boldsymbol{P}\{X_i = \cdot\}$ for each $i \in X$, see [4]. On the other hand, for stochastic networks with loops, messages may not converge and, if they do converge, beliefs may not be equal to marginal probabilities. An example for Cayley trees will be given later. In order to study the BP algorithm with the lazy update rule for a network X with loops, Weiss [10] introduced the concept of *unwrapped networks* (*computation trees* in [9]), which are associating infinite trees T_k, $k \in X$. T_k is the limit of increasing finite trees $\{T_k^{(n)}\}$, $n = 1, 2, \ldots$, defined as follows, see Fig. 1.

1) Let $N_i = 0$, $i \neq k$, and $N_k = 1$. For convenience, let $T_k^{(0)} = \{k^{(1)}\}$ where $k^{(1)}$ is a copy of k.
2) Let $\{i, j, \ldots\} = \partial k$, $N_i = N_j = \cdots = 1$ and $i^{(1)}, j^{(1)}, \ldots$ be copies of $i, j, \ldots$ respectively. The first unwrapped network $T_k^{(1)}$ consists of nodes $k^{(1)}, i^{(1)}, j^{(1)}, \ldots$ and corresponding edges $(k^{(1)}, i^{(1)}), (k^{(1)}, j^{(1)}), \ldots$.
3) If the n-th unwrapped network $T_k^{(n)}$ is defined, the next unwrapped network $T_k^{(n+1)}$ is defined to be $T_k^{(n)}$ augmented by new nodes and edges repeating the following steps:
   a) For each edge $(r^{(\ell)}, s^{(m)})$ of $T_k^{(n)}$ with $r^{(\ell)} \notin T_k^{(n-1)}$, let $i, j, \ldots$ be the nodes $\partial r \setminus \{s\}$ (if non-empty).
   b) Let $N_i \leftarrow N_i + 1, N_j \leftarrow N_j + 1, \ldots$ and $i^{(N_i)}, j^{(N_j)}, \ldots$ be new copies of $i, j, \ldots$ respectively. Add new nodes $i^{(N_i)}, j^{(N_j)}, \ldots$ and corresponding edges $(i^{(N_i)}, r^{(\ell)})$, $(j^{(N_j)}, r^{(\ell)})$, ... to $T_k^{(n)}$.

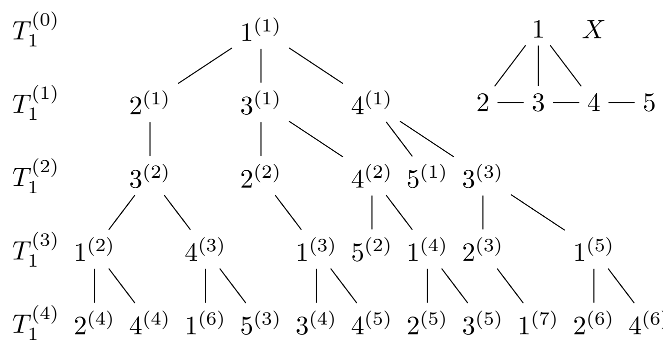

Fig. 1. A network X with loops (right) and its unwrapped networks $T_1^{(n)}$ up to the fourth (left).

The state space E_i is associated with each node $i^{(n)} \in T_k$ and let $\phi_{i^{(n)}j^{(m)}} = \phi_{ij}$ and $\phi_{i^{(n)}} = \phi_i$. If X has no loops, T_k is the same as X except labeling of nodes. It is easily seen that the message $m_{jk}^{(n)}(x_k)$ which is the result of the n-th BP update with the lazy update rule on X starting from k is equal to $m_{j^{(1)}k^{(1)}}^{T_k^{(n)}}(x_k)$, the result of the n-th BP update of messages performed on $T_k^{(n)}$, that is, on T_k starting from $k^{(1)}$. Therefore, the limiting message heading for k, if exists, is the same for both X and T_k. Tatikonda and Jordan [9] formulated the convergence of messages for unwrapped networks as the one of Markov random fields introduced in the next section.

III. Markov random fields and Markov chains on a tree

Let T be a finite or infinite tree. A finite state space E_i is associated with each $i \in T$. A *configuration* $x = \{x_i\}$ is an element of $E^T = \prod_{i \in T} E_i$. Its restriction to a subset

Λ is denoted by x_Λ. Let B be the set of *bonds* (i,j), that is, pairs of neighboring nodes. $\mathcal{F}$ is the set of non-empty finite subtrees of T. For each $(i,j) \in B$, let ij denote the associated oriented edge which points from i to j and $\overrightarrow{B}$ denote the set of all oriented edges. Any $i,j \in T$ is combined by a (unique) path $((i,k_1),(k_1,k_2),\ldots,(k_n,j))$ of bonds where $i,k_1,\ldots,k_n,j$ are all different. The associated sequence $(ik_1,k_1k_2,\ldots,k_nj)$ is called an oriented path from i to j. For each bond $b=(i,j)$, a two-body potential function $\Phi_b = \Phi_{ij}$ is a finite-valued function on $E_i \times E_j$. It is symmetric $\Phi_{ij}(x_i,x_j) = \Phi_{ji}(x_j,x_i)$. One-body potential functions Φ_i, $i \in T$, is a finite-valued function on E_i. For each $\Lambda \in \mathcal{F}$, let $B_\Lambda = \{b \in B : b \subset \Lambda\}$ and $\overrightarrow{B}_\Lambda$ be the corresponding oriented edges. Let $E^\Lambda = \prod_{i\in\Lambda} E_i$ and $\partial\Lambda = (\cup_{i\in\Lambda}\partial i)\backslash\Lambda$. The space of configurations E^T is equipped with the discrete topology and let $\mathcal{B}$ be its (standard) Borel σ-algebra. $\mathcal{B}_\Lambda$, $\Lambda \in \mathcal{F}$, is the (standard) Borel sub-σ-algebra on E^Λ.

A *Gibbs specification* is a system $\{\mu_\Lambda(\cdot \mid \xi) : \Lambda \in \mathcal{F}, \xi \in E^T\}$ of probability measures defined by

$$\mu_\Lambda(x \mid \xi) = \frac{1}{Z}\exp\left\{-\sum_{(i,j)\in B_\Lambda}\Phi_{ij}(x_i,x_j) - \sum_{i\in\Lambda}\Phi_i(x_i) - \sum_{i\in\Lambda, j\in\partial\Lambda}\Phi_{ij}(x_i,\xi_j)\right\}$$

for all $\Lambda \in \mathcal{F}$, where $x \in E^\Lambda$ and $Z = Z_{\Lambda,\xi}$ is the normalizing constant called the *partition function*. ξ is called a *boundary condition*. Since $\mu_\Lambda(x \mid \xi)$ is dependent on ξ only through $\xi_{\partial\Lambda}$, it is, in particular, called a *Markov specification*. We also need following specifications without boundary condition:

$$\mu_\Lambda(x) = \frac{1}{Z}\exp\left\{-\sum_{(i,j)\in B_\Lambda}\Phi_{ij}(x_i,x_j) - \sum_{i\in\Lambda}\Phi_i(x_i)\right\}.$$

A probability measure μ on $(E^T,\mathcal{B})$ is called a *Gibbs random field* with potentials $\Phi = \{\Phi_{ij},\Phi_i\}$ if it satisfies the following DLR (*Dobrushin-Lanford-Ruelle*) equations:

$$\mu(x \mid \mathcal{B}_{T\backslash\Lambda})(\xi) = \mu_\Lambda(x \mid \xi), \quad \xi \in E^{\partial\Lambda},$$

for all $\Lambda \in \mathcal{F}$, where $x \in E^\Lambda$ is canonically embedded into E^T as $x \times E^{T\backslash\Lambda}$. Since $\mu(x \mid \mathcal{B}_{T\backslash\Lambda}) = \mu(x \mid \mathcal{B}_{\partial\Lambda})$, such μ is called a *Markov random field*. Let $\mathcal{G}_\Phi$ denote the set of all Markov random fields for potentials Φ. *Transition matrices* P_{ij} are defined as

$$P_{ij}(\xi_i,\xi_j) = \mu(\sigma_j = \xi_j \mid \sigma_i = \xi_i) \quad \mu\text{-a.s.}$$

for $ij \in \overrightarrow{B}$, $(\xi_i,\xi_j) \in E_i \times E_j$. *Transfer matrices* are defined as

$$\begin{aligned} Q_{ij}(\xi_i,\xi_j) &= Q_{ji}(\xi_j,\xi_i) = Q_b(\xi_b) \\ &= e^{-\Phi_{ij}(\xi_i,\xi_j) - |\partial i|^{-1}\Phi_{\{i\}}(\xi_i) - |\partial j|^{-1}\Phi_{\{j\}}(\xi_j)} \end{aligned}$$

for $b=(i,j) \in B$, $(\xi_i,\xi_j) \in E_i \times E_j$. A family $\{\ell_{ij} : ij \in \overrightarrow{B}\}$ of vectors $\ell_{ij} \in (0,1)^{E_i}$ is called a *boundary law* for $\{Q_{ij} : ij \in \overrightarrow{B}\}$ if, for each $ij \in \overrightarrow{B}$, there is a number $c_{ij} > 0$ such that

$$\begin{aligned} \ell_{ij}(x_i) &= c_{ij}\prod_{k\in\partial i\backslash\{j\}} \ell_{ki}Q_{ki}(x_i) \\ &= c_{ij}\prod_{k\in\partial i\backslash\{j\}}\sum_{x_k}\ell_{ki}(x_k)Q_{ki}(x_k,x_i). \end{aligned}$$

In the theory of Markov random fields, it is known that there exist multiple Markov random fields for a Gibbs specification in some cases. In that case, it is said that a *phase transition* occurs. When a phase transition occurs, it is also shown that there exist multiple boundary laws.

Tatikonda and Jordan [9] showed a relation between boundary laws and messages for unwrapped networks. Therefore, they related the convergence of the BP algorithm to the lack of phase transitions for models on the corresponding unwrapped networks. In addition, if message updates converge, they showed the beliefs give the marginal probabilities of the Markov random fields on the corresponding unwrapped networks, which are not original networks.

IV. Update rules and expansion of unwrapped networks

So far, only the lazy update rule is discussed. However, the formulation as in [9] can be also applied to the impatient update rule almost without modifications. For example, let consider the simplest network X with a loop in Fig. 2. In this example, the impatient update rule becomes as follows:

$$\begin{aligned} m_{12}^{(n+1)}(x_2) &\propto \sum_{x_1} Q_{12}(x_1,x_2)m_{31}^{(n)}(x_1) \\ m_{21}^{(n+1)}(x_1) &\propto \sum_{x_2} Q_{21}(x_2,x_1)m_{32}^{(n)}(x_2) \\ m_{13}^{(n+1)}(x_3) &\propto \sum_{x_1} Q_{13}(x_1,x_3)m_{21}^{(n+1)}(x_1) \\ m_{31}^{(n+1)}(x_1) &\propto \sum_{x_3} Q_{31}(x_3,x_1)m_{23}^{(n)}(x_3) \\ m_{23}^{(n+1)}(x_3) &\propto \sum_{x_2} Q_{23}(x_2,x_3)m_{12}^{(n+1)}(x_2) \\ m_{32}^{(n+1)}(x_2) &\propto \sum_{x_3} Q_{32}(x_3,x_2)m_{13}^{(n+1)}(x_3). \end{aligned}$$

The unwrapped networks starting from the node 1 associated with both update rules are shown in Fig. 2.

It is seen that the tree associated with the impatient update rule is essentially the same with the one associated with the lazy update rule and therefore the results in [9] are also valid.

Actually, the expansion pattern of unwrapped networks is dependent on update rules. An expansion pattern $\{T^{(n)}\}$ corresponds to a subsequence of $\{\mu_\Lambda\}$ of the Gibbs specification. Note that there is a possibility that subsequences may converge even though a phase transition occurs and then the convergence property may depend on the update rule. Moreover, if messages converge, from Tatikonda and Jordan [9] and the theory of Markov random fields, beliefs are the marginal probability of an "extremal" Markov random field.

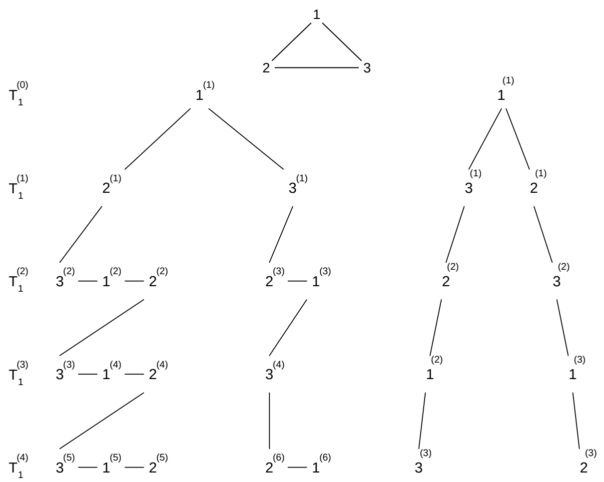

Fig. 2. A network X with a loop (top) and its unwrapped networks corresponding to the impatient update rule (left) and to the lazy update rule (right).

On the other hand, the spread speed of unwrapped networks with the impatient update rule is generally faster than the one with the lazy update. According to the fact that under the sufficient condition introduced in Tatikonda and Jordan [9], the convergence rate depends on the spread speed of unwrapped networks. Therefore, it is expected that convergence rate of algorithm with the impatient update rule is likely faster than the one with the lazy update.

V. Numerical experiments

We reviewed convergence of the calculation of BP algorithm is related to the phase transition phenomenon for unwrapped networks corresponding to the original stochastic network. Nevertheless, when there exists a phase transition, its convergence is still unclear.

In this section, We checked numerically the convergence and accuracy of beliefs between the impatient update rule and the lazy update rule for Ising models on a Cayley tree, which is known to show phase transitions for certain parameters.

An infinite tree is called the *Cayley tree* $CT(d)$ of degree d if $|\partial i| = d+1$ for every node i. It is the unwrapped network of some complete graph. Fig. 3 shows a complete graph with 4 nodes and the corresponding unwrapped network $CT(2)$.

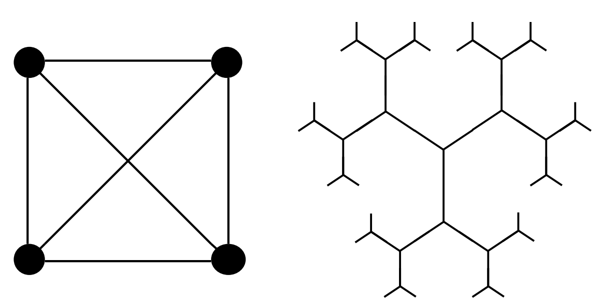

Fig. 3. Complete graph of 4 nodes (left) and Cayley tree of degree 2 (right).

Let $E_i = \{-1, 1\}$ for $i \in CT(d)$. We consider following *Ising potentials* on $CT(d)$

$$\Phi_A^{J,h}(x) = \begin{cases} -Jx_ix_j & \text{if } A = \{i,j\}, \\ -hx_i & \text{if } A = \{i\}, \\ 0 & \text{otherwise} \end{cases}$$

and let μ_{Jh} be a corresponding Markov random field. If $J > 0$ (resp. $J < 0$), it is called *ferromagnetic* (resp. *antiferromagnitic*). These models are exceptional in the sense that the nearly complete condition whether phase transitions occur or not is known, see [3]. Fig. 4 shows two phase transition regions.

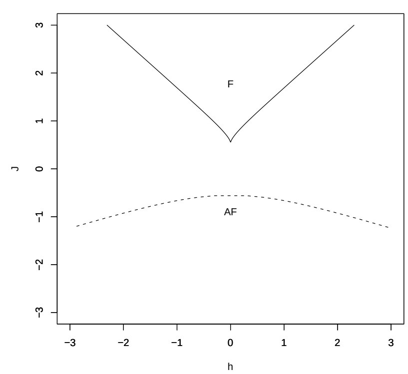

Fig. 4. The phase transition region for the Ising model on CT(2).

The region AF is open. The region F includes its boundary except for the singular point at $J = J(d) \equiv \frac{1}{2}\log\frac{d+1}{d-1}$. Note that Ising models are symmetric with respect to the transform $(x, h) \mapsto (-x, -h)$. In particular, if $h = 0$, they are symmetric with respect to the transform $x \mapsto -x$ and all the one-node marginal probabilities are uniform ($= 1/2$).

Let X be a finite graph with loops and let its unwrapped network be the Cayley tree $CT(2)$. The probability on X is defined by

$$p(x) = \frac{1}{Z}\exp\left[\sum_{i\sim j}\left\{Jx_ix_j + \left(\frac{h}{3}\right)(x_i + x_j)\right\}\right].$$

For each $(J, h) = (\frac{a}{10}, \frac{b}{10})$, $a, b = 0, \pm 1, \ldots, \pm 30$, iteration times till convergence up to 10000 were calculated. We used $1/2$ for initial messages. Message were assumed to converge if $\max_{x\in\{-1,1\}}\sum_{i\sim j}|m_{ij}^{(n+1)}(x) - m_{ij}^{(n)}(x)| \le 10^{-6}$ for some n.

Fig. 5 shows the common logarithm of the iteration number. For $h = 0$, two iterations were enough. For $J \ge 0$, all calculations with both update rules stopped within 50 times. In particular, it includes F and shows that the absence of phase transitions are by no means necessary for the BP convergence. It is seen that the number of iteration increases as (J, h) come close to the singular point $(J(2), 0)$. All calculations with both update rules for $J < 0$ outside AF also stopped. Again the iteration number increases as (J, h) come close to the boundary of AF. It is shown that for the lazy update rule, the iteration numbers exceeded 1500 for some parameters near AF but for the impatient update rule, iterations up to 30 were sufficient. Inside AF except for $h = 0$, all calculations for the lazy update rule exceeded 10000 and did not seem to converge. However, it is shown that, for some parameters around the boundary of AF, the calculation with the impatient update rule did converge within 100 iterations, while the lazy update rules never converged. It is said that the convergence with the impatient update rule is always faster than the one with the lazy update rule.

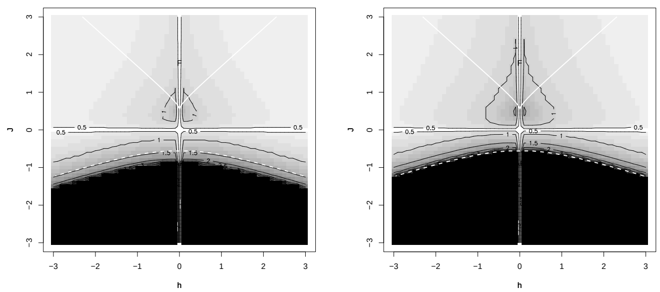

Fig. 5. Common logarithms of numbers of iterations till convergence up to 10000. Images and contours. The impatient update rule case (left) and the lazy update rule case (right). Boundaries of F and AF are shown by white.

We also examined the accuracy of beliefs. We dealt only with the case of the impatient update rule. Let p_{Jh} be the correct marginal probability for $x_i = 1$ and b_{Jh} be the corresponding belief for each (J, h). Fig. 6 shows p_{Jh} for each (J, h).

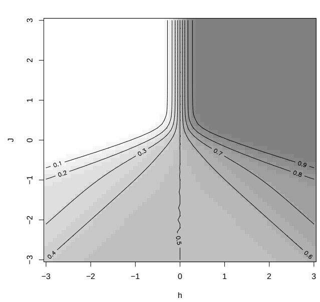

Fig. 6. True marginal probabilities $p_{Jh}(x_1 = 1)$.

Fig. 7 shows absolute and relative errors. In both figures, the white part in AF is the region where the iteration exceeded 10000 times. It is seen that, beliefs are close to true marginals if the BP algorithm converge. Also, it is seen that relative errors increase as (J, h) come close to and into AF and F and are large in F.

In the literature, it is sometimes stressed that, although beliefs may differ from true marginals, states with the highest value coincide. In fact, this is true in the present experiment for all cases where BP iterations converged.

What is a difference between convergent and non-convergent case? To see this, we examined the behavior of the value of belief as message updates perform. Fig. 8 shows the behavior of beliefs for parameters $J = -0.97, -0.99, -0.993, -1, -1.2$ and -1.4 with $h = 1, d = 2$ up to 1000 iterations. For $J \leq -1$, the calculation did not converge within 10000 iterations. It is shown that the convergence rate increases as the parameter come close

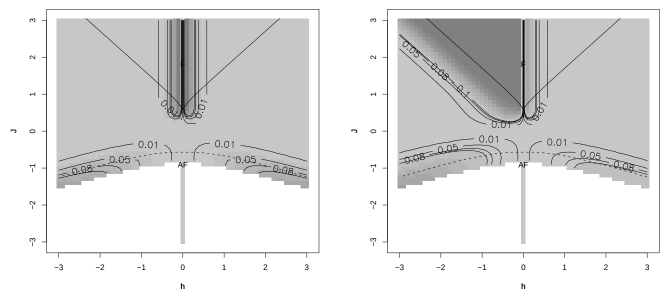

Fig. 7. Absolute errors (left) and relative errors (right) between true marginal probabilities and corresponding beliefs. Boundaries of F and AF are imposed.

to some critical point and as in the critical region, the calculations never seem to stop. It is interesting to note that there are convergent, chaotic and periodic behaviors appeared as parameters varied.

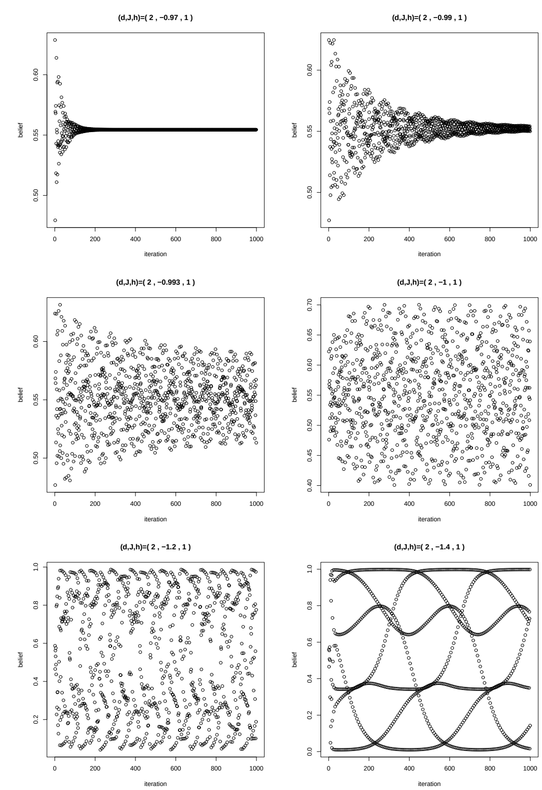

Fig. 8. Behavior of beliefs till 1000 iteration for parameter $J = -0.97, -0.99, -0, 993, -1, -1.2$ and -1.4 with $(d, h) = (2, 1)$ (from top left to bottom right).

We finally show a result about the influence of initial messages. We used an uniformly distributed number on the interval $(0, 1)$ to each initial message and check the iteration time up to 100 for each (J, h). Fig. 9 shows the variance of 100 samples. It shows the convergence depends on initial messages especially for the parameters which have large iteration time in Fig. 5.

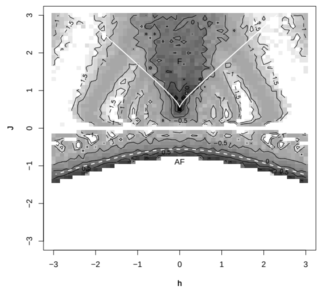

Fig. 9. Common logarithm of variances of iteration up to 100.

VI. Concluding remark

We review the convergence of BP algorithm and show the convergence depends on update rules and check difference of convergence between the lazy update rule and the impatient update rule for simple models.

Numerical experiments suggests that calculations with the impatient update rule converge faster than the one with the lazy update rule and that the former has a wider region of convergence, even in phase transition regions. It is shown that when message updates converge, the beliefs are good approximations of the true marginal probabilities. We also show the failure to converge shows the existence of phase transitions and the dependency of initial message for iteration time.

In the rest of this section, we give a remark about a sufficient condition.

Tatikonda and Jordan [9] referred a well-known sufficient condition for the lack of phase transitions

$$c(\Phi) \equiv \sup_{i \in S} \sum_{A: i \in A} (|A| - 1)\delta(\Phi_A) < 2, \tag{6}$$

where $\delta(f) \equiv \sup_{\xi, \eta \in \Omega} |f(\xi) - f(\eta)|$. In the case of Ising models on CT(d), $c(\Phi) = 2(d+1)|J|$ and the condition is satisfied for $|J| < 1/(d+1)$. Since $|A| - 1 = 0$ for $|A| = 1$, the condition does not have the influence of one-body potentials.

It is easily checked in the case of Ising models on Cayley trees. Let $J(d)$ defined in Section V and $f(x) = \dfrac{1}{2}\log\dfrac{x+1}{x-1} - \dfrac{1}{x+1}$. Since $f'(x) < 0$ for $x > 1$, $f(2) > 0$ and $\lim_{x \to \infty} f(x) = 0$, $J(d) > 1/(d+1),\ d = 2, 3, \cdots$. Hence the region where the uniqueness condition is satisfied is a broad horizontal band between F and AF. Compared with the region where no phase transition occurs the condition is seen fairly restrictive.

References

[1] Cowell, R.G., Dawid, A.P., Lauritzen, S.L. and Spiegehalter, D.J. (1999). *Probabilistic Networks and Expert Systems*, Springer Verlag, New York, Berlin, Heidelberg.

[2] Frey, B.J. (1998). *Graphical Models for Pattern Classification, Data Compression and Channel Coding,* MIT press, Cambridge.

[3] Georgii, H.-O. (1988). *Gibbs Measures and Phase Transitions*, Walter de Gruyter, Berlin · New York.

[4] Jensen, F. (1996) *An Introduction to Bayesian Networks*, UCL Press, London.

[5] McEliece, R.J., MackKay, D.J.C. and Cheng, J.F. (1998). Turbo Decodng as an Instance of Pearl's "Belief Propagation" Algorithm, *IEEE Journal on Selected Areas in Communication*, **16(2)** 140-152 Springer Verlag, New York, Berlin, Heidelberg.

[6] Murphy, K.P., Weiss, Y. and Jordan, M.I. (1999). Loopy belief propagation for approximate inference: an empirical study, *Proceedings of the 15th Conference on Uncertainty in Artificial Intelligence*, Morgan Kaufmann, San Francisco.

[7] Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference,* Morgan Kaufmann, San Francisco.

[8] Ruelle, D. (1978). *Thermodynamic Formalism*, Addison Wesley, New York.

[9] Tatikonda, S.C. and Jordan, M.I. (2002). Loopy Belief Propagation and Gibbs Measures, *Proceedings of the 18th Conference on Uncertainty in Artificial Intelligence,* Morgan Kaufmann, San Francisco.

[10] Weiss, Y. (2000). *Correctness of Local Probability Propagation in Graphical Models with Loops*, Neural Computation, **12**, 1-41.

[11] Weiss, Y. and Freeman, W.T. (2001). *Correctness of Belief Propagation in Gaussian Graphical Models of Arbitrary Topology,* Neural Computation, **12**, 2173-2200.

[12] Yedidia, J.S., Freeman, W.T. and Weiss, Y. (2002). Constructing Free Energy Approximations and Generalized Belief Propagation Algorithms, Mitsubishi Electric Research Laboratories Technical Reports TR2002-35, Cambridge.