

Sequential Minimization of the Bethe Free Energy of Ising Spin Systems

Yukinori TONOSAKI and Yoshiyuki KABASHIMA

*Department of Computational Intelligence and Systems Science, Tokyo Institute of Technology,
Yokohama 226-8502, Japan*

Received December 1, 2006; final version accepted December 27, 2006

A technique for minimizing the Bethe free energy of Ising spin systems is presented. The technique is based on a property of the Bethe free energy that diagonal elements of the Hessian are generally positive. This implies that solving the extremum condition with respect to a single element with fixing others at each update yields a unique solution which can be easily found by efficient algorithms such as a bisection method and reduces the value of the Bethe free energy. The proposed method, therefore, iterates sequential minimization with respect to a single element, which probably leads to convergence to a local minimum. Practical relevance of the scheme is shown by an application to a problem of multidimensional probabilistic reasoning that arises in a modern wireless communication system.

KEYWORDS: Bethe free energy, belief propagation, convex-concave computational procedure, sequential minimization

I. Introduction

Probabilistic reasoning or efficient computation of expectations in massive statistical models is one of the major problems in modern information sciences. When an objective system can be pictorially represented by a graph free from loops, an efficient algorithm termed the belief propagation (BP) offers the exact values of expectations in a practically feasible time [1]. However, as the class of systems that can be represented by loop-free graphs are limited, much effort is being made recently for developing efficient approximation algorithms that can be applied to a wider class of systems [2–5].

Much attention is being paid for a family of mean field methods from statistical mechanics as a basis of developing such approximation algorithms. In particular, a scheme termed the Bethe approximation [6, 7] has been studied extensively since a close link to BP was reported several years ago [8]. An approximate solution of this method, which is characterized by the extremum condition of a certain cost function termed the Bethe free energy, can be identified with a fixed point of the BP dynamics that is applied to loopy graphs [2]. This link potentiated development of a novel algorithm for finding a solution of the Bethe approximation which is guaranteed to converge [3]. The new algorithm, which is often referred to as the convex-concave computational procedure (CCCP), is derived by constructing a procedure for solving the stationary solution of the BP dynamics in such a way that descent of the Bethe free energy is guaranteed at each update. Construction of CCCP is instructive for understanding why BP does not necessarily converge to a fixed point when employed in loopy graphs. In addition, this algorithm is of practical relevance in several applications exhibiting better performance than BP with consumption of computational cost of times [9, 10]. However, it is noteworthy that CCCP is not the unique algorithm for minimizing the Bethe free energy although a close relationship to BP is prominent in research of probabilistic reasoning. This implies that it is still important to explore other possible algorithms that carry out this approximation in a practically feasible time scale in order to provide for potentially wide applications of the Bethe method.

The present study is promoted by such motivation. More specifically, we propose an algorithm that searches a solution of the Bethe approximation sequentially minimizing the Bethe free energy for systems that are composed of binary variables (Ising spins). The algorithm is based on a distinctive property of the Bethe free energy that diagonal elements of the Hessian are generally positive. This indicates that the solving the extremum condition with respect to a single element with fixing others at each update yields a unique solution, which can be easily searched by efficient algorithms such as a bisection method reducing the value of the Bethe free energy. The proposed method, therefore, iterates sequential minimization with respect to a single element, which probably leads to convergence to a local minimum. Although our algorithm is currently limited to the systems of Ising spins, extension to general cases is probably possible, which will be examined in the future research.

This paper is organized as follows. In the next section, we introduce a formalism of the Bethe approximation based on the concept of the Bethe free energy. In addition, BP and CCCP are briefly reviewed as computationally feasible algorithms performing this approximation. In Section 3, we show that diagonal elements of the Hessian of the Bethe free energy are generally positive. Utilizing this property, we present an algorithm for sequentially minimizing the Bethe free energy, which can be employed for finding a solution of the Bethe approximation in a practically feasible

time. In Section 4, practical relevance of the developed scheme is examined by applying to a multidimensional probabilistic reasoning problem that arises in a wireless communication. The final section is devoted to summary.

II. Bethe Approximation: Bethe Free Energy, BP and CCCP

As a basis for the proposed algorithm, let us consider a joint distribution of N dimensional vector of Ising spins $\mathbf{S} = (S_1, S_2, \dots, S_N)$ ($S_l = \{+1, -1\}$, $l = 1, 2, \dots, N$),

$$P(\mathbf{S}) = Z^{-1} \prod_{\mu=1}^M \psi_{\mu}(\mathbf{S}_{\mu}) \exp \left[\sum_{l=1}^N h_l S_l \right], \quad (1)$$

where $\psi_{\mu}(\mathbf{S}_{\mu})$ is termed the clique evidence, which depends on a certain subset of multiple components (clique) \mathbf{S}_{μ} , and h_l is a local external field. $Z = \sum_{\mathbf{S}} \prod_{\mu=1}^M \psi_{\mu}(\mathbf{S}_{\mu}) \exp[\sum_{l=1}^N h_l S_l]$ is the partition function. In such systems, assessment of marginal probabilities

$$P(S_l) = \sum_{\mathbf{S} \setminus S_l} P(\mathbf{S}) \quad (2)$$

is computationally difficult in general, where $X \setminus Y$ represents a subset of X from which Y is excluded. Therefore, developing computationally tractable algorithms that assess an accurate approximation of Eq. (1) is important.

The Kullback-Leibler divergence from a test distribution $Q(\mathbf{S})$ to the objective one (1), which is defined as

$$KL(Q|P) = \sum_{\mathbf{S}} Q(\mathbf{S}) \ln \frac{Q(\mathbf{S})}{P(\mathbf{S})}, \quad (3)$$

offers a useful guideline for development of such an algorithm. $KL(Q|P)$ is non-negative and minimized to zero if and only if $Q(\mathbf{S}) = P(\mathbf{S})$, which implies that minimization of $KL(Q|P)$ with respect to $Q(\mathbf{S})$ always yields the correct distribution $P(\mathbf{S}) = Q(\mathbf{S})$. This property is, however, not useful for directly assessing Eq. (2). Instead, application of this minimization principle or its approximation to a family of tractable test distribution systematically leads to various potentially effective algorithms.

The Bethe approximation can be formulated along this guideline. For this, we rewrite Eq. (3) utilizing the variational free energy $\mathcal{F}(Q)$ as

$$\begin{aligned} \mathcal{F}(Q) &= \sum_{\mathbf{S}} Q(\mathbf{S}) \ln \frac{Q(\mathbf{S})}{\prod_{\mu=1}^M \psi_{\mu}(\mathbf{S}_{\mu}) \exp[\sum_{l=1}^N h_l S_l]} \\ &= -\ln Z + KL(Q|P) \geq -\ln Z, \end{aligned} \quad (4)$$

which implies that minimization of $\mathcal{F}(Q)$ yields the correct distribution $Q(\mathbf{S}) = P(\mathbf{S})$. The Bethe scheme is offered by approximating $\mathcal{F}(Q)$ with the Bethe free energy $\mathcal{F}_{\text{Bethe}}(\{b_{\mu}\}, \{b_l\})$, as

$$\begin{aligned} \mathcal{F}(Q) &\simeq \mathcal{F}_{\text{Bethe}}(\{b_{\mu}\}, \{b_l\}) \\ &= \sum_{\mu=1}^M \sum_{\mathbf{S}_{\mu}} b_{\mu}(\mathbf{S}_{\mu}) \ln \frac{b_{\mu}(\mathbf{S}_{\mu})}{\psi_{\mu}(\mathbf{S}_{\mu}) \exp[\sum_{l \in \mathcal{L}(\mu)} h_l S_l]} + \sum_{l=1}^N (1 - C_l) \sum_{S_l} b_l(S_l) \ln \frac{b_l(S_l)}{\exp[h_l S_l]}, \end{aligned} \quad (5)$$

where $\mathcal{L}(\mu)$ denotes the set of element indices that directly relate to clique μ and C_l is the number of cliques to which element S_l is directly connected. The components of the test distribution $b_{\mu}(\mathbf{S}_{\mu})$ and $b_l(S_l)$ are termed the beliefs, which correspond to the marginals $\sum_{\mathbf{S} \setminus \mathbf{S}_{\mu}} P(\mathbf{S})$ and $\sum_{\mathbf{S} \setminus S_l} P(\mathbf{S})$, respectively. In the case of Ising spin systems that we are focusing on, one can express the single site marginal as $b_l(S_l) = (1 + m_l S_l)/2$ utilizing the expectation parameter $m_l = \sum_{S_l = \pm 1} S_l b_l(S_l)$ without loss of generality. We adopt this expression hereafter. Note that the marginals are reduced from the identical distribution $P(\mathbf{S})$, the reducibility condition

$$\sum_{\mathbf{S}_{\mu}} S_l b_{\mu}(\mathbf{S}_{\mu}) = \sum_{S_l} S_l b_l(S_l) = m_l, \quad (6)$$

must hold when S_l is an element of \mathbf{S}_{μ} .

If the variable dependence in Eq. (1) is pictorially expressed by a loop-free graph, Eq. (5) under the constraint of Eq. (6) exactly agrees to the correct variational free energy $\mathcal{F}(Q)$ for the test distribution

$$Q(\mathbf{S}) = \frac{\prod_{\mu=1}^M b_{\mu}(\mathbf{S}_{\mu})}{\prod_{l=1}^N \left(\frac{1 + m_l S_l}{2} \right)^{C_l - 1}}. \quad (7)$$

In such cases, minimization of the Bethe free energy (5) offers the exact assessment of Eq. (2). Unfortunately, Eqs. (4) and (5) do not accord in general. However, this implies that a stationary point of Eq. (5) probably offers a good approximation of Eq. (2) when influences of loops in the graph can be regarded as weak. This is an intuitive

justification of the Bethe approximation.

Extremizing Eq. (4) with respect to beliefs adding terms $\sum_{\mu, l \in \mathcal{L}(\mu)} \theta_{\mu l} (m_l - \sum_{S_\mu} S_l P(S_l))$ to impose the constraint (6) yields

$$b_\mu(\mathbf{S}_\mu) = \frac{\psi_\mu(\mathbf{S}_\mu) \prod_{l \in \mathcal{L}(\mu)} e^{(h_l + \theta_{\mu l}) S_l}}{\sum_{S_\mu} \psi_\mu(\mathbf{S}_\mu) \prod_{l \in \mathcal{L}(\mu)} e^{(h_l + \theta_{\mu l}) S_l}}, \quad (8)$$

$$\tanh^{-1}(m_l) = h_l + \sum_{\mu \in \mathcal{M}(l)} \frac{\theta_{\mu l}}{C_l - 1}, \quad (9)$$

where $\mathcal{M}(l)$ denotes the set of clique indices that are directly linked to element index l . Inserting Eqs. (8) and (9) to Eq. (6) yields the stationary point condition of the Lagrange multipliers, which can be expressed as

$$\hat{\theta}_{\mu l} = \tanh^{-1} \left(\frac{\sum_{S_\mu} S_l \psi_\mu(\mathbf{S}_\mu) e^{\sum_{j \in \mathcal{L}(\mu) \setminus \{l\}} (h_j + \theta_{\mu j}) S_j}}{\sum_{S_\mu} \psi_\mu(\mathbf{S}_\mu) e^{\sum_{j \in \mathcal{L}(\mu) \setminus \{l\}} (h_j + \theta_{\mu j}) S_j}} \right), \quad (10)$$

$$\theta_{\mu l} = \sum_{v \in \mathcal{M}(l) \setminus \mu} \hat{\theta}_{v l}. \quad (11)$$

BP can be regarded as the natural iteration of Eqs. (10) and (11). The approximate expectation is assessed as $m_l = \tanh(h_l + \sum_{\mu \in \mathcal{M}(l)} \hat{\theta}_{\mu l})$ utilizing the convergent solution of these equations.

Although BP empirically exhibits an excellent ability of quickly finding the stationary solution when employed under appropriate conditions, convergence to a local minimum of Eq. (5) is not necessarily guaranteed, which may undermine the reliability of the algorithm. In order to resolve such problems, Yuille proposed CCCP, for which descent of Eq. (5) is ensured at each update [3]. This is represented as a double loop algorithm

$$\textbf{Outer Loop: } m_l^{t+1} := \tanh \left(h_l + \sum_{\mu \in \mathcal{M}(l)} (\tanh^{-1}(m_l^t) - \theta_{\mu l} - h_l) \right), \quad (12)$$

where $\theta_{\mu l}$ is provided as the fixed point value of the inner loop update

$$\textbf{Inner Loop: } \theta_{\mu l}^{\tau+1} := \frac{\tanh^{-1}(m_l^t) - \hat{\theta}_{\mu l}^\tau - h_l}{2} + \sum_{v \in \mathcal{M}(l) \setminus \mu} \frac{\tanh^{-1}(m_l^t) - \theta_{v l}^\tau - h_l}{2}, \quad (13)$$

where a single pair of indices μ and l is chosen at each inner loop update indexed by τ and $\hat{\theta}_{\mu l}^\tau$ denotes the value of Eq. (10) evaluated utilizing values of $\theta_{\mu l}$ at inner loop time τ .

Unlike BP, CCCP is guaranteed to converge to a certain local minimum of Eq. (5). In addition, it is reported that considerably better performance than that of BP can be obtained by spending computational cost of only several times when conditions are optimally tuned in applications to decoding of low density parity check codes [9] and multiuser detection of code division multiple access systems [10]. However, necessary time for convergence of the inner loop empirically grows significantly in such situations that the Bethe free energy (5) can possess many local minima [10], which is probably linked to the concept of the replica symmetry breaking (RSB) [11, 12]. This implies that CCCP does not necessarily yield satisfactory results in practical situations and, therefore, quest for other methods is still important.

III. Sequential Minimization

The algorithm that we present relies on an inequality with respect to the diagonal elements of the Hessian of the Bethe free energy

$$\frac{\partial^2 \mathcal{F}_{\text{Bethe}}(\mathbf{m})}{\partial m_l^2} \geq 0, \quad (14)$$

which holds for any Ising spin systems and any set of expectation parameters $\mathbf{m} = (m_1, m_2, \dots, m_N)$, where $\mathcal{F}_{\text{Bethe}}(\mathbf{m}) = \min_{\{b_\mu\}} \{ \mathcal{F}_{\text{Bethe}}(\{b_\mu\}, \{b_l\}) \mid \sum_{S_l = \pm 1} S_l b_l(S_l) = m_l \ (l = 1, 2, \dots, N) \}$. This implies that for any element m_l , solving the extremum condition

$$\frac{\partial \mathcal{F}_{\text{Bethe}}(\mathbf{m})}{\partial m_l} = \tanh^{-1}(m_l) - h_l + \sum_{\mu \in \mathcal{M}(l)} (\theta_{\mu l} + h_l - \tanh^{-1}(m_l)) = 0, \quad (15)$$

where Lagrange multipliers $\theta_{\mu l}$ are determined for given \mathbf{m} so as to satisfy Eq. (6) with respect to m_l fixing the values of other elements $\mathbf{m} \setminus m_l$ provides a unique solution, which can be found by a certain efficient scheme such as a bisection method. Further, update of m_l to the solution guarantees descent of the value of the Bethe free energy. This indicates that sequentially updating m_l in any order leads to convergence to a certain solution which is probably a local minimum of $\mathcal{F}_{\text{Bethe}}(\mathbf{m})$ if $\mathcal{F}_{\text{Bethe}}(\mathbf{m})$ is bounded from below by a certain finite value, which holds in most cases [3]. For comparison, flowcharts of BP, CCCP and the current algorithm (SEQ) are shown in Fig. 1.

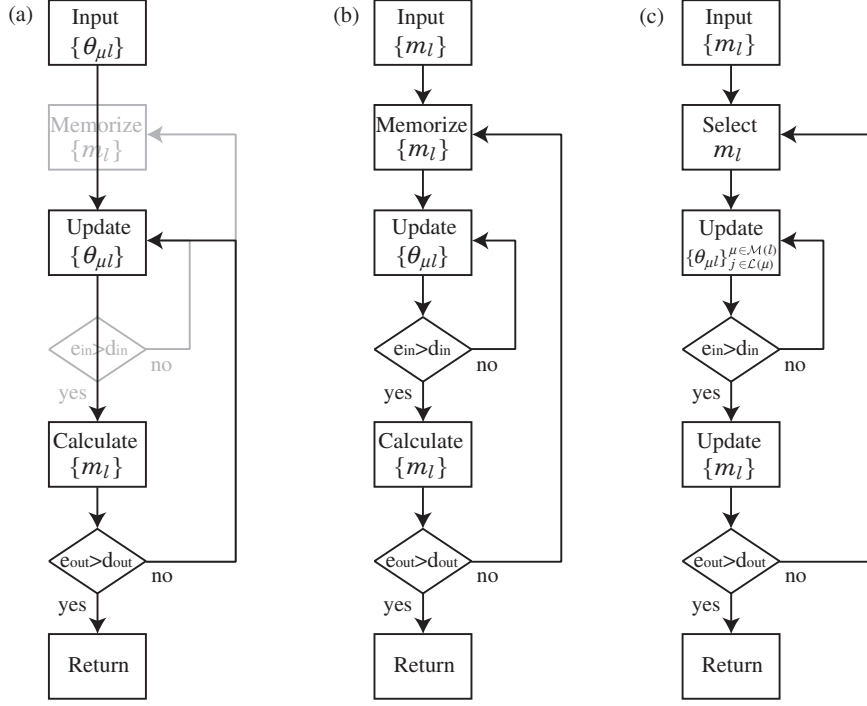


Fig. 1. Flowcharts of (a): BP, (b): CCCP and (c): the sequential scheme (SEQ). d_{out} denotes a deviation between successive updates while ϵ_{out} is a predetermined tolerance criterion. (a): In BP, the extremum condition of the Bethe free energy is solved with respect to a set of the Lagrange multipliers $\{\theta_{\mu l}\}$. (b): CCCP constitutes a double loop dynamics, where the set $\{\theta_{\mu l}\}$ are determined by an iterative manner in the inner loop, the convergent solution of which updates the spin averages $\{m_l\}$ in the outer loop. (c): SEQ solves the extremum condition with respect to a single element m_l sequentially. This requires determining Lagrange multipliers $\theta_{\mu l}$ of $\mu \in \mathcal{M}(l)$ at each update.

In order to prove inequality (14), we explicitly express the left hand side as

$$\frac{\partial^2 \mathcal{F}_{\text{Bethe}}(\mathbf{m})}{\partial m_l^2} = \frac{1}{1 - m_l^2} + \sum_{\mu \in \mathcal{M}(l)} \left(\frac{\partial \theta_{\mu l}}{\partial m_l} - \frac{1}{1 - m_l^2} \right). \quad (16)$$

For assessing $\partial \theta_{\mu l} / \partial m_l$, we utilize a relation

$$\frac{\partial m_l}{\partial \theta_{\mu k}} = \begin{cases} \langle S_l S_k \rangle_{\mu} - \langle S_l \rangle_{\mu} \langle S_k \rangle_{\mu}, & \text{if } l, k \in \mathcal{L}(\mu), \\ 0, & \text{otherwise,} \end{cases} \quad (17)$$

where $\langle \dots \rangle_{\mu}$ indicates average over Eq. (8), which represents the joint distribution of \mathbf{S}_{μ} . Let us denote the covariance matrix of \mathbf{S}_{μ} as $C^{\mu} = (C_{lk}^{\mu}) \equiv (\langle S_l S_k \rangle_{\mu} - \langle S_l \rangle_{\mu} \langle S_k \rangle_{\mu})$ and its inverse $A^{\mu} = (A_{lk}^{\mu}) \equiv (C^{\mu})^{-1}$, where $l, k \in \mathcal{L}(\mu)$. Using this, $\partial \theta_{\mu l} / \partial m_l$ is provided as

$$\frac{\partial \theta_{\mu l}}{\partial m_l} = A_{ll}^{\mu}. \quad (18)$$

Note that due to general properties of covariance matrices, any sub-matrices of C^{μ} and A^{μ} that are composed of elements of these matrices by restricting range of indices to any subset of $\mathcal{L}(\mu)$ are symmetric and positive definite.

In order to evaluate Eq. (18), we utilize a property of a Gaussian distribution of dimensionality $|\mathcal{L}(\mu)|$

$$P(\mathbf{z}_{\mu}) = \frac{|\det A^{\mu}|^{1/2}}{(2\pi)^{|\mathcal{L}(\mu)|/2}} \exp \left[-\frac{\mathbf{z}_{\mu}^T A^{\mu} \mathbf{z}_{\mu}}{2} \right], \quad (19)$$

where $\mathbf{z}_{\mu} = (z_l)$ and $l \in \mathcal{L}(\mu)$. The covariance matrix of this distribution accords to C^{μ} and, in particular, $\int d\mathbf{z}_{\mu} z_l^2 P(\mathbf{z}_{\mu}) = \langle S_l^2 \rangle_{\mu} - \langle S_l \rangle_{\mu}^2 = 1 - m_l^2$. This implies that marginal distribution with respect to z_l is assessed as

$$\begin{aligned} \int \prod_{j \in \mathcal{L}(\mu) \setminus l} dz_j P(\mathbf{z}_{\mu}) &= \frac{|\det A^{\mu}|^{1/2}}{\sqrt{2\pi} |\det A^{\mu \setminus l}|^{1/2}} \exp \left[-\frac{A_{ll}^{\mu} - \sum_{j, k \in \mathcal{L}(\mu) \setminus l} (A^{\mu \setminus l})_{jk}^{-1} A_{jl}^{\mu} A_{kl}^{\mu}}{2} z_l^2 \right] \\ &= \frac{1}{\sqrt{2\pi(1 - m_l^2)}} \exp \left[-\frac{z_l^2}{2(1 - m_l^2)} \right], \end{aligned} \quad (20)$$

where $A^{\mu \setminus l}$ represents the cofactor matrix of A^{μ} with respect to ll element and $(A^{\mu \setminus l})_{jk}^{-1}$ denotes jk element of inverse

matrix $(A^{\mu \setminus l})^{-1}$. This relation yields an inequality

$$\frac{\partial \theta_{\mu l}}{\partial m_l} - \frac{1}{1 - m_l^2} = A_{ll}^{\mu} - \frac{1}{1 - m_l^2} = \sum_{j,k \in \mathcal{L}(\mu) \setminus l} (A^{\mu \setminus l})_{jk}^{-1} A_{jl}^{\mu} A_{kl}^{\mu} \geq 0, \quad (21)$$

since the right hand side represents a quadratic form of positive definite matrix $(A^{\mu \setminus l})^{-1}$ due to properties of covariance matrices. Inserting Eq. (21) into Eq. (16) offers Eq. (14).

Three issues are noteworthy here. The first one is that finding a solution of Eq. (15) at each update is not computationally difficult as long as the size of clique μ , $|\mathcal{L}(\mu)|$, is of the order of unity. This is because Eq. (6), which is employed to determine Lagrange multipliers $\theta_{\mu} = (\theta_{\mu l})$ ($l \in \mathcal{L}(\mu)$) for \mathbf{m} , can be regarded as the stationary condition of a convex function

$$g_{\mu}(\theta_{\mu} | \mathbf{m}) = \ln \left[\sum_{\mathbf{S}_{\mu}} \psi_{\mu}(\mathbf{S}_{\mu}) e^{\sum_{l \in \mathcal{L}(\mu)} (h_l + \theta_{\mu l}) S_l} \right] - \sum_{l \in \mathcal{L}(\mu)} \theta_{\mu l} m_l. \quad (22)$$

Minimization of this function can be quickly carried out utilizing efficient techniques for optimizing convex functions as long as Eq. (22) is assessed with a computational cost of $O(1)$. This is the case when $|\mathcal{L}(\mu)| \sim O(1)$. As $|\mathcal{L}(\mu)|$ grows, cost for solving Eq. (15) in general increases exponentially and, therefore, performing our sequential scheme becomes difficult. However, even in such cases, there is an example in which our scheme can still be performed in a practically feasible time scale with aids of the central limit theorem and Taylor's expansion, which is shown in the next section. The second one is concerning physical implication of positivity of Eq. (21). In order to illustrate this, let us consider the Sherrington–Kirkpatrick (SK) model [13], which corresponds to the case of $\psi_{\mu}(\mathbf{S}_{\mu}) = \exp[\beta J_{lk} S_l S_k]$ where $J_{lk} \sim \mathcal{N}(J_0/N, J^2/N)$, β denotes the inverse temperature and a pair of non-ordered indices (lk) is identified with a clique index μ . For this model, perturbation expansion offers a relation $\theta_{\mu l} \simeq \tanh^{-1}(m_l) - \beta J_{lk} m_k + (\beta J_{lk})^2 (1 - m_k^2) m_l$. This yields

$$\frac{\partial \theta_{\mu l}}{\partial m_l} - \frac{1}{1 - m_l^2} \simeq (\beta J_{lk})^2 (1 - m_k^2) \geq 0, \quad (23)$$

which, after being summed up over index μ (or k), represents the strength of an effective self-interaction frequently referred to as the Onsager's reaction term [14, 15]. In many body systems under interaction, each element effectively interacts with itself by *positive* feedback effect created through interactions with other elements, which is the origin of positivity of Eq. (21). In the Bethe approximation, such effect is virtually cancelled by locally approximating the manner of interaction with graphs of loop-free structure, which is reduced to subtraction of the Onsager reaction term from local fields in the case of the SK model. The final issue is that descent of the value of Eq. (5) by solving Eq. (15) with respect to m_l does not necessarily mean convergence to a local minimum. This is because even if an objective function does not increase against perturbation with respect to a single element at a certain point, it can be reduced by changing multiple elements simultaneously to an appropriate direction when the Hessian of the point possesses negative eigen values. However, as far as we experimentally investigated, the Hessians of all the convergent solutions obtained by randomly permuting order of elements to update were positive definite. This implies that our sequential scheme practically serves as a minimization method of the Bethe free energy.

IV. Application to CDMA Multiuser Detection

In order to examine practical relevance, we applied the proposed method to the multiuser detection problem, which arises in the code division multiple access (CDMA) scheme used in a modern wireless communication system.

In a general scenario of the CDMA system, bit symbols of K users, $\mathbf{b}^0 = (b_k^0) = \{+1, -1\}^K$, are transmitted to a single base station being modulated by spreading sequences $\mathbf{s}_k = (s_{\mu k}) = \{+1, -1\}^N$, which are assigned to each user k . At the base station, mixtures of the modulated signals and noises are received. Under a simplifying assumption that both of the symbol and chip timings are synchronized across users, signal power is perfectly controlled to unity per each user and symbol period and the channel can be modeled by an additive white Gaussian noise of variance σ_0^2 , the received signals are represented as

$$r_{\mu} = \frac{1}{\sqrt{N}} \sum_{k=1}^K s_{\mu k} b_k^0 + \sigma_0 \eta_{\mu}. \quad (24)$$

The base station, therefore, has to demodulate $\mathbf{r} = (r_{\mu})$ to detect users' symbols $\mathbf{b}^0 = (b_k^0)$ with knowledge of sequences $\{\mathbf{s}_k\}$, which is termed the user detection.

The user detection can be regarded as a multidimensional probabilistic reasoning problem. A general argument of the Bayesian statistics indicates that the optimal detection scheme to minimize the bit error rate P_b , which represents the probability of wrongly estimating bit symbols in the component-wise manner, is offered as

$$\hat{b}_k = \operatorname{argmax}_{b_k \in \{+1, -1\}} \left\{ \sum_{\mathbf{b} \setminus b_k} P(\mathbf{b} | \mathbf{r}, \{s_k\}) \right\}. \quad (25)$$

Here, the posterior probability is provided as

$$P(\mathbf{b} | \mathbf{r}, \{s_k\}) = \frac{\prod_{\mu=1}^N \exp[-(2\sigma^2)^{-1}(r_\mu - N^{-1/2} \sum_{k=1}^K s_{\mu k} b_k)^2]}{\sum_{\mathbf{b}} \prod_{\mu=1}^N \exp[-(2\sigma^2)^{-1}(r_\mu - N^{-1/2} \sum_{k=1}^K s_{\mu k} b_k)^2]}, \quad (26)$$

where the uniform prior $P(\mathbf{b}) = 2^{-K}$ is assumed and σ^2 denotes the noise variance assumed at the base station. As knowledge of all users' sequences is necessary for assessing Eq. (25), detection scheme of this type is often referred to as the multiuser detection.

Computational cost for exactly performing the multiuser detection in general grows exponentially with respect to the number of users K . In order to practically resolve this difficulty, we employ our sequential minimization scheme introducing the Bethe free energy to Eq. (26). For this purpose, we define clique evidences as $\psi_\mu(\mathbf{b}) = \exp[-(2\sigma^2)^{-1}(r_\mu - N^{-1/2} \sum_{k=1}^K s_{\mu k} b_k)^2]$ and set $h_k = 0$ for $k = 1, 2, \dots, K$. This implies that the size of clique μ is K and, therefore, evaluating the Lagrange multipliers θ_μ for given \mathbf{m} is computationally difficult as K grows. However, when $s_{\mu k}$ is independently and identically drawn from $P(s) = (1/2)(\delta(s-1) + \delta(s+1))$, which is at least the case in a basic model [16], summation over \mathbf{b} in Eq. (22) can be well approximated with one dimensional integral with respect to a Gaussian measure $\mathcal{N}(\langle \Delta_\mu \rangle_{\setminus \mu}, \beta(1 - Q_{\setminus \mu}))$ because of the central limit theorem, where $\langle \Delta_\mu \rangle_{\setminus \mu} = N^{-1/2} \sum_{k=1}^N s_{\mu k} \tanh(\theta_{\mu k})$, $Q_{\setminus \mu} = K^{-1} \sum_{k=1}^K \tanh^2(\theta_{\mu k})$ and $\beta = K/N$, which yields

$$g_\mu(\theta_\mu | \mathbf{m}) \simeq -\frac{1}{2} \ln(\sigma^2 + \beta(1 - Q_{\setminus \mu})) - \frac{(r_\mu - \langle \Delta_\mu \rangle_{\setminus \mu})^2}{2(\sigma^2 + \beta(1 - Q_{\setminus \mu}))} + \sum_{k=1}^K (\ln(2 \cosh(\theta_{\mu k})) - \theta_{\mu k} m_k). \quad (27)$$

For large N and K with keeping $\beta \sim O(1)$, minimization of Eq. (27) with respect to $\theta_{\mu k}$ utilizing the Taylor expansion provides

$$\theta_{\mu k} \simeq \tanh^{-1}(m_k) - \frac{s_{\mu k}}{\sqrt{N}\sigma^2} \left(r_\mu - \sum_{j=1}^K \frac{s_{\mu j}}{\sqrt{N}} m_j \right) - \frac{m_k}{N(\sigma^2 + \beta(1 - Q))}, \quad (28)$$

where $Q_{\setminus \mu}$ is approximated with $Q = K^{-1} \sum_{k=1}^K m_k^2$ utilizing the law of large numbers. Inserting this into Eq. (15) offers an expression of the extremum condition

$$\tanh^{-1}(m_k) + \frac{\beta(1 - Q)m_k}{\sigma^2(\sigma^2 + \beta(1 - Q))} - \frac{1}{\sigma^2} \sum_{\mu=1}^N \frac{s_{\mu k}}{\sqrt{N}} \left(r_\mu - \sum_{j \neq k} \frac{s_{\mu j}}{\sqrt{N}} m_j \right) = 0, \quad (29)$$

which is to be solved sequentially with respect to m_k at each update. Equation (29) corresponds to the Thouless–Anderson–Palmer equation of the current problem [15, 17]. As a bisection method can be employed for finding a solution of Eq. (29) with respect to m_k , computational cost for performing one sequential step which indicates solving m_k for $k = 1, 2, \dots, K$ in a random order scales as $O(KN)$, which can be regarded as efficient [17, 18].

In Figs. 2(a) and (b), P_b numerically evaluated via 10000 experiments utilizing the proposed sequential scheme is plotted versus various noise variances σ_0^2 in the case of $N = 50$ and $N = 100$ for $\beta = 0.5$. In the experiments, the

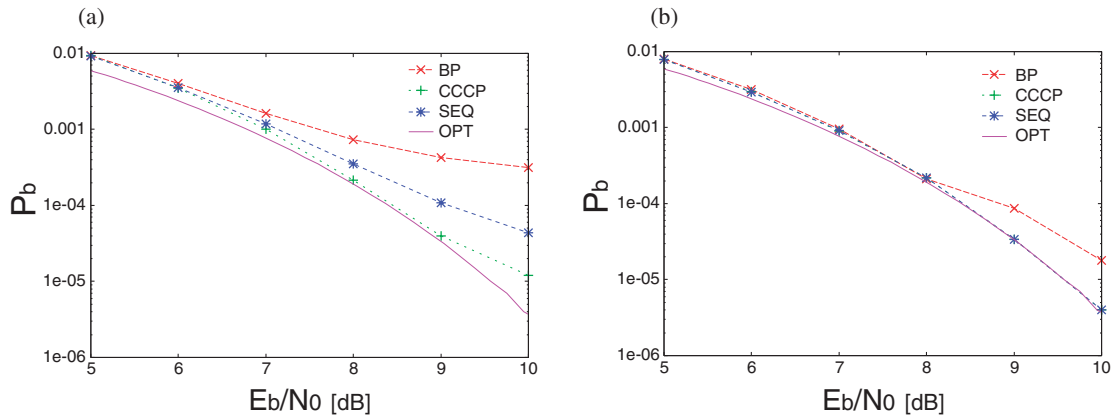


Fig. 2. Bit error rate P_b for the sequential scheme (SEQ), CCCP and BP in the cases of $N = 50$ (a) and $N = 100$ (b). The horizontal axis represents $E_b/N_0 = 10 \log_{10}(1/(2\sigma_0^2))$. The markers denotes experimental values estimated by 10000 numerical simulations in which the correct noise parameters $\sigma^2 = \sigma_0^2$ were used. OPT indicates performance obtained when the channel is accessed by only a single user.

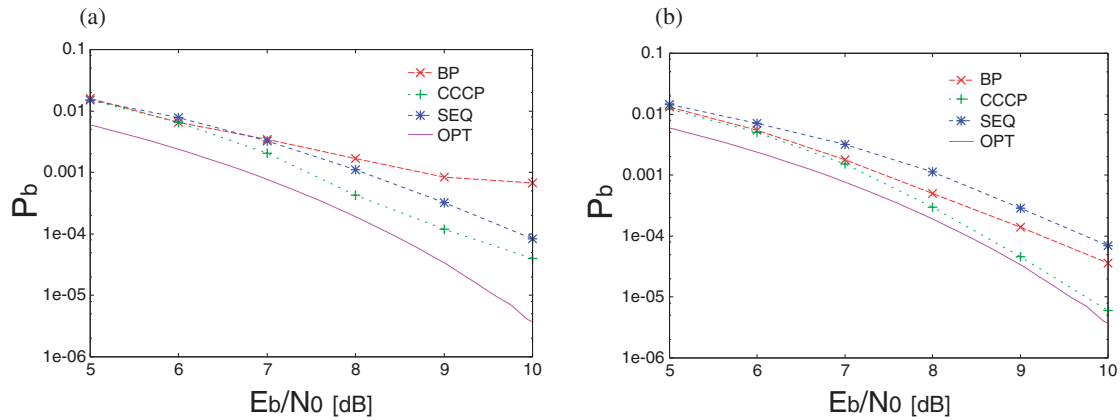


Fig. 3. Results for cases of parameter mismatch $\sigma^2 = 0.01\sigma_0^2$.

correct parameters $\sigma^2 = \sigma_0^2$ were used. For comparison, results for BP and CCCP-based algorithms [10, 17] are also presented. These figures indicate that the current scheme (SEQ) exhibits performance close to the optimal in the case of $N = 100$ although performance for $N = 50$ is considerably worse than that for CCCP. However, our scheme can still be practically useful because CCCP requires a more computational cost of $O(N^2K)$ per update than that of the current scheme ($O(NK)$), which practically limits applicability of CCCP to relatively small systems. Figures 3(a) and (b) show P_b in the case of parameter mismatch $\sigma^2 = 0.01\sigma_0^2$. Other conditions are same as those in Figs. 2(a) and (b). These indicate that performance of the current scheme is not necessarily good when a too small parameter is used. When σ^2 is set below a critical value for given β and σ_0^2 , the de Almeida–Thouless stability [19] can be broken in this system in the limit of $N, K \rightarrow \infty$ [16, 17], which implies that the Bethe free energy can possess many local minima. Deterioration of performance of the current scheme may be because the local minima are adequately found. Such property is preferred in examination of a complex structure of RSB in spin glass research.

V. Summary

In summary, we have developed a method for performing the Bethe approximation in Ising spin systems. The method is based on a property of the Bethe free energy that diagonal elements of the Hessian are generally positive. This guarantees that solving the extremum condition with respect to a single element at each update yields a unique solution, which can be quickly found by efficient methods, and update to the solution generally decreases the Bethe free energy. Therefore, sequentially solving the extremum condition with respect to a single element probably converges to a local minimum. Property of this scheme has been examined by application to the CDMA multiuser detection problem, which indicates that performance close to the optimal can be achieved for a case of mediate size with computational cost of $O(NK)$ per sequential update although the performance considerably deteriorates when the Bethe free energy can have many local minima. This may be an evidence that the current scheme can adequately find various local minima.

Application to a problem of spin glasses can be found in [20].

Acknowledgments

This work was partially supported by Grant-in-Aids from JSPS/MEXT, Japan, Nos. 17340116 and 18079006.

REFERENCES

- [1] Pearl, J., *Probabilistic Reasoning in Intelligent Systems: Network of Plausible Inference*, Morgan Kaufmann, San Francisco (1988).
- [2] Yedidia, J. S., Freeman, W. T., and Weiss, Y., *Advances in Neural Information Processing Systems*, **13**: 689–695 (2000).
- [3] Yuille, A. L., *Neural Comput.*, **14**: 1691–1722 (2002).
- [4] Tanaka, K., *IEICE Transactions on Information and Systems*, **E86-D**: 1228–1242 (2003).
- [5] Welling, M., and Teh, Y. W., *Neural Comput.*, **16**: 197–221 (2004).
- [6] Bethe, H., *Proc. Roy. Soc. (London)*, **A150**: 552–575 (1935).
- [7] Kikuchi, R., *Phys. Rev.*, **81**: 988 (1951).
- [8] Kabashima, Y., and Saad, D., *Europhys Lett.*, **44**: 668–674 (1998).
- [9] Shibuya, T., and Sakaniwa, K., *IEICE Trans. Fundamentals*, **E86-A**: 2601–2606 (2003).
- [10] Tonosaki, Y., and Kabashima, Y., *IEICE Transactions on Information and Systems*, **J89-D** (Japanese): 1049–1060 (2006).
- [11] Mézard, M., Parisi, G., and Virasoro, M. A., *Spin Glass Theory and Beyond*, World Scientific, Singapore (1987).
- [12] Fischer, K. H., and Hertz, J. A., *Spin Glasses*, Cambridge University Press, Cambridge (1991).

- [13] Sherrington, D., and Kirkpatrick, S., *Phys. Rev. Lett.*, **35**: 1792 (1975).
- [14] Morita, T., and Horiguchi, T., *Solid State Commun.*, **19**: 833–835 (1976).
- [15] Thouless, D. J., Anderson, P. W., and Palmer, R. G., *Philos. Mag.*, **35**: 593–601 (1977).
- [16] Tanaka, T., *IEEE Trans. Inf. Theor.*, **48**: 2888–2910 (2002).
- [17] Kabashima, Y., *J. Phys. A*, **37**: 11111–11121 (2003).
- [18] Varanasi, M. K., and Aazhang, B., *IEEE Trans. Commun.*, **38**: 509–519 (1990).
- [19] de Almeida, J. R. L., and Thouless, D. J., *J. Phys. A*, **11**: 983–990 (1978).
- [20] Tonosaki, Y., Takeda, K., and Kabashima, Y., *Phys. Rev. B*, **75**: 094405 (2007).