Exactness of belief propagation for some graphical models with loops

Michael Chertkov
Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA
E-mail: chertkov@lanl.gov

Received 25 July 2008
Accepted 10 September 2008
Published 17 October 2008

Online at stacks.iop.org/JSTAT/2008/P10016
doi:10.1088/1742-5468/2008/10/P10016

Abstract. It is well known that an arbitrary graphical model of statistical inference defined on a tree, i.e. on a graph without loops, is solved exactly and efficiently by an iterative belief propagation (BP) algorithm convergent to the unique minimum of the so-called Bethe free energy functional. For a general graphical model on a loopy graph, the functional may show multiple minima, the iterative BP algorithm may converge to one of the minima or may not converge at all, and the global minimum of the Bethe free energy functional is not guaranteed to correspond to the optimal maximum likelihood (ML) solution in the zero-temperature limit. However, there are exceptions to this general rule, discussed by Kolmogorov and Wainwright (2005) and by Bayati et al (2006, 2008) in two different contexts, where the zero-temperature version of the BP algorithm finds the ML solution for special models on graphs with loops. These two models share a key feature: their ML solutions can be found by an efficient linear programming (LP) algorithm with a totally uni-modular (TUM) matrix of constraints. Generalizing the two models, we consider a class of graphical models reducible in the zero-temperature limit to LP with TUM constraints. Assuming that a gedanken algorithm, g-BP, for finding the global minimum of the Bethe free energy is available, we show that in the limit of zero temperature, g-BP outputs the ML solution. Our consideration is based on equivalence established between gapless linear programming (LP) relaxation of the graphical model in the $T \to 0$ limit and the respective LP version of the Bethe free energy minimization.

Keywords: spin glasses (theory), analysis of algorithms, exact results, message-passing algorithms

ArXiv ePrint: 0801.0341
1. Introduction

Belief propagation (BP) is an algorithm finding an ML solution or marginal probabilities on a graph without loops, a tree. The algorithm was introduced in [8] as an efficient heuristic for decoding of sparse (so-called graphical) codes and it was independently considered in the context of graphical models of artificial intelligence [16]. Originally the algorithm was primarily thought of as an iterative procedure [24, 23], inspired by earlier works of [3] and [17] in statistical physics, and suggested to use a more fundamental notion of the Bethe free energy. Extrema of the Bethe free energy represent fixed points of the iterative BP algorithm on graphs with cycles. Equations describing the stationary points of the Bethe free energy and the fixed points of the iterative BP form are called belief propagation, Bethe–Peierls, or BP equations.

The significance of BP, understood as an algorithm looking for a minimum of the Bethe free energy, was further elucidated within the framework dubbed loop calculus [5, 6]. It was shown that an algorithm finding an extremum of the Bethe free energy is not just an approximation/heuristics in the loopy case, but also allows explicit reconstruction of the exact inference in terms of a series, where each term corresponds to a loop on the graph.

If the graphical model is dense, there are many loops, and thus many contributions to the loop series. However, not all loops are equal. Thus, considering models characterized by the same graph but different factor functions or local weights (exact definitions will follow) one expects strong sensitivity of an individual loop contribution (and its significance within the loop series) to the factor functions. In this context it is of interest to study the following question: are there graphical models, defined on an arbitrary graph but with specially tuned factor functions, such that BP provides exact inference?
A positive answer to this question was given, independently and for two different models, by [12] and [2]. It was shown in [12] that for a graphical model defined on an arbitrary graph in terms of binary variables with pairwise sub-modular interaction, a properly defined version of BP (linear programming relaxation underlying the tree-reweighted method of [20]) yields a globally optimal maximum likelihood solution. This model is equivalent to the ferromagnetic random field Ising (FRFI) model popular in statistical physics of disordered systems; see e.g. [10]. The maximum weight matching problem on a bipartite graph was analysed in [2] and later in [18], where it was shown that, in spite of the fact that the underlying graph has many short cycles, an algorithm of BP type does converge to the correct ML assignment. This consideration was also extended to the problem of weighted b-matchings on an arbitrary graph (which is yet another problem solvable exactly by BP) in [1,11]. Closely related general results, discussing convexified versions of Bethe free energy and an iterative convex BP scheme converging to the respective LP, were reported in [22,9].

In this paper we use the Bethe free energy approach of [23] to suggest a complementary and unifying explanation for these remarkable, and somewhat surprising, results of [12,2]. In two subsequent sections we consider two models, FRFI discussed in [12] and a binary model with totally uni-modular (TUM) constraints generalizing the weight matching problem considered in [2]. Statistical weights are defined for both models in terms of a characteristic temperature, $T$. Our strategy in dealing with both models is illustrated in figure 1. It consists of the following three steps.

- Starting from the original setting we first go anticlockwise, getting an integer programming (IP) formulation for the ML, $T \to 0$, version of the problem. The most important feature of the two models is that the LP relaxation of the respective IP, shown as LP-A in figure 1, is tight/exact. In both cases this reduction from IP to LP is exact due to the total uni-modularity (TUM) feature of the underlying matrix of constraints.

- Then we return to the original setting and start moving clockwise (see figure 1), first to the Bethe free energy formulation of the problem. We call the gedanken algorithm, finding the global minima of the Bethe free energy, g-BP$^1$. In the $T \to 0$ limit the Bethe free energy turns into the respective self-energy (the entropy term multiplied by temperature is irrelevant) which is a linear functional of beliefs. Thus one gets to an LP problem here as well, the one shown as LP-B in figure 1. This transformation from g-BP to LP-B is analogous to the similar relation between g-BP and LP decoding introduced in the coding theory in [7,19].

- Finally, we show that LP-A and LP-B are identical, thus demonstrating that g-BP in the $T \to 0$ limit outputs the ML solution.

Note that convexity of the Bethe free energy at finite temperature, playing the key role in the analysis of [12,22,9], is not a required part our consideration. Moreover, the Bethe free energy of the binary model with TUM constraints is generally not convex.

---

$^1$ The Bethe free energy is non-convex; therefore finding the global minimum at a finite temperature is not necessarily straightforward. Acknowledging the importance of the problem, we will not discuss in this manuscript plausibility and details of the respective iterative algorithm convergent to the global minimum of the Bethe free energy. We refer the interested reader to comprehensive discussion of such iterative schemes in the general context in [22,9] and for the FRFI model and maximum weight matching model in [12] and in [2,18] respectively.
2. Ferromagnetic random field Ising model

Consider an undirected graph $\mathcal{G}$, consisting of $n$ vertices, $\mathcal{V} = \{1, \ldots, n\}$, and weighted edges, $\mathcal{E}$, with the weight matrix $(J_{ij}|i, j = 1, \ldots, n)$ such that whenever the two vertices are connected by an edge, i.e. $i \in j$ or $j \in i$, $J_{ij} > 0$, and $J_{ij} = 0$, otherwise. It is also useful to introduce the directed version of the graph, $\mathcal{G}_d$, where any undirected edge of $\mathcal{G}$ is replaced by two directed edges of $\mathcal{G}_d$, with the weights $J_{i\rightarrow j} = J_{j\rightarrow i} = J_{ij}/2$ respectively. The ferromagnetic random field Ising (FRFI) model is defined by the following statistical weight associated with any configuration of $\sigma = (\sigma_i = \pm 1| i = 1, \ldots, n)$ on $\mathcal{V}$:

$$p(\sigma) = Z^{-1} \exp \left( \frac{1}{2T} \sum_{(i, j) \in \mathcal{G}} J_{ij} \sigma_i \sigma_j + \frac{1}{T} \sum_i h_i \sigma_i \right)$$

$$= Z^{-1} \exp \left( \frac{1}{2T} \sum_{(i\rightarrow j) \in \mathcal{G}_d} J_{i\rightarrow j} \sigma_i \sigma_j + \frac{1}{T} \sum_i h_i \sigma_i \right),$$

(1)

where $h_i$ can be positive or negative; $T$ is the temperature; $Z$ is the partition function, enforcing the probability normalization condition, $\sum_\sigma p(\sigma) = 1$; and $(i, j)/(i \rightarrow j)$ marks an undirected/directed edge of $\mathcal{G}/\mathcal{G}_d$ connecting the two neighbours $i$ and $j$.

2.1. From FRFI to the ‘min-cut’ problem

The maximum likelihood (ground state) solution of equation (1) turns into the problem of quadratic integer programming

$$\min_\sigma \left( \frac{1}{2} \sum_{(i\rightarrow j) \in \mathcal{G}_d} J_{i\rightarrow j} \sigma_i \sigma_j - \sum_i h_i \sigma_i \right) \quad \forall \ i \in \mathcal{G}_d; \sigma_i = \pm 1,$$

(2)
It is well known that any sub-modular energy function (and the quadratic function in equation (2) is of this type) can be minimized in polynomial time by reducing the task to the maximum flow/‘min-cut’ problem \[4,10\]. In this section we will reproduce these known results.

To unify linear and quadratic terms in equation (2), one constructs a new graph, \(G'_d\), adding two new nodes to \(G_d\): the source (s) and destination (d), with \(\sigma_s = +1\) and \(\sigma_d = -1\) respectively. The (s)-node is linked to all the nodes of \(G_d\) with \(h_i > 0\), while any node of \(G_d\) with \(h_i < 0\) is linked to (d). Weights of the newly introduced directed edges of \(G'_d\) are

\[
J_{s\rightarrow i} = 2h_i, \quad \text{if } h_i > 0; \quad J_{i\rightarrow d} = 2|h_i|, \quad \text{if } h_i < 0.
\]

This results in the following version of equation (2):

\[
\min_{\sigma} \left( -\frac{1}{2} \sum_{(i\rightarrow j)\in G'_d} J_{i\rightarrow j}\sigma_i\sigma_j \right) \quad \forall i\in G_d: \sigma_i = \pm 1; \sigma_s = +1; \sigma_d = -1
\]

Reduction from quadratic integer programming (4) to an integer linear programming is the next step. This is achieved via transformation to the edge variables,

\[
\forall (i\rightarrow j)\in G_d: \quad \eta_{i\rightarrow j} = \begin{cases} 1, & \sigma_i = 1, \sigma_j = -1 \\ 0, & \text{otherwise.} \end{cases}
\]

The relations can also be restated:

\[
\forall (i\rightarrow j), (j\rightarrow i)\in G_d: \quad \sigma_i\sigma_j + \sigma_j\sigma_i = 2 - 4(\eta_{i\rightarrow j} + \eta_{j\rightarrow i}) \quad \text{(6)}
\]

\[
\forall (d\rightarrow i), (j\rightarrow t)\in G_d: \quad \sigma_i\sigma_d + \sigma_d\sigma_i = 1 - 2\eta_{d\rightarrow i}, \quad \sigma_s\sigma_j = 1 - 2\eta_{j\rightarrow s}. \quad \text{(7)}
\]

Therefore, taking into account that \(J_{i\rightarrow j} = J_{j\rightarrow i}\) for any \((i\rightarrow j), (j\rightarrow i)\in G_d\), substituting equations (5)–(7) into (4) and changing variables from \(\sigma_i = \pm 1\) to \(p_i = (1 - \sigma_i)/2 = 0, 1\), one arrives at

\[
-\frac{1}{2} \sum_{(i\rightarrow j)\in G'_d} J_{i\rightarrow j} + \min_{\{\eta,p\}} \sum_{(i\rightarrow j)\in G'_d} J_{i\rightarrow j}\eta_{i\rightarrow j} \quad \forall i\in G'_d: p_i = 0; \eta_i = 0; p_d = 1; \eta_d = 0. \quad \text{(8)}
\]

This expression is nothing but the integer programming formulation of the famous ‘min-cut’ problem, calculating the minimum weight cut splitting all the nodes of the directed graph into two parts such that the group including the source node has all variables in the 0 state while the other group, including the destination node, has all variables in the 1 state.

Any \(\{\eta, p\}\) configuration which satisfies conditions in equation (8) requires that either \(\eta_{i\rightarrow j} = 0\) and \(\eta_{j\rightarrow i} = 1\) or \(\eta_{i\rightarrow j} = 1\) and \(\eta_{j\rightarrow i} = 0\) for any pair of directed edges \((i\rightarrow j), (j\rightarrow i)\in G_d\). This suggests that equation (8) can be restated in terms of the undirected graph \(G'\), equivalent to the original \(G\) supplemented by the source and destination vertices and edges with the following positive weights:

\[
J_{si} = 2h_i, \quad \text{if } h_i > 0; \quad J_{id} = 2|h_i|, \quad \text{if } h_i < 0. \quad \text{(9)}
\]

doi:10.1088/1742-5468/2008/10/P10016
One derives the following undirected version of equation (8):

$$\frac{1}{2} \sum_{(i,j) \in G'} J_{ij} + \min_{\{\eta, p\}} \left( \sum_{(i,j) \in G'} J_{ij} \eta_{i \rightarrow j} \right) \left\{ \begin{array}{l}
\forall (i,j) \in G' : p_i = 0; p_j = 1; \\
\forall (i,j) \in G' : p_i - p_j + \eta_{ij} = 0; p_i = 1
\end{array} \right. \right.$$

(10)

The min-cut problem (10) is solvable in polynomial time. This means, in particular, that the solution of the integer programming equation (10) and the solution of the respective relaxed LP-A,

$$\frac{1}{2} \sum_{(i,j) \in G'} J_{ij} + \min_{\{\eta, p\}} \left( \sum_{(i,j) \in G'} J_{ij} \eta_{ij} \right) \left\{ \begin{array}{l}
\forall (i,j) \in G', 1 \geq p_i \geq 0; \\
\forall (i,j) \in G' : p_i - p_j + \eta_{ij} \geq 0; p_i = 1
\end{array} \right. \right.$$

(11)

are identical. The tightness of the relaxation is discussed in, e.g., [15]. (See chapter 6.1 and specifically theorems 6.1, 6.2 in [15].) Also, this observation is closely related to the fact that the matrix of constraints in the ‘max-flow’ problem, which is dual to equation (10), is totally uni-modular (TUM), i.e. such that any square minor of the matrix has a determinant which is 0, +1 or -1. (See e.g. chapter 13.2 of [15] for discussion of the TUM IP/LP problems.)

2.2. Bethe free energy and belief propagation for FRFI

Discussing the FRFI model defined in equation (1) and following the general heuristic approach to the graphical models, suggested in [23], one introduces beliefs, i.e. estimated probabilities, for vertices and edges, $b_i(\sigma_i)$, $b_{ij}(\sigma_i, \sigma_j)$, related to each other according to

$$\forall i \quad \text{and} \quad \forall j \in i : \quad b_i(\sigma_i) = \sum_{\sigma_j} b_{ij}(\sigma_i, \sigma_j),$$

(12)

and also satisfying the obvious normalization condition

$$\forall i : \quad \sum_{\sigma_i} b_i(\sigma_i) = 1.$$  

(13)

Then the Bethe free energy functional of the beliefs is defined as

$$F = E - TS, \quad E = \frac{1}{2} \sum_{(i,j)} \sum_{\sigma_i, \sigma_j} b_{ij}(\sigma_i, \sigma_j) J_{ij} \sigma_i \sigma_j - \sum_i \sum_{\sigma_i} b_i(\sigma_i) h_i \sigma_i,$$

(14)

$$S = \sum_{(i,j)} \sum_{\sigma_i, \sigma_j} b_{ij}(\sigma_i, \sigma_j) \ln b_{ij}(\sigma_i, \sigma_j) - \sum_i \sum_{\sigma_i} b_i(\sigma_i) \ln b_i(\sigma_i).$$

(15)

Introducing Lagrangian multipliers associated with the constraints (12), (13), one defines the Lagrangian functional

$$\mathcal{L} = F + \sum_{i} \sum_{j \in i} \sum_{\sigma_i} \eta_{ij}(\sigma_i) \left( b_i(\sigma_i) - \sum_{\sigma_j} b_{ij}(\sigma_i, \sigma_j) \right) + \sum_{i} \lambda_i \left( \sum_{\sigma_i} b_i(\sigma_i) - 1 \right).$$

(16)

Looking for the stationary point of the Lagrangian over all the parameters (the beliefs and the Lagrangian multipliers) will define the belief propagation (BP) equations. Iterative solution of the BP equations constitutes the celebrated BP algorithm, which is often used as an efficient heuristic for estimating marginal probabilities in sparse graphical models.

doi:10.1088/1742-5468/2008/10/P10016
2.2.1. Ground state. In the $T \to 0$ limit the entropy terms in the expression for the Bethe free energy in equations (14) can be neglected and the task of finding the absolute minimum of the Bethe free energy functional turns into minimization of the self-energy, $E$ from equation (14), under the set of constraints (12), (13). Both the optimization functional and the constraints are linear in the beliefs; therefore one gets here the following linear programming optimization:

$$
\min_{\{b_i; b_{ij}\}} \left( - \sum_{(i,j) \in G} \sum_j b_{ij}(\sigma_i, \sigma_j) \frac{J_{ij}}{2} \sigma_i \sigma_j - \sum_i \sum_{\sigma_i} b_i(\sigma_i) h_i(\sigma_i) \right) \bigg|_{b_i(\sigma_i) = \sum_{\sigma_j} b_{ij}(\sigma_i, \sigma_j), \sum_{\sigma_j} b_{ij}(\sigma_i, \sigma_j) = 1}^{\forall \sigma_i, \forall (i,j) \in G'},
$$

where it is also assumed that all the beliefs are positive and smaller than or equal to unity (as we are looking only for physically sensible solutions of the optimization problem).

Making the transformation from the original graph $G$ to its extended version, $G'$, i.e. introducing new edges with weights defined in equations (9), and requiring that the spin values of the source/destination are fixed to $\pm 1$, equations (9) and (14), under the set of constraints (12), (13). Both the optimization functional and the constraints are linear in the beliefs; therefore one gets here the following linear programming optimization:

$$
\min_{\{b_i; b_{ij}\}} \left( - \sum_{(i,j) \in G'} \sum_j b_{ij}(\sigma_i, \sigma_j) \frac{J_{ij}}{2} \sigma_i \sigma_j \right) \bigg|_{b_i(\sigma_i) = \sum_{\sigma_j} b_{ij}(\sigma_i, \sigma_j), \sum_{\sigma_j} b_{ij}(\sigma_i, \sigma_j) = 1}^{\forall \sigma_i, \forall (i,j) \in G'},
$$

defined as the probabilities of observing the edge $(i, j)$ either in the state $(+, +)$ or in the state $(-, +)$. Thus, by construction, $1 \geq \mu_{ij} \geq 0$. The $\mu_{ij}$ variables defined at different edges are related to each other through local beliefs, $\pi_i = b_i(-) = 1 - b_i(+)$, which all satisfy, $0 \leq \pi_i \leq 1$. Taking all these observations into account one rewrites equation (18) as

$$
-\frac{1}{2} \sum_{(i,j) \in G'} J_{ij} + \min_{\{\mu_i\}} \sum_{(i,j) \in G'} J_{ij} \mu_{ij} \bigg|_{\forall \sigma_i, \forall (i,j) \in G', \forall \eta_j : \sigma_i \pi_j + \mu_{ij} \geq 0; \sum_{\eta_j} \eta_j \geq 0; \pi_i \geq 0; \pi_i \geq 1,}^{\forall \sigma_i : \pi_i \geq 0}.
$$

One finds that, up to an obvious change of variables from $\mu$ to $\eta$ and from $\pi$ to $p$, the LP-B of equation (20) is identical to the LP-A (11). According to theorem 6.1 of [15], solutions of equation (20), or equation (11), are integers, $\forall (i, j) \in \Gamma$, $\mu_{ij}, \eta_{ij} = 0, 1$ and $\forall \sigma_i : \pi_i = 0, 1$.

Summarizing, it was just shown that as $T \to 0$ the BP solution of the FRFI model, understood as the global minimum of the Bethe free energy, is also the ML solution of the model.

doi:10.1088/1742-5468/2008/10/P10016
3. Binary model with totally uni-modular constraints

Consider $N$ binary variables combined in the vector $\sigma = (\sigma_i = 0, 1|i = 1, \ldots, N)$, and associate the following normalized probability with any possible value of the vector:

$$p(\sigma) = Z^{-1} \exp \left( -T^{-1} \sum_i h_i \sigma_i \right) \prod_\alpha \delta \left( \sum_i J_{\alpha i} \sigma_i, m_\alpha \right), \quad (21)$$

where $\delta(x, y)$ is 1 if $x = y$ and it is 0 otherwise; $\alpha = 1, \ldots, M$; matrix $J \equiv (J_{\alpha i} = 0, 1|i = 1, \ldots, N; \alpha = 1, \ldots, M)$ is totally uni-modular (TUM), i.e. the determinant of any square minor of the matrix is 0, 1 or $-1$; the vector $m = (m_\alpha | \alpha = 1, \ldots, M)$ is constructed from positive integers, so $\forall \alpha$, $m_\alpha \leq q_\alpha \equiv \sum_i J_{\alpha i}$. The partition function $Z$ is introduced in equation (21) to guarantee normalization, $\sum_\sigma p(\sigma) = 1$.

The model equation (21) can be viewed as a graphical model defined on the bipartite graph consisting of ‘bits’, $\{i\}$, and ‘checks’, $\{\alpha\}$. Also, there may be other graphical interpretations. Thus, for the weighted matching problem, e.g. as studied in [2], the binary variables in the formulation of equation (21) are associated with edges of the complete bipartite graph. (In this case of the weighted matching, one can show that the resulting matrix of constraints is indeed TUM.)

3.1. Efficient ML solution

We, first of all, observe that the problem of finding the maximum likelihood of equation (21) is equivalent to the following integer programming (IP):

$$\min_\sigma \sum_i h_i \sigma_i \quad \left\{ \begin{array}{l} \forall i: \sigma_i = 0, 1 \setminus \forall \alpha: \sum_i J_{\alpha i} \sigma_i = m_\alpha \end{array} \right. \quad (22)$$

Relaxing the IP to the respective LP-A, with $\sigma_i = 0, 1$ changed to $s_i = [0; 1]$,

$$\min_\sigma \sum_i h_i s_i \quad \left\{ \begin{array}{l} \forall i: 0 \leq s_i \leq 1 \setminus \forall \alpha: \sum_i J_{\alpha i} s_i = m_\alpha \end{array} \right. \quad (23)$$

One finds that the relaxation is tight. In other words, the solutions of the IP problem and the LP problem are exactly equivalent. This is due to the theorem (see e.g. theorem 13.1 of [15]) stating that if $J$ is TUM and $m$ is integer, then all feasible solutions of the LP problem are integer.

3.2. Bethe free energy and BP

Here we discuss the Bethe free energy/belief propagation (BP) approach to the model defined in equation (21). The Bethe free energy functional is

$$F = E - TS, \quad E = \sum_i h_i b_i(1),$$

$$S = \sum_\alpha \sum_{\sigma_\alpha} b_\alpha(\sigma_\alpha) \ln b_\alpha(\sigma_\alpha) - \sum_i (q_\alpha - 1) b_i(\sigma_i) \ln b_i(\sigma_i), \quad (24)$$
where a vector \( \mathbf{\sigma}_\alpha \equiv (\sigma_i \mid \forall i \text{ s.t. } J_{ai} = 1; \sum_i J_{ai}\sigma_i = m_\alpha) \) defines the set of allowed configurations of variables marked by index \( i \) associated and consistent with the given constrained \( \alpha \). For any given \( m_\alpha \) the number of such allowed vectors/ configurations of \( \mathbf{\sigma}_\alpha \) is \( C_{m_\alpha}^{m_\alpha} = m_\alpha! / (m_\alpha - q_\alpha)! q_\alpha! \). As usual, \( b_\alpha(\mathbf{\sigma}_\alpha) \) and \( b_i(\sigma_i) \) are beliefs (estimations for the respective probabilities) associated with the variables and the constraints. The two types of beliefs are related to each other via the following compatibility constraints:

\[
\forall i \quad \text{and} \quad \forall \alpha \text{ s.t. } J_{ai} = 1 : \quad b_i(\sigma_i) = \sum_{\mathbf{\sigma}_\alpha \setminus \sigma_i} b_\alpha(\mathbf{\sigma}_\alpha),
\]

and one should also impose the normalization constraint

\[
\forall i : \quad \sum_{\sigma_i} b_i(\sigma_i) = 1.
\]

Incorporating the compatibility and normalization constraints in the form of Lagrangian multipliers into the variational functional one derives the Lagrangian

\[
\mathcal{L} = F + \sum_i \sum_{\alpha \ni i} \mu_{\alpha i}(\sigma_i) \left( b_i(\sigma_i) - \sum_{\mathbf{\sigma}_\alpha \setminus \sigma_i} b_\alpha(\mathbf{\sigma}_\alpha) \right) + \sum_i \lambda_i(\sigma_i) \left( \sum_{\sigma_i} b_i(\sigma_i) - 1 \right).
\]

Looking for the stationary points of the Lagrangian with respect to all the beliefs and the Lagrangian multipliers, \( \lambda, \mu \), one arrives at the belief propagation equations for the problem.

### 3.3. \( T \to 0 \) limit of the Bethe free energy

In the \( T \to 0 \) limit the entropy term in equation (24) can be dropped and the problem turns into a minimization of the LP type:

\[
\min_{\{b_i,b_\alpha\}} \sum_i b_i(1) \left\{ \forall i \quad \text{and} \quad \forall \alpha \text{ s.t. } J_{ai} = 1 : \quad b_i(\sigma_i) = \sum_{\mathbf{\sigma}_\alpha \setminus \sigma_i} b_\alpha(\mathbf{\sigma}_\alpha) \right. \quad \forall i : \quad \sum_{\sigma_i} b_i(\sigma_i) = 1 \}
\]

It is straightforward to verify that the beliefs associated with \( \alpha \) could be completely removed from equation (28), and the LP problem can be restated solely in terms of the \( i \)-related variables, \( b_i \equiv b_i(1) = 1 - b_i(0) \).

Let us illustrate this point with an example of a single \( \alpha \) constraint with \( m_\alpha = 2 \) and \( q_\alpha = 3 \). Then the set of allowed \( \alpha \)-beliefs are

\[
b_\alpha(1,1,0), b_\alpha(1,0,1), b_\alpha(0,1,1),
\]

and the respective set of relations (25) between \( \beta_1, \beta_2, \beta_3 \) associated with the check \( \alpha \) and the \( \alpha \) beliefs are

\[
\beta_1 = b_\alpha(1,1,0) + b_\alpha(1,0,1), \quad \beta_2 = b_\alpha(1,1,0) + b_\alpha(0,1,1),
\]

\[
\beta_3 = b_\alpha(1,0,1) + b_\alpha(0,1,1).
\]

On the other hand the normalization condition, restated in terms of the \( \alpha \)-beliefs (29), is

\[
b_\alpha(1,1,0) + b_\alpha(1,0,1) + b_\alpha(0,1,1) = 1.
\]
Summing equations (30) and accounting for equation (31), one finds
\[ \beta_1 + \beta_2 + \beta_3 = 2. \]  
(32)

In general, one finds that the relation between \( \beta \) variables associated with an \( \alpha \)-constraint is
\[ \sum_i J_{\alpha i} \beta_i = m_\alpha. \]  
(33)

One derives that equation (28) reduces to a simpler LP-B problem stated solely in terms of the \( \beta \) variables:
\[
\min_{\beta_i} \sum_i h_i \beta_i \quad \left| \begin{array}{l}
\forall i: 0 \leq \beta_i \leq 1 \\
\forall \alpha: \sum_i J_{\alpha i} \beta_i = m_\alpha
\end{array} \right.
\]  
(34)

Furthermore, one observes that, up to re-definition of \( \beta_i \) to \( s_i \), equation (34) is equivalent to equation (17). In other words, we just showed that the \( T \to 0 \) solution of the BP equations, understood as the global minimum of the Bethe free energy, is tight, i.e. it gives exactly the ML solution of the binary model (21).

As a side remark, one notes that it is suggestive to start exploration of the Bethe free energy at finite \( T \) from the LP solution discussed above. It might be especially useful to initiate BP with the (easy to get) LP solution when the Bethe free energy optimization at finite \( T \) is non-convex.

4. Summary and path forward

In this work we discussed easy problems when a zero-temperature BP scheme generates an exact ML result. We argued that this special feature of BP is due to the fact that the related LP optimization is tight (i.e. the LP outputs the ML solution as well). Our consideration was based on the flexibility and convenience provided by the so-called Bethe free energy formulation, naturally relating BP and LP. The results were illustrated with two examples, the FRFI model and the perfect matching model. Also, we briefly discussed a broader class of easy examples related to LP with a TUM matrix of constraints.

We conclude briefly, mentioning some future challenges which follow from our analysis.

- It is useful to continue further exploration of other models of statistical inference with loops allowing computationally efficient optimal solutions. Thus, it would be interesting to find examples of ‘easy’ non-binary problems, and also ones which allow efficient and optimal finite temperature evaluation of marginal probabilities or partition functions. In this context, one mentions the exactness of BP marginals at any temperature known to hold for the continuous variable Gaussian model on an arbitrary graph [21,13] and also the recently established, \( T \to 0 \), relation between an iterative algorithm of BP type and the quadratic optimization problem [14].

- Probably the most intriguing future challenge is that of analysing problems that are not computationally easy, but still close, in some metric, to easy problems. Thus, the models discussed above but considered at finite, not zero, temperature may not allow an explicit efficient solution. Similarly, perturbation of the FRFI model with some
number of graph local frustrations (e.g. some number of randomly thrown negative $J_{ij}$ violating the TUM feature of the model) sets up another ‘close to easy’ problem of theoretical and applied interest. As suggested in [12], BP can be utilized as an efficient heuristic in these ‘close to easy’ cases. Note that in this case finding minima of the Bethe free energy may be a challenge, and the problem turns into the quest for an efficient algorithm for the optimization of non-convex functions [25, 26]. Here novel BP convexification ideas developed in [20, 12, 22, 9] might be helpful. Notice also that the loop calculus approach of [5, 6] is another useful tool which may come handy in perturbative and non-perturbative analysis of these ‘close to easy’ problems.

- BP is the algorithm of choice for decoding of error-correction codes stated in terms of sparse graphs [8]. On the other hand, the above discussion suggests that for BP to decode optimally, or close to optimally, the graphical structure should not necessarily be sparse. Therefore, an intriguing question is: can one design a class of dense codes decoded optimally (or close to optimally) by an algorithm of BP type?

Acknowledgments

The author acknowledges inspiring discussions with V Chernyak, M Vergassola, D Shah, B Shraiman and M Wainwright. The work was carried out under the auspices of the National Nuclear Security Administration of the US Department of Energy at Los Alamos National Laboratory under Contract No DE-AC52-06NA25396. The author also acknowledges the Weston Visiting Professorship Programme supporting his stay at the Weizmann Institute, where the work was completed.

References

[1] Bayati M, Borgs C, Chayes J and Zecchina R, Belief-propagation for weighted b-matchings on arbitrary graphs and its relation to linear programs with integer solutions, 2007 arXiv:0709.1190
[2] Bayati M, Shah D and Sharma M, Max-product for maximum weight matching: Convergence, correctness, and LP duality, 2008 IEEE Trans. Inform. Theory 54 1241
[3] Bethe H A, Statistical theory of superlattices, 1935 Proc. R. Soc. A 150 552
[4] Boros E and Hammer P L, Pseudo-boolean optimization, 2002 Discrete Appl. Math. 123 155
[5] Chertkov M and Chernyak V, Loop calculus in statistical physics and information science, 2006 Phys. Rev. E 73 065102(R)
[6] Chertkov M and Chernyak V, Loop series for discrete statistical models on graphs, 2006 J. Stat. Mech. P06009
[7] Feldman J, Wainwright M and Karger D R, Using linear programming to decode binary linear codes, 2005 IEEE Trans. Inform. Theory 51 954
[8] Gallager R G, 1963 Low Density Parity Check Codes (Cambridge, MA: MIT Press)
[9] Globerson A and Jaakola T, Fixing max-product: convergent message-passing algorithms for map lp-relaxations, 2007 Proc. NIPS
[10] Hartmann A and Rieger H, 2002 Optimization Algorithms in Physics (Berlin: Wiley VCH)
[11] Huang B and Jebraa T, Loopy belief propagation for bipartite maximum weight b-matching, 2007 AISTATS: Proc. of Artificial Intelligence and Statistics
[12] Kolmogorov V and Wainwright M J, On the optimality of tree-reweighted max-product message-passing, 2005 Proc. 21st Ann. Conf. on Uncertainty in Artificial Intelligence (UAI-05) (Arlington, VA: AUAI Press) pp 316–32
[13] Malioutov D M, Johnson J K and Willsky A S, Walk-sums and belief propagation in gaussian graphical models, 2006 J. Mach. Learn. Res. 7 2031
Exactness of belief propagation for some graphical models with loops

[14] Moallemi C C and Van Roy B, *Convergence of the min-sum message passing algorithm for quadratic optimization*, 2006 arXiv:cs/0603058

[15] Papadimitriou H and Steiglitz I, 1998 *Combinatorial Optimization: Algorithms and Complexity* (New York: Dover)

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

[17] Peierls H A, *Ising's model of ferromagnetism*, 1936 Proc. Cambridge Phil. Soc. 32 477

[18] Sanghavi S, Malioutov D M and Willsky A, *Linear programming analysis of loopy belief propagation for weighted matching*, 2007 Proc. NIPS

[19] Wainwright M J and Jordan M I, *Graphical models, exponential families, and variational inference*, 2003 Technical Report 649 UC Berkeley, Department of Statistics

[20] Wainwright M J, Jaakkola T S and Willsky A S, *Tree-based reparametrization framework for approximate estimation on graphs with cycles*, 2003 IEEE Trans. Inform. Theory 49 1120

[21] Weiss Y and Freeman W T, *Correctness of belief propagation in Gaussian graphical models of arbitrary topology*, 2001 Neural Comput. 13 2173

[22] Weiss Y, Yanover C and Melzer T, *Map estimation, linear programming and belief propagation with convex free energies*, 2007 Proc. UAI

[23] Yedidia J S, Freeman W T and Weiss Y, *Constructing free-energy approximations and generalized belief propagation algorithms*, 2005 IEEE Trans. Inform. Theory 51 2282

[24] Yedidia J S, Freeman W T and Weiss Y, 2001 *Generalized Belief Propagation* vol 13 (Cambridge, MA: MIT Press) pp 689–95

[25] Yuille A L, *Cccp algorithms to minimize the Bethe and Kikuchi free energies: convergent alternatives to belief propagation*, 2002 Neural Comput. 14 1691

[26] Yuille A L and Rangarajan A, *The concave–convex procedure*, 2003 Neural Comput. 15 915

doi:10.1088/1742-5468/2008/10/P10016