Convergence and Correctness of Max-Product Belief Propagation for Linear Programming*

Sejun Park Jinwoo Shin†

Abstract

The max-product Belief Propagation (BP) is a popular message-passing heuristic for approximating a maximum-a-posteriori (MAP) assignment in a joint distribution represented by a graphical model (GM). In the past years, it has been shown that BP can solve a few classes of Linear Programming (LP) formulations to combinatorial optimization problems including maximum weight matching, shortest path and network flow, i.e., BP can be used as a message-passing solver for certain combinatorial optimizations. However, those LPs and corresponding BP analysis are very sensitive to underlying problem setups, and it has been not clear what extent these results can be generalized to. In this paper, we obtain a generic criteria that BP converges to the optimal solution of given LP, and show that it is satisfied in LP formulations associated to many classical combinatorial optimization problems including maximum weight perfect matching, shortest path, traveling salesman, cycle packing, vertex/edge cover and network flow.

1 Introduction

The max-product belief propagation (BP) is the most popular heuristic for approximating a maximum-a-posteriori (MAP) assignment[1] of given Graphical model (GM) [28, 17, 16, 25], where its performance has been not well understood in loopy GMs, i.e., GM with cycles. Nevertheless, BP often shows remarkable performances even on loopy GM. Distributed implementation, associated ease of programming and strong parallelization potential are the main reasons for the growing popularity of the BP algorithm. For example, several software architectures for implementing parallel BPs were recently proposed [14, 10, 15].

In the past years, there have been made extensive research efforts to understand BP performances on loopy GMs under connections to combinatorial optimization [3, 20, 11, 19, 2, 23, 18, 9, 6, 1, 21]. In particular, it has been studied about the BP convergence to the correct answer under a few classes of loopy GM formulations of combinatorial optimization problems: matching [3, 20, 11, 19], perfect matching [2], matching with odd cycles [23], shortest path [18] and network flow [9]. The important common feature of these instances is that BP converges to a correct MAP assignment if the Linear Programming (LP) relaxation of the MAP inference problem is tight, i.e., it has no integrality gap. In other words,

*The preliminary conference version of this work was presented at Conference on Uncertainty in Artificial Intelligence (UAI), 2015

†S. Park and J. Shin are with School of Electrical Engineering, Korea Advanced Institute of Science & Technology, Republic of Korea, Daejeon 34141, Republic of Korea. Email: sejun.park@kaist.ac.kr, jinwoos@kaist.ac.kr

[1] In general, MAP is NP-hard to compute exactly [7].

BP can be used as an efficient distributed solver for those LPs, and is presumably a better choice than classical centralized LP solvers such as simplex methods [8], interior point methods [24] and ellipsoid methods [13] for large-scale inputs. However, these theoretical results on BP are very sensitive to underlying structural properties depending on specific problems and it is not clear what extent they can be generalized to, e.g., the BP analysis for matching problems [3, 20, 11, 19] does not extend to even for perfect matching ones [2]. In this paper, we overcome such technical difficulties for enhancing the power of BP as a LP solver.

1.1 Contribution

We establish a generic criteria for GM formulations of given LP so that BP converges to the optimal LP solution. By product, it also provides a sufficient condition for unique BP fixed point. As one can naturally expect given prior results, one of our conditions requires the LP tightness. Our main contribution is finding other sufficient generic conditions so that BP converges to the correct MAP assignment of GM. First of all, our generic criteria can rediscover all prior BP results on this line, including matching [3, 20, 11], perfect matching [2], matching with odd cycles [23] and shortest path [18], i.e., we provide a unified framework on establishing the convergence and correctness of BPs in relation to associated LPs. Furthermore, we provide new instances under our framework: we show that BP can solve LP formulations associated to other popular combinatorial optimizations including perfect matching with odd cycles, traveling salesman, cycle packing, network flow and vertex/edge cover, which are not known in the literature. Here, we remark that the same network flow problem was already studied using BP by Gamarnik et al. [9]. However, our BP is different from theirs and much simpler to implement/analyze: the authors study BP on continuous GMs, and we do BP on discrete GMs. While most prior known BP results on this line focused on the case when the associated LP has an integral solution, the proposed criteria naturally guides the BP design to compute fractional LP solutions as well (see Section 3.3 and Section 3.5 for details).

Our proof technique is built on that of [20] where the authors construct an alternating path in the computational tree induced by BP to analyze its performance for the maximum weight matching problem. Such a trick needs specialized case studies depending on the associated LP when the path reaches a leaf of the tree, and this is one of main reasons why it is not easy to generalize to other problems beyond matching. The main technical contribution of this paper is providing a way to avoid the issue in the BP analysis via carefully analyzing associated LP polytopes. The main appeals of our results are providing not only tools on BP analysis, but also guidelines on BP design for its high performance, i.e., one can carefully design a BP given LP so that it satisfies the proposed criteria. Our results provide not only new tools on BP analysis and design, but also new directions on efficient distributed (and parallel) solvers for large-scale LPs and combinatorial optimization problems.

1.2 Organization

In Section 2, we introduce necessary backgrounds for the BP algorithm. In Section 3, we provide the main result of the paper as well as its several concrete applications to popular combinatorial optimizations. Proofs are presented in Section 4 and Section 5.

2 Preliminaries

2.1 Graphical Model

A joint distribution of n (binary) random variables $Z = [Z_i] \in \{0,1\}^n$ is called a Graphical Model (GM) if it factorizes as follows: for $z = [z_i] \in \Omega^n$,

$$\Pr[Z = z] \;\propto\; \prod_{i \in \{1,\ldots,n\}} \psi_i(z_i) \prod_{\alpha \in F} \psi_\alpha(z_\alpha),$$

where $\{\psi_i, \psi_\alpha\}$ are (given) non-negative functions, the so-called factors; F is a collection of subsets

$$F = \{\alpha_1, \alpha_2, ..., \alpha_k\} \subset 2^{\{1,2,...,n\}}$$

(each α_j is a subset of $\{1,2,\ldots,n\}$ with $|\alpha_j| \geq 2$); z_α is the projection of z onto dimensions included in α.[2] In particular, ψ_i is called a variable factor. Figure 1 depicts the graphical relation between factors F and variables z.

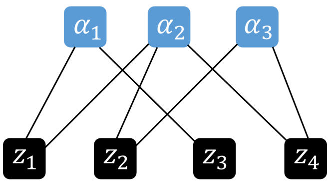

Figure 1: Factor graph for the graphical model with $F = \{\alpha_1, \alpha_2, \alpha_3\}$ and $n = 4$: $\Pr[z] \propto \psi_{\alpha_1}(z_1, z_3)\psi_{\alpha_2}(z_1, z_2, z_4)\psi_{\alpha_3}(z_2, z_3, z_4)$. Each α_j selects a subset of z, e.g., α_1 selects $\{z_1, z_3\}$.

Assignment z^* is called a maximum-a-posteriori (MAP) assignment if z^* satisfies $z^* = \arg\max_{z \in \{0,1\}^n} \Pr[z]$. This means that computing a MAP assignment requires comparing $\Pr[z]$ for all possible z, which is typically computationally intractable (i.e., NP-hard) unless the induced bipartite graph of factors F and variables z, so-called factor graph, has a bounded treewidth [7].

2.2 Max-Product Belief Propagation

The (max-product) Belief Propagation (BP) algorithms are popular heuristics for approximating the MAP assignment in a graphical model. BP is an iterative procedure; at each iteration t, there are four messages

$$\{m^t_{\alpha \to i}(c), m^t_{i \to \alpha}(c) : c \in \{0,1\}\}$$

between every variable z_i and every associated $\alpha \in F_i$, where $F_i := \{\alpha \in F : i \in \alpha\}$; that is, F_i is a subset of F such that all α in F_i include the i^{th} position of z for any given z. Then, messages are updated as follows:

$$m^{t+1}_{\alpha \to i}(c) \;=\; \max_{z_\alpha : z_i = c} \psi_\alpha(z_\alpha) \prod_{j \in \alpha \setminus i} m^t_{j \to \alpha}(z_j) \qquad (1)$$

$$m^{t+1}_{i \to \alpha}(c) \;=\; \psi_i(c) \prod_{\alpha' \in F_i \setminus \alpha} m^t_{\alpha' \to i}(c). \qquad (2)$$

[2] For example, if $z = [0,1,0]$ and $\alpha = \{1,3\}$, then $z_\alpha = [0,0]$.

First, we note that each z_i only sends messages to F_i; that is, z_i sends messages to α_j only if α_j selects/includes i. The outer-term in the message computation (1) is maximized over all possible $z_\alpha \in \{0,1\}^{|\alpha|}$ with $z_i = c$. The inner-term is a product that only depends on the variables z_j (excluding z_i) that are connected to α. The message-update (2) from a variable z_i to a factor ψ_α is a product which considers all messages received by ψ_α in the previous iteration, except for the message sent by z_i itself.

One can reduce the complexity of messages by combining (1) and (2) as:

$$m_{i\to\alpha}^{t+1}(c) = \psi_i(c) \prod_{\alpha'\in F_i\setminus\alpha} \max_{z_{\alpha'}:z_i=c} \psi_{\alpha'}(z_{\alpha'}) \prod_{j\in\alpha'\setminus i} m_{j\to\alpha'}^t(z_j),$$

which we analyze in this paper. Finally, given a set of messages $\{m_{i\to\alpha}(c), m_{\alpha\to i}(c) : c \in \{0,1\}\}$, the so-called BP marginal beliefs are computed as follows:

$$b_i[z_i] \;=\; \prod_{\alpha\in F_i} m_{\alpha\to i}(z_i). \tag{3}$$

Then, the BP algorithm outputs $z^{BP} = [z_i^{BP}]$ as

$$z_i^{BP} = \begin{cases} 1 & \text{if } b_i[1] > b_i[0] \\ ? & \text{if } b_i[1] = b_i[0] \\ 0 & \text{if } b_i[1] < b_i[0] \end{cases}.$$

It is known that z^{BP} converges to a MAP assignment after a large enough number of iterations, if the factor graph is a tree and the MAP assignment is unique. However, if the graph has loops in it, the BP algorithm has no guarantee to find a MAP assignment in general.

3 Convergence and Correctness of Belief Propagation

3.1 Convergence and Correctness Criteria of BP

In this section, we provide the main result of this paper: a convergence and correctness criteria of BP. Consider the following GM: for $x = [x_i] \in \{0,1\}^n$ and $w = [w_i] \in \mathbb{R}^n$,

$$\Pr[X = x] \;\propto\; \prod_i e^{-w_i x_i} \prod_{\alpha\in F} \psi_\alpha(x_\alpha), \tag{4}$$

where F is the set of non-variable factors and the factor function ψ_α for $\alpha \in F$ is defined as

$$\psi_\alpha(x_\alpha) = \begin{cases} 1 & \text{if } A_\alpha x_\alpha \geq b_\alpha \\ 0 & \text{otherwise} \end{cases},$$

for some matrices A_α and vectors b_α. Now we consider the Linear Programming (LP) corresponding the above GM:

$$\begin{aligned} \text{minimize} \quad & w \cdot x \\ \text{subject to} \quad & A_\alpha x \geq b_\alpha \text{ for all } \alpha \in F \\ & x = [x_i] \in [0,1]^n \end{aligned} \tag{5}$$

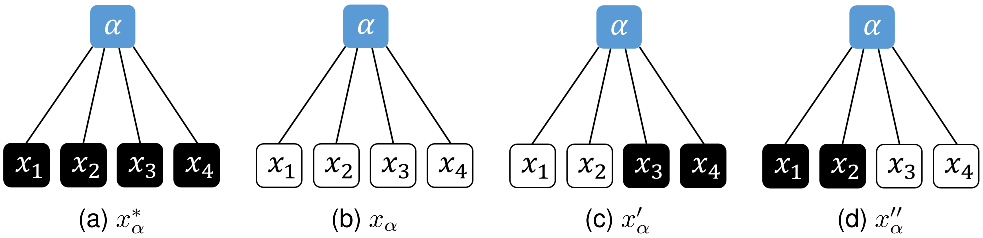

Figure 2: Illustration of Condition *C3* of Theorem 1 under $i = 1, \gamma = \{2\}$ and $\psi_\alpha(x_\alpha) = 1$ (i.e., say ψ_α satisfies) if and only if $\sum_{i\in\alpha} x_i = 1$. All four variables $x^*_\alpha, x_\alpha, x'_\alpha, x''_\alpha$ must satisfy ψ_α. For example, let $x^*_\alpha = (1, 0, 0, 0)$ and $x_\alpha = (0, 1, 0, 0)$. Then, both $x'_\alpha = (x^*_1, x^*_2, x_3, x_4)$ and $x''_\alpha = (x_1, x_2, x^*_3, x^*_4)$ satisfy ψ_α.

To simplify the notation, we often use $Ax \geq b$ with $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$ which includes all inequalities $A_\alpha x \geq b_\alpha$ and $x \in [0, 1]^n$. Without loss of generality, we assume that $\|A_{i*}\|_2 = 1$ for all $i = 1, 2, \ldots, m$ throughout this paper, where A_{i*} is the i-th row of A. Similarly, we denote A_{*i} as the i-th column of A. One can easily observe that the MAP assignments for GM (4) corresponds to the (optimal) solution of LP (5) if the LP has an integral solution $x^* \in \{0, 1\}^n$. Furthermore, if the solution of LP (5) is unique, the there exists a positive constant ρ satisfying the following identity:

$$\rho := \inf_{x\in\mathcal{P}\setminus x^*} \frac{w \cdot x - w \cdot x^*}{\|x - x^*\|_1} > 0.$$

Using the notation and observation, we establish the following sufficient conditions so that the max-product BP can indeed find the LP solution.

Theorem 1 *Suppose the following conditions hold:*

C1. LP (5) has a unique integral solution $x^* \in \{0, 1\}^n$, i.e., it is tight.

C2. For every $i \in \{1, 2, \ldots, n\}$, the number of factors associated with x_i is at most two, i.e., $|F_i| \leq 2$.

C3. For every factor ψ_α, every $x_\alpha \in \{0, 1\}^{|\alpha|}$ with $\psi_\alpha(x_\alpha) = 1$, and every $i \in \alpha$ with $x_i \neq x^*_i$, there exists $\gamma \subset \alpha$ such that

$$|\{j \in \{i\} \cup \gamma : |F_j| = 2\}| \leq 2$$

$$\psi_\alpha(x'_\alpha) = 1, \qquad \text{where } x'_k = \begin{cases} x_k & \text{if } k \notin \{i\} \cup \gamma \\ x^*_k & \text{otherwise} \end{cases}.$$

$$\psi_\alpha(x''_\alpha) = 1, \qquad \text{where } x''_k = \begin{cases} x_k & \text{if } k \in \{i\} \cup \gamma \\ x^*_k & \text{otherwise} \end{cases}.$$

Then the max-product BP on GM (4) with arbitrary initial message converges to the solution of LP (5) in $\left(\frac{w_{\max}}{\rho} + 1\right) K$ iterations, where[3]

$$K = \max_{\xi\subset\{1,\ldots,m\}:|\xi|=n,\det(A_\xi)\neq 0} \|(A_\xi)^{-1}\mathbf{1}\|_1 \qquad \text{and} \qquad w_{\max} = \max_j |w_j|.$$

[3] A_ξ is a square matrix consisting of rows of A corresponding to the row index set ξ, and $\mathbf{1}$ is the vector consisting of ones.

We refer Figure 2 illustrating Condition *C3*. Since Theorem 1 holds for arbitrary initial messages, the conditions *C1, C2, C3* also provides the uniqueness of BP fixed points, as stated in what follows.

Corollary 2 *The max-product BP on GM* (4) *has a unique fixed point if conditions C1, C2, C3 hold.*

The conditions *C2, C3* are typically easy to check given GM (4) and the uniqueness in *C1* can be easily guaranteed via adding random noises, On the other hand, the integral property in *C1* requires to analyze LP (5), where it has been extensively studied in the field of combinatorial optimization [22]. Nevertheless, Theorem 1 provides important guidelines to design BP algorithms, irrespectively of the LP analysis.

In the following sections, we introduce concrete instances of LPs satisfying the conditions of Theorem 1 so that BP correctly converges to its optimal solution. Specifically, we consider LP formulations associated to several combinatorial optimization problems including shortest path, maximum weight perfect matching, traveling salesman, maximum weight disjoint vertex cycle packing, vertex/edge cover and network flow. We note that the shortest path result, i.e., Corollary 3, is known [18], where we rediscover it as a corollary of Theorem 1. Our other results, i.e., Corollaries 4-10, are new and what we first establish in this paper.

Furthermore, all examples we consider in the following sections have the following common feature: each row of A contains non-zero entries $A_{ij_1}, \dots, A_{ij_k}$ of the same absolute value, i.e., $|A_{ij_1}| = \cdots = |A_{ij_k}|$, and each column of A does at most two non-zero entries. For such A, one can prove that $K \le n^{2.5}$. To see this, consider a 'scaled' matrix $\widetilde{A}$ such that $\widetilde{A}_{i*} = c_i A_{i*}$ for some constant c_i so that every entry of $\widetilde{A}$ is in $\{0, \pm 1\}$. Since we assume $\|A_{i*}\|_2 = 1$, we have $c_i \le \sqrt{n}$. Then, for any $n \times n$ invertible submatrix $\widetilde{A}_\xi$ of $\widetilde{A}$, it is known [5] that every entry of $(\widetilde{A}_\xi)^{-1}$ is in $\{0, \pm 1/2, \pm 1\}$. This observation leads to the following bound on K:

$$
\begin{aligned}
K &= \max_{\xi \subset \{1,\dots,m\} : |\xi| = n, \det(A_\xi) \neq 0} \|(A_\xi)^{-1}\mathbf{1}\|_1 \\
&\le \max_{\xi \subset \{1,\dots,m\} : |\xi| = n, \det(A_\xi) \neq 0} \big(\max_i c_i\big) \|(\widetilde{A}_\xi)^{-1}\mathbf{1}\|_1 \\
&\le n^{2.5}
\end{aligned}
$$

since $((A_\xi)^{-1})_{*i} = c_i((\widetilde{A}_\xi)^{-1})_{*i}$ and $c_i \le \sqrt{n}$.

3.2 Example I: Shortest Path

Given a directed graph $G = (V, E)$ and non-negative edge weights $w = [w_e : e \in E] \in \mathbb{R}_+^{|E|}$, the shortest path problem is to find the shortest path from the source s to the destination t: it minimizes the sum of edge weights along the path. One can naturally design the following LP for this problem:

$$
\begin{aligned}
&\text{minimize} && w \cdot x \\
&\text{subject to} && \sum_{e \in \delta^o(v)} x_e - \sum_{e \in \delta^i(v)} x_e = \begin{cases} 1 & \text{if } v = s \\ -1 & \text{if } v = t \\ 0 & \text{otherwise} \end{cases} \qquad (6) \\
& && x = [x_e] \in [0,1]^{|E|}.
\end{aligned}
$$

where $\delta^i(v), \delta^o(v)$ are sets of incoming, outgoing edges of v. It is known that the above LP always has an integral solution, i.e., the shortest path from s to t. We consider the following GM for LP (6):

$$\Pr[X = x] \;\propto\; \prod_{e\in E} e^{-w_e x_e} \prod_{v\in V} \psi_v(x_{\delta(v)}), \tag{7}$$

where $\delta(v) = \delta^i(v) \cup \delta^o(v)$ and the factor function ψ_v is defined as

$$\psi_v(x_{\delta(v)}) = \begin{cases} 1 & \text{if } \sum_{e\in\delta^o(v)} x_e - \sum_{e\in\delta^i(v)} x_e \\ & \qquad = \begin{cases} 1 & \text{if } v = s \\ -1 & \text{if } v = t \\ 0 & \text{otherwise} \end{cases} \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (7), one can easily check Conditions *C2, C3* of Theorem 1 hold and derive the following corollary whose formal proof is presented in Section 5.1.

Corollary 3 *If the shortest path from s to t, i.e., the solution of the shortest path LP* (6)*, is unique, then the max-product BP on GM* (7) *converges in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.*

The uniqueness condition in the above corollary is easy to guarantee by adding small random noises to edge weights.

3.3 Example II: Maximum Weight Perfect Matching

Given an undirected graph $G = (V, E)$ and non-negative edge weights $w = [w_e : e \in E] \in \mathbb{R}_+^{|E|}$ on edges, the maximum weight perfect matching problem is to find a set of edges such that each vertex is connected to exactly one edge in the set and the sum of edge weights in the set is maximized. One can naturally design the following LP for this problem:

$$\begin{aligned} \text{maximize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e\in\delta(v)} x_e = 1 \\ & x = [x_e] \in [0,1]^{|E|}. \end{aligned} \tag{8}$$

where $\delta(v)$ is the set of edges connected to a vertex v. If the above LP has an integral solution, it corresponds to the solution of the maximum weight perfect matching problem.

It is known that the maximum weight matching LP (8) always has a half-integral solution $x^* \in \{0, \frac{1}{2}, 1\}^{|E|}$. We will design BP for obtaining the half-integral solution. To this end, duplicate each edge e to e_1, e_2 and define a new graph $G' = (V, E')$ where $E' = \{e_1, e_2 : e \in E\}$. Then, we suggest the following equivalent LP that always have an integral solution:

$$\begin{aligned} \text{maximize} \quad & w' \cdot x \\ \text{subject to} \quad & \sum_{e_i\in\delta(v)} x_{e_i} = 2 \\ & x = [x_{e_i}] \in [0,1]^{|E'|}. \end{aligned} \tag{9}$$

where $w'_{e_1} = w'_{e_2} = w_e$. One can easily observe that solving LP (9) is equivalent to solving LP (8) due to our construction of G' and w'. Now, construct the following GM for LP (9):

$$\Pr[X = x] \;\propto\; \prod_{e_i \in E'} e^{w'_{e_i} x_{e_i}} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{10}$$

where the factor function ψ_v is defined as

$$\psi_v(x_{\delta(v)}) = \begin{cases} 1 & \text{if } \sum_{e_i \in \delta(v)} x_{e_i} = 2 \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (10), we derive the following corollary of Theorem 1 whose formal proof is presented in Section 5.2.

Corollary 4 *If the solution of the maximum weight perfect matching LP* (9) *is unique, then the max-product BP on GM* (10) *converges in* $O(w_{\max}|E|^{2.5}/\rho)$ *iterations.*

Again, the uniqueness condition in the above corollary is easy to guarantee by adding small random noises to edge weights $[w'_{e_i}]$. We note that it is known [2] that BP converges to the unique and integral solution of LP (8), while Corollary 4 implies that BP can solve it without the integrality condition of LP (8) by solving GM (10). We note that one can easily obtain a similar result for the maximum weight (non-perfect) matching problem, where we omit the details in this paper.

3.4 Example III: Maximum Weight Perfect Matching with Odd Cycles

In previous section we prove that BP converges to the optimal (possibly, fractional) solution of LP (9), equivalently LP (8). One can add odd cycle (also called Blossom) constraints and make those LPs tight i.e. solves the maximum weight perfect matching problem:

$$\begin{aligned} \text{maximize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e \in \delta(v)} x_e = 1, \quad \forall v \in V \\ & \sum_{e \in C} x_e \le \frac{|C|-1}{2}, \quad \forall C \in \mathcal{C}, \\ & x = [x_e] \in [0,1]^{|E|}. \end{aligned} \tag{11}$$

where $\mathcal{C}$ is a set of odd cycles in G. The authors [23] study BP for solving LP (11) by replacing $\sum_{e \in \delta(v)} x_e = 1$ by $\sum_{e \in \delta(v)} x_e \le 1$, i.e., for the maximum weight (non-perfect) matching problem. Using Theorem 1, one can extend the result to the maximum weight perfect matching problem, i.e., solving LP (11). To this end, we follow the approach [23] and construct the following graph $G' = (V', E')$ and weight $w' = [w'_e : e \in E'] \in \mathbb{R}^{|E'|}$ given set $\mathcal{C}$ of disjoint odd cycles:

$$\begin{aligned} V' &= V \cup \{v_C : C \in \mathcal{C}\} \\ E' &= \{(u, v_C) : u \in C, C \in \mathcal{C}\} \cup E \setminus \{e \in C : C \in \mathcal{C}\} \end{aligned}$$

$$w'_e = \begin{cases} \frac{1}{2} \sum_{e' \in E(C)} (-1)^{d_C(u,e')} w_{e'} & \text{if } e = (u, v_C) \\ & \text{for some } C \in \mathcal{C}, \\ w_e & \text{otherwise} \end{cases}$$

where $d_C(u, e')$ is the graph distance between u, e' in cycle C. Then, LP (11) is equivalent to the following LP:

$$\begin{aligned}
\text{maximize} \quad & w' \cdot y \\
\text{subject to} \quad & \sum_{e \in \delta(v)} y_e = 1, && \forall\, v \in V \\
& \sum_{u \in V(C)} (-1)^{d_C(u,e)} y_{(v_C,u)} \in [0,2], \ \forall e \in E(C) \\
& \sum_{e \in \delta(v_C)} y_e \le |C| - 1, && \forall C \in \mathcal{C} \\
& y = [y_e] \in [0,1]^{|E'|}.
\end{aligned} \tag{12}$$

Now, we construct the following GM from the above LP:

$$\Pr[Y = y] \;\propto\; \prod_{e \in E} e^{w_e y_e} \prod_{v \in V} \psi_v(y_{\delta(v)}) \prod_{C \in \mathcal{C}} \psi_C(y_{\delta(v_C)}), \tag{13}$$

where the factor function ψ_v, ψ_C is defined as

$$\psi_v(y_{\delta(v)}) = \begin{cases} 1 & \text{if } \sum_{e \in \delta(v)} y_e = 1 \\ 0 & \text{otherwise} \end{cases},$$

$$\psi_C(y_{\delta(v_C)}) = \begin{cases} 1 & \text{if } \sum_{u \in V(C)} (-1)^{d_C(u,e)} y_{(v_C,u)} \in \{0,2\} \\ & \quad \sum_{e \in \delta(v_C)} y_e \le |C| - 1 \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (13), we derive the following corollary of Theorem 1 whose formal proof is presented in Section 5.3.

Corollary 5 *If the solution of the maximum weight perfect matching with odd cycles LP (12) is unique and integral, then the max-product BP on GM (13) converges in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.*

We again emphasize that a similar result for the maximum weight (non-perfect) matching problem was established in [23]. However, the proof technique in the paper does not extend to the perfect matching problem. This is in essence because presumably the perfect matching problem is harder than the non-perfect matching one. Under the proposed generic criteria of Theorem 1, we overcome the technical difficulty.

3.5 Example IV: Vertex Cover

Given an undirected graph $G = (V, E)$ and non-negative integer vertex weights $b = [b_v : v \in V] \in \mathbb{Z}_+^{|V|}$, the vertex cover problem is to find a set of vertices minimizes the sum of vertex weights in the set such that each edge is connected to at least one vertex in it. This problem is one of Karp's 21 NP-complete problems [12]. The associated LP formulation to the vertex cover problem is as follows:

$$\begin{aligned}
\text{minimize} \quad & b \cdot y \\
\text{subject to} \quad & y_u + y_v \ge 1 \\
& y = [y_v] \in [0,1]^{|V|}.
\end{aligned} \tag{14}$$

However, if we design a GM from the above LP, it does not satisfy conditions in Theorem 1. Instead, we will show that BP can solve the following dual LP:

$$\begin{aligned} \text{maximize} \quad & \sum_{e \in E} x_e \\ \text{subject to} \quad & \sum_{e \in \delta(v)} x_e \le b_v \\ & x = [x_e] \in \mathbb{R}_+^{|E|}. \end{aligned} \tag{15}$$

Note that the above LP always has a half-integral solution. As we did in Section 3.3, one can duplicate edges, i.e., $E' = \{e_1, \dots, e_{2b_{\max}} : e \in E\}$ with $b_{\max} = \max_v b_v$, and design the following equivalent LP having an integral solution:

$$\begin{aligned} \text{maximize} \quad & w' \cdot x \\ \text{subject to} \quad & \sum_{e_i \in \delta(v)} x_{e_i} \le 2b_v, \quad \forall v \in V \\ & x = [x_{e_i}] \in [0,1]^{|E'|} \end{aligned}, \tag{16}$$

where $w'_{e_i} = w_e$ for $e \in E$ and its copy $e_i \in E'$. From the above LP, we can construct the following GM:

$$\Pr[X = x] \propto \prod_{e_i \in E'} e^{w'_{e_i} x_{e_i}} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{17}$$

where the factor function ψ_v is defined as

$$\psi_v(x_{\delta(v)}) = \begin{cases} 1 & \text{if } \sum_{e_i \in \delta(v)} x_{e_i} \le 2b_v \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (17), we derive the following corollary of Theorem 1 whose formal proof is presented in Section 5.4.

Corollary 6 *If the solution of the vertex cover dual LP* (16) *is unique, then the max-product BP on GM* (17) *converges in* $O(w_{\max}|E|^{2.5}/\rho)$ *iterations.*

Again, the uniqueness condition in the above corollary is easy to guarantee by adding small random noises to edge weights $[w'_{e_i}]$. We further remark that if the solution of the primal LP (14) is integral, then it can be easily found from the solution of the dual LP (16) using the strictly complementary slackness condition [4] .

3.6 Example V: Edge Cover

Given an undirected graph $G = (V, E)$ and non-negative edge weights $w = [w_e : e \in E] \in \mathbb{R}_+^{|E|}$ on edges, the minimum weight edge cover problem is to find a set of edges such that each vertex is connected to at least one edge in the set and the sum of edge weights in the set is minimized. One can naturally design the following LP for this problem:

$$\begin{aligned} \text{minimize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e \in \delta(v)} x_e \ge 1 \\ & x = [x_e] \in [0,1]^{|E|}. \end{aligned} \tag{18}$$

where $\delta(v)$ is the set of edges connected to a vertex v. If the above LP has an integral solution, it corresponds to the solution of the minimum weight edge cover problem.

Similarly as the case of matching, it is known that the minimum weight edge cover LP (18) always has a half-integral solution $x^* \in \{0, \frac{1}{2}, 1\}^{|E|}$. We will design BP for obtaining the half-integral solution. To this end, duplicate each edge e to e_1, e_2 and define a new graph $G' = (V, E')$ where $E' = \{e_1, e_2 : e \in E\}$. Then, we suggest the following equivalent LP that always have an integral solution:

$$\begin{aligned} \text{minimize} \quad & w' \cdot x \\ \text{subject to} \quad & \sum_{e_i \in \delta(v)} x_{e_i} \geq 2 \\ & x = [x_{e_i}] \in [0,1]^{|E'|}. \end{aligned} \tag{19}$$

where $w'_{e_1} = w'_{e_2} = w_e$. One can easily observe that solving LP (19) is equivalent to solving LP (18) due to our construction of G' and w'. Now, construct the following GM for LP (19):

$$\Pr[X = x] \;\propto\; \prod_{e_i \in E'} e^{-w'_{e_i} x_{e_i}} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{20}$$

where the factor function ψ_v is defined as

$$\psi_v(x_{\delta(v)}) = \begin{cases} 1 & \text{if } \sum_{e_i \in \delta(v)} x_{e_i} \geq 2 \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (20), we derive the following corollary of Theorem 1 whose formal proof is presented in Section 5.5.

Corollary 7 *If the solution of the minimum weight edge cover LP (9) is unique, then the max-product BP on GM (20) converges in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.*

Again, the uniqueness condition in the above corollary is easy to guarantee by adding small random noises to edge weights $[w'_{e_i}]$.

3.7 Example VI: Traveling Salesman

Given a directed graph $G = (V, E)$ and non-negative edge weights $w = [w_e : e \in E] \in \mathbb{R}_+^{|E|}$, the traveling salesman problem (TSP) is to find the minimum weight Hamiltonian cycle in G. The natural LP formulation to TSP is the following:

$$\begin{aligned} \text{minimize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e \in \delta(v)} x_e = 2 \\ & x = [x_e] \in [0,1]^{|E|}. \end{aligned} \tag{21}$$

From the above LP, one can construct the following GM:

$$\Pr[X = x] \;\propto\; \prod_{e \in E} e^{-w_e x_e} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{22}$$

where the factor function ψ_v is defined as

$$\psi_v(x_{\delta(v)}) = \begin{cases} 1 & \text{if } \sum_{e\in\delta(v)} x_e = 2 \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (22), we derive the following corollary of Theorem 1 whose formal proof is presented in Section 5.6.

Corollary 8 *If the solution of the traveling salesman LP* (21) *is unique and integral, then the max-product BP on GM* (22) *converges in* $O(w_{\max}|E|^{2.5}/\rho)$ *iterations.*

Again, the uniqueness condition in the above corollary is easy to guarantee by adding small random noises to edge weights.

3.8 Example VII: Maximum Weight Cycle Packing

Given an undirected graph $G = (V, E)$ and non-negative edge weights $w = [w_e : e \in E] \in \mathbb{R}_+^{|E|}$, the maximum weight vertex disjoint cycle packing problem is to find the maximum weight set of cycles with no common vertex. It is easy to observe that it is equivalent to find a subgraph maximizing the sum of edge weights on it such that each vertex of the subgraph has degree 2 or 0. The natural LP formulation to this problem is following:

$$\begin{aligned} &\text{maximize} && w \cdot x \\ &\text{subject to} && \sum_{e\in\delta(v)} x_e = 2y_v \\ &&& x = [x_e] \in [0,1]^{|E|}, y = [y_v] \in [0,1]^{|V|}. \end{aligned} \tag{23}$$

From the above LP, one can construct the following GM:

$$\Pr[X = x, Y = y] \propto \prod_{e\in E} e^{w_e x_e} \prod_{v\in V} \psi_v(x_{\delta(v)}, y_v), \tag{24}$$

where the factor function ψ_v is defined as

$$\psi_v(x_{\delta(v)}, y_v) = \begin{cases} 1 & \text{if } \sum_{e\in\delta(v)} x_e = 2y_v \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (24), we derive the following corollary of Theorem 1 whose formal proof is presented in Section 5.7.

Corollary 9 *If the solution of maximum weight vertex disjoint cycle packing LP* (23) *is unique and integral, then the max-product BP on GM* (24) *converges in* $O(w_{\max}|E|^{2.5}/\rho)$ *iterations.*

Again, the uniqueness condition in the above corollary is easy to guarantee by adding small random noises to edge weights.

3.9 Example VIII: Minimum Cost Network Flow

Given a directed graph $G = (V, E)$, supply/demand $d = [d_v] \in \mathbb{Z}^{|E|}$ and capacity $c = [c_e : e \in E] \in \mathbb{Z}_+^{|E|}$, the minimum cost network flow problem can be forumlated by the following LP.

$$\begin{aligned} \text{minimize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e \in \delta^o(v)} x_e - \sum_{e \in \delta^i(v)} x_e = d_v \\ & x_e \le c_e \\ & x = [x_e] \in \mathbb{R}_+^{|E|}, \end{aligned} \tag{25}$$

where $\delta^i(v), \delta^o(v)$ are the set of incoming, outgoing edges of v. It is known that the above LP always has an integral solution. We will design BP for obtaining the solution of LP (25). To this end, duplicate each edge e to $e_1, \dots, e_{c_e}$ and define a new graph $G' = (V, E')$ where $E' = \{e_1, \dots, e_{c_e} : e \in E\}$. Then, we suggest the following equivalent LP that always have an integral solution:

$$\begin{aligned} \text{minimize} \quad & w' \cdot x \\ \text{subject to} \quad & \sum_{e_i \in \delta^o(v)} x_{e_i} - \sum_{e_i \in \delta^i(v)} x_{e_i} = d_v \\ & x = [x_{e_i}] \in [0, 1]^{|E'|}. \end{aligned} \tag{26}$$

where $w'_{e_1} = \cdots = w'_{e_{c_e}} = w_e$. One can easily observe that solvin LP (25) is equivalent to solving LP (26) due to our construction of G' and w'. Now, construct the following GM for LP (26):

$$\Pr[X = x] \;\propto\; \prod_{e_i \in E'} e^{-w'_{e_i} x_{e_i}} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{27}$$

where the factor function ψ_v is defined as

$$\psi_v(x_{\delta(v)}) = \begin{cases} 1 & \text{if } \sum_{e_i \in \delta^o(v)} x_{e_i} - \sum_{e_i \in \delta^i(v)} x_{e_i} = d_v \\ 0 & \text{otherwise} \end{cases}.$$

For the above GM (27), one can easily check Conditions *C2, C3* of Theorem 1 hold and derive the following corollary whose formal proof is presented in Section 5.8.

Corollary 10 *If the shortest path from s to t, i.e., the solution of the network flow LP (25), is unique, then the max-product BP on GM (27) converges in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.*

Gamarnik et al. [9] also studied the convergence and correct of BP on the minimum cost network flow problem. However, they studied BP on GM of continuous variables while our analysis is for BP on GM of binary variables. For practical purposes, the latter is easier to run than the former.

4 Proof of Theorem 1

To begin with, we define some necessary notation. We let $\mathcal{P}$ denote the polytope of feasible solutions of LP (5):

$$\mathcal{P} := \{x \in [0,1]^n \;:\; \psi_\alpha(x_\alpha) = 1,\ \forall\, \alpha \in F\}\,.$$

Similarly, $\mathcal{P}_\alpha$ is defined as

$$\mathcal{P}_\alpha := \left\{x \in [0,1]^{|\alpha|} \;:\; \psi_\alpha(x_\alpha) = 1\right\}.$$

Now, we state the following key technical lemma.

Lemma 11 *There exist universal constants $\eta > 0$ for LP (5) such that if $z \in [0,1]^n$ and $0 < \varepsilon < \eta$ satisfy the followings:*

P1. There exist at most two violated factors for z, i.e., $|\{\alpha \in F \,:\, z_\alpha \notin \mathcal{P}_\alpha\}| \leq 2$.

P2. For each violated factor α, there exists $i \in \alpha$ such that $z^\dagger_\alpha \in \mathcal{P}_\alpha$, where $z^\dagger = z + \varepsilon e_i$ or $z^\dagger = z - \varepsilon e_i$ where $e_i \in \{0,1\}^n$ is the unit vector whose i-th coordinate is 1,

then there exists $z^\ddagger \in \mathcal{P}$ such that $\|z - z^\ddagger\|_1 \leq \varepsilon K$.[4]

The proof of Lemma 11 is presented in Section 4.1. Now, from Condition *C1*, it follows that there exists $\rho > 0$ such that

$$\rho := \inf_{x \in \mathcal{P}\setminus x^*} \frac{w \cdot x - w \cdot x^*}{\|x - x^*\|_1} > 0.$$

We let $\widehat{x}^t \in \{0,1,?\}^n$ denote the BP estimate at the t-th iteration for the MAP computation. We will show that under Conditions *C1*-*C3*,

$$\widehat{x}^t = x^*, \qquad \text{for} \quad t > \left(\frac{w_{\max}}{\rho} + 1\right) K.$$

Suppose the above statement is false, i.e., there exists $i \in \{1,2,\ldots,n\}$ such that $\widehat{x}^t_i \neq x^*_i$ for $t > \left(\frac{w_{\max}}{\rho} + 1\right) K$. Under the assumption, we will reach a contradiction. To this end, we construct a tree-structured GM $T_i(t)$, popularly known as the computational tree [27], as follows:

1. Add $y_i \in \{0,1\}$ as the root variable with variable factor function $e^{-w_i y_i}$.
2. For each leaf variable y_j and for each $\alpha \in F_j$ and ψ_α is not associated with y_j in the current tree-structured GM, add a factor function ψ_α as a child of y_j.
3. For each leaf factor ψ_α and for each variable y_k such that $k \in \alpha$ and y_k is not associated with ψ_α in the current tree-structured GM, add a variable y_k as a child of ψ_α with variable factor function $e^{-w_k y_k}$.
4. Repeat Step 2, 3 t times.

[4] K is defined in Theorem 1.

Suppose the initial messages of BP are set by 1, i.e., $m_{j\to\alpha}(\cdot)^0 = 1$. Then, if $\widehat{x}_i^t \in \{0, ?\}$, it is known [26] that there exists a MAP configuration $y^{\mathtt{MAP}}$ on $T_i(t)$ with $y_i^{\mathtt{MAP}} = 0$ at the root variable. A similar conclusion also holds for the case $\widehat{x}_i^t \in \{1, ?\}$. For other initial messages, one can guarantee the same property under changing weights of leaf variables of the tree-structured GM. Specifically, for a leaf variable k with $|F_k = \{\alpha_1, \alpha_2\}| = 2$ and α_1 being its parent factor in $T_i(t)$, one can reset its variable factor by $e^{-w'_k y_k}$, where

$$w'_k = w_k - \log \frac{\max_{z_{\alpha_2}:z_k=1} \psi_{\alpha_2}(z_{\alpha_2}) \Pi_{j\in\alpha_2\setminus k} m^0_{j\to\alpha_2}(z_j)}{\max_{z_{\alpha_2}:z_k=0} \psi_{\alpha_2}(z_{\alpha_2}) \Pi_{j\in\alpha_2\setminus k} m^0_{j\to\alpha_2}(z_j)}. \tag{28}$$

This is the reason why our proof of Theorem 1 goes through for arbitrary initial messages. For notational convenience, we present the proof for the standard initial message of $m^0_{j\to\alpha}(\cdot) = 1$, where it can be naturally generalized to other initial messages using (28).

Now we construct a new valid assignment $y^{\mathtt{NEW}}$ on the computational tree $T_i(t)$ as follows:

1. Initially, set $y^{\mathtt{NEW}} \leftarrow y^{\mathtt{MAP}}$.
2. Update the value of the root variable of $T_i(t)$ by $y_i^{\mathtt{NEW}} \leftarrow x_i^*$.
3. For each child factor ψ_α of root $i \in \alpha$, choose $\gamma \subset \alpha$ according to Condition *C3* and update the associated variable by $y_j^{\mathtt{NEW}} \leftarrow x_j^* \;\; \forall j \in \gamma$.
4. Repeat Step 2,3 recursively by substituting $T_i(t)$ by the subtree of $T_i(t)$ of root $j \in \gamma$ until the process stops (i.e., $i = j$) or the leaf of $T_i(t)$ is reached (i.e., i does not have a child).

One can notice that the set of revised variables in Step 2 of the above procedure forms a path structure Q in the tree-structured GM. Define ζ_j and κ_j be the number of copies of x_j in path Q with $x_j^* = 1$ and $x_j^* = 0$, respectively, where $\zeta = [\zeta_j], \kappa = [\kappa_j] \in \mathbb{Z}_+^n$. Then, from our construction of $y^{\mathtt{NEW}}$, one can observe that

$$w \cdot y^{\mathtt{MAP}} - w \cdot y^{\mathtt{NEW}} = w \cdot (\kappa - \zeta).$$

We consider three cases: (a) no end of the path Q touches a leaf of $T_i(t)$, (b) only one end of the path Q touches a leaf of $T_i(t)$, and (c) both ends of the path Q touch leaves of $T_i(t)$. First, consider the case (a). If we set $z = x^* + \varepsilon(\kappa - \zeta)$ where $0 < \varepsilon < \frac{1}{2t}$, then due to our construction of $y^{\mathtt{NEW}}$ utilizing Condition *C3*, one can observe $z \in \mathcal{P}$. However, since x^* is the unique optimum of LP (5), we have

$$w \cdot y^{\mathtt{MAP}} - w \cdot y^{\mathtt{NEW}} = \frac{1}{\varepsilon}(w \cdot z - w \cdot x^*) > 0,$$

which contradicts to the fact that $y^{\mathtt{MAP}}$ is a MAP configuration. Next, consider the case (c), where the case (b) can be argued in a similar manner. In this case, we use Lemma 11 by setting $z = x^* + \varepsilon(\kappa - \zeta)$ where $0 < \varepsilon < \min\left\{\frac{1}{2t}, \eta\right\}$ and one can check that z satisfies Conditions *P1, P2* of Lemma 11 due to Conditions *C2, C3*. Hence, from Lemma 11, there exists $z^\ddagger \in \mathcal{P}$ such that

$$\|z^\ddagger - z\|_1 \le \varepsilon K \quad \text{and} \quad \|z^\ddagger - x^*\|_1 \ge \varepsilon(\|\zeta\|_1 + \|\kappa\|_1 - K) \ge \varepsilon(t - K).$$

Hence, it follows that

$$
\begin{aligned}
0 < \rho &\le \frac{w \cdot z^{\ddagger} - w \cdot x^*}{\|z^{\ddagger} - x^*\|_1} \\
&\le \frac{w \cdot z + \varepsilon w_{\max} K - w \cdot x^*}{\varepsilon(t-K)} \\
&= \frac{\varepsilon w \cdot (\kappa - \zeta) + \varepsilon w_{\max} K}{\varepsilon(t-K)} \\
&= \frac{w \cdot (\kappa - \zeta) + w_{\max} K}{t-K}.
\end{aligned}
$$

Furthermore, if $t > \left(\frac{w_{\max}}{\rho} + 1\right) K$, the above inequality implies that

$$
\begin{aligned}
w \cdot y^{\texttt{MAP}} - w \cdot y^{\texttt{NEW}} &= w \cdot (\kappa - \zeta) \\
&\ge \rho t - (w_{\max} + \rho) K \; > \; 0.
\end{aligned}
$$

This is the contradiction to the fact that $y^{\texttt{MAP}}$ is a MAP configuration. This completes the proof of Theorem 1.

4.1 Proof of Lemma 11

We first define $\mathcal{P}_\varepsilon = \{x \, : \, Ax \ge b - \varepsilon \mathbf{1}\}$, where $\mathbf{1}$ is the vector of ones. Then, one can check that $z \in \mathcal{P}_\varepsilon$ for z, ε satisfying conditions of Lemma 11. Now we aim to achieve the following inequality

$$
\texttt{dist}(\mathcal{P}, \mathcal{P}_\varepsilon) := \max_{x \in \mathcal{P}_\varepsilon} \min_{y \in \mathcal{P}} \|x - y\|_1 \le \varepsilon K,
$$

which leads to the conclusion of Lemma 11. To this end, for $\xi \subset [1, 2, \dots, m]$ with $|\xi| = n$, we again let A_ξ be the square sub-matrix of A by choosing ξ-th rows of A and b_ξ is the n-dimensional subvector of b corresponding ξ. Using this notation, we first prove the following claim.

Claim 12 *If A_ξ is invertible and $v_\xi := (A_\xi)^{-1} b_\xi \in \mathcal{P}$, then v_ξ is a vertex of polytope $\mathcal{P}$.*

Proof. Suppose v_ξ is not a vertex of $\mathcal{P}$, i.e. there exist $x, y \in \mathcal{P}$ such that $x \neq y$ and $v_\xi = \lambda x + (1-\lambda) y$ for some $\lambda \in (0, 1/2]$. Under the assumption, we will reach a contradiction. Since $\mathcal{P}$ is a convex set,

$$
\frac{3\lambda}{2} x + \left(1 - \frac{3\lambda}{2}\right) y \in \mathcal{P}. \tag{29}
$$

However, as A_ξ is invertible,

$$
A_\xi \left(\frac{3\lambda}{2} x + \left(1 - \frac{3\lambda}{2}\right) y \right) \neq b_\xi. \tag{30}
$$

From (29) and (30), there exists a row vector A_{i*} of A_ξ and the corresponding entry b_i of b_ξ such that

$$
A_{i*} \cdot \left(\frac{3\lambda}{2} x + \left(1 - \frac{3\lambda}{2}\right) y \right) > b_i.
$$

Using the above inequality and $A_{i*} \cdot (\lambda x + (1-\lambda)y) = b_i$, one can conclude that

$$A_{i*} \cdot \left(\frac{\lambda}{2} x + \left(1 - \frac{\lambda}{2}\right) y \right) < b_i,$$

which contradict to $\frac{\lambda}{2} x + \left(1 - \frac{\lambda}{2}\right) y \in \mathcal{P}$. This completes the proof of Claim 12. $\square$

We also note that if v is a vertex of polytope $\mathcal{P}$, there exists ξ such that A_ξ is invertible and $v = (A_\xi)^{-1} b_\xi$. We define the following notation:

$$\mathcal{I} = \{\xi : (A_\xi)^{-1} b_\xi \in \mathcal{P}\} \quad \text{and} \quad \mathcal{I}_\varepsilon = \{\xi : (A_\xi)^{-1}(b_\xi - \varepsilon \mathbf{1}) \in \mathcal{P}_\varepsilon\},$$

where Claim 12 implies that $\{v_\xi := (A_\xi)^{-1} b_\xi : \xi \in \mathcal{I}\}$ and $\{u_{\xi,\varepsilon} := (A_\xi)^{-1}(b_\xi - \varepsilon \mathbf{1}) : \xi \in \mathcal{I}_\varepsilon\}$ are sets of vertices of $\mathcal{P}$ and $\mathcal{P}_\varepsilon$, respectively. Using the notation, we show the following claim.

Claim 13 *There exists $\eta > 0$ such that $\mathcal{I}_\varepsilon \subset \mathcal{I}$ for all $\varepsilon \in (0, \eta)$.*

Proof. Suppose $\eta > 0$ satisfying the conclusion of Claim 13 does not exist. Then, there exists a strictly decreasing sequence $\{\varepsilon_k > 0 : k = 1, 2, \dots\}$ converges to 0 such that $\mathcal{I}_{\varepsilon_k} \cap \{\xi : \xi \notin \mathcal{I}\} \neq \emptyset$. Since $|\{\xi : \xi \subset [1, 2, \dots, m]\}| < \infty$, there exists ξ' such that

$$|\mathcal{K} := \{k : \xi' \in \mathcal{I}_{\varepsilon_k} \cap \{\xi : \xi \notin \mathcal{I}\}\}| = \infty. \tag{31}$$

For any $k \in \mathcal{K}$, observe that the sequence $\{u_{\xi', \varepsilon_\ell} : \ell \geq k, \ell \in \mathcal{K}\}$ converges to $v_{\xi'}$. Furthermore, all points in the sequence are in $\mathcal{P}_{\varepsilon_k}$ since $\mathcal{P}_{\varepsilon_\ell} \subset \mathcal{P}_{\varepsilon_k}$ for any $\ell \geq k$. Therefore, one can conclude that $v_{\xi'} \in \mathcal{P}_{\varepsilon_k}$ for all $k \in \mathcal{K}$, where we additionally use the fact that $\mathcal{P}_{\varepsilon_k}$ is a closed set. Because $\mathcal{P} = \bigcap_{k \in \mathcal{K}} \mathcal{P}_{\varepsilon_k}$, it must be that $v_{\xi'} \in \mathcal{P}$, i.e., $v_{\xi'}$ must be a vertex of $\mathcal{P}$ from Claim 12. This contradicts to the fact $\xi' \in \{\xi : \xi \notin \mathcal{I}\}$. This completes the proof of Claim 13. $\square$

From the above claim, we observe that any $x \in \mathcal{P}_\varepsilon$ can be expressed as a convex combination of $\{u_{\xi,\varepsilon} : \xi \in \mathcal{I}\}$, i.e., $x = \sum_{\xi \in \mathcal{I}} \lambda_\xi u_{\xi,\varepsilon}$ with $\sum_{\xi \in \mathcal{I}} \lambda_\xi = 1$ and $\lambda_\xi \geq 0$. For all $\varepsilon \in (0, \eta)$ for $\eta > 0$ in Claim 13, one can conclude that

$$\begin{aligned}
\mathtt{dist}(\mathcal{P}, \mathcal{P}_\varepsilon) &\leq \max_{x \in \mathcal{P}_\varepsilon} \| \sum_{\xi \in \mathcal{I}} \lambda_\xi u_{\xi,\varepsilon} - \sum_{\xi \in \mathcal{I}} \lambda_\xi v_\xi \|_1 \\
&= \max_{x \in \mathcal{P}_\varepsilon} \varepsilon \| \sum_{\xi \in \mathcal{I}} \lambda_\xi (A_\xi)^{-1} \mathbf{1} \|_1 \\
&\leq \varepsilon \max_{\xi \in \mathcal{I}} \|(A_\xi)^{-1} \mathbf{1}\|_1 \\
&\leq \varepsilon K.
\end{aligned}$$

This completes the proof of Lemma 11.

5 Proofs of Corollaries

5.1 Proof of Corollary 3

Since LP (6) always has an integral solution, it suffices to show that the max-product BP on GM (7) converges to the solution of LP. The proof of Corollary 3 can be done by using Theorem 1. From GM (7), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. Suppose there are v and $x_{\delta(v)}$ with $\psi_v(x_{\delta(v)}) = 1$. Consider the case when there is $e \in \delta^i(v)$ with $x_e = 1 \neq x^*_e$. If $e' \in \delta^i(v)$ with $x_{e'} = 0 \neq x^*_{e'}$ exists, choose such e'. If not, choose $e' \in \delta^o(v)$ with $x_{e'} = 1 \neq x^*_{e'}$. On the other hand, consider when there is $e \in \delta^i(v)$ with $x_e = 0 \neq x^*_e$. If $e' \in \delta^o(v)$ with $x_{e'} = 1 \neq x^*_{e'}$ exists, choose such e'. If not, choose $e' \in \delta^i(v)$ with $x_{e'} = 0 \neq x^*_{e'}$. Then,

$$\psi_v(x'_{\delta(v)}) = 1, \qquad \text{where } x'_{e''} = \begin{cases} x_{e''} & \text{if } e'' \neq e, e' \\ x^*_{e''} & \text{otherwise} \end{cases}.$$

$$\psi_v(x''_{\delta(v)}) = 1, \qquad \text{where } x''_{e''} = \begin{cases} x_{e''} & \text{if } e'' = e, e' \\ x^*_{e''} & \text{otherwise} \end{cases}.$$

We can apply similar argument for the case when $e \in \delta^o(v)$, $v = s$ or t. From Theorem 1, we can conclude that if the solution of LP (6) is unique, the max-product BP on GM (7) converges to the solution of LP (6) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

5.2 Proof of Corollary 4

The proof of Corollary 4 can be done by using Theorem 1. From GM (10), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. Suppose there are v and $x_{\delta(v)}$ with $\psi_v(x_{\delta(v)}) = 1$. Consider the case when there is $e_i \in \delta(v)$ with $x_{e_i} = 1 \neq x^*_{e_i}$. Then, there is $e'_j \in \delta(v)$ with $x_{e'_j} = 0 \neq x^*_{e'_j}$. Choose such e'_j. On the other hand, consider when there is $e_i \in \delta(v)$ with $x_{e_i} = 0 \neq x^*_{e_i}$. Then, there is $e'_j \in \delta(v)$ with $x_{e'_j} = 1 \neq x^*_{e'_j}$. Choose such e'_j. Then,

$$\psi_v(x'_{\delta(v)}) = 1, \qquad \text{where } x'_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k \neq e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

$$\psi_v(x''_{\delta(v)}) = 1, \qquad \text{where } x''_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k = e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

From Theorem 1, we can conclude that if the solution of LP (9) is unique, the max-product BP on GM (10) converges to the solution of LP (9) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

5.3 Proof of Corollary 5

The proof of Corollary 5 can be done by using Theorem 1. From GM (13), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. For $v \in V$, we can apply same argument as the maximum weight matching case. Suppose there are v_C and $y_{\delta(v_C)}$ with $\psi_C(y_{\delta(v_C)}) = 1$. Consider the case when there is $(u_1, v_C) \in \delta(v_C)$ with $y_{(u_1,v_C)} = 1 \neq y^*_{(u_1,v_C)}$. As a feasible solution $y_{\delta(v_C)}$ forms a disjoint even paths [23], check edges along the path contains u_1. If there is $u_2 \in V(C)$ in the path with $y_{(u_2,v_C)} = 1 \neq y^*_{(u_2,v_C)}$ exists, choose such (u_1, v_C). If not, choose $(u_2, v_C) \in V(C)$ with $y_{(u_2,v_C)} = 0 \neq y^*_{(u_2,v_C)}$ at the end of the path. On the other hand, consider the case when there is $(u_1, v_C) \in \delta(v_C)$ with $y_{(u_1,v_C)} = 0 \neq y^*_{(u_1,v_C)}$. As a feasible solution $y_{\delta(v_C)}$ form a disjoint even paths, check edges along the path contains u_1. If there is $u_2 \in V(C)$ in the path with $y_{(u_2,v_C)} = 0 \neq y^*_{(u_2,v_C)}$ exists, choose such (u_1, v_C). If not, choose

$(u_2, v_C) \in V(C)$ with $y_{(u_2,v_C)} = 1 \neq y^*_{(u_2,v_C)}$ at the end of the path. Then, from disjoint even paths point of view, we can check that

$$\begin{aligned}
&\psi_C(y'_{\delta(v_C)}) = 1, \\
&\qquad \text{where } y'_{(u,v_C)} = \begin{cases} y_{(u,v_C)} & \text{if } u \neq u_1, u_2 \\ y^*_{(u,v_C)} & \text{otherwise} \end{cases}. \\
&\psi_C(y''_{\delta(v_C)}) = 1, \\
&\qquad \text{where } y''_{(u,v_C)} = \begin{cases} y_{(u,v_C)} & \text{if } u = u_1, u_2 \\ y^*_{(u,v_C)} & \text{otherwise} \end{cases}.
\end{aligned}$$

From Theorem 1, we can conclude that if the solution of LP (12) is unique and integral, the max-product BP on GM (13) converges to the solution of LP (12) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

5.4 Proof of Corollary 6

The proof of Corollary can be done by using Theorem 1. From GM (17), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. Suppose there are v and $x_{\delta(v)}$ with $\psi_v(x_{\delta(v)}) = 1$. Consider the case when there is $e_i \in \delta(v)$ with $x_{e_i} = 1 \neq x^*_{e_i}$. If there is $e'_j \in \delta(v)$ with $x_{e'_j} = 0 \neq x^*_{e'_j}$, choose such e'_j. If not, choose $e'_j = e_i$ On the other hand, consider when there is $e_i \in \delta(v)$ with $x_{e_i} = 0 \neq x^*_{e_i}$. If there is $e'_j \in \delta(v)$ with $x_{e'_j} = 1 \neq x^*_{e'_j}$, choose such e'_j. If not, choose $e'_j = e_i$ Then,

$$\psi_v(x'_{\delta(v)}) = 1, \qquad \text{where } x'_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k \neq e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

$$\psi_v(x''_{\delta(v)}) = 1, \qquad \text{where } x''_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k = e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

From Theorem 1, we can conclude that if the solution of LP (16) is unique, the max-product BP on GM (17) converges to the solution of LP (16) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

5.5 Proof of Corollary 7

The proof of Corollary can be done by using Theorem 1. From GM (20), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. Suppose there are v and $x_{\delta(v)}$ with $\psi_v(x_{\delta(v)}) = 1$. Consider the case when there is $e_i \in \delta(v)$ with $x_{e_i} = 1 \neq x^*_{e_i}$. If there is $e'_j \in \delta(v)$ with $x_{e'_j} = 0 \neq x^*_{e'_j}$, choose such e'_j. If not, choose $e'_j = e_i$ On the other hand, consider when there is $e_i \in \delta(v)$ with $x_{e_i} = 0 \neq x^*_{e_i}$. If there is $e'_j \in \delta(v)$ with $x_{e'_j} = 1 \neq x^*_{e'_j}$, choose such e'_j. If not, choose $e'_j = e_i$ Then,

$$\psi_v(x'_{\delta(v)}) = 1, \qquad \text{where } x'_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k \neq e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

$$\psi_v(x''_{\delta(v)}) = 1, \qquad \text{where } x''_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k = e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

From Theorem 1, we can conclude that if the solution of LP (19) is unique, the max-product BP on GM (20) converges to the solution of LP (19) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

5.6 Proof of Corollary 8

The proof of Corollary 8 can be done by using Theorem 1. From GM (22), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. Suppose there are v and $x_{\delta(v)}$ with $\psi_v(x_{\delta(v)}) = 1$. Consider the case when there is $e \in \delta(v)$ with $x_e = 1 \neq x^*_e$. By formulation of GM, there exists $e' \in \delta(v)$ with $x_{e'} = 0 \neq x^*_{e'}$. Choose such e'. On the other hand, consider when there is $e \in \delta(v)$ with $x_e = 0 \neq x^*_e$. There exists $e' \in \delta(v)$ with $x_{e'} = 1 \neq x^*_{e'}$. Choose such e'. Then,

$$\psi_v(x'_{\delta(v)}) = 1, \qquad \text{where } x'_{e''} = \begin{cases} x_{e''} & \text{if } e'' \neq e, e' \\ x^*_{e''} & \text{otherwise} \end{cases}.$$

$$\psi_v(x''_{\delta(v)}) = 1, \qquad \text{where } x''_{e''} = \begin{cases} x_{e''} & \text{if } e'' = e, e'' \\ x^*_{e''} & \text{otherwise} \end{cases}.$$

From Theorem 1, we can conclude that if the solution of LP (21) is unique and integral, the max-product BP on GM (22) converges to the solution of LP (21) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

5.7 Proof of Corollary 9

The proof of Corollary 9 can be done by using Theorem 1. From GM (24), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. Suppose there are v and $x_{\delta(v)}$ with $\psi_v(x_{\delta(v)}) = 1$. Consider the case when there is $e \in \delta(v)$ with $x_e = 1 \neq x^*_e$. If $e' \in \delta(v)$ with $x_{e'} = 0 \neq x^*_{e'}$ exists. Choose such e'. If not, there exists $e' \in \delta(v)$ with $x_{e'} = 1 \neq x^*_{e'}$. Choose such e'. On the other hand, consider when there is $e \in \delta(v)$ with $x_e = 0 \neq x^*_e$. If $e' \in \delta(v)$ with $x_{e'} = 1 \neq x^*_{e'}$ exists. Choose such e'. If not, there exists $e' \in \delta(v)$ with $x_{e'} = 0 \neq x^*_{e'}$. Choose such e'. Then,

$$\psi_v(x'_{\delta(v)}) = 1, \quad \text{where} \left\{ \begin{array}{l} x'_{e''} = \begin{cases} x_{e''} & \text{if } e'' \neq e, e' \\ x^*_{e''} & \text{otherwise} \end{cases} \\ y'_v = y^*_v \end{array} \right. .$$

$$\psi_v(x''_{\delta(v)}) = 1, \quad \text{where} \left\{ \begin{array}{l} x''_{e''} = \begin{cases} x_{e''} & \text{if } e'' = e, e'' \\ x^*_{e''} & \text{otherwise} \end{cases} \\ y'_v = y_v \end{array} \right. .$$

Case of y variable can be done in similar manner. From Theorem 1, we can conclude that if the solution of LP (23) is unique and integral, the max-product BP on GM (24) converges to the solution of LP (23) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

5.8 Proof of Corollary 10

The proof of Corollary can be done by using Theorem 1. From GM (27), each variable is connected to two factors (*C2* of Theorem 1). Now, lets check *C3* of Theorem 1. Suppose there are v and $x_{\delta(v)}$ with $\psi_v(x_{\delta(v)}) = 1$. Consider the case when there is $e_i \in \delta^o(v)$ with $x_{e_i} = 1 \neq x^*_{e_i}$. If there is $e_j \in \delta(v)^i$ with $x_{e_j} = 1 \neq x^*_{e_j}$, choose such e_j. Otherwise, there is $e_j \in \delta(v)^o$ with $x_{e_j} = 0 \neq x^*_{e_j}$. Choose such e_j. On the other hand, consider when there is $e_i \in \delta^o(v)$ with $x_{e_i} = 0 \neq x^*_{e_i}$. If there is $e_j \in \delta(v)^i$ with $x_{e_j} = 0 \neq x^*_{e_j}$, choose such e_j. Otherwise, there is $e_j \in \delta(v)^o$ with $x_{e_j} = 1 \neq x^*_{e_j}$. Choose such e_j. One can apply similar argument for $x_i \in \delta(v)^i$. Then,

$$\psi_v(x'_{\delta(v)}) = 1, \qquad \text{where } x'_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k \neq e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

$$\psi_v(x''_{\delta(v)}) = 1, \qquad \text{where } x''_{e''_k} = \begin{cases} x_{e''_k} & \text{if } e''_k = e_i, e'_j \\ x^*_{e''_k} & \text{otherwise} \end{cases}.$$

From Theorem 1, we can conclude that if the solution of LP (26) is unique, the max-product BP on GM (27) converges to the solution of LP (26) in $O(w_{\max}|E|^{2.5}/\rho)$ iterations.

6 Conclusion

The BP algorithm has been the most popular algorithm for solving inference problems arising graphical models, where its distributed implementation, associated ease of programming and strong parallelization potential are the main reasons for its growing popularity. In this paper, we aim for designing BP algorithms solving LPs, and provide sufficient conditions for its correctness and convergence. We believe that our results provide new interesting directions on designing efficient distributed (and parallel) solvers for large-scale LPs.

References

[1] Antar Bandyopadhyay and David Gamarnik. Counting without sampling: new algorithms for enumeration problems using statistical physics. In *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, pages 890–899. Society for Industrial and Applied Mathematics, 2006.

[2] Mohsen Bayati, Christian Borgs, Jennifer Chayes, and Riccardo Zecchina. Belief propagation for weighted b-matchings on arbitrary graphs and its relation to linear programs with integer solutions. *SIAM Journal on Discrete Mathematics*, 25(2):989–1011, 2011.

[3] Mohsen Bayati, Devavrat Shah, and Mayank Sharma. Maximum weight matching via max-product belief propagation. In *Information Theory, 2005. ISIT 2005. Proceedings. International Symposium on*, pages 1763–1767. IEEE, 2005.

[4] Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume 6. Athena Scientific Belmont, MA, 1997.

[5] Ethan D Bolker and Thomas Zaslavsky. A simple algorithm that proves half-integrality of bidirected network programming. *Networks*, 48(1):36–38, 2006.

[6] Venkat Chandrasekaran, Misha Chertkov, David Gamarnik, Devavrat Shah, and Jinwoo Shin. Counting independent sets using the bethe approximation. *SIAM Journal on Discrete Mathematics*, 25(2):1012–1034, 2011.

[7] Venkat Chandrasekaran, Nathan Srebro, and Prahladh Harsha. Complexity of inference in graphical models. In *UAI 2008, Proceedings of the 24th Conference in Uncertainty in Artificial Intelligence*, pages 70–78. AUAI Press, 2008.

[8] George B Dantzig. *Linear programming and extensions*. Princeton university press, 1998.

[9] David Gamarnik, Devavrat Shah, and Yehua Wei. Belief propagation for min-cost network flow: Convergence and correctness. *Operations Research*, 60(2):410–428, 2012.

[10] Joseph Gonzalez, Yucheng Low, and Carlos Guestrin. Parallel splash belief propagation. Technical report, DTIC Document, 2010.

[11] Bert C Huang and Tony Jebara. Loopy belief propagation for bipartite maximum weight b-matching. In *International Conference on Artificial Intelligence and Statistics*, pages 195–202, 2007.

[12] Richard M Karp. *Reducibility among combinatorial problems*. Springer, 1972.

[13] Leonid G Khachiyan. Polynomial algorithms in linear programming. *USSR Computational Mathematics and Mathematical Physics*, 20(1):53–72, 1980.

[14] Yucheng Low, Joseph Gonzalez, Aapo Kyrola, Danny Bickson, Carlos Guestrin, and Joseph M. Hellerstein. Graphlab: A new framework for parallel machine learning. In *UAI 2010, Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence*, pages 340–349. AUAI Press, 2010.

[15] Nam Ma, Yinglong Xia, and Viktor K Prasanna. Task parallel implementation of belief propagation in factor graphs. In *Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2012 IEEE 26th International*, pages 1944–1953. IEEE, 2012.

[16] Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

[17] Tom Richardson and Ruediger Urbanke. *Modern coding theory*. Cambridge University Press, 2008.

[18] Nicholas Ruozzi and Sekhar Tatikonda. st paths using the min-sum algorithm. In *Communication, Control, and Computing, 2008 46th Annual Allerton Conference on*, pages 918–921. IEEE, 2008.

[19] Justin Salez and Devavrat Shah. Belief propagation: an asymptotically optimal algorithm for the random assignment problem. *Mathematics of Operations Research*, 34(2):468–480, 2009.

[20] Sujay Sanghavi, Dmitry Malioutov, and Alan Willsky. Belief propagation and lp relaxation for weighted matching in general graphs. *Information Theory, IEEE Transactions on*, 57(4):2203–2212, 2011.

[21] Sujay Sanghavi, Devavrat Shah, and Alan S Willsky. Message passing for maximum weight independent set. *Information Theory, IEEE Transactions on*, 55(11):4822–4834, 2009.

[22] Alexander Schrijver. *Combinatorial optimization: polyhedra and efficiency*, volume 24. Springer, 2003.

[23] Jinwoo Shin, Andrew E Gelfand, and Misha Chertkov. A graphical transformation for belief propagation: Maximum weight matchings and odd-sized cycles. In *Advances in Neural Information Processing Systems*, pages 2022–2030, 2013.

[24] George B Dantzig Mukund N Thapa. Linear programming. 2003.

[25] Martin J Wainwright and Michael I Jordan. Graphical models, exponential families, and variational inference. *Foundations and Trends® in Machine Learning*, 1(1-2):1–305, 2008.

[26] Yair Weiss. Belief propagation and revision in networks with loops. 1997.

[27] Yair Weiss and William T Freeman. On the optimality of solutions of the max-product belief-propagation algorithm in arbitrary graphs. *Information Theory, IEEE Transactions on*, 47(2):736–744, 2001.

[28] Jonathan S Yedidia, William T Freeman, and Yair Weiss. Constructing free-energy approximations and generalized belief propagation algorithms. *Information Theory, IEEE Transactions on*, 51(7):2282–2312, 2005.