Max-Product Belief Propagation for Linear Programming: Applications to Combinatorial Optimization

Sejun Park
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology
sejun.park@kaist.ac.kr

Jinwoo Shin
Department of Electrical Engineering
Korea Advanced Institute of Science and Technology
jinwoos@kaist.ac.kr

Abstract

Max-product belief propagation (BP) is a popular message-passing algorithm for computing a maximum-a-posteriori (MAP) assignment in a joint distribution represented by a graphical model (GM). It has been shown that BP can solve a few classes of Linear Programming (LP) formulations to combinatorial optimization problems including maximum weight matching and shortest path, i.e., BP can be a distributed solver for certain LPs. 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 and vertex cover. More importantly, our criteria can guide the BP design to compute fractional LP solutions, while most prior results focus on integral ones. Our results provide new tools on BP analysis and new directions on efficient solvers for large-scale LPs.

1 INTRODUCTION

Graphical model (GM) has been one of powerful paradigms for succinct representations of joint probability distributions in variety of scientific fields (Yedidia et al., 2005; Richardson and Urbanke, 2008; Mezard and Montanari, 2009; Wainwright and Jordan, 2008). GM represents a joint distribution of some random vector to a graph structured model where each vertex corresponds to a random variable and each edge captures to a conditional dependency between random variables. In many applications involving GMs, finding maximum-a-posteriori (MAP) assignment in GM is an important inference task, which is known to be computationally intractable (i.e., NP-hard) in general (Chandrasekaran et al., 2008). Max-product belief propagation (BP) is the most popular heuristic for approximating a MAP assignment of given GM, where its performance has been not well understood in loopy GMs. 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 (Low et al., 2010; Gonzalez et al., 2010; Ma et al., 2012) by different research groups in machine learning communities.

In the past years, there have been made extensive research efforts to understand BP performances on loopy GMs behind its empirical success. Several characterizations of the max-product BP fixed points have been proposed (Weiss and Freeman, 2001; Vinyals et al., 2010), whereas they do not guarantee the BP convergence in general. It has also been studied about the BP convergence to the correct answer, in particular, under a few classes of loopy GM formulations of combinatorial optimization problems: matching (Bayati et al., 2005; Sanghavi et al., 2011; Huang and Jebara, 2007; Salez and Shah, 2009), perfect matching (Bayati et al., 2011), matching with odd cycles (Shin et al., 2013) and shortest path (Ruozzi and Tatikonda, 2008). 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, BP can be used an efficient distributed solver for those LPs, and is presumably of better choice than classical centralized LP solvers such as simplex methods (Dantzig, 1998), interior point methods (Thapa, 2003) and ellipsoid methods (Khachiyan, 1980) 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 (Bayati et al., 2005; Sanghavi et al., 2011; Huang and Jebara, 2007; Salez and Shah, 2009) are not extended to even for perfect matching ones (Bayati et al., 2011). In this paper, we overcome such technical difficulties for enhancing the power of BP as a LP solver.

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 a 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 (Bayati et al., 2005; Sanghavi et al., 2011; Huang and Jebara, 2007), perfect matching (Bayati et al., 2011), matching with odd cycles (Shin et al., 2013) and shortest path (Ruozzi and Tatikonda, 2008), 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 and vertex cover, which are not known in the literature. 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 4.2 and Section 4.4 for details).

Our proof technique is built upon on that of Sanghavi et al. (2011) 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. We run such a BP for solving the famous traveling saleman problem (TSP), and our experiments show that BP outperforms other popular heuristics (see Section 5). 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.

Organization. In Section 2, we introduce necessary backgrounds for the BP algorithm. In Section 3, we provide the main result of the paper, and several concrete applications to popular combinatorial optimizations are described in Section 4. In Section 5, we show empirical performances of BP algorithms for solving TSP.

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 $z = [z_i] \in \{0,1\}^n$,

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

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

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

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

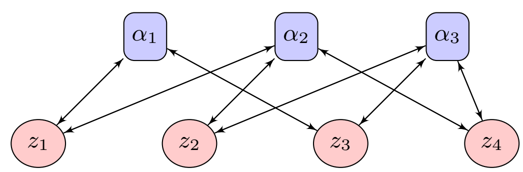

Figure 1: Factor graph for the graphical model $\Pr[z] \propto \psi_{\alpha_1}(z_1, z_3)\psi_{\alpha_2}(z_1, z_2, z_4)\psi_{\alpha_3}(z_2, z_3, z_4)$, i.e., $F = \{\alpha_1, \alpha_2, \alpha_3\}$ and $n = 4$. Each α_j selects a subset of z. For example, α_1 selects $\{z_1, z_3\}$.

Assignment z^* is called a maximum-a-posteriori (MAP) assignment if $z^* = \arg\max_{z \in \{0,1\}^n} \Pr[z]$. This means that computing a MAP assignment requires us to compare $\Pr[z]$ 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 (Chandrasekaran et al., 2008).

2.2 MAX-PRODUCT BELIEF PROPAGATION

The (max-product) BP algorithm is a popular heuristic for approximating the MAP assignment in GM. BP is implemented iteratively; at each iteration t, BP maintains 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 are associated with z_i. The messages

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

are updated as follows:

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

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

Where 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 variable z_i to factor ψ_α is a product containing all messages received by z_i in the previous iteration, except for the message sent by ψ_α itself.

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

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

The BP fixed-point of messages is defined as $m^{t+1} = m^t$ under the above updating rule. 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:

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

This BP algorithm outputs $z^{BP} = [z^{BP}_i]$ where

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

It is known that z^{BP} converges to a MAP assignment after a sufficient number of iterations, if the factor graph is a tree and the MAP assignment is unique. However, if the graph contains cycles, the BP algorithm is not guaranteed to converge a MAP assignment in general.

3 CONVERGENCE AND CORRECTNESS OF BELIEF PROPAGATION

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 $w = [w_i] \in \mathbb{R}^n$,

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

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

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

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

$$\begin{aligned} \text{minimize} \quad & w \cdot x \\ \text{subject to} \quad & \psi_\alpha(x_\alpha) = 1, \quad \forall \alpha \in F \\ & x = [x_i] \in [0,1]^n. \end{aligned} \quad (5)$$

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 $x^* \in \{0,1\}^n$. As stated in the following theorem, we establish other sufficient conditions so that the max-product BP can indeed find the LP solution.

Theorem 1 *The max-product BP on GM (4) with arbitrary initial message converges to the solution of LP (5) if 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., $|F_i| \le 2$.

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 $x_i \ne x^*_i$, there exists $\gamma \subset \alpha$ such that

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

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

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

Since Theorem 1 holds for arbitrary initial messages, the conditions *C1, C2, C3* also provides the uniqueness of BP fixed-points in term of marginal beliefs, as follows.

Corollary 2 *The BP fixed-points of GM (4) have the same marginal beliefs 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, where we provide several concrete examples in Section 4. 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 (Schrijver, 2003). Nevertheless, Theorem 1 provides important guidelines to design BP algorithms, irrespectively of the LP analysis. For example, in Section 5, we report empirical performances of BP following the above guideline for solving the traveling salesman problem, without relying on whether the corresponding LP has an integral solution or not.

3.1 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

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

We first state the following key technical lemma.

Lemma 3 *There exist universal constants $K, \eta > 0$ for LP (5) such that if $z \in [0,1]^n$ and $0 < \varepsilon < \eta$ satisfy the followings:*

1. *There exist at most two violated factors for z, i.e., $|\{\alpha \in F \,:\, z_\alpha \notin \mathcal{P}_\alpha\}| \leq 2$.*
2. *For each violated factor α, there exist $i \in \alpha$ such that $z^\dagger_\alpha \in \mathcal{P}_\alpha$, where $z^\dagger = z + \varepsilon e_i$ or $z^\dagger = z - \varepsilon e_i$ and $e_i \in \{0,1\}^n$ is the unit vector whose i-th coordinate is 1,*

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

The proof of Lemma 3 is presented in Section 3.2. Now, from Condition *C1*, it follows that there exists $\rho > 0$ such that

$$\rho := \inf_{x \in \mathcal{P}\setminus x^*} \frac{w\cdot x - w \cdot x^*}{\|x - x^*\|_1} > 0. \tag{6}$$

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*,

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

where $w_{\max} = \max_j |w_j|$ and K is the universal constant in Lemma 3. 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.

Now we construct a tree-structured GM $T_i(t)$, popularly known as the computational tree (Weiss and Freeman, 2001), 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.

Suppose the initial messages of BP are set by 1, i.e., $m_{j\to\alpha}(\cdot)^0 = 1$. Then, if $x^*_i \neq \widehat{x}^t_i$, it is known (Weiss, 1997) that there exists a MAP configuration y^{MAP} on $T_i(t)$ with $y^{MAP}_i \neq x^*_i$ at the root variable. 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 $T_i(t)$, we reset its variable factor by $e^{-w'_k y_k}$, where

$$w'_k = w_k - \log \frac{\max_{z_{\alpha_2}: z_k = 1} \psi_{\alpha_2}(z_{\alpha_2}) \Pi_{j \in \alpha_2 \setminus k} m^0_{j\to\alpha_2}(z_j)}{\max_{z_{\alpha_2}: z_k = 0} \psi_{\alpha_2}(z_{\alpha_2}) \Pi_{j \in \alpha_2 \setminus k} m^0_{j\to\alpha_2}(z_j)}. \tag{7}$$

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 (7).

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

1. Initially, set $y^{NEW} \leftarrow y^{MAP}$.
2. Update the value of the root variable of $T_i(t)$ by $y^{NEW}_i \leftarrow x^*_i$.
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 $y^{NEW}_j \leftarrow x^*_j \;\; \forall j \in \gamma$.
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. We first, consider the case that both ends of the path Q touch leaves of $T_i(t)$, where other cases can be argued in a similar manner. 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^{NEW}, one can observe that

$$\begin{aligned} y^{NEW} &= y^{MAP} + \zeta - \kappa \\ w\cdot y^{MAP} - w\cdot y^{NEW} &= w \cdot (\kappa - \zeta). \end{aligned}$$

If we set $z = x^* + \varepsilon(\kappa - \zeta)$ where $0 < \varepsilon < \min\{1/2t, \eta\}$, then one can check that z satisfies the conditions of Lemma 3 using Conditions *C2, C3*. Hence, from Lemma 3, there exists $z^\ddagger \in \mathcal{P}$ such that

$$\begin{aligned} &\|z^\ddagger - z\|_1 \leq \varepsilon K \\ &\|z^\ddagger - x^*\|_1 \geq \varepsilon(\|\zeta\|_1 + \|\kappa\|_1 - K) \geq \varepsilon(t - K). \end{aligned}$$

where $z = x^* + \varepsilon(\kappa - \zeta)$. Hence, it follows that

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

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

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

This contradicts to the fact that y^{MAP} is a MAP configuration. This completes the proof of Theorem 1.

3.2 PROOF OF LEMMA 3

One can write $\mathcal{P} = \{x \,:\, Ax \geq b\} \subset [0,1]^n$ for some matrix $A \in \mathbb{R}^{m \times n}$ and vector $b \in \mathbb{R}^m$, where without loss of generality, we can assume that $\|A_i\|_2 = 1$ where $\{A_i\}$ is the set of row vectors of A. We define

$$\mathcal{P}_\varepsilon = \{x \,:\, Ax \geq b - \varepsilon \mathbf{1}\},$$

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 3. Now we aim for finding a universal constant K satisfying

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

which leads to the conclusion of Lemma 3.

To this end, for $\xi \subset [1, 2, \dots, m]$ with $|\xi| = n$, we 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. Throughout the proof, we only consider ξ such that A_ξ is invertible. Using this notation, we first claim the following.

Claim 4 *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,

$$\frac{3\lambda}{2} x + \left(1 - \frac{3\lambda}{2}\right) y \in \mathcal{P}. \tag{8}$$

However, as A_ξ is invertible,

$$A_\xi \left(\frac{3\lambda}{2} x + \left(1 - \frac{3\lambda}{2}\right) y\right) \neq b_\xi. \tag{9}$$

From (8) and (9), there exists a row vector A_i of A_ξ and the corresponding element 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 4. □

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:

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

where Claim 4 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 5 *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 5 does not exist. Then, there exists a strictly decreasing sequence $\{\varepsilon_k > 0 : k = 1, 2, \dots\}$ converges to 0 such that $\mathcal{I}_{\varepsilon_k} - \mathcal{I} \neq \emptyset$. Since $|\{\xi : \xi \subset [1, 2, \dots, m]\}| < \infty$, there exists ξ' such that

$$|\mathcal{K} := \{k \,:\, \xi' \in \mathcal{I}_{\varepsilon_k} - \mathcal{I}\}| = \infty. \tag{10}$$

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 4. This contradicts to the fact $\xi' \notin \mathcal{I}$. This completes the proof of Claim 5. □

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., $x = \sum_{\xi \in I} \lambda_\xi u_{\xi,\varepsilon}$ with $\sum_{\xi \in I} \lambda_\xi = 1$ and $\lambda_\xi \geq 0$. For all $\varepsilon \in (0, \eta)$ for $\eta > 0$ in Claim 5, one can conclude that

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

where we choose $K = \max_\xi \|A_\xi^{-1} \mathbf{1}\|_1$. This completes the proof of Lemma 3.

4 APPLICATIONS OF THEOREM 1 TO COMBINATORIAL OPTIMIZATION

In this section, 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 and vertex cover. We note that the shortest path result, Corollary 6, is known (Ruozzi and Tatikonda, 2008), where we rediscover it as a corollary of Theorem 1. Our other results, Corollaries 7-11, are new and what we first establish in this paper.

4.1 SHORTEST PATH

Given 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:

$$\begin{aligned} \text{minimize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e \in \delta^o(v)} x_e - \sum_{e \in \delta^i(v)} x_e \\ & = \begin{cases} 1 & \text{if } v = s \\ -1 & \text{if } v = t \quad \forall v \in V \\ 0 & \text{otherwise} \end{cases} \\ & x = [x_e] \in [0,1]^{|E|}. \end{aligned} \tag{11}$$

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, i.e., the shortest path from s to t. We consider the following GM for LP (11):

$$\Pr[X = x] \propto \prod_{e \in E} e^{-w_e x_e} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{12}$$

where the factor function ψ_v is defined as

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

For the above GM (12), one can easily check Conditions *C2, C3* of Theorem 1 hold and derive the following corollary whose formal proof is presented in the supplementary material due to the space constraint.

Corollary 6 *If the shortest path from s to t, i.e., the solution of the shortest path LP (11), is unique, then the max-product BP on GM (12) converges to it.*

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

4.2 MAXIMUM WEIGHT PERFECT MATCHING

Given 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:

$$\begin{aligned} \text{maximize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e \in \delta(v)} x_e = 1, \quad \forall v \in V \\ & x = [x_e] \in [0,1]^{|E|}. \end{aligned} \tag{13}$$

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 (13) 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:

$$\begin{aligned} \text{maximize} \quad & w' \cdot x \\ \text{subject to} \quad & \sum_{e_i \in \delta(v)} x_{e_i} = 2 \qquad \forall v \in V \\ & x = [x_{e_i}] \in [0,1]^{|E'|}. \end{aligned} \tag{14}$$

where $w'_{e_1} = w'_{e_2} = w_e$. One can easily observe that solving LP (14) is equivalent to solving LP (13) due to our construction of G' and w'. Now, construct the following GM for LP (14):

$$\Pr[X = x] \propto \prod_{e_i \in E'} e^{w'_{e_i} x_{e_i}} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{15}$$

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 (15), we derive the following corollary of Theorem 1 whose formal proof is presented in the supplementary material due to the space constraint.

Corollary 7 *If the solution of the maximum weight perfect matching LP (14) is unique, then the max-product BP on GM (15) converges it.*

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 (Bayati et al., 2011) that BP converges to the unique and integral solution of LP (13), while Corollary 7 implies that BP can solve it without the integrality condition. 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.

4.3 MAXIMUM WEIGHT PERFECT MATCHING WITH ODD CYCLES

In previous section we prove that BP converges to the optimal (possibly, fractional) solution of LP (14), equivalently LP (13). One can add odd cycle (also called Blossom) constraints and make those LPs tight i.e. solves the maximum weight perfect matching problem:

$$\begin{aligned} \text{maximize} \quad & w \cdot x \\ \text{subject to} \quad & \sum_{e \in \delta(v)} x_e = 1, \quad \forall\, v \in V \\ & \sum_{e \in C} x_e \leq \frac{|C|-1}{2}, \quad \forall C \in \mathcal{C}, \\ & x = [x_e] \in [0,1]^{|E|}. \end{aligned} \tag{16}$$

where $\mathcal{C}$ is a set of odd cycles in G. The authors (Shin et al., 2013) study BP for solving LP (16) by replacing $\sum_{e\in\delta(v)} x_e = 1$ by $\sum_{e\in\delta(v)} x_e \leq 1$, 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 (16). To this end, we follow the approach (Shin et al., 2013) 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 (16) is equivalent to the following LP:

$$\begin{aligned} \text{maximize} \quad & w' \cdot y \\ \text{subject to} \quad & \sum_{e\in\delta(v)} y_e = 1, && \forall\, v \in V \\ & \sum_{u\in V(C)} (-1)^{d_C(u,e)} y_{(v_C,u)} \in [0,2], && \forall e \in E(C) \\ & \sum_{e\in\delta(v_C)} y_e \leq |C| - 1, && \forall C \in \mathcal{C} \\ & y = [y_e] \in [0,1]^{|E'|}. \end{aligned} \tag{17}$$

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

$$\Pr[Y = y] \propto \prod_{e\in E} e^{w_e y_e} \prod_{v\in V} \psi_v(y_{\delta(v)}) \prod_{C\in\mathcal{C}} \psi_C(y_{\delta(v_C)}), \tag{18}$$

where the factor function ψ_v, ψ_C is defined as

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

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

For the above GM (18), we derive the following corollary of Theorem 1 whose formal proof is presented in the supplementary material due to the space constraint.

Corollary 8 *If the solution of the maximum weight perfect matching with odd cycles LP (17) is unique and integral, then the max-product BP on GM (18) converges to it.*

We again emphasize that a similar result for the maximum weight (non-perfect) matching problem was established in (Shin et al., 2013). 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.

4.4 VERTEX COVER

Given 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 (Karp, 1972). The associated LP formulation to the vertex cover problem is as follows:

$$\begin{aligned} \text{minimize} \quad & b \cdot y \\ \text{subject to} \quad & y_u + y_v \geq 1, (u,v) \in E \\ & y = [y_v] \in [0,1]^{|V|}. \end{aligned} \tag{19}$$

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:

$$\begin{aligned} \text{maximize} \quad & \sum_{e\in E} x_e \\ \text{subject to} \quad & \sum_{e\in\delta(v)} x_e \leq b_v, \quad \forall\, v \in V \\ & x = [x_e] \in \mathbb{R}_+^{|E|}. \end{aligned} \tag{20}$$

Note that the above LP always has a half-integral solution. As we did in Section 4.2, 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:

$$\begin{array}{ll} \text{maximize} & w' \cdot x \\ \text{subject to} & \sum_{e_i \in \delta(v)} x_{e_i} \le 2b_v, \quad \forall v \in V \\ & x = [x_{e_i}] \in [0,1]^{|E'|} \end{array}, \tag{21}$$

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:

$$\Pr[X = x] \propto \prod_{e_i \in E'} e^{w'_{e_i} x_{e_i}} \prod_{v \in V} \psi_v(x_{\delta(v)}), \tag{22}$$

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 (22), we derive the following corollary of Theorem 1 whose formal proof is presented in the supplementary material due to the space constraint.

Corollary 9 *If the solution of the vertex cover dual LP (21) is unique, then the max-product BP on GM (22) converges it.*

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 (19) is integral, then it can be easily found from the solution of the dual LP (21) using the strictly complementary slackness condition (Bertsimas and Tsitsiklis, 1997) .

4.5 TRAVELING SALESMAN

Given 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 following:

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

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

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

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 (24), we derive the following corollary of Theorem 1 whose formal proof is presented in the supplementary material due to the space constraint.

Corollary 10 *If the solution of the traveling salesman LP (23) is unique and integral, then the max-product BP on GM (24) converges it.*

Again, the uniqueness condition in the above corollary is easy to guarantee by adding small random noises to edge weights. In Section 5, we show the empirical performance of the max-product BP on GM (24) for solving TSP without relying on the integrality condition in Corollary 10.

4.6 MAXIMUM WEIGHT CYCLE PACKING

Given 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:

$$\begin{array}{ll} \text{maximize} & w \cdot x \\ \text{subject to} & \sum_{e \in \delta(v)} x_e = 2y_v \\ & x = [x_e] \in [0,1]^{|E|}, y = [y_v] \in [0,1]^{|V|}. \end{array} \tag{25}$$

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

$$\Pr[X = x, Y = y] \propto \prod_{e \in E} e^{w_e x_e} \prod_{v \in V} \psi_v(x_{\delta(v)}, y_v), \tag{26}$$

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 (26), we derive the following corollary of Theorem 1 whose formal proof is presented in the supplementary material due to the space constraint.

Corollary 11 *If the solution of maximum weight vertex disjoint cycle packing LP (25) is unique and integral, then the max-product BP on GM (26) converges it.*

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

5 EXPERIMENTAL RESULTS FOR TRAVELING SALESMAN PROBLEM

In this section, we report empirical performances of BP on GM (24) for solving the traveling salesman problem (TSP)

Table 1: Experimental results for small size complete graph and each number is the average among 100 samples. For example, Greedy+BP means that the Greedy algorithm using edge weights as BP beliefs as we describe in Section 5. The left value is the approximation ratio, i.e., the average weight ratio between the heuristic solution and the exact solution. The right value is the average weight of the heuristic solutions. The last row is a ratio of tight TSP LP (23).

Size	5	10	15	20	25
Greedy	1.07 / 1.84	1.20 / 2.25	1.33 / 2.58	1.51 / 2.85	1.51 / 3.04
Greedy+BP	1.00 / 1.75	1.05 / 1.98	1.13 / 2.23	1.19 / 2.27	1.21 / 2.43
Christofides	1.38 / 1.85	1.38 / 2.56	1.67 / 3.20	1.99 / 3.75	2.16 / 4.32
Christofides+BP	1.00 / 1.75	1.09 / 2.07	1.23 / 2.43	1.30 / 2.50	1.45 / 2.90
Insertion	1.03 / 1.79	1.29 / 2.38	1.53 / 2.95	1.72 / 3.26	1.89 / 3.77
Insertion+BP	1.00 / 1.75	1.29 / 2.39	1.52 / 2.97	1.79 / 3.38	1.94 / 3.89
N-Neighbor	1.07 / 1.84	1.27 / 2.39	1.42 / 2.74	1.55 / 2.96	1.64 / 3.30
N-Neighbor+BP	1.00 / 1.75	1.05 / 1.98	1.13 / 2.23	1.15 / 2.21	1.20 / 2.40
2-Opt	1.00 / 1.75	1.08 / 2.04	1.12 / 2.21	1.24 / 2.36	1.28 / 2.57
2-Opt+BP	1.00 / 1.75	1.04 / 1.96	1.07 / 2.11	1.11 / 2.13	1.16 / 2.34
Tight LPs	100%	93%	88%	87%	84%

Table 2: Experimental results for sparse Erdos-Renyi graph with fixed average vertex degrees and each number is the average among 1000 samples. The left value is the ratio that a heuristic finds the Hamiltonian cycle without penalty edges. The right value is the average weight of the heuristic solutions.

Size	100			200		
Degree	10	25	50	10	25	50
Greedy	0% / 7729.43	0.3% / 2841.98	13% / 1259.08	0% / 15619.9	0% / 5828.88	0.3% / 2766.07
Greedy+BP	14% / 1612.82	21% / 1110.27	44% / 622.488	6.4% / 2314.95	10% / 1687.29	16% / 1198.48
Christoifeds	0% / 19527.3	0% / 16114.3	0% / 10763.7	0% / 41382.5	0% / 37297.0	0% / 32023.1
Christofides+BP	14% / 2415.73	20% / 1663.47	34% / 965.775	6.1% / 3586.77	9.2% / 2876.35	12% / 2183.80
Insertion	0% / 12739.2	84% / 198.099	100% / 14.2655	0% / 34801.6	0.9% / 3780.71	99% / 44.1293
Insertion+BP	0% / 13029.0	76% / 283.766	100% / 14.6964	0% / 34146.7	0.3% / 4349.11	99% / 41.2176
N-Neighbor	0% / 9312.77	0% / 3385.14	7.6% / 1531.83	0% / 19090.7	0% / 7383.23	0.3% / 3484.82
N-Neighbor+BP	16% / 1206.95	26% / 824.232	50% / 509.349	6.9% / 1782.17	12% / 1170.38	24% / 888.421
2-Opt	34% / 1078.03	100% / 14.6873	100% / 7.36289	2% / 3522.78	100% / 35.8421	100% / 18.6147
2-Opt+BP	76% / 293.450	100% / 13.5773	100% / 6.53995	33% / 1088.79	100% / 34.7768	100% / 17.4883
Tight LPs	62%	62.3%	63%	52.2%	55%	52.2%

that is presumably the most famous one in combinatorial optimization. In particular, we design the following BP-guided heuristic for solving TSP:

1. Run BP for a fixed number of iterations, say 100, and calculate the BP marginal beliefs (3).
2. Run the known TSP heuristic using edge weights as $\log \frac{b[0]}{b[1]}$ using BP margianl beliefs instead of the original weights.

For TSP heuristic in Step 2, we use Greedy, Christoifeds, Insertion, N-Neighbor and 2-Opt provided by the LEMON graph library (Dezső et al., 2011). We first perform the experiments on the complete graphs of size 5, 10, 15, 20, 25 and random edge weight in $(0, 1)$ to measure approximation qualities of heuristics, where it is reported in Table 1. Second, we consider the sparse Erdos-Renyi random graph of size 100, 200 and random edge weight in $(0, 1)$. Then, we make it a complete graph by adding non-existing edges with penalty edge weight 1000.[2] For these random instances, we report performance of various heuristics in Table 2. Our experiments show that BP boosts performances of known TSP heuristics in overall, where BP is very easy to code and does not hurt the simplicity of heuristics.

[2]This is to ensure that a Hamiltonian cycle always exists.

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.

Acknowledgements.

We would like to acknowledge the support of the AOARD project, FA2386-14-1-4058.

References

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.

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.

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

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.

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

Balázs Dezső, Alpár Jüttner, and Péter Kovács. Lemon–an open source c++ graph template library. *Electronic Notes in Theoretical Computer Science*, 264(5):23–45, 2011.

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

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.

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

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

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.

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.

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

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

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.

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.

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.

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

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.

George B Dantzig Mukund N Thapa. Linear programming. 2003.

Meritxell Vinyals, Alessandro Farinelli, Juan A Rodríguez-aguilar, et al. Worst-case bounds on the quality of max-product fixed-points. In *Advances in Neural Information Processing Systems*, pages 2325–2333, 2010.

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.

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

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.

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.