Belief Propagation for Min-Cost Network Flow: Convergence & Correctness

The MIT Faculty has made this article openly available. Please share how this access benefits you. Your story matters.

<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>As Published</td>
<td><a href="http://dl.acm.org/citation.cfm?id=1873625">http://dl.acm.org/citation.cfm?id=1873625</a></td>
</tr>
<tr>
<td>Publisher</td>
<td>Society for Industrial and Applied Mathematics</td>
</tr>
<tr>
<td>Version</td>
<td>Final published version</td>
</tr>
<tr>
<td>Accessed</td>
<td>Mon Nov 05 12:14:31 EST 2018</td>
</tr>
<tr>
<td>Citable Link</td>
<td><a href="http://hdl.handle.net/1721.1/73521">http://hdl.handle.net/1721.1/73521</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>Article is made available in accordance with the publisher’s policy and may be subject to US copyright law. Please refer to the publisher’s site for terms of use.</td>
</tr>
<tr>
<td>Detailed Terms</td>
<td></td>
</tr>
</tbody>
</table>
Belief Propagation for Min-cost Network Flow: Convergence & Correctness

David Gamarnik* Devavrat Shah‡ Yehua Wei ‡

Abstract
We formulate a Belief Propagation (BP) algorithm in the context of the capacitated minimum-cost network flow problem ($\text{MCF}$). Unlike most of the instances of BP studied in the past, the messages of BP in the context of this problem are piecewise-linear functions. We prove that BP converges to the optimal solution in pseudo-polynomial time, provided that the optimal solution is unique and the problem input is integral. Moreover, we present a simple modification of the BP algorithm which gives a fully polynomial-time randomized approximation scheme (FPRAS) for $\text{MCF}$. This is the first instance where BP is proved to have fully-polynomial running time.

1 Introduction
The Markov random field or graphical model provides a succinct representation for capturing the dependency structure between a collection of random variables. In the recent years, the need for large scale statistical inference has made graphical models the representation of choice in many statistical applications. There are two key inference problems for a graphical model of interest. The first problem is the computation of marginal distribution of a random variable. This problem is (computationally) equivalent to the computation of partition function in statistical physics or counting the number of combinatorial objects (e.g., independent sets) in computer science. The second problem is that of finding the mode of the distribution, i.e., the assignment with maximum likelihood (ML). For a constrained optimization problem, when the constraints are modeled through a graphical model and probability is proportional to the cost of the assignment, an ML assignment is the same as an optimal solution. Both of these questions, in general, are computationally hard.

Belief Propagation (BP) is an “umbrella” heuristic designed for these two problems. Its version for the first problem is known as the “sum-product algorithm” and for the second problem is known as the “max-product algorithm”. Both versions of the BP algorithm are iterative, simple and message-passing in nature. In this paper, our interest lies in the correctness and convergence properties of the max-product version of BP when applied to the minimum-cost network flow problems ($\text{MCF}$), an important class of linear (or more generally convex) optimization problems. In the rest of the paper, we will use the term BP to refer to the max-product version unless specified otherwise.

The BP algorithm is essentially an approximation of the dynamic programming assuming that underlying graphical model is a tree [8], [22], [16]. Specifically, assuming that the graphical model is a tree, one can obtain a natural parallel iterative version of the dynamic programming in which variable nodes pass messages between each other along edges of the graphical model. Somewhat surprisingly, this seemingly naive BP heuristic has become quite popular in practice [2], [9], [11], [17]. In our opinion, there are two primary reasons for the popularity of BP. First, like dynamic programming, it is generically applicable, easy to understand and implementation-friendly due to its iterative, simple and message-passing nature. Second, in many practical scenarios, the performance of BP is surprisingly good [21] [22]. On one hand, for an optimist, this unexpected success of BP provides a hope for its being a genuinely much more powerful algorithm than what we know thus far (e.g., better than primal-dual methods). On the other hand, a skeptic would demand a systematic understanding of the limitations (and strengths) of BP, in order to caution a practitioner in using it. Thus, irrespective of the perspective of an algorithmic theorist, rigorous understanding of BP is very important.

Prior work. Despite these compelling reasons, only recently have we witnessed an explosion of research for theoretical understanding of the performance of the BP algorithm in the context of various combinatorial optimization problems, both tractable and intractable (NP-hard) versions. To begin with, Bayati, Shah and Sharma [6] considered the performance of BP for the the problem of finding maximum weight matching in a bipartite graph. They established that BP converges in pseudo-polynomial time to the optimal solution when the optimal solution is unique [6]. Bayati et al. [4] as well as Sanghavi et al. [18] generalized this result.
2 BP for Linear Programming Problem

Suppose we are given an LP problem in the standard form:

\[
\begin{align*}
\text{(LP)} \quad & \quad \min c^T x \\
\text{s.t.} \quad & \quad Ax = b, \\
& \quad x \geq 0, x \in \mathbb{R}^n
\end{align*}
\]

where \( A \) is a real \( m \times n \) matrix, \( b \) is a real vector of dimension \( m \), and \( c \) is a real vector of dimension \( n \). We define its factor graph, \( F_{LP} \), to be a bipartite graph, with factor nodes \( v_1, v_2, \ldots, v_m \) and variable nodes \( e_1, e_2, \ldots, e_n \), such that \( e_i \) and \( v_j \) are adjacent if and only if \( a_{ij} \neq 0 \). For example, the graph shown in Figure 1, is the factor graph of the LP problem:

\[
\begin{align*}
\min & \quad 4x_1 + x_2 - 3x_4 + x_5 + 2x_6 - x_7 \\
\text{s.t.} & \quad x_1 + 9x_3 - x_4 + 5x_7 = 10 \\
& \quad -2x_1 + x_2 + x_3 - 3x_4 + x_5 - x_6 + x_7 = 3 \\
& \quad x_1 + x_6 = 2 \\
& \quad x \geq 0, x \in \mathbb{R}^7.
\end{align*}
\]

In \( F_{LP} \), for every \( 1 \leq i \leq m \), let \( E_i = \{ e_j | a_{ij} \neq 0 \} \); and let \( V_j = \{ v_i | a_{ij} \neq 0 \} \). For each factor node \( v_i \), define the factor function \( \psi_i : \mathbb{R}^{|E_i|} \rightarrow \mathbb{R} (\mathbb{R} = \mathbb{R} \cup \infty) \) as:

\[
\psi_i(z) = \begin{cases} 
0 & \text{if } \sum_{j \in E_i} a_{ij} \cdot z_j = b_j \\
\infty & \text{otherwise}
\end{cases}
\]

Essentially, \( \psi_i \) represents the \( i \)-th equality constraints in \( LP \). If we let \( x_{E_i} \) be the subvector of \( x \) with indices in \( E_i \), then \( \psi_i(x_{E_i}) < \infty \) if and only if \( x \) does not violate the \( i \)-th equality constraint in \( LP \). Also, for any variable node \( e_j \), define variable function \( \phi_j : \mathbb{R} \rightarrow \mathbb{R} \) as:

\[
\phi_j(z) = \begin{cases} 
c_j \cdot z & \text{if } z \geq 0 \\
\infty & \text{otherwise}
\end{cases}
\]

Notice that problem \( LP \) is equivalent to \( \min_{x \in \mathbb{R}^n} \sum_{i=1}^m \psi_i(x_{E_i}) + \sum_{j=1}^n \phi_j(x_j) \), which is defined on \( F_{LP} \). As BP is defined on graphical models, we can now formulate the BP algorithm as an iterative message-passing algorithm on \( F_{LP} \). It is formally presented below as Algorithm 1. The idea of the algorithm is that at the \( t \)-th iteration, every factor vertex \( v_i \) sends a message function \( m_{v_i \rightarrow e_j}(z) \) to each one of its neighbors \( e_j \), where \( m_{v_i \rightarrow e_j}(z) \) is \( v_i \)'s estimate of the cost if \( x_j \) takes value \( z \), based on the messages \( v_i \) received at time \( t \) from its neighbors. Additionally, every variable vertex \( e_j \) sends a message function \( m_{e_j \rightarrow v_i}(z) \) to each one of its neighbors \( v_i \), where \( m_{e_j \rightarrow v_i}(z) \) is \( e_j \)'s estimate of the cost if \( x_j \) takes value \( z \), based on the messages \( e_j \) received at time \( t - 1 \) from all of its neighbors except \( v_i \). After \( N \) iterations, the algorithm calculates the "belief" at variable node \( e_j \) using the messages sent to vertex \( e_j \) at time \( N \) for each \( j \).

In Algorithm 1, one basic question is the computation procedure of message functions of the form \( m_{v_i \rightarrow e_j}(z) \) and \( m_{e_j \rightarrow v_i}(z) \). We claim that every message function is a convex piecewise-linear function, which allows us to encode it in terms of a finite vector describing the break points and slopes of its linear pieces. In subsection 3.1, we describe in detail how \( m_{v_i \rightarrow e_j}(z) \) can be computed.

In general, the max-product behaves badly when solving an LP instance without a unique optimal solution [6], [4]. Yet, even with the assumption that an
Algorithm 1 BP for LP

1: Initialize $t = 0$, and for each variable node $v_j$, create a factor to variable message $m_{v_j}^0(z) = 0$, $\forall v_j \in V_t$.
2: for $t = 1, 2, ..., N$ do
3: \hspace{0.5cm} $\forall v_j \in V_t$, where $v_i \in V_j$, set $m_{v_i}^t(z)$ to
4: \hspace{1cm} $\phi_j(z) + \sum_{v_i \in V_j \setminus v_i} m_{v_i}^{t-1}(z)$, \hspace{0.5cm} $\forall z \in R$.
5: \hspace{0.5cm} $\forall e_j \in E_t$, set $m_{v_i}^t(z)$ to
6: \hspace{1cm} $\min_{z \in R^{[v_i] \setminus z}, z_j = z} \left\{ \psi_i(z) + \sum_{e_j \in E_i \setminus e_j} m_{v_i}^{t-1}(z_j) \right\}$, \hspace{0.5cm} $\forall z \in R$.
7: \hspace{0.5cm} $t := t + 1$
8: end for
9: Return $\hat{x}^N$.

LP problem has a unique optimal solution, in general $\hat{x}^N$ in Algorithm 1 does not converge to the unique optimal solution as $N$ grows large. One such instance (LP-relaxation of an instance of the maximum-weight independent set problem) which was earlier described in [19]:

$$\min \sum_{i=1}^{3} 2x_i + \sum_{j=1}^{3} 3y_j$$

s.t. $x_i + y_j + z_{ij} = 1$, $\forall 1 \leq i \leq 3, 1 \leq j \leq 3$

$x, y, z \geq 0$

3 BP Algorithm for Min-Cost Network Flow Problem

We start off this section by defining the capacitated min-cost network flow problem ($\mathcal{MCF}$). Given a directed graph $G = (V, E)$, let $V$, $E$ denote the set of vertices and arcs respectively, and let $|V| = n$, $|E| = m$. For any vertex $v \in V$, let $E_v$ be the set of arcs incident to $v$, and for any $e \in E_v$, let $\Delta(v, e) = 1$ if $e$ is an out-arc of $v$ (i.e., arc $e = (v, w)$, for some $w \in V$), and $\Delta(v, e) = -1$ if $e$ is an in-arc of $v$ (i.e., arc $e = (w, v)$, for some $w \in V$). $\mathcal{MCF}$ on $G$ is formulated as follows [7],[1]:

$$\min \sum_{e \in E} c_e x_e,$$

(\mathcal{MCF})

$$\Delta(v, e)x_e = b_v, \forall v \in V \text{ (demand constraints)}$$

$$0 \leq x_e \leq u_e, \forall e \in E \text{ (flow constraints)}$$

where $c_e \geq 0$, $u_e \geq 0$, $c_e \in R$, $u_e \in R$, for each $e \in E$, and $b_v \in R$ for each $v \in V$. As $\mathcal{MCF}$ is a special class of LP, we can view each arc $e \in E$ as a variable vertex and each $v \in V$ as a factor vertex, and define functions $\psi_v$, $\phi_e$, $\forall v \in V$, $\forall e \in E$:

$$\psi_v(z) = \begin{cases} 0 & \text{if } \sum_{e \in E_v} \Delta(v, e)z_e = b_v \\ \infty & \text{otherwise} \end{cases}$$

$$\phi_e(z) = \begin{cases} c_e z & \text{if } 0 \leq z \leq u_e \\ \infty & \text{otherwise} \end{cases}$$

Algorithm 2 BP for Network Flow

1: Initialize $t = 0$. For each $e \in E$, suppose $e = (v, w)$.
2: Initialize messages $m_{v}^{0}(z) = 0$, $\forall z \in R$ and $m_{w}^{0}(z) = 0$, $\forall z \in R$.
3: for $t = 1, 2, 3, ..., N$ do
4: \hspace{0.5cm} $\forall v \in E$, let $e = (v, w)$, update messages:
5: \hspace{1cm} $m_{v}^{t}(z) = \phi_e(z)$
6: \hspace{1cm} + $\min_{z \in R^{[v] \setminus z}, z_j = z} \left\{ \psi_i(z) + \sum_{e_j \in E_i \setminus e_j} m_{v}^{t-1}(z_j) \right\}$, \hspace{0.5cm} $\forall z \in R$
7: \hspace{1cm} $m_{w}^{t}(z) = \phi_e(z)$
8: \hspace{1cm} + $\min_{z \in R^{[w] \setminus z}, z_j = z} \left\{ \psi_i(z) + \sum_{e_j \in E_i \setminus e_j} m_{w}^{t-1}(z_j) \right\}$, \hspace{0.5cm} $\forall z \in R$.
9: \hspace{0.5cm} $t := t + 1$
10: end for
11: $\forall v \in E$, let $e = (v, w)$, and set the “belief” function:
12: \hspace{0.5cm} $n_e^N(z) = m_{v}^{N}(z) + m_{w}^{N}(z) - \phi_e(z)$
13: Calculate the “belief” by finding $\hat{x}_e^N = \arg\min n_e^N(z)$ for each $e \in E$.
14: Return $\hat{x}_e^N$, which is the guess of the optimal solution of $\mathcal{MCF}$.
Clearly, solving $\mathcal{MCF}$ is equivalent to solving
\[
\min_{\psi_B} \{ \sum_{v \in V} \psi_v(x_{E_v}) + \sum_{e \in E} \phi_e(x_e) \}.
\]
We can then apply BP algorithm for LP (Algorithm 1) for $\mathcal{MCF}$.
Because of the special structure of $\mathcal{MCF}$, each variable node is adjacent
to exactly two factor nodes. This allows us to skip the step of update messages
$m^t_{e \rightarrow v}$, and present a simplified version of BP algorithm for $\mathcal{MCF}$,
which we refer to as Algorithm 2.

To understand Algorithm 2 at a high level, each arc can be thought of as an agent, which is trying to figure out its own flow while meeting the conservation constraints at its endpoints. Each link maintains an estimate of its “local cost” as a function of its flow (thus this estimate is a function, not a single number). At each time step an arc updates its function as follows: the cost of assigning $x$ units of flow to link $e$ is the cost of pushing $x$ units of flow through $e$ plus the minimum-cost way of assigning flow to neighboring edges (w.r.t. the functions they computed last time step) to restore flow conservation at the endpoints of $e$.

Before formally stating theorem of convergence of BP for $\mathcal{MCF}$, we first give the definition of a residual network $\mathcal{G}$ [1]. Define $G(x)$ to be the residual network of $G$ and flow $x$ as: $G(x)$ has the same vertex set as $G$, and $\forall (v, w) \in E$, if $x_v < u_{vw}$, then $e$ is an arc in $G(x)$ with cost $c^r_e = c_e$, also, if $x_v > 0$, then there is an arc $e' = (w, v)$ in $G(x)$ with cost $c^r_{e'} = -c_e$, now, let $\delta(x) = \min_{C \in C} \{ c^+(C) = \sum_{e \in C} c^r_e \}$ where $C$ is the set of directed cycles in $G(x)$. Note that if $x^*$ is a unique optimal solution of $\mathcal{MCF}$ with directed graph $G$, then $\delta(x^*) > 0$ in $G(x^*)$.

**Theorem 3.1.** Suppose $\mathcal{MCF}$ has a unique optimal solution $x^*$. Define $L$ to be the maximum cost of a simple directed path in $G(x^*)$. Then for any $N \geq (\lceil \frac{L}{2\delta(x^*)} \rceil + 1)n$, we have $\hat{x}^N = x^*$.

Thus, by Theorem 3.1, the BP algorithm finds the unique optimal solution of $\mathcal{MCF}$ in at most $(\lceil \frac{L}{2\delta(x^*)} \rceil + 1)n$ iterations.

### 3.1 Computing/Encoding Message Functions

First, we formally define a convex piecewise-linear function:

**Definition 3.1.** A function $f$ is a convex piecewise-linear function if for some finite set of reals, $a_0 < a_1 < \ldots < a_n$, (allowing $a_0 = -\infty$ and $a_n = \infty$), we have:

\[
f(z) = \begin{cases} c_1(z - a_1) + f(a_1) & \text{if } z \in [a_0, a_1] \\ c_i+1(z - a_i) + f(a_i) & \text{if } z \in [a_i, a_{i+1}], 1 \leq i < n \\ \infty & \text{otherwise} \end{cases}
\]

where $c_1 < c_2 < \ldots < c_n$ and $f(a_1) \in \mathbb{R}$. We define $a_0, a_1, \ldots, a_n$ as the vertices of $f$, $p(f) = n$ ($p(f)$ denotes the number of linear pieces in $f$) and $c_i(z - a_{i-1}) + f(a_{i-1})$ for $z \in [a_{i-1}, a_i]$ the $i$th linear piece of $f$.

Clearly, if $f$ is a convex piecewise-linear function, then we can store all the “information” about $f$ in a finite vector of size $O(p(f))$.

**Property 3.1.** Let $f_1(x)$, $f_2(x)$ be two convex piecewise-linear functions. Then, $f_1(ax + b)$, $cf_1(x) + df_2(x)$ are also convex piecewise-linear functions, for any real numbers $a$, $b$, $c$ and $d$, where $c \geq 0, d \geq 0$.

**Definition 3.2.** Let $S = \{ f_1, f_2, \ldots, f_k \}$ be a set of convex piecewise-linear functions, and let $\Psi_t : \mathbb{R}^k \rightarrow \mathbb{R}$ be:

\[
\Psi_t(x) = \begin{cases} 0 & \text{if } \sum_{i=1}^{k} x_i = t \\ \infty & \text{otherwise} \end{cases}
\]

We say $I_S(t) = \min_{x \in \mathbb{R}^k} \{ \psi_t(x) + \sum_{i=1}^{k} f_i(x_i) \}$ is an interpolation of $S$.

**Lemma 3.1.** Let $f_1$, $f_2$ be two convex piecewise-linear functions, and suppose $z_1^*$, $z_2^*$ are vertices of $f_1$, $f_2$ and $z_1^* = \arg \min f_1(z)$, $z_2^* = \arg \min f_2(z)$. Let $S = \{ f_1, f_2 \}$, then $I_S(t)$ (the interpolation of $S$) is a convex piecewise-linear function and it can be computed in $O(p(f_1) + p(f_2))$ operations.

**Proof.** We prove this lemma by describing a procedure to construct $I_S(t)$. The idea behind construction of $I_S(t)$ is essentially to “stitch” together the linear pieces.
of $f_1$ and $f_2$. Specifically, let $g(t)$ be the function which is defined only at $z_1^* + z_2^*$, with $g(z_1^* + z_2^*) = f_1(z_1^*) + f_2(z_2^*)$, and let $L_1 = z_1^*, L_2 = z_2^*, U_1 = z_1^*, U_2 = z_2^*$. We will construct $g$ iteratively and eventually have $g(t) = I_S(t)$. The construction is described as follows: at each iteration, let $X_1$ (and $X_2$) be the linear piece of $f_1$ (and $f_2$) at the left side of $L_1$ (and $L_2$). Then choose the linear piece with the larger slope from $\{X_1, X_2\}$, and “stitch” this piece onto the left side of the left endpoints of $g$. If $P_1$ is chosen, update $L_1$ to the vertex which is on the left end of $P_1$. As an example, consider $f_1$ and $f_2$ in Figure 2, then $z_1^* = 1$ and $z_2^* = 0$ are vertices of $f_1$ and $f_2$ such that $z_1^* = \arg \min f_1(z), z_2^* = \arg \min f_2(z)$. Note that the linear piece $X_1$ in the procedure is labeled as $P_1$ on the graph, while $X_2$ does not exist (since there is no linear piece for $f_2$ on the right side of $z_2$). Hence, we “stitch” $P_1$ to the left side of $g$, and update $L_1$ to 0.

Similarly, let $Y_1$ ($Y_2$) be the linear piece of $f_1$ ($f_2$) at the right side of $U_1$ ($U_2$), then choose the linear piece with the smaller slope and “stitch” this piece onto the right side of the right endpoints of $g$. If $Q_1$ is chosen, update $U_1$ to the vertex which is on the right side of $P_1$. Again, we use $f_1$ and $f_2$ in Figure 2 as an illustration. Then the linear piece $Y_1$ in the procedure is labeled as $P_2$, while $Y_2$ is labeled as $P_3$. As $P_2$ has a lower slope than $P_3$, we “stitch” $P_2$ to the right side of $g$ and update $U_1$ to 2.

Repeat this procedure until both $L_1$ (and $L_2$) and $U_1$ (and $U_2$) are the left most (and right most) endpoints of $f_1$ (and $f_2$), or both the endpoints of $g$ are infinity. See Figure 2 and Figure 3 as an illustration of interpolation of two functions.

Note that the total number of iterations is bounded by $O(p(f_1) + p(f_2))$, and each iteration takes at most constant number of operations. By construction, it is clear that $g$ is a convex piecewise-linear function. Also, $g(z_1^* + z_2^*) = f_1(z_1^*) + f_2(z_2^*)$, and by the way we constructed $g$, we must have $g(t) \leq \{\Psi_t(x) + f_1(x_1) + f_2(x_2)\}$ for any $t \in \mathbb{R}$. Therefore, we have $g = I_S$, and we are done.

Remark. In the case where either of $f_1$ or $f_2$ do not have global minimal, the interpolation can be still obtained in $O(p(f_1) + p(f_2))$ operations, but it involves tedious analysis of different cases. As reader will notice, the implementation of Algorithm 2 does not require this case and hence we don’t discuss it here.

Theorem 3.2. Given a set $S$ consisting of convex piecewise-linear functions, $I_S(t)$ is a convex piecewise-linear function. Moreover, suppose $|S| = k$, and $P = \sum_{f \in S} p(f)$. Then, $I_S(t)$ can be computed in $O(P \cdot \log k)$ operations.

Proof. For the sake of simplicity, assume $k$ is divisible by 2. Let $S_1 = \{f_1, f_2\}, S_2 = \{f_3, f_4\}, \ldots, S_{\frac{k}{2}} = \{f_{k-1}, f_k\}$, and $S' = \{I_{S_1}, I_{S_2}, \ldots, I_{S_{\frac{k}{2}}}\}$. Then one can observe that $I'_{S} = I_S$, by definition of $I_S$. By Lemma 3.1, each function $S'$ is piecewise-linear, and $S'$ can be computed in $O(P)$ operations. Consider changing $S$ to $S'$ as a procedure of decreasing the number of convex piecewise-linear functions. This procedure reduces the number by a factor of 2 each time, and consumes $O(P)$ operations. Hence, it takes $O(\log k)$ procedures to reduce set $S$ into a single convex piecewise-linear function. And hence, computing $I_S(t)$ takes $O(P \cdot \log k)$ operations.

Definition 3.3. Let $S = \{f_1, f_2, \ldots, f_k\}$ be a set of convex piecewise-linear functions, $a \in \mathbb{R}^k$, and let $\Psi_t : \mathbb{R}^k \rightarrow \mathbb{R}$ be:

$$\Psi_t(x) = \begin{cases} 0 & \text{if } \sum_{i=1}^{k} a_i x_i = t, \\ \infty & \text{otherwise} \end{cases}, \quad \forall v \in V$$

We say $I_S^k(t) = \min_{x \in \mathbb{R}^k} \{\Psi_t(x) + \sum_{i=1}^{k} f_i(x_i)\}$ is a scaled interpolation of $S$.

Corollary 3.1. Given any set of convex piecewise-linear functions $S$, $I_S^k(t)$ is a convex piecewise-linear function. Suppose $|S| = k$, and $P = \sum_{f \in S} p(f)$, then, the corners and slopes of $I_S(t)$ can be found in $O(P \cdot \log k)$ operations.

Proof. Let $S = \{f_1(x), f_2(x), \ldots, f_k(x)\}$, and let $S' = \{f_1'(x) = f_1(a_1 x), f_2'(x) = f_2(a_2 x), \ldots, f_k'(x) = f_k(a_k x)\}$. Then we have $I_S^k(t) = I_{S'}(t)$, and the result follows immediately from Theorem 3.2.

Corollary 3.2. For any nonnegative integer $t$, $e \in E$ such that $e = (v, w)$ (or $e = (w, v)$), then $m_{v \rightarrow v} (t)$, a message function at $t$-th iteration of Algorithm 2, is a convex piecewise-linear function.

Proof. We show this by doing induction on $t$. For $t = 0$, $m_{v \rightarrow v} (t)$ is a convex piecewise-linear function by definition. For $t > 0$, let us recall that $m_{v \rightarrow v} (t) = \phi_{v}(z) + \min_{z \in \mathbb{R} \setminus E_{\{v\}}} \{\psi_{w}(\bar{z}) + \sum_{e \in E_{\{v\}}} m_{v \rightarrow w}^{-1}(\bar{z})\}$, for $z \in \mathbb{R}$.

By induction hypothesis, any $m_{v \rightarrow w}^{-1}(t)$ is a piecewise-linear convex function. Suppose $S = \{m_{v \rightarrow w}^{-1}(\bar{z})\}$, $\bar{c} \in E_{\{w\}} \setminus e$, and $a = \Delta(v, e) a_v$ for any $e \in E_{\{w\}} \setminus e$. Then, $g(z) = \min_{z \in \mathbb{R} \setminus E_{\{w\}}} \{\psi_{w}(\bar{z}) + \sum_{e \in E_{\{w\}}} m_{v \rightarrow w}^{-1}(\bar{z})\}$ is equal to $I_S^k(z \bar{c} + d)$ for some real number $c$ and $d$ (i.e., a shifted scaled interpolation of $S$). By Corollary 3.1, $g(z)$ is a convex piecewise-linear function, and $\phi_{v}$ is a convex piecewise-linear function. Therefore, we have that $m_{v \rightarrow v} (t) = g + \phi_{v}$ is a convex piecewise-linear function.
message functions, and thus can also be encoded in terms of functions in Algorithm 1 are also convex piecewise-linear functions. We would like to note the result that the slopes of its linear pieces infinite number of iterations.

Using similar argument, we can see that the message function of the slopes for the linear pieces of message functions can be computed as piecewise-linear convex interpolation of functions \( m_{e^{-w}} \) is integral, all vertices of \( m_{e^{-w}} \) are integral, and all of their entries are bounded by a constant \( K \).

To see \( MCF' \) is indeed a \( MCF \), observe that for each \( v \in V \), if we split \( v \) into two vertices \( v_{in} \) and \( v_{out} \), where \( v_{in} \) is incident to all in-arcs of \( v \) with \( b_{v_{in}} = 0 \), and \( v_{out} \) is incident to all out-arcs of \( v \) with \( b_{v_{out}} = b_{v} \); while create an arc from \( v_{in} \) to \( v_{out} \), set the capacity of the arc to be \( \tilde{u}_{v} \), and cost of the arc to be 0. Let the new graph be \( G' \), then, the \( MCF \) on \( G' \) is equivalent to \( MCF' \). Although we can use Algorithm 2 to solve the \( MCF \) on the new graph, we would like to define functions \( \psi_{v} \), \( \phi_{v} \), \( \forall v \in V \), \( \forall e \in E \):

$$
\psi_{v}(z) = \begin{cases} 
0 & \text{if } \sum_{e \in E_{u}} \Delta(v,e)z_{e} = b_{v} \\
\infty & \text{otherwise}
\end{cases}
$$

$$
\phi_{e}(z) = \begin{cases} 
c_{e}z & \text{if } 0 \leq z \leq u_{e} \\
\infty & \text{otherwise}
\end{cases}
$$

and apply Algorithm 2 with defined \( \psi \) and \( \phi \). When we update message functions \( m_{e^{-w}} \) for all \( e \in E_{w} \), the inequality \( \sum_{e \in \delta(u)} x_{e} \leq \tilde{u}_{w} \), allow us to only look at the \( \tilde{u}_{w} \) linear pieces from message functions \( m_{e^{-w}}^{t-1} \) for all but constant number of \( e \in E_{w} \). This allows us to further trim the running time of \( e \) for \( MCF' \), and we will present this more formally in Section 6.1.

4 Convergence of BP for \( MCF \)

Before proving Theorem 3.1, we define the computation trees, and then establish the connection between computation trees and BP. Using this connection, we proceed to prove the main result of our paper.
4.1 Computation Tree and BP Let $T^N_e$, the $N$-computation tree corresponding to variable $x_e$, be defined as follows:

1. Let $V(T^N_e)$ be the vertex set of $T^N_e$ and $E(T^N_e)$ be the arc set of $T^N_e$. Let $\Gamma(.)$ be the map which maps each vertex $v' \in V(T^N_e)$ to a vertex $v \in V$. And $\Gamma$ preserves the arcs (i.e. For each $v' = (v'_a, v'_b) \in E(T^N_e)$, $\Gamma(v') = (\Gamma(v'_a), \Gamma(v'_b)) \in E$).

2. Divide $V(T^N_e)$ into $N + 1$ levels, on the 0-th level, we have a “root” arc $r$, and $\Gamma(r) = e$. And a vertex $v'$ is on the $t$-th level of $T^N_e$ if $v'$ is $t$ arcs apart from either nodes of $r$.

3. For any $v' \in V(T^N_e)$, let $E_{v'}$ be the set of all arcs incident to $v'$. If $v'$ is on the $t$-th level of $T^N_e$, and $t < N$, then $\Gamma(E_{v'}) = E_{\Gamma(v')}$ (i.e. $\Gamma$ preserves the neighborhood of $v'$).

4. For every vertex $v'$ that is on the $N$-th level of $T^N_e$, $v'$ is incident to exactly one arc in $T^N_e$.

In other literatures, $T^N_e$ is known as the “unwrapped tree” of $G$ rooted at $e$. Figure 4 gives an example of a computation tree. We make a note that the definition of computation trees we have introduced is slightly different from the definition in other papers [4] [6] [18], although the analysis and insight for computation trees is very similar.

Let $V^o(T^N_e)$ be the set of all the vertices which are not on the $N$-th level of $T^N_e$. Consider the problem

$$\begin{align*}
\min & \sum_{f \in E(T^N_e)} c_{\Gamma(f)} x_f \\
\text{s.t.} & \sum_{f \in E_v} \Delta(v, f) x_f = b_{\Gamma(v)}, \quad \forall v \in V^o(T^N_e) \\
& 0 \leq x_f \leq u_{\Gamma(f)}, \quad f \in E(T^N_e)
\end{align*}$$

(\text{MCF}^N_e)

Loosely speaking, $\text{MCF}^N_e$ is almost an $\text{MCF}$ on $T^N_e$. There is a flow constraint for every arc $e \in E(T^N_e)$, and a demand/supply constraint for every node $v \in V^o(T^N_e)$. Now, we state the lemma which exhibits the connection between BP and the computation trees.

**Lemma 4.1.** $\text{MCF}^N_e$ has an optimal solution $y^*$, satisfying $y^*_e = x^*_e$ ($r$ is the root of $T^N_e$, $\Gamma(r) = e$).

This result is not too surprising as BP is exact on trees. A thorough analysis can show that the message functions in BP are essentially the implementation of an algorithm using the “bottom up” (aka dynamic programming) approach, starting from the bottom level of $T^N_i$. A more formal, combinatorial proof of this lemma also exists, which is a rather technical argument using induction. Proofs for similar statements can be found in [4] and [18].

4.2 Proof of the Main Theorem

**Proof.** [Proof of Theorem 3.1] Suppose $\exists e_0 \in E$ and $N \geq \left\lceil \frac{\log |V^o(T^N_e)|}{\log 2} \right\rceil + 1$ such that $x^*_N \neq x^*_{e_0}$. By Lemma 4.1, we can find an optimal solution $y^*$ for $\text{MCF}^N_e$, where $y^*_r = x^*_{e_0}$. Now, without loss of generality, we assume $y^*_r > x^*_{e_0}$. Let $r = (v_\alpha, v_\beta)$, because $y^*$ is feasible for $\text{MCF}^{N-1}_{e_0}$ and $x^*$ is feasible for $\text{MCF}$, we have

$$b_{\Gamma(v_\alpha)} = \sum_{e \in E_{v_\alpha}} \Delta(v_\alpha, e) y^*_e = y^*_r + \sum_{e \in E_{v_\alpha}} \Delta(v_\alpha, e) y^*_e$$

Then, we can find arc $e'_1$ incident to $v_\alpha$, $e'_1 \neq r$, such that $y^*_{e'_1} > x^*_{\Gamma(e'_1)}$ if $e'_1$ has the same orientation as $r$, and $y^*_{e'_1} < x^*_{\Gamma(e'_1)}$ if $e'_1$ has the opposite orientation as $r$. Similarly, we can find some arc $e'_{-1}$ incident to $v_\beta$ satisfying the same condition. Let $v_{\alpha_1}, v_{\alpha_{-1}}$ be the vertices on $e'_1, e'_{-1}$ which are one level below $v_\alpha, v_\beta$ on $T^N_e$, then, we can find some arc $e'_2, e'_{-2}$ satisfying the same condition. If we continue this all the way down to the leaf nodes of $T^N_{e_0}$, we can find the set of arcs $(e'_{-N}, e'_{-N+1}, ..., e'_{-1}, e'_1, ..., e'_N)$ such that

$$y^*_{e'_i} > x^*_{\Gamma(e'_i)} \iff e'_i \text{ has the same orientation as } r$$

$$y^*_{e'_i} < x^*_{\Gamma(e'_i)} \iff e'_i \text{ has the opposite orientation as } r$$

Let $X = \{e'_{-N}, e'_{-N+1}, ..., e'_{-1}, e'_0 = r, e'_1, ..., e'_N\}$. For any $e' = (v_p, v_q) \in X$, let $Aug(e') = (v_p, v_q)$ if $y^*_{e'} > x^*_{\Gamma(e')}$, and $Aug(e') = (v_q, v_p)$ if $y^*_{e'} < x^*_{\Gamma(e')}$. Then, each $\Gamma(Aug(e'))$ is an arc on the residual graph $G(x^*)$, and $W = (Aug(e'_{-N}), Aug(e'_{-N+1}), ..., Aug(e'_0), ..., Aug(e'_N))$ is a directed path on $T^N_{e_0}$. We call $W$ the *augmenting path* of $y^*$ with respect to $x^*$. Also, $\Gamma(W)$ is a directed
walk on \( G(x^*) \). Now we can decompose \( \Gamma(W) \) into a simple directed path \( P(W) \), and a set of simple directed cycles \( \text{Cyc}(W) \). Because each simple directed cycle/path on \( G(x^*) \) can have at most \( n \) arcs, and \( W \) has \( 2N + 1 \) arcs, where \( N \geq \left( \frac{1}{\delta(x^*)} \right) + 1 \), we have that \( |\text{Cyc}(W)| > \frac{1}{\delta(x^*)} \). Then, we have

\[
c^*(W) = \sum_{C \in \text{Cyc}(W)} c^*(C) + c^*(P(W))
\]

\[
\geq \sum_{C \in \text{Cyc}(W)} c^*(C) - L
\]

\[
> \frac{L}{\delta(x^*)} \cdot \delta(x^*) - L
\]

\[
= 0
\]

Let \( \text{Forw} = \{e | e \in X, y^*_e > x^*_T(e)\} \), \( \text{Back} = \{e | e \in P, y^*_e < x^*_T(e)\} \). Since both \( \text{Forw} \) and \( \text{Back} \) are finite, we can find \( \lambda > 0 \) such that \( y^*_e - \lambda \geq x^*_T(e), \forall e \in \text{Forw} \)

and \( y^*_e + \lambda \leq x^*_T(e), \forall e \in \text{Back} \). Let \( \tilde{y}_e \in \mathbb{R}^{|E(T^N_n)|} \) such that

\[
\tilde{y}_e = \begin{cases} 
  y^*_e - \lambda & : e \in \text{Forw} \\
  y^*_e + \lambda & : e \in \text{Back} \\
  0 & : \text{otherwise}
\end{cases}
\]

We can think \( \tilde{y} \) as pushing \( \lambda \) units of flow on \( W \) for \( y^* \). Since for each \( e \in \text{Forw} \), \( y^*_e - \lambda \geq x^*_T(e) \geq 0 \), and for each \( e \in \text{Back} \), \( y^*_e + \lambda \leq x^*_T(e) \leq u_T(e) \), \( \tilde{y} \) satisfy all the flow constraints. Also, because \( \text{Forw} = \{e | e \in X, e \ has \ the \ same \ orientation \ as \ r\} \), and \( \text{Back} = \{e | e \in X, e \ has \ the \ opposite \ orientation \ as \ r\} \), we have that for any \( v \in V \), \( \sum_{e \in E_v} \Delta(v, e) \tilde{y}_e = \sum_{e \in E_v} \Delta(v, e) y^*_e = b_T(e) \), which implies \( \tilde{y} \) satisfies all the demand/supply constraints. Therefore, \( \tilde{y} \) is feasible solution for \( (\mathcal{MC}_F^N) \). But

\[
\sum_{e \in E(T^N_n)} c_{T(e)}(\tilde{y}_e) - \sum_{e \in E(T^N_n)} c_{T(e)}(y^*_e)
\]

\[
= \sum_{e \in \text{Forw}} c_{T(e)}(\lambda) - \sum_{e \in \text{Back}} c_{T(e)}(\lambda)
\]

\[
= c^*(W) \lambda
\]

\[
> 0
\]

This contradicts the optimality of \( y^* \), and completes the proof of Theorem 3.1.

**Proof.** The proof of this corollary uses the same idea of finding an augmenting path on the computation tree as in Theorem 3.1.

\((\Rightarrow)\): Suppose \( \mathcal{MC}_F \) has a unique optimal solution. Let \( y \) be the optimal solution \( \mathcal{MC}_F^N \), when the variable at \( r, y_r \) is fixed to be \( z^*_r - 1 \). Then, we can find an augmenting path \( W \) of length \( 2n^2c_{\max} \). Then \( W \) can be decomposed into at least \( 2n^2c_{\max} \) disjoint cycles and one path. Since each cycle has a cost of at most \(-\delta(x^*)\), which is at least -1 as \( \mathcal{MC}_F \) has integral data. Hence, once we push 1 unit of flow of \( y \) on \( W \), we can decrease the cost of \( y \) by at least \( n^2c_{\max} \). Hence \( n^2N(z^*_e - 1) + n^2c_{\max} < n^2N(z^*_e) \). Similarly, we can also show that \( n^2N(z^*_e) < n^2N(z^*_e + 1) + n^2c_{\max} \).

\((\Leftarrow)\): Suppose \( \mathcal{MC}_F \) does not have a unique optimal solution. Let \( x^* \) be an optimal solution of \( \mathcal{MC}_F \), and \( x^* \) be the optimal solution \( \mathcal{MC}_F^N \). Again, we can find an augmenting path \( W \) of \( y \) with respect to \( x^* \), and can decompose \( W \) into at least \( 2n^2c_{\max} \) disjoint cycles and one path \( P \). The cost of \( P \) is at most \( (n - 1)c_{\max} \), while each cycle has a non-positive cost. Hence, when we push 1 unit of flow of \( y \) on \( W \), we increase the cost of \( y \) by at most \( (n - 1)c_{\max} \). Therefore, depends on the orientation of \( r \) on \( W \), either \( n^2N(z^*_e + 1) - n^2c_{\max} \leq n^2N(z^*_e) \) or \( n^2N(z^*_e + 1) - n^2c_{\max} > n^2N(z^*_e) \).

Theorem 4.1 shows that BP can be used to detect the existence of an unique optimal solution for \( \mathcal{MC}_F \).

## 5 Extension of BP on Convex-Cost Network Flow Problems

In this section, we discuss the extension of Theorem 3.1 to convex-cost network flow problem (or convex flow problem). Convex flow problem is defined on graph \( G = (V, E) \) as follows:

\[
\text{min} \sum_{e \in E} c_e(x_e) \quad \text{(CP)}
\]

\[
\sum_{e \in E_v} \Delta(v, e)x_e = b_v, \quad \forall v \in V
\]

\[
0 \leq x_e \leq u_e, \quad \forall e \in E
\]

where each \( c_e \) is a convex piecewise-linear function. Notice that if we define \( \psi \) exactly same as we did for \( \mathcal{MC}_F \), and for each \( e \in E \), define

\[
\phi_e(z) = \begin{cases} 
  c_e(z) & \text{if } 0 \leq z \leq u_e \\
  \infty & \text{otherwise}
\end{cases}
\]

then, we can apply Algorithm 2 on the graph \( G \) with functions \( \psi \) and \( \phi \), and hence obtain BP on convex flow problem.
Now, we give the definition of a residual graph on convex cost flow problem. Suppose \( x \) is a feasible solution for \((CP)\), let \( G(x) \) be the residual graph for \( G \) and \( x^* \) defined as follows: \( \forall e = (v_\alpha, v_\beta) \in E, \) if \( x_\alpha < u_\alpha \), then \( e \) is also an arc in \( G(x) \) with cost \( c^e_\alpha = \lim_{t \to 0} \frac{c(e_\alpha + t) - c(e_\alpha)}{t} \) (the value of the slope of \( c_\alpha \) at the right side of \( x_\alpha \); if \( x_\alpha > 0 \), then there is an arc \( e' = (v_\beta, v_\alpha) \) in \( G(x) \) with cost \( c^e_\beta = \lim_{t \to 0} \frac{c(e_\beta - t) - c(e_\beta)}{t} \) (negative value of the slope of \( c_\beta \) at the left side of \( x_\beta \)). Finally, let
\[
\delta(x) = \min_{C \in \mathcal{C}} \left\{ \sum_{e \in \mathcal{C}} c^e \right\},
\]
where \( \mathcal{C} \) is the set of directed cycles in \( G(x) \).

**Theorem 5.1.** Suppose \( x^* \) is the unique optimal solution for \((CP)\). Let \( L \) be the maximum cost of a simple directed path in \( G(x^*) \), and by uniqueness of optimal solution, \( \delta(x^*) > 0 \). Then, for any \( N \geq \left(\frac{1}{2\delta(x^*)}\right) + 1 \), \( \bar{x}^N = x^* \). Namely, the BP algorithm converges to the optimal solution in at most \( \left(\frac{1}{2\delta(x^*)}\right) + 1 \) iterations.

The proof of Theorem 5.1 is almost identical to the proof of Theorem 3.1. Theorem 5.1 is an illustration of the power of BP, that not only it can solve \( \mathcal{MCF} \), but possibly many other variants of \( \mathcal{MCF} \) as well.

### 6 Running Time of BP on \( \mathcal{MCF} \)

In the next two sections, we will always assume vectors \( c, u \) and \( b \) in problem \( \mathcal{MCF} \) are integral. This assumption is not restrictive in practice, as the \( \mathcal{MCF} \) solved in practice usually have rational data sets [1]. We define \( c_{\max} = \max_{e \in E} \{c_e\} \), and provide an upper bound for the running time of BP on \( \mathcal{MCF} \) in terms \( m \) (the number of arcs in \( G \)), \( n \) (the number of nodes in \( G \)), and \( c_{\max} \).

**Lemma 6.1.** Suppose \( \mathcal{MCF} \) has integral data, in Algorithm 2, the number of operations to update all the messages at iteration \( t \) is \( O(nm \log n \cdot t_{c_{\max}}) \).

**Proof.** At iteration \( t \), fix a valid message function \( m_{t-1}^{t} \), and we update it as:
\[
m_{t-1}^{t}(\bar{z}) = \phi_e(\bar{z}) + \min_{\bar{z}_e \in \mathbb{R}^{|E_w|}, \bar{z}_e = \bar{z}} \left\{ \psi_w(\bar{z}) + \sum_{\bar{z}_e \in E_u \setminus e} m_{t-1}^{t}(\bar{z}_e) \right\}
\]

As \( \mathcal{MCF} \) has integral data, by Corollary 3.3, any messages of the form \( m_{t-1}^{t} \) has integral slopes, and the absolute values of the slopes are bounded by \( (t - 1)c_{\max} \). This implies for each \( m_{t-1}^{t} \), the number of linear pieces of \( m_{t-1}^{t} \) is bounded by \( 2(t - 1)c_{\max} \).

As the size of \( E_w \) is at most \( n \), by Corollary 3.1, \( g(z) = \min_{\bar{z} \in \mathbb{R}^{|E_w|}, \bar{z}_e = z} \left\{ \psi_w(\bar{z}) + \sum_{\bar{z}_e \in E_u \setminus e} m_{t-1}^{t}(\bar{z}_e) \right\} \) can be calculated in \( O(\log n \cdot nt_{c_{\max}}) \) operations. As \( m_{t-1}^{t}(\bar{z}) = g(z) + \phi_e(z) \), we have that \( m_{t-1}^{t} \) can be also calculated in \( O(\log n \cdot nt_{c_{\max}}) \) operations. Since at each iteration, we update \( 2m \) message functions, the total number of computations at iteration \( t \) can be bounded by \( O(nm \log n \cdot t_{c_{\max}}) \).

**Theorem 6.1.** Given \( \mathcal{MCF} \) has a unique optimal solution \( x^* \) and integral data, the BP algorithm finds the unique optimal solution of \( \mathcal{MCF} \) in \( O(n^2 m \log n \cdot c_{\max}^{-1}) \) operations.

**Proof.** Because \( c \) is integral, the value \( \delta(x^*) \) in Theorem 3.1 is also integral. Recall that \( L \) is the maximum cost of a simple directed path in \( G(x^*) \). Since simple directed path has at most \( n - 1 \) arcs, \( L \) is bounded by \( nc_{\max} \). Thus, by Theorem 3.1, the BP algorithm (Algorithm 2) converges to the optimal solution of \( \mathcal{MCF} \) after \( O(n^2 c_{\max}) \) iterations. Combine this with Lemma 6.1, we have that BP converges to the optimal solution in \( O(n^2 c_{\max} \cdot nm \log n \cdot n^2 c_{\max} \cdot c_{\max}) = O(n^5 m \log n \cdot c_{\max}^{-1}) \) operations.

We would like to point out that for \( \mathcal{MCF} \) with a unique optimal solution, Algorithm 2 can take an exponential number of iterations to converge. Consider the \( \mathcal{MCF} \) on the directed graph \( G \) shown in Figure 5. Take a large positive integer \( D \), set \( c_{e_1} = c_{e_2} = D, c_{e_3} = 2D - 1, b_{v_1} = 1, b_{v_2} = 0 \) and \( b_{v_3} = -1 \). It can be checked that \( \bar{x}^N \) alternates between 1 and -1 when \( 2N + 1 < \frac{2D}{2D - 1} \). This means that BP algorithm takes at least \( \Omega(D) \) iterations to converge. Since the input size of a large \( D \) is just \( \log(D) \), we have that Algorithm 2 for \( \mathcal{MCF} \) does not converge to the unique optimal solution in polynomial time.

### 6.1 Runtime of BP on \( \mathcal{MCF}' \)

Here we analyze the run time of BP on \( \mathcal{MCF}' \), which is defined in Section 3.2. We show that BP performs much better on this special class of \( \mathcal{MCF} \). Specifically, we state the following theorem:

**Theorem 6.2.** If \( \mathcal{MCF}' \) has a unique optimal solution, Algorithm 2 for \( \mathcal{MCF}' \) finds the unique optimal solution in \( O(n^4) \) operations.
First, we state the convergence result of BP for $\mathcal{MCF}'$ which is reminiscent of Theorem 3.1.

**Theorem 6.3.** Suppose $\mathcal{MCF}'$ has a unique optimal solution $x^*$. Define $L$ to be the maximum cost of a simple directed path in $G(x^*)$. Then, Algorithm 2 for $\mathcal{MCF}'$ converges to the optimal solution in at most $(\frac{L}{2K}) + 1)n$ iterations.

The proof of this theorem is exactly the same as the proof of Theorem 3.1, and is hence omitted. Now, we apply this result to compute the running time of Algorithm 2 for $\mathcal{MCF}'$.

**Proof.** [Proof of Theorem 6.2] For the simplicity of the proof, we assume that every linear piece in a message function has unit length. This assumption is not restrictive, as each linear piece in general has integral vertices (from Corollary 3.4), so we can always break a bigger linear piece into many unit length linear pieces with the same slopes. At iteration $t \geq 1$, recall $m_{t-1,v}(z)$ is defined as:

$$\phi_t(z) + \min_{\tilde{z} \in \mathbf{R}^{E_{\text{out}}}, \tilde{z}_v = z} \left\{ \psi_w(\tilde{z}) + \sum_{e \in E_{\text{in}}} m_{t-1,w}(\tilde{z}_e) \right\}$$

We claim that each $m_{t-1,v}$ can be constructed using the following information:

1. $\tilde{u}_w$ smallest linear pieces in $\{m_{t-1,w}|e \in \delta(w) \setminus e\}$, where $\delta(w)$ denotes all in-arcs of $w$;
2. $(\tilde{u}_w - b_w)$ smallest linear pieces in $\{m_{t-1,w}|e \in \gamma(w) \setminus e\}$ where $\gamma(w)$ denotes all out-arcs of $w$;
3. The value of $\sum_{e \in E_w} m_{t-1,w}(0)$.

Observe that once all three sets of information is given, for any integer $0 \leq z \leq u_e$, one can find $m_{t-1,v}(z)$ in a constant number of operations. As $u_e$ is also bounded by a constant, we have that $m_{t-1,v}$ can be computed in constant time when information (1), (2) and (3) are given.

Now, since $|E_w|$ is bounded by $n$, obtaining each one of information (1), (2) or (3) for a specific message $m_{t-1,v}$, takes $O(n)$ operations. When we update all the messages in the set $\{m_{t-1,v}|e \in E_w\}$, note that every $e \in E_w$ except for a constant number of them, information (1) for updating $m_{t-1,v}$ is equivalent to $\tilde{u}_w$ smallest pieces in $\{m_{t-1,v}|\tilde{e} \in \delta(w)\}$, for some $\tilde{e} \in \delta(w)$; and information (2) for updating $m_{t-1,v}$ is equivalent $\tilde{u}_w - b_w$ smallest linear pieces in $m_{t-1,v}$ for some $\tilde{e} \in \gamma(w)$. In fact, the exceptions are precisely those $e$ such that $m_{t-1,v}$ contain at least one of those $\tilde{u}_w$ smallest linear pieces in $\{m_{t-1,v}|e \in \delta(w)\}$ or one of those $(\tilde{u}_w - b_w)$ smallest linear pieces in $\{m_{t-1,v}|e \in \gamma(w)\}$. Moreover, update any message in the set $\{m_{t-1,v}|e \in E_w\}$ has information (3) to be exactly $\sum_{e \in E_w} m_{t-1,v}(0)$. Hence, to update all messages in the set $\{m_{t-1,v}|e \in E_w\}$, we can obtain information (1), (2), (3) for all the messages in $\{m_{t-1,v}|e \in E_w\}$ in $O(n)$ operations, and hence can compute all messages in $\{m_{t-1,v}|e \in E_w\}$ in $O(n)$ operations.

Therefore, updating all the message functions at every iteration takes $O(n^2)$ operations. By Theorem 6.2, since $L$ is bounded by $nK$, $\delta(x^*) \geq 1$, then $(\frac{L}{2K}) + 1)n$ is bounded by $O(n^2)$. This concludes that Algorithm 2 finds a unique optimal solution in $O(n^2)$ operations.

Note both the shortest-path problem and bipartite matching problem belongs to the class of $\mathcal{MCF}'$, where $\tilde{u}$, $b$, $u$ are all bounded by $2$.

### 7 FPRAS Implementation

In this section, we provide a fully polynomial-time randomized approximation scheme (FPRAS) for $\mathcal{MCF}$ using BP as a subroutine. First, we describe the main idea of our approximation scheme.

As Theorem 6.1 indicated, in order to come up with an efficient approximation scheme using BP (Algorithm 2), we need to get around the following difficulties:

1. The convergence of BP requires $\mathcal{MCF}$ to have a unique optimal solution.
2. The running time of BP is polynomial in $m$, $n$ and $c_{\text{max}}$.

For an instance of $\mathcal{MCF}$, our strategy for finding an efficient $(1+\epsilon)$ approximation scheme is to find a modified cost vector $\tilde{c}$ (let $\mathcal{MCF}'$ be the problem with the modified cost vector) which satisfies the following properties:

1. The $\mathcal{MCF}'$ has a unique optimal solution.
2. $\tilde{c}_{\text{max}}$ is polynomial in $m$, $n$ and $\frac{1}{\epsilon}$.
3. The optimal solution of $\mathcal{MCF}'$ is a “near optimal” solution of $\mathcal{MCF}$. The term “near optimal” is rather fuzzy and we will address this question later in the section.

First, in order to find a modified $\mathcal{MCF}'$ with a unique optimal solution, we first state a result which is a variant of the Isolation Lemma introduced in [15].

#### 7.1 Variation of Isolation Lemma

**Theorem 7.1.** Let $\mathcal{MCF}$ be an instance of min-cost flow problem with underlying graph $G = (V,E)$, demand vector $b$, constraint vector $u$. Let its cost vector $c$ be generated as follows: for each $e \in E$, $\tilde{c}_e$ is chosen
independently and uniformly over \( N_e \), where \( N_e \) is a discrete set of \( 4m \) positive numbers (\( m = |E| \)). Then, the probability that \( \mathcal{MCF} \) has a unique optimal solution is at least \( \frac{1}{2} \).

**Proof.** Fix an arc \( c_1 \in E \), and fix \( \bar{c}_e \) for all \( e \in E \setminus e_1 \). Suppose that there exists \( \alpha > 0 \), such that when \( \bar{c}_{e_1} = \alpha \), \( \mathcal{MCF} \) have optimal solutions \( x^*, x^{**} \), where \( x^*_1 = 0 \) and \( x^{**}_1 > 0 \). Then, if \( \bar{c}_{e_1} > \alpha \), for any feasible solution \( x \) of \( \mathcal{MCF} \), where \( x_{e_1} > 0 \), we have

\[
\sum_{e \in E} \bar{c}_e x_e^* = \sum_{e \in E, e \neq e_1} \bar{c}_e x_e^* \\
\leq \sum_{e \in E, e \neq e_1} \bar{c}_e x_e + x_{e_1} \alpha \\
< \sum_{e \in E} \bar{c}_e x_e
\]

And if \( \bar{c}_{e_1} < \alpha \), for any feasible solution \( x \) of \( \mathcal{MCF} \) where \( x_{e_1} = 0 \), we have

\[
\sum_{e \in E} \bar{c}_e x_e^{**} = \sum_{e \in E, e \neq e_1} \bar{c}_e x_e^{**} + \alpha x_{e_1}^{**} \\
\leq \sum_{e \in E, e \neq e_1} \bar{c}_e x_e + \alpha x_{e_1} \\
= \sum_{e \in E} \bar{c}_e x_e
\]

This implies there exists at most one value for \( \alpha \), such that if \( \bar{c}_{e_1} = \alpha \) then \( \mathcal{MCF} \) have optimal solutions \( x^*, x^{**} \), where \( x^*_1 = 0 \) and \( x^{**}_1 > 0 \). Similarly, we can also deduce that there exists at most one value for \( \beta \), such that if \( \bar{c}_{e_1} = \beta \), \( \mathcal{MCF} \) have optimal solutions \( x^*, x^{**} \), where \( x^*_1 < u_{e_1} \) and \( x^{**}_1 = u_{e_1} \).

Let \( OS \) denote the set of all optimal solutions of \( \mathcal{MCF} \), and let \( D(e) \) be the condition of either: \( 0 = x_e, \forall x \in OS \), or \( 0 < x_e < u_e, \forall x \in OS \) or \( x_e = u_e, \forall x \in OS \). Since \( \bar{c}_{e_1} \) has \( 4m \) possible values, where each value is chosen with equal probability, we conclude that the probability of \( D(e_1) \) is satisfied is at least \( \frac{4m-2}{4m} = \frac{2m-1}{2m} \). By the union bound of probability, we have that the probability of \( D(e) \) is satisfied for all \( e \in E \) is at least \( 1 - \sum_{e \in E} \frac{1}{2m} = \frac{1}{2} \). Now, we state the following lemma:

**Lemma 7.1.** If \( \forall e \in E \), condition \( D(e) \) is satisfied, then \( \mathcal{MCF} \) have a unique optimal solution.

Then by Lemma 7.1, we conclude that the probability that \( \mathcal{MCF} \) has a unique optimal solution is at least \( \frac{1}{2} \).

**Proof.** [Proof of Lemma 7.1] Suppose \( x^* \) and \( x^{**} \) are two distinct optimal solutions of \( \mathcal{MCF} \). Let \( d = x^{**} - x^* \), then \( x^* + \lambda d \) is an optimal solution of \( \mathcal{MCF} \) iff \( 0 \leq (x^* + \lambda d)_e \leq u_e, \forall e \in E \). As \( \bar{c}_{e_1} > 0 \) for any \( e \in E \), and \( \overline{c^T}d = \overline{c^T}x^{**} - \overline{c^T}x^* = 0 \), there exists some \( \lambda' \) such that \( d_{\lambda'} < 0 \). Let \( \lambda^* = \sup \{ \lambda | x^* + \lambda d \) is an optimal solution of \( \mathcal{MCF} \} \), since \( d_{\lambda'} < 0 \), \( \lambda^* \) is bounded; and since \( x^* + d = x^{**} \), \( \lambda^* > 0 \). By optimality of \( \lambda^* \), there must exists some \( \lambda'' \) such that either \( (x^* + \lambda^* d_{\lambda''}) = 0 \) or \( u_e \). Since \( \lambda^* > 0 \), \( x^*_{\lambda''} \neq (x^* + \lambda^* d_{\lambda''})_e \), this contradicts the assumption that \( D(e') \) is satisfied. Thus, \( \mathcal{MCF} \) must have a unique optimal solution.

We note that Theorem 7.1 can be easily modified for LP in standard form.

**Corollary 7.1.** Let \( \mathcal{LP} \) be an LP problem with constraint \( Ax = b \), where \( A \) is an \( m \times n \) matrix, \( b \in \mathbb{R}^m \). The cost vector \( \overline{c} \) of \( \mathcal{LP} \) is generated as follows: for each \( e \in E \), \( \bar{c}_e \) is chosen independently and uniformly over \( N_e \), where \( N_e \) is a discrete set of \( 2n \) elements. Then, the probability that \( \mathcal{LP} \) has a unique optimal solution is at least \( \frac{1}{4} \).

### 7.2 Finding the Correct Modified Cost Vector \( \bar{c} \)

Next, we provide a randomly generate \( \bar{c} \) with the desired properties stated in the beginning of this section. Let \( X : E \rightarrow \{1, 2, ..., 4m\} \) be a random function where for each \( e \in E \), \( X(e) \) is chosen independently and uniformly over the range. Let \( t = \frac{4m}{\min \{\bar{c}_e \}} \), and generate \( \bar{c} \) as: for each \( e \in E \), let \( \bar{c}_e = 4m \cdot \frac{1}{t} + X(e) \). Then, for any \( \bar{c} \) generated at random, \( \bar{c}_{\text{max}} \) is polynomial in \( m, n \) and \( \frac{1}{t} \), and by Theorem 7.1, the probability of \( \mathcal{MCF} \) having a unique optimal solution is greater than \( \frac{1}{4} \).

Now, we introduce algorithm \( \text{APRXMT}(\mathcal{MCF}, \epsilon) \), which works as follows: select a random \( \bar{c} \); try to solve \( \mathcal{MCF} \) using BP. If BP discovers that \( \mathcal{MCF} \) has no unique optimal solution, then we start the procedure by selecting another \( \bar{c} \) at random, otherwise, return the unique optimal solution found by BP. Formally, we present \( \text{APRXMT}(\mathcal{MCF}, \epsilon) \) as Algorithm 3.

**Corollary 7.2.** The expected number of operations for \( \text{APRXMT}(\mathcal{MCF}, \epsilon) \) to terminate is \( O(\frac{m^{1/2} \log n}{\epsilon^3}) \).

**Proof.** With Theorem 7.1, when we call \( \text{APRXMT}(\mathcal{MCF}, \epsilon) \), the expected number of \( \mathcal{MCF} \) BP tried to solve is bounded by 2. For each selection of \( \bar{c} \), we run Algorithm 2 for \( 2\bar{c}_{\text{max}} n^2 \) iterations. As \( \bar{c}_{\text{max}} = O(\frac{1}{\epsilon^3}) \), by Lemma 6.1, the expected number of operations for \( \text{APRXMT}(\mathcal{MCF}, \epsilon) \) to terminate is \( O(\frac{c_{\text{max}}^3 n^5 m \log n}{\epsilon^3}) \).

Now let \( \bar{c} \) be a randomly chosen vector such that \( \mathcal{MCF} \) has a unique optimal solution \( x^* \). The next thing we want is to show that \( x^* \) is a “near optimal” solution of \( \mathcal{MCF} \). To accomplish this, let \( e' = \arg \max\{e_c\} \) and define
Algorithm 3 \textsc{APRXMT}(\textit{MCF}, \epsilon)

\begin{enumerate}
\item Let $t = \frac{\text{maxc}}{4mn}$, for any $e \in E$, assign $\tilde{c}_e = 4m \cdot \left\lceil \frac{t}{4} \right\rceil + p_e$, where $p_e$ is an integer chosen independently, uniformly random from \{1, 2, ..., 4m\}.
\item Let $\textit{MCF}$ be the problem with modified cost $\tilde{c}$.
\item Run Algorithm 2 on $\textit{MCF}$ for $N = 2\tilde{c}_{\text{max}}n^2$ iterations.
\item Use Corollary 4.1 to determine if $\textit{MCF}$ has a unique solution.
\item If $\textit{MCF}$ does not have a unique solution then
\begin{enumerate}
\item Terminate and return $x^2 = \hat{x}^N$, where $\hat{x}^N$ if the “belief” vector found in Algorithm 2.
\end{enumerate}
\end{enumerate}

\begin{align*}
\min_{e \in E} \sum_{e \in E} c_e x_e & \quad (\textit{MCF}) \\
\sum_{e \in E} \Delta(v, e)x_e = b_v, \quad \forall v \in V \quad \text{(demand constraints)} \\
x_{e'} = x_{e}^2, \quad 0 \leq x_e \leq u_e, \quad \forall e \in E \quad \text{(flow constraints)}
\end{align*}

Suppose $x^3$ is an optimal solution for $(\textit{MCF})$, and $x^1$ is an optimal solution of $\textit{MCF}$, then we have that $\epsilon'k \leq 0$. And

\begin{align*}
\tilde{c}_e = 4m \left\lceil \frac{c_e}{t} \right\rceil + p_e, \quad 1 \leq p_e \leq 4m, \\
\implies \tilde{c}_e, 4mc_e \leq 4m \left\lceil \frac{c_e}{t} \right\rceil + 4m \left( \frac{c_e}{t} \right) + 1 \right\rceil, \\
\implies |\tilde{c}_e - 4mc_e| \leq 4m, \\
\implies \sum c |4mc_e - \tilde{c}_e||k_e| \leq \sum |4m||k_e| \leq 4mn,
\end{align*}

but
\begin{align*}
\frac{4mc_e'}{t} \leq \frac{4mc_k'}{t} - \epsilon'k \leq \sum \frac{4mc_e}{t} - \tilde{c}_e||k_e|, \\
\text{thus, we have} \quad \frac{4mc_e'}{t} \leq 4mn,
\end{align*}

$\implies \epsilon'k \leq nt$.

Clearly, $x^1 + \sum_{k \in K'} k$ satisfy the demand/supply constraints of $\textit{MCF}$. Since $\min\{x_{k'}, x_{k'}^2\} \leq x^1 + \sum_{k \in K'} k \leq \max\{x^1, x^2\}$ for all $e \in E$, and each $k \in K'$, we have that $x^1 + \sum_{k \in K'} k$ is a feasible solution for $\textit{MCF}$. Since $x^3$ is the optimal solution of $\textit{MCF}$, $c^T x^3 \leq c^T x^1 + \sum_{k \in K'} c^T k \leq c^T x^1 + |K'|nt$. Since $|K'| = |x_{e'}^2 - x_{e'}^2|$, we have $c^T x^3 \leq |x_{e'}^2 - x_{e'}^2|nt$.

\textbf{Corollary 7.3}. $c^T x^3 \leq (1 + \frac{\epsilon}{2m})c^T x^1$, for any $\epsilon \leq 2$.

\textbf{Proof}.

By Lemma 7.2,\begin{align*}
\frac{c^T x^3 - c^T x^1}{c^T x^1} \leq \frac{|x_{e'}^2 - x_{e'}|nt}{c^T x^1} \\
\leq \frac{|x_{e'}^2 - x_{e'}^2|nt}{|x_{e'}^2 - x_{e'}^2||c_e^2|} \\
= \epsilon \frac{4m}{4m} = \epsilon, \text{ as } t = \frac{c_{e'}^2}{4m}.
\end{align*}

Thus,
\begin{align*}
\frac{c^T x^3 - c^T x^1}{c^T x^3} \leq \frac{\epsilon}{4m} \\
\implies c^T x^3 \leq \frac{1}{1 - \frac{\epsilon}{4m}}c^T x^1 \\
\implies c^T x^3 \leq (1 + \frac{\epsilon}{2m})c^T x^1
\end{align*}

where the last inequality holds because $(1 - \frac{\epsilon}{4m})(1 + \frac{\epsilon}{2m}) = 1 + \frac{\epsilon}{4m} - \frac{\epsilon^2}{8m} \geq 1, \text{ as } \frac{\epsilon^2}{8m} \leq 1$.

\textbf{7.3 A (1 + $\epsilon$) Approximation Scheme} Loosely speaking, Corollary 7.3 shows that $x^2$ at arc $e'$ is “near optimal”, since fixing the flow at arc $e'$ to $x_{e'}^2$ helps us in finding a feasible solution of $\textit{MCF}$ which is close to optimal. This leads us to approximate algorithm $\text{AS}(\textit{MCF}, \epsilon)$ (see Algorithm 4); at each iteration, it uses $\text{APRXMT}$ (Algorithm 3), and iteratively fixes the flow at the arc with the largest cost.
In this paper, we formulated belief propagation (BP) for \( \mathcal{MCF} \) running time of BP for existing algorithms for \( \mathcal{MCF} \) as an optimization solver. Although the result from [6], and provides new insights for under-
the optimal solution is unique. This result generalizes
existing algorithms for \( \mathcal{MCF} \) for other variants of network flow problems. One such
heuristic, and one should be able to modify it to
work in the absence of a unique optimal solution.
In the approximation scheme, the heuristic of fixing
values on variables while running BP is commonly
known as ‘decimation’ (see [14]). To the best of
out knowledge, this is the first disciplined, provable
instance of decimation procedure. The ‘decimation’ in
our scheme is extremely conservative, and a natural
question is if there exists a more aggressive ‘decimation’
heuristic which can improve the running time.

Acknowledgements

While working on this paper, D. Gamarnik was partially
supported by NSF Project CMMI-0726733; D. Shah was
supported in parts by NSF EMT Project CCF 0829893
and NSF CAREER Project CNS 0546590; and Y. Wei
was partially supported NSERC PGS M. The authors
would also like to thank the anonymous referees for the
helpful comments.

References

passing algorithms and improved lp decoding. 2009.
arbitrary graphs and its relation to linear programs
analysis of the cavity equations for the mini-
mum spanning tree. Journal of Mathematical Physics,
maximum weight matching: Convergence, correctness,
and lp duality. IEEE Transaction on Information
Optimization, pages 289–290. Athena Scientific, third
thesis, Massachusetts Institute of Technology, Cam-
bridge, MA, 1963.
[9] G. B. Horn. Iterative Decoding and Pseudocode-
words. PhD thesis, California Institute of Technology,
[10] D. Malioutov, J. K. Johnson, and A. S. Willsky. \( \star \) Walk-sums and belief propagation in gaussian graphi-
and algorithmic solution of random satisfiability problems.
[12] C. Moallemi and B. V. Roy. Convergence of min-
sum message passing for convex optimization. In 45th

Algorithm 4 AS(\( \mathcal{MCF}, \epsilon \))

1: Let \( G = (V, E) \) be the underlying directed graph of
\( \mathcal{MCF} \), \( m = |E| \), and \( n = |V| \).
2: while \( \mathcal{MCF} \) contains at least 1 arcs do
3: Run APRXMT(\( \mathcal{MCF}, \epsilon \)), let \( x^2 \) be the solution returned.
4: Find \( (i', j') = e' = \arg \max_{e \in E} \{ c_e \} \), modify \( \mathcal{MCF} \) by fix the flow on arc \( e' \) by \( x^2 \), and change the demands/supply on nodes \( i', j' \) accordingly.
5: end while


