Uniformly Reweighted Belief Propagation: A Factor Graph Approach

Henk Wymeersch
Chalmers University of Technology
Gothenburg, Sweden
henkw@chalmers.se

Federico Penna
Politecnico di Torino
Torino, Italy
federico.penna@polito.it

Vladimir Savić
Universidad Politecnica de Madrid
Madrid, Spain
vladimir@gaps.ssr.upm.es

***Abstract*—Tree-reweighted belief propagation is a message passing method that has certain advantages compared to traditional belief propagation (BP). However, it fails to outperform BP in a consistent manner, does not lend itself well to distributed implementation, and has not been applied to distributions with higher-order interactions. We propose a method called uniformly-reweighted belief propagation that mitigates these drawbacks. After having shown in previous works that this method can substantially outperform BP in distributed inference with pairwise interaction models, in this paper we extend it to higher-order interactions and apply it to LDPC decoding, leading performance gains over BP.**

I. Introduction

Belief propagation (BP) [1] is a powerful method to perform approximate inference, and has found applications in a wide variety of fields [2]. Furthermore, it provides new interpretations of existing algorithms, such as the Viterbi algorithm [3]. BP can be seen as a message passing algorithm on a graphical model. When this graphical model, which represents the factorization of an un-normalized distribution, is cycle-free (i.e., a tree), BP is guaranteed to converge and can provide marginal distributions as well as the normalization constant of the distribution (also known as the partition function). Moreover, BP can be interpreted as an iterative scheme to find stationary points of a convex variational optimization problem [4]. Unfortunately, when the graphical model contains cycles, the variational problem is no longer guaranteed to be convex, so that BP may end up in a local optimum, or fail to converge altogether.

To address the problem of non-convexity, variations of BP have been proposed that are provably convex, or allow to compute a lower or upper bound of the original optimization problem. One such variation is tree-reweighted BP (TRW-BP), which corresponds to a convex upper approximation of the original objective function (to be introduced later) with relaxed constraints [5]. In its original formulation, the authors only provided explicit message passing rules for TRW-BP in the case of a particular graphical model (a Markov random field) with a particular factorization (pairwise interactions). The extension to higher-order interaction was described through hypergraphs, though no explicit message passing rules were derived. Explicit rules for higher-order interactions were derived in [6] for fractional BP, of which TRW-BP was shown to be a special case. TRW-BP has been observed to outperform BP in some, though not all cases [5]. Moreover, TRW-BP requires an optimization of a distribution over spanning trees, making it unsuitable for distributed inference problems (e.g., over wireless networks). In our prior work on distributed positioning [7] and distributed detection [8], we have proposed a novel message passing scheme, uniformly-reweighted belief propagation (URW-BP), which encompasses both BP as well TRW-BP over uniform graphs, and does not involve an optimization of a distribution over spanning trees. This method was limited to distributions with pairwise interactions.

In this paper, we extend URW-BP to higher-order interaction. Rather than relying on hypergraphs as in [5], we employ factor graphs. We formulate a variational problem and determine the stationary points of the Lagragian, leading to explicit message passing rules for higher-order interactions. We also apply these new message passing rules to the example of decoding an LDPC code, and make connections with another recently proposed decoder [9].

II. The Inference Problem

We consider an a posteriori distribution of the following form

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x})}{p(\mathbf{y})}, \tag{1}$$

where $\mathbf{y}$ is a (fixed) observation and $\mathbf{x} = [x_1, x_2, \dots, x_N]$ is the unobserved variable of interest. We will assume that x_n is a discrete random variable defined over a finite set. Both the likelihood function $p(\mathbf{y}|\mathbf{x})$ as well as the prior $p(\mathbf{x})$ are assumed to be known. The value $p(\mathbf{y})$ is unknown, and in sometimes referred to as the partition function. Moreover, we assume a factorization exists of the form

$$p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) = \prod_{n=1}^{N} \phi_n(x_n) \prod_{l=1}^{L} \psi_l(x_{C_l}), \tag{2}$$

where $x_{C_l} \subset \mathbf{x}$ contains at least two variables, indexed by the set $C_l \subset \{1, \dots, N\}$. Typical goals in inference include (i) determining the marginal a posteriori distributions, $p(x_n|\mathbf{y})$, for every component x_n; (ii) computing $p(\mathbf{y})$.

This research is supported, in part, by the FPU fellowship from Spanish Ministry of Science and Innovation, ICT project FP7-ICT-2009-4-248894-WHERE-2, program CONSOLIDER-INGENIO 2010 under grant CSD2008-00010 COMONSENS, and the European Research Council, under grant COOPNET No. 258418.

For a given $\mathbf{y}$ and a given $\mathbf{x}$, evaluating $p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$ is straightforward. However, brute-force computation of $p(x_n|\mathbf{y})$ or $p(\mathbf{y})$ is intractable, due to the high-dimensional nature of $\mathbf{x}$. A factorization of the form (2) can be conveniently expressed by a graphical model, such as a factor graph or a Markov random field. By executing a message passing algorithm (e.g., BP or TRW-BP) on this graphical model, one can approximate $p(x_n|\mathbf{y})$ as well as $p(\mathbf{y})$ [4], [5].

III. Pairwise Interactions: TRW-BP and BP

When there are only pairwise interactions, every C_l contains exactly 2 elements. For notational convenience, we denote $\psi_l(x_m, x_n)$ by $\psi_{mn}(x_m, x_n)$. As an example, consider a factorization

$$p(\mathbf{x}|\mathbf{y}) \propto \phi_1(x_1)\phi_2(x_2)\phi_3(x_3)\psi_{12}(x_1,x_2)\psi_{23}(x_2,x_3).$$

The corresponding Markov random field is depicted in Fig. 1. We define "neighbor set" $\mathcal{N}_m$ of variable X_m as follows: $X_n \in \mathcal{N}_m$ if and only if there exists a factor $\psi_{mn}(x_m, x_n)$. Then, the message from X_m to a neighboring variable X_n is given by [5, eq. (39)]

$$M_{mn}(x_n) \propto \sum_{x_m} \phi_m(x_m)\psi_{mn}^{1/\rho_{mn}}(x_n,x_m)\frac{\prod_{k\in\mathcal{N}_m\setminus\{n\}} M_{km}^{\rho_{km}}(x_m)}{M_{nm}^{1-\rho_{mn}}(x_m)}, \tag{3}$$

where $\rho_{nm} \equiv \rho_{mn}$ is the so-called edge appearance probability (EAP) of factor ψ_{nm}. The terminology of EAP is due to the representation of ψ_{nm} as an edge in the Markov random field representation of ψ_{nm} in (2). The so-called beliefs, which are an approximation of the a posteriori marginals $p(x_n|\mathbf{y})$, are given by

$$b_n(x_n) \propto \phi_n(x_n)\prod_{m\in\mathcal{N}_n} M_{mn}^{\rho_{nm}}(x_n). \tag{4}$$

Equations (3)–(4) are iterated until the beliefs converge. The set of valid EAPs is non-trivial: given a Markov random field graph G and the set $\mathfrak{T}(G)$ of all possible trees, we can introduce a distribution over the trees: $0 \le \rho(T) \le 1$, for $T \in \mathfrak{T}(G)$, with $\sum_{T\in\mathfrak{T}(G)} \rho(T) = 1$. For a given distribution $\rho(T)$, the EAP of edge (n, m) is then given by

$$\rho_{nm} = \sum_{T\in\mathfrak{T}(G)} \rho(T) \times \mathbb{I}\{(n,m)\in T\},$$

where $\mathbb{I}\{\cdot\}$ is the indicator function. Note that when G is a tree, $\rho_{nm} = 1$, for all edges (n, m). When the graph G contains cycles, $\rho_{nm} < 1$ for at least one edge. BP can now be interpreted as a variation of TRW-BP, where $\rho_{nm} = 1$, for all edges (n, m), irrespective of the structure of G. Note that this is not a valid choice of EAPs for a graph with cycles.

IV. Extension to Higher-Order Interactions

A. Factor Graphs

When some of the cliques C_l contain 3 or more elements, a factor graph can more elegantly capture those higher-order interactions than a Markov random field. A factor graph is a bi-partite graph where we distinguish between variable vertices and factor vertices. A factor vertex and a variable vertex are connected by an edge when the the corresponding variable appears in the corresponding factor (see Fig. 1). When ψ_l has x_n as a variable, we will write $l \in \mathcal{N}_n$.

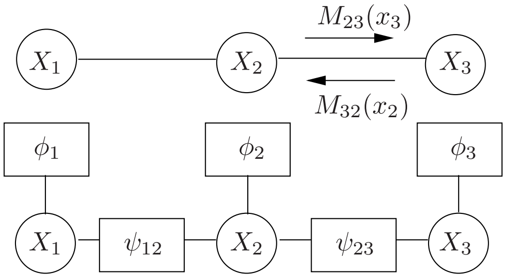

Figure 1. Markov random field (top) and factor graph (bottom) of the factorization $\phi_1(x_1)\phi_2(x_2)\phi_3(x_3)\psi_{12}(x_1,x_2)\psi_{23}(x_2,x_3)$.

We now provide a derivation of TRW-BP message passing rules by following a variational approach, resulting in explicit message expressions. Then, we specialize this to our proposed method, URW-BP.

B. Variational Interpretation

Given any distribution $b(\mathbf{x})$, the Kullback-Leibler divergence between $b(\mathbf{x})$ and $p(\mathbf{x}|\mathbf{y})$ is given by

$$\mathrm{KL}\,(b||p) = \sum_{\mathbf{x}} b(\mathbf{x})\log\frac{b(\mathbf{x})}{p(\mathbf{x}|\mathbf{y})} \ge 0. \tag{5}$$

If we insert (2) into (5), and perform some straightforward manipulations, we can rewrite this inequality as a variational problem:

$$\log p(\mathbf{y}) = \max_{b\in\mathbb{M}(G)}\Big\{\mathcal{H}(b) + \sum_{n=1}^{N}\sum_{x_n} b_n(x_n)\log\phi_n(x_n) + \sum_{l=1}^{L}\sum_{x_{C_l}} b_{C_l}(x_{C_l})\log\psi_l(x_{C_l})\Big\}, \tag{6}$$

where $\mathcal{H}(b)$ denotes the entropy of the distribution $b(\mathbf{x})$, $\mathcal{H}(b) = -\sum_{\mathbf{x}} b(\mathbf{x})\log b(\mathbf{x})$, and $\mathbb{M}(G)$ is the so-called marginal polytope, which is the set of marginal distributions $b_k(x_k)$ and $b_{C_l}(x_{C_l})$ that can be related to a valid distribution, which factorizes according to the same factor graph G, induced by (2). Note that the solution to (6) is $b(\mathbf{x}) = p(\mathbf{x}|\mathbf{y})$ with corresponding maximum equal to $\log p(\mathbf{y})$. The optimization problem (6) turns out to be convex, but, unless G is a tree, intractable: (i) the number of constraints to describe $\mathbb{M}(G)$ is intractable (ii) computing the entropy for any $b(\mathbf{x})$ is intractable.

C. Standard Belief Propagation

In standard BP, a tractable solution is achieved by (i) relaxing $\mathbb{M}(G)$ to the local polytope $\mathbb{L}(G)$, the set of marginal distributions $b_k(x_k)$ and $b_{C_l}(x_{C_l})$ that are normalized, non-negative, and mutually consistent, but not necessarily correspond to a valid global distribution; (ii) approximating the entropy $\mathcal{H}(b)$ by the so-called *Bethe* entropy

$$\mathcal{H}_{\text{Bethe}}(b) = \sum_{n=1}^{N} \mathcal{H}(b_n) - \sum_{l=1}^{L} \mathcal{I}_{C_l}(b_{C_l}),$$

where $\mathcal{H}(b_n)$ is the entropy of $b_n(x_n)$ and $\mathcal{I}_{C_l}(b_{C_l})$ is a mutual information term defined as

$$\mathcal{I}_{C_l}(b_{C_l}) = \sum_{x_{C_l}} b_{C_l}(x_{C_l}) \log \frac{b_{C_l}(x_{C_l})}{\prod_{m\in C_l} b_m(x_m)}.$$

Substitution into (6), expressing the stationary conditions by setting the derivative of the Lagrangian to zero leads to the well-known BP message passing rules [4]:

- Message from variable vertex X_n to factor vertex ψ_l, $n \in C_l$:
$$\mu_{X_n\to\psi_l}(x_n) = \phi_n(x_n) \prod_{k\in\mathcal{N}_n\setminus\{l\}} \mu_{\psi_k\to X_n}(x_n). \tag{7}$$
- Message from factor vertex[1] ψ_l to variable vertex X_n, $n \in C_l$:
$$\mu_{\psi_l\to X_n}(x_n) = \sum_{\sim x_n} \psi_l(x_{C_l}) \prod_{m\in C_l\setminus\{n\}} \mu_{X_m\to\psi_l}(x_m), \tag{8}$$
where $\sum_{\sim x_n}$ refers to the summation over all variables in x_{C_l}, except x_n.
- Belief of variable x_n:
$$b_n(x_n) \propto \phi_n(x_n) \prod_{l\in\mathcal{N}_n} \mu_{\psi_l\to X_n}(x_n). \tag{9}$$

Equations (7)–(9) are iterated until the beliefs converge. When there are only pairwise interactions, it is easy to show that (7)–(9) is equivalent with (3)–(4), when $\rho_{nm} = 1$, for all (n, m).

D. Tree-reweighted Belief Propagation

1) Formulation: Starting from a factor graph of a factorization of (2), we can now generate a tree by removing a suitable subset of factor vertices ψ_l, along with all connected edges. For any such tree T, we can express the corresponding entropy as

$$\mathcal{H}(b;T) = \sum_{n=1}^{N} \mathcal{H}(b_n) - \sum_{C_l\in T} \mathcal{I}_{C_l}(b_{C_l}),$$

where $\sum_{C_l\in T}$ denotes a summation over all factor vertices that remain in the tree. For any tree and a fixed $b(\cdot)$, we know that $\mathcal{H}(b;T) \geq \mathcal{H}(b)$ [10]. We can now introduce a distribution $p(T)$ over such trees, as well as factor appearance probabilities:

$$\rho_l = \sum_{T\in\mathfrak{T}(G)} p(T)\mathbb{I}\left\{C_l \in T\right\}.$$

Given a distribution over trees, a convex upper bound on $\mathcal{H}(b)$ is given by

$$\mathcal{H}(b;\boldsymbol{\rho}) = \sum_{n=1}^{N} \mathcal{H}(b_n) - \sum_{l=1}^{L} \rho_l \mathcal{I}_{C_l}(b_{C_l}).$$

[1]For mathematical convenience, we only compute messages from factor vertices corresponding to factors with at least 2 variables.

We finally arrive to the following variational problem, by relaxing $\mathbb{M}(G)$ to $\mathbb{L}(G)$, and replacing $\mathcal{H}(b)$ by $\mathcal{H}(b;\boldsymbol{\rho})$:

$$\log p(\mathbf{y}) \geq \arg\max_{b\in\mathbb{L}(G)} \Bigg\{ \mathcal{H}(b;\boldsymbol{\rho}) + \sum_{n=1}^{N}\sum_{x_n} b_n(x_n)\log\phi_n(x_n) + \sum_{l=1}^{L}\sum_{x_{C_l}} b_{C_l}(x_{C_l})\log\psi_l(x_{C_l}) \Bigg\}. \tag{10}$$

2) Message Passing Rules: Message update rules are then obtained by expressing the Lagrangian and setting the derivative to zero. We state here the final results; the complete derivation is presented in the Appendix.

- Message from variable vertex X_n to factor vertex ψ_l, $n \in C_l$:
$$\mu_{X_n\to\psi_l}(x_n) = \phi_n(x_n)\mu^{\rho_l-1}_{\psi_l\to X_n}(x_n) \prod_{k\in\mathcal{N}_n\setminus\{l\}} \mu^{\rho_k}_{\psi_k\to X_n}(x_n). \tag{11}$$
- Message from factor vertex ψ_l to variable vertex X_n, $n \in C_l$:
$$\mu_{\psi_l\to X_n}(x_n) = \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \prod_{m\in C_l\setminus\{n\}} \mu_{X_m\to\psi_l}(x_m). \tag{12}$$
- Belief of variable x_n:
$$b_n(x_n) \quad\propto\quad \phi_n(x_n) \prod_{l\in\mathcal{N}_n} \mu^{\rho_l}_{\psi_l\to X_n}(x_n). \tag{13}$$

As we would expect, when $\rho_l = 1$, $\forall l$, we find equations (7)–(9) again. Moreover, when there are only pairwise interactions, we find (3)–(4). It should be noted that equivalent message passing rules were presented in [6] in the context of fractional belief propagation.

E. Uniformly Reweighted Belief Propagation

It is now straightforward to introduce URW-BP. In principle, TRW-BP allows us optimize over all possible distributions $\boldsymbol{\rho}$. Nevertheless, we are not guaranteed that this will lead to better performance than BP. Here we restrict our attention to fixed $\rho_l = \rho$, $\forall l$. This scheme is motivated by the fact that in graphs that are roughly uniform in structure with roughly equal properties of the factor vertices, we expect that a uniform distribution over trees would be optimal, leading to $\boldsymbol{\rho} = \rho\mathbf{1}$, for some $0 < \rho \leq 1$ [10], with $\rho = 1$ when the factor graph is a tree. Hence, for such uniform graphs, we expect URW-BP to achieve the best performance among BP and TRW-BP.

V. Numerical Example

A. Model

We consider a practical example for which, to the best of our knowledge, TRW-BP has not been applied before: decoding of an LDPC code. Consider an LDPC code with an $L \times N$ sparse parity check matrix $\mathbf{H}$. Let $\mathbf{x}$ denote the transmitted codeword

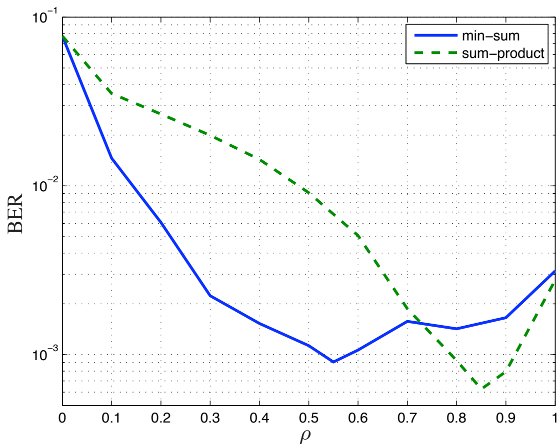

Figure 2. Performance of URW-BP after 20 decoding iterations as a function of the scalar parameter ρ for LDPC code, at a fixed SNR ($E_b/N_0 = 3$ dB). BP corresponds to $\rho = 1$.

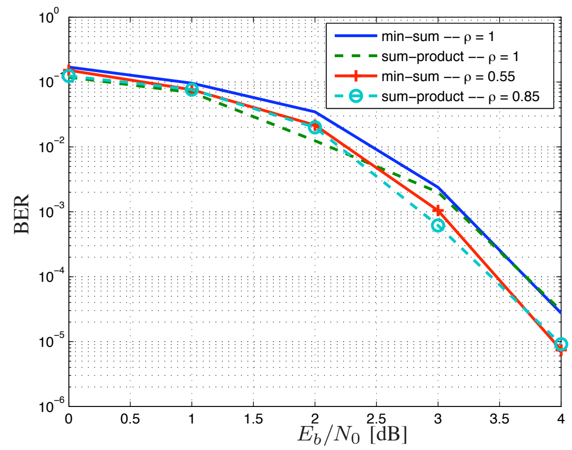

Figure 3. Performance of BP and URW-BP as a function of the SNR for LDPC code after 20 decoding iterations.

and $\mathbf{y}$ the observation over a memoryless channel, with known $p(y_n|x_n)$. The a posteriori distribution of interest is now

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &\propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}) \\ &= \prod_{n=1}^{N} p(y_n|x_n)\mathbb{I}\{\mathbf{Hx} = \mathbf{0}\} \\ &= \prod_{n=1}^{N} p(y_n|x_n) \prod_{l=1}^{L} \mathbb{I}\left\{\sum_{m \in C_l} x_m = 0\right\}, \end{aligned}$$

where the summation is in the binary field, and C_l is the index set corresponding to the non-zero elements of the lth row in $\mathbf{H}$. Clearly, we can make the association $\phi_n(x_n) \leftrightarrow p(y_n|x_n)$ and $\psi_l(x_{C_l}) \leftrightarrow \mathbb{I}\{\sum_{m\in C_l} x_m = 0\}$.

B. Message Passing Rules

The message from variable node X_n to check node ψ_l, expressed as a log-likelihood ratio (LLR), is given by

$$\lambda_{X_n \to \psi_l} = \lambda_{\text{ch},n} + \sum_{k \in \mathcal{N}_n} \rho_k \lambda_{\psi_k \to X_n} - \lambda_{\psi_l \to X_n},$$

where

$$\lambda_{\text{ch},n} = \log \frac{p(y_n|x_n = 1)}{p(y_n|x_n = 0)}.$$

The messages from check nodes to variable nodes are unchanged with respect to standard BP, since here $\psi_l^{1/\rho_l}(x_{C_l}) = \psi_l(x_{C_l})$, irrespective of ρ_l. We will consider two types of messages from check nodes to variable nodes: *sum-product* messages (in the log-domain, see [2, eq. (22)]) and simplified *min-sum* messages (see [9, eq. (14)]). Finally, the beliefs in LLR format are given by

$$\lambda_{b,n} = \lambda_{\text{ch},n} + \sum_{k \in \mathcal{N}_n} \rho_k \lambda_{\psi_k \to X_n}.$$

Interestingly, in its min-sum version, URW-BP is equivalent to a recently proposed LDPC decoder from [9], based on the divide and concur algorithm, provided that ρ is set to $2Z$, where Z was introduced in [9, eq. (15)].

C. Performance Example

In Fig. 2, we plot the bit error rate (BER) performance[2] of a rate $1/2$ LDPC code with $N = 256$ and $L = 128$ at a fixed SNR as a function of ρ. We observe that $\rho = 1$ does not yield the best performance and that the global minimum in the BER is achieved by $\rho \approx 0.85$ and $\rho \approx 0.55$, for sum-product and min-sum, respectively. The optimal value of ρ for min-sum is lower because the min-sum rule tends to overshoot the LLRs more than sum-product. In Fig. 3 we compare the performance of BP and URW-BP (with $\rho = 0.85$ and $\rho = 0.55$ for sum-product and min-sum, respectively) in terms of BER vs. SNR for the same LDPC codes. We see that URW-BP outperforms BP especially at high SNR. The difference in BER between standard BP and URW-BP is not large, but still significant considering that we have used a real LDPC code, designed such that loops are long and have limited impact on BP decoding. The performance gap can be much greater in case of non-optimized graph configurations, i.e., with many short loops.

VI. Conclusions

Motivated by promising results in distributed processing, we have derived uniformly reweighted belief propagation (URW-BP) for distributions with higher-order interactions. We have adopted a factor graph model, as this allows an elegant representation of higher-order interactions, as well as the message passing interpretation. The message passing rules (11), (12), and (13) easily lead to URW-BP. URW-BP combines the potential improved performance of TRW-BP with the distributed nature of BP. We have applied these new message passing rules to LDPC decoding and have shown that URW-BP can outperform BP, even for an LDPC code without short cycles.

[2]For all decoding algorithms, we applied the following stopping rule: when a valid codeword is found at a certain iteration, no further iterations are performed.

APPENDIX

Here we provide the derivation leading to the message passing rules for TRW-BP with higher-order interactions, following the reasoning from [10, Section 4.1.3].

The Local Polytope: We first describe the local polytope $\mathbb{L}(G)$. Any element $b \in \mathbb{L}(G)$ is described by single-variable beliefs $b_n(x_n), n = 1, \ldots, N$ and higher-order beliefs $b_l(x_{C_l})$, $l = 1, \ldots, L$ and must satisfy the following conditions: non-negativity, normalization, and consistency: $b_n(x_n) \geq 0$, $b_{C_l}(x_{C_l}) \geq 0$, $\sum_{x_n} b_n(x_n) = 1$, $\sum_{x_{C_l}} b_{C_l}(x_{C_l}) = 1$, $\sum_{\sim x_n} b_{C_l}(x_{C_l}) = b_n(x_n)$, $n \in C_l$. We will enforce the consistency constraints explicitly with Lagrange multipliers $\lambda_{C_l}(x_n)$, and the non-negativity and normalization constraints implicitly.

The Lagrangian: The Lagrangian is given by

$$\begin{aligned}\mathcal{L}(b, \lambda) &= \mathcal{H}(b; \boldsymbol{\rho}) + \sum_{n=1}^{N} \sum_{x_n} b_n(x_n) \log \phi_n(x_n) \\ &+ \sum_{l=1}^{L} \sum_{x_{C_l}} b_{C_l}(x_{C_l}) \log \psi_l(x_{C_l}) \\ &+ \sum_{l=1}^{L} \left\{ \sum_{m \in C_l} \sum_{x_m} \lambda_{C_l}(x_m) \left[b_m(x_m) - \sum_{\sim x_m} b_{C_l}(x_{C_l}) \right] \right\},\end{aligned}$$

subject to non-negativity and normalization constraints.

Stationary Points: We now determine the stationary points of the Lagrangian. The derivatives with respect to the beliefs are given by

$$\frac{\partial \mathcal{L}(b, \lambda)}{\partial b_n(x_n)} = \kappa_n - \log \frac{b_n(x_n)}{\phi_n(x_n)} + \sum_{l \in \mathcal{N}_n} \lambda_{C_l}(x_n) \tag{14}$$

and

$$\frac{\partial \mathcal{L}(b, \lambda)}{\partial b_{C_l}(x_{C_l})} = \tag{15}$$

$$\kappa_{C_l} - \rho_l \log \frac{b_{C_l}(x_{C_l})}{\prod_{m \in C_l} b_m(x_m)} + \log \psi_l(x_{C_l}) - \sum_{m \in C_l} \lambda_{C_l}(x_m),$$

where κ_n and κ_{C_l} are constants. Setting the derivatives to zero and taking exponentials leads to

$$b_n(x_n) \propto \phi_n(x_n) \prod_{l \in \mathcal{N}_i} \exp\left(\lambda_{C_l}(x_n)\right) \tag{16}$$

$$b_{C_l}(x_{C_l}) \propto \psi_l^{1/\rho_l}(x_{C_l}) \prod_{m \in C_l} \frac{b_m(x_m)}{\exp\left(\frac{\lambda_{C_l}(x_m)}{\rho_l}\right)}. \tag{17}$$

We introduce

$$\mu_{\psi_l \to m}(x_m) = \exp\left(\frac{\lambda_{C_l}(x_m)}{\rho_l}\right), \tag{18}$$

so that (16) and (17) can be written as

$$b_n(x_n) \propto \phi_n(x_n) \prod_{l \in \mathcal{N}_n} \mu_{\psi_l \to X_n}^{\rho_l}(x_n) \tag{19}$$

$$b_{C_l}(x_{C_l}) \propto \psi_l^{1/\rho_l}(x_{C_l}) \prod_{m \in C_l} \frac{b_m(x_m)}{\mu_{\psi_l \to X_m}(x_m)}. \tag{20}$$

Observe that (19) is exactly (13). Note that normalization and non-negativity are inherently satisfied.

Message Passing Rules: Finally, we express the consistency constraints to reveal the message passing rules. Essentially, our objective is to write a valid expression that does not contain beliefs. Summing out all variables except x_n in (20) yields

$$\begin{aligned}b_n(x_n) &= \frac{b_n(x_n)}{\mu_{\psi_l \to X_n}(x_n)} \\ &\times \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \prod_{m \in C_l \setminus \{n\}} \frac{b_m(x_m)}{\mu_{\psi_l \to X_m}(x_m)},\end{aligned} \tag{21}$$

so that, after canceling out $b_n(x_n)$ and moving $\mu_{\psi_l \to X_n}(x_n)$ to the left hand side, we find that

$$\begin{aligned}\mu_{\psi_l \to X_n}(x_n) &= \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \\ &\times \prod_{m \in C_l \setminus \{n\}} \frac{\phi_m(x_m) \prod_{k \in \mathcal{N}_m} \mu_{\psi_k \to X_m}^{\rho_k}(x_m)}{\mu_{\psi_l \to X_m}(x_m)}.\end{aligned} \tag{22}$$

Introducing the message from variable vertex X_m to factor vertex ψ_l,

$$\mu_{X_m \to \psi_l}(x_m) = \frac{\phi_m(x_m) \prod_{k \in \mathcal{N}_m} \mu_{\psi_k \to X_m}^{\rho_k}(x_m)}{\mu_{\psi_l \to X_m}(x_m)} \tag{23}$$

we immediately find that

$$\mu_{\psi_l \to X_n}(x_n) = \sum_{\sim x_n} \psi_l^{1/\rho_l}(x_{C_l}) \prod_{m \in C_l \setminus \{n\}} \mu_{X_m \to \psi_l}(x_m). \tag{24}$$

Here, (24) is exactly (12) and (23) is by construction the same as (11).

REFERENCES

[1] J. Pearl, "Probabilistic Reasoning in Intelligent Systems. Networks of plausible inference," *Morgan Kaufmann*, 1988.

[2] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Transactions on Information Theory*, vol. 47, no. 2, pp. 498–519, Feb. 2001.

[3] H. A. Loeliger, "An introduction to factor graphs," *Signal Processing Magazine, IEEE*, vol. 21, no. 1, pp. pp. 28–41, Jan. 2004.

[4] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2282–2312, July 2005.

[5] M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky, "A new class of upper bounds on the log partition function," *IEEE Transactions on Information Theory*, vol. 51, no. 7, pp. 2313–2335, July 2005.

[6] T. Minka, "Divergence measures and message passing," *Microsoft Research Technical Report MSR-TR-2005-173*, Dec. 2005.

[7] V. Savic, H. Wymeersch, F. Penna, and S. Zazo, "Optimized edge appearance probability for cooperative localization based on tree-reweighted nonparametric belief propagation," in *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, May 22–27 2011.

[8] F. Penna, V. Savic, and H. Wymeersch, "Uniformly reweighted belief propagation for distributed bayesian hypothesis testing," in *Proc. IEEE International Workshop on Statistical Signal Processing (SSP)*, June 2011.

[9] J. S. Yedidia, Yige Wang, and S. C. Draper, "Divide and concur and difference-map BP decoders for LDPC codes," *IEEE Transactions on Information Theory*, vol. 57, no. 2, pp. 786 –802, Feb. 2011.

[10] M. J. Wainwright and M. I. Jordan, *Graphical models, exponential families, and variational inference*, Foundations and Trends in Machine Learning. Now Publishers Inc., 2007.