Supplementary Material: Message-Passing for Approximate MAP Inference with Latent Variables

Appendix A: Variational Derivation for Hybrid Message Passing Algorithm

The Marginal-MAP problem can be formulated by

$$\operatorname*{argmax}_{\bar{x}\in\mathcal{X}} \log p(\bar{x}) = \operatorname*{argmax}_{\bar{x}\in\mathcal{X}} A(\theta_{\bar{x}}) \tag{1}$$

where $\mathcal{X}$ is the set of all possible assignments. The corresponding max-marginals satisfy

$$\mu_{\text{max}} \in M_{\bar{x}} = \{\mu_{\bar{x}} | \text{for any } x' \in \mathcal{X}, \mu_{\bar{x}}(x') = 1 \text{ iff } x' = \bar{x}\}$$

So solving Marginal-MAP problem is equivalent to solving the following optimization problem:

$$\max_{\bar{x}\in\mathcal{X}} \sup_{\mu_{\text{other}}\in M(G_{\bar{x}})} \langle\theta,\mu\rangle + H_{Bethe}(\mu) \approx \sup_{\mu_{\max}\in M_{\bar{x}}} \sup_{\mu_{\text{other}}\in M(G_{\bar{x}})} \langle\theta,\mu\rangle + H_{Bethe}(\mu) \tag{2}$$

μ_{other} contains all the marginals except $\mu_{\max}$. By relaxing μ_{sum}s to satisfy only normalization and marginalization conditions, and only relaxing the constraints on $\mu_{\text{sum}\to\text{max}}$ (Here we distinguish between $\mu_{\text{sum}\to\text{max}}$ and $\mu_{\text{max}\to\text{sum}}$ by the direction of messages). Define

$$L_z(G_{\bar{x}}) = \left\{ \mu \geq 0 \;\middle|\; \begin{array}{c} \sum_{z_s} \mu_s(z_s) = 1, \mu_s(x_s) = 1 \text{ iff } x_s = \bar{x}_s, \\ \sum_{z_t} \mu_{st}(v_s, z_t) = \mu_s(v_s), \\ \sum_{z_s} \mu_{st}(z_s, v_t) = \mu_t(v_t), \\ \mu_{st}(x_s, z_t) = \mu_t(z_t) \text{ iff } x_s = \bar{x}_s, \\ \mu_{st}(x_s, x_t) = 1 \text{ iff } x_s = \bar{x}_s, x_t = \bar{x}_t. \end{array} \right\}$$

On $M_{\bar{x}} \times L_z$, we only allow x to take integral solutions and fixed the corresponding assignment for pairwise marginals $\mu_{\text{max}\to\text{sum}}$, but we allow $\mu_{\text{sum}\to\text{max}}$ to assign partial probability to max nodes. In 2, the Bethe entropy terms can be written as (H is the entropy and I is mutual information)

$$H_{Bethe}(\mu) = H_{\mu_{\text{max}}} + H_{\mu_{\text{sum}}} - I_{\mu_{\text{max}}\to\mu_{\text{max}}} - I_{\mu_{\text{sum}}\to\mu_{\text{sum}}} - I_{\mu_{\text{max}}\to\mu_{\text{sum}}} - I_{\mu_{\text{sum}}\to\mu_{\text{max}}}$$

We have

- Entropy of max nodes $H_{\mu_{\text{max}}} = H_s(\mu_s) = 0$, $\forall s \in X$.
- Mutual information between max nodes $I_{\mu_{\text{max}}\to\mu_{\text{max}}} = I_{st}(x_s, x_t) = 0$, $\forall s, t \in X$.
- Mutual information from max node to sum node

$$\begin{aligned} I_{\mu_{\text{max}}\to\mu_{\text{sum}}} &= I_{st}(x_s, z_t) && (3)\\ &= \sum_{(x_s,z_t)\in\mathcal{X}_s\times\mathcal{Z}_t} \mu_{st}(x_s, z_t) \log \frac{\mu_{st}(x_s, z_t)}{\mu_s(x_s)\mu_t(z_t)} && (4)\\ &= \sum_{z_t\in\mathcal{Z}_t} \mu_{st}(x^*, z_t) \log \frac{\mu_{st}(x^*, z_t)}{\mu_s(x^*)\mu_t(z_t)} && (5)\\ &= 0, \qquad \forall s \in X, t \in Z && (6) \end{aligned}$$

where x^* is the assigned state of x at node s.

- Entropy for sum nodes, Mutual information between sum nodes, and from sum node to max node is nonzero as we allow sum nodes assign probabilities to max nodes.

Now, we do the LP-relaxation on max nodes and the relaxed optimization problem on new domain $L(G)$ is

$$\sup_{\mu \in L(G)} \langle \mu, \theta \rangle + H(\mu_{\text{sum}}) - I(\mu_{\text{sum}\to\text{sum}}) - I(\mu_{\text{sum}\to\text{max}}) \tag{7}$$

where

$$L(G) = \left\{ \mu \geq 0 \;\middle|\; \begin{array}{c} \sum_{v_s} \mu_s(v_s) = 1, \\ \sum_{v_t} \mu_{st}(v_s, v_t) = \mu_s(v_s), \\ \sum_{v_s} \mu_{st}(v_s, v_t) = \mu_t(v_t). \end{array} \right\}$$

Let $s, t \in V$, λ_{ss} be a Lagrangian multiplier associated with the normalization constraint $C_{ss}(\mu) = 0$, where

$$C_{ss}(\mu) = 1 - \sum_{v_s} \mu_s(v_s)$$

.

Similarly, we define the constraints for each direction $t \to s$ for every possible edge.

$$C_{ts}(v_s; \mu) = \mu_s(v_s) - \sum_{v_t} \mu_{st}(v_s, v_t)$$

and its Lagrange multiplier $\lambda_{ts}(v_s)$. Then the Lagrangian for the Bethe variational problem (BVP) is

$$\begin{aligned} L(\mu, \lambda; \theta) &= \sup_{\mu \in L(G)} \langle \mu, \theta \rangle + H(\mu_{\text{sum}}) - I(\mu_{\text{sum}\to\text{sum}}) - I(\mu_{\text{sum}\to\text{max}}) \\ &\quad + \sum_{s \in V} \lambda_{ss} C_{ss}(\mu) + \sum_{(s,t) \in E} \left[\sum_{v_s} \lambda_{ts}(v_s) C_{ts}(v_s; \mu) + \sum_{y_t} \lambda_{st}(v_t) C_{st}(v_t; \mu) \right] \end{aligned} \tag{8}$$

Now we show that the partial derivative with respect to different μs are identical to the derivatives in the standard sum/max product formulations (depending on the node type).

Taking the derivative with respect to the node marginals of:

Sum Nodes

For sum node s, denote its (pseudo-)marginal as μ_s.

$$\begin{aligned} \nabla_{\mu_s} L &= \nabla_{\mu_s} \{ \langle \theta_{\text{sum}}, \mu_{\text{sum}} \rangle + H(\mu_{\text{sum}}) + \sum_{s \in V} \lambda_{ss} C_{ss}(\mu) + \sum_{(s,t) \in E} [\sum_{v_s} \lambda_{ts}(v_s) C_{ts}(v_s; \mu) + \sum_{v_t} \lambda_{st}(v_t) C_{st}(v_t; \mu)] \} \\ &= \theta_s(z_s) + \lambda_{ss} + \sum_{t \in N(s)} \lambda_{ts}(z_s) - \log \mu_s(z_s) \end{aligned} \tag{9}$$

In [1] (page 84, eq. (4.19)), using our notation, the Lagrangian for the BVP to marginal problem is

$$L_{sum}(\mu, \lambda; \theta) = \langle \theta, \mu \rangle + H(\mu) + \sum_{s \in Z} \lambda_{ss} C_{ss}(\mu) + \sum_{(s,t) \in E} \left[\sum_{z_s} \lambda_{ts}(z_s) C_{ts}(z_s; \mu) + \sum_{z_t} \lambda_{st}(z_t) C_{st}(z_t; \mu) \right] \tag{10}$$

$$\nabla_{\mu_s} L_{sum} = \nabla_{\mu_s} L$$

So, for sum nodes, the derivative of node marginals is the same as those for the BVP to marginal problem. Similar arguments hold for the pairwise marginals between sum nodes, and sum$\to$max node pair. Thus we have the identical equations to Eq.(4.23), Eq(4.24)[1]:

$$\mu_s(z_s) = \kappa \exp(\theta_s(z_s)) \prod_{u \in N(s)} M_{st}(z_t) \tag{11}$$

$$\mu_{st}(z_s, z_t) = \kappa' \exp(\theta_{st}(z_s, z_t) + \theta_s(z_s) + \theta_t(z_t)) \prod_{u \in N(s) \backslash t} M_{us}(z_s) \prod_{u \in N(t) \backslash s} M_{ut}(z_t) \tag{12}$$

For any type of node t in the graph, its marginal can be represented as

$$\mu_t(v_t) = \kappa \exp\{\theta_t(v_t)\} \prod_{s \in N(t)} M_{st}(v_t)$$

(We will show this for max node in the following subsection.) Similarly, taking the derivative w.r.t. μ_{st}, $s \in Z$ and $t \in X$ and combine with Eq (12),

$$\mu_{st}(z_s, v_t) = \kappa' \exp(\theta_{st}(z_s, v_t) + \theta_s(z_s) + \theta_t(v_t)) \prod_{u \in N(s)\backslash t} M_{us}(z_s) \prod_{u \in N(t)\backslash s} M_{ut}(v_t) \quad (13)$$

By applying the marginalization constraints $\sum_{z_s} \mu_{st}(z_s, v_t) = \mu_t(v_t)$, we get the message from sum node t to any node s:

$$M_{ts}(v_s) \leftarrow \kappa_1 \sum_{z'_t \in \mathcal{Z}_t} \left\{ \exp[\theta_{st}(v_s, z'_t) + \theta_t(z'_t)] \prod_{u \in N(t)\backslash s} M_{ut}(z'_t) \right\}$$

Max Nodes

For the derivative w.r.t. node marginals of max nodes μ_s, $s \in X$.

$$\nabla_{\mu_s} L = \theta_s(x_s) + \lambda_{ss} + \sum_{t \in N(s)} \lambda_{ts}(x_s) = \nabla_{\mu_s} L_{max}$$

where L_{max} is the Lagrangian of pure MAP problem. Similarly, we also have

$$\nabla_{\mu_{\text{max}\to\text{max}}} L = \nabla_{\mu_{\text{max}\to\text{max}}} L_{\text{max}} = \nabla_{\mu_{\text{max}\to\text{sum}}} L$$

It is identical to follow [1] (Section 8.2) to check that max marginal:

$$\mu_s(x_s) = \kappa \exp(\theta_s(x_s)) \prod_{u \in N(s)} M_{st}(x_t)$$

and the fixed point of max messages:

$$M_{ts}(v_s) \leftarrow \kappa_2 \max_{x'_t \in \mathcal{X}_t} \left\{ \exp[\theta_{st}(v_s, x'_t) + \theta_t(x'_t)] \prod_{u \in N(t)\backslash s} M_{ut}(x'_t) \right\}$$

provide a solution to the (partial) problem $(\mu_{\text{max}}, \mu_{\text{max}\to\mu_{\text{max}}}, \mu_{\text{max}\to\mu_{\text{sum}}})$. In conclusion, the hybrid message passing gives an approximation to the Marginal-MAP problem.

Appendix B: EM via Message Passing

B.1 EM Objective

Note here that standard EM does $\max_\theta \sum_z p_\theta(x|z)$ for fixed x, but we want $\max_x \sum_z p_\theta(x|z)$ for fixed θ. The derivation is nearly identical and we do not write down the dependency on θ for convenience. Define $F(\tilde{p}, x) = \mathbb{E}_{\tilde{p}}[\log p(x, z)] + H(\tilde{p}(z))$ based on the following routine application of Jensen's inequality:

$$\begin{aligned}
\log p(x) &= \log \sum_z p(x, z) \\
&= \log \sum_z \tilde{p}(z) \frac{1}{\tilde{p}(z)} p(x, z) \\
&\geq \sum_z \tilde{p}(z) \log \left[\frac{1}{\tilde{p}(z)} p(x, z) \right] \\
&= \sum_z \tilde{p}(z) \log p(x, z) - \sum_z \tilde{p}(z) \log p(z) \qquad (14)
\end{aligned}$$

So $\log p(x) \geq \mathbb{E}_{\tilde{p}(z)} \log p(x, z) - \mathbb{E}_{\tilde{p}(z)} \log \tilde{p}(z) = \mathbb{E}_{\tilde{p}}[\log p(x, z)] + H(\tilde{p}(z))$

B.2 Proof[1] of Proposition 1

Proposition 1. *With the value of x fixed in function F, the unique solution to maximizing $F(\tilde{p}, x)$ is given by $\tilde{p}(z) = p(z|x)$.*

Proof. Since $\tilde{p}(z)$ is a distribution over z, $\sum_z \tilde{p}(z) = 1$. This is a constrained optimization problem, so the Lagrangian is

$$L(\tilde{p}, x) = F(\tilde{p}, x) - \lambda \left(\sum_z \tilde{p}(z) - 1 \right)$$

λ is the Lagrange multiplier. So at the maximum $\tilde{p}(z)$, the derivative of L with respect to the components of $\tilde{p}$ should be zero. Then we have where

$$\lambda = \log p(x, z) - \log \tilde{p}(z) - 1$$

This indicates $\tilde{p}(z) \propto p(x, z)$ and given the constraint that $\sum_z \tilde{p}(z) = 1$, the unique solution to this optimization problem is $\tilde{p}(z) = p(z|x)$. (Note x is fixed here.) □

B.3 Proof of Proposition 2

Proposition 2. *If $\tilde{p}(z) = p(z|x)$, then $F(\tilde{p}, x) = \log p(x) = \log \sum_z p(x, z)$.*

Proof.

$$\begin{aligned} F(\tilde{p}, x) &= \mathbb{E}_{\tilde{p}(z)} \left[\log p(x, z)\right] + H(\tilde{p}) \\ &= \mathbb{E}_{\tilde{p}(z)} \left[\log p(x, z)\right] - \mathbb{E}_{\tilde{p}(z)}[\log p(z|x)] \\ &= \mathbb{E}_{\tilde{p}(z)}[\log p(x, z) - \log p(z|x)] \\ &= \mathbb{E}_{\tilde{p}(z)}[\log p(x)] \\ &= \log p(x) \end{aligned} \tag{15}$$

□

B.4 Derivation of EM via Message Passing

E-step: Estimate $\tilde{p}(z) = p(z|x)$ given x.

M-step: Consider the conditional,

$$p_\theta(x \mid z) = \frac{\exp\left[\langle \theta, \phi(x, z) \rangle - A(\theta)\right]}{\sum_{x'} \exp\left[\langle \theta, \phi(x', z) \rangle - A(\theta)\right]} = \exp\left[\langle \theta, \phi(x, z) \rangle - B_z(\theta)\right]$$

where $B_z(\theta) = \log \sum_x \exp\left[\langle \theta, \phi(x, z) \rangle\right]$.

Fixing x in $p(z|x)$ to be $\bar{x}$, the assignment given by the previous M-step,

$$\begin{aligned} \max_x \mathbb{E}_{z \sim p_\theta(z \mid \bar{x})} \log p_\theta(x, z) &= \max_x \mathbb{E}_{z \sim p_\theta(z \mid \bar{x})} \log p_\theta(x \mid z) \\ &= \max_x \sum_z p(z \mid \bar{x}) [\langle \theta, \phi(x, z) \rangle - B_z(\theta)] \\ &= \max_x \sum_z p(z \mid \bar{x}) \langle \theta, \phi(x, z) \rangle \end{aligned} \tag{16}$$

We use the shorthand notation of overcomplete representation of sufficient statistics in the following table.

[1] The proofs in B.2 and B.3 are almost identical to Lemma 1 and 2 in [2] page 4-5.

SYMBOL	EXPRESSION
Θ_x	$\sum_{s\in X}\sum_i \theta_{s;i}\mathbb{I}_{s;i}(x_s)$
Θ_z	$\sum_{s\in Z}\sum_i \theta_{s;i}\mathbb{I}_{s;i}(z_s)$
Θ_{xx}	$\sum_{(s,t)\in E, s,t\in X}\sum_{(i,j)} \theta_{st;ij}\mathbb{I}_{st;ij}(x_s, x_t)$
Θ_{zz}	$\sum_{(s,t)\in E, s,t\in Z}\sum_{(i,j)} \theta_{st;ij}\mathbb{I}_{st;ij}(z_s, z_t)$
Θ_{xz}	$\sum_{(s,t)\in E, s\in X, t\in Z}\sum_{(i,j)} \theta_{st;ij}\mathbb{I}_{st;ij}(x_s, z_t)$

Then

$$\begin{aligned}
\sum_z p(z \mid \bar{x})\langle\theta, \phi(x,z)\rangle &= \sum_z p(z \mid \bar{x})\left[\Theta_x + \Theta_z + \Theta_{xx} + \Theta_{zz} + \Theta_{xz}\right] \\
&= \Theta_x + \Theta_{xx} + \sum_z p(z \mid \bar{x})\Theta_{xz} + C \\
&\approx \Theta_x + \Theta_{xx} + \sum_{(s,t)\in E, s\in X, t\in Z}\sum_{(i,j)} \theta_{st;ij}\mathbb{I}_{s;i}(x_s)\mu_{t;j} + C \\
&= \sum_{s\in X, i}\left[\theta_{s;i} + \sum_{t\in Z, j}\mu_{t;j}\theta_{st;ij}\right]\mathbb{I}_{s;i}(x_s) \\
&+ \sum_{(s,t)\in E, s,t\in X}\sum_{(i,j)} \theta_{st;ij}\mathbb{I}_{st;ij}(x_s, x_t) + C
\end{aligned} \tag{17}$$

where C subsumes the terms irrelevant to the maximization over x. μ_t is the pseudo-marginal of node t given $\bar{x}$, so we get an approximation instead of an equality, and we use the fact that $\sum_z p(z \mid \bar{x}) = 1$. Then, the M-step amounts to running the max product algorithm with potentials on x nodes modified according to Eq (17).

Appendix C: Hybrid Tree-Reweighted Message Passing Algorithm

On loopy graphs, we can also apply the hybrid scheme and get a hybrid tree-reweighted message passing algorithm (which we use in the experiments on the loopy graphs and the protein data). It is sketched in Algorithm 1.

Appendix D: Related Work and Discussion

On a high-level, the Marginal-MAP problem can be seen as doing a search over the space of all possible assignments over the max nodes of the graph, having done a variable elimination over the sum nodes. Several heuristics exists to perform this search step. Typical approaches under this category include methods such as branch-and-bound and beam-search [3, 4]. However these methods are designed to give only the MAP estimates for the max nodes whereas our hybrid message-passing algorithm provides both the MAP estimates for the max nodes and marginals for the sum nodes.

In principle, we note however that the Expectation-Maximization algorithm [5] can also be used if the *marginal* posterior in the E-step can be computed in closed-form. If not, then Monte-Carlo simulations can be used to estimate the expectation. Alternatively, [6, 7] proposed an MCMC-based algorithm for *direct* maximization of marginal posterior distributions by introducing an artificially augmented probability model, whose sampling gives marginal-MAP estimates of the variables of interest. Recently, [8] proposed a Sequential Monte Carlo based approach (similar to simulated annealing) which is much less sensitive to initialization than EM/MCMC algorithms.

In our work, we take a different approach and show how message-passing algorithms for graphical models can be used to obtain marginal-MAP estimates in a variational framework [1]. A lot of work has gone into improving the standard sum-product and max-product algorithms [9, 10], and there is no apparent reason why such advances cannot also improve our hybrid message passing algorithm.

Algorithm 1 Hybrid Tree-Reweighted Message-Passing Algorithm

Inputs: Graph $G = (V, E)$, $V = X \cup Z$, potentials θ_v, $s \in V$ and θ_{st}, $(s,t) \in E$.

1. Initialize the messages to some arbitrary value.
2. (Greedily) Find a set of spanning trees $\mathcal{T}$ that covers G, and compute the edge appearance probability w_{st}, for $(s,t) \in E$.
3. For each node s in G, do the following until messages converge (or maximum number of iterations reached)
   - If $s \in X$, update messages by
   $$M_{ts}(v_s) \leftarrow \kappa \max_{x_t' \in \mathcal{X}_t} \left\{ \exp[\frac{1}{w_{st}}\theta_{st}(v_s, x_t') + \theta_t(x_t')] \frac{\prod_{u \in N(t) \setminus s} M_{ut}^{w_{ut}}(x_t')}{M_{st}^{1-w_{st}}(x_t')} \right\} \quad (18)$$
   - If $s \in Z$, update messages by
   $$M_{ts}(v_s) \leftarrow \kappa \sum_{z_t' \in \mathcal{X}_t} \left\{ \exp[\frac{1}{w_{st}}\theta_{st}(v_s, z_t') + \theta_t(z_t')] \frac{\prod_{u \in N(t) \setminus s} M_{ut}^{w_{ut}}(z_t')}{M_{st}^{1-w_{st}}(z_t')} \right\} \quad (19)$$
4. Compute the local belief for each node s. $\mu_s(v_s) = \kappa \exp\{\theta_s(v_s)\} \prod_{t \in N(s)} M_{ts}^{w_{ts}}(v_s)$
5. For all $x_s \in X$, return $\arg\max_{x_s \in \mathcal{X}_s} \mu_s(x_s)$
6. For all $z_s \in Z$, return $\mu_s(z_s)$.

Before concluding, we discuss another recent work by Liu and Ihler [11] which is most similar in spirit to our work, and is a simultaneous development. Liu and Ihler [11] proposed a method based optimizing a variational objective on specific graph structures. Here we would highlight the similarities and the differences between their work and ours.

In particular, we define the log-partition function for the marginal-MAP problem by fixing the $\boldsymbol{x}$ assignments to $\bar{\boldsymbol{x}}$ and constructing a new graph $G_{\bar{\boldsymbol{x}}}$. Liu and Ihler [11], on the other hand, use the conditional entropy (conditioned on the max nodes) to define the free energy term in the log partition function. Both ways of defining the log partition function are equivalent because fixing the $\boldsymbol{x}$ assignments amounts to conditioning.

In our approach, we then define the marginal polytope w.r.t. the MAP $\boldsymbol{x}$ assignments we are seeking, and propose relaxations that give us our final variational objective (11). We then derive a hybrid message-passing algorithm for this variational objective (details in appendix A), in a way similar to how standard sum and max product algorithms are derived (Wainright and Jordan, 2008).

In contrast, Liu and Ihler [11] propose relaxations of their variational objective to solve the marginal MAP probel but their relaxations also require specific constraints on the structure of the graphs to ensure local or global optima (e.g., sum nodes forming a tree).

Liu and Ihler [11] also proposed a hybrid algorithm similar to ours except for one difference: in our case, a max-product message is sent from a max node to sum node; in their case this message is defined by a set of ”mixed-marginals”, which requires solving an extra local MAP problem.

Another important difference is the way the connection to EM is shown. Liu and Ihler [11] do it from variational principles and derive an EM algorithm which solves their original variational objective in an alternating fashion for sum and max nodes. In contrast, we do so by deriving a “message passing variant” of the EM algorithm to solve the marginal-MAP problem, and then showing how the messages in the E and M steps of this algorithm are akin to the messages passed in our hybrid message passing algorithm.

Establishing the guarantee about convergence [12] of the hybrid message-passing algorithm we describe in this paper is another open issue, and we hope that related theoretic developments (e.g., [13, 14]) will be able to shed more light on this.

References

[1] M. J. Wainwright and M. I. Jordan. Graphical Models, Exponential Families, and Variational Inference. *Foundations and Trends in Machine Learning*, 2008.

[2] Radford M. Neal and Geoffrey E. Hinton. A View of the EM Algorithm that Justifies Incremental, Sparse, and Other Variants. In *Learning in graphical models*, pages 355–368, 1999.

[3] James D. Park. MAP Complexity Results and Approximation Methods. In *UAI*, 2002.

[4] D. Koller and N. Friedman. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press, 2009.

[5] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum Likelihood from Incomplete Data via the EM algorithm. *Journal of The Royal Statistica Society*, 1977.

[6] A. Doucet, S. J. Godsill, and C. P. Robert. Marginal Maximum a Posteriori Estimation using Markov chain Monte Carlo. *Statistics and Computing*, 2002.

[7] Changhe Yuan, Tsai-Ching Lu, and Marek Druzdzel. Annealed map. In *UAI*, 2004.

[8] A. M. Johansen, A. Doucet, and M. Davy. Particle Methods for Maximum Likelihood Estimation in Latent Variable Models. *Statistics and Computing*, 2008.

[9] David Sontag, Talya Meltzer, Amir Globerson, Tommi Jaakkola, and Yair Weiss. Tightening LP Relaxations for MAP using Message Passing. In *UAI*, 2008.

[10] Talya Meltzer, Amir Globerson, and Yair Weiss. Convergent message passing algorithms - a unifying view. In *UAI*, 2009.

[11] Qiang Liu and Alexander Ihler. Variational algorithms for marginal map. In *UAI*, 2011.

[12] Judea Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1988.

[13] Tanya Roosta, Martin J. Wainwright, and Shankar Sastry. Convergence analysis of reweighted sum-product algorithms. In *IEEE Transactions on Signal Processing*, 2008.

[14] Alexander T. Ihler, John W. Fischer III, and Alan S. Willsky. Loopy belief propagation: Convergence and effects of message errors. *Journal of Machine Learning Research*, 2005.