Convergent and Correct Message Passing Algorithms

Nicholas Ruozzi and Sekhar Tatikonda
Yale University
Overview

- A simple derivation of a new family of message passing algorithms

- Conditions under which the parameters guarantee correctness upon convergence

- An asynchronous algorithm that generalizes the notion of “bound minimizing” algorithms [Meltzer, Globerson, Weiss 2009]

- A simple choice of parameters that guarantees both convergence and correctness of the asynchronous algorithm
Previous Work

- Convergent message passing algorithms:
  - Serial TRMP [Kolmogorov 2006]
  - MPLP [Globerson and Jaakkola 2007]
  - Max-sum diffusion [Werner 2007]*
  - Norm-product BP [Hazan and Shashua 2008]

- Upon convergence, any assignment that simultaneously minimizes all of the beliefs must minimize the objective function
Outline

- Review of the min-sum algorithm
- The splitting algorithm
  - Correctness
  - Convergence
- Conclusion
Min-Sum

\[ f(x_1, \ldots, x_n) = \sum_i \phi_i(x_i) + \sum_\alpha \psi_\alpha(x_\alpha) \]

- Minimize an objective function that factorizes as a sum of potentials
- \( \alpha \in \mathcal{A} \) (some multiset whose elements are subsets of the variables)
Corresponding Graph

\[
f(x_1, x_2, x_3) = \phi_1 + \phi_2 + \phi_3 + \psi_{23} + \psi_{12}
\]
Min-Sum

- Messages at time $t$:

$$m^t_{i \rightarrow \alpha}(x_i) = \kappa + \phi_i(x_i) + \sum_{\beta \in \partial i \setminus \alpha} m^{t-1}_{\beta \rightarrow i}(x_i)$$

$$m^t_{\alpha \rightarrow i}(x_i) = \kappa + \min_{x_{\alpha \setminus i}} \left[ \psi_{\alpha}(x_{\alpha}) + \sum_{k \in \alpha \setminus i} m^{t-1}_{k \rightarrow \alpha}(x_k) \right]$$

- Normalization factor is arbitrary for each message

- Initial messages must be finite
Computing Beliefs

- At each step, construct a set of beliefs:

\[
\begin{align*}
    b^t_i(x_i) &= \kappa + \phi_i(x_i) + \sum_{\alpha \in \partial i} m^t_{\alpha \rightarrow i}(x_i) \\
    b^t_\alpha(x_\alpha) &= \kappa + \psi_\alpha(x_\alpha) + \sum_{i \in \alpha} m^t_{i \rightarrow \alpha}(x_i)
\end{align*}
\]

- Estimate the optimal assignment as

\[x^t_i = \arg\min_{x_i} b^t_i(x_i)\]

- Converges if we obtain a fixed point of the message updates
Outline

- Review of the min-sum algorithm

- The splitting algorithm
  - Correctness
  - Convergence

- Conclusion
“Splitting” Heuristic

Original Graph

Split the (1,2) factor
Reduce the potentials on each edge by 1/2

\[ \psi'_{12} = \frac{\psi_{12}}{2} \]

\[ \psi''_{12} = \frac{\psi_{12}}{2} \]
Consider the message passed across one of the split edges:

\[ m_{1 \rightarrow (1,2)'}(x_1) = \phi_1(x_1) + m_{(3,1) \rightarrow 1}(x_1) + m_{(1,2)'' \rightarrow 1}(x_1) \]

If the initial messages are the same for the copied edges, then, at each time step, the messages passed across the duplicated edges are the same:

\[ m^t_{1 \rightarrow (1,2)'}(x_1) = m^t_{1 \rightarrow (1,2)''}(x_1) \]
Min-Sum with Splitting

- Consider the message passed across one of the split edges:

  \[ m_{1 \rightarrow (1,2)}(x_1) = \phi_1(x_1) + m_{(3,1) \rightarrow 1}(x_1) + m_{(1,2) \rightarrow 1}(x_1) \]

- If the initial messages are the same for the copied edges, then, at each time step, the messages passed across the duplicated edges are the same:

  \[ m_{1 \rightarrow (1,2)}^t(x_1) = m_{1 \rightarrow (1,2)}^{t'}(x_1) \]
Min-Sum with Splitting

- We can split any of the factors $c_\alpha$ times and divide the potentials by a factor of $c_\alpha$

$$m_{i \rightarrow \alpha}^t(x_i) = \kappa + \phi_i(x_i) + (c_\alpha - 1)m_{\alpha \rightarrow i}^{t-1}(x_i) + \sum_{\beta \in \partial i \setminus \alpha} c_\beta m_{\beta \rightarrow i}^{t-1}(x_i)$$

$$m_{\alpha \rightarrow i}^t(x_i) = \kappa + \min_{x_\alpha \setminus i} \left[ \frac{\psi_\alpha(x_\alpha)}{c_\alpha} + \sum_{k \in \alpha \setminus i} m_{k \rightarrow \alpha}^{t-1}(x_k) \right]$$

- Messages are passed on the original factor graph

- Update equation is similar to TRMP
General Splitting Algorithm

- Splitting the variable nodes and the factor nodes:

\[ m_{i \rightarrow \alpha}^t(x_i) = \kappa + \frac{\phi_i(x_i)}{c_i} + (c_{\alpha} - 1)m_{\alpha \rightarrow i}^{t-1}(x_i) + \sum_{\beta \in \partial i \setminus \alpha} c_{\beta}m_{\beta \rightarrow i}^{t-1}(x_i) \]

\[ m_{\alpha \rightarrow i}^t(x_i) = \kappa + \min_{x_{\alpha \setminus i}} \left[ \psi_{\alpha}(x_{\alpha}) \frac{c_{\alpha}}{c_{\alpha}} + (c_{i} - 1)m_{i \rightarrow \alpha}^{t-1}(x_i) + \sum_{k \in \alpha \setminus i} c_{k}m_{k \rightarrow \alpha}^{t-1}(x_k) \right] \]

- Recover min-sum if all parameters are chosen to be equal to one
Beliefs

- **Corresponding beliefs:**

  \[
  b^t_i(x_i) = \kappa + \frac{\phi_i(x_i)}{c_i} + \sum_{\alpha \in \partial i} c_\alpha m^t_{\alpha \rightarrow i}(x_i)
  \]

  \[
  b^t_\alpha(x_\alpha) = \kappa + \frac{\psi_\alpha(x_\alpha)}{c_\alpha} + \sum_{k \in \alpha} c_k \left[ \frac{\phi_k(x_k)}{c_k} + (c_\alpha - 1)m^t_{\alpha \rightarrow k}(x_k) \right] + \sum_{\beta \in \partial k \setminus \alpha} c_\beta m^t_{\beta \rightarrow k}(x_k)
  \]

- **Not the same as before**
Outline

- Review of the min-sum algorithm
- The splitting algorithm
  - Correctness
  - Convergence
- Conclusion
Admissibility and Min-consistency

- A vector of beliefs, $b$, is **admissible** for a function $f$ if

  \[ f(x) = \kappa + \sum_i c_i b_i(x_i) + \sum_\alpha c_\alpha \left[ b_\alpha(x_\alpha) - \sum_{k \in \alpha} c_k b_k(x_k) \right] \]

  Makes sense for any non-zero, real-valued parameters

- A vector of beliefs, $b$, is **min-consistent** if for all $\alpha$ and all $i \in \alpha$:

  \[ \min_{x_\alpha \setminus i} b_\alpha(x_\alpha) = \kappa + b_i(x_i) \]
Conical Combinations

- If there is a unique $x^*$ that simultaneously minimizes each belief, does it minimize $f$?

$$f(x) = \kappa + \sum_i c_i b_i(x_i) + \sum_{\alpha} c_\alpha \left[ b_\alpha(x_\alpha) - \sum_{k \in \alpha} c_k b_k(x_k) \right]$$

- $f$ can be written as a **conical combination** of the beliefs if there exists a vector, $d$, of non-negative reals such that

$$f(x) = \sum_{i, \alpha: i \in \alpha} d_{i\alpha} (b_\alpha(x_\alpha) - b_i(x_i)) + \sum_\alpha d_{\alpha \alpha} b_\alpha(x_\alpha) + \sum_i d_{ii} b_i(x_i)$$
Global Optimality

**Theorem:** Given an objective function, $f$, suppose that

- $b$ is a vector of admissible and min-consistent beliefs for the function $f$
- $c$ is chosen such that $f$ can be written as a conical combination of the beliefs
- There exists an $x^*$ that simultaneously minimizes all of the beliefs

then $x^*$ minimizes the objective function
Global Optimality

- Examples of globally optimal parameters:
  - **TRMP:** $c_i = 1$ and $c_\alpha$ is the “edge appearance probability”
  - **MPLP:** $c_\alpha = 1$ and $c_i = 1/2$
  - **Others:** $c_\alpha > 0$ and $c_i(1 - \sum_{\alpha \in \partial i} c_\alpha) \geq 0$
  - Weaker conditions on $c$ guarantee local optimality (wrt the Hamming distance)
Outline

- Review of the min-sum algorithm
- The splitting algorithm
  - Correctness
  - Convergence
- Conclusion
Notion of Convergence

- Let $c$ be chosen as: $c_i = 1$

$$c_\alpha = 1 / \max_i |\partial i|$$

- Consider the lower bound:

$$\min_x f(x) \geq \sum_i c_i (1 - \sum_{\alpha \in \partial i} c_\alpha \min_{x_i} b_i(x_i)) + \sum_\alpha c_\alpha \min_{x_\alpha} b_\alpha(x_\alpha)$$

$$\equiv LB(m)$$

- We will say that the algorithm has converged if this lower bound cannot be improved by further iteration.
Convergence

- For certain c, there is an asynchronous message passing schedule that is guaranteed not to decrease this bound.

- All “bound minimizing” algorithms have similar guarantees [Meltzer, Globerson, Weiss 2009]
  - Ensure min-consistency over one subtree of the graph at a time.
Asynchronous Algorithm

- Update the blue edges then the red edges
- After this update beliefs are min-consistent with respect to the variable $x_1$
Asynchronous Algorithm

- Update the blue edges then the red edges
- After this update beliefs are min-consistent with respect to the variable $x_1$
Asynchronous Convergence

- Order the variables and perform the one step update for a different variable at each step of the algorithm
  - This cannot decrease the lower bound
  - This update can be thought of as a coordinate ascent algorithm in an appropriate dual space
An Example: MWIS

- 100x100 grid graph
- Weights randomly chosen between 0 and 1
- Convergence is monotone as predicted
Conclusion

- A simple derivation of a new family of message passing algorithms
- Conditions under which the parameters guarantee correctness upon convergence
- An asynchronous algorithm that generalizes the notion of “bound minimizing” algorithms [Meltzer, Globerson, Weiss 2009]
- A simple choice of parameters that guarantees both convergence and correctness of the asynchronous algorithm
Questions?

Preprint available online at: http://arxiv.org/abs/1002.3239
Other Results

- We can extend the computation tree story to this new algorithm.

- Can use graph covers to understand when the algorithm can converge to a collection of beliefs that admit a unique estimate.

- The lower bound is always tight for pairwise binary graphical models.
  - New proof that uses graph covers instead of duality.