Message Passing for Estimation and Decision in Graphical Models

Qiang Liu, Alexander Ihler

UC Irvine
Graphical models

\[ p(x) = p(x_1, x_2, \cdots, x_N) \]

\[ = \frac{1}{Z} \prod_{\alpha \in \mathcal{I}} \psi_\alpha(x_\alpha) \quad \text{(factorization form)} \]
Graphical models

\[ p(x) = p(x_1, x_2, \ldots, x_N) \]

\[ = \frac{1}{Z} \prod_{\alpha \in \mathcal{I}} \psi_\alpha(x_\alpha) \quad \text{(factorization form)} \]

\[ = \frac{1}{Z} \exp \left[ \sum_{\alpha \in \mathcal{I}} \theta_\alpha(x_\alpha) \right] \quad \text{(exponential form, } \theta_\alpha = \log \psi_\alpha \text{)} \]
Graphical models

\[ p(x) = p(x_1, x_2, \cdots, x_N) \]

\[ = \frac{1}{Z} \prod_{\alpha \in I} \psi_{\alpha}(x_{\alpha}) \quad \text{(factorization form)} \]

\[ = \frac{1}{Z} \exp \left[ \sum_{\alpha \in I} \theta_{\alpha}(x_{\alpha}) \right] \quad \text{(exponential form, } \theta_{\alpha} = \log \psi_{\alpha} \text{)} \]

Bayesian Network

\[ p(x) = \prod_{i \in V} p(x_i | x_{\text{pa}_i}) \]
Graphical models

\[ p(\mathbf{x}) = p(x_1, x_2, \ldots, x_N) \]

\[ = \frac{1}{Z} \prod_{\alpha \in \mathcal{I}} \psi_\alpha(x_\alpha) \]  
(factorization form)

\[ = \frac{1}{Z} \exp \left[ \sum_{\alpha \in \mathcal{I}} \theta_\alpha(x_\alpha) \right] \]  
(exponential form, \( \theta_\alpha = \log \psi_\alpha \))

---

Bayesian Network

\[ p(\mathbf{x}) = \prod_{i \in V} p(x_i | x_{\text{pa}_i}) \]

Markov Random Field

\[ p(\mathbf{x}) = \frac{1}{Z} \prod_{i \in V} \psi_i(x_i) \prod_{(ij) \in E} \psi_{ij}(x_i, x_j) \]  
(pairwise models)
"Inference"

- Inference = answering queries about the graphical model

Models

Queries

\[
p(\mathbf{x}) \propto \prod_{\alpha \in \mathcal{I}} \psi(\mathbf{x}_\alpha)
\]
Queries of Graphical Models

- **Max-Inference**

- **Sum-Inference**

- **Mixed-Inference** (Marginal MAP)

- **Decision Making**

called maximum a posteriori estimator (MAP) or most probable explanation (MPE)

\[ \mathbf{x}^* = \arg \max_{\mathbf{x}} p(\mathbf{x}) = \arg \max_{\mathbf{x}} \prod_{\alpha} \psi(\mathbf{x}_\alpha) \]
Calculating the marginal probabilities or the partition function:

\[ Z = \sum_{x} \prod_{\alpha} \psi(x_{\alpha}) \quad \text{or} \quad b_i(x_i) = \frac{1}{Z} \sum_{x/x_i} \prod_{\alpha} \psi(x_{\alpha}) \]
Find the a partial configuration that maximizes its marginal distribution (marginal MAP or simply called MAP)

\[ x^*_B = \arg \max_{x_B} p(x_B) = \arg \max_{x_B} \sum_{x_A} \prod_{\alpha} \psi(x_\alpha) \]

where \( A \cup B = V \)

Queries of Graphical Models

- Max-Inference
- Sum-Inference
- Mixed-Inference (Marginal MAP)
- Decision Making
Queries of Graphical Models

- Max-Inference
- Sum-Inference
- Mixed-Inference (Marginal MAP)
- Decision Making

- Find the a partial configuration that maximizes its marginal distribution (marginal MAP or simply called MAP)

\[
\mathbf{x}_B^* = \arg \max_{\mathbf{x}_B} \mathcal{p}(\mathbf{x}_B) = \arg \max_{\mathbf{x}_B} \sum_{\mathbf{x}_A} \prod_{\alpha} \psi(\mathbf{x}_\alpha)
\]

where \( A \cup B = V \)

Why marginal MAP?

- Interested in partial configuration

- Hidden Variable Models
  - Deep belief nets, latent tree models, hidden-state CRF …

- Bayesian approaches
  - Maximum-Likelihood v.s. Bayesian inference
Queries of Graphical Models

- Max-Inference
- Sum-Inference
- Mixed-Inference (Marginal MAP)
- Decision Making

Find the partial configuration that maximizes its marginal distribution (marginal MAP or simply called MAP)

\[ x^*_B = \arg \max_{x_B} p(x_B) = \arg \max_{x_B} \sum_{x_A} \prod_{\alpha} \psi(x_\alpha) \]

where \( A \cup B = V \)

**Marginal MAP Challenging!**

- NP-hard even on tree structured graphs

Example from D. Koller and N. Friedman (2009)
Queries of Graphical Models

- Max-Inference
- Sum-Inference
- Mixed-Inference (Marginal MAP)
- Decision Making

Find the a partial configuration that maximizes its marginal distribution (marginal MAP or simply called MAP)

\[ x^*_B = \arg \max_{x_B} p(x_B) = \arg \max_{x_B} \sum_{x_A} \prod_{\alpha} \psi(x_\alpha) \]

where \( A \cup B = V \)

Marginal MAP Challenging!

- NP-hard even on tree structured graphs

\[ \sum \max \neq \max \sum \]

Example from D. Koller and N. Friedman (2009)
Find the a partial configuration that maximizes its marginal distribution (marginal MAP or simply called MAP)

\[ x_B^* = \arg \max_{x_B} p(x_B) = \arg \max_{x_B} \sum_{x_A} \prod_{\alpha} \psi(x_{\alpha}) \]

where \( A \cup B = V \)

**Marginal MAP Challenging!**

- NP-hard even on tree structured graphs

Example from D. Koller and N. Friedman (2009)
Queries of Graphical Models

- Max-Inference
- Sum-Inference
- Mixed-Inference (Marginal MAP)
- Decision Making (stochastic optimization)

Even more complicated ...

Sequential decision making, influence diagram (with perfect recall), stochastic optimization (see Liu & Ihler UAI 2012)

$$\max \sum_{x_{B_m}} \ldots \max \sum_{x_{B_1}} \prod_{\alpha} \psi(x_{\alpha})$$

max: decision variables
sum: random variables
A Toy Example (MAP v.s. marginal MAP)

- \( x \in \{\text{rainy}, \text{sunny}\} \): the weather of Irvine
- \( y \in \{\text{walk}, \text{drive}\} \): whether I walk or drive to office

\[
p(x, y) = \begin{array}{|c|c|c|}
\hline
 & \text{walk} & \text{drive} \\
\hline
\text{rainy} & 0.05 & 0.35 \\
\hline
\text{sunny} & 0.3 & 0.3 \\
\hline
\end{array}
\]

Query:
Is Irvine more likely to be sunny?
### A Toy Example (MAP v.s. marginal MAP)

- $x \in \{\text{rainy, sunny}\}$: the weather of Irvine
- $y \in \{\text{walk, drive}\}$: whether I walk or drive to office

#### $p(x, y)$

<table>
<thead>
<tr>
<th></th>
<th>walk</th>
<th>drive</th>
</tr>
</thead>
<tbody>
<tr>
<td>rainy</td>
<td>0.05</td>
<td>0.35</td>
</tr>
<tr>
<td>sunny</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

**Marginal MAP:**

$$x^* = \arg \max_x p(x)$$

$$x^* = \text{sunny}$$

**Query:**

Is Irvine more likely to be sunny?
A Toy Example (MAP v.s. marginal MAP)

- $x \in \{\text{rainy, sunny}\}$: the weather of Irvine
- $y \in \{\text{walk, drive}\}$: whether I walk or drive to office

<table>
<thead>
<tr>
<th></th>
<th>walk</th>
<th>drive</th>
</tr>
</thead>
<tbody>
<tr>
<td>rainy</td>
<td>0.05</td>
<td>0.35</td>
</tr>
<tr>
<td>sunny</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

\[
p(x, y)
\]

\[
\text{Marginal MAP:} \quad x^* = \arg \max_x p(x) \quad \Rightarrow \quad x^* = \text{sunny}
\]

\[
\text{MAP:} \quad (x^*, y^*) = \arg \max_{x,y} p(x, y) \quad \Rightarrow \quad (x^*, y^*) = [\text{rainy, drive}]
\]

Query:
Is Irvine more likely to be sunny?

Paradox: Why does my behavior influence the weather!
A Toy Example (MAP v.s. marginal MAP)

- $x \in \{\text{rainy, sunny}\}$: the weather of Irvine
- $y \in \{\text{walk, drive}\}$: whether I walk or drive to office

Query:
Is Irvine more likely to be sunny?

Paradox: Why does my behavior influence the weather!

<table>
<thead>
<tr>
<th></th>
<th>walk</th>
<th>drive</th>
</tr>
</thead>
<tbody>
<tr>
<td>rainy</td>
<td>0.05</td>
<td>0.35</td>
</tr>
<tr>
<td>sunny</td>
<td>0.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>

MAP: $(x^*, y^*) = \arg\max_{x,y} p(x, y)$

$\Rightarrow (x^*, y^*) = [\text{rainy, drive}]$

Marginal MAP: $x^* = \arg\max_x p(x)$

$\Rightarrow x^* = \text{sunny}$

**DO NOT use joint MAP when only subset of variables are of direct interest!**
Different type of queries

<table>
<thead>
<tr>
<th>Type</th>
<th>Formula</th>
</tr>
</thead>
<tbody>
<tr>
<td>Max-Inference</td>
<td>$\max_{\mathbf{x}} \prod_{\alpha} \psi(\mathbf{x}_\alpha)$</td>
</tr>
<tr>
<td>Sum-Inference</td>
<td>$\sum_{\mathbf{x}} \prod_{\alpha} \psi(\mathbf{x}_\alpha)$</td>
</tr>
<tr>
<td>Mix-Inference</td>
<td>$\max_{\mathbf{x}<em>B} \sum</em>{\mathbf{x}<em>A} \prod</em>{\alpha} \psi(\mathbf{x}_\alpha)$</td>
</tr>
<tr>
<td>Sequential Decision Making</td>
<td>$\max_{\mathbf{x}<em>{B_m}} \sum</em>{\mathbf{x}<em>{A_m}} \cdots \max</em>{\mathbf{x}<em>{B_1}} \sum</em>{\mathbf{x}<em>{A_1}} \prod</em>{\alpha} \psi(\mathbf{x}_\alpha)$</td>
</tr>
</tbody>
</table>
Variational Algorithms

Replacing “elimination” with optimization over distributions
Variational Algorithms

Replacing “elimination” with optimization over distributions

Maximization:

\[
\max_x \theta(x) = \max_{q \in \mathbb{M}} \mathbb{E}_q[\theta(x)] \quad \text{(maximum: } q = 1(x^*)\text{)}
\]

\[\mathbb{M}: \text{The set of joint distributions on } x.\]
Variational Algorithms

Replacing “elimination” with optimization over distributions

- **Maximization:**
  \[
  \max_x \theta(x) = \max_{q \in \mathcal{M}} \mathbb{E}_q[\theta(x)]
  \]  
  (maximum: \( q = 1(x^\*) \))

  \( \mathcal{M} \): The set of joint distributions on \( x \).

- **Marginalization** (Wainwright & Jordan 08):
  \[
  \log Z = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x|q) \}
  \]  
  (maximum: \( q = p \))
Variational Algorithms

Replacing “elimination” with optimization over distributions

Maximization:

\[ \max_x \theta(x) = \max_{q \in \mathcal{M}} \mathbb{E}_q[\theta(x)] \quad \text{(maximum: } q = 1(x^*)) \]

\[ \mathcal{M} : \text{The set of joint distributions on } x. \]

Marginalization (Wainwright & Jordan 08):

\[ \log Z = \max_{q \in \mathcal{M}} \left\{ \mathbb{E}_q[\theta(x)] + H(x|q) \right\} \quad \text{(maximum: } q = p) \]

Proof: Using Kullback-Leibler divergence,

\[ p = \arg \min_q \{KL(q||p)\}, \quad KL(q||p) \geq 0 \]
Variational Algorithms

Replacing “elimination” with optimization over distributions

- **Maximization:**
  \[
  \max_x \theta(x) = \max_{q \in \mathcal{M}} \mathbb{E}_q[\theta(x)] \quad \text{(maximum: } q = \delta(x^*))
  \]

  \(\mathcal{M}\): The set of joint distributions on \(x\).

- **Marginalization** (Wainwright & Jordan 08):
  \[
  \log Z = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x|q) \} \quad \text{(maximum: } q = p)\]

- \(\mathcal{M}\) and the entropy are **intractable**!

- Variational approximation
Variational Algorithms

Replacing “elimination” with optimization over distributions

Maximization:

$$\max_{\mathbf{x}} \theta(\mathbf{x}) = \max_{q \in \mathcal{M}} \mathbb{E}_q[\theta(\mathbf{x})] \approx \max_{q \in \mathcal{L}} \mathbb{E}_q[\theta(\mathbf{x})]$$

\[\mathcal{L}: \text{The local consistent polytope.}\]

Marginalization (Wainwright & Jordan 08):

$$\log Z = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x|q) \} \quad \text{(maximum: } q = p)$$
Variational Algorithms

Replacing “elimination” with optimization over distributions

Maximization:
\[
\max_{\mathbf{x}} \theta(\mathbf{x}) = \max_{q \in \mathbb{M}} \mathbb{E}_q[\theta(\mathbf{x})] \approx \max_{q \in \mathbb{L}} \mathbb{E}_q[\theta(\mathbf{x})]
\]

\(\mathbb{L}\): The local consistent polytope.

Marginalization (Wainwright & Jordan 08):

\[
\log Z = \max_{q \in \mathbb{M}} \left\{ \mathbb{E}_q[\theta(\mathbf{x})] + H(\mathbf{x}|q) \right\} \quad \text{(maximum: } q = p)\]

\[
\log Z \approx \max_{q \in \mathbb{L}} \left\{ E_q[\theta(\mathbf{x})] + \sum_i H_i(q_i) - \sum_{ij} I_{ij}(q_{ij}) \right\}
\]

Bethe approximation
Variational Algorithms

Replacing “elimination” with optimization over distributions

Maximization:

\[
\max_{\mathbf{x}} \theta(\mathbf{x}) = \max_{q \in \mathcal{M}} \mathbb{E}_q [\theta(\mathbf{x})] \approx \max_{q \in \mathcal{L}} \mathbb{E}_q [\theta(\mathbf{x})]
\]

\[\mathcal{L}: \text{The local consistent polytope.}\]

Marginalization (Wainwright & Jordan 08):

\[
\log Z = \max_{q \in \mathcal{M}} \{\mathbb{E}_q [\theta(\mathbf{x})] + H(\mathbf{x}|q)\} \quad (\text{maximum: } q = p)
\]

\[
\log Z \approx \max_{q \in \mathcal{L}} \{\mathbb{E}_{\mathcal{L}} [\theta(\mathbf{x})] + \sum_{i} H_i(q_i) - \sum_{ij} I_{ij}(q_{ij})\}
\]

Solving gives loopy belief propagation algorithms [Yedidia et al. 2001]
<table>
<thead>
<tr>
<th></th>
<th>Primal form</th>
<th>Dual form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum-Inference</td>
<td>$\sum_x \exp(\theta(x))$</td>
<td>$\max_{q \in \mathcal{M}} { \mathbb{E}_q[\theta(x)] + H_x(q) }$ (Loopy BP, TRBP/negTRBP, mean field, Weighted mini-bucket ...)</td>
</tr>
<tr>
<td>Max-Inference</td>
<td>$\max_x { \exp(\theta(x)) }$</td>
<td>$\max_{q \in \mathcal{M}} \mathbb{E}_q[\theta(x)]$ (Max-prod LBP, TRBP, dual decomposition)</td>
</tr>
<tr>
<td>Marginal MAP</td>
<td>$\max_{x_B} \sum_{x_A} \exp(\theta(x))$</td>
<td></td>
</tr>
<tr>
<td>Sequential</td>
<td>$\max_{x_{Bm}} \cdots \max_{x_{B1}} \sum_{x_{A1}} \exp(\theta(x))$</td>
<td></td>
</tr>
</tbody>
</table>
Variational Representation of Marginal MAP

- For marginal MAP, \( Z_{AB} = \max_{x_B} \sum_{x_A} \exp(\theta(x)) \)

\[
\log Z_{AB} = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x_A|x_B;q) \}
\]
For marginal MAP,

\[ Z_{AB} = \max_{x_B} \sum_{x_A} \exp(\theta(x)) \]

\[
\log Z_{AB} = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x_A|x_B; q) \}
\]

**Intuition:** “truncating” the entropies of the max nodes.

**Conditional entropy:**

\[
H(x_A|x_B; q) = H(x; q) - H(x_B; q)
\]
Intuition: “truncating” the entropies of the max nodes.

Conditional entropy:
\[ H(x_A|x_B; q) = H(x; q) - H(x_B; q) \]

For marginal MAP, \[ Z_{AB} = \max_{x_B} \sum_{x_A} \exp(\theta(x)) \]
\[ \log Z_{AB} = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x_A|x_B; q) \} \]

Generalizes the dualities of max- and sum-inference:
\[ H(x_A|x_B; q) = \begin{cases} 0 & \text{if } B = V \text{ (max-inference)} \\ H(x; q) & \text{if } B = \emptyset \text{ (sum-inference)} \end{cases} \]
Variational Representation of Marginal MAP

- For marginal MAP, \( Z_{AB} = \max_{x_B} \sum_{x_A} \exp(\theta(x)) \)

\[
\log Z_{AB} = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x_A|x_B; q) \}
\]

**Intuition:** “truncating” the entropies of the max nodes.

Conditional entropy:
\[
H(x_A|x_B; q) = H(x; q) - H(x_B; q)
\]

- The optimal \( x_B^* \) is obtained by

\[
x_i^* = \arg \max_{x_i} \{ q^*(x_i) \}, \quad \forall i \in B.
\]
Variational Representation of Marginal MAP

For marginal MAP, \( Z_{AB} = \max_{x_B} \sum_{x_A} \exp(\theta(x)) \)

\[
\log Z_{AB} = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + H(x_A|x_B; q) \}
\]

**Intuition:** “truncating” the entropies of the max nodes.

Conditional entropy:

\( H(x_A|x_B; q) = H(x; q) - H(x_B; q) \)

Proof: Using the non-negativity of conditional KL divergence:

\[
\mathbb{E}_{q(x_B)}[\text{KL}(q(x_A|x_B)||p(x_A|x_B))] \geq 0, \quad \forall q \in \mathcal{M}
\]

See the paper for details
Variational Representation of Sequential Decision

- Sequential decision making:

\[ Z_{seq} = \max_{x_{B_m}} \sum_{x_{A_m}} \cdots \max_{x_{B_1}} \sum_{x_{A_1}} \exp(\theta(x)) \]

\[ \log Z_{seq} = \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta(x)] + \sum_{i=1}^{m} H(x_{A_i} | x_{pa(A_i)}; q) \} \]

\[ pa(A_i) = B_i \cup A_{i+1} \cup B_{i+1} \cup \cdots \cup A_m \cup B_m \]

- Generalizes the results of marginalization, MAP and marginal MAP (by entropic chain rule)
Even More Generally ...

- Powered-sum:

\[ \bigoplus_x^w f(x) \overset{\text{def}}{=} (\sum_x f(x)^{1/w})^w \]

- Zero-temperature limit:

\[ \lim_{w \to 0^+} \bigoplus_x^w f(x) = \max_x f(x) \quad \lim_{w \to 0^-} \bigoplus_x^w f(x) = \min_x f(x) \]

\[ \frac{1}{\bigoplus_x} f(x) = \sum_x f(x) \]

- Powered-sum generalizes both sum and max operators
Consider \( Z_w = \bigoplus_{x_n} \cdots \bigoplus_{x_2} \bigoplus_{x_1} \exp(\theta(x)) \)

\[
\log Z_w = \max_{q \in \mathcal{M}} \left\{ \mathbb{E}_q[\theta(x)] + \sum_{i=1}^{n} w_i H(x_i | x_{\text{pa}(i)}; q) \right\}
\]

Reduces to the earlier results when \( w_i = 1 \) or \( 0^+ \)

Useful for approximate inference (Liu&Ihler ICML 2011)

Useful when deriving annealed algorithm and proximal point algorithm (this work and Liu&Ihler UAI 2012)
<table>
<thead>
<tr>
<th></th>
<th>Primal form</th>
<th>Dual form</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sum-Inference</td>
<td>$\sum_x \exp(\theta(x))$</td>
<td>$\max_{q \in \mathcal{M}} { \mathbb{E}_q[\theta(x)] + H_x(q) }$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Loopy BP, TRBP/negTRBP, mean field, Weighted mini-bucket …)</td>
</tr>
<tr>
<td>Max-Inference</td>
<td>$\max_x { \exp(\theta(x)) }$</td>
<td>$\max_{q \in \mathcal{M}} \mathbb{E}_q[\theta(x)]$</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(Max-prod LBP, TRBP, dual decomposition)</td>
</tr>
<tr>
<td>Marginal MAP</td>
<td>$\max_{x_B} \sum_{x_A} \exp(\theta(x))$</td>
<td>$\max_{q \in \mathcal{M}} { \mathbb{E}<em>q[\theta(x)] + \sum</em>{i=1}^m H(X_{A_i}</td>
</tr>
<tr>
<td>Sequential Decision Making</td>
<td>$\max_{x_{B_m}} \sum_{x_{A_m}} \cdots \max_{x_{B_1}} \sum_{x_{A_1}} \exp(\theta(x))$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Variational Approximation (Marginal MAP)

- **Bethe approximation:**

\[
H(x) \approx \sum_{i \in A \cup B} H_i - \sum_{(ij) \in E} I_{ij}, \quad H(x_B) \approx \sum_{i \in B} H_i - \sum_{(ij) \in E_B} I_{ij}
\]

\[
H(x_A | x_B) = H(x) - H(x_B) \approx \sum_{i \in A} H_i - \sum_{(ij) \in E \setminus E_B} I_{ij}
\]

**Intuition:** “truncating” all the entropy and mutual information terms involves only with max

\[
\text{max (B)} \quad \text{sum (A)}
\]
Variational Approximation (Marginal MAP)

- **Bethe approximation:**

  \[ H(x) \approx \sum_{i \in A \cup B} H_i - \sum_{(ij) \in E} I_{ij}, \quad H(x_B) \approx \sum_{i \in B} H_i - \sum_{(ij) \in E_B} I_{ij} \]

  \[ H(x_A | x_B) = H(x) - H(x_B) \approx \sum_{i \in A} H_i - \sum_{(ij) \in E \setminus E_B} I_{ij} \]

- **Approximate variational form:**

  \[ \log Z_{AB} \approx \max_{q \in \mathbb{L}} \{ \mathbb{E}_q[\theta(x)] + \sum_{i \in A} H_i - \sum_{(ij) \in E \setminus E_B} I_{ij} \} \]

- **Intuition:** “truncating” all the entropy and mutual information terms involves only with max

- **Solving using Lagrangian multiplier method** gives a loopy BP algorithm for marginal MAP!
“Mixed”-product BP (marginal MAP)

\[ \text{sum } \Rightarrow \text{sum } \cup \text{max } : \]
(Sum-product)

\[ \text{max } \Rightarrow \text{max } : \]
(Max-product)

\[ \text{max } \Rightarrow \text{sum } : \]
(Argmax-product)

\[ m_{i \rightarrow j} \propto \sum_{x_i} \psi_{i j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

where \( \mathcal{X}_i^* = \text{arg max}_{x_i} b_i(x_i), \quad b_i(x_i) \propto \psi_i \prod_k m_{k \rightarrow i} \)
“Mixed”-product BP (marginal MAP)

\[ \text{sum} \Rightarrow \text{sum} \cup \text{max} : \]

(Sum-product)

\[ m_{i \rightarrow j} \propto \sum_{x_i} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

(max ⇒ max :

(Max-product)

\[ m_{i \rightarrow j} \propto \max_{x_i} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

(max ⇒ sum :

(Argmax-product)

\[ m_{i \rightarrow j} \propto \sum_{x_i \in \mathcal{X}_i^*} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

where \( \mathcal{X}_i^* = \arg \max_{x_i} b_i(x_i) \),

\[ b_i(x_i) \propto \psi_i \prod_k m_{k \rightarrow i} \]

Restrict the sum to the set of argmax values.

Solving a “local” marginal MAP problem based on the current belief \( b_i(x_i) \).
“Mixed”-product BP (marginal MAP)

\[ \text{sum} \Rightarrow \text{sum} \cup \text{max} : \]
\[ (\text{Sum-product}) \]
\[ \text{max} \Rightarrow \text{max} : \]
\[ (\text{Max-product}) \]
\[ \text{max} \Rightarrow \text{sum} : \]
\[ (\text{Argmax-product}) \]

\[ m_{i \rightarrow j} \propto \sum_{x_i} \psi_{ij} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

\[ m_{i \rightarrow j} \propto \max_{x_i} \psi_{ij} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

\[ m_{i \rightarrow j} \propto \sum_{x_i \in \mathcal{X}_i^*} \psi_{ij} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

where \( \mathcal{X}_i^* = \arg \max_{x_i} b_i(x_i) \),
\[ b_i(x_i) \propto \psi_i \prod_k m_{k \rightarrow i} \]

* Decoding solution:

\[ x_i^* = \arg \max_{x_i} b_i(x_i), \forall i \in B \]
“Mixed”-product BP (marginal MAP)

\[ \text{sum} \Rightarrow \text{sum} \cup \text{max} : \]
\[ \text{(Sum-product)} \]
\[ \text{max} \Rightarrow \text{max} : \]
\[ \text{(Max-product)} \]
\[ \text{max} \Rightarrow \text{sum} : \]
\[ \text{(Argmax-product)} \]

\[ m_{i \rightarrow j} \propto \sum_{x_i} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]
\[ m_{i \rightarrow j} \propto \max_{x_i} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]
\[ m_{i \rightarrow j} \propto \sum_{x_i \in \mathcal{X}_i^*} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

where \( \mathcal{X}_i^* = \arg \max_{x_i} b_i(x_i), \quad b_i(x_i) \propto \psi_i \prod_k m_{k \rightarrow i} \)

- “Simultaneously” solves max & sum subproblems
- fast, no inner loops
“Mixed”-product BP (marginal MAP)

\[ \text{sum} \Rightarrow \text{sum} \cup \text{max} : \]
\[(\text{Sum-product})\]

\[ m_{i \rightarrow j} \propto \sum_{x_i} \psi_{ij} \psi_{i} \prod_{k \neq j} m_{k \rightarrow i} \]

\[ \text{max} \Rightarrow \text{max} : \]
\[(\text{Max-product})\]

\[ m_{i \rightarrow j} \propto \max_{x_i} \psi_{ij} \psi_{i} \prod_{k \neq j} m_{k \rightarrow i} \]

\[ \text{max} \Rightarrow \text{sum} : \]
\[(\text{Argmax-product})\]

\[ m_{i \rightarrow j} \propto \sum_{x_i \in \mathcal{X}_i^*} \psi_{ij} \psi_{i} \prod_{k \neq j} m_{k \rightarrow i} \]

where \( \mathcal{X}_i^* = \arg \max_{x_i} b_i(x_i) \), \( b_i(x_i) \propto \psi_{i} \prod_{k} m_{k \rightarrow i} \)

- “Simultaneously” solves max & sum subproblems
- fast, no inner loops

- Strong local optimality Guarantees
  - Greedy algorithms: coordinate-wise local optimality
  - Mixed-Product BP: tree block-wise local optimality
A similar hybrid message passing algorithm proposed by Jiang & Daume III 2011:

\[ \text{sum} \Rightarrow \text{sum} \cup \text{sum} : \quad \text{(Sum-product)} \]
\[ m_{i \rightarrow j} \propto \sum_{x_i} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

\[ \text{max} \Rightarrow \text{sum} \cup \text{max} : \quad \text{(Max-product)} \]
\[ m_{i \rightarrow j} \propto \max_{x_i} \psi_{i,j} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]
A similar hybrid message passing algorithm proposed by Jiang & Daume III 2011:

- \[ \text{sum} \Rightarrow \text{sum} \cup \text{sum} : \] (Sum-product)
  \[ m_{i \rightarrow j} \propto \sum_{x_i} \psi_{ij} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

- \[ \text{max} \Rightarrow \text{sum} \cup \text{max} : \] (Max-product)
  \[ m_{i \rightarrow j} \propto \max_{x_i} \psi_{ij} \psi_i \prod_{k \neq j} m_{k \rightarrow i} \]

---

**Mixed-product BP**

- \( \text{max} (B) \)
- \( \text{sum} (A) \)

**Jiang’s Method**

- \( \text{max} (B) \)
- \( \text{sum} (A) \)

---

Argmax-product is critical!
Mixed-product v.s. Jiang’s Algorithm

True solution:

\[ x_4^* = \arg \max_{x_4} \max_{x_3} \sum_{x_2} \sum_{x_1} [\exp(\theta(x))]. \]
Mixed-product v.s. Jiang’s Algorithm

- **True solution:**
  \[ x^*_4 = \arg\max_{x_4} \max_{x_3} \sum_{x_2} \sum_{x_1} [\exp(\theta(x))] \]

- **Jiang’s algorithm:**
  \[ x^*_4 = \arg\max_{x_4} \sum_{x_2} \sum_{x_1} \max_{x_3} [\exp(\theta(x))] \]

Exchanging the max / sum order
Mixed-product v.s. Jiang’s Algorithm

- **True solution:**
  \[ x_4^* = \arg \max_{x_4} \max_{x_3} \sum_{x_2} \sum_{x_1} \exp(\theta(x)) \].

- **Jiang’s algorithm:**
  \[ x_4^* = \arg \max_{x_4} \sum_{x_2} \sum_{x_1} \max_{x_3} \exp(\theta(x)) \],

- **Mixed-product BP:**
  \[ x_4^{t+1} = \arg \max_{x_4} \sum_{x_2} \sum_{x_1} \exp(\theta([x_3^{t}, x_{-3}])) \]

**Diagram:**
- **Max** order.
- **Sum** order.
- Exchanging the max / sum order.
- Equivalent to coordinate descent.
The concave-convex procedure (CCCP) (Yuille & Rangarajan 03):

Consider optimization problem,

$$\min_q \left\{ f(q) \equiv f_+(q) - f_-(q) \right\}$$
The concave-convex procedure (CCCP) \textit{(Yuille & Rangarajan 03)}:

Consider optimization problem,

\[
\min_q \{ f(q) \equiv f_+(q) - f_-(q) \}
\]

Concave-convex procedure:

For iteration $t$

\[
q^{t+1} = \arg \max_q \{ f_+(q) - \langle \nabla f_-(q^t), q \rangle \}
\]

End
Our problem:

\[
\max_{q \in \mathcal{M}} \{ f(q) \equiv \mathbb{E}_q[\theta(x)] + H(x; q) - H(x_B; q) \}
\]

\[
f_+(q) \quad f_-(q) \approx \mathbb{E}_q[-\log q^*(x)] + \text{const}
\]
Our problem:

\[
\max_{q \in \mathcal{M}} \{ f(q) \equiv \mathbb{E}_q[\theta(x)] + H(x; q) - H(x_B; q) \} \\
\]

\[
f_+ (q) \quad \quad f_- (q) \approx \mathbb{E}_q[-\log q^t(x)] + \text{const}
\]

Concave-convex procedure:

For iteration t

\[
\theta^t(x) = \theta(x) + \log q^t(x)
\]

\[
q^{t+1} = \arg \max_{q \in \mathcal{M}} \{ \mathbb{E}_q[\theta^t(x) + H_x(q)] \}
\]

End
Our problem:

\[
\max_{q \in \mathcal{M}} \left\{ f(q) = \mathbb{E}_q[\theta(x)] + H(x; q) - H(x_B; q) \right\}
\]

\[
f_+(q) \quad f_-(q) \approx \mathbb{E}_q[-\log q^t(x)] + \text{const}
\]

Concave-convex procedure:

For iteration t

\[
\theta^t(x) = \theta(x) + \log q^t(x)
\]

\[
q^{t+1} = \arg \max_{q \in \mathcal{M}} \left\{ \mathbb{E}_q[\theta^t(x)] + H_x(q) \right\}
\]

End
Convergent Algorithms

More general decomposition:

\[
\max_{q \in \mathbb{M}} \{ f(q) \equiv \mathbb{E}_q[\theta(x)] + H(x_A|x_B) + w^t H(x_B) - w^t H(x_B) \} + f_+(q) + f_-(q)
\]

Proximal point algorithm:

For iteration \(t\), \((w^t = 1/t)\)

\[
\theta^t(x) = \theta(x) + w^t \log q^t(x)
\]

\[
q^{t+1} = \arg \max_{q \in \mathbb{M}} \{ \mathbb{E}_q[\theta^t(x)] + H(x_A|x_B) + w^t H(x_B) \}
\]

End
Experiments (Marginal MAP)

- Hidden Markov Chain (of length 10):

  \[ \text{max} \ (B) \]

  \[ \text{sum} \ (A) \]

Figure 3: Results on the hidden Markov chain of Fig. 1. The max-product BP and sum-product BP are among the worst of the tested algorithms, indicating the danger of approximating mixed-inference by pure max- or sum-inferences. Interestingly, the performances of max-product BP and sum-product BP have been known for TRW-BP on standard max- and sum- inference. EM performs similarly to (sometimes worse than) Jiang's method. We note that our TRW approximation gives much less accurate solutions than the mix-product (Bethe) and leveraging our results to learn hidden variables models for data.

Figure 6: (a) A marginal MAP problem defined on a 10 pow 4 grid, where the max parts are fully disconnected

(b) The approximate relative errors of different algorithms as a function of coupling strength

Thanks Arthur Choi for help on Samlam!
Experiments (Marginal MAP)

- 10X10 Grid (with tree structured max part)

![Diagram of a 10X10 grid with labeled max (B) and sum (A) nodes.]

Future directions include improving the performance of the truncated TRW approximation by optimizing weights, deriving optimality conditions that may be applicable even approximately, opening new doors for developing efficient algorithms. In particular, we show that our proposed "mixed-product" BP admires appealing theoretical properties and performs surprisingly well in practice.

We have presented a general variational framework for solving marginal MAP problems when the sum component does not form a tree, studying the convergent properties of mixed-product by optimizing weights, deriving optimality conditions that may be applicable even approximately, opening new doors for developing efficient algorithms. In particular, we show that our proposed "mixed-product" BP admires appealing theoretical properties and performs surprisingly well in practice.
Experiments (Marginal MAP)

- 10X10 Grid (with tree structured sum part)

```
max (B): green
sum (A): orange
```

<table>
<thead>
<tr>
<th>Algorithms</th>
<th>Line Style</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mix-product (Bethe)</td>
<td>blue square</td>
</tr>
<tr>
<td>Proximal (Bethe)</td>
<td>green square</td>
</tr>
<tr>
<td>Proximal (TRW)</td>
<td>blue circle</td>
</tr>
<tr>
<td>Proximal (TRW, upper bound)</td>
<td>blue triangle</td>
</tr>
<tr>
<td>Max-product</td>
<td>red square</td>
</tr>
<tr>
<td>Sum-product</td>
<td>red triangle</td>
</tr>
<tr>
<td>Jiang’s method</td>
<td>pink square</td>
</tr>
<tr>
<td>EM</td>
<td>cyan line</td>
</tr>
<tr>
<td>Local Search (SamIam)</td>
<td>brown star</td>
</tr>
</tbody>
</table>

Figure 5: (a) A marginal MAP problem defined on a 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Future directions include improving the performance of the truncated TRW approximation by optimizing weights, deriving optimality conditions that may be applicable even approximately, opening new doors for developing efficient algorithms. In particular, we have presented a general variational framework for solving marginal MAP problems. We leverage our results to learn hidden variables models for data.

10. Conclusion and Further Directions

Experiments (Marginal MAP)
Experiments (Marginal MAP)

- Diagnostic Bayesian Network (from UAI08 challenge):
  - randomly select a set of max nodes

Figure 7: The results on two diagnostic Bayesian networks (BNs) in UAI08 inference challenge. (a) the structure of diagnostic BN-2. (b)-(c) The performances of algorithms on these two BNs as the percentage of max nodes increases. Averaged on 100 random trials.

Acknowledgments
We thank Arthur Choi for providing help on SamIam. This work was supported in part by NSF grant IIS-1065618 and a Microsoft Research Ph.D Fellowship.

References
Conclusions

- An extension of the variational inference framework

- **Mixed-product BP**
  - Solve sum & max “simultaneously”
  - Fast, scalable to large models
  - Strong local optimality guarantees

- **CCCP & proximal point algorithms**
  - Iteratively solving marginalization problems
  - Convergent & better performance
Conclusions

- An extension of the variational inference framework

- Mixed-product BP
  - Solve sum & max “simultaneously”
  - Fast, scalable to large models
  - Strong local optimality guarantees

- CCCP & proximal point algorithms
  - Iteratively solving marginalization problems
  - Convergent & better performance

Thanks

Acknowledgements: Work supported in part by NSF and a Microsoft Research Fellowship.
Experiments

- Hidden Markov Chain (of length 10):

![Diagram of a latent tree model](fig:rand_tree_result)

**Figure 4**: (a) A typical latent tree model, whose leaf nodes are taken to be max nodes (shaped). (b) The approximate relative energy errors of different algorithms as a function of coupling strength.

**Figure 5**: (a) A marginal MAP problem defined on a 10-dimensional Ising grid, but with max / sum pattern approximately, opening new doors for developing efficient algorithms.

**Figure 6**: (a) A marginal MAP problem defined on a 10-dimensional Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

The regular max-product BP and sum-product BP are among the worst of the tested algorithms, indicating the danger of approximating mixed-inference by pure max- or sum-product algorithms. In particular, we show that our proposed "mixed-product" BP admires appealing theoretical properties and performs surprisingly well in practice.
Experiments

Hidden Markov Chain (of length 10):

max (B)

sum (A)

Coupling Strength

Percentage of Correct Solutions

Relative Errc.
Experiments

- 10X10 Grid (with tree structured max part)

We have presented a general variational framework for solving marginal MAP problems approximately, opening new doors for developing efficient algorithms. In particular, we show that our proposed "mixed-product" BP admires appealing theoretical properties and performs surprisingly well in practice.

Future directions include improving the performance of the truncated TRW approximation by optimizing weights, deriving optimality conditions that may be applicable even when the sum component does not form a tree, studying the convergent properties of mixed-product BP, and leveraging our results to learn hidden variables models for data.

Experiments

- 10X10 Grid (with tree structured max part)

Figure 5: (a) A marginal MAP problem defined on a 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 Ising grid, but with max / product BP, and leveraging our results to learn hidden variables models for data.

Approximate Relative Error

<table>
<thead>
<tr>
<th>Coupling Strength</th>
<th>Mix–product (Bethe)</th>
<th>Proximal (Bethe)</th>
<th>Proximal (TRW)</th>
<th>Proximal (TRW, upper bound)</th>
<th>Max–product</th>
<th>Sum–product</th>
<th>Jiang's method</th>
<th>EM</th>
<th>Local Search (SamIam)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>1.5</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Approximate Relative Error
Experiments

10X10 Grid (with tree structured max part)

Future directions include improving the performance of the truncated TRW approximation by optimizing weights, deriving optimality conditions that may be applicable even approximately, opening new doors for developing efficient algorithms.

Liu and Ihler's method performs surprisingly well in practice.

We have presented a general variational framework for solving marginal MAP problems as a function of coupling strength.

**Experiments**

- **10X10 Grid (with tree structured max part)**

**Approximate Relative Error**

![Graph showing approximate relative error against coupling strength](fig:chessboard_result)

**Conclusion and Further Directions**

We have presented a general variational framework for solving marginal MAP problems as a function of coupling strength.

Future directions include improving the performance of the truncated TRW approximation by optimizing weights, deriving optimality conditions that may be applicable even approximately, opening new doors for developing efficient algorithms.
10X10 Grid (with tree structured sum part)
Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.

Experiments

- 10X10 Grid (with tree structured sum part)

max (B):  

sum (A):  

Future directions include improving the performance of the truncated TRW approximation as a function of coupling strength.

Figure 6: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, with shaded sum nodes and unshaded max nodes; note that the sum part is a loopy graph, while max part is fully disconnected. (b) The approximate relative errors of different algorithms as a function of coupling strength.

Figure 5: (a) A marginal MAP problem defined on a 10 X 10 Ising grid, but with max / sum part exactly opposite to that in Fig. 6.