Mixing Sum-Product and Max-Product to Tighten log-Partition Upper Bounds

Tamir Hazan
Toyota Technological Institute at Chicago

Joint work: Amnon Shashua
Inference in Graphical Models
Inference in Graphical Models

- Infer label for every pixel $x_i \in \{\text{person, bus, airplane, bg}\}$
Inference in Graphical Models

- Infer label for every pixel $x_i \in \{\text{person, bus, airplane, bg}\}$

- Exponential complexity: Use local (graph) information
Inference in Graphical Models

- Infer label for every pixel $x_i \in \{\text{person, bus, airplane, bg}\}$

- Exponential complexity: Use local (graph) information

- Pixel Information $\theta_i(x_i)$
Inference in Graphical Models

- Infer label for every pixel \( x_i \in \{ \text{person, bus, airplane, bg} \} \)

- Exponential complexity: Use local (graph) information

- Pixel Information \( \theta_i(x_i) \)

- Pairwise \( \theta_{i,j}(x_i, x_j) \)
Inference in Graphical Models

- Infer label for every pixel \( x_i \in \{\text{person, bus, airplane, bg}\} \)

- Exponential complexity: Use local (graph) information

- Pixel Information \( \theta_i(x_i) \)

- Pairwise \( \theta_{i,j}(x_i, x_j) \)

- Region \( \theta_r(x_r) \quad r \subset \{1, \ldots, \text{total nodes}\} \)
Inference in Graphical Models

Input: local weights

$$\psi_\alpha(x_\alpha) = \exp \theta_\alpha(x_\alpha)$$

$\alpha$ represents $i, (i, j), r$
Inference in Graphical Models

- **Input: local weights**
  \[ \psi_\alpha(x_\alpha) = \exp \theta_\alpha(x_\alpha) \]
  \( \alpha \) represents \( i, (i, j), r \)

1) The best scoring configuration (MAP)

\[
\text{argmax}_{x_1, \ldots, x_n} \prod_{\alpha} \psi_\alpha(x_\alpha)
\]
Inference in Graphical Models

- Input: local weights
  \[ \psi_\alpha(x_\alpha) = \exp \theta_\alpha(x_\alpha) \]
  \( \alpha \) represents \( i, (i, j), r \)

1) The best scoring configuration (MAP)
\[
\arg\max_{x_1, \ldots, x_n} \prod_{\alpha} \psi_\alpha(x_\alpha)
\]

2) The relative weight over all configuration (marginals)
- partition function \( Z = \sum_{x_1, \ldots, x_n} \prod_{\alpha} \psi_\alpha(x_\alpha) \)
- weight of partial assignment \( Z(x_\beta) = \sum_{x_1, \ldots, x_n \setminus x_\beta} \prod_{\alpha} \psi_\alpha(x_\alpha) \)
Inference in Graphical Models

\[ \prod_{\alpha} \psi_{\alpha}(x_{\alpha}) \]
Inference in Graphical Models

\[ \prod_{\alpha} \psi_{\alpha}(x_{\alpha}) \]

- \( \psi_3 \)
- \( \psi_{2,3,4} \)
- \( \psi_{3,5} \)
- \( \psi_{1,2,3} \)
- \( \psi_{2,3} \)
Inference in Graphical Models

$$\prod_{\alpha} \psi_\alpha(x_\alpha)$$

Belief Propagation
Belief Propagation
Belief Propagation

\[ \prod_{\alpha} \psi_\alpha(x_\alpha) \]
Belief Propagation
Belief Propagation

**child-to-parent**

\[ n_{\beta \rightarrow \alpha}(x_\beta) = \psi_\beta(x_\beta) \prod_{p \in P(\beta) \setminus \alpha} m_{\beta \rightarrow p}(x_\beta) \]

**parent-to-child**

\[ m_{\alpha \rightarrow \beta}(x_\beta) = \sum_{x_\alpha \setminus x_\beta} \psi_\alpha(x_\alpha) \prod_{c \in C(\alpha) \setminus \beta} n_{c \rightarrow \alpha}(x_c) \]

\[ Z = \sum_x \prod_\alpha \psi_\alpha(x_\alpha) \]

**Belief Propagation**

\[ \prod_\alpha \psi_\alpha(x_\alpha) \]
Inference in Graphical Models

$$\prod_{\alpha} \psi_{\alpha}(x_{\alpha}) \rightarrow \text{argmax}_x \prod_{\alpha} \psi_{\alpha}(x_{\alpha})$$

Belief Propagation

child-to-parent

$$n_{\beta \rightarrow \alpha}(x_{\beta}) = \psi_{\beta}(x_{\beta}) \prod_{p \in P(\beta) \setminus \alpha} m_{\beta \rightarrow p}(x_{\beta})$$

max-product

$$m_{\alpha \rightarrow \beta}(x_{\beta}) = \max_{x_{\alpha \setminus \beta}} \psi_{\alpha}(x_{\alpha}) \prod_{c \in C(\alpha) \setminus \beta} n_{c \rightarrow \alpha}(x_{c})$$

parent-to-child
Inference as Optimization
Inference as Optimization

\[
\begin{align*}
\text{MAP} & = \sum_{x_{\alpha} \setminus x_{\beta}} \max_{\alpha, x_{\alpha}} \sum \ b_{\alpha}(x_{\alpha}) \theta_{\alpha}(x_{\alpha}) \\
\beta_{\alpha}(x_{\alpha}) & \text{ are probability distributions, } b_{\alpha}(x_{\alpha}) \in \{0, 1\}
\end{align*}
\]
Inference as Optimization

**MAP**

\[
\sum_{x_\alpha \setminus x_\beta} b_{\alpha}(x_\alpha) = b_{\beta}(x_\beta) \max_{\alpha, x_\alpha} \sum b_{\alpha}(x_\alpha) \theta_{\alpha}(x_\alpha)
\]

- \(b_{\alpha}(x_\alpha)\) are probability distributions, \(b_{\alpha}(x_\alpha) \in \{0, 1\}\)

\[
\sum_{x_1} b_{1,2,3}(x_1, x_2, x_3) = b_{2,3}(x_2, x_3)
\]
Inference as Optimization

**MAP**

\[
\sum_{x_\alpha \setminus x_\beta} \max_{\alpha, x_\alpha} \sum b_\alpha(x_\alpha) \theta_\alpha(x_\alpha)
\]

- \(b_\alpha(x_\alpha)\) are probability distributions, \(b_\alpha(x_\alpha) \in \{0, 1\}\)

- LP-relaxation: ignores integral constraints.

Saturday, December 11, 2010
Inference as Optimization

**MAP**

\[
\sum_{x_\alpha \setminus x_\beta} \max_{\alpha, x_\alpha} \sum b_\alpha(x_\alpha) \theta_\alpha(x_\alpha)
\]

- \( b_\alpha(x_\alpha) \) are probability distributions, \( b_\alpha(x_\alpha) \in \{0, 1\} \)

\[
\sum_{x_2, x_4} b_{2,3,4,5}(x_2, x_3, x_4, x_5) = b_{3,5}(x_3, x_5)
\]

- LP-relaxation: ignores integral constraints.
- Add marginalization constraints on larger regions (region with n nodes guarantee integral solution)
Inference as Optimization

log-partition \[ \max \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_\alpha c_\alpha H(b_\alpha) \]
Inference as Optimization

log-partition

\[
\max \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_\alpha c_\alpha H(b_\alpha)
\]
Inference as Optimization

log-partition

\[
\max \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha)
\]

\(c_{3,5} = 1\)
\(c_3 = 0\)
\(c_{1,2,3} = 1\)
\(c_{2,3,4} = 1\)
\(c_{2,3} = 0\)

- Upper bound the log-partition with entropies
Inference as Optimization

log-partition

\[
\max \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha)
\]

- Upper bound the log-partition with entropies
- Add larger regions
Inference as Optimization

\[
\text{max} \quad \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha)
\]

- Upper bound the log-partition with entropies
- Add larger regions

1) Add consistency constraints.
Inference as Optimization

log-partition \[ \max \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha) \]

- Upper bound the log-partition with entropies
- Add larger regions

1) Add consistency constraints.
2) Tighten the entropy.
Related Work

\[
\max \sum_{\alpha, x_{\alpha}} b_{\alpha}(x_{\alpha}) \theta_{\alpha}(x_{\alpha}) + \epsilon \sum_{\alpha} c_{\alpha} H(b_{\alpha})
\]
Related Work

\[
\max \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha)
\]

\(\epsilon = 1\)

- sum-product Gallager 63, Pearl 88, \(c_\alpha = 1 - |P(\alpha)|\)
- sum-TRBP (Wainwright et al 05, \(c_\alpha \leq 0\))
- Convex message-passing (Heskes 07, \(c_\alpha > 0\))
- TCBO (Meltzer 09)
Related Work

\[
\max_{\alpha, x_\alpha} \sum \alpha \beta_\alpha (x_\alpha) \theta_\alpha (x_\alpha) + \epsilon \sum \alpha c_\alpha H(b_\alpha)
\]

- \( \epsilon = 1 \)
  - sum-product Gallager 63, Pearl 88, \( c_\alpha = 1 - |P(\alpha)| \), sum-TRBP (Wainwright et al 05, \( c_\alpha \leq 0 \)), Convex message-passing (Heskes 07, \( c_\alpha > 0 \)), TCBO (Meltzer 09).

- \( \epsilon = 0 \)
  - max-product, max-TRBP, MSD (Schlesinger 75, Koster 98, Werner 07), TRW-S (Kolmogorov 05), NMPLP, MPLP (Globerson et al 07), TBCD (Sontag 09), TCBO (Meltzer 09), OPD (Batra 10), Komodakis et al 10.
Related Work

\[
\max_{\alpha, x_\alpha} \sum_{\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha)
\]

- \(\epsilon = 1\)
  - sum-product Gallager 63, Pearl 88, \(c_\alpha = 1 - |P(\alpha)|\), sum-TRBP (Wainwright et al 05, \(c_\alpha \leq 0\)), Convex message-passing (Heskes 07, \(c_\alpha > 0\)), TCBO (Meltzer 09).

- \(\epsilon = 0\)
  - max-product, max-TRBP, MSD (Schlesinger 75, Koster 98, Werner 07), TRW-S (Kolmogorov 05), NMPLP, MPLP (Globerson et al 07), TBCD (Sontag 09), TCBO (Meltzer 09), OPD (Batra 10), Komodakis et al 10.

- Temperature methods (\(\epsilon \to 0\)) Weiss et al 07, Johnson et al 07, Jojic 10)
  - Proximal methods (Ravikumar 08).
\[ \sum_{x_\alpha \setminus x_\beta} \max b_\alpha(x_\alpha) = b_\beta(x_\beta) \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha) \]
\[
\sum_{x_\alpha \setminus x_\beta} \max b_\alpha(x_\alpha) = b_\beta(x_\beta) \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha)
\]

- convex and non-convex belief propagation. parent message depends on the \((1/\epsilon c_\alpha)\)-norm
convex and non-convex belief propagation.

parent message depends on the $(1/\epsilon c_\alpha)$-norm
convex and non-convex belief propagation.

parent message depends on the $(1/\epsilon c_\alpha)$-norm

convex belief propagation for some $c_\alpha \leq 0$.

New messages over “entropy graph”.

$$\max \sum_{x_\alpha \neq x_\beta} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_\alpha c_\alpha H(b_\alpha)$$
Duality

strictly convex $\iff$ smooth

$$L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \to \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right)$$
Duality

strictly convex $\longleftrightarrow$ smooth

$$L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_\alpha c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right)$$

$$- \left( \sum_{x_\beta} b_\beta(x_\beta) \theta_\beta(x_\beta) + \epsilon c_\beta H(b_\beta) \right)$$

$$\sum_{\alpha \in P(\beta), x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha \in P(\beta)} c_\alpha H(b_\alpha)$$
Duality

strictly convex $\iff$ smooth

\[
L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right)
\]

slope $= \lambda$

Saturday, December 11, 2010
Duality

strictly convex \longleftrightarrow \text{ smooth}

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]
Duality

strictly convex $\leftrightarrow$ smooth

$$L() = \sum_{\alpha,x_{\alpha}} b_{\alpha}(x_{\alpha})\theta_{\alpha}(x_{\alpha}) + \epsilon \sum_{\alpha} c_{\alpha} H(b_{\alpha}) + \sum_{\beta,\alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_{\beta}) \left( \sum_{x_{\alpha} \setminus x_{\beta}} b_{\alpha}(x_{\alpha}) - b_{\beta}(x_{\beta}) \right)$$

Re-parameterization

$$b^*_{\alpha}(x_{\alpha}) \propto \exp \left( \frac{\theta_{\alpha}(x_{\alpha}) + \sum_{c \in C(\alpha)} \lambda_{c \rightarrow \alpha}(x_{c}) - \sum_{p \in P(\alpha)} \lambda_{\alpha \rightarrow p}(x_{\alpha})}{\epsilon c_{\alpha}} \right)$$
Duality

strictly convex $\iff$ smooth

\[
L() = \sum_{\alpha,x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\beta,\alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right)
\]

Re-parameterization

\[
b^*_\alpha(x_\alpha) \propto \exp \left( \theta_\alpha(x_\alpha) + \sum_{c \in C(\alpha)} \lambda_{c \rightarrow \alpha}(x_c) - \sum_{p \in P(\alpha)} \lambda_{\alpha \rightarrow p}(x_\alpha) \right)
\]

Non-convex: $c_\alpha \leq 0$

saddle points

Saturday, December 11, 2010
Duality

linear program (MAP) \longleftrightarrow \text{non-smooth}

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]
Duality

linear program (MAP) $\iff$ non-smooth

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \to \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]
Duality
linear program (MAP) ↔ non-smooth

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \to \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]

Re-parameterization

\[
\text{support}(b^*_\alpha(x_\alpha)) \subset \text{argmax}_{x_\alpha} \left\{ \theta_\alpha(x_\alpha) + \sum_{c \in C(\alpha)} \lambda_{c \to \alpha}(x_c) + \sum_{p \in P(\alpha)} \lambda_{\alpha \to p}(x_\alpha) \right\}
\]

max-marginals
(Danskin Theorem, e.g. no-ties=integral)
Duality

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]

- **Dual function**

  \[ \max_{b_\alpha(x_\alpha)} \text{ is probability} \]

  \[ L(b_\alpha(x_\alpha), \lambda_{\beta \rightarrow \alpha}(x_\beta)) \]
Duality

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]

**Dual function**

\( b_\alpha(x_\alpha) \) is probability

\[ \max \quad L(b_\alpha(x_\alpha), \lambda_{\beta \rightarrow \alpha}(x_\beta)) \]

\[ \min_{\lambda_{\beta \rightarrow \alpha}(x_\beta)} \sum_{\alpha} \ln \left\| \exp \left( \theta_\alpha(x_\alpha) + \sum_{c \in C(\alpha)} \lambda_{c \rightarrow \alpha}(x_c) - \sum_{p \in P(\alpha)} \lambda_{\alpha \rightarrow p}(x_\alpha) \right) \right\|_1 / \epsilon_{c_\alpha} \]
Duality

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_\alpha c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]

- **Dual function**
  
  \[ \max_{b_\alpha(x_\alpha)} \quad \text{subject to} \quad b_\alpha(x_\alpha) \text{ is probability} \quad L(b_\alpha(x_\alpha), \lambda_{\beta \rightarrow \alpha}(x_\beta)) \]

- **MAP - piecewise linear.**

  \[ \min_{\lambda_{\beta \rightarrow \alpha}(x_\beta)} \sum_\alpha \ln \left\| \exp \left( \theta_\alpha(x_\alpha) + \sum_{c \in C(\alpha)} \lambda_{c \rightarrow \alpha}(x_c) - \sum_{p \in P(\alpha)} \lambda_{\alpha \rightarrow p}(x_\alpha) \right) \right\|^{1/\epsilon c_\alpha} \]
Duality

\[ L() = \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \epsilon \sum_\alpha c_\alpha H(b_\alpha) + \sum_{\beta, \alpha \in P(\beta)} \lambda_{\beta \rightarrow \alpha}(x_\beta) \left( \sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) - b_\beta(x_\beta) \right) \]

- Dual function

\[ \text{max} \quad b_\alpha(x_\alpha) \text{ is probability} \quad L(b_\alpha(x_\alpha), \lambda_{\beta \rightarrow \alpha}(x_\beta)) \]

\[ \min_{\lambda_{\beta \rightarrow \alpha}(x_\beta)} \sum_\alpha \ln \left\| \exp \left( \theta_\alpha(x_\alpha) + \sum_{c \in C(\alpha)} \lambda_{c \rightarrow \alpha}(x_c) - \sum_{p \in P(\alpha)} \lambda_{\alpha \rightarrow p}(x_\alpha) \right) \right\|_1 / \epsilon c_\alpha \]

- MAP - piecewise linear.

- Low temperature - smooth approximation (soft-max)
The Norm-Product

\[ m_{\alpha \to \beta}(x_{\beta}) = \left\| \psi_{\alpha}(x_{\alpha}) \prod_{c \in C(\alpha) \setminus \beta} n_{c \to \alpha}(x_{c}) \prod_{p \in P(\alpha)} n_{\alpha \to p}^{-1}(x_{\alpha}) \right\|_{1/\epsilon c_{\alpha}} \]

\[ n_{\beta \to \alpha}(x_{\beta}) = \left( \psi_{\beta}(x_{\beta}) \prod_{p \in P(\beta)} m_{p \to \beta}(x_{\beta}) \prod_{c \in C(\beta)} n_{c \to \beta}(x_{c}) \right)^{c_{\alpha} / \hat{c}_{\beta}} / m_{\alpha \to \beta}(x_{\beta}) \]
The Norm-Product

\[ m_{\alpha \rightarrow \beta}(x_\beta) = \left\| \psi_\alpha(x_\alpha) \prod_{c \in C(\alpha) \setminus \beta} n_{c \rightarrow \alpha}(x_c) \prod_{p \in P(\alpha)} n_{c \rightarrow p}^{-1}(x_\alpha) \right\| \frac{1}{\epsilon c_\alpha} \]

\[ n_{\beta \rightarrow \alpha}(x_\beta) = \left( \psi_\beta(x_\beta) \prod_{p \in P(\beta)} m_{p \rightarrow \beta}(x_\beta) \prod_{c \in C(\beta)} n_{c \rightarrow \beta}(x_c) \right)^{c_\alpha / \hat{c}_\beta} \frac{1}{m_{\alpha \rightarrow \beta}(x_\beta)} \]

- In the pairwise case
The Norm-Product

\[
m_{\alpha \rightarrow \beta}(x_\beta) = \left| \psi_\alpha(x_\alpha) \prod_{c \in C(\alpha) \setminus \beta} n_{c \rightarrow \alpha}(x_c) \prod_{p \in P(\alpha)} n_{c \rightarrow p}^{-1}(x_\alpha) \right|^{1/c_\alpha}
\]

\[
n_{\beta \rightarrow \alpha}(x_\beta) = \left( \psi_\beta(x_\beta) \prod_{p \in P(\beta)} m_{p \rightarrow \beta}(x_\beta) \prod_{c \in C(\beta)} n_{c \rightarrow \beta}(x_c) \right)^{c_\alpha / c_\beta} / m_{\alpha \rightarrow \beta}(x_\beta)
\]

- In the pairwise case
- In the MAP (\(\epsilon = 0\)) the entropy weight \(c_\alpha\) counts
convex and non-convex belief propagation. Parent message depends on the $(1/\epsilon c_{\alpha})$-norm

convex belief propagation for some $c_{\alpha} \leq 0$. New messages over “entropy graph”.
convex and non-convex belief propagation. Parent message depends on the $(1/\epsilon c_\alpha)$-norm.

- $c_\alpha = 0$
- $c_\alpha = 1$

New messages over "entropy graph".

convex belief propagation for some $c_\alpha \leq 0$. 
Tree Based Upper Bounds

\[
\text{log-partition} \sum_{x_\alpha \setminus x_\beta} \max_{b_\alpha(x_\alpha) = b_\beta(x_\beta)} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_\alpha c_\alpha H(b_\alpha)
\]
Tree Based Upper Bounds

\[
\log\text{-partition} \sum_{x_\alpha \in x_\beta} \max \sum_{\alpha, x_\alpha} b_{\alpha}(x_\alpha) \theta_{\alpha}(x_\alpha) + \sum_{\alpha} c_{\alpha} H(b_{\alpha})
\]

\[
c_{2,3,4,5} = 1
c_{2,3,4} = 0
c_{3,5} = 0
c_{3} = 0
\]

\[
c_{1,2,3} = 1
c_{2,3} = 0
\]
Tree Based Upper Bounds

log-partition
\[
\sum_{x_\alpha \setminus x_\beta} \max b_\alpha(x_\alpha) = b_\beta(x_\beta) \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha)
\]

\[
c_{2,3,4,5} = 1
c_{2,3,4} = 0
c_{3,5} = 0
c_3 = 0
c_{1,2,3} = 1
c_{2,3} = 0
\]

Wainwright et al: Bethe entropy \( c_\alpha = 1 - |P(\alpha)| \) on spanning-tree upper bounds the log-partition.
Wainwright et al: Bethe entropy $c_{\alpha} = 1 - |P(\alpha)|$ on spanning-tree upper bounds the log-partition.
Tree Based Upper Bounds

\[
\log\text{-}partition \quad \sum_{x_\alpha \setminus x_\beta} \max_{b_\alpha(x_\alpha) = b_\beta(x_\beta)} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha)
\]

\[
c_{2,3,4,5} = 1,
c_{3,5} = 0,
c_3 = 0,
c_{1,2,3} = 1,
c_{2,3} = -1
\]

- Wainwright et al: Bethe entropy \( c_\alpha = 1 - |P(\alpha)| \) on spanning-tree upper bounds the log-partition.
Tree Based Upper Bounds

\[
\log\text{-partition} \quad \max_{x_\alpha \setminus x_\beta} \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_\alpha c_\alpha H(b_\alpha)
\]

\[
c_{2,3,4,5} = 1
\]

\[
c_{2,3,4} = 0
\]

\[
c_{2,3} = -1
\]

\[
c_{3} = 0
\]

\[
c_{3,5} = 0
\]

\[
c_{1,2,3} = 1
\]

\text{Wainwright et al: Bethe entropy } c_\alpha = 1 - |P(\alpha)| \text{ on spanning-tree upper bounds the log-partition.}

\text{Mixed signs but concave (over the purple constraints).}
Tree Based Upper Bounds

log-partition

\[
\sum_{x_\alpha \setminus x_\beta} \max b_\alpha(x_\alpha) = b_\beta(x_\beta) \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha)
\]

- Wainwright et al: Bethe entropy \( c_\alpha = 1 - |P(\alpha)| \) on spanning-tree upper bounds the log-partition.
- Mixed signs but concave (over the purple constraints).

Pakzad & Anantharam, Heskes

\[
H(b_{1,2,3}) + (H(b_{2,3,4,5}) - H(b_{2,3}))
\]

conditional entropy is concave

\[c_2,3,4,5 = 1\]
\[c_3,5 = 0\]
\[c_3 = 0\]
\[c_1,2,3 = 1\]
\[c_{2,3} = -1\]
\[c_{2,3,4} = 0\]
Tree Based Upper Bounds

$c_{2,3,4,5} = 1$

$2,3,4,5$

$3,5$

$c_{3,5} = 0$

$3$

$c_3 = 0$

$1,2,3$

$c_{1,2,3} = 1$

$2,3$

$c_{2,3} = -1$

$2,3,4$

$c_{2,3,4} = 0$
Tree Based Upper Bounds

\[ c_{2,3,4,5} = 1 \]

\[ c_{2,3,4} = 0 \]

consistency messages
Tree Based Upper Bounds

$c_{2,3,4,5} = 1$

$c_{3,5} = 0$
$c_3 = 0$

$c_{1,2,3} = 1$
$c_{2,3} = -1$

$c_{2,3,4} = 0$

consistency messages

conditional entropy messages

Saturday, December 11, 2010
Duality

\[
\max \quad \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\alpha, \beta} c_{\alpha, \beta} (H(b_\alpha) - H(b_\beta))
\]

s.t. \[ b_\alpha(x_\alpha) \text{ is probability} \]
\[
\sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) = b_\beta(x_\beta)
\]
\[
\sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) = b_\beta(x_\beta)
\]
\[
b_\alpha(x_\alpha) = b_\alpha(x_\alpha)
\]
Duality

\[
\max \sum_{\alpha, x_\alpha} b_\alpha(x_\alpha) \theta_\alpha(x_\alpha) + \sum_{\alpha} c_\alpha H(b_\alpha) + \sum_{\alpha, \beta} c_{\alpha, \beta} (H(b_\alpha) - H(b_\beta))
\]

s.t. \( b_\alpha(x_\alpha) \) is probability

\[
\sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) = b_\beta(x_\beta)
\]

\[
\sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) = b_\beta(x_\beta)
\]

- Dual function

\[
q() = \max \; L(b_\alpha(x_\alpha), b_\alpha(x_\alpha), \lambda_{\beta \rightarrow \alpha}(x_\beta), \nu_{\beta \rightarrow \alpha}(x_\alpha))
\]

\[
b_\alpha(x_\alpha) \; b_\alpha(x_\alpha) \; \text{is probability}
\]

\[
\sum_{x_\alpha \setminus x_\beta} b_\alpha(x_\alpha) = b_\beta(x_\beta)
\]
The Norm-Product
The Norm-Product

2, 3, 4, 5

2, 3, 4

2, 3

1, 2, 3
The Norm-Product

\[ m_{\alpha \rightarrow \beta}(x_\beta) = \left\| \psi_\alpha(x_\alpha) \prod_{c \in C(\alpha) \setminus \beta} (n_{c \rightarrow \alpha}(x_c) \cdot \hat{n}_{c \rightarrow \alpha}(x_\alpha)) \prod_{p \in P(\alpha)} n_{\alpha \rightarrow p}^{-1}(x_\alpha) \right\|_{1/\epsilon(c_\alpha + c_\alpha, \beta)} \]
The Norm-Product

\[ n_{\beta \rightarrow \alpha}(x_{\beta}) = \left( \psi_{\beta}(x_{\beta}) \prod_{p \in P(\beta)} m_{p \rightarrow \beta}(x_{\beta}) \prod_{c \in C(\beta)} (n_{c \rightarrow \beta}(x_{c}) \cdot \hat{n}_{c \rightarrow \beta}(x_{\beta})) \right)^{c_{\alpha} / \hat{c}_{\beta}} / m_{\alpha \rightarrow \beta}(x_{\beta}) \]
The Norm-Product

\[ \hat{n}_{\beta \rightarrow \alpha}(x_\beta) = \left( \frac{m_{\alpha \rightarrow \beta}(x_\beta)}{\psi_\alpha(x_\alpha)} \prod_{c \in C(\alpha) \setminus \beta} (n_{c \rightarrow \alpha}(x_c) \cdot \hat{n}_{c \rightarrow \beta}(x_\alpha)) \prod_{p \in P(\alpha)} n_{\alpha \rightarrow p}^{-1}(x_\alpha) \right)^{c_{\alpha, \beta}/(c_\alpha + c_{\alpha, \beta})} \]
The Norm-Product

Non-convex

Norm-product
\[ \epsilon, c_\alpha, c_\alpha, \beta \]

convex
The Norm-Product

$\epsilon = 1, c_\alpha = 1 - |P(\alpha)|, c_{\alpha,\beta} = 0$

sum-product

Norm-product $\epsilon, c_\alpha, c_\alpha, \beta$

non-convex

convex
The Norm-Product

$\epsilon = 1, c_\alpha = 1 - |P(\alpha)|, c_{\alpha,\beta} = 0$

**Sum-product**

$\epsilon = 0, c_\alpha = 1 - |P(\alpha)|, c_{\alpha,\beta} = 0$

**Max-product**

**Norm-product**

$\epsilon, C_\alpha, C_\beta, \beta$

**Convex**

**Non-convex**
The Norm-Product

$\epsilon = 1, c_\alpha = 1 - |P(\alpha)|, c_\alpha, \beta = 0$

sum-product

TRBP $c_\alpha \leq 0$

Norm-product $\epsilon, c_\alpha, c_\alpha, \beta$

$\epsilon = 0, c_\alpha = 1 - |P(\alpha)|, c_\alpha, \beta = 0$

max-product

non-convex

convex
The Norm-Product

\[ \epsilon = 1, c_\alpha = 1 - |P(\alpha)|, c_\alpha, \beta = 0 \]

sum-product

\[ \epsilon, c_\alpha, c_\alpha, \beta \geq 0 \]

TRBP

\[ c_\alpha \leq 0 \]

Norm-product \[ \epsilon, c_\alpha, c_\alpha, \beta \]

\[ \epsilon = 0, c_\alpha = 1 - |P(\alpha)|, c_\alpha, \beta = 0 \]

max-product

convex

non-convex

convergent TRBP
The Norm-Product

\[ \epsilon = 1, c_\alpha = 1 - |P(\alpha)|, c_{\alpha,\beta} = 0 \]

sum-product

\[ \epsilon = 0, c_\alpha = 1 - |P(\alpha)|, c_{\alpha,\beta} = 0 \]

max-product

TRBP \( c_\alpha \leq 0 \)

\[ \epsilon, c_\alpha, c_{\alpha,\beta} \geq 0 \]

convergent TRBP

Heskes, TCBO, Convex-sum-product

\[ \epsilon = 1 \]
The Norm-Product

\[ \epsilon = 1, c_\alpha = 1 - |P(\alpha)|, c_{\alpha,\beta} = 0 \]

**sum-product**

**TRBP**

\[ c_\alpha \leq 0 \]

**convergent TRBP**

\[ \epsilon, c_\alpha, c_{\alpha,\beta} \geq 0 \]

Heskes, TCBO, Convex-sum-product

\[ \epsilon = 1 \]

**Norm-product**

\[ \epsilon, c_\alpha, c_{\alpha}, c_{\alpha,\beta} \]

**max-product**

\[ \epsilon = 0, c_\alpha = 1 - |P(\alpha)|, c_{\alpha,\beta} = 0 \]

MPLP, TCBO, Convex-max-product

\[ \epsilon = 0 \]

**non-convex**

**convex**
The Norm-Product

\[ \epsilon = 1, \ c_\alpha = 1 - |P(\alpha)|, \ c_{\alpha,\beta} = 0 \]

sum-product

**TRBP**

\[ c_\alpha \leq 0 \]

\[ \epsilon, \ c_\alpha, \ c_{\alpha,\beta} \geq 0 \]

convergent TRBP

\[ \epsilon = 1 \]

Heskes, TCBO, Convex-sum-product

\[ \epsilon \rightarrow 0 \]

Heskes, TCBO, LP-solver

\[ \epsilon = 0 \]

MPLP, TCBO, Convex-max-product

max-product

TRBP

\[ c_\alpha \leq 0 \]

non-convex

convex

Saturday, December 11, 2010
Open Problems
Open Problems

- MAP
Open Problems

MAP

1) How the entropy weights can be used?
Open Problems

1) How the entropy weights can be used?

2) Rounding schemes (Grothendieck constant?)
Open Problems

1) How the entropy weights can be used?
2) Rounding schemes (Grothendieck constant?)

- log-partition

MAP
Open Problems

- MAP

1) How the entropy weights can be used?
2) Rounding schemes (Grothendieck constant?)

- log-partition

1) Extending tree-bounds to regions (Submodular bounds?)
Open Problems

- MAP
  1) How the entropy weights can be used?
  2) Rounding schemes (Grothendieck constant?)
- log-partition
  1) Extending tree-bounds to regions (Submodular bounds?)
  2) How to set the entropy weights for good approximations?
Open Problems

MAP

1) How the entropy weights can be used?
2) Rounding schemes (Grothendieck constant?)

log-partition

1) Extending tree-bounds to regions (Submodular bounds?)
2) How to set the entropy weights for good approximations?

soft-max?
Open Problems

• MAP

1) How the entropy weights can be used?

2) Rounding schemes (Grothendieck constant?)

• log-partition

1) Extending tree-bounds to regions (Submodular bounds?)

2) How to set the entropy weights for good approximations?

• soft-max?

Thank you