Reasoning and Decision in Probabilistic Graphical models – A Unified Approach

Qiang Liu

Thesis Defense

Dissertation Committee :

Alexander Ihler (chair)

Rina Dechter

Padhraic Symth

Outline

- Background:
  - Probabilistic graphical models
  - Inference & decision tasks
- Main Results:
  - A unified variational representation
  - Derive lots of efficient approximation algorithms
  - Experiments
- Conclusion & Future Directions

Probabilistic Graphical Models

- High dimensional probability:

$$p(\boldsymbol{x}) = p(x_1, x_2, \ldots, x_n) = \frac{1}{Z} \prod_k \psi(x_{\alpha_k})$$

e.g., $x \in \{0,1\}^n$ $\qquad$ Z: normalization constant

– Example: $p(x) = \frac{1}{Z} \psi_1(x_1) \psi_3(x_3) \psi_{12}(x_1, x_2) \psi_{23}(x_2, x_3)$

	0	1
X1	9	1

X1 \ X2	0	1
0	8	3
1	1	9

Probabilistic Graphical Models

- High dimensional probability:

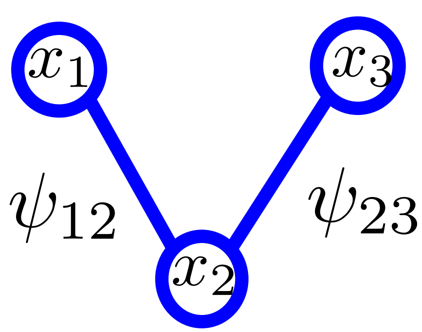
$$p(\boldsymbol{x}) = p(x_1, x_2, \ldots, x_n) = \frac{1}{Z} \prod_k \psi(x_{\alpha_k})$$

e.g., $x \in \{0, 1\}^n$

Z: normalization constant

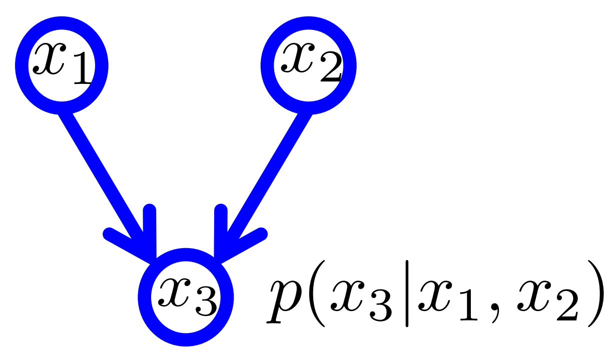
Undirected graph:

(Markov random field)



Directed graph:

(Bayesian Network)

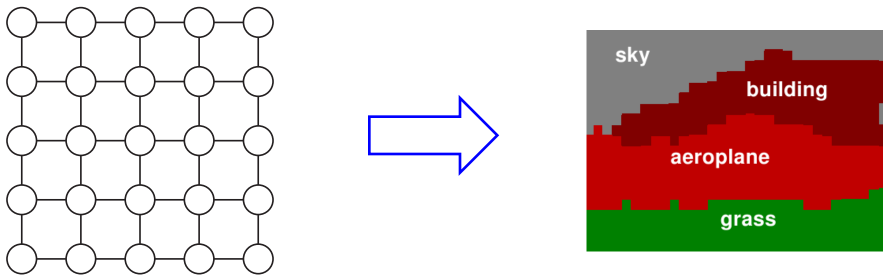


Key Inference Types

- **Optimization (max)**
- Marginalization (sum)
- Marginal MAP (max-sum)
- Decision Making

- Most probable configuration (Maximum *a posteriori* (MAP) estimate):

$$x^* = \arg\max_x p(x) = \arg\max_x \prod_k \psi_k(x_{\alpha_k})$$
$$x \in \{0,1\}^n$$

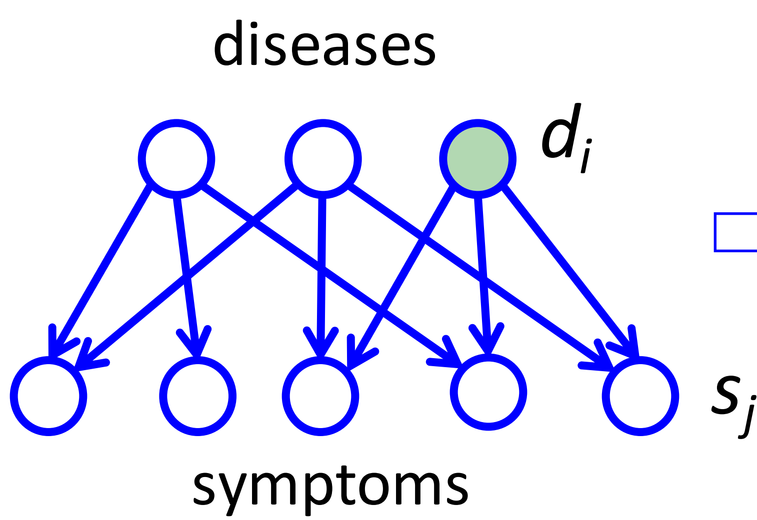


Key Inference Types

- Optimization (max)
- Marginalization (sum)
- Marginal MAP (max-sum)
- Decision Making

- Normalization constant (a.k.a. partition function), or marginal probabilities:

$$Z = \sum_x \prod_k \psi_k(x_{\alpha_k}) \quad \text{or} \quad p(x_i) = \frac{1}{Z} \sum_{x \setminus x_i} \prod_k \psi_k(x_{\alpha_k})$$



$$\Rightarrow \quad p(d_i | s)$$

Key Inference Types

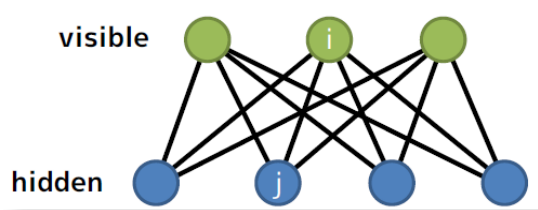
- Optimization (max)
- Marginalization (sum)
- Marginal MAP (max-sum)
- Decision Making

- Maximize marginal probability, or expected objective w.r.t. “latent variables”:

$$x_B^* = \arg\max_{x_B} \sum_{x_A} p([x_A, x_B]) = \arg\max_{x_B} \sum_{x_A} \prod_k \psi_k(x_{\alpha_k})$$

$$x = [x_A, x_B]$$

- Latent variables:



- Missing labels:

Black: missing labels

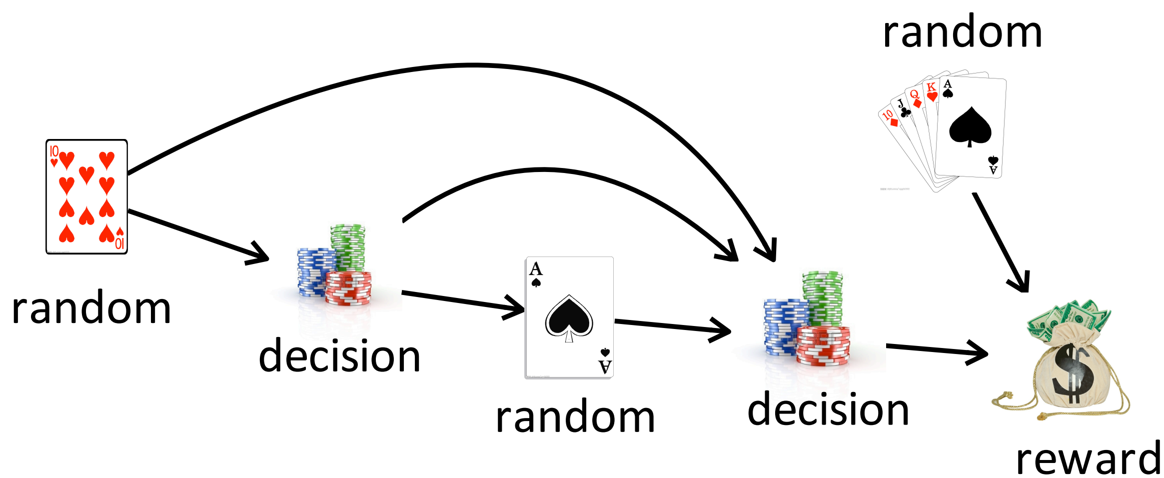
Key Inference Types

- Optimization (max)
- Marginalization (sum)
- Marginal MAP (max-sum)
- **Decision Making**

- Sequential Decision under uncertainty (stochastic programming):

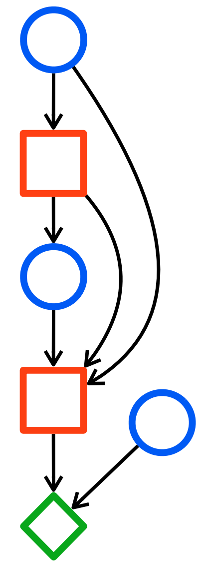
$$\max_{x_{D_1}} \sum_{x_{R_1}} \cdots \max_{x_{D_k}} \sum_{x_{R_k}} \prod_k \psi_k(x_{\alpha_k})$$

  - Sum: average random variables
  - Max: optimize decision variables



Key Inference Types

- Optimization (max)
- Marginalization (sum)
- Marginal MAP (max-sum)
- Decision Making



- Represented as influence diagrams (a.k.a. decision networks):

○: Random variables x_R

□: Decision variables x_D

◇: Utility function: $U([x_R, x_D])$

Maximize expected utility (MEU):

$$\max_{\delta_D} \mathbb{E}(U(x) \mid \delta_D), \quad \delta_D \text{ : decision policies}$$

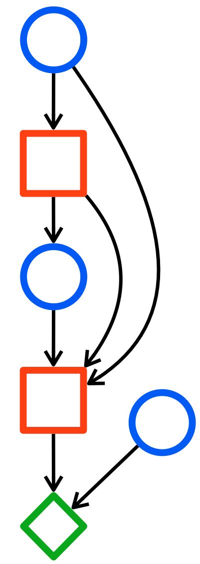
$$= \max_{x_{D_1}} \sum_{x_{R_1}} \cdots \max_{x_{D_k}} \sum_{x_{R_k}} U(x) p(x_R | x_D)$$

(Assume: single agent; perfect memory)

Key Inference Types

- Optimization (max)
- Marginalization (sum)
- Marginal MAP (max-sum)
- Decision Making

- Represented as influence diagrams (a.k.a. decision networks):



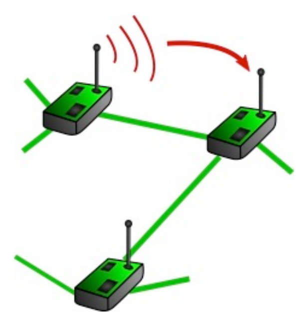
○: Random variables x_R

□: Decision variables x_D

◇: Utility function: $U([x_R, x_D])$

Maximize expected utility (MEU):

$$\max_{\delta_D} \mathbb{E}(U(x) \mid \delta_D), \quad \delta_D \text{ : decision policies}$$



(Multi-agent; limited communication: no sequential elimination form)

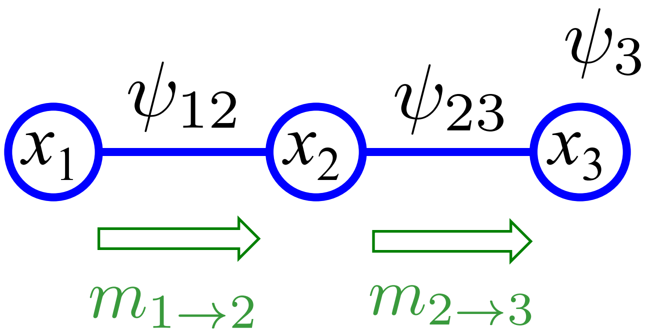
Key Inference Types

Maximization	$\max_x \prod_k \psi_k(x_{\alpha_k})$
Sum/Counting	$\sum_x \prod_k \psi_k(x_{\alpha_k})$
Max-Sum	$\max_{x_B} \sum_{x_A} \prod_k \psi_k(x_{\alpha_k})$
Decision Making (sequential)	$\max_{x_{B_1}} \sum_{x_{A_1}} \cdots \max_{x_{B_k}} \sum_{x_{A_k}} \prod_k \psi_k(x_{\alpha_k})$
Decision Making (multi-agent)	 No sequential elimination form

Difficult!

Pure sum or max are (relatively) "easy"

- Dynamic programming.



"message passing"

$$Z = \sum_{x_1,x_2,x_3} \psi_3(x_3)\psi_{23}(x_2,x_3)\psi_{12}(x_1,x_2)$$

$$= \sum_{x_2,x_3} \psi_3(x_3)\psi_{23}(x_2,x_3)\underbrace{\sum_{x_1}\psi_{12}(x_1,x_2)}$$

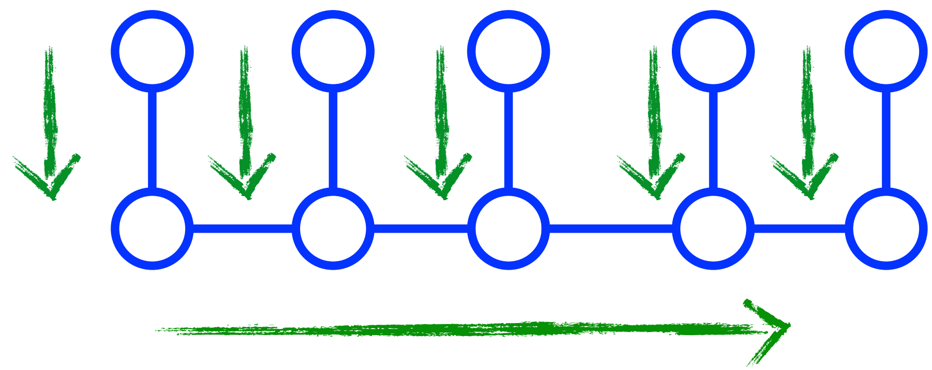
$$= \sum_{x_2,x_3} \psi_3(x_3)\psi_{23}(x_2,x_3)m_{1\to 2}(x_2)$$

$$= \sum_{x_3} \psi_3(x_3)\underbrace{\sum_{x_2}\psi_{23}(x_2,x_3)m_{1\to 2}(x_2)}$$

$$= \sum_{x_3} \psi_3(x_3)m_{2\to 3}(x_3)$$

Pure sum or max are (relatively) easy!

- On general trees
  - Linear complexity



  - Sum or max-product Belief Propagation (BP):

$$m_{i\to j}(x_j) \propto \begin{array}{c} \max\limits_{x} \\ \text{or} \\ \sum\limits_{x} \end{array} \psi_{ij}(x_i, x_j) \prod_{k \neq j} m_{k\to i}(x_i)$$

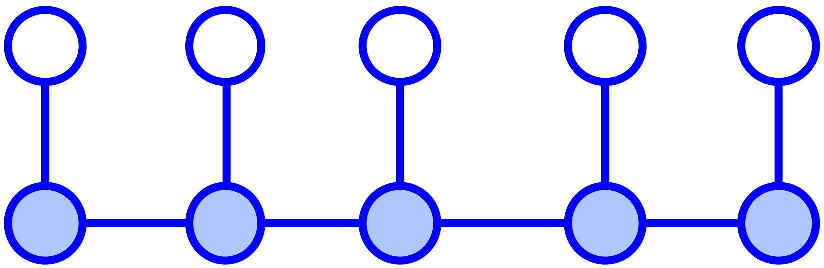
"Message passing"

Marginal MAP & Decision are challenging!

- NP-hard even on trees:

max: x_B

sum: x_A



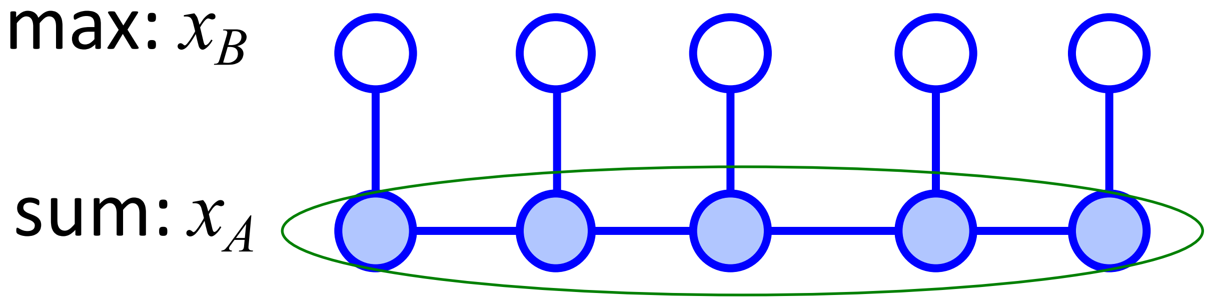
Koller & Friedman 09

$$\max_{x_B} \sum_{x_A} p([x_A, x_B])$$

  - Can not exchange max and sum ($\max \sum \neq \sum \max$)

Marginal MAP & Decision are challenging!

- NP-hard even on trees:



$$\max_{x_B} \sum_{x_A} p([x_A, x_B])$$

$$= \max_{x_B} \hat{\psi}(x_B)$$

- Can not exchange max and sum ($\max \sum \neq \sum \max$)

Use Joint Optimization instead?

$$[x_A^*, x_B^*] = \arg\max_x p([x_A, x_B])$$

Use Joint Optimization instead?

$$[x_A^*, {\color{blue}x_B^*}] = \arg\max_x p([x_A, x_B])$$

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

$p(x, y)$

	walk	drive
rainy	0.05	0.35
sunny	0.3	0.3

Use Joint Optimization instead?

$$[x_A^*, x_B^*] = \arg\max_x p([x_A, x_B])$$

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

$p(x, y)$

	walk	drive
rainy	0.05	0.35
sunny	0.3	0.3

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

$\Rightarrow x^* = \text{sunny}$ ✓

▶ Paradox: My behavior influence the weather!

Use Joint Optimization instead?

$$[x_A^*, x_B^*] = \arg\max_x p([x_A, x_B])$$

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

$p(x, y)$

	walk	drive
rainy	0.05	0.35
sunny	0.3	0.3

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

$\Rightarrow x^* = \text{sunny}$ ✓

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

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

▶ Paradox: My behavior influence the weather!

Two Types of Approximate Methods

- Elimination-based approximation:
  - Directly approximate the sum & max operators
  - Mini-bucket, weighted mini-bucket (ICML11)

- Variational approximation:
  - Reframe into continuous (hopefully linear or convex) optimization
  - Approximating the continuous optimization

Elimination form → continuous optimization (hopefully linear or convex) → All kinds of approximations

Outline

- Background:
  - Probabilistic graphical models
  - Inference & decision tasks

- **Main Results:**
  - A unified variational representation
  - Derive lots of efficient approximation algorithms
  - Experiments

- Conclusion & Future Directions

Variational Algorithms

- Reframe into continuous (hopefully linear or convex) optimization
  - Notation:

$$p(x) = \frac{1}{Z} \prod_k \psi_k(x_{\alpha_k}) = \frac{1}{Z} \exp(\theta(x))$$

$$\Longrightarrow \quad \theta(x) = \log \prod_k \psi_k(x_{\alpha_k})$$

Variational Algorithms

- Reframe into continuous (hopefully linear or convex) optimization

$$\boxed{\theta(x) = \log \prod_k \psi_k(x_{\alpha_k})}$$

  - Maximization

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

$\mathbb{M}$: The set of joint distributions on $\boldsymbol{x}$:

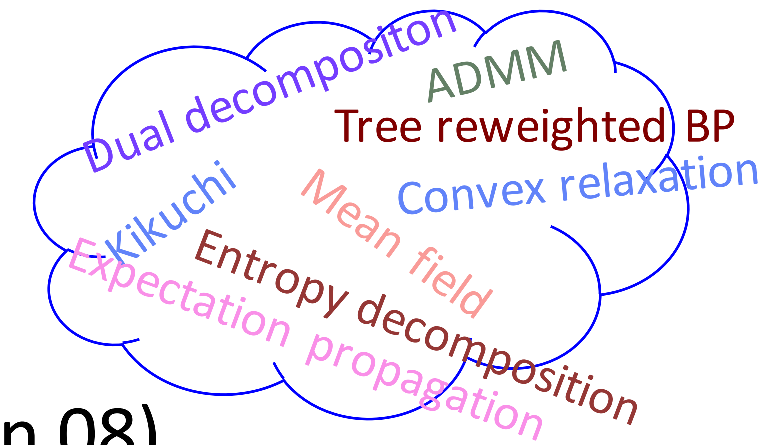
  - Summation:

$$\log \sum_x \exp(\theta(\boldsymbol{x})) = \max_{q \in \mathbb{M}} \{\mathbb{E}_q[\theta(\boldsymbol{x})] + \underline{H(\boldsymbol{x}\ ;\ q)}\} \quad \text{(maximum: } q = p)$$

Entropy: $-\sum_x q(x) \log q(x)$

  - $\mathbb{M}$ and $H(x\ ;\ q)$ are intractable!

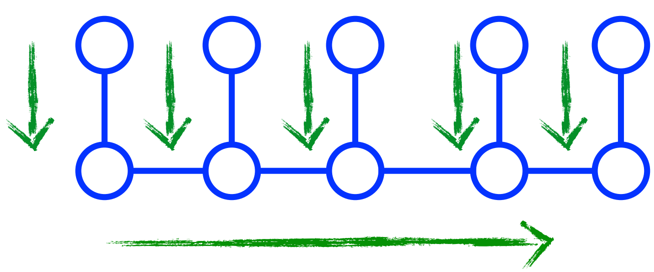
Variational Approximation



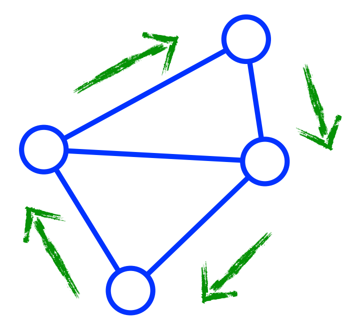
- Approximate $\mathbb{M}$ and $H(x\,;\,q)$
- Derives lots of "message passing" algorithms (e.g., Wainwright & Jordan 08)
  – e.g., Bethe approximation yields loopy BP (Yedidia et al 2000)

$$m_{i\to j}(x_j) \propto \begin{matrix} \max_x \\ \sum_x \end{matrix} \psi_{ij}(x_i, x_j) \prod_{k \neq j} m_{k\to i}(x_i)$$

– *Exact on trees*



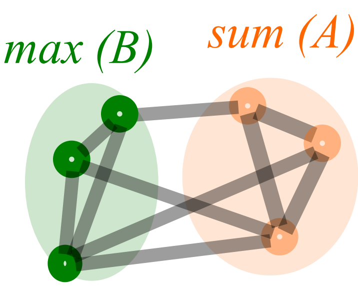
– *Approximate on loopy graphs (Pearl 88)*



	Primal Form	Dual Form
Max	$\max_{\boldsymbol{x}} \exp(\theta(\boldsymbol{x}))$	$\max_{q \in \mathbb{M}} \mathbb{E}_q[\theta(\boldsymbol{x})]$
Sum	$\sum_{\boldsymbol{x}} \exp(\theta(\boldsymbol{x}))$	$\max_{q \in \mathbb{M}} \{\mathbb{E}_q[\theta(\boldsymbol{x})] + H(\boldsymbol{x}; q)\}$
Max-sum	$\max_{\boldsymbol{x}_B} \sum_{\boldsymbol{x}_A} \exp(\theta(\boldsymbol{x}))$	
Decision Making (perfect recall)	$\max_{x_{B_1}} \sum_{x_{A_1}} \cdots \max_{x_{B_k}} \sum_{x_{A_k}} \exp(\theta(x))$	
Decision Making (imperfect recall)		

Variational Representations (UAI11, JMLR13)

– For marginal MAP: $Z_{AB} = \max_{x_B} \sum_{x_A} \exp(\theta(\boldsymbol{x}))$



$$\log Z_{AB} = \max_{q \in \mathbb{M}} \left\{ \mathbb{E}_q[\theta(\boldsymbol{x})] + H(\boldsymbol{x}_A | \boldsymbol{x}_B \; ; \; q) \right\}$$

$$= \max_{q \in \mathbb{M}} \left\{ \mathbb{E}_q[\theta(\boldsymbol{x})] + H(\boldsymbol{x} \; ; \; q) - H(\boldsymbol{x}_B \; ; \; q) \right\}$$

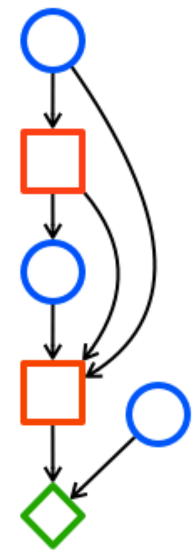
Same as fully summation

"truncating" the entropies of the max nodes; enforcing deterministic choice

Joint max $(A = \emptyset)$ $\Rightarrow$ $\max_{q \in \mathbb{M}} \{ \mathbb{E}_q[\theta(\boldsymbol{x})] \}$

Joint sum $(B = \emptyset)$ $\Rightarrow$ $\max_{q \in \mathbb{M}} \{ \mathbb{E}_q[\theta(\boldsymbol{x})] + H(\boldsymbol{x} \; ; \; q) \}$

Variational Representations (UAI12)



– Sequential decision making:

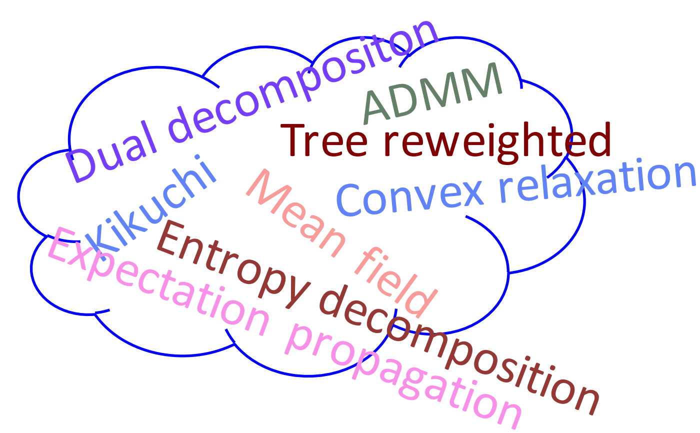
$$Z = \max_{D_1} \sum_{R_1} \cdots \max_{D_k} \sum_{R_k} \exp(\theta(\boldsymbol{x}))$$

$$\log Z = \max_{q \in \mathbb{M}} \{ \mathbb{E}_q[\theta(\boldsymbol{x})] + H(\boldsymbol{x} \; ; \; q) - \sum_i H(x_{D_i} | x_{\mathrm{pa}(D_i)} \; ; \; q) \}$$

Same as fully summation

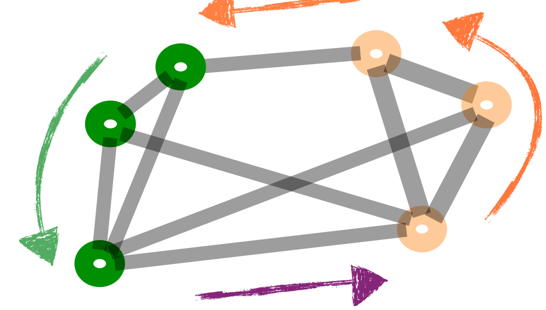
“truncating” conditional entropies of decision nodes

– Directly extends to more general (multi-agent) decision making (details in UAI12)

	Primal Form	Dual Form
Max	$\max_{\boldsymbol{x}} \exp(\theta(\boldsymbol{x}))$	
Sum	$\sum_{\boldsymbol{x}} \exp(\theta(\boldsymbol{x}))$	
Max-sum	$\max_{\boldsymbol{x}_B} \sum_{\boldsymbol{x}_A} \exp(\theta(\boldsymbol{x}))$	$\max_{q\in\mathbb{M}} \{\mathbb{E}_q[\theta(\boldsymbol{x})] + H(\boldsymbol{x}\,;\,q) - \sum_i H(x_{A_i} \mid x_{\mathrm{pa}(A_i)}\,;\,q)\}$
Decision Making (sequential)	$\max_{x_{B_1}} \sum_{x_{A_1}} \cdots \max_{x_{B_k}} \sum_{x_{A_k}} \exp(\theta(x))$	Derive all kinds of algorithms:
Decision Making (multi-agent)		

"Mixed-product" BP (for marginal MAP)

- "Mixed-product" BP: sends different messages between different types of nodes



$Max \Rightarrow Max$: (Max-Product) $$m_{i\to j} \propto \max_{x_i} \psi_{ij} \prod_{k \neq j} m_{k\to i}$$

$Sum \Rightarrow Sum \cup Max$: (Sum-Product) $$m_{i\to j} \propto \sum_{x_i} \psi_{ij} \prod_{k \neq j} m_{k\to i}$$

$Max \Rightarrow Sum$: (Argmax-Product) $$m_{i\to j} \propto \sum_{x_i \in \mathcal{X}_i^*} \psi_{ij} \prod_{k \neq j} m_{k\to i}$$

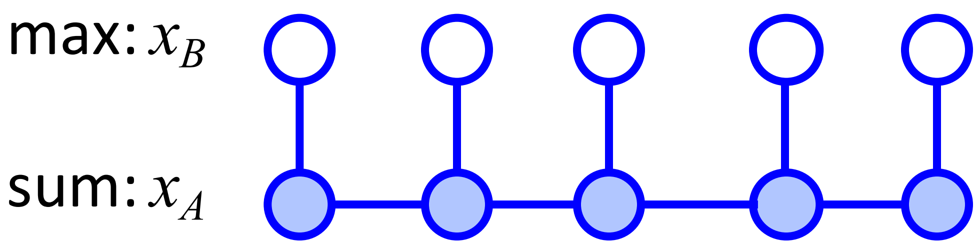
where $\mathcal{X}_i^* = \arg\max_{x_i} \psi_i \prod_k m_{k\to i}$.

- Intuition
  - Optimize x_i, pass the result to the sum part

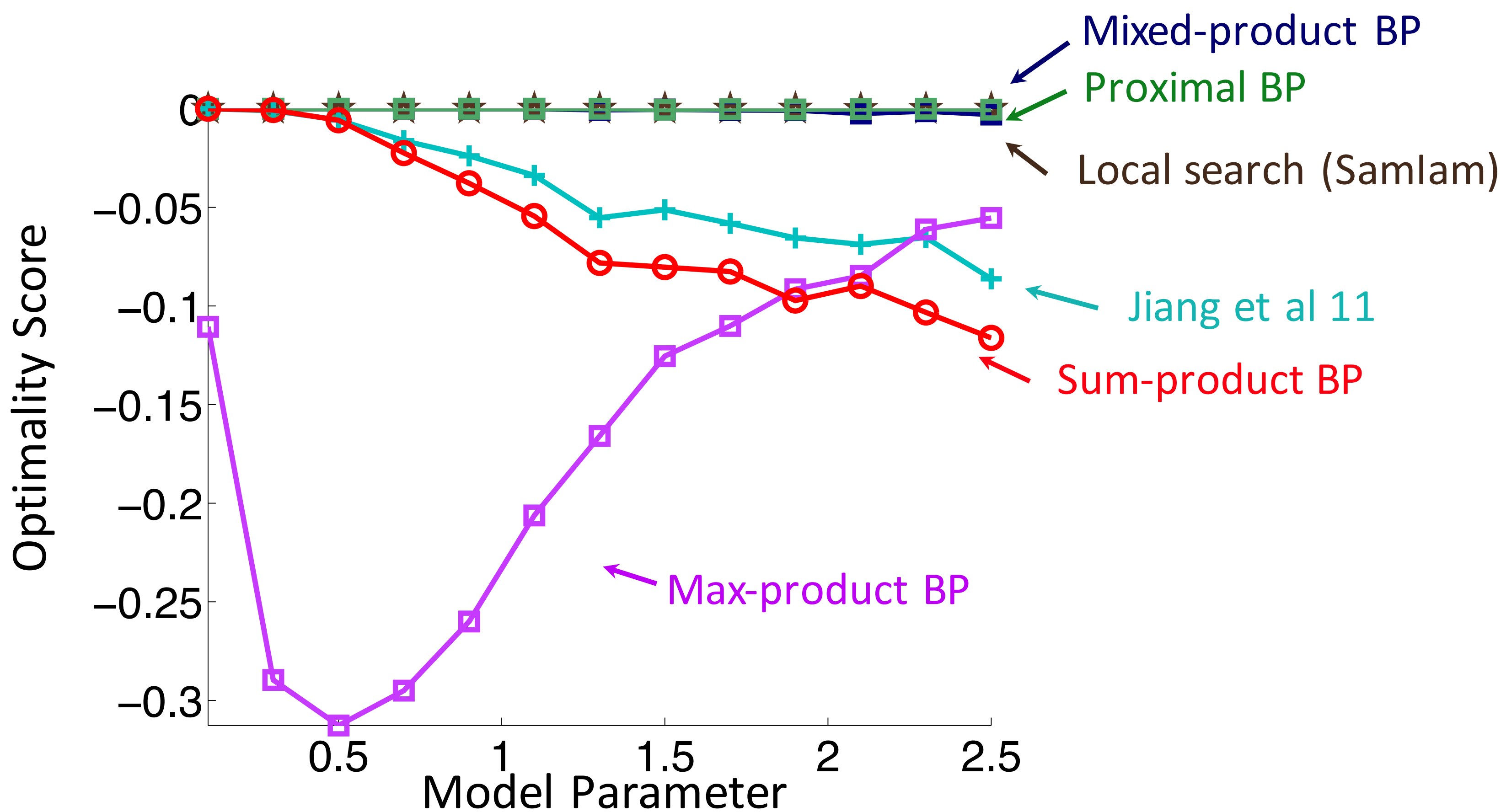
- Optimality guarantees (details in UAI11,UAI12, JMRL13)

Experiments (marginal MAP)

- Hidden Markov Chain (of length 10):



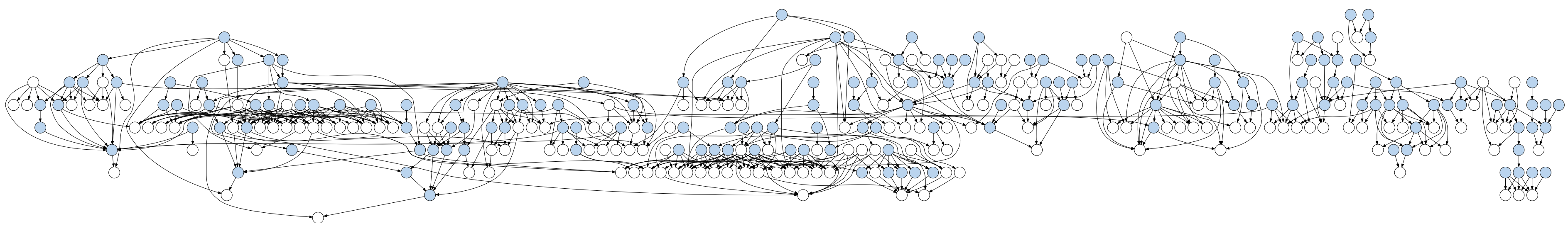
$$\max_{x_B} \sum_{x_A} p([x_A, x_B])$$

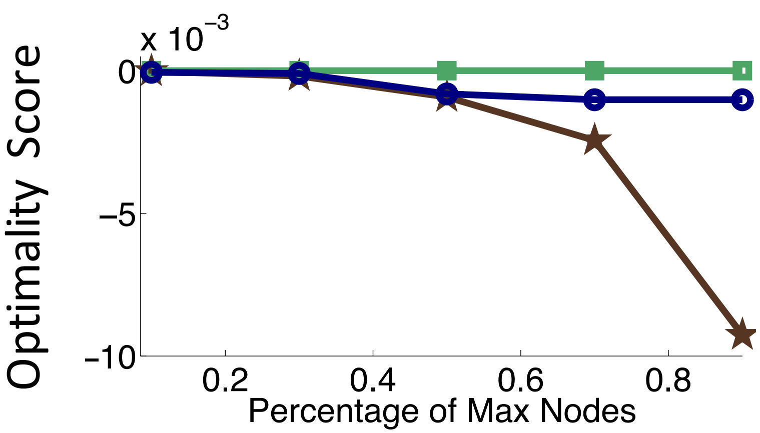


Experiments (marginal MAP)

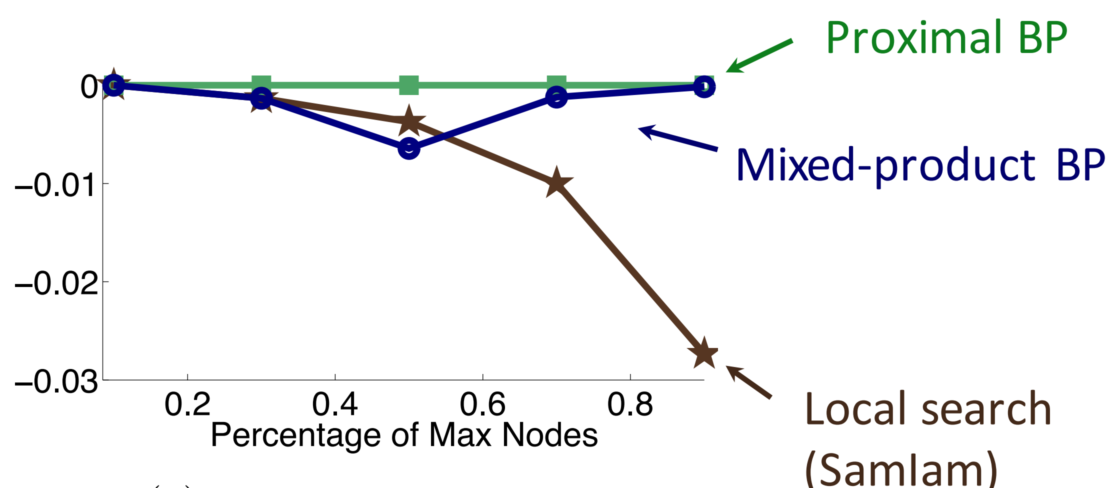
- Two diagnostic Bayesian networks from UAI08 challenge:
  - ~200-300 nodes, ~300-600 edges
  - Randomly select a set of max nodes

max (B): ○
sum (A): ●





(b). Diagnostic BN-1



(c). Diagnostic BN-2

Structured Prediction with Missing labels

(ICML14, experiments done by Wei Ping)

Building
Grass
Sky
Missing

Microsoft Research Cambridge Dataset (Winn et al. 05)

Accuracy (%):

MSRC Data	MSSVM	LSSVM	HCRFs
Building	**72.4**	70.7	71.7
Grass	**89.7**	88.9	88.3
Sky	**88.3**	85.6	88.2
Tree	**71.9**	71.0	70.1
Car	**70.8**	69.4	70.2

Marginal MAP

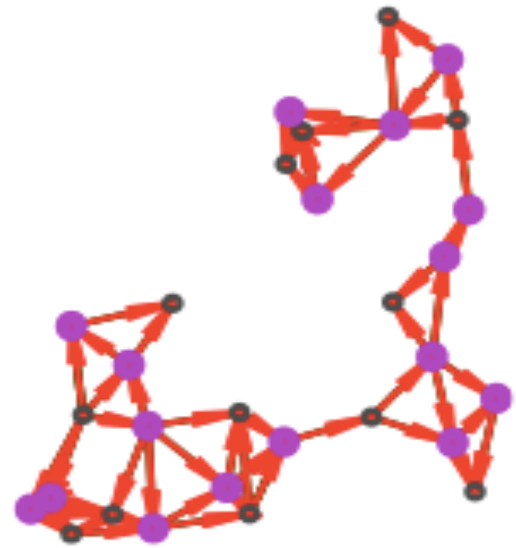
$$x = \arg\max_x \sum_y f(x, y)$$

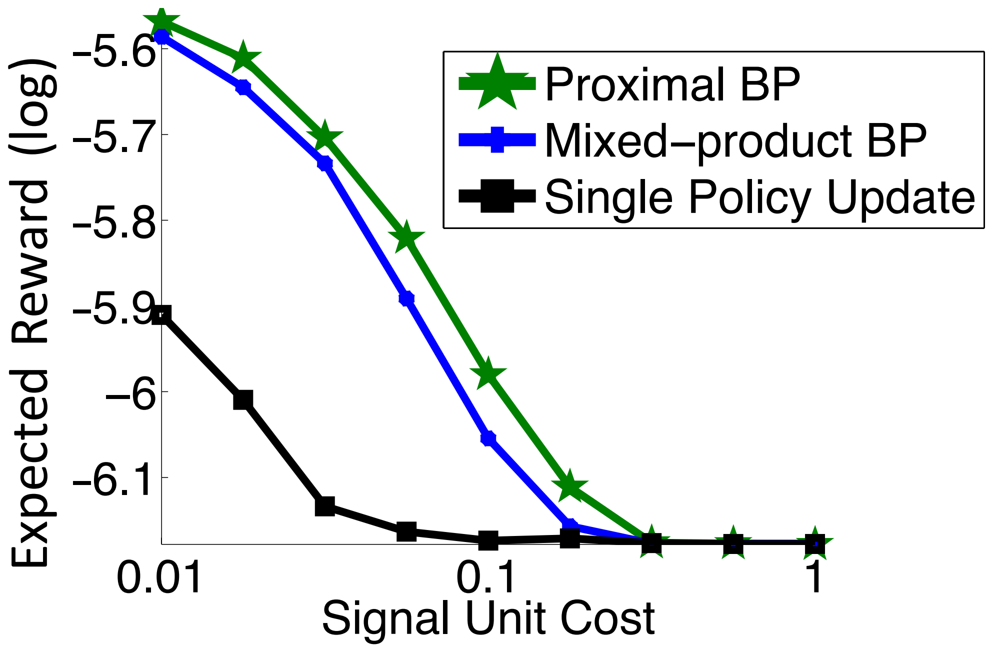
Joint MAP

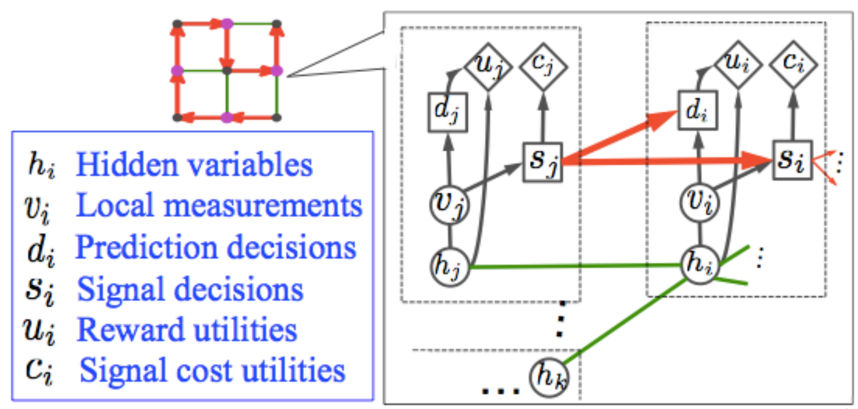
$$[x, y] = \arg\max_{x,y} f(x, y)$$

Collaborative Decision Making (UAI12)

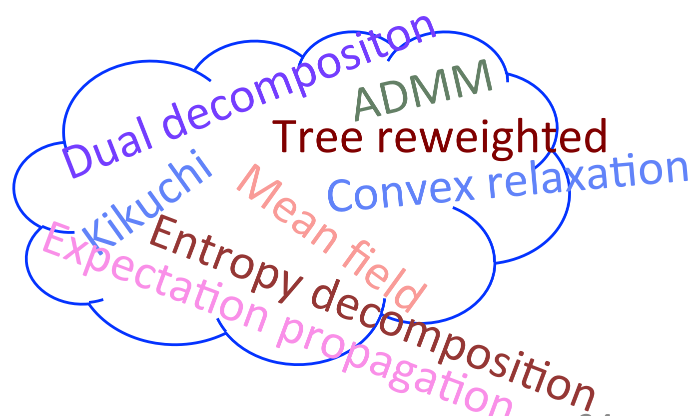
- Also derive message passing algorithms for decision
- Distributed detection:
  - Sensors make local predictions
  - Send signals to help other sensors
  - Reward = accuracy - signal cost







Conclusion

	Primal Form	Dual Form
Max	$\max_{\boldsymbol{x}} \exp(\theta(\boldsymbol{x}))$	$\max_{q \in \mathbb{M}} \{\mathbb{E}_q[\theta(\boldsymbol{x})] + H(\boldsymbol{x}\ ;\ q) - \sum_i H(x_{A_i} \mid x_{\mathrm{pa}(A_i)}\ ;\ q)\}$ Derive all kinds of algorithms: 
Sum	$\sum_{\boldsymbol{x}} \exp(\theta(\boldsymbol{x}))$	
Max-sum	$\max_{\boldsymbol{x}_B} \sum_{\boldsymbol{x}_A} \exp(\theta(\boldsymbol{x}))$	
Decision Making (sequential)	$\max_{x_{B_1}} \sum_{x_{A_1}} \cdots \max_{x_{B_k}} \sum_{x_{A_k}} \exp(\theta(x))$	
Decision Making (multi-agent)		

Thank you!