What is IJGP?

- IJGP is an approximate algorithm for belief updating in Bayesian networks.
- IJGP is a version of joint-tree clustering which is both 
  *anytime* and *iterative*.
- IJGP applies message passing along a joint-graph, rather than a joint-tree.
- Empirical evaluation shows that IJGP is almost always superior to other approximate schemes (IBP, MC).
Outline

- IBP - Iterative Belief Propagation
- MC - Mini Clustering
- IJGP - Iterative Join Graph Propagation
- Empirical evaluation
- Conclusion
Belief propagation is exact for poly-trees

IBP - applying BP iteratively to cyclic networks

No guarantees for convergence

Works well for many coding networks
Mini-Clustering - MC

Cluster Tree Elimination

1. $A \overset{B}{\bigcirc} C$
   - $p(a), p(b|a), p(c|a,b)$

2. $B \bigcirc D$
   - $p(d|b), h_{(1,2)}(b,c)$

3. $B \overset{E}{\bigcirc} F$
   - $p(e|b,f)$

4. $E \bigcirc F \overset{G}{\bigcirc}$
   - $p(g|e,f)$

$h_{(1,2)}(b,c) = \sum_a p(a) \cdot p(b|a) \cdot p(c|a,b)$

$h_{(1,2)}(b,c) = \sum_a p(a) \cdot p(b|a) \cdot p(c|a,b)$

$h_{(2,3)}(b,f) = \sum_{c,d} p(d|b) \cdot h_{(1,2)}^1(b,c) \cdot p(f|c,d)$

$h_{(2,3)}^1(b) = \sum_{c,d} p(d|b) \cdot h_{(1,2)}^1(b,c)$

$h_{(2,3)}^2(f) = \sum_{c,d} p(f|c,d)$
Cluster Tree Elimination

Mini-Clustering, i=3
IJGP - Motivation

- IBP is applied to a loopy network iteratively
  - not an anytime algorithm
  - when it converges, it converges very fast

- MC applies bounded inference along a tree decomposition
  - MC is an anytime algorithm controlled by i-bound

- IJGP combines:
  - the iterative feature of IBP
  - the anytime feature of MC
IJGP - The basic idea

- Apply Cluster Tree Elimination to any join-graph
- We commit to graphs that are minimal I-maps
- Avoid cycles as long as I-mapness is not violated
- Result: use *minimal arc-labeled* join-graphs
IJGP - Example

a) Belief network

a) The graph IBP works on
Arc-minimal join-graph
Minimal arc-labeled join-graph
Join-graph decompositions

a) Minimal arc-labeled join graph

b) Join-graph obtained by collapsing nodes of graph a)

c) Minimal arc-labeled join graph
a) Minimal arc-labeled join graph

a) Tree decomposition
Join-graphs

more accuracy

less complexity
Minimal arc-labeled decomposition

- Use a DFS algorithm to eliminate cycles relative to each variable

a) Fragment of an arc-labeled join-graph

a) Shrinking labels to make it a minimal arc-labeled join-graph
Minimal arc-labeled:
sep(1,2)={D, E}
elim(1,2)={A, B, C}

Non-minimal arc-labeled:
sep(1,2)={C, D, E}
elim(1,2)={A, B}

\[
h_{(1,2)}(de) = \sum_{a,b,c} p(a) p(c) p(b \mid ac) p(d \mid abe) p(e \mid bc) h_{(3,1)}(bc)
\]

\[
h_{(1,2)}(cde) = \sum_{a,b} p(a) p(c) p(b \mid ac) p(d \mid abe) p(e \mid bc) h_{(3,1)}(bc)
\]
Bounded decompositions

- We want arc-labeled decompositions such that:
  - the cluster size (internal width) is bounded by $i$ (the accuracy parameter)
  - the width of the decomposition as a graph (external width) is as small as possible

- Possible approaches to build decompositions:
  - partition-based algorithms - inspired by the mini-bucket decomposition
  - grouping-based algorithms
Partition-based algorithms

G: (GFE)
E: (EBF)\rightarrow (EF)
F: (FCD)\rightarrow (BF)
D: (DB)\rightarrow (CD)
C: (CAB)\rightarrow (CB)
B: (BA)\rightarrow (AB)\rightarrow (B)
A: \rightarrow (A)

a) schematic mini-bucket(i), i=3

b) arc-labeled join-graph decomposition
IJGP properties

- IJGP($i$) applies BP to min arc-labeled join-graph, whose cluster size is bounded by $i$

- On join-trees IJGP finds exact beliefs

- IJGP is a Generalized Belief Propagation algorithm (Yedidia, Freeman, Weiss 2001)

- Complexity of one iteration:
  - time: $O(deg \cdot (n+N) \cdot d^{i+1})$
  - space: $O(N \cdot d^{\theta})$
Empirical evaluation

- **Algorithms:**
  - Exact
  - IBP
  - MC
  - IJGP

- **Measures:**
  - Absolute error
  - Relative error
  - Kullback-Leibler (KL) distance
  - Bit Error Rate
  - Time

- **Networks (all variables are binary):**
  - Random networks
  - Grid networks (MxM)
  - CPCS 54, 360, 422
  - Coding networks
Random networks - KL at convergence

Random networks, N=50, K=2, P=3, evid=0, w*=16, 100 instances

KL distance

evidence=0

Random networks, N=50, K=2, P=3, evid=5, w*=16, 100 instances

KL distance

evidence=5
Random networks - KL vs. iterations

Random networks, N=50, K=2, P=3, evid=0, w*=16, 100 instances

Number of iterations
0 5 10 15 20 25 30 35
KL distance
1e-5
1e-4
1e-3
1e-2

IJGP(2)
IJGP(10)
IBP

Random networks, N=50, K=2, P=3, evid=5, w*=16, 100 instances

Number of iterations
0 5 10 15 20 25 30 35
KL distance
1e-5
1e-4
1e-3
1e-2
1e-1

IJGP(2)
IJGP(10)
IBP

evidence=0
evidence=5
Random networks - Time

Random networks, N=50, K=2, P=3, evid=5, w*=16, 100 instances

Time (seconds)

IJGP 20 it

MC

IBP 10 it

i-bound
Grid 81 – KL distance at convergence

Grid network, N=81, K=2, evid=5, w*=12, 100 instances

KL distance

IJGP
MC
IBP
Grid network, $N=81$, $K=2$, evid=5, $w^*=12$, 100 instances

Number of iterations

0 5 10 15 20 25

KL distance

$10^{-5}$
$10^{-4}$
$10^{-3}$
$10^{-2}$
$10^{-1}$

IJGP(2)
IJGP(8)
IBP
Grid 81 – Time vs. i-bound

Grid network, N=81, K=2, evid=5, w*=12, 100 instances

Time (seconds)

IJGP 10 it (at convergence)
MC
IBP 10 it (at convergence)
# Coding networks

<table>
<thead>
<tr>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>( \sigma )</th>
<th>IBP</th>
</tr>
</thead>
<tbody>
<tr>
<td>IJGP</td>
<td>0.00005</td>
<td>0.00005</td>
<td>0.00005</td>
<td>0.00005</td>
<td>0.00005</td>
<td>0.00005</td>
<td>0.00005</td>
<td>0.00005</td>
<td>0.00005</td>
</tr>
<tr>
<td>MC</td>
<td>0.00501</td>
<td>0.00800</td>
<td>0.00586</td>
<td>0.00462</td>
<td>0.00392</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IJGP</td>
<td>0.00062</td>
<td>0.00062</td>
<td>0.00062</td>
<td>0.00062</td>
<td>0.00062</td>
<td>0.00062</td>
<td>0.00062</td>
<td>0.00062</td>
<td>0.00064</td>
</tr>
<tr>
<td>MC</td>
<td>0.02170</td>
<td>0.02968</td>
<td>0.02492</td>
<td>0.02048</td>
<td>0.01840</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IJGP</td>
<td>0.00238</td>
<td>0.00238</td>
<td>0.00238</td>
<td>0.00238</td>
<td>0.00238</td>
<td>0.00238</td>
<td>0.00238</td>
<td>0.00238</td>
<td>0.00242</td>
</tr>
<tr>
<td>MC</td>
<td>0.04018</td>
<td>0.05004</td>
<td>0.04480</td>
<td>0.03878</td>
<td>0.03558</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IJGP</td>
<td>0.01202</td>
<td>0.01188</td>
<td>0.01194</td>
<td>0.01210</td>
<td>0.01192</td>
<td>0.01220</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>0.08726</td>
<td>0.09762</td>
<td>0.09272</td>
<td>0.08766</td>
<td>0.08334</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IJGP</td>
<td>0.07664</td>
<td>0.07498</td>
<td>0.07524</td>
<td>0.07578</td>
<td>0.07554</td>
<td>0.07816</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>0.15396</td>
<td>0.16048</td>
<td>0.15710</td>
<td>0.15452</td>
<td>0.15180</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IJGP</td>
<td>0.19070</td>
<td>0.19056</td>
<td>0.19016</td>
<td>0.19030</td>
<td>0.19056</td>
<td>0.19142</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MC</td>
<td>0.21890</td>
<td>0.22056</td>
<td>0.21928</td>
<td>0.21904</td>
<td>0.21830</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Bit Error Rate**

**Time**

\( N=400, P=4, 500 \text{ instances}, 30 \text{ iterations}, w^*=43 \)
Coding networks - Time

Coding, N=400, 500 instances, 30 iterations, w*=43
CPCS 422 – KL distance

Evidence = 0

Evidence = 30
CPCS 422 – KL vs. iterations

CPCS 422, evid=0, w*=23, 1instance

CPCS 422, evid=30, w*=23, 1instance

evidence=0
evidence=30
CPCS 422 – Relative error

CPCS 422, evid=30, w*=23, 1 instance

Relative error

0.01
0.1
1
10
100

IJGP at convergence
MC
IBP at convergence

evidence=30
CPCS 422 – Time vs. i-bound

CPCS 422, evid=0, w*=23, 1instance

Time (seconds)

IJGP 30 it (at convergence)
MC
IBP 10 it (at convergence)

i-bound

evidence=0
Conclusion

- IJGP borrows the iterative feature from IBP and the anytime virtues of bounded inference from MC

- Empirical evaluation showed the potential of IJGP, which improves with iteration and most of the time with i-bound, and scales up to large networks

- IJGP is almost always superior, often by a high margin, to IBP and MC

- Based on all our experiments, we think that IJGP provides a practical breakthrough to the task of belief updating