Feedback Message Passing for Inference in Gaussian Graphical Models

Ying Liu
Venkat Chandrasekaran, Animashree Anandkumar and Alan Willsky

Stochastic Systems Group,
Laboratory for Information and Decision Systems,
Massachusetts Institute of Technology

ISIT, Austin Texas, June 18, 2010

Gaussian Graphical Models

Gaussian Graphical Models

- The probability density of a Gaussian graphical model can be written as

$$p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T J \mathbf{x} + \mathbf{h}^T \mathbf{x}\}$$

where J is called the information matrix and $\mathbf{h}$ is called the potential vector.

Gaussian Graphical Models

- The probability density of a Gaussian graphical model can be written as
$$p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T J\mathbf{x} + \mathbf{h}^T\mathbf{x}\}$$
where J is called the information matrix and $\mathbf{h}$ is called the potential vector.
- For a valid model, J is symmetric and positive semidefinite.

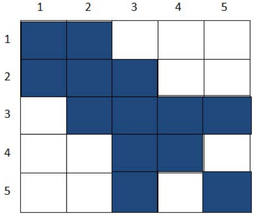
Gaussian Graphical Models

- The probability density of a Gaussian graphical model can be written as

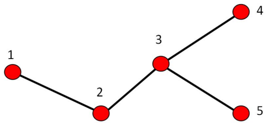
$$p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T J \mathbf{x} + \mathbf{h}^T \mathbf{x}\}$$

where J is called the information matrix and $\mathbf{h}$ is called the potential vector.
- For a valid model, J is symmetric and positive semidefinite.
- An information matrix J is sparse or Markov with respect to a graph if $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$: $\forall (i,j) \notin \mathcal{E}$, $J_{ij} = 0$.

Matrix Structure



Graph Structure



Inference Problem and Applications

$$p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T J\mathbf{x} + \mathbf{h}^T\mathbf{x}\}$$

Inference Problem and Applications

$$p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T J\mathbf{x} + \mathbf{h}^T\mathbf{x}\}$$

means $\boldsymbol{\mu} = J^{-1}\mathbf{h}$ and variances $\mathrm{diag}\{\Sigma = J^{-1}\}$

Inference Problem and Applications

$$p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T J\mathbf{x} + \mathbf{h}^T\mathbf{x}\}$$

means $\boldsymbol{\mu} = J^{-1}\mathbf{h}$ and variances $\mathrm{diag}\{\Sigma = J^{-1}\}$

- Solving this problem in general has $\mathcal{O}(n^3)$(fastest $\mathcal{O}(n^{2.376})$) time complexity, which is intractable for very large scale models.

Inference Problem and Applications

$$p(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T J\mathbf{x} + \mathbf{h}^T\mathbf{x}\}$$

means $\boldsymbol{\mu} = J^{-1}\mathbf{h}$ and variances $\mathrm{diag}\{\Sigma = J^{-1}\}$

- Solving this problem in general has $\mathcal{O}(n^3)$(fastest $\mathcal{O}(n^{2.376})$) time complexity, which is intractable for very large scale models.
- Applications: gene regulatory networks, medical diagnostics, oceanography, and communication systems

Related Work

Related Work

- Belief propagation on trees: linear time complexity, exactness

Related Work

- Belief propagation on trees: linear time complexity, exactness
- Loopy belief propagation for graphs with cycles.

Related Work

- Belief propagation on trees: linear time complexity, exactness
- Loopy belief propagation for graphs with cycles.
    - LBP performs reasonably well for certain loopy graphs (Murphy et al, Crick et al).

Related Work

- Belief propagation on trees: linear time complexity, exactness
- Loopy belief propagation for graphs with cycles.
  - LBP performs reasonably well for certain loopy graphs (Murphy et al, Crick et al).
  - Convergence and accuracy not guaranteed in general (Ihler et al, Weiss et al)

Related Work

- Belief propagation on trees: linear time complexity, exactness
- Loopy belief propagation for graphs with cycles.
  - LBP performs reasonably well for certain loopy graphs (Murphy et al, Crick et al).
  - Convergence and accuracy not guaranteed in general (Ihler et al, Weiss et al)
  - For Gaussian graphical models, if LBP converges, the means are correct and the variances are generally incorrect (Weiss et al)

Related Work

- Belief propagation on trees: linear time complexity, exactness
- Loopy belief propagation for graphs with cycles.
  - LBP performs reasonably well for certain loopy graphs (Murphy et al, Crick et al).
  - Convergence and accuracy not guaranteed in general (Ihler et al, Weiss et al)
  - For Gaussian graphical models, if LBP converges, the means are correct and the variances are generally incorrect (Weiss et al)
  - Walk-sum analysis framework (Malioutov et al)

Related Work

- Belief propagation on trees: linear time complexity, exactness
- Loopy belief propagation for graphs with cycles.
  - LBP performs reasonably well for certain loopy graphs (Murphy et al, Crick et al).
  - Convergence and accuracy not guaranteed in general (Ihler et al, Weiss et al)
  - For Gaussian graphical models, if LBP converges, the means are correct and the variances are generally incorrect (Weiss et al)
  - Walk-sum analysis framework (Malioutov et al)
- Generalized BP (Yedidia et al.), embedded trees (Sudderth et al.), inference by tractable subgraphs (Chandrasekaran et al.)

Main Results

Main Results

- *Exact Feedback Message Passing*
  Exact Solution: $\mathcal{O}(k^2 n)$, where k is the size of the "feedback nodes" and n is the number of nodes.

Main Results

- *Exact Feedback Message Passing*
  Exact Solution: $\mathcal{O}(k^2 n)$, where k is the size of the "feedback nodes" and n is the number of nodes.
- *Approximate Feedback Message Passing*
  Approximate Solution: trade-off complexity and accuracy by selecting a proper set of "feedback nodes" of bounded size. Walk-sum interpretation.

Main Results

- *Exact Feedback Message Passing*
  Exact Solution: $\mathcal{O}(k^2 n)$, where k is the size of the "feedback nodes" and n is the number of nodes.
- *Approximate Feedback Message Passing*
  Approximate Solution: trade-off complexity and accuracy by selecting a proper set of "feedback nodes" of bounded size. Walk-sum interpretation.
- High level idea: run common BP/LBP on non-feedback nodes; special message passing scheme for feedback nodes.

Main Results

- *Exact Feedback Message Passing*
  Exact Solution: $\mathcal{O}(k^2 n)$, where k is the size of the "feedback nodes" and n is the number of nodes.
- *Approximate Feedback Message Passing*
  Approximate Solution: trade-off complexity and accuracy by selecting a proper set of "feedback nodes" of bounded size. Walk-sum interpretation.
- High level idea: run common BP/LBP on non-feedback nodes; special message passing scheme for feedback nodes.
  1. Obtain inference results for feedback nodes first.

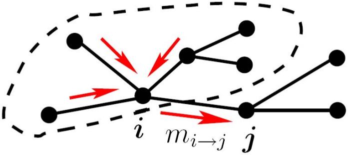
Main Results

- *Exact Feedback Message Passing*
  Exact Solution: $\mathcal{O}(k^2 n)$, where k is the size of the "feedback nodes" and n is the number of nodes.
- *Approximate Feedback Message Passing*
  Approximate Solution: trade-off complexity and accuracy by selecting a proper set of "feedback nodes" of bounded size. Walk-sum interpretation.
- High level idea: run common BP/LBP on non-feedback nodes; special message passing scheme for feedback nodes.
  1. Obtain inference results for feedback nodes first.
  2. Make corrections for the non-feedback nodes afterward.

Gaussian Belief Propagation

1. Message Passing

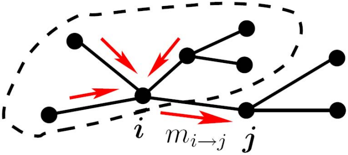
$$\forall j \in \mathcal{N}(i), \quad \Delta J_{i\to j}$$
$$\Delta h_{i\to j}$$



Gaussian Belief Propagation

1. Message Passing

$$\forall j \in \mathcal{N}(i), \quad \Delta J_{i \to j}$$
$$\Delta h_{i \to j}$$



2. Marginal Computation

$$\forall i \in \mathcal{V}, \quad \hat{J}_i = J_{ii} + \sum_{k \in \mathcal{N}(i)} \Delta J_{k \to i} \qquad \hat{h}_i = h_i + \sum_{k \in \mathcal{N}(i)} \Delta h_{k \to i}$$

$$\mu_i = \hat{J}_i^{-1} \hat{h}_i \qquad \mathbf{Var}\{i\} = \hat{J}_i^{-1}$$

.

Loopy Belief Propagation

- Message update scheme: completely local, no header information and suffers from the cyclic effects.

Loopy Belief Propagation

- Message update scheme: completely local, no header information and suffers from the cyclic effects.
- More memory and multiple messages?

Loopy Belief Propagation

- Message update scheme: completely local, no header information and suffers from the cyclic effects.
- More memory and multiple messages?
- Sacrifice some distributiveness for better convergence and accuracy?

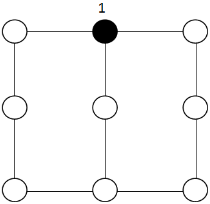
Loopy Belief Propagation

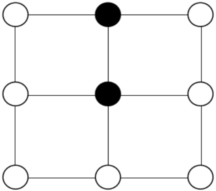
- Message update scheme: completely local, no header information and suffers from the cyclic effects.
- More memory and multiple messages?
- Sacrifice some distributiveness for better convergence and accuracy?
- Some special nodes?

Feedback Vertex Set

Feedback Vertex Set

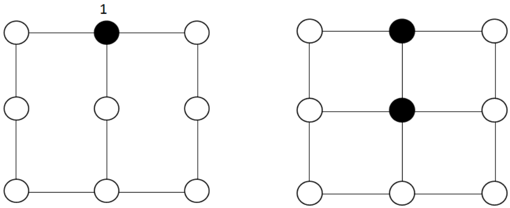
- Feedback vertex set (FVS) is a set of nodes whose removal results in a cycle-free graph.





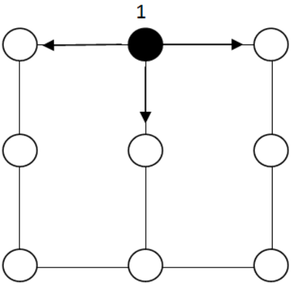
Feedback Vertex Set

- Feedback vertex set (FVS) is a set of nodes whose removal results in a cycle-free graph.



- In practice, a pseudo-FVS (a small subset of the FVS) may be sufficient for convergence and accuracy.

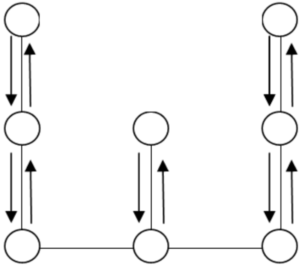
Exact Inference: a Single Feedback Node Case



- Extra potential vector $\mathbf{h}^1$,

$$h_j^1 = \begin{cases} 0 & j \notin \mathcal{N}(1) \\ J_{1j} & j \in \mathcal{N}(1) \end{cases}$$

Exact Inference: a Single Feedback Node Case (cont')



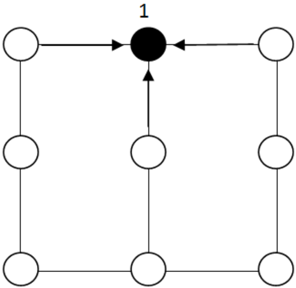
- Run belief propagation on $\mathcal{T}$. The messages are

$$\Delta J_{i\to j}^{\mathcal{T}} \quad \Delta h_{i\to j}^{\mathcal{T}} \quad \Delta h_{i\to j}^{1}$$

We obtain partial variance, partial mean, and feedback gain:

$$\mathrm{Var}^{\mathcal{T}}\{i\} \quad \mu_i^{\mathcal{T}} \quad g_i^1$$

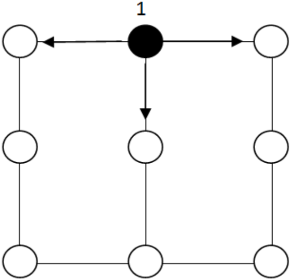
Exact Inference: a Single Feedback Node Case (cont')



$$\mathrm{Var}\{1\} = (J_{11} - \sum_{k \in \mathcal{N}(1)} J_{1k} g_k^1)^{-1}$$

$$\mu_1 = \mathrm{Var}\{1\}(h_1 - \sum_{j \in \mathcal{N}(1)} J_{1j} \mu_j^{\mathcal{T}})$$

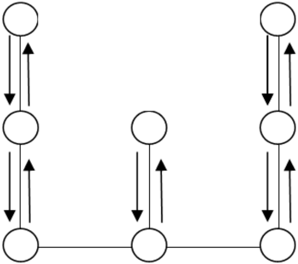
Exact Inference: a Single Feedback Node Case (cont')



- Node 1 tells its neighbors to make revisions on their node potentials.

$$\widetilde{h}_j = h_j - J_{1j}\mu_1,\ \forall j \in \mathcal{N}(1) \qquad \widetilde{h}_j = h_j,\ \forall j \notin \mathcal{N}(1)$$

Exact Inference: a Single Feedback Node Case (cont')



- Run BP on $\mathcal{T}$ with revised node potentials $\widetilde{\mathbf{h}}$ to obtain exact means.
- The exact variances can be achieved as

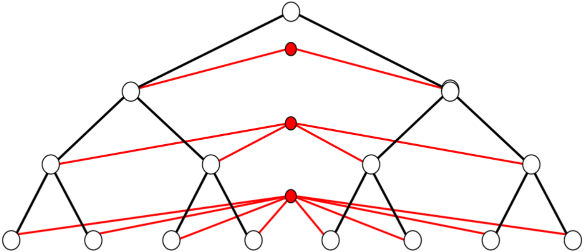
$$\mathrm{Var}\{i\} = \mathrm{Var}^{\mathcal{T}}\{i\} + \mathrm{Var}\{1\}(g_i^1)^2, \quad \forall i \in \mathcal{T}.$$

Exact Inference: Multiple Feedback Nodes Case

- With size k FVS, run BP with k extra messages and add more correction terms. $\mathcal{O}(k^2 n)$

Exact Inference: Multiple Feedback Nodes Case

- With size k FVS, run BP with k extra messages and add more correction terms. $\mathcal{O}(k^2 n)$
- Example:
  Exact Inference: $\mathcal{O}((\log n)^2 n)$



Approximate Feedback Message Passing

Approximate Feedback Message Passing

- Full FVS $\Rightarrow$ pseudo-FVS

Approximate Feedback Message Passing

- Full FVS $\Rightarrow$ pseudo-FVS
- Approximate inference among the tree-like part.

Approximate Feedback Message Passing

- Full FVS $\Rightarrow$ pseudo-FVS
- Approximate inference among the tree-like part.
- Exact inference among the feedback nodes.

Approximate Inference: Theoretical Results

Approximate Inference: Theoretical Results

- The spectral radius $\rho_{\mathcal{T}} < 1$ for the remaining graph $\mathcal{T}$ is a sufficient condition for convergence

Approximate Inference: Theoretical Results

- The spectral radius $\rho_{\mathcal{T}} < 1$ for the remaining graph $\mathcal{T}$ is a sufficient condition for convergence
- When it converges, feedback nodes get exact means and variances.

Approximate Inference: Theoretical Results

- The spectral radius $\rho_{\mathcal{T}} < 1$ for the remaining graph $\mathcal{T}$ is a sufficient condition for convergence
- When it converges, feedback nodes get exact means and variances.
- When it converges, non-feedback nodes get exact means but inaccurate variances (capturing a strictly larger set of walks).

Approximate Inference: Theoretical Results

- The spectral radius $\rho_{\mathcal{T}} < 1$ for the remaining graph $\mathcal{T}$ is a sufficient condition for convergence
- When it converges, feedback nodes get exact means and variances.
- When it converges, non-feedback nodes get exact means but inaccurate variances (capturing a strictly larger set of walks).
- For attractive models (where $J_{ij} \leq 0$ for $i \neq j$), better lower bounds of the variances.

Selecting a pseudo-FVS of Bounded Size

- Two goals: better convergence and better accuracy

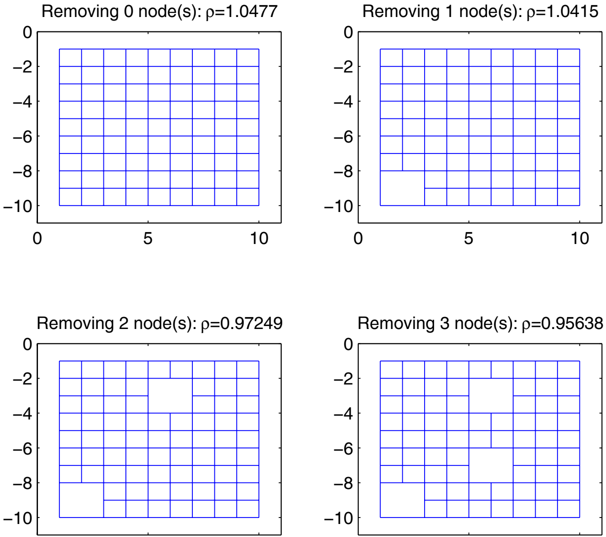
Selecting a pseudo-FVS of Bounded Size

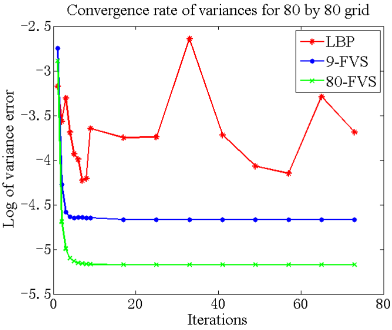
- Two goals: better convergence and better accuracy
- 1. $s(i) = \sum_{j \in \mathcal{N}(i)} |J_{ij}|$
  2. $s(i) = \sum_{l,k \in \mathcal{N}(i), l<k} |J_{il} J_{ik}|$

Selecting a pseudo-FVS of Bounded Size

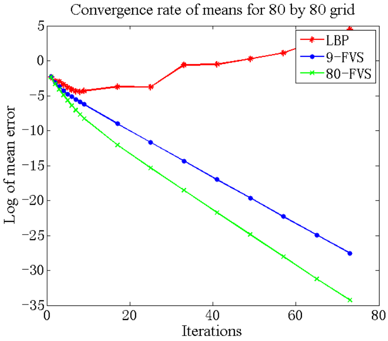
- Two goals: better convergence and better accuracy
- 1. $s(i) = \sum_{j \in \mathcal{N}(i)} |J_{ij}|$
  2. $s(i) = \sum_{l,k \in \mathcal{N}(i), l<k} |J_{il} J_{ik}|$
- Pick up one node with the largest score $s(i)$ at one step and continue with the remaining graph

Numerical Results





(o) Iterations versus variance errors



(p) Iterations versus mean errors

Figure: Inference errors of a 80×80 grid graph

- Empirically, $k = \mathcal{O}(\log n)$ seems to be sufficient.

Conclusions and Future Research

Conclusions

Conclusions and Future Research

Conclusions

- Exact feedback message passing: $\mathcal{O}(k^2 n)$

Conclusions and Future Research

Conclusions

- Exact feedback message passing: $\mathcal{O}(k^2 n)$
- Approximate feedback message passing: trade-off complexity and accuracy

Conclusions and Future Research

Conclusions

- Exact feedback message passing: $\mathcal{O}(k^2 n)$
- Approximate feedback message passing: trade-off complexity and accuracy

Future Research

- Performance on random graphs
- Computing the partition function

Conclusions and Future Research

Conclusions

- Exact feedback message passing: $\mathcal{O}(k^2 n)$
- Approximate feedback message passing: trade-off complexity and accuracy

Future Research

- Performance on random graphs
- Computing the partition function
- Corresponding structural learning problem

Questions and Comments?

- Thank you!