Orbit-Product Representation and Correction of Gaussian Belief Propagation

Jason Johnson[a], Vladimir Chernyak[b,a], Michael Chertkov[a]

[a]Los Alamos National Laboratory, Los Alamos, New Mexico

[b]Wayne State University, Detroit, Michigan

International Conference on Machine Learning

Montreal, Quebec

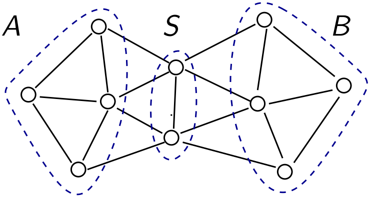
June 16, 2009

Graphical Models

A *graphical model* is a multivariate probability distribution that is expressed in terms of interactions among subsets of variables (e.g. pairwise interactions on the edges of a graph G).

$$P(x) = \frac{1}{Z} \prod_{i \in V} \psi_i(x_i) \prod_{\{i,j\} \in G} \psi_{ij}(x_i, x_j)$$

Markov property:



$$P(x_A, x_B | x_S) = P(x_A|x_S)P(x_B|x_S)$$

Given the potential functions ψ, the goal of *inference* is to compute marginals $P(x_i) = \sum_{x_{V \setminus i}} P(x)$ or the normalization constant Z, which is generally difficult in large, complex graphical models.

Gaussian Graphical Model

Information form of Gaussian density.

$$P(x) \propto \exp\left\{-\frac{1}{2}x^T Jx + h^T x\right\}$$

Inference corresponds to calculation of mean vector $\mu = J^{-1}h$, covariance matrix $K = J^{-1}$ or determinant $Z = \det J^{-1}$.

Gaussian graphical model: sparse J matrix

$$J_{ij} \neq 0 \text{ if and only if } \{i, j\} \in G$$

Potentials:

$$\begin{aligned} \psi_i(x_i) &= e^{-\frac{1}{2}J_{ii}x_i^2 + h_i x_i} \\ \psi_{ij}(x_i, x_j) &= e^{-J_{ij}x_i x_j} \end{aligned}$$

Marginals $P(x_i)$ specified by means μ_i and variances K_{ii}.

Belief Propagation

Belief Propagation iteratively updates a set of *messages* $\mu_{i \to j}(x_j)$ defined on directed edges of the graph G using the rule:

$$\mu_{i \to j}(x_j) \propto \sum_{x_i} \psi_i(x_i) \prod_{k \in N(i) \setminus j} \mu_{k \to i}(x_i) \psi(x_i, x_j)$$

Iterate message updates until converges to a fixed point.

Marginal Estimates: combine messages at a node

$$P(x_i) = \frac{1}{Z_i} \underbrace{\psi_i(x_i) \prod_{k \in N(i)} \mu_{k \to i}(x_i)}_{\tilde{\psi}_i(x_i)}$$

Belief Propagation II

Pairwise Estimates (on edges of graph):

$$P(x_i, x_j) = \frac{1}{Z_{ij}} \tilde{\psi}_i(x_i) \tilde{\psi}_j(x_j) \underbrace{\frac{\psi(x_i, x_j)}{\mu_{i \to j}(x_j) \mu_{j \to i}(x_i)}}_{\tilde{\psi}_{ij}(x_i, x_j)}$$

Estimate of Normalization Constant:

$$Z^{\text{bp}} = \prod_{i \in V} Z_i \prod_{\{i,j\} \in G} \frac{Z_{ij}}{Z_i Z_j}$$

BP fixed point is *saddle point* of RHS with respect to messages/reparameterizations.

In trees, BP converges in finite number of steps and is exact (equivalent to variable elimination).

Gaussian Belief Propagation (GaBP)

Messages $\mu_{i\to j}(x_j) \propto \exp\{\frac{1}{2}\alpha_{i\to j}x_j^2 + \beta_{i\to j}x_j\}$.

BP fixed-point equations reduce to:

$$\begin{aligned}
\alpha_{i\to j} &= J_{ij}^2(J_{ii} - \alpha_{i\setminus j})^{-1} \\
\beta_{i\to j} &= -J_{ij}(J_{ii} - \alpha_{i\setminus j})^{-1}(h_i + \beta_{i\setminus j})
\end{aligned}$$

where $\alpha_{i\setminus j} = \sum_{k\in N(i)\setminus j} \alpha_{k\to i}$ and $\beta_{i\setminus j} = \sum_{k\in N(i)\setminus j} \alpha_{k\to i}$.

Marginals specified by:

$$K_i^{\mathrm{bp}} = (J_{ii} - \sum_{k\in N(i)} \alpha_{k\to i})^{-1}$$

$$\mu_i^{\mathrm{bp}} = K_i^{\mathrm{bp}}(h_i + \sum_{k\in N(i)} \beta_{k\to i})$$

Gaussian BP Determinant Estimate

Estimates of pairwise covariance on edges:

$$K^{\mathrm{bp}}_{(ij)} = \begin{pmatrix} J_{ii} - \alpha_{i\setminus j} & J_{ij} \\ J_{ij} & J_{jj} - \alpha_{j\setminus i} \end{pmatrix}^{-1}$$

Estimate of $Z \triangleq \det K = \det J^{-1}$:

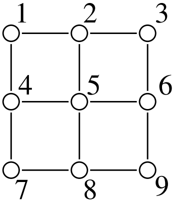
$$Z^{\mathrm{bp}} = \prod_{i \in V} Z_i \prod_{\{i,j\} \in G} \frac{Z_{ij}}{Z_i Z_j}$$

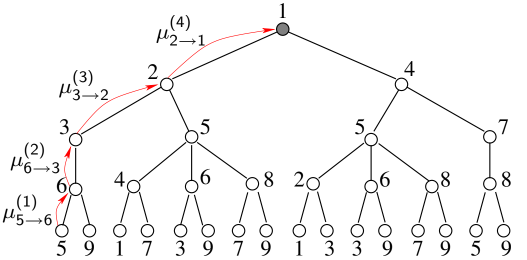
where $Z_i = K^{\mathrm{bp}}_i$ and $Z_{ij} = \det K^{\mathrm{bp}}_{(ij)}$.

Exact in tree models (equivalent to Gaussian elimination), approximate in loopy models.

The BP Computation Tree

BP marginal estimates are equivalent to the exact marginal in a tree-structured model [Weiss & Freeman].





The BP messages correspond to upwards variable elimination steps in this computation tree.

Neumann Series and Walk-Sums[†]

Let $J = I - R$. If $\rho(R) < 1$ then $(I - R)^{-1} = \sum_{L=0}^{\infty} R^L$.

Walk-Sum interpretation of inference:

$$K_{ij} = \sum_{L=0}^{\infty} \sum_{w:i \xrightarrow{L} j} R^w \stackrel{?}{=} \sum_{w:i \to j} R^w$$

$$\mu_i = \sum_j h_j \sum_{L=0}^{\infty} \sum_{w:j \xrightarrow{L} i} R^w \stackrel{?}{=} \sum_{w:* \to i} h_* R^w$$

Walk-Summable if $\sum_{w:i \to j} |R^w|$ converges for all i, j. Absolute convergence implies convergence of walk-sums (to same value) for arbitrary orderings and partitions of the set of walks. Equivalent to $\rho(|R|) < 1$.

[†]Prior work with D. Malioutov and A. Willsky (NIPS, JMLR).

Zeta Function and Orbit-Product

What about the determinant?

Definition of Orbits:

- A walk is *closed* if it begins and ends at same vertex.
- It is *primitive* if does not repeat a shorter walk.
- Two primitive walks are *equivalent* if one is a cyclic shift of the other.
- Define *orbits* $\ell \in \mathcal{L}$ of G to be equivalence classes of closed, primitive walks.

Theorem. Let $Z \triangleq \det(I - R)^{-1}$. If $\rho(|R|) < 1$ then

$$Z = \prod_{\ell} (1 - R^{\ell})^{-1} \triangleq \prod_{\ell} Z_{\ell}.$$

Closely resembles definition of *zeta functions* in graph theory.

Walk-Sum Interpretation of GaBP†

Combine interpretation of BP as exact inference on computation tree with walk-sum interpretation of Gaussian inference in trees:

- complete walk-sum for the means
- incomplete walk-sum for the variances
- messages represent walk-sums in subtrees of computation tree

†Prior work with D. Malioutov and A. Willsky (NIPS,JMLR).

Z_{bp} as Totally-Backtracking Orbit-Product

Classification of Orbits:

- Orbit is *reducible* if it contains backtracking steps ...$(ij)(ji)$..., else it is *irreducible* (or *backtrackless*).
- Every orbit ℓ has a unique irreducible core $\gamma = \Gamma(\ell)$ obtained by iteratively deleting pairs of backtracking steps until no more remain. Let $\mathcal{L}_\gamma$ denote the set of all orbits that reduce to γ.
- Orbit is *totally backtracking* (or *trivial*) if it reduces to the empty orbit $\Gamma(\ell) = \emptyset$, else it is *non-trivial*.

Theorem. If $\rho(|R|) < 1$ then Z^{bp} (defined earlier) is equal to the totally-backtracking orbit-product:

$$Z^{\mathrm{bp}} = \prod_{\ell \in \mathcal{L}_\emptyset} Z_\ell$$

Orbit-Product Correction and Error Bound

Orbit-product correction to Z^{bp}:

$$Z = Z^{\mathrm{bp}} \prod_{\ell \notin \mathcal{L}_\emptyset} Z_\ell$$

Error Bound: missing orbits must all involve cycles of the graph...

$$\left| \log \frac{Z}{Z^{\mathrm{bp}}} \right| \leq \frac{\rho^g}{g(1-\rho)}$$

where $\rho \triangleq \rho(|R|) < 1$ and g is girth of the graph (length of shortest cycle).

Reduction to Backtrackless Orbit-Product Correction

We may reduce the orbit-product correction to one over just backtrackless orbits γ

$$Z = Z_{\mathrm{bp}} \prod_{\ell} Z_{\ell} = Z_{\mathrm{bp}} \prod_{\gamma} \underbrace{\left(\prod_{\ell \in \mathcal{L}(\gamma)} Z_{\ell} \right)}_{Z'_{\gamma}}$$

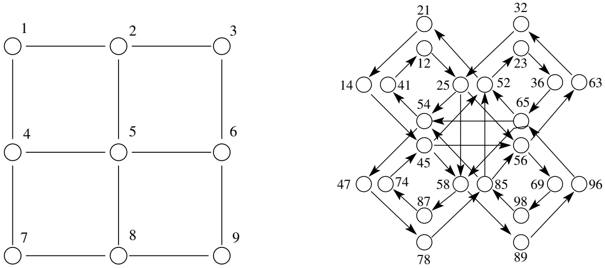
with modified orbit-factors Z'_{γ} based on GaBP

$$Z'_{\gamma} = (1 - \prod_{(ij) \in \gamma} r'_{ij})^{-1} \quad \text{where} \quad r'_{ij} \triangleq (1 - \alpha_{i \backslash j})^{-1} r_{ij}$$

The factor $(1 - \alpha_{i \backslash j})^{-1}$ serves to reconstruct totally-backtracking walks at each point i along the backtrackless orbit γ.

Backtrackless Determinant Correction

Define backtrackless graph G' of G as follows: nodes of G' correspond to directed edges of G, edges $(ij) \to (jk)$ for $k \neq i$.



Let R' be adjacency matrix of G' with modified edge-weights r' based on GaBP. Then,

$$Z = Z_{\mathrm{bp}} \det(I - R')^{-1}$$

Block-Resummation Method

Let $\mathcal{B}$ be a collection of subsets of nodes (*blocks*) $B \subset V$ such that if $A, B \in \mathcal{B}$ the $A \cap B \in \mathcal{B}$. Define $n_B = 1 - \sum_{B' \supsetneq B} n_{B'}$.

To capture all orbits covered by any block (without over-counting) we calculate the estimate:

$$Z_{\mathcal{B}} \triangleq \prod_B Z_B^{n_B} \triangleq \prod_B (\det(I - R_B)^{-1})^{n_B}$$

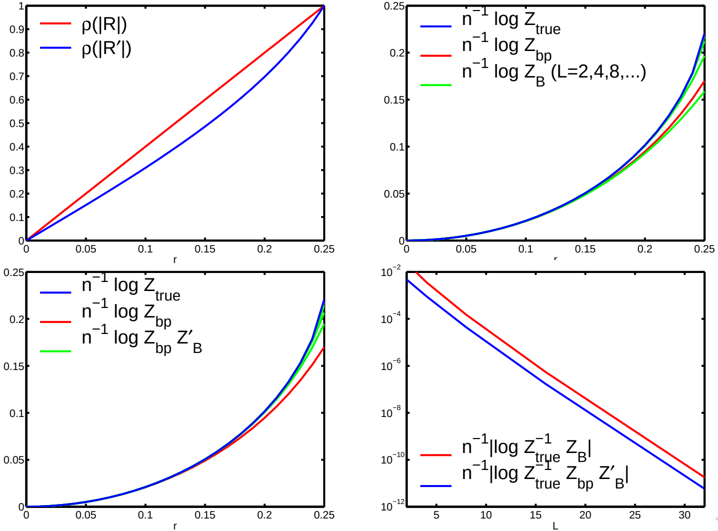
Error Bounds. Select blocks to cover all orbits up to length L. Then,

$$\left| \frac{1}{n} \log \frac{Z_{\mathcal{B}}}{Z} \right| \leq \frac{\rho^L}{L(1-\rho)}$$

Similar approach to estimate $Z' \triangleq \det(I - R')^{-1}$ from sub-matrices of R'. Error controlled by $\rho' \leq \rho$.

Example: 2-D Grids

256×256 Periodic Grid, uniform edge weights $r \in [0, .25]$.
Blocks: $L \times L$, $L \times \frac{L}{2}$, $\frac{L}{2} \times L$ and $\frac{L}{2} \times \frac{L}{2}$ shifted by $\frac{L}{2}$.
Test with $L = 2, 4, 8, 16, 32$.



Conclusion and Future Work

Graphical view of inference in walk-summable Gaussian graphical models give a very intuitive framework for understanding iterative inference algorithms and approximation methods.

Future Work:

- Extension to Generalized Belief Propagation (iterative message-passing between blocks).
- Extension to Non-Walksummable Models: compute corrections to inference based on nearest walk-summable model.
- Boot-Strapping GaBP using powers of a matrix.
- Multiscale resummation methods to approximate long orbits from coarse-grained model.