On the minima of Bethe Free Energy in Gaussian Distributions

Yu Nishiyama and Sumio Watanabe

Tokyo Institute of Technology, Japan

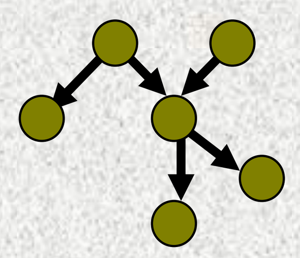
Abstract

Belief propagation (BP) is an effective algorithm for computing marginals of high-dimensional distributions. Loopy belief propagation (LBP), which is applied to graphical models with cycles, is known not to be guaranteed to converge and, if it does, it computes approximate marginals. Fixed-points of LBP are known to accord with extrema of Bethe free energy.

We focus on the Bethe free energy applied to Gaussian graphical models (GGM) and analytically clarify the accuracy of LBP. For a single cycle, we derive the minimum of the Bethe free energy and show the quantity that determines accuracy of LBP. For arbitrary topological graph, we also show the principal quantity that decides accuracy of LBP.

Probabilistic information processing

Bayesian networks

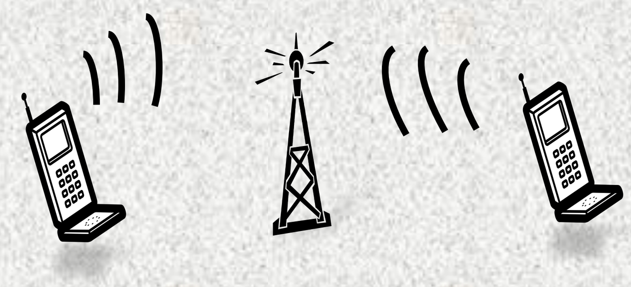


error correcting codes

noise decode

000101 → 000111 → 000101

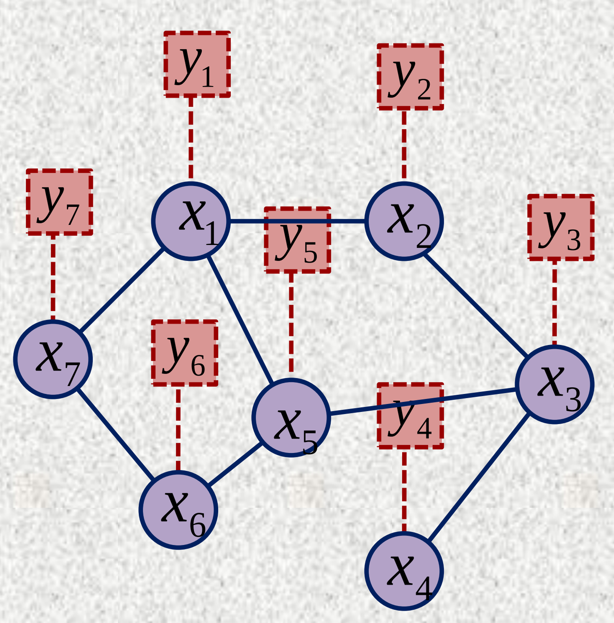
mobile communications



probabilistic image processing

degrade image —restore→ restored image

Probabilistic inference



$G=\{V,E\}$

$\mathbf{x}=\{x_1,\cdots,x_d\}$: hidden variables

$\mathbf{y}=\{y_1,\cdots,y_d\}$: observed variables

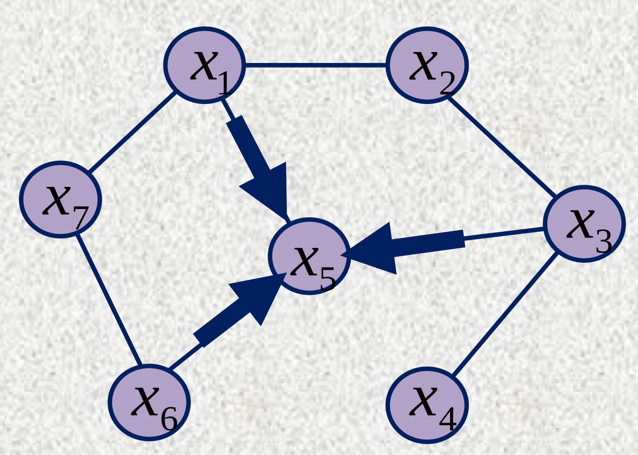
We consider Gaussian graphical models such that

$$P(\mathbf{x},\mathbf{y})\propto \exp\{-\frac{1}{2}(\mathbf{x}^T,\mathbf{y}^T)\begin{pmatrix}V_{xx} & V_{xy}\\ V_{yx} & V_{yy}\end{pmatrix}\begin{pmatrix}\mathbf{x}\\ \mathbf{y}\end{pmatrix}\}.$$

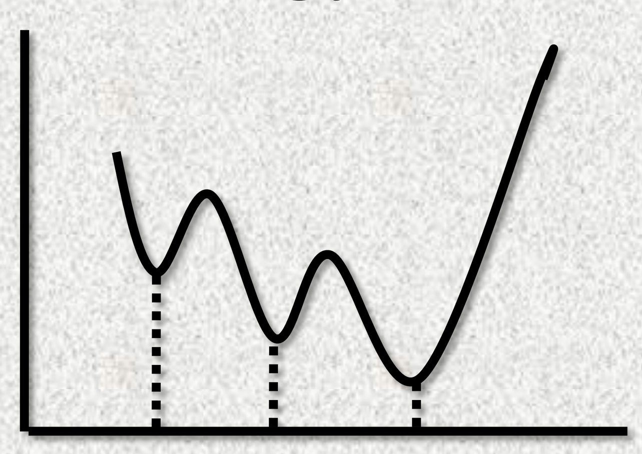
Focusing probabilistic inference problem is to compute posterior mean $\mu_{x_i|\mathbf{y}}$ and variance $\Sigma_{x_i|\mathbf{y}}$ of a single hidden node i , given data $\mathbf{y}=\{y_1,\cdots,y_d\}$.

Y. Weiss, *Neural Computation* **13**(10), 2173-2200, 2001.

Belief propagation (BP) & Bethe free energy



Bethe free energy



Optimization techniques

- CCCP (A. L. Yuille, 2002)
- SEQ (Tonosaki, 2007)
- ⋮

Beliefs computed by BP ⟷ Extrema of Bethe free energy

BP algorithm

Target probability distribution

$$P(\mathbf{x},\mathbf{y}) = \frac{1}{Z}\prod_{\{ij\}\in E}\psi_{ij}(x_i,x_j)\prod_{i\in V}\psi_{ii}(x_i,y_i)$$

Message update rules

$$M_{i\to j}^{(t+1)}(x_j) \leftarrow \frac{1}{\tilde{Z}_{ij}}\int \psi_{ij}(x_i,x_j)\psi_{ii}(x_i,y_i)\prod_{k\in N_i\setminus\{j\}} M_{k\to i}^{(t)}(x_i)dx_i$$

$$M_{j\to i}^{(t+1)}(x_i) \leftarrow \frac{1}{\tilde{Z}_{ji}}\int \psi_{ij}(x_i,x_j)\psi_{jj}(x_j,y_j)\prod_{k\in N_j\setminus\{i\}} M_{k\to j}^{(t)}(x_j)dx_j \qquad \{ij\}\in E$$

Belief update rules

$$b_i^{(t)}(x_i) \leftarrow \frac{\psi_{ii}(x_i,y_i)}{Z_i}\prod_{k\in N_i} M_{k\to i}^{(t)}(x_i) \qquad i\in V$$

$$b_{ij}^{(t)}(x_i,x_j) \leftarrow \frac{\psi_{ii}(x_i,y_i)}{Z_{ij}}\prod_{k\in N_i\setminus\{j\}} M_{k\to i}^{(t)}(x_i)\prod_{k\in N_j\setminus\{i\}} M_{k\to j}^{(t)}(x_j) \qquad \{ij\}\in E$$

Gaussian BP algorithm

We set messages and beliefs as Gaussians

$$M_{i\to j}(x_j) = \sqrt{\frac{\lambda_{i\to j}}{2\pi}}\exp\{-\frac{\lambda_{i\to j}}{2}(x_j-\mu_{i\to j})^2\} \qquad b_i(x_i) = \sqrt{\frac{\Lambda_i}{2\pi}}\exp\{-\frac{\Lambda_i}{2}(x_i-\mu_i)^2\} \qquad i\in V$$

$$M_{j\to i}(x_i) = \sqrt{\frac{\lambda_{j\to i}}{2\pi}}\exp\{-\frac{\lambda_{j\to i}}{2}(x_i-\mu_{j\to i})^2\} \quad \{ij\}\in E \qquad b_{ij}(x_i,x_j) = \sqrt{\frac{|\Lambda_{ij}|}{(2\pi)^2}}\exp\{-\frac{1}{2}(x_i-\mu_i,x_j-\mu_j)\Lambda_{ij}(x_i-\mu_i,x_j-\mu_j)^T\} \quad \{ij\}\in E$$

Target probability distribution

$$P(\mathbf{x},\mathbf{y}) = \frac{1}{Z}\prod_{\{ij\}\in E}\exp\{-\frac{1}{2}(x_i,x_j)V_{ij}\begin{pmatrix}x_i\\x_j\end{pmatrix}\}\prod_{i\in V}\exp\{-\frac{1}{2}(x_i,y_i)V_{ii}\begin{pmatrix}x_i\\y_i\end{pmatrix}\} \qquad V_{ij} = \begin{pmatrix}\rho_{i\to j}s_{ii} & s_{ij}\\ s_{ji} & \rho_{j\to i}s_{jj}\end{pmatrix} \qquad (V_{ii})_{1,1} = \rho_{i\to i}s_{ii}$$

Message update rules

$$\tilde{\lambda}_{i\to j}^{(t+1)} \leftarrow -\frac{s_{ij}^2}{s_{ii}+\sum_{k\in N(i)\setminus\{j\}}\tilde{\lambda}_{k\to i}^{(t)}}$$

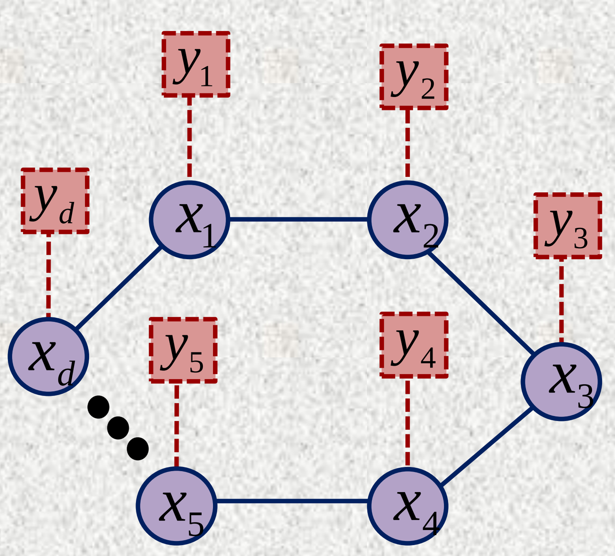
$$\tilde{\lambda}_{j\to i}^{(t+1)} \leftarrow -\frac{s_{ji}^2}{s_{jj}+\sum_{k\in N(j)\setminus\{i\}}\tilde{\lambda}_{k\to j}^{(t)}} \qquad \{ij\}\in E$$

Belief update rules

$$\Lambda_i^{(t)} \leftarrow s_{ii} + \sum_{k\in N(i)}\tilde{\lambda}_{k\to i}^{(t)} \qquad i\in V$$

$$\Lambda_{ij}^{(t)} \leftarrow \begin{pmatrix} s_{ii}+\sum_{k\in N(i)\setminus\{j\}}\tilde{\lambda}_{k\to i}^{(t)} & s_{ij}\\ s_{ji} & s_{jj}+\sum_{k\in N_j\setminus\{i\}}\tilde{\lambda}_{k\to j}^{(t)}\end{pmatrix} \qquad \{ij\}\in E$$

Single cycle



$$G = \{V, E\}$$

True inverse variances

$$\frac{\det S}{\Delta_{ii}} \qquad i\in V$$

Theorem 1 (messages on a single cycle)

When graph G is a single cycle, the fixed-points of messages are given by

$$\lambda_{i\to i+1}^{*} = \frac{s_{i,i+1}\Delta_{i,i+1} + (2\rho_{i+1\to i}-1)s_{i+1,i+1}\Delta_{i+1,i+1} - s_{i+2,i+1}\Delta_{i+2,i+1} \pm\sqrt{D}}{2\Delta_{i+1,i+1}},$$

$$\lambda_{i\to i-1}^{*} = \frac{s_{i,i-1}\Delta_{i,i-1} + (2\rho_{i-1\to i}-1)s_{i-1,i-1}\Delta_{i-1,i-1} - s_{i-2,i-1}\Delta_{i-2,i-1} \pm\sqrt{D}}{2\Delta_{i-1,i-1}}, \qquad i\in V,$$

$$D = (\det S)^2 + (-1)^d 4 s_{12}s_{23}s_{34}\cdots s_{d-1,d}s_{d,1}\det S,$$

where Δ_{ij} is the (i, j)th cofactor of a matrix $S = V_{xx}$.

Theorem 2 (beliefs on a single cycle)

When graph G is a single cycle, the minimum of Bethe free energy is given by

$$\Lambda_i^{*} = \frac{\det S}{\Delta_{ii}}\sqrt{1+\varepsilon}, \qquad \Lambda_{i,i+1}^{*} = \begin{pmatrix}\frac{E_{i,i+1}}{\Delta_{ii}} & s_{i,i+1}\\ s_{i,i+1} & \frac{E_{i,i+1}}{\Delta_{i+1,i+1}}\end{pmatrix}, \qquad i\in V,$$

where

$$\varepsilon = \frac{(-1)^d 4 s_{12}s_{23}s_{34}\cdots s_{d-1,d}s_{d,1}}{\det S}, \qquad E_{i,i+1} = \frac{\det S+\sqrt{D}}{2} - s_{i,i+1}\Delta_{i,i+1}.$$

Accuracy of LBP is decided by the quantity ε, which is explicitly calculated by matrix $S = V_{xx}$.

Condition 1 (LBP convergence)

When graph G is a single cycle, the condition

$$\varepsilon = \frac{(-1)^d 4 s_{12} s_{23} s_{34} \cdots s_{d-1,d} s_{d,1}}{\det S} \geq -1$$

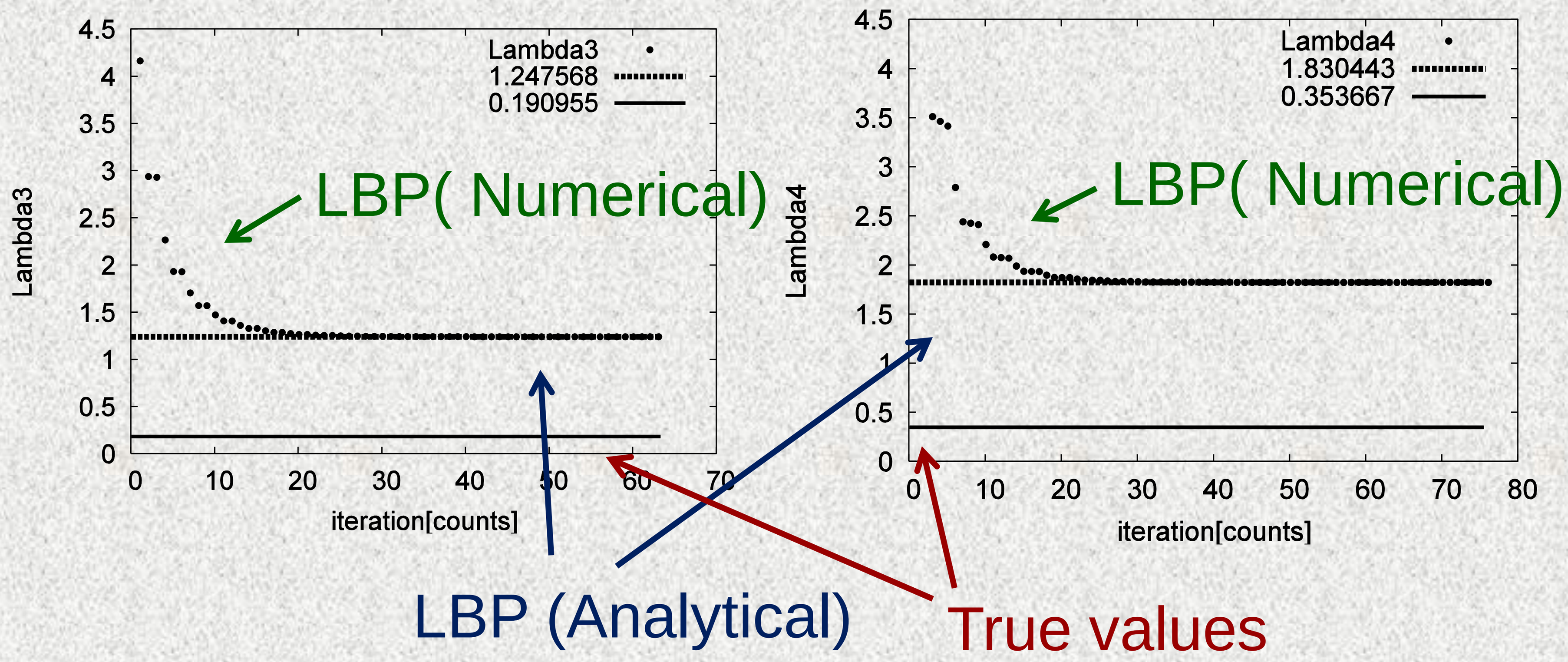
is a necessary condition for LBP convergence.

For example, when positive definite matrix is

$$S = \begin{pmatrix} 10 & 9 & 1 \\ 9 & 10 & 5 \\ 1 & 5 & 10 \end{pmatrix},$$

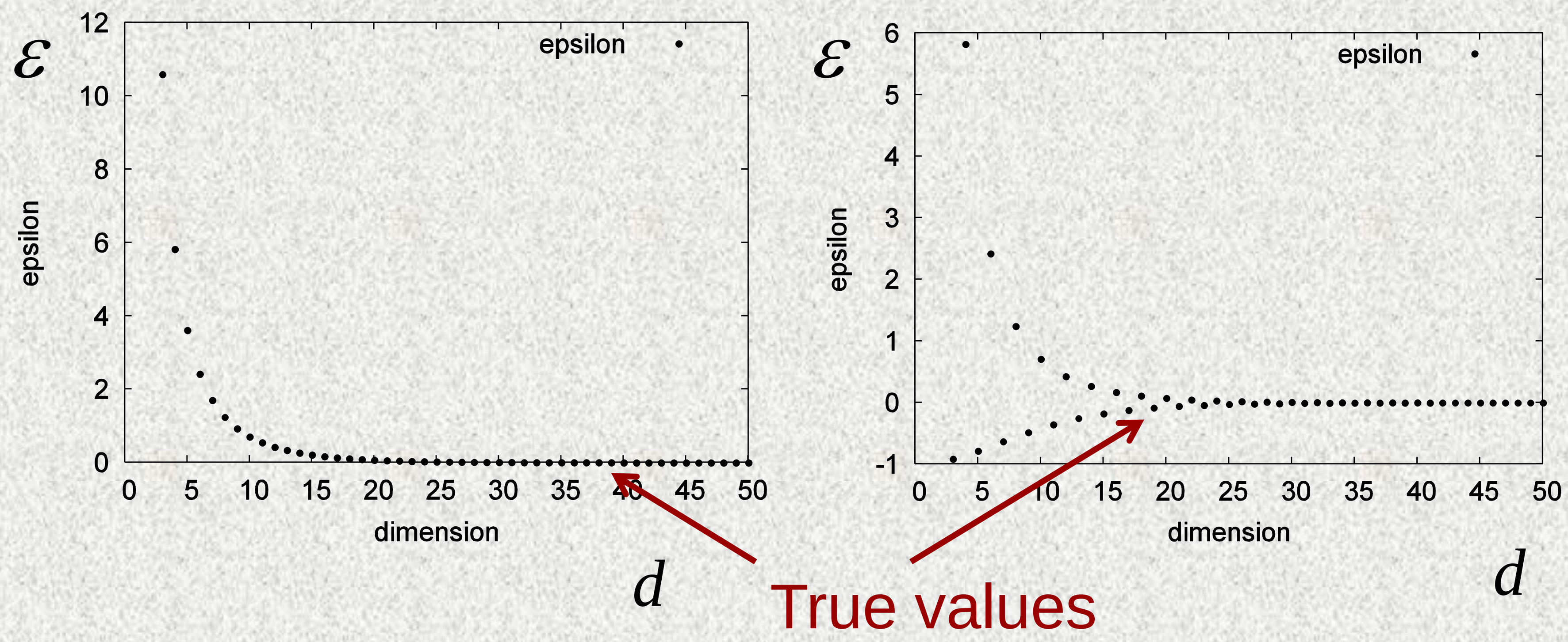
ε is calculated as $\varepsilon = -9$.

Numerical experiments



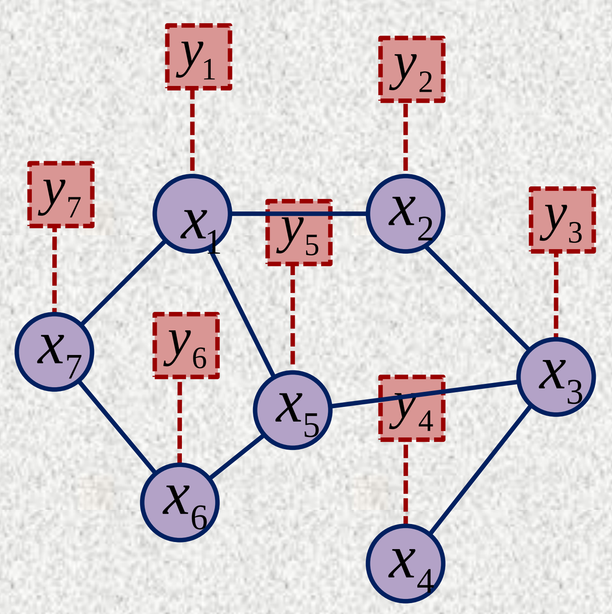
$$S_1 = \begin{pmatrix} 10 & -9.9 & 1 \\ -9.9 & 10 & 0.4 \\ 1 & 0.4 & 10 \end{pmatrix} \qquad S_2 = \begin{pmatrix} 10 & 9 & 0 & 1 \\ 9 & 10 & 4 & 0 \\ 0 & 4 & 10 & 1.9 \\ 1 & 0 & 1.9 & 10 \end{pmatrix}$$

Behavior of the quantity ε



$$S_3(d) = \begin{pmatrix} 10 & -4.9 & & & -4.9 \\ -4.9 & 10 & -4.9 & & \\ & -4.9 & 10 & \ddots & \\ & & \ddots & \ddots & -4.9 \\ -4.9 & & & -4.9 & 10 \end{pmatrix} \qquad S_4(d) = \begin{pmatrix} 10 & 4.9 & & & 4.9 \\ 4.9 & 10 & 4.9 & & \\ & 4.9 & 10 & \ddots & \\ & & \ddots & \ddots & 4.9 \\ 4.9 & & & 4.9 & 10 \end{pmatrix}$$

Arbitrary topological graph



$$G = \{V, E\}$$

$$S = \begin{pmatrix} s_{11} & ss_{12} & ss_{13} & ss_{14} & ss_{15} \\ ss_{21} & s_{22} & ss_{23} & ss_{24} & ss_{25} \\ ss_{31} & ss_{32} & s_{33} & ss_{34} & ss_{35} \\ ss_{41} & ss_{42} & ss_{43} & s_{44} & ss_{45} \\ ss_{51} & ss_{52} & ss_{53} & ss_{54} & s_{55} \end{pmatrix}$$

$$= S_d + sS_o \quad (s << 1)$$

Theorem 3 (a LBP fixed-point)

When graph G is an arbitrary topological graph, a LBP fixed-point is expanded as

$$\Lambda_i^*(s) = s_{ii} - \sum_{j=1(\neq i)}^{d} \frac{s_{ji}^2}{s_{jj}} s^2 + O(s^4), \qquad i \in V.$$

The true inverse variance is expanded as

$$\frac{\det S}{\Delta_{ii}} = s_{ii} - \sum_{j=1(\neq i)}^{d} \frac{s_{ji}^2}{s_{jj}} s^2 + \delta_i s^3 + O(s^4), \qquad i \in V,$$

where

$$\delta_i = 2 s_{ii} \sum_{\substack{(j,k) \\ \neq i}} \bar{s}_{ij} \bar{s}_{ik} \bar{s}_{jk}. \quad \left(\bar{s}_{ij} = \frac{s_{ij}}{\sqrt{s_{ii}}\sqrt{s_{jj}}}\right)$$

When correlations between variable nodes are small $(s << 1)$, δ_i mainly decides accuracy of LBP.

Summary

For loopy belief propagation applied to Gaussian graphical models, we derived estimated posterior variances and analytically clarified accuracy of LBP. For graphs of a single cycle, the quantity ε determines accuracy of LBP and yields a necessary condition for LBP convergence. For arbitrary topological graphs having small correlations, δ_i decides accuracy of LBP. Using δ_i, a straightforward correction of LBP algorithm can be constructed.