Convergence Analysis of Belief Propagation for Pairwise Linear Gaussian Models

Jian Du†
Joint Work with Shaodan Ma◊, Yik-Chung Wu†, Soummya Kar†, and José M. F. Moura†

Carnegie Mellon University†
University of Macau◊
The University of Hong Kong†

Acknowledgements: NSF under grants CCF-1513936.
Belief Propagation (BP) on Trees

- Computing *marginal distributions/modes* efficiently by exploiting the distributive law.

\[
q(x_1) \propto \int ... \int f_A(x_1) f_B(x_1, x_2, x_5) f_C(x_2, x_3) f_D(x_2, x_4) \, dx_2 \, dx_3 \, dx_4 \, dx_5 \, dx_6
\]

\[
= f_A(x_1) \int_{x_2} \int_{x_5} f_B(x_1, x_2, x_5) f_E(x_5) \left( \int_{x_3} f_C(x_2, x_3) \, dx_3 \right) \left( \int_{x_4} f_D(x_2, x_4) \, dx_4 \right) \, dx_5 \, dx_2
\]

**Message from variable to factor:**
1. \(m_{f_A \rightarrow 1}(x_1)\)
2. \(m_{f_B \rightarrow 1}(x_1)\)
3. \(m_{f_B \rightarrow 2}(x_2)\)

**Message from factor to variable:**
1. \(m_{f_C \rightarrow 2}(x_2)\)
3. \(m_{f_D \rightarrow 2}(x_2)\)

**Marginal distribution:**
\[
q(x_1) \propto \text{(1)} \times \text{(3)}
\]
Message Definition

- **Message from variable to factor:**

\[
m_{j \rightarrow f_n}(x_j) := p(x_j) \prod_{f_k \in B(j) \setminus f_n} m_{f_k \rightarrow j}(x_j)
\]

  e.g., \(m_{2 \rightarrow f_B}(x_2) = m_{f_D \rightarrow 2}(x_2)m_{f_C \rightarrow 2}(x_2)\)

- **Message from factor to variable:**

\[
m_{f_n \rightarrow i}(x_i) := \int \cdots \int f_n \prod_{j \in B(f_n) \setminus i} m_{j \rightarrow f_n}(x_j) d\{x_j\}_{j \in (f_n) \setminus i}
\]

  e.g., \(m_{f_B \rightarrow 1}(x_1) = \int \int f_B m_{2 \rightarrow f_B}(x_2)m_{5 \rightarrow f_B}(x_5)\) \(d x_2 d x_5\)

- **Marginal distribution**

\[
b(x_i) \propto \prod_{f_n \in B(x_i)} m_{f_n \rightarrow i}(x_i)
\]

  e.g., \(b(x_1) \propto m_{f_A \rightarrow 1}(x_1)m_{f_B \rightarrow 1}(x_1)\)
BP on Graph with Loops

- Distributive law may NOT be exploited on graph with loops

\[ q(x_1) \propto \int \cdots \int f_A(x_1, x_2, x_3, x_4) f_B(x_1, x_2, x_3, x_4) \times f_C(x_1, x_2, x_3, x_4) \text{d}x_2 \text{d}x_3 \text{d}x_4 \]

- Use same message updating rule \textit{in parallel}
  - Message from \textit{variable} to \textit{factor}:
    \[ m_{j \rightarrow f_n}^{(l)}(x_j) = p(x_j) \prod_{f_k \in \mathcal{B}(j) \setminus f_n} m_{f_k \rightarrow j}^{(l-1)}(x_j) \]
  - Message from \textit{factor} to \textit{variable}:
    \[ m_{f_n \rightarrow i}^{(l)}(x_i) = \int \cdots \int f_n \prod_{j \in \mathcal{B}(f_n) \setminus i} m_{j \rightarrow f_n}^{(l)}(x_j) \text{d}\{x_j\}_{j \in (f_n) \setminus i} \]
  - Approximate marginal distribution:
    \[ b^{(l)}(x_i) \propto \prod_{f_n \in \mathcal{B}(x_i)} m_{f_n \rightarrow i}^{(l-1)}(x_i) \]

Will \(b^{(l)}\) converge? Where will it converge? Convergence rate?
Pairwise Models: GMRF & Linear Gaussian Model

1) GMRF

\[ q(x) \propto \exp \left\{ -\frac{1}{2} x^T J x + h^T x \right\} \]

\[ f_i(x_i) = \exp \left( -\frac{1}{2} J_{i,i} x_i^2 + h_i x_i \right) \]

\[ f_{i,j}(x_i, x_j) = \exp \left( -x_i J_{i,j} x_j \right) \]

2) Pairwise Linear Gaussian Model

\[ f_i(x_j) \sim \mathcal{N}(x_j | 0, W_j) \]

\[ f_{i,j}(x_i, x_j) = \mathcal{N} \left( y_{i,j} | A_{i,j} x_i + A_{i,j} x_j, R_{i,j} \right) \]
**BP on GMRF**

- A joint Gaussian distribution function can always be written as:

\[
p(x) \propto \exp \left( -\frac{1}{2} x^T J x + h^T x \right)
\]

\[
= \prod_{i \in \mathcal{V}} \exp \left( -\frac{1}{2} J_{i,i} x_i^2 + h_i x_i \right) \prod_{(i,j) \in \mathcal{E}_{\text{MRF}}} \exp \left( -x_i J_{i,j} x_j \right)
\]

\[\equiv f_i(x_i) \quad \text{and} \quad \equiv f_{i,j}(x_{i,j})\]

- A sufficient convergence condition, given by the spectrum radius is obtained:

\[
\rho(|I - J|) < 1 \iff \text{Walk-summable (BP converges on GMRF)}
\]

---

BP in Linear Gaussian Model

- In a general connected network, the local observations at every node \( n \in \mathcal{V} \) are in the form of

\[
\mathbf{y}_{i,j} = \mathbf{A}_{j,i} \mathbf{x}_i + \mathbf{A}_{i,j} \mathbf{x}_j + \mathbf{z}_{i,j} \\
\mathbf{z}_{i,j} \sim \mathcal{N}(\mathbf{z}_{i,j}|0, \mathbf{R}_{i,j}) \\
p(\mathbf{x}_i) \sim \mathcal{N}(\mathbf{x}_i|0, \mathbf{W}_i)
\]

- **The joint posterior distribution of** \([\mathbf{x}_1, \ldots, \mathbf{x}_{|\mathcal{V}|}]^T\)

\[
p(\mathbf{x}) p(\mathbf{y}|\mathbf{x}) = \prod_{i \in \mathcal{V}} p(\mathbf{x}_i) \prod_{i \in \mathcal{V}} p(\mathbf{y}_{i,j}|\mathbf{x}_i, \mathbf{x}_j, \{i, j\} \in \mathcal{E}_{\text{Net}}) \triangleq f_i \quad \triangleq f_{i,j}
\]

- **Applications for distributed estimation:**
  - distributed power state estimation,
  - distributed localization/synchronization, etc.
BP Updating Equation in Linear Gaussian Model

- The general expression for message updating from variable node to factor node is

\[
\left[ C_{j \rightarrow f_{i,j}}^{(\ell)} \right]^{-1} = W_j^{-1} + \sum_{f_{k,j} \in B(j) \setminus f_{i,j}} \left[ C_{f_{k,j} \rightarrow j}^{(\ell-1)} \right]^{-1}
\]

\[
v_{j \rightarrow f_{i,j}}^{(\ell)} = C_{j \rightarrow f_{i,j}}^{(\ell)} \left[ \sum_{f_{k,j} \in B(j) \setminus f_{i,j}} \left[ C_{f_{k,j} \rightarrow j}^{(\ell-1)} \right]^{-1} v_{f_{k,j} \rightarrow j}^{(\ell-1)} \right]
\]

- The message from factor node to variable node is

\[
\left[ C_{f_{i,j} \rightarrow i}^{(\ell)} \right]^{-1} = A_{j,i}^T \left[ R_{i,j} + A_{i,j} C_{j \rightarrow f_{i,j}}^{(\ell)} A_{i,j}^T \right]^{-1} A_{j,i}.
\]

\[
v_{f_{i,j} \rightarrow i}^{(\ell)} = A_{j,i}^T \left[ R_{i,j} + A_{i,j} C_{j \rightarrow f_{i,j}}^{(\ell)} A_{i,j}^T \right]^{-1} \left( y_{i,j} - A_{i,j} v_{j \rightarrow f_{i,j}}^{(\ell)} \right)
\]

like a Kalman Gain

local innovation

cooperative innovation
Convergence Property

\[
[C_{f_{i,j}\rightarrow i}^{(\ell)}]^{-1} = A_{j,i}^T \left[ R_{i,j} + A_{i,j} \left[ W_j^{-1} + \sum_{f_k,j \in B(j) \setminus f_{i,j}} [C_{f_{k,j}\rightarrow j}^{(l-1)}]^{-1} \right]^{-1} A_{i,j} \right]^{-1} A_{j,i}
\]

Theorem 1. The matrix sequence \( \left\{ [C_{f_{i,j}\rightarrow i}^{(0)}]^{-1} \right\}_{l=0,1,...} \) converges to a unique positive definite matrix for any initial covariance matrix \( [C_{f_{i,j}\rightarrow i}^{(0)}]^{-1} \preceq 0 \) for all \( i,j \in V \)
Convergence Rate

- Convergence rate with respect to part metric:

  **Part (Birkhoff) Metric:** For arbitrary square matrices $X$ and $Y$ with the same dimension, if there exists $\alpha \geq 1$ such that $\alpha X \succeq Y \succeq \alpha^{-1}X$, $X$ and $Y$ are called the parts, and $d(X, Y) \triangleq \inf \{ \log \alpha : \alpha X \succeq Y \succeq \alpha^{-1}X, \alpha \geq 1 \}$ defines a metric called the part metric.

  $$d \left( C^{(\ell)}, C^* \right) < \beta^\ell d \left( C^{(0)}, C^* \right), \quad 0 < \beta < 1.$$  

**Theorem 2.** With the initial covariance matrix set to be an arbitrary p.s.d. matrix, i.e., $\left[ C_{fi,j \to i}^{(0)} \right]^{-1} \succeq 0$, the sequence $\left\{ \left[ C_{fi,j \to i}^{(0)} \right]^{-1} \right\}_{l=0}^{\infty}$ converges at a geometric rate with respect to the part metric.
Convergence Rate

- Convergence rate with respect to part metric:
  \[ d \left( C^{(\ell)}, C^* \right) < \beta^\ell d \left( C^{(0)}, C^* \right), \quad 0 < \beta < 1. \]

- From Part metric to monotone norm
  \[ \| J^{(\ell)} - J^* \| \leq \left( 2 \exp \left\{ d \left( J^{(\ell)}, J^* \right) \right\} - \exp \left\{ -d \left( J^{(\ell)}, J^* \right) \right\} - 1 \right) \min \left\{ \| J^{(\ell)} \|, \| J^* \| \right\}. \]

  \[ \| J^{(\ell)} - J^* \| < 2 \zeta \exp \left\{ c^\ell d_0 \right\}, \quad c < 1. \]

Theorem 3. With the initial covariance matrix set to be an arbitrary p.s.d. matrix, i.e., \( \left[ C_{f_n \to i}^{(0)} \right]^{-1} \geq 0 \), the sequence \( \left\{ C^{(\ell)} \right\}_{\ell=0,1,...} \) converges at a double exponential rate in terms of the monotone norm.
Convergence Property

- Updating equation

\[ v^{(\ell)}_{j \rightarrow f_{i,j}} = b_{j \rightarrow f_{i,j}} - C^*_{j \rightarrow f_{i,j}} \sum_{f_{k,j} \in B(j) \setminus f_{i,j}} C^*_{f_{k,j} \rightarrow j} M_{k,j} A_{j,k} v^{(\ell)}_{k \rightarrow f_{k,j}}, \]

\[ v^{(\ell)} = -Qv^{(\ell-1)} + b. \]

Theorem 4. The belief mean converges to the optimal value if and only if \( \rho(Q) < 1 \). The matrix \( Q \) satisfies \( v^{(\ell)} = Qv^{(\ell-1)} + b \) with \( v^{(\ell)} \) being a vector containing all the \( v^{(\ell)}_{j \rightarrow f_n} \).

- A *distributed* sufficient convergence condition

Theorem 5. The belief mean converges to the optimal value if the spectrum radius of block diagonal (each block’s dim. equals the corresponding variable’s dim.) of \( QQ^T \) is smaller than 1.
LGM subsumes Walk-Summable GMRF (I)

- A joint Gaussian distribution function can always be written as:

\[
p(x) \propto \exp \left( -\frac{1}{2} x^T J x + h^T x \right) = \prod_{i \in V} \exp \left( -\frac{1}{2} J_{i,i} x_i^2 + h_i x_i \right) \prod_{(i,j) \in E_{\text{MRF}}} \exp \left( -x_i J_{i,j} x_j \right) \]

\[\triangleq f_i(x_i) \triangleq f_{i,j}(x_{i,j})\]

\[
\rho(|I - J|) < 1 \iff \text{Walk-summable (BP converges on GMRF)}
\]

- Converged linear Gaussian model subsumes walk-summable GMRF.

Valid Models ($J > 0$)

Linear Gaussian Models

Walk-summable GMRF

Diagonally dominant
LGM subsumes Walk-Summable GMRF (II)

- Distributive law may NOT be exploited on graph with loops

\[ p(x) \propto \exp \left( -\frac{1}{2} x^T J x + h^T x \right) \]

\[ \propto \exp \left\{ -\frac{1}{2} x^T (J - \omega I) x - \frac{1}{2} \omega x^T x + h^T x \right\} \]

\[ = \exp \left\{ -\frac{1}{2} (V^T x)^T (V^T x) - \frac{1}{2} (\omega x^T x - 2h^T x) \right\} \]

\[ \propto \exp \left\{ -\frac{1}{2} \sum_{n=1}^{M} (V_{n,n_i} x_{n_i} + V_{n,n_j} x_{n_j})^2 - \frac{1}{2} \sum_{n=1}^{M} \omega(x_n - \frac{h_n}{\omega})^2 \right\} \]

- Let \( 0 < \omega < 1 \) and \( \omega \) is smaller than the minimum eigenvalue of \( I - |R| \).

\[ q(x) \propto \prod_{n=1}^{M} \mathcal{N}(x_n | \frac{1}{\omega} h_n, \frac{1}{\omega}) \prod_{n=1}^{M} \mathcal{N}(0 | V_{n,n_i} x_{n_i} + V_{n,n_j} x_{n_j}, 1) \cdot \]

Moallemi and Roy, TIT 2009
LGM subsumes Walk-Summable GMRF (III)

- H-Matrices

Definition $H$-Matrices (Boman et al. 2015): A matrix $X$ is an $H$-matrix if all eigenvalues of the matrix $H(X)$, where $[H(X)]_{i,i} = |X_{i,i}|$, and $[H(X)]_{i,j} = -|X_{i,j}|$ have positive real parts.

Factor width at most 2 factorization (Boman et al. 2015): A symmetric H-matrix $X$ that has non-negative diagonals can always be factorized as $X = VV^T$, where $V$ is a real matrix with each column of $V$ containing at most 2 non-zeros.

$(1 - \omega)I - R$ is an $H$ matrix.

$(1 - \omega)I - R = J - \omega I = VV^T$, where each column of $V$ contains at most 2 non-zeros.
Conclusion

- Existing convergence analysis of GMRF can not be used for distributed inference in linear Gaussian models.

- Belief covariance of BP in pairwise linear Gaussian models
  - converges to a unique positive definite matrix for arbitrary positive definite initial value.
  - converges at a doubly exponential rate.

- Necessary and sufficient convergence condition of belief mean
- The convergence condition of pairwise linear Gaussian models subsumes walk-summable GMRF.

Thank you!
\[ J = \begin{bmatrix}
1 & \frac{1}{3\sqrt{2}} & \frac{1}{\sqrt{3}} & \frac{\sqrt{2}}{3} \\
\frac{1}{3\sqrt{2}} & 1 & 0 & \frac{1}{3} \\
\frac{1}{\sqrt{3}} & 0 & 1 & \frac{1}{\sqrt{6}} \\
\frac{\sqrt{2}}{3} & \frac{1}{3} & \frac{1}{\sqrt{6}} & 1
\end{bmatrix}. \]

The eigenvalues of \( I - |R| \) to 4 decimal places are \(-0.0754\), \(0.9712\), \(1.4780\), and \(1.6262\). According to the walk-summable definition in, it is non walk-summable and the convergence condition is inconclusive as to whether Gaussian BP converges.
References