Convergence Analysis for Guassian Belief Propagation: Dynamic Behaviour of Marginal Covariances

Tianju Sui[1],Damián Marelli[2] and Minyue Fu[1,2]
[1]*School of Control Science and Engineering, Zhejiang University, China.*
[2]*School of Electrical Engineering and Computer Science, University of Newcastle, Australia.*

***Abstract*— Despite of its wide success in many distributed statistical learning applications, the well-known Gaussian *belief propagation* (BP) algorithm still lacks sufficient understanding at the theoretical level. This paper studies the convergence of Gaussian BP by analyzing the dynamic behaviour of the marginal covariances. We show, under a mild technical assumption, that the information matrices (i.e., the inverses of marginal covariances) are guaranteed to converge exponentially to positive-definite matrices. The convergence rate is explicitly characterized. This result is a key step to the understanding of the dynamic behaviour of the BP iterations.**

Keywords: Belief propagation, Gaussian belief propagation, statistical learning, distributed state estimation.

I. Introduction

Pearl's Belief Propagation, or *Belief Propagation* (BP) for short, is a well-celebrated algorithm for solving distributed optimization problems. Originally proposed by Pearl [1] in 1982, BP (also known as *sum-product message passing*), is a *message passing* algorithm for computing marginal PDFs on Bayesian networks (directed and acyclic graphs) and Markov random fields (undirected and cyclic graphs). Since its introduction, BP has been widely accepted as a powerful distributed algorithm in many scientific and engineering fields, including artificial intelligence, information theory, applied mathematics, signal processing and control systems [2]. Renowned applications of BP include low-density parity-check codes and turbo codes for digital communications [3]–[5], free energy approximation for statistical learning [2], combinatorial optimization [6] and computer vision [7], [8].

For acycilc graphs (i.e., graphs without loops), it is known that BP converges in a finite number of iterations, and the correct marginals will be produced [1]. For cyclic graphs (i.e., graphs with loops), BP is not guaranteed to converge in general, and even if it does, it does not calculate the correct marginals. Nevertheless, the wonderful and mysterious feature of BP is that for most applications, BP delivers amazingly good approximations for the marginals, despite the existence of loops [9]–[11].

Gaussian BP is the BP algorithm specialized to Gaussian distributions. The algorithm computes iteratively the mean and variance (or covariance) of each marginal. Gaussian BP has been successfully applied in low complexity detection and estimation problems arising in communication systems, state estimation for large-scale linear system, sparse Bayesian learning, estimation in Gaussian graphical model, distributed beam forming, inter-cell interference mitigation, distributed synchronization and localization in wireless sensor networks, distributed energy efficient self-deployment in mobile sensor networks, distributed rate control in Ad Hoc networks, distributed network utility maximization, and large-scale sparse Bayesian learning [12]–[14].

BP's excellent performances have inspired *many* researchers over the last 20 years to study its theoretical properties. The fundamental questions are: 1) For a cyclic network graph, under what conditions will BP iterations converge? 2) Upon convergence, how accurate are the approximate marginals? Viewing BP iterations as a dynamic process, these questions amount to its stability and performance analysis. For a general cyclic graph, [15]–[19] studied the convergence condition for BP and [16], [19], [20] worked on the accuracy analysis. However, these references only give partial answers, and the conditions given in these references are mainly applicable to discrete random variables only. Several conditions ensuring the convergence of the marginals under a designated initialization set have been proposed [21]–[25]. But several major drawbacks exist. Firstly, Convergence conditions are too difficult to check. For example, [24] requires to run a semi-definite programming (SDP) to check if the variances offered by BP converge. The convergence condition for the mean in [25] requires the evaluation of the spectral radius of an infinite dimensional matrix, which is impossible in practice. Secondly, the convergence analysis is done only for scalar systems (i.e., the state of each node is a scalar[1]). Finally, although [21] discusses the accuracy of BP for vector systems, estimation errors are not quantified.

This paper aims at taking the first step towards answering the above questions. More specifically, we study the convergence of the information matrices (the inverse of the marginal covariance matrices) of the Gaussian BP iterations under the general setting of a cyclic graph and vector system (i.e., each node has a random vector). Our study is conducted through the problem of distributed state estimation for a networked linear system with additive Gaussian noises, using the weighted least-squares (WLS) criterion. It turns out that Gaussian BP can be described through this setting, and this equivalence gives strong motivation about applying Gaussian BP in distributed state estimation. Viewing BP iterations as a dynamic process for the marginal mean and error covari-

[1]Since the state components for each node are not independent for a vector system, results for scalar systems are not applicable to vector systems.

ance for each node, we show that, under a mild technical condition, the information matrices always converge and their steady state are bounded from both below and above. Moreover, we show that the convergence rate to the steady state is exponential and this rate is explicitly characterized.

II. Problem Formulation

A. The general BP algorithm

The BP algorithm is concerned with a system represented by a *bipartite graph* with I *variable nodes* and V *factor nodes*, as depicted in Fig. 1. Each variable node i is associated with a random variable $x_i \in \mathbf{R}^{n_i}$ and each factor node v is connected to a subset of variable nodes, $\mathcal{F}_v \subset \{1, 2, \ldots, I\}$. Denoting the joint (or global) variable by $X = \{x_i : i = 1, 2, \ldots, I\}$, it is assumed that its joint PDF $f(X)$ can be expressed in a factor form:

$$f(X) = \prod_{v=1}^{V} f_v\left(X_v\right),$$

where $X_v = \{x_i : i \in \mathcal{F}_v\}, v = 1, 2, \ldots, V$. Each $f_v(X_v)$ represents a piece of partial "knowledge" about X.

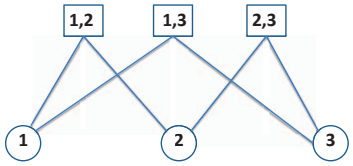

Fig. 1. Bipartite graph: circles = variable nodes; squares = factor nodes

The goal of BP is to compute, at each node i, the marginal $g_i(x_i)$ of $f(X)$, which is defined by

$$g_i(x_i) = \int f(X) d(X \setminus x_i), \tag{1}$$

where $X \setminus x_i$ is the set obtained from X by removing x_i. The algorithm resorts to iterative computation and local communication between connected variable nodes and factor nodes. More specifically, the algorithm starts by each factor node v sending to each variable node $i \in \mathcal{F}_v$ the following marginal pdf (called *message*)

$$m_{v\to i}^{(0)}(x_i) = \int f_v(X_v) d(X_v \setminus x_i). \tag{2}$$

Then, at each iteration $k = 1, 2, \ldots$, each variable node i sends to be every connected factor node v the following message:

$$m_{i\to v}^{(k)}(x_i) = \prod_{w \in \mathcal{N}_i \setminus v} m_{w\to i}^{(k)}(x_i), \tag{3}$$

where $\mathcal{N}_i$ is the set of factor nodes connected to node i. Similarly, each factor node v sends to every connected variable node i the following message:

$$m_{v\to i}^{(k)}(x_i) = \int f_v(X_v) \prod_{j \in \mathcal{F}_v \setminus i} m_{j\to v}^{(k-1)}(x_j) d(X_v \setminus x_i). \tag{4}$$

The desired marginal at node i and iteration N is estimated:

$$g_i^{(N)}(x_i) = \prod_{w \in \mathcal{N}_i} m_{w\to i}^{(N)}(x_i), \tag{5}$$

modulo a constant scalar to make its integral equal 1.

B. Gaussian BP and Distributed WLS Estimation

Gaussian BP refers to the special but extremely important case of BP where the functions $f_v(X_v)$ are Gaussian distributions. Following the tradition in the Gaussian BP literature [18], [19], [21], [24], [25], each factor node v is assumed to be connected to either one or two variable nodes, i.e, each $\mathcal{F}_v$ contains at most two elements[2].

The significance of Gaussian BP lies in its connection with distributed WLS estimation. Consider a system with I unknown variables $x_1, x_2, \ldots, x_I$. Associated with the system are two kinds of measurements, the so-called *self measurement* for node i,

$$z_i = C_i x_i + w_i, \tag{6}$$

and (pair-wise) *Joint measurement* between nodes i and j,

$$z_{i,j} = C_{i,j} x_i + C_{j,i} x_j + w_{i,j}. \tag{7}$$

In the above, the matrices $C_i, C_{i,j}$ and $C_{j,i}$ are known; w_i and $w_{i,j}$ are independent measurement noises with zero-mean Gaussian variables with known covariances $R_i > 0$ and $R_{i,j} > 0$, respectively. Note that 1) the factor node (i, j) is unordered, i.e., $(i, j) = (j, i)$; 2) $z_{i,j} = z_{j,i}$ and $w_{i,j} = w_{j,i}$; 3) It is not necessary for all variable nodes to have self measurements or all variable node pairs to have joint measurements. In fact, joint measurements are typically sparse for large graphs.

The problem of distributed WLS estimation is to compute the maximum likelihood (ML) estimate for each x_i and the corresponding estimation error covariance using a fully distributed algorithm. It is clear that the likelihood functions given by the self and joint measurements are, respectively,

$$f_i(x_i) = p\left(z_i | x_i\right) \sim \mathcal{N}(z_i - C_i x_i, R_i), \tag{8}$$

$$\begin{aligned} f_{i,j}(x_i, x_j) &= p\left(z_{i,j} | x_i, x_j\right) \\ &\sim \mathcal{N}(z_{i,j} - C_{i,j} x_i - C_{j,i} x_j, R_{ij}), \end{aligned} \tag{9}$$

where $\mathcal{N}(\mu, \Sigma)$ stands for a Gaussian PDF with mean μ and covariance Σ. The joint likelihood function for X becomes

$$f(X) = \prod_i f_i(x_i) \prod_{(i,j)} f_{i,j}(x_i, x_j). \tag{10}$$

Therefore, the maximum likelihood function for each x_i is given by (1), which is exactly the task of BP.

Returning to the BP iterations (2)-(4), it is straightforward to see that, at each iteration k, a factor node with a single variable node i will not get any update and will always send the same message $f_i(x_i)$ to node x_i. Hence, we can remove these factor nodes from the iterations (3)-(4), and directly add

[2]It suffices to consider this type of graphs because any Bayesian network can be converted into an undirected graph with pairwise cliques by adding cluster nodes for all parent nodes that share a common child; see [21].

the factor $f_i(x_i)$ into (4). Subsequently, the iterations (3)-(4) can be rewritten as

$$m^{(0)}_{i,j\to i}(x_i) = \int p\left(z_{i,j}|x_i,x_j\right) dx_j \tag{11}$$

$$m^{(k)}_{i\to i,j}\left(x_i\right) = p\left(z_i|x_i\right) \prod_{w\in\mathcal{N}_i\setminus j} m^{(k)}_{i,w\to i}\left(x_i\right), \tag{12}$$

$$m^{(k)}_{i,j\to i}\left(x_i\right) = \int p\left(z_{i,j}|x_i,x_j\right) m^{(k-1)}_{j\to i,j}\left(x_j\right) dx_j. \tag{13}$$

Also, the marginal estimate (5) at iteration N becomes

$$\begin{aligned} g^{(N)}_i(x_i) &= p\left(z_i|x_i\right) \prod_{k\in\mathcal{N}_i} m^{(N)}_{i,k\to i}\left(x_i\right) \\ &\sim \mathcal{N}(\hat{x}_i(N), \Sigma_i(N)) \end{aligned} \tag{14}$$

for some mean $\hat{x}_i(N)$ and covariance $\Sigma_i(N)$.

Defining the *information vector* and *information matrix*:

$$\alpha_i(N) = Q_i(N)\hat{x}_i(N); \quad Q_i(N) = \Sigma_i^{-1}(N), \tag{15}$$

which we call *information parameters* collectively, it is clear that they will fully characterize the marginal estimate $g^{(N)}_i(x_i)$. Similarly, $m_{i,j\to i}\left(x_i\right)$ and $m_{i\to i,j}\left(x_i\right)$ are also Gaussian distributed, thus they can also be fully characterized by their information parameters $\alpha_{i,j\to i}(N)$, $Q_{i,j\to i}(N)$, $\alpha_{i\to i,j}(N)$ and $Q_{i\to i,j}(N)$. Using this representation, the recursions (12)-(13) can be implemented by updating the information parameters, which is done by Algorithm 1.

Denote $\Omega_{i,j} = Q_{i\to i,j}(1)$, $i = 1, 2, \ldots, I$ and $j \in \mathcal{N}_i$, for simplicity. It is easy to verify that

$$\Omega_{i,j} = C_i^T R_i^{-1} C_i + \sum_{w\in\mathcal{N}_i\setminus j} C_{i,w}^T R_{i,w}^{-1} C_{i,w}.$$

It follows that $\Omega_{i,j} > 0$, for every (i,j), is needed for Algorithm 1. Throughout this paper, we will make the following slightly stronger assumption than the above.

Assumption 1: For all $i = 1, 2, \ldots, I$ and $j \in \mathcal{N}_i$,

$$\Omega_{i,j} = C_i^T R_i^{-1} C_i + \sum_{w\in\mathcal{N}_i\setminus j} C_{i,w}^T R_{i,w}^{-1} C_{i,w} > C_{i,j}^T R_{i.j}^{-1} C_{i,j}.$$

Remark 1: Roughly speaking, (1) means that, for each node i, the information contribution from any single neighbouring node j (i.e., $C_{i,j}^T R_{i,j}^{-1} C_{i,j}$) is strictly smaller than that from node i (i.e., $C_i^T R_i^{-1} C_i$) and all other neighbouring nodes w (i.e., $C_{i,w}^T R_{i,w}^{-1} C_{i,w}$).

C. Problem Statement

The canonical graph $\mathcal{G}$ has a node associated with each variable node $i = 1, \ldots, I$. Also, nodes i and j are connected by an edge if there exists a factor node (i,j), i.e., $j \in \mathcal{N}_i$. It is assumed throughout the paper that $\mathcal{G}$ is a connected undirected graph.

It is well known that BP (thus Gaussian BP) converges to the correct marginals in a finite number of iterations when $\mathcal{G}$ is acyclic [1]. The fundamental challenge in the study of BP is to understand how the algorithm performs for cyclic graphs. As mentioned in Section I, the goal of this paper is to provide conditions to guarantee the convergence of the marginal covariances in Gaussian BP iterations when the induced bipartite graph is cyclic.

Algorithm 1 Gaussian BP for Distributed WLS Estimation

1) **Initialization:** At time $k = 0$, factor node (i,j) sends to each connected variable node i:

$$\alpha_{i,j\to i}(0) = C_{i,j}^T R_{i,j}^{-1} z_{i,j}, \tag{16}$$

$$\tag{17}$$

$$Q_{i,j\to i}(0) = C_{i,j}^T R_{i,j}^{-1} C_{i,j}. \tag{18}$$

2) **Main loop:** At time $k = 1, 2, \cdots$, do:

2.1) Each variable node i computes

$$\begin{aligned} \alpha_i(k) &= C_i^T R_i^{-1} z_i + \sum_{j\in\mathcal{N}_i} \alpha_{i,j\to i}(k-1), \\ Q_i(k) &= C_i^T R_i^{-1} C_i + \sum_{j\in\mathcal{N}_i} Q_{i,j\to i}(k-1), \end{aligned} \tag{19}$$

and (if required at this iteration)

$$\begin{aligned} \hat{x}_i(k) &= Q_i^{-1}(k)\alpha_i(k), \\ \Sigma_i(k) &= Q_i^{-1}(k). \end{aligned} \tag{20}$$

2.2) Each variable node i sends to factor node (i,j) with $j \in \mathcal{N}_i$:

$$\begin{aligned} \alpha_{i\to i,j}(k) &= \alpha_i(k) - \alpha_{i,j\to i}(k-1), \\ Q_{i\to i,j}(k) &= Q_i(k) - Q_{i,j\to i}(k-1), \end{aligned} \tag{21}$$

2.3) Each factor node (i,j) sends to each connected variable node j:

$$\begin{aligned} \alpha_{i,j\to j}(k) &= C_{j,i}^T R_{i,j\to j}^{-1}(k) z_{i,j\to j}(k), \\ Q_{i,j\to j}(k) &= C_{j,i}^T R_{i,j\to j}^{-1}(k) C_{j,i}, \end{aligned} \tag{22}$$

where

$$\begin{aligned} z_{i,j\to j}(k) &= z_{i,j} - C_{i,j} Q_{i\to i,j}^{-1}(k)\alpha_{i\to i,j}(k), \\ R_{i,j\to j}(k) &= R_{i,j} + C_{i,j} Q_{i\to i,j}^{-1}(k) C_{i,j}^T. \end{aligned} \tag{23}$$

III. Convergence of Information Matrices

In this section, we provide our main result which shows that the information matrices $Q_i(k)$ always converge exponentially to a positive definite matrix, under Assumption 1. The rate of convergence is also characterized. This result is fundamental, not only on its own, but also for carrying out further analysis on the convergence of the BP estimates.

To derive the main result, we will first provide two lemmas. The first lemma characterises the monotonicity properties for $Q_{i\to i,j}(k)$, $Q_{i,j\to j}(k)$ and $R_{i,j\to j}(k)$. The second lemma states the boundedness of $R_{i,j\to j}(k)$ and positive definiteness of $Q_{i\to i,j}(k)$ at steady state. These results then lead to the first theorem which characterizes the exponential convergence of $Q_{i\to i,j}(k)$. This will then lead to the second theorem (our main result), which characterizes the exponential convergence of $Q_i(k)$.

Lemma 1: . The following monotonicity properties hold

for Gaussian BP: For any $k \in \mathbb{N}$, $1 \leq i \leq I$ and $j \in \mathcal{N}_i$,

$$\begin{aligned} Q_{i\to i,j}(k+1) &\leq Q_{i\to i,j}(k); \\ Q_{i,j\to j}(k+1) &\leq Q_{i,j\to j}(k); \\ R_{i,j\to j}(k+1) &\geq R_{i,j\to j}(k). \end{aligned} \tag{24}$$

In particular, $Q_{i\to i,j}(k) \leq \Omega_{i,j}$ for all $k \geq 1$.

Lemma 2: The following limits hold for Gaussian BP under Assumption 1: For every $1 \leq i \leq I$ and $j \in \mathcal{N}_i$, we have

$$\begin{aligned} Q_{i\to i,j}(\infty) &= \lim_{k\to\infty} Q_{i\to i,j}(k) > 0; \\ R_{i,j\to j}(\infty) &= \lim_{k\to\infty} R_{i,j\to j}(k) < \infty. \end{aligned}$$

Next, we give the main result on convergence. Define

$$\begin{aligned} \Delta Q_{i\to i,j}(k) &= Q_{i\to i,j}^{-1/2}(\infty) Q_{i\to i,j}(k) Q_{i\to i,j}^{-1/2}(\infty) - I; \\ \Delta Q_{i,j\to j}(k) &= Q_{i,j\to j}^{-1/2}(\infty) Q_{i,j\to j}(k) Q_{i,j\to j}^{-1/2}(\infty) - I; \\ \Delta R_{i,j\to j}(k) &= R_{i,j\to j}^{-1/2}(\infty) R_{i,j\to j}(k) R_{i,j\to j}^{-1/2}(\infty) - I. \end{aligned}$$

Also, let constants $\rho > 0$ and $\alpha > 0$ be defined as follows:

$$\rho = \max_{i,j} \|R_{i,j\to j}^{-1/2}(\infty) C_{i,j} Q_{i\to i,j}^{-1}(\infty) C_{i,j}^T R_{i,j\to j}^{-1/2}(\infty)\|, \tag{25}$$

$$\alpha = \max_{i,j} \|Q_{i\to i,j}^{-1/2}(\infty) \Omega_{i,j} Q_{i\to i,j}^{-1/2}(\infty) - I\|. \tag{26}$$

Note that $\rho < 1$ follows from $R_{i,j} > 0$ and

$$R_{i,j\to j}(\infty) = R_{i,j} + C_{i,j} Q_{i\to i,j}^{-1}(\infty) C_{i,j}^T.$$

Theorem 1: Suppose Gaussian BP is applied under Assumption 1. Then, we have, for every node i, its connecting node j and all $k \in \mathbb{N}$, we have

$$0 \leq \Delta Q_{i\to i,j}(k) \leq \alpha \rho^{k-1} I. \tag{27}$$

Since ultimately we are only interested in the information matrices $Q_i(k)$, we get the following result from Theorem 1.

Theorem 2: Under Assumption 1, it holds that $Q_i(k) \to Q_i(\infty) > 0$, as $k \to \infty$, for every node i of $\mathcal{G}$ and all $k \in \mathcal{N}$. Moreover, by defining

$$\Delta Q_i(k) = Q_i^{-1/2}(\infty) Q_i(k) Q_i^{-1/2}(\infty) - I,$$

it holds, for very node i of $\mathcal{G}$ and all $k \in \mathcal{N}$, that

$$0 \leq \Delta Q_i(k) \leq \alpha \rho^{k-1} I. \tag{28}$$

Remark 2: The above result shows that the information matrices converge to their steady state exponentially. It is straightforward to see that their inverses, i.e., marginal covariances, will enjoy the same property. The details are not provided due to space limit.

IV. Conclusion

In this paper, we have studied the dynamic behaviour of the Gaussian BP iterations through the convergence analysis of the information matrices. We believe that our result is a crucial for the full understanding of the dynamic behaviour of Gaussian BP iterations. In particular, this result will be fundamental for studying the convergence of the BP estimates and their accuracies, which will be our next task.

References

[1] J. Pearl, *Probabilistic Reasoning in Intelligent Systems*. Morgan Kaufman, 1988.

[2] A. Braunstein, R. Mézard, and R. Zecchina, "Survey propagation: An algorithm for satisfiability," *Random Structures & Algorithms*, vol. 27, no. 2, pp. 201–226, 2005.

[3] R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng, "Turbo decoding as an instance of pearl's belief propagation algorithm," *Selected Areas in Communications, IEEE Journal on*, vol. 16, no. 2, pp. 140–152, 1998.

[4] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-codes," *Communications, IEEE Transactions on*, vol. 44, no. 10, pp. 1261–1271, 1996.

[5] M. Fu, "Stochastic analysis of turbo decoding," *IEEE Transactions on Information Theory*, vol. 54, no. 1, pp. 81–100, 2005.

[6] A. Braunstein and R. Zecchina, "Survey propagation as local equilibrium equations," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2004, no. 06, p. P06007, 2004.

[7] J. Sun, N.-N. Zheng, and H.-Y. Shum, "Stereo matching using belief propagation," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 25, no. 7, pp. 787–800, 2003.

[8] K. Tanaka, "Statistical-mechanical approach to image processing," *Journal of Physics A: Mathematical and General*, vol. 35, no. 37, p. R81, 2002.

[9] B. J. Frey, "Graphical models for pattern classification, data compression and channel coding," 1998.

[10] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: An empirical study," in *Proceedings of the Fifteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1999, pp. 467–475.

[11] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, "Learning low-level vision," *International journal of computer vision*, vol. 40, no. 1, pp. 25–47, 2000.

[12] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multi–user detection," *arXiv preprint cs/0510044*, 2005.

[13] Q. Guo and D. Huang, "Em-based joint channel estimation and detection for frequency selective channels using Gaussian message passing," *Signal Processing, IEEE Transactions on*, vol. 59, no. 8, pp. 4030–4035, 2011.

[14] X. Tai, Z. Lin, M. Fu, and Y. Sun, "A new distributed state estimation technique for power networks," in *American Control Conference*, 2013, pp. 3338–3343.

[15] S. C. Tatikonda and M. I. Jordan, "Loopy belief propagation and gibbs measures," in *Proceedings of the Eighteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 2002, pp. 493–500.

[16] S. C. Tatikonda, "Convergence of the sum-product algorithm," in *Information Theory Workshop, 2003. Proceedings. 2003 IEEE*. IEEE, 2003, pp. 222–225.

[17] A. T. Ihler, J. W. Fisher, and A. S. Willsky, "Message errors in belief propagation," in *Advances in Neural Information Processing Systems*, 2004, pp. 609–616.

[18] J. M. Mooij and H. J. Kappen, "Sufficient conditions for convergence of the sum–product algorithm," *Information Theory, IEEE Transactions on*, vol. 53, no. 12, pp. 4422–4437, 2007.

[19] A. T. Ihler, J. Iii, and A. S. Willsky, "Loopy belief propagation: Convergence and effects of message errors," in *Journal of Machine Learning Research*, 2005, pp. 905–936.

[20] T. Heskes, "On the uniqueness of loopy belief propagation fixed points," *Neural Computation*, vol. 16, no. 11, pp. 2379–2413, 2004.

[21] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in gaussian graphical models of arbitrary topology," *Neural computation*, vol. 13, no. 10, pp. 2173–2200, 2001.

[22] C. C. Moallemi and B. Van Roy, "Convergence of min-sum message-passing for convex optimization," *IEEE Transactions on Information Theory*, vol. 56, no. 4, pp. 2041–2050, 2010.

[23] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *The Journal of Machine Learning Research*, vol. 7, pp. 2031–2064, 2006.

[24] Q. Su and Y.-C. Wu, "Convergence analysis of the variance in Gaussian belief propagation," *IEEE Transactions on Signal Processing*, vol. 62, pp. 5119–5131, 2014.

[25] ——, "On convergence conditions of Gaussian belief propagation," *IEEE Transactions on Signal Processing*, vol. 63, pp. 1144–1155, 2015.