Convergence Rate Analysis of Gaussian Belief Propagation for Markov Networks

Zhaorong Zhang and Minyue Fu, *Fellow, IEEE*

Abstract—Gaussian belief propagation algorithm (GaBP) is one of the most important distributed algorithms in signal processing and statistical learning involving Markov networks. It is well known that the algorithm correctly computes marginal density functions from a high dimensional joint density function over a Markov network in a finite number of iterations when the underlying Gaussian graph is acyclic. It is also known more recently that the algorithm produces correct marginal means asymptotically for cyclic Gaussian graphs under the condition of walk summability (or generalised diagonal dominance). This paper extends this convergence result further by showing that the convergence is exponential under the generalised diagonal dominance condition, and provides a simple bound for the convergence rate. Our results are derived by combining the known walk summability approach for asymptotic convergence analysis with the control systems approach for stability analysis.

Index Terms—Belief propagation, distributed algorithm, distributed estimation, Gaussian belief propagation, Markov networks.

I. INTRODUCTION

BELIEF propagation (BP) algorithm is a well-celebrated distributed algorithm for Markov networks that has been widely utilized in many disciplines, ranging from statistical learning and artificial intelligence to distributed estimation, distributed optimisation, networked control and digital communications [1]–[13].

Initially introduced by Pearl [1] in 1988, the BP algorithm is also known as Pearl's algorithm, message-passing algorithm and sum-product algorithm. It is designed to compute the marginal probability densities of random variables from the joint probability density function over a large Markov network with sparse connections among individual random variables. The significance of the algorithm stems from the facts that it

is fully distributed (i.e., only local information is needed for iteration computation) and that a wide range of application problems can be formulated as a BP problem. It is well known that the BP algorithm produces correct marginal probability densities in a finite number of iterations when the underlying graph for the joint density function is acyclic (i.e., no cycles or loops). But the properties of the algorithm for cyclic (loopy) graphs have been a major research topic over several decades.

The Gaussian BP algorithm (GaBP), a special version of the BP algorithm for Markov networks with Gaussian distributions (also known as Gaussian graphical model), has received special attention for the study of its convergence properties. In [2], it was shown that GaBP produces asymptotically the correct marginal means under the assumption that the joint information matrix is diagonal dominance. It was relaxed in [3] that the same asymptotic convergence holds when the joint information matrix is walk-summable, which is equivalent to the condition of generalised diagonal dominance. The convergence property in [3] was generalised by [5] to allow an alternative decomposition of the optimizing function and more flexible message initialization. In [7], [8], necessary and sufficient conditions for asymptotic convergence of GaBP are studied. In [5], convergence properties of the BP algorithm for convex optimisation (including quadratic optimisation) are studied. A pertinent result in [5] is a bound on the convergence rate of the BP algorithm under a scaled diagonal dominance assumption and a particular decomposition of optimizing function. Many other interesting properties of GaBP can be found in, e.g., [14]–[16] and the references therein.

The purpose of this paper is to study the convergence rate of GaBP. Under the generalised diagonal dominance condition, we provide a simple bound for the exponential convergence rate of the marginal means. This bound is simply the spectral radius of the matrix related to the information matrix. This bound also coincides with that in [5] but under weaker conditions (see details later). Our results are derived by combining the walk summability approach in [3] for asymptotic convergence analysis with the control systems approach for stability analysis.

In the rest of the paper, we introduce GaBP in Section II and discuss the walk summability condition in Section III, followed by convergence rate analysis in Section IV, illustrating examples in Section V and conclusions in Section VI.

II. PROBLEM FORMULATION

A Gaussian graphical model is a Markov network with Gaussian distributions, characterised by an undirected graph

Z. R. Zhang is with the School of Electrical Engineering and Computer Science, The University of Newcastle, University Drive, NSW 2308, Australia (e-mail: zhaorong.zhang@uon.edu.au).

M. Y. Fu is with the School of Electrical Engineering and Computer Science, The University of Newcastle, University Drive, NSW 2308, Australia, and also with the School of Automation, Guangdong University of Technology, and Guangdong Key Laboratory of Intelligent Decision and Cooperative Control, Guangzhou 510006, China (e-mail: minyue.fu@newcastle.edu.au).
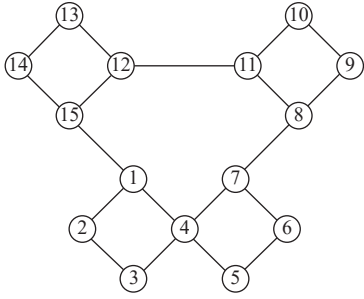
Fig. 1.   An example of Markov network.

$\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$, where $\mathcal{V} = \{1, 2, \ldots, n\}$ represents the set of nodes and $\mathcal{E}$ is the set of edges (or unordered pairs $\{i, j\} \subset \mathcal{V}$), with each node $i \in \mathcal{V}$ being associated with a random variable x_i. Fig. 1 shows an example of Markov network. The joint probability density for $x = \text{col}\{x_1, x_2, \ldots, x_n\}$ is given by the following Gaussian density function

$$p(x) \propto \exp\left[-\frac{1}{2}x^T A x + b^T x\right] \tag{1}$$

where $A = \{a_{ij}\}$ is a sparse information matrix with $a_{ij} = 0$ for all $\{i, j\} \notin \mathcal{E}$, which is a symmetric and positive definite matrix, and b is the potential vector. It is straightforward to verify that the mean vector $\mu = \mathbb{E}\{x\}$ and covariance matrix $P = \mathbb{E}\{(x - \mu)(x - \mu)^T\}$ are given respectively by

$$\mu = A^{-1}b, \quad P = A^{-1}. \tag{2}$$

The problem of concern is for each node $i \in \mathcal{V}$ to compute, in a distributed fashion, the marginal density function $p_i(x_i)$ of x_i, defined by

$$p_i(x_i) = \int_{x_{-i}} p(x) dx_{-i} \tag{3}$$

where x_{-i} denotes the vector x with the component x_i removed. It is well known that this amounts to computing the marginal mean μ_i (the ith term of μ) and marginal variance p_{ii} (the ith diagonal term of P).

Using the Gaussian graphical model, $p(x)$ can be factorized into

$$p(x) \propto \prod_{i \in \mathcal{V}} \phi_i(x_i) \prod_{\{i,j\} \in \mathcal{E}} \phi(x_i, x_j) \tag{4}$$

with

$$\phi_i(x_i) = \exp\left[-\frac{1}{2}a_{ii}x_i^2 + b_i x_i\right]$$
$$\phi_{ij}(x_i, x_j) = \exp[-a_{ij}x_i x_j]. \tag{5}$$

The BP algorithm is an iterative message-passing algorithm for computing $p_i(x_i)$. In each iteration k, each node $i \in \mathcal{V}$ computes and transmits to each node $j \in \mathcal{N}_i$ (the set of neighbouring nodes of i) the message $m_{i \to j}^{(k)}(x_j)$

$$m_{i \to j}^{(k)}(x_j) = \int \phi_{ij}(x_i, x_j)\phi_i(x_i) \prod_{v \in \mathcal{N}_i \setminus j} m_{v \to i}^{(k-1)}(x_i) dx_i,$$

where $m_{v \to i}^{(k-1)}$ is the message node i receives from its neighbouring node v in iteration $k - 1$. This results in the marginal density $p_i(x_i)$ to be estimated in iteration k as

$$p_i^{(k)}(x_i) \propto \phi_i(x_i) \prod_{v \in \mathcal{N}_i} m_{v \to i}^{(k-1)}(x_i).$$

For a Gaussian graphical model, the message $m_{i \to j}^{(k)}(x_j)$ can be expressed as

$$m_{i \to j}^{(k)}(x_j) \propto \exp\left[-\frac{1}{2}a_{i \to j}(k)x_j^2 + b_{i \to j}(k)x_j\right].$$

This results in GaBP below

$$a_{i \to j}(k) = -\frac{a_{ij}a_{ji}}{a_{i \to j}(k)}, \quad b_{i \to j}(k) = -\frac{a_{ji}b_{i \to j}(k)}{a_{i \to j}(k)}$$

with

$$a_{i \to j}(k) = a_{ii} + \sum_{v \in \mathcal{N}_i \setminus j} a_{v \to i}(k - 1)$$

$$b_{i \to j}(k) = b_i + \sum_{v \in \mathcal{N}_i \setminus j} b_{v \to i}(k - 1).$$

The initialization is done by taking $a_{i \to j}(0) = a_{ii}$ and $b_{i \to j}(0) = b_i$. The marginal mean and marginal variance of $p_i^{(k)}(x_i)$ are then given by, respectively

$$\mu_i(k) = \frac{b_i + \sum_{v \in \mathcal{N}_i} b_{v \to i}(k - 1)}{a_{ii} + \sum_{v \in \mathcal{N}_i} a_{v \to i}(k - 1)} \tag{6}$$

$$p_{ii}(k) = \frac{1}{a_{ii} + \sum_{v \in \mathcal{N}_i} a_{v \to i}(k - 1)}. \tag{7}$$

It is well known [2] that, when the graph $\mathcal{G}$ is acyclic, GaBP converges in d iterations with $\mu_i(k) = \mu_i$ and $p_{ii}(k) = p_{ii}$ for all i, where d is the diameter of $\mathcal{G}$ (i.e., the largest distance between any two nodes in $\mathcal{G}$). Actually, for each node i, d_i iterations are sufficient to yield the above convergence, where d_i is the largest distance from any node in $\mathcal{G}$ to node i [2]. (The distance of two nodes is the minimum path length between the nodes.)

For cyclic (or loopy) graphs, GaBP produces the correct marginal means asymptotically under certain conditions. In particular, it has been established in [3] that $\mu_i(k)$ converges to u_i for all i asymptotically under the so-called walk summability condition. This condition is also known to be equivalent to requiring the matrix A to be generalised diagonally dominant [17].

The goal of this paper is to study the convergence rate of GaBP under the same walk summability condition.

III. WALK SUMMABILITY

Walk-sum analysis is an elegant approach introduced in [3] (and their earlier references thereof) for studying the convergence of GaBP. Here we provide a quick summary of this approach.

Given a matrix $R = \{r_{ij}\} \in \mathbb{R}^{n \times n}$ and its induced graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, a walk w in the graph is a node sequence

$$w = (w_0, w_1, \ldots, w_l), \ \forall \, w_i \in \mathcal{V}, \ (w_i, w_{i+1}) \in \mathcal{E} \tag{8}$$

and its length is l. The weight of the walk is defined to be

$$\phi(w) = \prod_{i=0}^{l-1} r_{w_i w_{i+1}}. \tag{9}$$

As a convention, a single node $i \in \mathcal{V}$ is regarded as a special (zero-length) walk with its weight $\phi(i) = 1$. A walk w from node i to j is also denoted by $w : i \rightarrow j$, and such a walk with length l is denoted by $w : i \overset{l}{\rightarrow} j$. The set of all walks from node i to node j is denoted by $\{i \rightarrow j\}$, and the set of all length l walks from node i to node j is denoted by $\{i \overset{l}{\rightarrow} j\}$. The walk-sum of a set of weights W is denoted by $\phi(W) = \sum_{w \in W} \phi(w)$.

The importance of walk sums is revealed in the relationship that (i, j)th element of matrix R^l is equal to

$$(R^l)_{ij} = \sum_{w_1,\ldots,w_{l-1}} r_{iw_1} r_{w_1 w_2} \ldots r_{w_{l-1} j} = \sum_{w:i \overset{l}{\rightarrow} j} \phi(w) \quad (10)$$

which can be verified by matrix multiplication. Now we give the definition of walk summability [3].

Definition 1: A matrix $A \in \mathbb{R}^{n \times n}$ with $a_{ii} = 1$ for all i is said to be walk-summable if all the walk-sums $\phi(\{i \rightarrow j\})$ converge absolutely, i.e., $\sum_{w:i \rightarrow j} |\phi(w)|$ converges for all i, j. This is the same as the unordered sum $\sum_{w:i \rightarrow j} \phi(w)$ is well defined (i.e., converges to the same value for every possible summation order) for all i, j. Further, a linear system $Ax = b$ is said to be walk-summable if A is walk-summable.

Defining $R = \{r_{ij}\} = I - A$ and $\bar{R} = \{|r_{ij}|\}$, the following properties are known for walk-summable systems [3].

Lemma 1 [3]: The following conditions are equivalent.
1) $A \in \mathbb{R}^{n \times n}$ with $a_{ii} = 1$ for all i is walk-summable;
2) $\sum_l \bar{R}^l$ converges;
3) $\rho(\bar{R}) < 1$;
4) $I - \bar{R} > 0$.
In addition, $\rho(R) \le \rho(\bar{R})$.

Using the walk-sum interpretation, the Gaussian variance P and mean μ in (2) can be expressed by walk sums under the assumption of walk summability [3]. More specifically, using Lemma 1, $\rho(R) \le \rho(\bar{R}) < 1$ which implies that

$$P = A^{-1} = (I - R)^{-1} = \sum_{l=0}^{\infty} R^l \quad (11)$$

and (10), we get

$$P_{ij} = \sum_{l=0}^{\infty} (R^l)_{ij} = \sum_{l=0}^{\infty} \sum_{w:i \overset{l}{\rightarrow} j} \phi(w) = \sum_{w:i \rightarrow j} \phi(w). \quad (12)$$

Similarly, using Lemma 1 again, we get $\mu = A^{-1} b = \sum_{l=0}^{\infty} R^l b$, which in turn implies

$$\mu_i = \left(\sum_{l=0}^{\infty} R^l b \right)_i$$
$$= \sum_{j=1}^{n} \sum_{l=0}^{\infty} (R^l)_{ij} b_j$$
$$= \sum_{j=1}^{n} \sum_{l=0}^{\infty} \sum_{w:j \overset{l}{\rightarrow} i} \phi(w) b_j$$
$$= \sum_{j=1}^{n} \sum_{w:j \rightarrow i} \phi(w) b_j. \quad (13)$$

The connection between walk summability and diagonal

dominance is revealed in the result below [17]. Recall [18] that a matrix $A = \{a_{ij}\}$ is called diagonally dominant if $a_{ii} > 0$ and $a_{ii} > \sum_{j \neq i} |a_{ij}|$ for all i.

Lemma 2 [3], [17]: A matrix $A \in \mathbb{R}^{n \times n}$ with $a_{ii} = 1$ for all i is walk-summable (i.e., $\rho(\bar{R}) < 1$) if and only if A is generalised diagonally dominant, i.e., there exists a diagonal matrix $D > 0$ such that $D^{-1}AD$ is diagonally dominant.

IV. CONVERGENCE RATE ANALYSIS

This section presents the main result of this paper on the convergence rate of GaBP. The key to this analysis is the so-called unwrapped tree graph proposed in [2], which is a computation tree graph, associated with the GaBP. Using this tool, the asymptotic convergence of GaBP was proved in [2] under the assumption of diagonal dominance. This tool was further used in [3] to relax the diagonal dominance assumption to walk summability (or equivalently, generalised diagonal dominance). Here we use the same tool for convergence rate analysis.

A. Unwrapped Tree Graph

Following the work of [2], we construct an unwrapped tree with depth $t > 0$ for a loopy graph $\mathcal{G}$ [2]. Take node i to be the root and then iterate the following procedure t times:
1) Find all leaves of the tree (start with the root);
2) For each leaf, find all the nodes in the loopy graph that neighbor this leaf node, except its parent node in the tree, and add all these nodes as the children to this leaf node.

The variables and weights for each node in the unwrapped tree are copied from the corresponding nodes in the loopy graph. It is clear that taking each node as root node will generate a different unwrapped tree. Fig. 2 shows the unwrapped tree around root node 1 for a loopy graph. Note, for example, that nodes $1', 1'', 1''', 1`, 1``, 1```$ all carry the same values b_1 and a_{11}. Similarly, if node $1'$ is the parent (or child) of node j' in the unwrapped tree, and node 1 and node j are a wrapped version of nodes 1 and j, then $a_{1'j'} = a_{1j}$ (or $a_{j'1'} = a_{j1}$). A similar comment applies to unwrapped b_i.
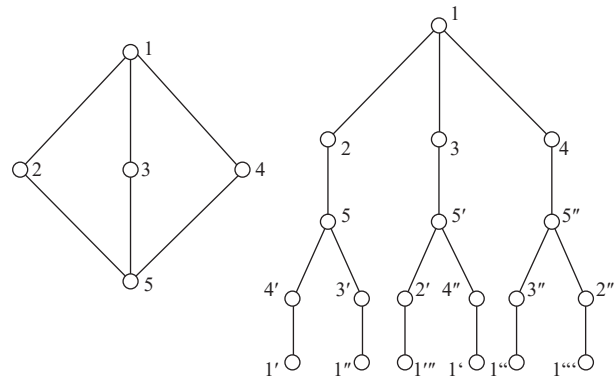

Fig. 2. Left: a loopy graph; Right: the unwrapped tree for root node 1 with 4 layers ($t = 4$).

List the nodes in the unwrapped tree in breadth first order, by starting from the root node, followed by the first layer (i.e., the children of the root node), then the second layer, etc.

Denote the unwrapped tree as $\mathcal{G}_i^{(t)} = \{\mathcal{V}_i^{(t)}, \mathcal{E}_i^{(t)}\}$ with the associated matrix $A_i^{(t)} = I - R_i^{(t)}$ and vector $b_i^{(t)}$. It is obvious that $\mathcal{G}_i^{(t)}$ is connected by construction.

We have the following key property.

Lemma 3 [3]: There is a one-to-one correspondence between finite-length walks in $\mathcal{G}$ that end at i, and walks in $\mathcal{G}_i^{(\infty)}$. That is, every finite-length walk in $\mathcal{G}$ has a counterpart in some $\mathcal{G}_i^{(k)}$ with some $i \in \mathcal{V}$ and some sufficiently large k, and every finite-length walk in $\mathcal{G}_i^{(k)}$ for any $i \in \mathcal{V}$ and $k \geq 0$ corresponds to a finite-length walk in $\mathcal{G}$.

B. Main Result

We first establish a relationship between $\mu_i(k)$ in (6) (obtained by GaBP) and the walks in $\mathcal{G}_i^{(k)}$.

Lemma 4: Under the assumption that the information matrix A in (1) is walk summable, we have, for any $i \in \mathcal{V}$ and $k \geq 0$

$$\mu_i(k) = \sum_{j=1}^{n} \sum_{w:\, j \to i | \mathcal{G}_i^{(k)}} \phi(w) b_j \qquad (14)$$

where $w : j \to i | \mathcal{G}_i^{(k)}$ denotes a walk from j to i inside the unwrapped graph $\mathcal{G}_i^{(k)}$.

Proof: Without loss of generality, we assume $i = 1$. For the unwrapped graph $\mathcal{G}_1^{(k)}$, consider the corresponding matrix $A_1^{(k)}$ and vector $b_1^{(k)}$. Define $z^{(k)} = (A_1^{(k)})^{-1} b_1^{(k)}$. Then, $z^{(k)}$ can be solved by applying GaBP on $\mathcal{G}_1^{(k)}$. As noted in Section II, since $\mathcal{G}_1^{(k)}$ is a tree graph, it is well known [2] that applying GaBP to $\mathcal{G}_1^k$ results in a correct solution for $z_1^{(k)}$ (the first component of $z^{(k)}$) in k iterations because every node in $\mathcal{G}_1^{(k)}$ is no more than k hops away from node 1. On the other hand, due to the fact that the parameters in $A_1^{(k)}$ and $b_1^{(k)}$ are all copied from A and b, applying GaBP to the original graph $\mathcal{G}$ for k iterations is identical to applying it to $\mathcal{G}_1(k)$. That is, $\mu_1^{(k)}$ in (6), which is obtained by applying GaBP on $\mathcal{G}$ for k iterations, is equal to $z_1^{(k)}$. Now, it is also known that every tree graph is walk-summable [3]. Thus, we can apply (13) to $\mathcal{G}_1^{(k)}$ to obtain

$$z_1^{(k)} = \sum_{j} \sum_{j \to 1 | \mathcal{G}_1^{(k)}} \phi(w) b_j.$$

Using $z_1^{(k)} = \mu_1(k)$, we have proved (14) for $i = 1$. Hence, the result in the lemma holds. ∎

Now we can state the main result, derived by combining the walk summability approach in [3] (i.e., Lemmas 1, 2 and 4) with the control systems approach. That is, we view the evolution of $\mu_i(k) - \mu_i$ as a dynamic system and examine its exponential stability property.

Theorem 1: Suppose the information matrix A in (1) is generalised diagonally dominant. Then, the convergence rate of GaBP is at least $\rho(\bar{R})$, i.e.,

$$|\mu_i(k) - \mu_i| \leq \rho(\bar{R})^k C \qquad (15)$$

for all $i \in \mathcal{V}$ and $k \geq 0$, where C is a constant (independent of k).

Proof: Firstly, using Lemma 2, we know that the assumption of general diagonal dominance is equivalent to assuming walk summability. In particular, A is invertible.

From (2), we have $\mu = A^{-1} b$. Using the walk summability condition and Lemma 1, we get (13), i.e.,

$$\mu_i = \sum_{j=1}^{n} \sum_{l=0}^{\infty} \sum_{w:\, j \xrightarrow{l} i} \phi(w) b_j.$$

On the other hand, $\mu_i(k)$ is given by (14), according to Lemma 4. Combining the above, we get

$$\mu_i(k) - \mu_i = \sum_{j=1}^{n} \sum_{l=0}^{\infty} \left(\sum_{w:\, j \xrightarrow{l} i | \mathcal{G}_i^{(k)}} \phi(w) b_j - \sum_{w:\, j \xrightarrow{l} i} \phi(w) b_j \right).$$

Denote by $W_i(k)$ the set of all the walks that end at node i with walk length greater than k, and by $\tilde{W}_i(k) \subset W_i(k)$ the subset of all the walks containing nodes not in $\mathcal{G}_k$. It is clear that every walk in $\tilde{W}_i(k)$ has length greater than k. Then, the above expression can be rewritten as

$$\mu_i(k) - \mu_i = \sum_{j=1}^{n} \sum_{l=0}^{\infty} \sum_{w:\, j \xrightarrow{l} i | \tilde{W}_i^{(k)}} \phi(w) b_j$$

$$= \sum_{j=1}^{n} \sum_{l=k+1}^{\infty} \sum_{w:\, j \xrightarrow{l} i | \tilde{W}_i^{(k)}} \phi(w) b_j.$$

It follows that

$$|\mu_i(k) - \mu_i| \leq \sum_{j=1}^{n} \sum_{l=k+1}^{\infty} \sum_{w:\, j \xrightarrow{l} i | \tilde{W}_i^{(k)}} |\phi(w)| |b_j|$$

$$\leq \sum_{j=1}^{n} \sum_{l=k+1}^{\infty} \sum_{w:\, j \xrightarrow{l} i} |\phi(w)| |b_j|$$

$$= \sum_{j=1}^{n} \sum_{l=k+1}^{\infty} (\bar{R}^l)_{ij} |b_j|$$

$$= \left(\sum_{l=k+1}^{\infty} \bar{R}^l |b| \right)_i$$

$$= \left(\bar{R}^k \sum_{l=1}^{\infty} \bar{R}^l |b| \right)_i$$

$$\leq \rho(\bar{R})^k C$$

where $C = \max_i (\sum_{l=1}^{\infty} \bar{R}^l |b|)_i$ is bounded due to the fact that $\rho(\bar{R}) < 1$. Hence, (15) holds for all i. ∎

Remark 1: The bound in Theorem 1 coincides with that in [5]. But our assumptions are weaker. More specifically, [5] requires the decomposition of $p(x)$ in (4) to be such that $\phi_{ij}(x_i, x_j)$ are convex, whereas the natural decomposition (5) does not have this property. It is true that if the system is generalised diagonally dominant, there exists a convex decomposition [3]. But how to search for such convex decomposition in a distributed way is a non-trivial exercise, and there is no such algorithm available. Also, our proof is simple and purely based on linear algebraic analysis, different from existing convergence analysis methods.

V. ILLUSTRATING EXAMPLES

To illustrate the convergence rate bound in Theorem 1, we give two loopy graphs in this section as examples. The first one is a 13-node graph with at most 7 neighbouring nodes for each node, as shown in Fig. 3. The second example is a 1000-node graph with at most 6 randomly selected neighbouring nodes for each node, as shown in Fig. 4. In both cases, the resulting matrix A is sparse. The parameters of the corresponding Gaussian density function $p(x) \propto \exp\left[-\frac{1}{2}x^T A x + b^T x\right]$ are chosen as follows: In the first example, every $a_{ij} = a_{ji}, i \neq j$, is a random value in $(-0.26, 0.26)$, and in the second example, every $a_{ij} = a_{ji}, i \neq j$, belongs to $(-0.165, 0.165)$ randomly. Additionally, $a_{ii} = 1$ and $b_i = i$, $i = 1, \ldots, n$. The values of a_{ij} are chosen to ensure diagonal dominance (checked numerically for each row after the random values are chosen), which in turn ensure walk summability.
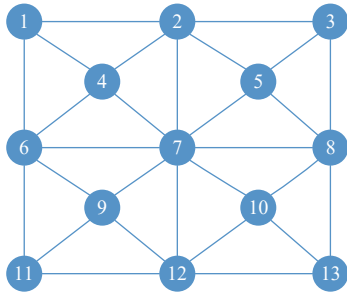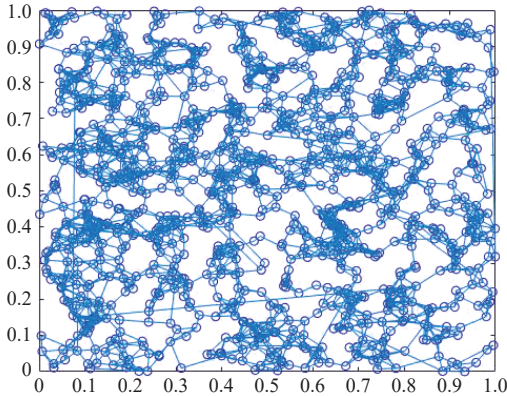

Fig. 3.   13-node graph.

Fig. 4.   1000-node graph.

The GaBP simulation results are shown in Figs. 5 and 6. The x-axis stands for iteration numbers and the y-axis stands for a log form of error between the true Gaussian mean μ and its estimate $\mu(k)$ calculated by GaBP, or more precisely, $\log_{10}(\sum_i (\mu_i(k) - \mu_i)^2/n)$. By Theorem 1, this term should decay at least linearly, i.e.,

$$\log_{10}\left(\sum_i (\mu_i(k) - \mu_i)^2/n\right) \leq \log_{10} C^2 + k \log_{10} \rho^2$$

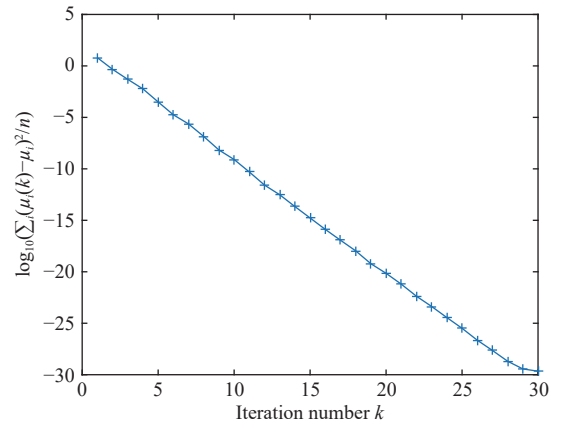where the term $\log_{10}\rho^2$ represents the slope.

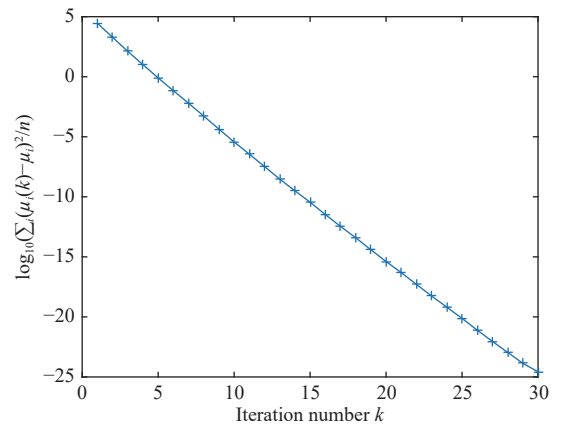Fig. 5.   GaBP iterations for the 13-node graph.


Fig. 6.   GaBP iterations for the 1000-node graph.

The simulation results for both two examples have shown that the error decreases exponentially with the increase of the iteration number. The slope for the 13-node example is measured to be roughly -1.0502, corresponding to the convergence rate of $10^{-1.0502/2} \approx 0.2985$. The slope for the 1000-node example is measured to be roughly -1.0642, corresponding to the convergence rate of $10^{-1.0642/2} \approx 0.2937$. In comparison, for the 13-node graph, the spectral radius of $\bar{R}$ is 0.6100; and for the 1000-node graph, the spectral radius of $\bar{R}$ is 0.9671. We see that in both examples, the $\rho(\bar{R})$ upper bounds the actual convergence rate of GaBP.

VI. CONCLUSIONS

In this paper, we have analysed the convergence property of GaBP for Markov networks and provided a simple bound on the convergence rate. The bound is characterised by $\rho(\bar{R})$ and is guaranteed to be less than 1 under the walk summability (or generalised diagonal dominance) assumption. This result gives a simple extension to the known asymptotic convergence property of GaBP under the same assumption [2], [3]. We see in the simulation results that the actual convergence rate is faster than predicted by $\rho(\bar{R})$. It would be interesting to see how this bound can be further improved. Other future directions include relaxing the walk summability assumption (possibly by following the work of [15]), and generalising GaBP to wider distributed estimation and distributed optimisation problems.

REFERENCES

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

[2] 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.

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

[4] D. Marelli and M. Fu, "Distributed weighted least-squares estimation with fast convergence for large-scale systems," *Automatica*, vol. 51, pp. 27–39, 2015.

[5] C. C. Moallemi and B. Van Roy, "Convergence of min-sum message-passing for quadratic optimization," *IEEE Trans. Inf. Theory*, vol. 55, no. 5, pp. 2413–2423, 2009.

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

[7] Q. Su and Y-C. Wu, "Convergence analysis of the variance in Gaussian belief propagation," *IEEE Trans. Signal Processing*, vol. 62, no. 19, pp. 5119–5131, 2014.

[8] Q. Su and Y-C. Wu, "On convergence conditions of Gaussian belief propagation," *IEEE Trans. Signal Processing*, vol. 63, no. 5, pp. 1144–1155, 2015.

[9] O. Shental, P. H. Siegel, J. K. Wolf, D. Bickson, and D. Dolev, "Gaussian belief propagation solver for systems of linear equations," in *Proc. IEEE Int. Symp. Information Theory*, vol. 4, pp. 1863–1867, 2008, Toronto, Canada.

[10] X.-M. Zhang, Q.-L. Han, X. H. Ge, D. R. Ding, L. Ding, D. Yue, and C. Peng, "Networked control systems: a survey of trends and techniques," *IEEE/CAA J. Autom. Sinica*, vol. 7, no. 1, pp. 1–17, Jan. 2020.

[11] H. F. Niu, A. Sahoo, C. Bhowmick, and S. Jagannathan, "An optimal hybrid learning approach for attack detection in linear networked control systems," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 6, pp. 1404–1416, Nov. 2019.

[12] B. Xia, W. H. Yuan, N. Xie, and C. H. Li, "A novel statistical manifold algorithm for position estimation," *IEEE/CAA J. Autom. Sinica*, vol. 6, no. 6, pp. 1513–1518, Nov. 2019.

[13] X. Tan and J. Li, "Computationally efficient sparse Bayesian learning via belief propagation," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2010–2021, Apr. 2010.

[14] N. Ruozzi and S. Tatikonda, "Message-passing algorithms for quadratic minimization," *J. Machine Learning Research*, vol. 14, pp. 2287–2314, Aug. 2013.

[15] N. Ruozzi and S. Tatikonda, "A new approach to Laplacian solvers and flow problems," *J. Machine Learning Research*, vol. 20, pp. 1–37, Feb. 2019.

[16] B. Cseke and T. M. Heskes, "Bounds on the Bethe free energy for Gaussian networks." in *Proc. 24th Conf. Uncertainty in Artificial Intelligence*, 2008, pp. 97–104.

[17] E. G. Boman, D. Chen, O. Parekh, and S. Toledo, "On factor width and symmetric H-matrices," *Linear Algebra and its Applications*, vol. 405, pp. 239–248, 2005.

[18] Y. Saad, *Iterative Methods for Sparse Linear Systems*, 2nd ed., SIAM, 2003.
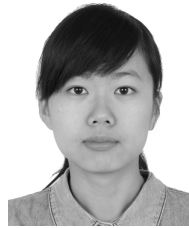
Zhaorong Zhang received the bachelor degree of engineering at Nanjing University of Science and Technology in 2016. She began the PhD study at University of Newcastle from Feb. 2017. Her research interest is distributed state estimation.

Minyue Fu (F'04) received the bachelor degree in electrical engineering from the University of Science and Technology of China in 1982, and the M.S. and Ph.D. degrees in electrical engineering from the University of Wisconsin-Madison in 1983 and 1987, respectively. From 1983 to 1987, he held a teaching assistantship and a research assistantship at the University of Wisconsin-Madison. From 1987 to 1989, he served as an Assistant Professor in the Department of Electrical and Computer Engineering, Wayne State University, USA. He joined the Department of Electrical and Computer Engineering, University of Newcastle, Australia, in 1989. Currently, he is a Chair Professor in electrical engineering in the School of Electrical Engineering and Computing. In addition, he has been a Visiting Associate Professor at the University of Iowa, a Visiting Professor at Nanyang Technological University, Singapore, Changjiang Professor at Shandong University, a Distinguished Scholar at Zhejiang University, and a Bai-ren Scholar at Guangdong University of Technology. He is a Fellow of the IEEE, Fellow of Engineers Australia, and Fellow of Chinese Association of Automation. His current research interests include networked control systems, multi-agent systems, and distributed estimation and control. He has been an Associate Editor for the *IEEE Transactions on Automatic Control, IEEE Transactions on Signal Processing, Automatica and Journal of Optimization and Engineering*.