CONVERGENCE ANALYSIS OF BELIEF PROPAGATION FOR PAIRWISE LINEAR GAUSSIAN MODELS

Jian Du[†] *Shaodan Ma*[⋆], *Yik-Chung Wu*[‡], *Soummya Kar*[†] *and José M. F. Moura*[†]

Electrical and Computer Engineering, Carnegie Mellon University[†], Pittsburgh, PA
Electrical and Computer Engineering, University of Macau[⋆], Macau
Electrical and Electronic Engineering, The University of Hong Kong[‡], Hong Kong

ABSTRACT

Gaussian belief propagation (BP) has been widely used for distributed inference in large-scale networks such as the smart grid, sensor networks, and social networks, where local measurements/observations are scattered over a wide geographical area. One particular case is when two neighboring agents share a common observation. For example, to estimate voltage in the direct current (DC) power flow model, the current measurement over a power line is proportional to the voltage difference between two neighboring buses. When applying the Gaussian BP algorithm to this type of problem, the convergence condition remains an open issue. In this paper, we analyze the convergence properties of Gaussian BP for this pairwise linear Gaussian model. We show analytically that the updating information matrix converges at a geometric rate to a unique positive definite matrix with arbitrary positive semidefinite initial value and further provide the necessary and sufficient convergence condition for the belief mean vector to the optimal estimate.

Index Terms— graphical model, belief propagation, large-scale networks, distributed inference, Markov random field.

1. INTRODUCTION

Gaussian belief propagation (BP) provides an efficiently distributed way to compute the marginal distribution from the joint distribution of unknown random variables, and it has been adopted in a variety of areas such as distributed power state estimation [1] in power networks, synchronization [2–4] in wireless communication networks [5,6], cooperative localization in distributed networks [7], factor analyzer network [8], sparse Bayesian learning [9], and peer-to-peer rating in social networks [10]. In one particular model of interested studied in [2–4, 7, 10, 11]), two neighboring agents share a common observation. In this paper, we name this type of model pairwise linear Gaussian models.

Although with great empirical success, the major challenge that hinders Gaussian BP to realize its full potential is the lack of theoretical guarantees of convergence in loopy networks. Sufficient convergence conditions for Gaussian BP have been developed in [12–14] when the underlying Gaussian distribution is expressed in terms of pairwise connections between scalar variables (also known as Markov random field (MRF)). However, as demonstrated in [15] the iterative equations for Gaussian BP on MRFs are different from that for distributed estimation problems such as in [1–3, 11, 16, 17], where linear measurements are involved. Therefore, the existing conditions and analysis methods in [12–14] are not applicable to distributed estimation problems. Though [15] gives the necessary and sufficient condition of BP for the Gaussian linear model, the type of observation allowed in [15] is not the most general in the sense that it does not allow two neighboring agents to share a common observation. In this paper, we focus particularly on the convergence analysis of BP for this pairwise linear Gaussian model. We show analytically that the updating of the information matrix converges at a geometric rate to a unique positive definite matrix with arbitrary positive semidefinite initial value and further provide the necessary and sufficient convergence condition for the updating belief mean vector to the optimal estimate.

Note that, in the setup of deterministic unknown parameter estimation, the distributed algorithm based on the consensus+innovations philosophy proposed in [18, 19] (see also the related family of diffusion algorithms [20]) converges to the optimal centralized estimator under the assumption of global observability of the (aggregate) sensing model and connectivity of the inter-agent communication network. In particular, these algorithms allow 1) the communication or message exchange network to be different from the physical coupling network, and 2) the communication network to have arbitrary network structure with cycles (as long as it is connected). The results in [18, 19] imply that the unknown variables $\mathbf{x}$ can be reconstructed completely at each agent in the network. For large-scale networks with high dimensional $\mathbf{x}$, it may be impractical to reconstruct $\mathbf{x}$ at every agent. In [21, section 3.4], the author developed approaches to address this problem, where each agent can reconstruct a set of unknown variables that should be larger than the set of variables that influence its local measurement. This paper studies a different distributed estimation problem when each agent estimates only its own unknown variables under pairwise independence condition of the unknown variables; this leads to lower dimensional data exchanges between neighbors.

2. COMPUTATION MODEL

Consider a general connected network of M agents, with $\mathcal{V} = \{1, \ldots, M\}$ denoting the set of agents, and $\mathcal{E}_{\text{Net}} \subset \mathcal{V} \times \mathcal{V}$ as the set of all undirect communication links in the network, i.e., if i and j are within the communication range, $(i, j) \in \mathcal{E}_{\text{Net}}$. The local observations, $\mathbf{y}_{i,j}$, between agents i and j are modeled by a pairwise Gaussian linear model:

$$\mathbf{y}_{i,j} = \mathbf{A}_{j,i}\mathbf{x}_i + \mathbf{A}_{i,j}\mathbf{x}_j + \mathbf{z}_{i,j}, \tag{1}$$

where $\mathbf{A}_{j,i}$ and $\mathbf{A}_{i,j}$ are the known coefficient matrices with full column rank, $\mathbf{x}_i$ and $\mathbf{x}_j$ are the local unknown vector parameters at agent i and j with dimension $N_i \times 1$ and $N_j \times 1$, and with the prior distribution $p(\mathbf{x}_i) \sim \mathcal{N}(\mathbf{x}_i|\mathbf{0}, \mathbf{W}_i)$ and $p(\mathbf{x}_j) \sim \mathcal{N}(\mathbf{x}_j|\mathbf{0}, \mathbf{W}_j)$ and $\mathbf{z}_{i,j}$ is the additive noise with distribution $\mathbf{z}_{i,j} \sim \mathcal{N}(\mathbf{z}_{i,j}|\mathbf{0}, \mathbf{R}_{i,j})$. It is assumed that $p(\mathbf{x}_i, \mathbf{x}_j) = p(\mathbf{x}_i)p(\mathbf{x}_j)$ and $p(\mathbf{z}_{i,j}, \mathbf{z}_{s,t}) = p(\mathbf{z}_{i,j})p(\mathbf{z}_{s,t})$ for $\{i, j\} \neq \{s, t\}$. The goal is to estimate $\mathbf{x}_i$, based on $\mathbf{y}_{i,j}$, $p(\mathbf{x}_i)$ and $p(\mathbf{z}_{i,j})$ for all $\mathbf{x}_i \in \mathcal{V}$. Note that in (1), $\mathbf{y}_{i,j} = \mathbf{y}_{j,i}$.

In centralized estimation, all the observations $\mathbf{y}_{i,j}$ at different agents are forwarded to a central processing unit. Define vectors $\mathbf{y}$, $\mathbf{x}$ and $\mathbf{z}$ as the stacking of $\mathbf{y}_{i,j}$, $\mathbf{x}_i$ and $\mathbf{z}_{i,j}$ in ascending order first with respect to i and then on j, respectively; then we obtain $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$, where $\mathbf{A}$ is constructed from $\mathbf{A}_{n,i}$, with specific arrangement depending on the network topology. Assuming $\mathbf{A}$ is a full column rank matrix, and since $\mathbf{z}$ is a Gaussian random vector, the optimal estimate $\hat{\mathbf{x}} \triangleq [\hat{\mathbf{x}}_1^T, \ldots, \hat{\mathbf{x}}_M^T]^T$ of $\mathbf{x}$ is given by

$$\hat{\mathbf{x}} = (\mathbf{W}^{-1} + \mathbf{A}^T\mathbf{R}^{-1}\mathbf{A})^{-1}\mathbf{A}^T\mathbf{R}^{-1}\mathbf{y}, \tag{2}$$

where $\mathbf{W}$ and $\mathbf{R}$ are block diagonal matrices containing $\mathbf{W}_{i,j}$ and $\mathbf{R}_{i,j}$ as their diagonal blocks, respectively. Although well-established, the drawbacks of the centralized estimation in large-scale networks include 1) the transmission of $\mathbf{y}_{i,j}$, $\mathbf{A}_{i,j}$ and $\mathbf{R}_{i,j}$ from peripheral agents to the computation center imposes huge communication overhead; 2) knowledge of the global network topology is needed in order to construct $\mathbf{A}$; and 3) the computation burden at the computation center scales up with the cubic of the dimension of the matrix inverse in (2) with complexity order $\mathcal{O}((\sum_{i=1}^{|\mathcal{V}|} N_i)^3)$.

The joint distribution $p(\mathbf{x})\, p(\mathbf{y}|\mathbf{x})$ is first written as the product of the prior distribution and the likelihood function as

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

To facilitate the derivation of the distributed inference algorithm, the factorization above is expressed in terms of a factor graph, where every variable vector $\mathbf{x}_i$ is represented by a variable node and the probability distribution of a vector variable or a group of vector variables is represented by a factor node. A variable node is connected to a factor node if the variable is involved in that particular factor. It involves two types of messages: One is the message from a factor node with function f to its neighboring variable node $\mathbf{x}_i$, defined as

$$m_{f\to i}^{(\ell)}(\mathbf{x}_i) = \int\cdots\int f \times \prod_{j\in\mathcal{B}(f)\setminus i} m_{j\to f}^{(\ell)}(\mathbf{x}_j)\, \mathrm{d}\{\mathbf{x}_n\}_{n\in\mathcal{B}(f)\setminus i}, \tag{3}$$

where $\mathcal{B}(f)$ denotes the set of neighboring variable nodes of factor node f on the factor graph. The other type of message is from factor node $\tilde{f}$, which denotes a likelihood function or prior distribution, to its neighboring variable node $\mathbf{x}_i$ and it is defined as

$$m_{j\to f}^{(\ell)}(\mathbf{x}_i) = \prod_{\tilde{f}\in\mathcal{B}(j)\setminus f} m_{\tilde{f}\to j}^{(\ell-1)}(\mathbf{x}_j), \tag{4}$$

where $\mathcal{B}(j)$ denotes the set of neighbouring factor nodes of $\mathbf{x}_j$, and $m_{\tilde{f}\to j}^{(\ell-1)}(\mathbf{x}_j)$ is the message from $\tilde{f}$ to $\mathbf{x}_j$ at time $l-1$. The process iterates between equations (4) and (3). At each iteration ℓ, the approximate marginal distribution, also named belief, on $\mathbf{x}_i$ is computed locally at $\mathbf{x}_i$ as

$$b_{\text{BP}}^{(\ell)}(\mathbf{x}_i) = \prod_{f\in\mathcal{B}(i)} m_{f\to i}^{(\ell)}(\mathbf{x}_i). \tag{5}$$

It can be shown that the message from factor node $f_{i,j}$ to variable node i is given by [15]

$$m_{f_{i,j}\to i}^{(\ell)}(\mathbf{x}_i) \propto \exp\Big\{ -\frac{1}{2}||\mathbf{x}_i - \mathbf{v}_{f_{i,j}\to i}^{(\ell)}||^2_{\mathbf{C}_{f_{i,j}\to i}^{(\ell)}} \Big\}, \tag{6}$$

where $\mathbf{C}_{f_{i,j}\to j}^{(\ell-1)}$ and $\mathbf{v}_{f_{i,j}\to j}^{(\ell-1)}$ are the message covariance matrix and mean vector received at variable node j at the $l-1$ iteration with

$$\Big[\mathbf{C}_{f_{i,j}\to i}^{(\ell)}\Big]^{-1} = \mathbf{A}_{j,i}^T \Big[\mathbf{R}_{i,j} + \mathbf{A}_{i,j}\mathbf{C}_{j\to f_{i,j}}^{(\ell)}\mathbf{A}_{i,j}^T\Big]^{-1}\mathbf{A}_{j,i}. \tag{7}$$

and

$$\mathbf{v}_{f_{i,j}\to i}^{(\ell)} = \mathbf{A}_{j,i}^T \Big[\mathbf{R}_{i,j} + \mathbf{A}_{i,j}\mathbf{C}_{j\to f_{i,j}}^{(\ell)}\mathbf{A}_{i,j}^T\Big]^{-1}\Big(\mathbf{y}_{i,j} - \mathbf{A}_{i,j}\mathbf{v}_{j\to f_{i,j}}^{(\ell)}\Big). \tag{8}$$

Furthermore, the general expression for the message from variable node j to factor node $f_{i,j}$ is

$$m_{j\to f_{i,j}}^{(\ell)}(\mathbf{x}_j) \propto \exp\Big\{ -\frac{1}{2}||\mathbf{x}_j - \mathbf{v}_{j\to f_{i,j}}^{(\ell)}||^2_{\mathbf{C}_{j\to f_{i,j}}^{(\ell)}} \Big\}, \tag{9}$$

where $\mathbf{C}_{j\to f_{i,j}}^{(\ell)}$ and $\mathbf{v}_{j\to f_{i,j}}^{(\ell)}$ are the message covariance matrix and mean vector received at variable node j at the ℓ-th iteration, with the information matrix computed as

$$\big[\mathbf{C}^{(\ell)}_{j\to f_{i,j}}\big]^{-1} = \mathbf{W}_j^{-1} + \sum_{f_{k,j}\in\mathcal{B}(j)\backslash f_{i,j}} \big[\mathbf{C}^{(\ell-1)}_{f_{k,j}\to j}\big]^{-1}. \quad (10)$$

and the mean vector is

$$\mathbf{v}^{(\ell)}_{j\to f_{i,j}} = \mathbf{C}^{(\ell)}_{j\to f_{i,j}}\left[\sum_{f_{k,j}\in\mathcal{B}(j)\backslash f_{i,j}} \big[\mathbf{C}^{(\ell-1)}_{f_{k,j}\to j}\big]^{-1}\mathbf{v}^{(\ell-1)}_{f_{k,j}\to j}\right], \quad (11)$$

Following Lemma 2 in [15], we know that setting the initial information matrix $[\mathbf{C}^{(0)}_{f_{k,j}\to i}]^{-1} \succeq \mathbf{0}$ for all $k \in \mathcal{V}$ and $j \in \mathcal{B}(k)$ guarantees $[\mathbf{C}^{(\ell)}_{j\to f_{i,j}}]^{-1} \succ \mathbf{0}$ for $l \geq 1$. Therefore, let the initial messages at factor node $f_{k,j}$ be in Gaussian function forms with covariance $[\mathbf{C}^{(0)}_{f_{k,j}\to j}]^{-1} \succeq \mathbf{0}$ for all $k \in \mathcal{V}$ and $j \in \mathcal{B}(f_{k,j})$. Then all the messages $m^{(\ell)}_{j\to f_{i,j}}(\mathbf{x}_j)$ and $m^{(\ell)}_{f_{i,j}\to i}(\mathbf{x}_i)$ exist and are in Gaussian form. Furthermore, during each round of message passing, each agent can compute the belief for $\mathbf{x}_i$ using (5), which can be easily shown to be

$$b_i^{(l)}(\mathbf{x}_i) \sim \mathcal{N}(\mathbf{x}_i|\boldsymbol{\mu}_i^{(l)}, \mathbf{P}_i^{(l)}), \quad (12)$$

with the inverse of the covariance matrix

$$\big[\mathbf{P}_i^{(l)}\big]^{-1} = \sum_{f_{i,j}\in\mathcal{B}(f_{i,j})} \big[\mathbf{C}^{(l)}_{f_{i,j}\to i}\big]^{-1}, \quad (13)$$

and mean vector

$$\boldsymbol{\mu}_i^{(l)} = \Big[\sum_{f_{i,j}\in\mathcal{B}(f_{i,j})} \big[\mathbf{C}^{(l)}_{f_{i,j}\to i}\big]^{-1}\Big]^{-1} \sum_{j\in\mathcal{B}(f_{i,j})} \big[\mathbf{C}^{(l)}_{f_{i,j}\to i}\big]^{-1}\mathbf{v}^{(l)}_{f_{i,j}\to i}. \quad (14)$$

The iterative algorithm based on BP is summarized as follows. The algorithm is started by setting the message from factor node to variable node as $m^{(0)}_{f_{i,j}\to i}(\mathbf{x}_i) = \mathcal{N}\left(\mathbf{x}_i; \boldsymbol{v}^{(0)}_{f_{i,j}\to i}, \boldsymbol{C}^{(0)}_{f_{i,j}\to i}\right)$ with a random initial vector $\boldsymbol{v}^{(0)}_{f_{i,j}\to i}$ and $\left[\boldsymbol{C}^{(0)}_{f_{i,j}\to i)}\right]^{-1} \succeq \mathbf{0}$. At each round of message exchange, every variable node computes the outgoing messages to factor nodes according to (10) and (11). After receiving the messages from its neighboring variable nodes, each factor node computes its outgoing messages according to (7) and (8). Such iteration is terminated when (14) converges (e.g., when $\|\boldsymbol{\mu}_i^{(\ell)} - \boldsymbol{\mu}_i^{(\ell-1)}\| < \eta$, where η is a threshold) or the maximum number of iterations is reached. Then the estimate of $\mathbf{x}_i$ of each node is obtained as in (14).

3. CONVERGENCE ANALYSIS

The challenge of deploying the BP algorithm for large-scale networks is determining whether it will converge. In particular, it is generally known that, if the factor graph contains cycles, the BP algorithm may diverge. Thus, determining convergence conditions for the BP algorithm is very important. Sufficient conditions for the convergence of Gaussian BP with scalar variable in loopy graphs are available in [12, 13] for Markov random fields. Unfortunately, as first pointed out in [15], the convergence analysis for the Gaussian Markov random field and for the Gaussian linear model are quite different due to different iteration equations. Though [15] gives the necessary and sufficient condition of BP for the Gaussian linear model, the type of observations allowed in [15] (e.g., equation (1) in [15]), is not the most general in the sense that it does not allow two neighboring agents to share a common observation as in equation (1) in this paper. In the following, we provide the convergence analysis of Gaussian BP for the pairwise linear Gaussian model.

Due to the recursively updating property of $m^{(\ell)}_{j\to f_{i,j}}(\mathbf{x}_j)$ and $m^{(\ell)}_{f_{i,j}\to i}(\mathbf{x}_i)$ in (9) and (6), the message evolution can be simplified by combining these two types of messages into a single one. By substituting $\big[\mathbf{C}^{(\ell)}_{j\to f_n}\big]^{-1}$ in (10) into (7), the updating of the message covariance matrix inverse, named message information matrix in the following, can be denoted as

$$\begin{aligned}
[\mathbf{C}^{(\ell)}_{f_{i,j}\to i}]^{-1} &= \mathbf{A}_{j,i}^T\big[\mathbf{R}_{i,j} + \mathbf{A}_{i,j}\big[\mathbf{W}_j^{-1} \\
&+ \sum_{f_{k,j}\in\mathcal{B}(j)\backslash f_{i,j}} \big[\mathbf{C}^{(\ell-1)}_{f_{k,j}\to j}\big]^{-1}\big]^{-1}\mathbf{A}_{i,j}^T\big]^{-1}\mathbf{A}_{j,i} \\
&\triangleq \mathcal{F}_{n\to i}\big(\{\big[\mathbf{C}^{(\ell-1)}_{f_{k,j}\to j}\big]^{-1}\}_{f_{k,j}\in\mathcal{B}(j)\backslash f_{i,j}}\big). \quad (15)
\end{aligned}$$

Observing that $\mathbf{C}^{(\ell)}_{f_{i,j}\to i}$ in (15) is independent of $\mathbf{v}^{(\ell)}_{f_{i,j}\to i}$, the other type of updating information, we first focus on the convergence property of $[\mathbf{C}^{(\ell)}_{f_n\to i}]^{-1}$.

To consider the updates of all message information matrices, we introduce the following definitions. Let $\mathbf{C}^{(\ell-1)} \triangleq \mathtt{Bdiag}(\{[\mathbf{C}^{(\ell-1)}_{f_{i,j}\to i}]^{-1}\}_{i\in\mathcal{V},\{i,j\}\in\mathcal{E}_{\text{Net}}}$ be a block diagonal matrix with diagonal blocks being the message information matrices in the network at time $l-1$ with index arranged in ascending order first on i and then on j. Using the definition of $\mathbf{C}^{(\ell-1)}$, the term $\sum_{f_{k,j}\in\mathcal{B}(j)\backslash f_{i,j}} \big[\mathbf{C}^{(\ell-1)}_{f_{k,j}\to j}\big]^{-1}$ in (15) can be written as $\boldsymbol{\Xi}_{i,j}\mathbf{C}^{(\ell-1)}\boldsymbol{\Xi}_{i,j}^T$, where $\boldsymbol{\Xi}_{i,j}$ selects appropriate components from $\mathbf{C}^{(\ell-1)}$ to form the summation.

$$\begin{aligned}
\left[\mathbf{C}^{(\ell)}_{f_{i,j}\to i}\right]^{-1} =& \mathbf{A}_{j,i}^T\big\{\mathbf{R}_{i,j} + \mathbf{A}_{i,j}[\mathbf{W}_j^{-1} \\
&+ \boldsymbol{\Xi}_{i,j}\mathbf{C}^{(\ell-1)}\boldsymbol{\Xi}_{i,j}^T]^{-1}\mathbf{A}_{i,j}^T\big\}^{-1}\mathbf{A}_{j,i}.
\end{aligned} \quad (16)$$

We define the function $\mathcal{G} \triangleq \{\mathcal{G}_{1\to k}, \ldots, \mathcal{G}_{n\to i}, \ldots, \mathcal{G}_{n\to M}\}$ that updates $\mathbf{C}^{(\ell)} = \mathcal{G}(\mathbf{C}^{(\ell-1)})$. Then, by stacking $\big[\mathbf{C}^{(\ell)}_{f_{i,j}\to i}\big]^{-1}$ on the left side of (16) for all n and i as the block diagonal matrix $\mathbf{C}^{(\ell)}$, we obtain

$$\begin{aligned}
\mathbf{C}^{(\ell)} &= \mathbf{A}^T\left[\mathbf{R} + \mathbf{H}\left(\mathbf{W} + \boldsymbol{\Xi}\mathbf{C}^{(\ell-1)}\boldsymbol{\Xi}^T\right)^{-1}\mathbf{H}^T\right]^{-1}\mathbf{A}, \\
&\triangleq \mathcal{G}(\mathbf{C}^{(\ell-1)}), \quad (17)
\end{aligned}$$

where $\mathbf{A}$, $\mathbf{R}$, $\mathbf{H}$, $\mathbf{W}$, and $\boldsymbol{\Xi}$ are block diagonal matrices with block elements $\mathbf{A}_{j,i}$, $\mathbf{R}_{i,j}$, $\mathbf{A}_{i,j}$, $\mathbf{W}_j$, and $\boldsymbol{\Xi}_{i,j}$, respectively, arranged in ascending order, first on n and then on i (i.e., the same order as $[\mathbf{C}^{(\ell)}_{f_n\to i}]^{-1}$ in $\mathbf{C}^{(\ell)}$). We first present properties of the updating operator $\mathcal{G}(\cdot)$, where the proof follows that in [15].

Property 1. *The updating operator $\mathcal{G}(\cdot)$ satisfies the following properties:*

P 1.1: $\mathcal{G}(\mathbf{C}^{(\ell)}) \succeq \mathcal{G}(\mathbf{C}^{(\ell-1)})$, if $\mathbf{C}^{(\ell)} \succeq \mathbf{C}^{(\ell-1)} \succeq \mathbf{0}$.
P 1.2: $\alpha\mathcal{G}(\mathbf{C}^{(\ell)}) \succ \mathcal{G}(\alpha\mathbf{C}^{(\ell)})$ and $\mathcal{G}(\alpha^{-1}\mathbf{C}^{(\ell)}) \succ \alpha^{-1}\mathcal{G}(\mathbf{C}^{(\ell)})$, if $\mathbf{C}^{(\ell)} \succ \mathbf{0}$ and $\alpha > 1$.
P 1.3: Define $\mathbf{U} \triangleq \mathbf{A}^T\mathbf{R}^{-1}\mathbf{A}$ and $\mathbf{L} \triangleq \mathbf{A}^T\left[\mathbf{R}+\mathbf{H}\mathbf{W}^{-1}\mathbf{H}^T\right]^{-1}\mathbf{A}$. With arbitrary $\mathbf{C}^{(0)} \succeq \mathbf{0}$, $\mathcal{G}(\mathbf{C}^{(\ell)})$ is bounded by $\mathbf{U} \succeq \mathcal{G}(\mathbf{C}^{(\ell)}) \succeq \mathbf{L} \succ \mathbf{0}$ for $l \geq 1$.

In this paper, $\mathbf{X} \succeq \mathbf{Y}$ ($\mathbf{X} \succ \mathbf{Y}$) means that $\mathbf{X}-\mathbf{Y}$ is positive semidefinite (definite). Note $\mathcal{G}$ is different from the function $\mathcal{F}$ in [22]. However, as demonstrated in [22], if a function $\mathcal{G}$ satisfies Property 1, we can establish the convergence property for $\mathbf{C}^{(\ell)}$ given by the following Theorem with detailed provided in [15].

Theorem 1. *With the initial covariance matrix set to be an arbitrary p.s.d. matrix, i.e., $[\mathbf{C}^{(0)}_{f_n\to i}]^{-1} \succeq \mathbf{0}$, the sequence $\{\mathbf{C}^{(\ell)}\}_{l=0,1,\ldots}$ converges at a double exponential rate to a unique p.d. matrix.*

Thus, if we choose $[\mathbf{C}^{(0)}_{f_{i,j}\to j}]^{-1} \succeq \mathbf{0}$ for all $j \in \mathcal{V}$ and $i \in \mathcal{B}(j)$, then $\left[\mathbf{C}^{(\ell)}_{f_{i,j}\to j}\right]^{-1}$ converges at a double exponential rate to a unique p.d. matrix $\left[\mathbf{C}^{*}_{f_{i,j}\to j}\right]^{-1}$. Furthermore, according to (10), $\left[\mathbf{C}^{(\ell)}_{j\to f_{i,j}}\right]^{-1}$ also converges to a p.d. matrix once $\left[\mathbf{C}^{(\ell-1)}_{f_{k,j}\to j}\right]^{-1}$ converges; the converged value is denoted by $\left[\mathbf{C}^{*}_{j\to f_{i,j}}\right]^{-1}$. Then, for arbitrary initial value $\mathbf{v}^{(0)}_{f_{k,j}\to j}$, the evolution of $\mathbf{v}^{(\ell)}_{j\to f_n}$ in (11) can be written in terms of the limit message information matrices as

$$\mathbf{v}^{(\ell)}_{j\to f_{i,j}} = \mathbf{C}^{*}_{j\to f_{i,j}}\left[\sum_{f_{k,j}\in\mathcal{B}(j)\setminus f_{i,j}}\left[\mathbf{C}^{*}_{f_{k,j}\to j}\right]^{-1}\mathbf{v}^{(\ell-1)}_{f_{k,j}\to j}\right]. \tag{18}$$

Using (8), and replacing indices j, i with k, j respectively, $\mathbf{v}^{(\ell-1)}_{f_{k,j}\to j}$ is given by

$$\begin{aligned}\mathbf{v}^{(\ell)}_{f_{k,j}\to j} =& \mathbf{A}^T_{k,j}\left[\mathbf{R}_{k,j}+\mathbf{A}_{j,k}\mathbf{C}^{*}_{k\to f_{k,j}}\mathbf{A}^T_{j,k}\right]^{-1} \\ &\times\left(\mathbf{y}_{k,j}-\mathbf{A}_{j,k}\mathbf{v}^{(\ell)}_{k\to f_{k,j}}\right).\end{aligned} \tag{19}$$

Putting (19) into (18), we have

$$\mathbf{v}^{(\ell)}_{j\to f_{i,j}} = \mathbf{b}_{j\to f_{i,j}} - \mathbf{C}^{*}_{j\to f_{i,j}}\sum_{f_{k,j}\in\mathcal{B}(j)\setminus f_{i,j}}\mathbf{C}^{*}_{f_{kj}\to j}\mathbf{M}_{k,j}\mathbf{A}_{j,k}\mathbf{v}^{(\ell)}_{k\to f_{k,j}}, \tag{20}$$

where $\mathbf{b}_{j\to f_{i,j}} = \mathbf{C}^{*}_{j\to f_{i,j}}\sum_{f_{k,j}\in\mathcal{B}(j)\setminus f_{i,j}}\mathbf{M}_{k,j}\mathbf{y}_k$ and $\mathbf{M}_{k,j} = \mathbf{A}^T_{k,j}\left[\mathbf{R}_{k,j}+\mathbf{A}_{j,k}\mathbf{C}^{*}_{k\to f_{k,j}}\mathbf{A}^T_{j,k}\right]^{-1}$. The above equation for all $j \in \mathcal{N}(i)$ cases can be further written in a compact form as

$$\mathbf{v}^{(\ell)}_j = \mathbf{b}_j - \mathbf{Q}_j\mathbf{v}^{(\ell-1)}, \tag{21}$$

with the column vector $\mathbf{v}^{(\ell)}_j$ containing all $\{\mathbf{v}^{(\ell)}_{j\to f_{i,j}}\}_{i\in\mathcal{N}(j)}$ as subvectors with ascending index on i. Similarly, $\mathbf{b}_j$ containing all $\{\mathbf{b}_{j\to f_{i,j}}\}_{i\in\mathcal{N}(j)}$ as subvectors with ascending index on i, and $\mathbf{v}^{(\ell-1)}$ containing $\mathbf{v}^{(\ell-1)}_{k\to f_{k,j}}$ for all $f_{k,j}\in\mathcal{B}(j)\setminus f_{i,j}$ as subvectors with ascending index first on z and then on k. The matrix $\mathbf{Q}_j$ is a block matrix with component blocks $\mathbf{0}$ and $\mathbf{C}^{*}_{j\to f_{i,j}}$ where $f_{k,j}\in\mathcal{B}(j)\setminus f_{i,j}$. We further define a diagonal block matrix $\mathbf{Q}$ as $\mathbf{Q} \triangleq \texttt{Bdiag}(\{[\mathbf{Q}_j]\}_{j\in\mathcal{V}}$ with increasing order on j, and $\mathbf{v}^{(\ell)}$ and $\mathbf{b}$ be the vectors containing $\mathbf{v}_j$ and $\mathbf{b}_j$, respectively, with the same stacking order as $\mathbf{Q}_j$. Following (21), we have

$$\mathbf{v}^{(\ell)} = -\mathbf{Q}\mathbf{v}^{(\ell-1)} + \mathbf{b}. \tag{22}$$

For this linear updating equation, it is well known that, for arbitrary initial value $\mathbf{v}^{(0)}$, $\mathbf{v}^{(\ell)}$ converges if and only if the spectral radius $\rho(\mathbf{Q}) < 1$. Note that an algorithmically we to check this condition in a distributed manner is provided in [23]. As convergence of $\mathbf{v}^{(\ell)}$ depends on the convergence of $\mathbf{C}^{(\ell)}$, we have the following result.

Theorem 2. *The vector sequence $\left\{\boldsymbol{v}^{(\ell)}\right\}_{l=0,1,\ldots}$ defined by (22) converges to a unique value for any initial value $\left\{\boldsymbol{v}^{(0)}\right\}$ and initial covariance matrix $\boldsymbol{C}^{(0)} \succeq \mathbf{0}$ if and only if $\rho(\mathbf{Q}) < 1$.*

According to (14), the convergence of $\boldsymbol{\mu}^{(l)}_i$ depends on $\left[\mathbf{C}^{(l)}_{f_{i,j}\to i}\right]^{-1}$ and $\mathbf{v}^{(l)}_{f_{i,j}\to i}$. As Theorem 1 shows that $\left[\mathbf{C}^{(l)}_{f_{i,j}\to i}\right]^{-1}$ is convergence guaranteed with arbitrary positive semidefinite initial value, the convergence condition of $\boldsymbol{\mu}^{(l)}_i$ is equivalent to the convergence of $\mathbf{v}^{(l)}_{f_{i,j}\to i}$. Moreover, as shown in [15], once $\boldsymbol{\mu}^{(l)}_i$ converges, it converges to $\hat{\mathbf{x}}_i$. We therefore conclude that the necessary and sufficient convergence condition of $\boldsymbol{\mu}^{(l)}_i$ to the optimal estimate is $\rho(\mathbf{Q}) < 1$.

4. CONCLUSION

In this paper, we have studied distributed inference using Gaussian belief propagation (BP) over networks with two neighboring agents sharing a common observation. We have analyzed the convergence property of the Gaussian BP algorithm for this particular model. We have shown analytically that, with arbitrary positive semidefinite matrix initialization, the message information matrix exchanged among agents converges at a geometric rate to a unique positive definite matrix. Moreover, we have presented the necessary and sufficient condition for convergence under which the belief mean vector converges to the optimal centralized estimate.

5. REFERENCES

[1] Y. Hu, A. Kuh, T. Yang, and A. Kavcic, "A belief propagation based power distribution system state estimator," *IEEE Comput. Intell. Mag.*, vol. 6, no. 3, pp. 36–46, 2011.

[2] J. Du and Y.-C. Wu, "Distributed clock skew and offset estimation in wireless sensor networks: Asynchronous algorithm and convergence analysis," *IEEE Trans. Wireless Commun.*, vol. 12, no. 11, pp. 5908–5917, Nov 2013.

[3] ——, "Network-wide distributed carrier frequency offsets estimation and compensation via belief propagation," *IEEE Trans. Signal Process.*, vol. 61, no. 23, pp. 5868–5877, 2013.

[4] ——, "Fully distributed clock skew and offset estimation in wireless sensor networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 4499–4503.

[5] Y. Zhou, H. Liu, Z. Pan, L. Tian, J. Shi, and G. Yang, "Two-stage cooperative multicast transmission with optimized power consumption and guaranteed coverage," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 274–284, 2014.

[6] Y. Zhou, H. Liu, Z. Pan, L. Tian, and J. Shi, "Spectral - and energy-efficient two-stage cooperative multicast for lte-advanced and beyond," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 34–41, 2014.

[7] H. Wymeersch, J. Lien, and M. Z. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427–450, 2009.

[8] B. J. Frey, "Local probability propagation for factor analysis," in *Neural Information Processing Systems (NIPS)*, Dec 1999, pp. 442–448.

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

[10] D. Bickson and D. Malkhi, "A unifying framework for rating users and data items in peer-to-peer and social networks," *Peer-to-Peer Networking and Applications (PPNA) Journal*, vol. 1, no. 2, pp. 93–103, 2008.

[11] J. Du, S. Ma, Y.-C. Wu, and H. V. Poor, "Distributed bayesian hybrid power state estimation with PMU synchronization errors," in *Global Communications Conference, 2014 IEEE*, 2014, pp. 3174–3179.

[12] 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, Mar. 2001.

[13] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *Journal of Machine Learning Research*, vol. 7, no. 2, pp. 2031–2064, Feb. 2006.

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

[15] J. Du, S. Ma, Y.-C. Wu, S. Kar, and J. M. F. Moura, "Convergence analysis of distributed inference with vector-valued Gaussian belief propagation," submitted for publication [Preprint Available]: https://users.ece.cmu.edu/~soummyak/GBP_convergence.

[16] B. L. Ng, J. Evans, S. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, no. 12, pp. 5491–5499, Dec 2008.

[17] J. Du and Y.-C. Wu, "Distributed cfos estimation and compensation in multi-cell cooperative networks," in *International Conference on Information and Communication Technology Convergence*, 2013.

[18] S. Kar and J. M. F. Moura, "Consensus+innovations distributed inference over networks: cooperation and sensing in networked systems," *IEEE Signal Process. Mag.*, vol. 30, no. 3, pp. 99–109, 2013.

[19] S. Kar, J. M. F. Moura, and H. Poor, "Distributed linear parameter estimation: asymptotically efficient adaptive strategies," *SIAM Journal on Control and Optimization*, vol. 51, no. 3, pp. 2200–2229, 2013.

[20] F. S. Cattivelli and A. H. Sayed, "Diffusion LMS strategies for distributed estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1035–1048, 2010.

[21] S. Kar, "Large scale networked dynamical systems: Distributed inference," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, Department of Electrical and Computer Engineering, June 2010.

[22] J. Du, S. Ma, Y.-C. Wu, S. Kar, and J. M. F. Moura, "Convergence analysis of the information matrix in Gaussian belief propagation," in *Proc. IEEE Acoustics, Speech and Signal Processing Conf. (ICASSP 2017)*, 2017.

[23] J. Du, , S. Kar, and J. M. F. Moura, "Distributed convergence verification for gaussian belief propagation," to appear in 2017 Asilomar Conference on Signals, Systems, and Computers.