

On Convergence Conditions of Gaussian Belief Propagation

Qinliang Su and Yik-Chung Wu

Abstract

In order to compute the marginal probability density function (PDF) with Gaussian belief propagation (BP), it is important to know whether it will converge in advance. By describing the message-passing process of Gaussian BP on the pairwise factor graph as a set of updating functions, the necessary and sufficient convergence condition of beliefs in synchronous Gaussian BP is first derived under a newly proposed initialization set. The proposed initialization set is proved to be largest among all currently known sets. Then, the necessary and sufficient convergence condition of beliefs in damped Gaussian BP is developed, with the allowable range of damping factor explicitly established. The results theoretically confirm the extensively reported conjecture that damping is helpful to improve the convergence of Gaussian BP. Under totally asynchronous scheduling, a sufficient convergence condition of beliefs is also derived for the same proposed initialization set. Relationships between the proposed convergence conditions and existing ones are established analytically. At last, numerical examples are presented to corroborate the established theories.

Index Terms

Gaussian belief propagation, loopy belief propagation, graphical model, factor graph, message passing, sum-product algorithm, convergence.

I. INTRODUCTION

Many problems in signal processing and machine learning eventually come to the issue of computing marginal probability density function (PDF) from a high dimensional joint PDF. In

Copyright (c) 2014 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

The authors are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, (e-mail: qlsu@eee.hku.hk, ycwu@eee.hku.hk).

general, the complexity of directly computing the marginal PDF could be very high. By passing messages among the neighboring nodes in factor graph, belief propagation (BP) provides an efficient way to compute the approximate marginal PDFs upon convergence [1]–[5].

In this paper, we focus on an important class of BP where the underlying joint PDF is Gaussian. In this case, the messages being passed in the factor graph maintain Gaussian form, and hence the updating of beliefs can be described by the updating of belief variance and belief mean. Due to the ability of efficiently computing true marginal mean [6], Gaussian BP has been successfully applied in many areas, such as MMSE multi-user detection, equalization and channel estimation in communication systems [7]–[9], fast solver for systems of linear equations [10], [11], sparse Bayesian learning in large-scale compressed sensing problem [12], and estimation on Gaussian graphical model [13], [14]. Moreover, because of the intrinsic distributed characteristic, Gaussian BP has been applied in many problems requiring distributed information processing, such as distributed beamforming [15], distributed utility maximization in large scale network [16], distributed synchronization and localization in wireless sensor networks [17]–[19], distributed energy efficient self-deployment in mobile sensor networks [20], distributed rate control in Ad Hoc networks [21], and inter-cell interference mitigation [22].

Gaussian BP can only work under the prerequisite that beliefs do converge. So far, several sufficient conditions ensuring the convergence of beliefs under a designated initialization set have been proposed, such as diagonal dominance [6], convex decomposition [23] and walk-summability [24]. These convergence conditions are derived to be applicable to all possible schedulings, thus are expected to be more conservative than the convergence condition for a particular scheduling. As synchronous scheduling is the most direct and widely applicable one, in this paper, synchronous scheduling is considered separately from asynchronous schedulings, resulting in two different convergence conditions in these two cases.

The contributions of this paper are summarized as follow.

- 1) For synchronous Gaussian BP, the necessary and sufficient convergence condition of beliefs under a newly proposed initialization set is derived. The proposed initialization set is proved to be the largest among all currently known results.
- 2) For damped Gaussian BP, the necessary and sufficient convergence condition of beliefs under the proposed initialization set is also derived. It is proved that this convergence condition is more relaxed than that of Gaussian BP. To the best of our knowledge, this is the first

convergence condition of damped Gaussian BP, and theoretically confirms the extensively reported conjecture that damping is helpful to improve the convergence of Gaussian BP [24]–[28]. A method on how to choose an appropriate damping factor is also proposed.

- 3) For asynchronous Gaussian BP, a sufficient condition is derived to guarantee the convergence of beliefs under all totally asynchronous schedulings.
- 4) Relationships between the proposed convergence conditions and existing ones are established analytically. The results demonstrate that the existing conditions are implied by the proposed ones.

The rest of this paper is organized as follows. Gaussian BP is reviewed in Section II. Section III analyzes the convergence condition of messages in Gaussian BP. Convergence conditions of beliefs in synchronous Gaussian BP, damped Gaussian BP and asynchronous Gaussian BP are derived in Section IV. Relationships between the proposed convergence conditions and existing ones are established in Section V. Numerical examples are presented in Section VI, which is followed by conclusions in Section VII.

The following notations are used throughout this paper. Symbols $\text{diag}(\mathbf{x})$ and $\text{Bdiag}(\mathbf{X}_1, \mathbf{X}_2, \dots)$ denote a diagonal matrix and block diagonal matrix with the elements of $\mathbf{x}$ and $\mathbf{X}_i$ located along the main diagonal, respectively. For two vectors, $\mathbf{x}_1 \geq \mathbf{x}_2$ and $\mathbf{x}_1 > \mathbf{x}_2$ mean the inequalities hold in all corresponding elements. Notations $\lambda_{\max}(\mathbf{G})$ and $\text{eig}_{\max}(\mathbf{G})$ represent the eigenvalue of matrix $\mathbf{G}$ with maximum module and its corresponding eigenvector, respectively. Notation $\lambda(\mathbf{G})$ means any eigenvalue of matrix $\mathbf{G}$, while $\rho(\mathbf{G})$ represents the spectral radius of $\mathbf{G}$. Symbols $|\cdot|$, $\Re(\cdot)$ and $\Im(\cdot)$ are the module, real part and imaginary part of a complex number, respectively.

II. GAUSSIAN BELIEF PROPAGATION

Consider a Gaussian PDF $f(\mathbf{x}) \propto \exp\{-\frac{1}{2}\mathbf{x}^T \mathbf{P} \mathbf{x} + \mathbf{h}^T \mathbf{x}\}$, where $\mathbf{x} = [x_1, x_2, \dots, x_N]^T$ is the random variable vector; $\mathbf{P} \succ 0$ is the precision matrix with p_{ij} being its (i, j) -th element; and $\mathbf{h} = [h_1, h_2, \dots, h_N]^T$. The Gaussian PDF can be written in a factorized form $f(\mathbf{x}) \propto \prod_{i=1}^N f_i(x_i) \prod_{j=1}^N \prod_{k=j+1}^N f_{jk}(x_j, x_k)$, where $f_i(x_i) = \exp\{-\frac{p_{ii}}{2}x_i^2 + h_i x_i\}$ and $f_{jk}(x_j, x_k) = \exp\{-p_{jk}x_j x_k\}$. Based on this expansion, a factor graph can be constructed by connecting each variable x_i with its associated factors $f_i(x_i)$ and $f_{ij}(x_i, x_j)$. Then, the messages of Gaussian BP

being passed from variable node i to node j are updated as

$$m_{i \rightarrow j}^d(x_i, t) \propto \prod_{k \in \mathcal{N}(i) \setminus j} m_{k \rightarrow i}^a(x_i, t) f_i(x_i), \quad (1)$$

$$m_{i \rightarrow j}^a(x_j, t+1) \propto \int m_{i \rightarrow j}^d(x_i, t) f_{ij}(x_i, x_j) dx_i, \quad (2)$$

where $(i, j) \in \mathcal{E} \triangleq \{(i, j) | i \neq j, p_{ij} \neq 0, \forall i, j \in \mathcal{V}\}$ with $\mathcal{V} \triangleq \{1, 2, \dots, N\}$; $\mathcal{N}(i) \triangleq \{j | j \neq i, p_{ij} \neq 0, \forall j \in \mathcal{V}\}$ represents the index set of neighboring variable nodes of node i ¹; and $\mathcal{N}(i) \setminus j$ is the set $\mathcal{N}(i)$ except node j . After obtaining the messages $m_{k \rightarrow i}^a(x_i, t)$, the belief at variable node i is computed as

$$b_i(x_i, t) \propto \prod_{k \in \mathcal{N}(i)} m_{k \rightarrow i}^a(x_i, t) f_i(x_i). \quad (3)$$

Without loss of generality, let the arriving message of node j at time t is in form of $m_{i \rightarrow j}^a(x_j, t) \propto \exp \left\{ -\frac{v_{i \rightarrow j}^a(t)}{2} x_j^2 + \beta_{i \rightarrow j}^a(t) x_j \right\}$, where $v_{i \rightarrow j}^a(t)$ and $\beta_{i \rightarrow j}^a(t)$ are the arriving precision and arriving linear coefficient, respectively. Inserting $m_{i \rightarrow j}^a(x_j, t)$ into (1), we obtain $m_{i \rightarrow j}^d(x_i, t) \propto \exp \left\{ -\frac{v_{i \rightarrow j}^d(t)}{2} x_i^2 + \beta_{i \rightarrow j}^d(t) x_i \right\}$, where

$$v_{i \rightarrow j}^d(t) = p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(t), \quad (4)$$

$$\beta_{i \rightarrow j}^d(t) = h_i + \sum_{k \in \mathcal{N}(i) \setminus j} \beta_{k \rightarrow i}^a(t) \quad (5)$$

are the departing precision and linear coefficient, respectively. Furthermore, substituting the departing message $m_{i \rightarrow j}^d(x_i, t)$ into (2), we obtain

$$m_{i \rightarrow j}^a(x_j, t+1) \propto \exp \left\{ \frac{p_{ij}^2}{2v_{i \rightarrow j}^d(t)} x_j^2 - \frac{p_{ij}\beta_{i \rightarrow j}^d(t)}{v_{i \rightarrow j}^d(t)} x_j \right\} \times \int \exp \left\{ -\frac{v_{i \rightarrow j}^d(t)}{2} \left(x_i - \frac{\beta_{i \rightarrow j}^d(t) - p_{ij}x_j}{v_{i \rightarrow j}^d(t)} \right)^2 \right\} dx_i. \quad (6)$$

If $v_{i \rightarrow j}^d(t) > 0$, the integration equals to a constant, and $m_{i \rightarrow j}^a(x_j, t+1) \propto \exp \left\{ \frac{p_{ij}^2}{2v_{i \rightarrow j}^d(t)} x_j^2 - \frac{p_{ij}\beta_{i \rightarrow j}^d(t)}{v_{i \rightarrow j}^d(t)} x_j \right\}$.

Thus, it can be obtained that

$$v_{i \rightarrow j}^a(t+1) = \begin{cases} \frac{-p_{ij}^2}{v_{i \rightarrow j}^d(t)}, & \text{if } v_{i \rightarrow j}^d(t) > 0, \\ \text{not defined,} & \text{otherwise,} \end{cases} \quad (7)$$

¹Notice that in a factor graph, the neighboring variable nodes i and j are not connected directly but through a factor node $f_{ij}(x_i, x_j)$.

$$\beta_{i \rightarrow j}^a(t+1) = \begin{cases} \frac{-p_{ij}\beta_{i \rightarrow j}^d(t)}{v_{i \rightarrow j}^d(t)}, & \text{if } v_{i \rightarrow j}^d(t) > 0, \\ \text{not defined,} & \text{otherwise.} \end{cases} \quad (8)$$

After obtaining $v_{k \rightarrow i}^a(t)$ and $\beta_{k \rightarrow i}^a(t)$, the belief variance and belief mean at node i and iteration t are computed as

$$\sigma_i^2(t) = \frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}, \quad (9)$$

$$\mu_i(t) = \frac{h_i + \sum_{k \in \mathcal{N}(i)} \beta_{k \rightarrow i}^a(t)}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}. \quad (10)$$

Now, we express the message update process in vector form, which will be useful for the rest of this paper. By substituting $v_{i \rightarrow j}^d(t) = p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(t)$ from (4) into (7) and then writing $v_{i \rightarrow j}^a(t+1)$ into a vector form, we obtain

$$\mathbf{v}^a(t+1) = \begin{cases} \mathbf{g}(\mathbf{v}^a(t)), & \text{if } \mathbf{v}^a(t) \in \mathcal{W}, \\ \text{not defined,} & \text{otherwise,} \end{cases} \quad (11)$$

where $\mathbf{v}^a(t)$ contains $v_{i \rightarrow j}^a(t)$ with $(i, j) \in \mathcal{E}$ arranged in ascending order first on j and then on i ; $\mathbf{g}(\cdot)$ is a vector-valued function containing components

$$g_{ij}(\mathbf{w}) \triangleq -\frac{p_{ij}^2}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}} \quad (12)$$

with $(i, j) \in \mathcal{E}$ arranged in ascending order first on j and then on i ; set $\mathcal{W}$ is defined as

$$\mathcal{W} \triangleq \left\{ \mathbf{w} \mid p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} > 0, \forall (i, j) \in \mathcal{E} \right\}; \quad (13)$$

and $\mathbf{w}$ contains w_{ij} with $(i, j) \in \mathcal{E}$ arranged in ascending order first on j and then on i . For convenience of understanding, the updating process of $v_{i \rightarrow j}^a(t+1)$ is illustrated in Fig. 1.

Similarly, by substituting $v_{i \rightarrow j}^d(t) = p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(t)$ from (4) and $\beta_{i \rightarrow j}^d(t) = h_i + \sum_{k \in \mathcal{N}(i) \setminus j} \beta_{k \rightarrow i}^a(t)$ from (5) into (8) and then writing $\beta_{i \rightarrow j}^a(t+1)$ into a vector form, we obtain

$$\beta^a(t+1) = \begin{cases} \mathbf{G}(t)\beta^a(t) + \mathbf{b}(t), & \text{if } \mathbf{v}^a(t) \in \mathcal{W}, \\ \text{not defined,} & \text{otherwise,} \end{cases} \quad (14)$$

where $\beta^a(t)$ is a vector containing $\beta_{i \rightarrow j}^a(t)$ with $(i, j) \in \mathcal{E}$ arranged in ascending order first on j then on i ; $\mathbf{G}(t)$ and $\mathbf{b}(t)$ are defined as

$$\mathbf{G}(t) \triangleq -\text{diag}^{-1}(\mathbf{A}\mathbf{v}^a(t) + \mathbf{u})\text{diag}(\mathbf{p})\mathbf{A}, \quad (15)$$

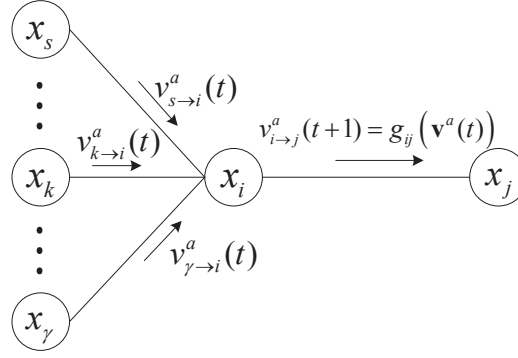


Fig. 1: Illustration of the updating process of $v_{i \rightarrow j}^a(t+1)$, where $g_{ij}(v^a(t)) = -\frac{p_{ij}^2}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(t)}$.

$$\mathbf{b}(t) \triangleq -\text{diag}^{-1}(\mathbf{A}\mathbf{v}^a(t) + \mathbf{u})\text{diag}(\mathbf{p})\boldsymbol{\xi}; \quad (16)$$

$\mathbf{A}$ is a $|\mathcal{E}| \times |\mathcal{E}|$ matrix defined such that $\mathbf{A}\mathbf{v}^a(t)$ is a column vector containing elements $\sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(t)$ with (i, j) arranged first on j and then on i ; $\mathbf{p}$ contains elements p_{ij} with $(i, j) \in \mathcal{E}$ arranged in ascending order first on j then on i ; $\mathbf{u} = [\mathbf{u}_1^T, \mathbf{u}_2^T, \dots, \mathbf{u}_N^T]^T$ with $\mathbf{u}_i$ being a column vector containing elements p_{jj} for all $j \in \mathcal{N}(i)$ arranged in ascending order; $\boldsymbol{\xi} = [\boldsymbol{\xi}_1^T, \boldsymbol{\xi}_2^T, \dots, \boldsymbol{\xi}_N^T]^T$ with $\boldsymbol{\xi}_i$ being a column vector containing elements h_j for all $j \in \mathcal{N}(i)$ arranged in ascending order.

Remark 1: The results in this paper are established under the prerequisite of pairwise factor graph, thus might not be applicable to other types of factor graphs. However, it should be emphasized that the pairwise factor graph is the most widely used one in Gaussian BP, and has been applied in the derivation of existing convergence conditions, e.g. diagonal dominance [6], convex decomposition [23] and walk-summability [24].

III. CONVERGENCE OF MESSAGES IN GAUSSIAN BP

In this section, we derive the convergence condition of messages, which are parameterized by $\mathbf{v}^a(t)$ and $\boldsymbol{\beta}^a(t)$. First, we present the convergence condition of $\mathbf{v}^a(t)$ only.

Theorem 1. [29] *Under any scheduling², $\mathbf{v}^a(t)$ converges to the same point for all $\mathbf{v}^a(0) \in \mathcal{A}$ if and only if $\mathcal{S}_1 \neq \emptyset$, where*

$$\mathcal{S}_1 \triangleq \{\mathbf{w} \mid \mathbf{w} \leq \mathbf{g}(\mathbf{w}) \text{ and } \mathbf{w} \in \mathcal{W}\}, \quad (17)$$

$$\mathcal{A} \triangleq \{\mathbf{w} \geq \mathbf{0}\} \cup \{\mathbf{w} \geq \mathbf{w}_0 \mid \mathbf{w}_0 \in \text{int}(\mathcal{S}_1)\} \cup \left\{ \mathbf{w} \geq \mathbf{w}_0 \mid \mathbf{w}_0 \in \mathcal{S}_1 \text{ and } \mathbf{w}_0 = \lim_{t \rightarrow \infty} \mathbf{g}^{(t)}(\mathbf{0}) \right\} \quad (18)$$

with $\mathbf{g}^{(t)}(\mathbf{w}) \triangleq \mathbf{g}(\mathbf{g}^{(t-1)}(\mathbf{w}))$ and $\mathbf{g}^{(0)}(\mathbf{w}) \triangleq \mathbf{w}$.

To determine whether $\mathcal{S}_1$ is $\emptyset$, we define the following semi-definite programming (SDP) problem:

$$\begin{aligned} \min_{\mathbf{w}} \quad & \|\mathbf{w}\|^2 \\ \text{s.t.} \quad & \begin{bmatrix} p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} & p_{ij} \\ p_{ij} & -w_{ij} \end{bmatrix} \succeq 0, \quad \forall (i, j) \in \mathcal{E}. \end{aligned} \quad (19)$$

The SDP problem (19) can be solved efficiently [31] by existing softwares, such as CVX [32] and SeDuMi [33], etc. Now, the following proposition can be obtained.

Proposition 1. $\mathcal{S}_1 \neq \emptyset$ if and only if (19) has a feasible solution.

Proof: Using Schur complement, the constraints in (19) can be equivalently written as

$$w_{ij} - g_{ij}(\mathbf{w}) \leq 0, \quad \forall (i, j) \in \mathcal{E}; \quad (20)$$

$$-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} \leq 0, \quad \forall (i, j) \in \mathcal{E}. \quad (21)$$

If $\mathcal{S}_1 \neq \emptyset$, there always exists a $\mathbf{w} \in \mathcal{S}_1$. From the definition of $\mathcal{S}_1$ in (17), we have $\mathbf{w} \leq \mathbf{g}(\mathbf{w})$ and $\mathbf{w} \in \mathcal{W}$. That is, $w_{ij} - g_{ij}(\mathbf{w}) \leq 0$ and $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} < 0$ for all $(i, j) \in \mathcal{E}$. It can be seen that the vector $\mathbf{w}$ satisfies the constraints (20) and (21), thus it is a feasible solution to (19).

On the other hand, if (19) has a feasible solution $\mathbf{w}$, then $\mathbf{w}$ must satisfy (20) and (21), that is, $w_{ij} - g_{ij}(\mathbf{w}) \leq 0$ and $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} \leq 0$. Now, suppose $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} = 0$. Obviously, in this case, the determinant of the matrix in the constraint of (19) equals to

²The schedule requires every component updates infinitely many times as t goes to infinity, but the order and frequency of update can be arbitrary.

$-p_{ij}^2 < 0$. Hence, the matrix in the constraint of (19) is not positive semi-definite, which contradicts to the prerequisite that $\mathbf{w}$ is a feasible solution of (19). Thus, we must have $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} \neq 0$. Combining with the established result $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} \leq 0$, we obtain $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki} < 0$. Together with the established result $w_{ij} - g_{ij}(\mathbf{w}) \leq 0$, from the definition of $\mathcal{S}_1$, we can infer that $\mathbf{w} \in \mathcal{S}_1$, and thus $\mathcal{S}_1 \neq \emptyset$. ■

In addition to determining whether $\mathcal{S}_1$ is $\emptyset$, the optimal solution of (19) also has an important meaning.

Proposition 2. Under $\mathcal{S}_1 \neq \emptyset$, the optimal solution of (19) $\mathbf{w}^* = \lim_{t \rightarrow \infty} \mathbf{v}^a(t)$ for $\mathbf{v}^a(0) \in \mathcal{A}$.

Proof: First, we establish two facts about $\mathbf{w}^*$. Since $\mathbf{w}^*$ is the optimal solution of (19), the constraint $w_{ij}^* - g_{ij}(\mathbf{w}^*) \leq 0$ in (20) holds for all $(i, j) \in \mathcal{E}$, which is equivalent to

$$\mathbf{w}^* \leq \mathbf{g}(\mathbf{w}^*). \quad (22)$$

Moreover, for the optimal solution $\mathbf{w}^*$, the constraint $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^* \leq 0$ in (21) also holds. Notice that if $p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^* = 0$, the function $g_{ij}(\mathbf{w}^*) = -\frac{p_{ij}^2}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*}$ becomes undefined. Thus, we have $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^* < 0$ for all $(i, j) \in \mathcal{E}$, or equivalently $\mathbf{w}^* \in \mathcal{W}$. Now, substituting $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^* < 0$ into (12) gives $g_{ij}(\mathbf{w}^*) < 0$ for all $(i, j) \in \mathcal{E}$, which is equivalent to $\mathbf{g}(\mathbf{w}^*) < \mathbf{0}$. Combining with $\mathbf{w}^* \leq \mathbf{g}(\mathbf{w}^*)$ gives

$$\mathbf{w}^* < \mathbf{0}. \quad (23)$$

Next, the first-order derivative of $g_{ij}(\mathbf{w})$ with respect to w_{ki} for $k \in \mathcal{N}(i) \setminus j$ is equal to

$$\frac{\partial g_{ij}(\cdot)}{\partial w_{ki}} = \frac{p_{ij}^2}{(p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki})^2}. \quad (24)$$

Obviously, it can be seen that $\frac{\partial g_{ij}(\cdot)}{\partial w_{ki}} > 0$. Together with the fact $g_{ij}(\mathbf{w})$ is continuous for $\mathbf{w} \in \mathcal{W}$, we can infer that $g_{ij}(\mathbf{w})$ is a monotonically increasing function for $\mathbf{w} \in \mathcal{W}$. Thus, by applying $\mathbf{g}(\cdot)$ on $\mathbf{w}^* < \mathbf{0}$ in (23), we obtain $\mathbf{g}(\mathbf{w}^*) \leq \mathbf{g}(\mathbf{0})$. Combining with the constraint $\mathbf{w}^* \leq \mathbf{g}(\mathbf{w}^*)$ in (22) gives $\mathbf{w}^* \leq \mathbf{g}(\mathbf{0})$. Then, applying $\mathbf{g}(\cdot)$ to $\mathbf{w}^* \leq \mathbf{g}(\mathbf{0})$ leads to $\mathbf{g}(\mathbf{w}^*) \leq \mathbf{g}^{(2)}(\mathbf{0})$. Due to $\mathbf{w}^* \leq \mathbf{g}(\mathbf{w}^*)$, we further have $\mathbf{w}^* \leq \mathbf{g}^{(2)}(\mathbf{0})$. By induction, it can be inferred that $\mathbf{w}^* \leq \mathbf{g}^{(t)}(\mathbf{0})$ for all $t \geq 0$. By taking the limit on both sides of $\mathbf{w}^* \leq \mathbf{g}^{(t)}(\mathbf{0})$, and due to $\mathbf{g}^{(t)}(\mathbf{0}) = \mathbf{v}^a(t)$ with $\mathbf{v}^a(0) = \mathbf{0}$, we obtain $\mathbf{w}^* \leq \lim_{t \rightarrow \infty} \mathbf{v}^a(t)$ under the initialization of $\mathbf{v}^a(0) = \mathbf{0}$. Since according

to Theorem 1, $\lim_{t \rightarrow \infty} \mathbf{v}^a(t)$ with $\mathbf{v}^a(0) = \mathbf{0}$ is the same as $\lim_{t \rightarrow \infty} \mathbf{v}^a(t)$ with $\mathbf{v}^a(0) \in \mathcal{A}$, we further have

$$\mathbf{w}^* \leq \lim_{t \rightarrow \infty} \mathbf{v}^a(t) \quad (25)$$

for any $\mathbf{v}^a(0) \in \mathcal{A}$.

Finally, since $\mathbf{v}^a(t)$ converges to $\lim_{t \rightarrow \infty} \mathbf{v}^a(t)$ for $\mathbf{v}^a(0) \in \mathcal{A}$, according to (11), this means that $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) = \mathbf{g} \left(\lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right)$ and $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) \in \mathcal{W}$. Writing the two conditions into scalar form, we obtain $\lim_{t \rightarrow \infty} v_{i \rightarrow j}^a(t) = g_{ij} \left(\lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right)$ and $p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} \lim_{t \rightarrow \infty} v_{i \rightarrow k}^a(t) > 0$ for all $(i, j) \in \mathcal{E}$. Comparing with the constraints (20) and (21), it can be easily seen that $\lim_{t \rightarrow \infty} \mathbf{v}^a(t)$ is a feasible solution of (19). Since $\mathbf{w}^*$ is the optimal solution of the minimization problem (19), we must have

$$\|\mathbf{w}^*\|^2 \leq \left\| \lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right\|^2. \quad (26)$$

On the other hand, substituting $p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} \lim_{t \rightarrow \infty} v_{i \rightarrow k}^a(t) > 0$ into (12), we obtain $g_{ij} \left(\lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right) < 0$ for all $(i, j) \in \mathcal{E}$, or equivalently $\mathbf{g} \left(\lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right) < \mathbf{0}$. Due to $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) = \mathbf{g} \left(\lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right)$, we have $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) < \mathbf{0}$. Combining with (25), it can be inferred that

$$\|\mathbf{w}^*\|^2 \geq \left\| \lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right\|^2. \quad (27)$$

Comparing (26) with (27), it can be seen that $\|\mathbf{w}^*\|^2 = \left\| \lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right\|^2$. Suppose $\mathbf{w}^* \neq \lim_{t \rightarrow \infty} \mathbf{v}^a(t)$. According to (25), it can be inferred that $\|\mathbf{w}^*\|^2 < \left\| \lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right\|^2$, which contradicts with the established result $\|\mathbf{w}^*\|^2 = \left\| \lim_{t \rightarrow \infty} \mathbf{v}^a(t) \right\|^2$. Therefore, we must have $\mathbf{w}^* = \lim_{t \rightarrow \infty} \mathbf{v}^a(t)$. ■

Now, we derive the convergence condition for the message parameters $(\mathbf{v}^a(t), \beta^a(t))$. If $\mathbf{v}^a(t)$ has already converged to $\mathbf{w}^*$, then $\beta^a(t)$ is updated as

$$\beta^a(t+1) = \mathbf{G}^* \beta^a(t) + \mathbf{b}^*, \quad (28)$$

where

$$\mathbf{G}^* \triangleq -\text{diag}^{-1}(\mathbf{A}\mathbf{w}^* + \mathbf{u})\text{diag}(\mathbf{p})\mathbf{A}, \quad (29)$$

$$\mathbf{b}^* \triangleq -\text{diag}^{-1}(\mathbf{A}\mathbf{w}^* + \mathbf{u})\text{diag}(\mathbf{p})\boldsymbol{\xi}. \quad (30)$$

It is well-known that $\beta^a(t)$ in (28) converges for all choices of $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$ if and only if $\rho(\mathbf{G}^*) < 1$ [30].

However, in Gaussian BP, $\mathbf{v}^a(t)$ and $\beta^a(t)$ are updated alternatively, rather than waiting for one quantity to converge before starting iterations of another. The following theorem shows

that even with simultaneous updating, the message parameters $(\mathbf{v}^a(t), \beta^a(t))$ converge under the same condition as if we wait for $\mathbf{v}^a(t)$ to converge first.

Theorem 2. *Message parameters $(\mathbf{v}^a(t), \beta^a(t))$ converge to the same point for all choices of $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$ if and only if $\mathcal{S}_1 \neq \emptyset$ and $\rho(\mathbf{G}^*) < 1$.*

Proof:

Sufficient Condition:

If $\mathcal{S}_1 \neq \emptyset$, according to Theorem 1, it is known that $\mathbf{v}^a(t)$ converges to the same point for all $\mathbf{v}^a(0) \in \mathcal{A}$. Thus, in order to prove the sufficiency, we only need to prove $\beta^a(t)$ also converges to the same point for all $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$.

Due to $\rho(\mathbf{G}^*) < 1$, we can infer that $\mathbf{I} - \mathbf{G}^*$ is invertible. Now, define the following vector

$$\phi \triangleq (\mathbf{I} - \mathbf{G}^*)^{-1} \mathbf{b}^*, \quad (31)$$

which can be equivalently written as $\phi = \mathbf{G}^* \phi + \mathbf{b}^*$. Subtracting it from (14) gives

$$\begin{aligned} \beta^a(t+1) - \phi &= \mathbf{G}(t) \beta^a(t) - \mathbf{G}^* \phi + \mathbf{b}(t) - \mathbf{b}^* \\ &= (\mathbf{G}^* + (\mathbf{G}(t) - \mathbf{G}^*)) (\beta^a(t) - \phi) + (\mathbf{G}(t) - \mathbf{G}^*) \phi + \mathbf{b}(t) - \mathbf{b}^*. \end{aligned} \quad (32)$$

According to (32), for any matrix norm $\|\cdot\|$, we have

$$\|\beta^a(t+1) - \phi\| \leq (\|\mathbf{G}^*\| + \|\mathbf{G}(t) - \mathbf{G}^*\|) \|\beta^a(t) - \phi\| + \|(\mathbf{G}(t) - \mathbf{G}^*) \phi + \mathbf{b}(t) - \mathbf{b}^*\|, \quad (33)$$

where $\|\cdot\|$ is the compatible vector norm associated with the matrix norm $\|\cdot\|$ [34, p. 297]. Denote $\rho(\mathbf{G}^*)$ as ρ . It is known that there always exists a specific matrix norm $\|\cdot\|_s$ such that $\|\mathbf{G}^*\|_s \leq \rho + \varepsilon_s$ for any $\varepsilon_s > 0$ [34]. Furthermore, due to $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) = \mathbf{w}^*$ from Proposition 2, it can be seen that $\lim_{t \rightarrow \infty} \mathbf{G}(t) = \mathbf{G}^*$, meaning that there exists an integer t_0 such that $\|\mathbf{G}(t) - \mathbf{G}^*\|_s \leq \varepsilon_G$ for all $t \geq t_0$ and any $\varepsilon_G > 0$. Thus, for $t \geq t_0$, we can derive from (33) that

$$\|\beta^a(t+1) - \phi\|_s \leq (\rho + \varepsilon_s + \varepsilon_G) \|\beta^a(t) - \phi\|_s + \|(\mathbf{G}(t) - \mathbf{G}^*) \phi + \mathbf{b}(t) - \mathbf{b}^*\|_s, \quad (34)$$

where $\|\cdot\|_s$ is the compatible vector norm of the matrix norm $\|\cdot\|_s$.

Now, we define another sequence $z(t)$ for $t \geq t_0$ as

$$\begin{aligned} z(t+1) &= (\rho + \varepsilon_s + \varepsilon_G)z(t) + \|(\mathbf{G}(t) - \mathbf{G}^*)\phi + \mathbf{b}(t) - \mathbf{b}^*\|_s \\ &= (\rho + \varepsilon_s + \varepsilon_G)^{t-t_0+1}z(t_0) + \sum_{i=t_0}^t (\rho + \varepsilon_s + \varepsilon_G)^{t-i} \|(\mathbf{G}(i) - \mathbf{G}^*)\phi + \mathbf{b}(i) - \mathbf{b}^*\|_s, \end{aligned} \quad (35)$$

where $z(t_0) \triangleq \|\beta^a(t_0) - \phi\|_s$. From (35), we further have

$$\begin{aligned} z(2t_1+1) &\leq (\rho + \varepsilon_s + \varepsilon_G)^{2t_1-t_0+1}z(t_0) + \sum_{i=t_0}^{t_1-1} (\rho + \varepsilon_s + \varepsilon_G)^{2t_1-i} \|(\mathbf{G}(i) - \mathbf{G}^*)\phi + \mathbf{b}(i) - \mathbf{b}^*\|_s \\ &\quad + \sum_{i=t_1}^{2t_1} (\rho + \varepsilon_s + \varepsilon_G)^{2t_1-i} \cdot \max_{2t_1 \geq t \geq t_1} \{ \|(\mathbf{G}(t) - \mathbf{G}^*)\phi + \mathbf{b}(t) - \mathbf{b}^*\|_s \}. \end{aligned} \quad (36)$$

Since $\mathbf{v}^a(t)$ converges for any $\mathbf{v}^a(0) \in \mathcal{A}$, according to (11), we must have $\mathbf{v}^a(t) \in \mathcal{W}$ for all $t \geq 0$, or equivalently in vector form

$$\mathbf{A}\mathbf{v}^a(t) + \mathbf{u} > \mathbf{0}, \quad \forall t \geq 0, \quad (37)$$

meaning that the entries in $\text{diag}^{-1}(\mathbf{A}\mathbf{v}^a(t) + \mathbf{u})$ are finite. Thus, $\mathbf{G}(t) = -\text{diag}^{-1}(\mathbf{A}\mathbf{v}^a(t) + \mathbf{u})\text{diag}(\mathbf{p})\mathbf{A}$ and $\mathbf{b}(t) = -\text{diag}^{-1}(\mathbf{A}\mathbf{v}^a(t) + \mathbf{u})\text{diag}(\mathbf{p})\boldsymbol{\xi}$ only contain finite entries, and so are $\mathbf{G}^* = \lim_{t \rightarrow \infty} \mathbf{G}(t)$ and $\mathbf{b}^* = \lim_{t \rightarrow \infty} \mathbf{b}(t)$. Therefore, the norm $\|(\mathbf{G}(t) - \mathbf{G}^*)\phi + \mathbf{b}(t) - \mathbf{b}^*\|_s$ for $t_0 \leq t \leq t_1$ is finite and can be upper bounded by a constant C_u . By applying the bound C_u to (36), we obtain

$$\begin{aligned} z(2t_1+1) &\leq (\rho + \varepsilon_s + \varepsilon_G)^{2t_1-t_0+1}z(t_0) + \frac{1 - (\rho + \varepsilon_s + \varepsilon_G)^{t_1-t_0}}{1 - (\rho + \varepsilon_s + \varepsilon_G)} (\rho + \varepsilon_s + \varepsilon_G)^{t_1+1}C_u \\ &\quad + \frac{1 - (\rho + \varepsilon_s + \varepsilon_G)^{t_1+1}}{1 - (\rho + \varepsilon_s + \varepsilon_G)} \cdot \max_{2t_1 \geq t \geq t_1} \{ \|(\mathbf{G}(t) - \mathbf{G}^*)\phi + \mathbf{b}(t) - \mathbf{b}^*\|_s \}. \end{aligned} \quad (38)$$

Due to $\lim_{t \rightarrow \infty} \{ \|(\mathbf{G}(t) - \mathbf{G}^*)\phi + \mathbf{b}(t) - \mathbf{b}^*\|_s \} = 0$, we have $\lim_{t_1 \rightarrow \infty} \max_{2t_1 \geq t \geq t_1} \{ \|(\mathbf{G}(t) - \mathbf{G}^*)\phi + \mathbf{b}(t) - \mathbf{b}^*\|_s \} = 0$. Thus, by choosing appropriate ε_s and ε_G so that $\rho + \varepsilon_s + \varepsilon_G < 1$, as $t_1 \rightarrow +\infty$, it can be inferred from (38) that $\lim_{t \rightarrow \infty} z(t) = 0$.

Subtracting (34) from (35) gives

$$(\rho + \varepsilon_\rho + \varepsilon_G)(z(t) - \|\beta^a(t) - \phi\|_s) \leq z(t+1) - \|\beta^a(t+1) - \phi\|_s. \quad (39)$$

Putting $t = t_0$ into (39), and since by definition $z(t_0) = \|\beta^a(t_0) - \phi\|_s$, it can be derived that $\|\beta^a(t_0+1) - \phi\|_s \leq z(t_0+1)$. In general, it can be recursively derived that $\|\beta^a(t) - \phi\|_s \leq z(t)$

for all $t \geq t_0$. Since $\lim_{t \rightarrow \infty} z(t) = 0$, we have $\lim_{t \rightarrow \infty} \beta^a(t) = \phi$, thus $\beta^a(t)$ converges to the same point.

Necessary Condition:

Since $\mathbf{v}^a(t)$ converges for all $\mathbf{v}^a(0) \in \mathcal{A}$, according to Theorem 1, we have $\mathcal{S}_1 \neq \emptyset$. Thus, to prove the necessity, we only need to prove $\rho(\mathbf{G}^*) < 1$ under the prerequisite of $\mathcal{S}_1 \neq \emptyset$.

First, from Proposition 2 and $\mathbf{0} \in \mathcal{A}$, we know that $\lim_{t \rightarrow \infty} \mathbf{g}^{(t)}(\mathbf{0}) = \mathbf{w}^*$. Together with $\mathbf{w}^* < \mathbf{0}$ from (23), set $\mathcal{A}$ defined in (18) reduces to $\mathcal{A} = \{\mathbf{w} \geq \mathbf{w}_0 | \mathbf{w}_0 \in \text{int}(\mathcal{S}_1)\} \cup \{\mathbf{w} \geq \mathbf{w}^*\}$. Thus, for any $\mathbf{v}^a(0) \in \mathcal{A}$, we can always find a $\gamma \in \text{int}(\mathcal{S}_1) \cup \{\mathbf{w}^*\}$ such that $\gamma \leq \mathbf{v}^a(0)$. By applying $\mathbf{g}(\cdot)$ to $\gamma \leq \mathbf{v}^a(0)$, from the monotonically increasing property of $\mathbf{g}(\cdot)$, we obtain

$$\mathbf{g}(\gamma) \leq \mathbf{v}^a(1). \quad (40)$$

If $\gamma \in \text{int}(\mathcal{S}_1)$, it is seen from (17) that $\gamma \leq \mathbf{g}(\gamma)$. On the other hand, if $\gamma = \mathbf{w}^*$, then $\gamma = \mathbf{g}(\gamma)$ since $\mathbf{w}^*$ is the converged point. Thus, with $\gamma \in \text{int}(\mathcal{S}_1) \cup \{\mathbf{w}^*\}$, we have $\gamma \leq \mathbf{g}(\gamma)$. Combining with (40) gives $\gamma \leq \mathbf{v}^a(1)$. By induction, it can be inferred that $\gamma \leq \mathbf{v}^a(t)$ for $t \geq 0$. Due to $\gamma \in \text{int}(\mathcal{S}_1) \cup \{\mathbf{w}^*\}$, from the definition of set $\mathcal{A}$ in (18), we have $\mathbf{v}^a(t) \in \mathcal{A}$ for all $t \geq 0$.

Next, consider two initializations $(\mathbf{v}^a(0), \beta_1^a(0))$ and $(\mathbf{v}^a(0), \beta_2^a(0))$ with $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta_1^a(0), \beta_2^a(0) \in \mathbb{R}^{|\mathcal{E}|}$. Due to $\beta_1^a(t+1) = \mathbf{G}(t)\beta_1^a(t) + \mathbf{b}(t)$ and $\beta_2^a(t+1) = \mathbf{G}(t)\beta_2^a(t) + \mathbf{b}(t)$ as given in (14), subtracting the two equations gives

$$\begin{aligned} \Delta\beta^a(t+1) &= \mathbf{G}(t)\Delta\beta^a(t) \\ &= \prod_{k=0}^t \mathbf{G}(k) \cdot \Delta\beta^a(0), \end{aligned} \quad (41)$$

where $\Delta\beta^a(t) = \beta_1^a(t) - \beta_2^a(t)$. Since $\beta_1^a(t)$ and $\beta_2^a(t)$ converges to the same point, then $\Delta\beta^a(t)$ converges to $\mathbf{0}$ for any $\mathbf{v}^a(0) \in \mathcal{A}$ and $\Delta\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$. Thus, according to (41), we can infer that $\lim_{t \rightarrow \infty} \prod_{k=0}^t \mathbf{G}(k) = \mathbf{0}$ for $\mathbf{v}^a(0) \in \mathcal{A}$. Similarly, for another $\bar{\mathbf{v}}^a(0) \in \mathcal{A}$, we also have $\lim_{t \rightarrow \infty} \prod_{k=0}^t \bar{\mathbf{G}}(k) = \mathbf{0}$, where $\bar{\mathbf{G}}(k) \triangleq -\text{diag}^{-1}(\mathbf{A}\bar{\mathbf{v}}^a(k) + \mathbf{u})\text{diag}(\mathbf{p})\mathbf{A}$ as defined in (15). Due to $\mathbf{v}^a(t_1) \in \mathcal{A}$ for any $t_1 \geq 0$ as proved above, we can choose $\bar{\mathbf{v}}^a(0) = \mathbf{v}^a(t_1)$, which means that $\bar{\mathbf{G}}(k) = \mathbf{G}(k+t_1)$. Substituting this result into $\lim_{t \rightarrow \infty} \prod_{k=0}^t \bar{\mathbf{G}}(k) = \mathbf{0}$, we obtain $\lim_{t \rightarrow \infty} \prod_{k=t_1}^{t+t_1} \mathbf{G}(k) = \mathbf{0}$ for all $t_1 \geq 0$. As $t_1 \rightarrow \infty$, using the fact $\lim_{k \rightarrow \infty} \mathbf{G}(k) = \mathbf{G}^*$, we obtain $\lim_{t \rightarrow \infty} \mathbf{G}^{*t} = \mathbf{0}$, which holds if and only if $\rho(\mathbf{G}^*) < 1$ [34, p. 298]. Thus, we have proved $\rho(\mathbf{G}^*) < 1$. ■

IV. CONVERGENCE OF BELIEFS IN GAUSSIAN BP

In this section, we will derive the convergence conditions of beliefs for synchronous Gaussian BP, damped Gaussian BP and asynchronous Gaussian BP, respectively.

A. Synchronous Gaussian BP

Theorem 3. *In synchronous Gaussian BP, belief parameters $(\sigma_i^2(t), \mu_i(t))$ converge to the same point for all choices of $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$ if and only if $\mathcal{S}_1 \neq \emptyset$, $\rho(\mathbf{G}^*) < 1$ and $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$.*

Proof:

Sufficient Condition:

If $\mathcal{S}_1 \neq \emptyset$, according to Proposition 2, it is known that $p_{ii} + \lim_{t \rightarrow \infty} \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t) = p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*$ for all $\mathbf{v}^a(0) \in \mathcal{A}$. Due to $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$, we can infer that belief variance $\sigma_i^2(t) = \frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}$ converges to $\frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*}$ for all $\mathbf{v}^a(0) \in \mathcal{A}$.

Next, if $\mathcal{S}_1 \neq \emptyset$ and $\rho(\mathbf{G}^*) < 1$, from Theorem 3, it can be inferred that $h_i + \lim_{t \rightarrow \infty} \sum_{k \in \mathcal{N}(i)} \beta_{k \rightarrow i}^a(t)$ exists and is unique for all $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$. Then, using the fact $\lim_{t \rightarrow \infty} \sigma_i^2(t) = \frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*}$ is unique for all $\mathbf{v}^a(0) \in \mathcal{A}$, we can infer that belief mean $\mu_i(t) = \frac{h_i + \sum_{k \in \mathcal{N}(i)} \beta_{k \rightarrow i}^a(t)}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}$ also converges to the same point for all $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$.

Necessary Condition:

First, we prove that if $\sigma_i^2(t)$ converges for all $\mathbf{v}^a(0) \in \mathcal{A}$, then $\mathcal{S}_1 \neq \emptyset$. For any $\mathbf{v}^a(0) \in \mathcal{A}$, from the definition of set $\mathcal{A}$ in (18), there exists a $\bar{\mathbf{v}}^a(0) \in \mathcal{A}$ satisfying $\mathbf{v}^a(0) \leq \bar{\mathbf{v}}^a(0)$ and $\mathbf{0} \leq \bar{\mathbf{v}}^a(0)$. Due to the monotonically increasing property of $\mathbf{g}(\cdot)$, applying $\mathbf{g}(\cdot)$ to $\mathbf{v}^a(0) \leq \bar{\mathbf{v}}^a(0)$ gives $\mathbf{v}^a(1) \leq \bar{\mathbf{v}}^a(1)$. On the other hand, substituting $\bar{\mathbf{v}}^a(0) \geq \mathbf{0}$ into (12) leads to $\bar{\mathbf{v}}^a(1) \leq \mathbf{0}$, and thereby $\bar{\mathbf{v}}^a(1) \leq \bar{\mathbf{v}}^a(0)$. Combining with $\mathbf{v}^a(1) \leq \bar{\mathbf{v}}^a(1)$, we obtain $\mathbf{v}^a(1) \leq \bar{\mathbf{v}}^a(1) \leq \bar{\mathbf{v}}^a(0)$. By applying $\mathbf{g}(\cdot)$ to the inequality repeatedly for $t \geq 0$ times, we obtain $\mathbf{v}^a(t+1) \leq \bar{\mathbf{v}}^a(t+1) \leq \bar{\mathbf{v}}^a(t)$. Due to $\bar{\mathbf{v}}^a(1) \leq \mathbf{0}$, we further have

$$\mathbf{v}^a(t+1) \leq \bar{\mathbf{v}}^a(t+1) \leq \bar{\mathbf{v}}^a(t) \leq \mathbf{0}, \quad \forall t \geq 1. \quad (42)$$

Since $\sigma_i^2(t)$ converges for any $\mathbf{v}^a(0) \in \mathcal{A}$, according to (11), we have $\mathbf{v}^a(t) \in \mathcal{W}$ for $t \geq 0$. From the definition of $\mathcal{W}$ in (13), this means $p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(t) > 0$ for all $(i, j) \in \mathcal{E}$, or equivalently $v_{\ell \rightarrow i}^a(t) > -p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j, \ell} v_{k \rightarrow i}^a(t)$. Using $\mathbf{v}^a(t) \leq \mathbf{0}$ for $t \geq 2$ from (42), we can

infer that $v_{\ell \rightarrow i}^a(t) > -p_{ii}$ for all $(\ell, i) \in \mathcal{E}$. Combining with $\mathbf{v}^a(t) \leq \bar{\mathbf{v}}^a(t)$ in (42), we can see that $\bar{\mathbf{v}}^a(t)$ is lower bounded. Together with the monotonically decreasing property of $\bar{\mathbf{v}}^a(t)$ from (42), we can infer that $\bar{\mathbf{v}}^a(t)$ converges, and thereby $\mathbf{g}\left(\lim_{t \rightarrow \infty} \bar{\mathbf{v}}^a(t)\right) = \lim_{t \rightarrow \infty} \bar{\mathbf{v}}^a(t)$. Furthermore, since $\bar{\mathbf{v}}^a(t) \geq \mathbf{v}^a(t)$ from (42) and $\mathbf{v}^a(t) \in \mathcal{W}$, it can be inferred from the definition of $\mathcal{W}$ in (13) that $\bar{\mathbf{v}}^a(t) \in \mathcal{W}$ for all $t \geq 0$, and thereby $\lim_{t \rightarrow \infty} \bar{\mathbf{v}}^a(t) \in \mathcal{W}$. Together with $\mathbf{g}\left(\lim_{t \rightarrow \infty} \bar{\mathbf{v}}^a(t)\right) = \lim_{t \rightarrow \infty} \bar{\mathbf{v}}^a(t)$, it can be seen from the definition of $\mathcal{S}_1$ in (17) that $\lim_{t \rightarrow \infty} \bar{\mathbf{v}}^a(t) \in \mathcal{S}_1$, and hence $\mathcal{S}_1 \neq \emptyset$.

Second, due to $\mathcal{S}_1 \neq \emptyset$, according to Proposition 2, it is known that $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) = \mathbf{w}^*$ for all $\mathbf{v}^a(0) \in \mathcal{A}$, and thus we have $p_{ii} + \lim_{t \rightarrow \infty} \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t) = p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*$. Since $\sigma_i^2(t) = \frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}$ converges for $\mathbf{v}^a(0) \in \mathcal{A}$, we can infer that $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$.

Finally, we prove $\rho(\mathbf{G}^*) < 1$. Due to $\mathcal{S}_1 \neq \emptyset$, then $p_{ii} + \lim_{t \rightarrow \infty} \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t) = p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*$ for all $\mathbf{v}^a(0) \in \mathcal{A}$. Thus, if $\mu_i(t) = \frac{h_i + \sum_{k \in \mathcal{N}(i)} \beta_{k \rightarrow i}^a(t)}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}$ converges to the same point, we can infer that $\lim_{t \rightarrow \infty} \sum_{k \in \mathcal{N}(i)} \beta_{k \rightarrow i}^a(t)$, or equivalently $\lim_{t \rightarrow \infty} \mathbf{E} \cdot \boldsymbol{\beta}^a(t)$ is unique for all $\mathbf{v}^a(0) \in \mathcal{A}$ and $\boldsymbol{\beta}^a(0) \in \mathbb{R}^{|\mathcal{E}|}$, where $\mathbf{E}$ is a $N \times |\mathcal{E}|$ matrix such that the i -th column of $\mathbf{E} \cdot \boldsymbol{\beta}^a(t)$ is equal to $\sum_{k \in \mathcal{N}(i)} \beta_{k \rightarrow i}^a(t)$.

Consider two initializations $(\mathbf{v}^a(0), \boldsymbol{\beta}_1^a(0))$ and $(\mathbf{v}^a(0), \boldsymbol{\beta}_2^a(0))$ with $\mathbf{v}^a(0) \in \mathcal{A}$ and $\boldsymbol{\beta}_1^a(0), \boldsymbol{\beta}_2^a(0) \in \mathbb{R}^{|\mathcal{E}|}$. Since $\mathbf{E} \cdot \boldsymbol{\beta}_1^a(t)$ and $\mathbf{E} \cdot \boldsymbol{\beta}_2^a(t)$ converges to the same point, we have $\mathbf{E}(\boldsymbol{\beta}_1^a(t) - \boldsymbol{\beta}_2^a(t))$ converges to $\mathbf{0}$, or equivalently

$$\lim_{t \rightarrow \infty} \mathbf{E} \cdot \Delta \boldsymbol{\beta}^a(t) = \mathbf{0} \quad (43)$$

for all $\mathbf{v}^a(0) \in \mathcal{A}$ and $\Delta \boldsymbol{\beta}^a(0) \in \mathbb{R}^{|\mathcal{E}|}$, where $\Delta \boldsymbol{\beta}^a(t) = \boldsymbol{\beta}_1^a(t) - \boldsymbol{\beta}_2^a(t)$. Substituting $\Delta \boldsymbol{\beta}^a(t) = \prod_{k=0}^{t-1} \mathbf{G}(k) \cdot \Delta \boldsymbol{\beta}^a(0)$ from (41) into (43) gives $\lim_{t \rightarrow \infty} \mathbf{E} \cdot \prod_{k=0}^{t-1} \mathbf{G}(k) \cdot \Delta \boldsymbol{\beta}^a(0) = \mathbf{0}$ for all $\mathbf{v}^a(0) \in \mathcal{A}$ and $\Delta \boldsymbol{\beta}^a(0) \in \mathbb{R}^{|\mathcal{E}|}$. Using the same arguments after (41), it can be inferred that

$$\lim_{t \rightarrow \infty} \mathbf{E} \cdot \mathbf{G}^{*t} = \mathbf{0}. \quad (44)$$

Now, we prove $\rho(\mathbf{G}^*) < 1$ by contradiction. Suppose $\rho(\mathbf{G}^*) \geq 1$. By multiplying $\text{eig}_{\max}(\mathbf{G}^*)$ on both sides of (44) and using the relation $\mathbf{G}^* \cdot \text{eig}_{\max}(\mathbf{G}^*) = \lambda_{\max}(\mathbf{G}^*) \cdot \text{eig}_{\max}(\mathbf{G}^*)$, we obtain $\lim_{t \rightarrow \infty} \lambda_{\max}^t(\mathbf{G}^*) \cdot \mathbf{E} \cdot \text{eig}_{\max}(\mathbf{G}^*) = \mathbf{0}$. Due to $\rho(\mathbf{G}^*) \geq 1$ by assumption, then $|\lambda_{\max}(\mathbf{G}^*)| \geq 1$, thus we must have

$$\mathbf{E} \cdot \text{eig}_{\max}(\mathbf{G}^*) = \mathbf{0}. \quad (45)$$

Writing (45) into a scalar form gives that $\sum_{k \in \mathcal{N}(i)} \mathbf{e}_{ki}^T \text{eig}_{\max}(\mathbf{G}^*) = 0$, or equivalently

$$\sum_{k \in \mathcal{N}(i) \setminus j} \mathbf{e}_{ki}^T \text{eig}_{\max}(\mathbf{G}^*) = -\mathbf{e}_{ji}^T \cdot \text{eig}_{\max}(\mathbf{G}^*), \quad (46)$$

where $\mathbf{e}_{ki}$ is a $|\mathcal{E}| \times 1$ vector with only one nonzero element 1 in a position such that $\mathbf{e}_{ki}^T \cdot \mathbf{v}^a(t) = v_{k \rightarrow i}^a(t)$. From the definitions of $\mathbf{A}$, $\mathbf{p}$ and $\mathbf{u}$ after (16), it is known that $\mathbf{e}_{ij}^T \mathbf{A} \text{eig}_{\max}(\mathbf{G}^*) = \sum_{k \in \mathcal{N}(i) \setminus j} \mathbf{e}_{ki}^T \text{eig}_{\max}(\mathbf{G}^*)$, $\mathbf{e}_{ij}^T \mathbf{p} = p_{ij}$ and $\mathbf{e}_{ij}^T (\mathbf{A} \mathbf{w}^* + \mathbf{u}) = p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*$. By applying these results to $\mathbf{G}^* = -\text{diag}^{-1}(\mathbf{A} \mathbf{w}^* + \mathbf{u}) \text{diag}(\mathbf{p}) \mathbf{A}$ in (29), we obtain $\mathbf{e}_{ij}^T \cdot \mathbf{G}^* \mathbf{y} = -\frac{p_{ij} \sum_{k \in \mathcal{N}(i) \setminus j} \mathbf{e}_{ki}^T \mathbf{y}}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*}$, where $\mathbf{y}$ is any vector with compatible dimension as $\mathbf{G}^*$. Then, substituting $\mathbf{y} = \text{eig}_{\max}(\mathbf{G}^*)$ into this equation and making use of (46) gives $\mathbf{e}_{ij}^T \cdot \mathbf{G}^* \text{eig}_{\max}(\mathbf{G}^*) = \frac{p_{ij} \mathbf{e}_{ji}^T \text{eig}_{\max}(\mathbf{G}^*)}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*}$. On the other hand, from the property of eigenvector, we also have $\mathbf{e}_{ij}^T \cdot \mathbf{G}^* \text{eig}_{\max}(\mathbf{G}^*) = \lambda_{\max}(\mathbf{G}^*) \cdot \mathbf{e}_{ij}^T \text{eig}_{\max}(\mathbf{G}^*)$. By equating the two equations, we obtain

$$\frac{p_{ij} \cdot \mathbf{e}_{ji}^T \text{eig}_{\max}(\mathbf{G}^*)}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*} = \lambda_{\max}(\mathbf{G}^*) \cdot \mathbf{e}_{ij}^T \text{eig}_{\max}(\mathbf{G}^*). \quad (47)$$

Reversing the positions of i and j , we also obtain

$$\frac{p_{ji} \cdot \mathbf{e}_{ij}^T \text{eig}_{\max}(\mathbf{G}^*)}{p_{jj} + \sum_{k \in \mathcal{N}(j) \setminus i} w_{kj}^*} = \lambda_{\max}(\mathbf{G}^*) \cdot \mathbf{e}_{ji}^T \text{eig}_{\max}(\mathbf{G}^*). \quad (48)$$

Multiplying (47) and (48) and eliminating the common terms gives

$$\begin{aligned} \lambda_{\max}^2(\mathbf{G}^*) &= \frac{p_{ij}^2}{(p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*)(p_{jj} + \sum_{k \in \mathcal{N}(j) \setminus i} w_{kj}^*)} \\ &= -\frac{w_{ji}^*}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*}, \end{aligned} \quad (49)$$

where the last equality is obtained directly from (12) with the fact that $g_{ij}(\mathbf{w}^*) = w_{ij}^*$.

It is known that if the variance $\sigma_i^2(t)$ converges, the converged variance $\lim_{t \rightarrow \infty} \sigma_i^2(t) = \frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*} > 0$ [29]. Thus, we have $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* > 0$, or equivalently $p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^* > -w_{ji}^*$. Together with (49), it can be obtained that $\lambda_{\max}^2(\mathbf{G}^*) = -\frac{w_{ji}^*}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*} < 1$, which contradicts with the assumption that $\rho(\mathbf{G}^*) \geq 1$. Thus, we have $\rho(\mathbf{G}^*) < 1$. ■

B. Damped Gaussian BP

In damped Gaussian BP, $\beta^a(t)$ is updated as [24]

$$\beta^a(t+1) = (1-d) \cdot \beta^a(t) + d \cdot (\mathbf{G}(t) \beta^a(t) + \mathbf{b}(t)), \quad (50)$$

where $d \neq 0$ is the damping factor. It is known that damped Gaussian BP shares the same fixed points as Gaussian BP [25], thus damped Gaussian BP can also be applied to compute the true marginal mean upon convergence [6]. Now, we present the convergence condition of beliefs in damped Gaussian BP.

Theorem 4. *In damped Gaussian BP, belief parameters $(\sigma_i^2(t), \mu_i(t))$ converge to the same point for all choices of $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$ under a nonzero damping factor d if and only if the three conditions hold: 1) $\mathcal{S}_1 \neq \emptyset$; 2) $\max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) < 1$ or $\min_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) > 1$; 3) $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$. Furthermore, the damping factor d should be chosen as*

$$d \in \begin{cases} \left(0, \min_{\lambda(\mathbf{G}^*)} 2 \cdot \frac{1 - \Re(\lambda(\mathbf{G}^*))}{|1 - \lambda(\mathbf{G}^*)|^2}\right), & \text{if } \max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) < 1 \\ \left(\max_{\lambda(\mathbf{G}^*)} 2 \cdot \frac{1 - \Re(\lambda(\mathbf{G}^*))}{|1 - \lambda(\mathbf{G}^*)|^2}, 0\right), & \text{if } \min_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) > 1. \end{cases} \quad (51)$$

Proof: Notice that (50) can be equivalently written as

$$\beta^a(t+1) = ((1-d) \cdot \mathbf{I} + d \cdot \mathbf{G}(t))\beta^a(t) + d \cdot \mathbf{b}(t), \quad (52)$$

which is in the same form as (14) in Gaussian BP. By applying Theorem 3, it can be inferred that belief parameters $(\sigma_i^2(t), \mu_i(t))$ of damped Gaussian BP converge for all choices of $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$ if and only if $\mathcal{S}_1 \neq \emptyset$, $\rho((1-d) \cdot \mathbf{I} + d \cdot \mathbf{G}^*) < 1$ and $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$.

Due to $|\lambda((1-d) \cdot \mathbf{I} + d \cdot \mathbf{G}^*)| = |1-d+d \cdot \lambda(\mathbf{G}^*)| = |(\lambda(\mathbf{G}^*)-1) \cdot d-1|$, we have $\rho^2((1-d) \cdot \mathbf{I} + d \cdot \mathbf{G}^*) = \max_{\lambda(\mathbf{G}^*)} |(1-\lambda(\mathbf{G}^*)) \cdot d-1|^2$. Thus, $\rho((1-d) \cdot \mathbf{I} + d \cdot \mathbf{G}^*) < 1$ is equivalent to

$$\max_{\lambda(\mathbf{G}^*)} \{|1-\lambda(\mathbf{G}^*)|^2 \cdot d^2 - 2(1-\Re(\lambda(\mathbf{G}^*))) \cdot d + 1\} < 1. \quad (53)$$

The maximum operation in (53) implies $|1-\lambda(\mathbf{G}^*)|^2 \cdot d^2 - 2(1-\Re(\lambda(\mathbf{G}^*))) \cdot d + 1 < 1$ for all $\lambda(\mathbf{G}^*)$, or equivalently $d(|1-\lambda(\mathbf{G}^*)|^2 \cdot d - 2(1-\Re(\lambda(\mathbf{G}^*)))) < 0$. The condition can only be satisfied under the two cases: 1) $d > 0$ and $|1-\lambda(\mathbf{G}^*)|^2 \cdot d - 2(1-\Re(\lambda(\mathbf{G}^*))) < 0$; and 2) $d < 0$ and $|1-\lambda(\mathbf{G}^*)|^2 \cdot d - 2(1-\Re(\lambda(\mathbf{G}^*))) > 0$. The conditions of the first case are equivalent to

$$0 < d < 2 \cdot \frac{1 - \Re(\lambda(\mathbf{G}^*))}{|1 - \lambda(\mathbf{G}^*)|^2}. \quad (54)$$

Obviously, the range of d in (54) is not empty if and only if $\Re(\lambda(\mathbf{G}^*)) < 1$. The conditions of the second case are equivalent to

$$2 \cdot \frac{1 - \Re(\lambda(\mathbf{G}^*))}{|1 - \lambda(\mathbf{G}^*)|^2} < d < 0. \quad (55)$$

Similarly, the range of d in (55) is not empty if and only if $\Re(\lambda(\mathbf{G}^*)) > 1$.

Since (54) and (55) should be satisfied for all $\lambda(\mathbf{G}^*)$, hence (53) is satisfied if and only if $\max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) < 1$ or $\min_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) > 1$. Furthermore, according to (54) and (55), it is obvious that the damping factor d satisfying (53) is given by (51). ■

Now, we reveal the relations between the convergence conditions of beliefs in Gaussian BP and damped Gaussian BP.

Proposition 3. *If the convergence condition of beliefs in Gaussian BP $\rho(\mathbf{G}^*) < 1$ holds, the convergence condition of beliefs in damped Gaussian BP $\max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) < 1$ holds as well, and $d = 1$ must be a damping factor that can guarantee convergence.*

Proof: Due to $\rho(\mathbf{G}^*) = \max_{\lambda(\mathbf{G}^*)} |\lambda(\mathbf{G}^*)|$, if the convergence condition of beliefs in Gaussian BP $\rho(\mathbf{G}^*) < 1$ holds, we have $\max_{\lambda(\mathbf{G}^*)} |\lambda(\mathbf{G}^*)| < 1$, and thereby

$$\max_{\lambda(\mathbf{G}^*)} |\lambda(\mathbf{G}^*)|^2 < 1. \quad (56)$$

Then, by noticing that $|\lambda(\mathbf{G}^*)|^2 = (\Re(\lambda(\mathbf{G}^*)))^2 + (\Im(\lambda(\mathbf{G}^*)))^2$, we can infer that $\max_{\lambda(\mathbf{G}^*)} (\Re(\lambda(\mathbf{G}^*)))^2 < 1$, and thereby $\max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*)) < 1$. That is, the convergence condition of beliefs in damped Gaussian BP holds.

Furthermore, from (56), we know that $|\lambda(\mathbf{G}^*)|^2 < 1$ holds for all $\lambda(\mathbf{G}^*)$. Adding $1 - 2\Re(\lambda(\mathbf{G}^*))$ on both sides of this equation gives $1 - 2\Re(\lambda(\mathbf{G}^*)) + |\lambda(\mathbf{G}^*)|^2 < 2 - 2\Re(\lambda(\mathbf{G}^*))$, or equivalently $\frac{2(1-\Re(\lambda(\mathbf{G}^*)))}{|1-\lambda(\mathbf{G}^*)|^2} > 1$. Since this holds for all $\lambda(\mathbf{G}^*)$, we can infer that $\min_{\lambda(\mathbf{G}^*)} 2 \cdot \frac{1-\Re(\lambda(\mathbf{G}^*))}{|1-\lambda(\mathbf{G}^*)|^2} > 1$. Comparing this result with the first case of (51), it is obvious that $d = 1$ is in the feasible range. ■

Proposition 3 reveals that the convergence condition of beliefs in Gaussian BP is always implied by that in damped Gaussian BP. Also notice that typically in damped BP [24]–[28], the range of d is confined in $(0, 1]$. Theorem 4 not only gives possibly wider range of d with convergence guarantee, but also avoids the trial-and-error for finding the d .

C. Asynchronous Gaussian BP

Modified from the synchronous updating of $\mathbf{v}^a(t)$ in (11) and $\beta^a(t)$ in (14), under asynchronous scheduling, $v_{i \rightarrow j}^a(t)$ and $\beta_{i \rightarrow j}^a(t)$ for $t \in \mathcal{T}_{i \rightarrow j}$ are updated as

$$v_{i \rightarrow j}^a(t+1) = -\frac{p_{ij}^2}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(\tau_{k \rightarrow i}^a(t))}, \quad (57)$$

$$\beta_{i \rightarrow j}^a(t+1) = -\frac{p_{ij} \left(h_i + \sum_{k \in \mathcal{N}(i) \setminus j} \beta_{k \rightarrow i}^a(\tau_{k \rightarrow i}^a(t)) \right)}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k \rightarrow i}^a(\tau_{k \rightarrow i}^a(t))}, \quad (58)$$

where $\mathcal{T}_{i \rightarrow j}$ is the set of time instants at which $v_{i \rightarrow j}^a(t)$ and $\beta_{i \rightarrow j}^a(t)$ are updated and $\tau_{k \rightarrow i}^a(t)$ satisfies the totally asynchronous scheduling defined below.

Definition 1. (*Totally Asynchronous Scheduling*) [30]: The sets $\mathcal{T}_{i \rightarrow j}$ are infinite, and $\tau_{i \rightarrow j}^a(t)$ satisfies $0 \leq \tau_{i \rightarrow j}^a(t) \leq t$ and $\lim_{t \rightarrow \infty} \tau_{i \rightarrow j}^a(t) = \infty$ for all $(i, j) \in \mathcal{E}$.

From the above definition, the updating of each message is executed independently at time instants $t \in \mathcal{T}_{i \rightarrow j}$ with $(i, j) \in \mathcal{E}$. Moreover, at each updating time instant, only the previously available information $v_{k \rightarrow i}^a(\tau_{k \rightarrow i}^a(t))$ and $\beta_{k \rightarrow i}^a(\tau_{k \rightarrow i}^a(t))$ at each node is needed. Now, the following theorem can be presented.

Theorem 5. In asynchronous Gaussian BP, if $\mathcal{S}_1 \neq \emptyset$, $\rho(|\mathbf{G}^*|) < 1$ and $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$, belief parameters $(\sigma_i^2(t), \mu_i(t))$ converge to the same point for all choices of $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$.

Proof: See Appendix A. ■

Due to $\rho(\mathbf{G}^*) \leq \rho(|\mathbf{G}^*|)$ [30], it can be seen that the convergence condition of asynchronous Gaussian BP in Theorem 5 is more stringent than that of synchronous Gaussian BP in Theorem 3.

V. RELATIONSHIPS WITH EXISTING CONVERGENCE CONDITIONS

In many existing convergence conditions, the original PDF $f(\mathbf{x})$ is first transformed into its normalized form

$$\tilde{f}(\tilde{\mathbf{x}}) \propto \exp \left\{ -\frac{1}{2} \tilde{\mathbf{x}}^T \tilde{\mathbf{P}} \tilde{\mathbf{x}} + \tilde{\mathbf{h}}^T \tilde{\mathbf{x}} \right\}, \quad (59)$$

where $\tilde{\mathbf{x}} = \text{diag}^{\frac{1}{2}}(\mathbf{P})\mathbf{x}$, $\tilde{\mathbf{P}} = \text{diag}^{-\frac{1}{2}}(\mathbf{P})\mathbf{P}\text{diag}^{-\frac{1}{2}}(\mathbf{P})$ and $\tilde{\mathbf{h}} = \text{diag}^{-\frac{1}{2}}(\mathbf{P})\mathbf{h}$. Then, Gaussian BP is carried out with $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{h}}$ using the updating equations in (4), (5), (7) and (8). Under the normalized $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{h}}$, we denote the corresponding message parameters as $\tilde{\mathbf{v}}^a(t)$ and $\tilde{\beta}^a(t)$. The corresponding belief variance and belief mean are calculated as $\tilde{\sigma}_i^2(t) = \frac{1}{1 + \sum_{k \in \mathcal{N}(i)} \tilde{v}_{k \rightarrow i}^a(t)}$ and $\tilde{\mu}_i(t) = \frac{\tilde{h}_i + \sum_{k \in \mathcal{N}(i)} \tilde{\beta}_{k \rightarrow i}^a(t)}{1 + \sum_{k \in \mathcal{N}(i)} \tilde{v}_{k \rightarrow i}^a(t)}$, respectively. Then, the belief variance and belief mean of the original PDF $f(\mathbf{x})$ can be recovered from $\sigma_i^2(t) = \frac{1}{p_{ii}} \tilde{\sigma}_i^2(t)$ and $\mu_i(t) = \frac{1}{\sqrt{p_{ii}}} \tilde{\mu}_i(t)$. With the normalized $\tilde{\mathbf{P}}$, the matrix $\tilde{\mathbf{G}}^*$ as well as sets $\tilde{\mathcal{S}}_1$ and $\tilde{\mathcal{A}}$ can be defined accordingly. The following proposition reveals that the convergence conditions proposed in this paper are invariant to normalization.

Proposition 4. *The following statements hold:*

- 1) $\tilde{\mathcal{S}}_1 \neq \emptyset$ if and only if $\mathcal{S}_1 \neq \emptyset$;
- 2) If $\mathcal{S}_1 \neq \emptyset$, then $\rho(\tilde{\mathbf{G}}^*) = \rho(\mathbf{G}^*)$ and $\rho(|\tilde{\mathbf{G}}^*|) = \rho(|\mathbf{G}^*|)$;
- 3) $1 + \sum_{k \in \mathcal{N}(i)} \tilde{w}_{ki}^* \neq 0$ if and only if $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$.

Proof: See Appendix B. ■

It is proved in [24] that if walk-summability $\rho(|\tilde{\mathbf{P}} - \mathbf{I}|) < 1$ (or equivalently $\rho(|\tilde{\mathbf{P}}| - \mathbf{I}) < 1$ since all diagonal elements in $\tilde{\mathbf{P}}$ are equal to one) is satisfied, the belief variance and belief mean can be computed with Gaussian BP under synchronous and asynchronous scheduling with message initialization $\tilde{\mathbf{v}}^a(0) = \mathbf{0}$ and $\tilde{\beta}^a(0) = \mathbf{0}$. Now, we show the relation between $\rho(|\tilde{\mathbf{P}}| - \mathbf{I})$ and the proposed conditions.

Theorem 6. *If $\rho(|\tilde{\mathbf{P}}| - \mathbf{I}) < 1$, then $\tilde{\mathcal{S}}_1 \neq \emptyset$, $\rho(|\tilde{\mathbf{G}}^*|) < 1$ and $1 + \sum_{k \in \mathcal{N}(i)} \tilde{w}_{ki}^* \neq 0$.*

Proof: As proved in [24, Proposition 13], walk-summability $\rho(|\tilde{\mathbf{P}}| - \mathbf{I}) < 1$ is equivalent to the pairwise normalizable condition, which is called the convex decomposition condition in [23]. Notice that $\tilde{\mathbf{x}}^T \tilde{\mathbf{P}} \tilde{\mathbf{x}}$ in the exponent of $\tilde{f}(\tilde{\mathbf{x}})$ in (59) can be expanded as $\sum_{i=1}^N (1 - \sum_{k \in \mathcal{N}(i)} \tilde{p}_{ki}^2 \theta_{ki}) \tilde{x}_i^2 + \frac{1}{2} \sum_{(i,j) \in \mathcal{E}} (\theta_{ji} \tilde{p}_{ij}^2 \tilde{x}_i^2 + 2\tilde{p}_{ij} \tilde{x}_i \tilde{x}_j + \theta_{ij} \tilde{p}_{ij}^2 \tilde{x}_j^2)$ for any $\theta_{ij} \in \mathbb{R}$. Convex decomposition requires the existence of θ_{ij} for all $(i, j) \in \mathcal{E}$ satisfying the two conditions:

$$\begin{bmatrix} \theta_{ji} \tilde{p}_{ij}^2 & \tilde{p}_{ij} \\ \tilde{p}_{ij} & \theta_{ij} \tilde{p}_{ij}^2 \end{bmatrix} \succeq 0, \quad \forall (i, j) \in \mathcal{E}, \quad (60)$$

$$1 - \sum_{k \in \mathcal{N}(i)} \tilde{p}_{ki}^2 \theta_{ki} > 0, \quad \forall i \in \mathcal{V}. \quad (61)$$

Now, we will prove that if convex decomposition conditions in (60) and (61) are satisfied, then $\tilde{\mathcal{S}}_1 \neq \emptyset$. From (60) and property of a positive semi-definite matrix, we can infer that $\theta_{ji} \tilde{p}_{ij}^2 \geq 0$ and $\theta_{ji} \theta_{ij} \tilde{p}_{ij}^4 - \tilde{p}_{ij}^2 \geq 0$, or equivalently

$$\theta_{ji} \geq 0, \quad (62)$$

$$\theta_{ji} \theta_{ij} \tilde{p}_{ij}^2 \geq 1. \quad (63)$$

If θ_{ij} or θ_{ji} equals 0, obviously, (63) cannot hold due to $\tilde{p}_{ij}^2 > 0$. Thus, we have $\theta_{ij} \neq 0$ and $\theta_{ji} \neq 0$. Putting this result into (62) and (63), we obtain $\theta_{ji} > 0$ and $\tilde{p}_{ij}^2 \theta_{ji} \geq \frac{1}{\theta_{ij}}$, respectively.

Substituting $\theta_{ji} > 0$ into (61)'s equivalent form $1 - \sum_{k \in \mathcal{N}(i) \setminus j} \tilde{p}_{ki}^2 \theta_{ki} > \tilde{p}_{ij}^2 \theta_{ji}$ gives

$$1 - \sum_{k \in \mathcal{N}(i) \setminus j} \tilde{p}_{ki}^2 \theta_{ki} > 0. \quad (64)$$

Further, applying $\tilde{p}_{ij}^2 \theta_{ji} \geq \frac{1}{\theta_{ij}}$ to $1 - \sum_{k \in \mathcal{N}(i) \setminus j} \tilde{p}_{ki}^2 \theta_{ki} > \tilde{p}_{ij}^2 \theta_{ji}$, we also obtain

$$-\tilde{p}_{ij}^2 \theta_{ij} < -\frac{\tilde{p}_{ij}^2}{1 - \sum_{k \in \mathcal{N}(i) \setminus j} \tilde{p}_{ki}^2 \theta_{ki}}. \quad (65)$$

By defining $c_{ij} = -\tilde{p}_{ij}^2 \theta_{ij}$, then (64) and (65) can be written as $1 + \sum_{k \in \mathcal{N}(i) \setminus j} c_{ki} > 0$ and $c_{ij} < -\frac{\tilde{p}_{ij}^2}{1 + \sum_{k \in \mathcal{N}(i) \setminus j} c_{ki}}$ for all $(i, j) \in \mathcal{E}$, which are exactly the conditions of $\tilde{\mathcal{S}}_1$ defined in (17). Therefore, we have $\mathbf{c} \in \tilde{\mathcal{S}}_1$ and $\tilde{\mathcal{S}}_1 \neq \emptyset$, where $\mathbf{c}$ contains c_{ij} with $(i, j) \in \mathcal{E}$ arranged in ascending order first on j and then on i .

Furthermore, it is proved in [23, Lemma 3] that if $\rho(|\tilde{\mathbf{P}}| - \mathbf{I}) < 1$, we must have $\rho(|\tilde{\mathbf{G}}^*|) < 1$. At last, since $\rho(|\tilde{\mathbf{P}}| - \mathbf{I}) < 1$ guarantees belief variance converge to $\lim_{t \rightarrow \infty} \sigma_i^2(t) = \frac{1}{1 + \sum_{k \in \mathcal{N}(i)} \tilde{w}_{ki}^*}$ under $\tilde{\mathbf{v}}^a(0) = \mathbf{0}$ [24], we must have $1 + \sum_{k \in \mathcal{N}(i)} \tilde{w}_{ki}^* \neq 0$. ■

A relaxed initialization set is also proposed in [23] under the convergence condition of convex decomposition, which is proved to be equivalent to walk-summability [24]. The initialization set in [23] is defined as $\mathcal{D} = \{\tilde{\mathbf{v}}^a(0) | -\text{diag}^{-2}(\tilde{\mathbf{p}})\tilde{\mathbf{v}}^a(0) \leq \boldsymbol{\theta}; \forall \boldsymbol{\theta} \text{ s.t. (60), (61)}\}$, where $\boldsymbol{\theta}$ contains elements θ_{ij} with $(i, j) \in \mathcal{E}$ arranged ascending order first on j and then on i . Now, we propose the following proposition.

Theorem 7. $\mathcal{D}$ is a subset of $\tilde{\mathcal{A}}$.

Proof: Notice that $\mathcal{D}$ can be written as $\mathcal{D} = \{\tilde{\mathbf{v}}^a(0) | \tilde{\mathbf{v}}^a(0) \geq -\text{diag}^2(\tilde{\mathbf{p}})\boldsymbol{\theta}; \forall \boldsymbol{\theta} \text{ s.t. (60), (61)}\}$. For any $\boldsymbol{\theta}$ satisfies the conditions of (60) and (61), we have (64) and (65) hold, which is equivalent to $\mathbf{c} = -\text{diag}^2(\tilde{\mathbf{p}})\boldsymbol{\theta} \in \text{int}(\tilde{\mathcal{S}}_1)$, where the boundary of $\tilde{\mathcal{S}}_1$ is excluded since equality sign is included in definition of $\tilde{\mathcal{S}}_1$ but not in (64) and (65). Thus, it can be seen that $\mathcal{D} \subseteq \{\tilde{\mathbf{v}}^a(0) \geq \mathbf{c} | \mathbf{c} \in \text{int}(\tilde{\mathcal{S}}_1)\}$. Then, from the definition of $\tilde{\mathcal{A}} = \{\mathbf{w} \geq \mathbf{0}\} \cup \left\{ \mathbf{w} \geq \mathbf{w}_0 \mid \mathbf{w}_0 \in \text{int}(\tilde{\mathcal{S}}_1) \right\} \cup \left\{ \mathbf{w} \geq \mathbf{w}_0 \mid \mathbf{w}_0 \in \tilde{\mathcal{S}}_1 \text{ and } \mathbf{w}_0 = \lim_{t \rightarrow \infty} \mathbf{g}^{(t)}(\mathbf{0}) \right\}$ in (18), it is obvious that $\mathcal{D}$ is a subset of $\tilde{\mathcal{A}}$. ■

VI. NUMERICAL EXAMPLES

In this section, numerical results are presented to illustrate the theories in this paper. The example is based on the linear coefficient $\mathbf{h} = [1, 1, \dots, 1]^T$ with length $N = 20$, and the

precision matrices $\mathbf{P}$ constructed as

$$p_{ij} = \begin{cases} 1, & \text{if } i = j, \\ \zeta \cdot a_{\text{mod}(i+j,10)+1}, & \text{if } i \neq j, \end{cases} \quad (66)$$

where ζ is a coefficient indicating the correlation strength among variables; and a_k is the k th element of the vector $\mathbf{a} = [0.13, 0.10, 0.71, -0.05, 0, 0.12, 0.07, 0.11, -0.02, -0.03]^T$. The varying of correlation strength ζ induces a series of matrices, and the positive definite constraint $\mathbf{P} \succ \mathbf{0}$ can be guaranteed when $\zeta < 0.5978$. Furthermore, numerical calculation shows that when $\zeta \leq 0.5859$, we have $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$ by solving the SDP problem in (19). Thus, when $\zeta \leq 0.5859$, to determine the convergence of beliefs, we only need to check whether $\rho(\mathbf{G}^*)$, $\max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*))$ and $\rho(|\mathbf{G}^*|)$ are satisfied for synchronous Gaussian BP, damped Gaussian BP and asynchronous Gaussian BP, respectively.

Fig. 2 illustrates how the spectral radii $\rho(|\tilde{\mathbf{P}}| - \mathbf{I})$, $\rho(|\mathbf{G}^*|)$, $\rho(\mathbf{G}^*)$ and the maximum real part of eigenvalues $\max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*))$ vary as a function of ζ . The curves are plotted for $\zeta \leq 0.5859$ which guarantees $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$. In Fig. 2, three critical values can be observed: $\zeta_a = 0.3945$, $\zeta_b = 0.4621$ and $\zeta_c = 0.5859$, which are labelled as A, B and C, respectively. It can be seen that $\zeta < \zeta_a$, $\zeta < \zeta_b$ and $\zeta < \zeta_c$ are the necessary and sufficient conditions for $\rho(|\mathbf{G}^*|) < 1$, $\rho(\mathbf{G}^*) < 1$ and $\mathcal{S}_1 \neq \emptyset$, respectively. Then, according to Theorem 5, if $\zeta < \zeta_a$, beliefs converge for all initialization $\mathbf{v}^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$ under any totally asynchronous scheduling. Furthermore, from Theorems 3 and 4, it can be obtained that $\zeta < \zeta_b$ and $\zeta < \zeta_c$ are the necessary and sufficient convergence conditions of beliefs in synchronous and damped Gaussian BP, respectively. Lastly, from Fig. 2, it is obvious that if $\rho(|\tilde{\mathbf{P}}| - \mathbf{I}) < 1$, then $\rho(|\mathbf{G}^*|) < 1$. This numerically corroborates the claim in Theorem 6 that walk-summability is implied by the convergence conditions proposed in this paper. Next, we inspect the convergence behaviors around the three special points A, B and C under two different initializations: $(\mathbf{v}^a(0) = \mathbf{0}, \beta^a(0) = \mathbf{0})$ and $(\mathbf{v}^a(0) = \mathbf{w}^*, \beta^a(0) = \mathbf{1} \times 0.01)$, where $\mathbf{w}^*$ is the optimal solution of (19) with $\zeta = 0.5858$.

First, consider $\zeta = 0.4600$ and 0.4700 , which are neighbors on opposite sides of the critical value $\zeta_b = 0.4621$. Fig. 3 illustrates how belief mean $\mu_1(t)$ computed with synchronous Gaussian BP evolves. When $\zeta = 0.4600 < \zeta_b$, it is observed from Fig. 3a that $\mu_1(t)$ converges to the same value under the two different initializations. On the other hand, when $\zeta = 0.4700 > \zeta_b$,

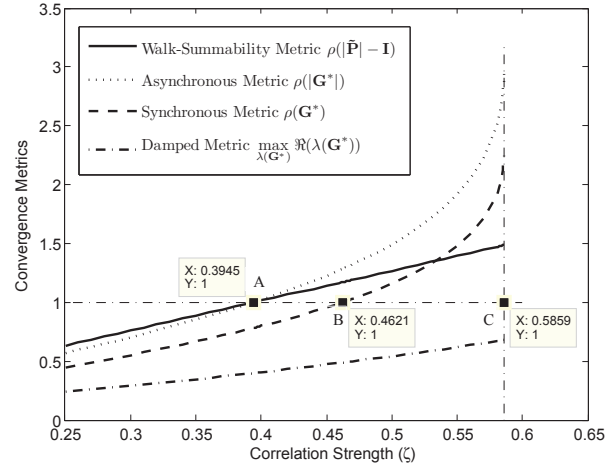


Fig. 2: Illustration of $\rho(|\tilde{\mathbf{P}}| - \mathbf{I})$, $\rho(|\mathbf{G}^*|)$, $\rho(\mathbf{G}^*)$ and $\max_{\lambda(\mathbf{G}^*)} \Re(\lambda(\mathbf{G}^*))$ as a function of ζ .

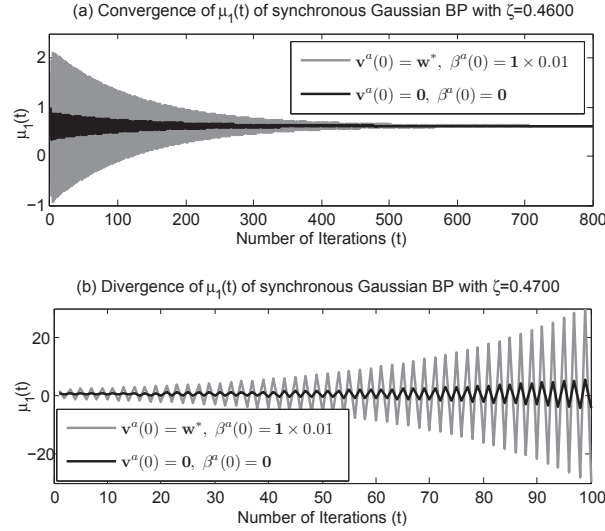


Fig. 3: Illustration of convergence and divergence of belief mean $\mu_1(t)$ in synchronous Gaussian BP under different initializations with ζ chosen around the critical value $\zeta_b = 0.4621$.

Fig. 3b demonstrates the divergence of $\mu_1(t)$. Thus, Fig. 3 verifies the necessary and sufficient convergence property of synchronous Gaussian BP in Theorem 3.

Second, consider $\zeta = 0.5858$ and 0.5860 , which are chosen to be slightly smaller and larger than $\zeta_c = 0.5859$, respectively. When $\zeta = 0.5858$ is slightly smaller than $\zeta_c = 0.5859$, Fig. 4a illustrates the convergence of $\mu_1(t)$ in damped Gaussian BP with damping factor $d = 0.61$,

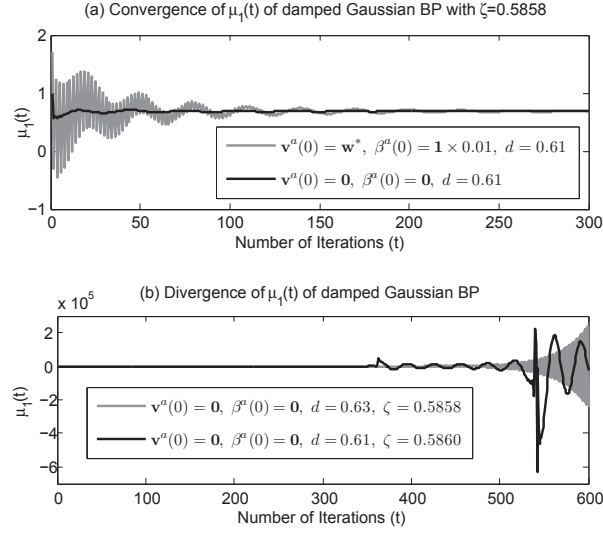


Fig. 4: Illustration of convergence and divergence of belief mean $\mu_1(t)$ in damped Gaussian BP under different initializations with ζ chosen around the critical value $\zeta_c = 0.5859$.

which locates within the required range $d \in (0, 0.6201)$ computed from (51). On the other hand, when the damping factor $d = 0.63$, which is outside of the range $(0, 0.6201)$, Fig. 4b shows that belief mean $\mu_1(t)$ does not converge. Moreover, when $\zeta = 0.5860$ which is slightly larger than $\zeta_c = 0.5859$, Fig. 4b illustrates that belief mean $\mu_1(t)$ cannot converge neither, even damping is applied. Obviously, Fig. 4 corroborates the necessary and sufficient convergence condition of damped Gaussian BP as well as the proposed range of damping factor in Theorem 4.

Finally, consider $\zeta = 0.3900$, which is slightly smaller than the critical value $\zeta_a = 0.3945$. To demonstrate the effects of asynchronous update, two schedulings with 2% and 10% chances of missing the exchanged messages are considered. When missing rate is 2%, Fig. 5a shows that the belief mean $\mu_1(t)$ converges to the same value under the two different initializations. Moreover, the convergence of $\mu_1(t)$ under the asynchronous scheduling with message missing rate equal to 10% is also observed in Fig. 5b. Obviously, Fig. 5 verifies the results in Theorem 5.

VII. CONCLUSIONS

In this paper, by describing the message-passing process as a set of updating equations, necessary and sufficient convergence conditions of beliefs in synchronous Gaussian BP and

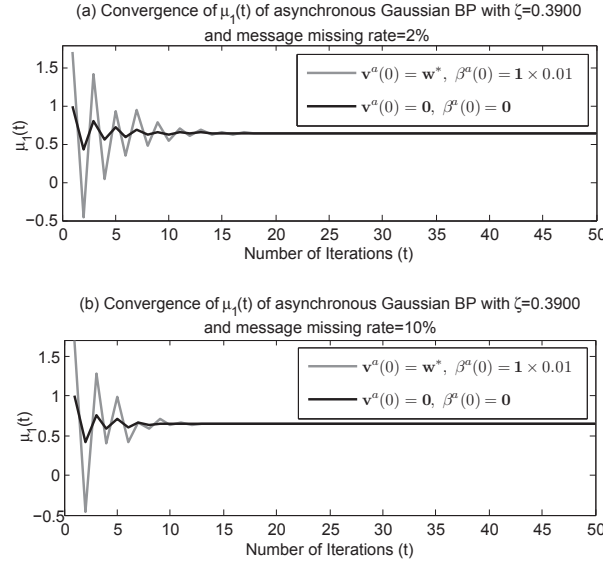


Fig. 5: Illustration of convergence of belief mean $\mu_1(t)$ under different initializations and message missing rates with ζ chosen smaller than the critical value $\zeta_a = 0.3945$.

damped Gaussian BP are derived under a newly proposed initialization set. The proposed initialization set is proved to be the largest among all currently known initialization sets. The results theoretically confirmed the extensively reported conjecture that damping is helpful to improve the convergence of Gaussian BP. Furthermore, under totally asynchronous scheduling, a sufficient convergence condition of beliefs is also derived for the same proposed initialization set. Relationships between the proposed convergence conditions and existing ones were established analytically, demonstrating that the existing convergence conditions are implied by the proposed ones. Numerical examples are presented to corroborate the established theories.

APPENDIX A

PROOF OF THEOREM 5

Under totally asynchronous scheduling, due to $\mathcal{S}_1 \neq \emptyset$, it is known from Theorem 1 that $\mathbf{v}^a(t)$ converges to the same point for all $\mathbf{v}^a(0) \in \mathcal{A}$ with $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) = \mathbf{w}^*$. Thus, if $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$, we can infer that belief variance $\sigma_i^2 = \frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}$ converges to $\frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*}$ for all $\mathbf{v}^a(0) \in \mathcal{A}$. Since $\lim_{t \rightarrow \infty} \frac{1}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}$ is unique, to prove that the belief mean $\mu_i(t) = \frac{h_i + \sum_{k \in \mathcal{N}(i)} \beta_{k \rightarrow i}^a(t)}{p_{ii} + \sum_{k \in \mathcal{N}(i)} v_{k \rightarrow i}^a(t)}$ converges to the same point, we only need to prove $\beta^a(t)$

converges to a unique value under totally asynchronous scheduling for all $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$.

First, define a sequence of sets $\mathcal{V}(m) = \{\mathbf{w} \mid \bar{\mathbf{v}}^a(m) \leq \mathbf{w} \leq \hat{\mathbf{v}}^a(m)\}$ with $\bar{\mathbf{v}}^a(0) \in \text{int}(\mathcal{S}_1) \cup \{\mathbf{w}^*\}$ and $\hat{\mathbf{v}}(0) \geq \mathbf{0}$. Due to $\bar{\mathbf{v}}^a(0) \in \text{int}(\mathcal{S}_1) \cup \{\mathbf{w}^*\}$, from the definition of $\mathcal{S}_1$ in (17), we have $\bar{\mathbf{v}}^a(0) \leq \mathbf{g}(\bar{\mathbf{v}}^a(0))$, or equivalently $\bar{\mathbf{v}}^a(0) \leq \bar{\mathbf{v}}^a(1)$. From the monotonically increasing property of $g_{ij}(\cdot)$ in (24), it can be inferred that $\mathbf{g}(\bar{\mathbf{v}}^a(0)) \leq \mathbf{g}(\bar{\mathbf{v}}^a(1))$, or equivalently $\bar{\mathbf{v}}^a(1) \leq \bar{\mathbf{v}}^a(2)$. By induction, we obtain $\bar{\mathbf{v}}^a(m) \leq \bar{\mathbf{v}}^a(m+1)$ for $m \geq 0$. On the other hand, due to $\hat{\mathbf{v}}(0) \geq \mathbf{0}$, substituting it into (12) gives $\hat{\mathbf{v}}(1) < \mathbf{0}$, and thereby $\hat{\mathbf{v}}(1) < \hat{\mathbf{v}}(0)$. From the monotonically increasing property of $g_{ij}(\cdot)$ in (24), it can be inferred that $\mathbf{g}(\hat{\mathbf{v}}(1)) \leq \mathbf{g}(\hat{\mathbf{v}}(0))$, or equivalently $\hat{\mathbf{v}}^a(2) \leq \hat{\mathbf{v}}^a(1)$. By induction, we have $\hat{\mathbf{v}}^a(m+1) \leq \hat{\mathbf{v}}^a(m)$ for all $m \geq 0$. Combining with $\bar{\mathbf{v}}^a(m) \leq \bar{\mathbf{v}}^a(m+1)$, we can infer that $\mathcal{V}(m+1) \subseteq \mathcal{V}(m)$. It is known from Theorem 1 that $\bar{\mathbf{v}}^a(m)$ and $\hat{\mathbf{v}}^a(m)$ both converge to $\mathbf{w}^*$, thus we have the set $\mathcal{V}(m)$ also converges to a single point $\mathbf{w}^*$.

Next, from (58), the updating function of $\beta^a(t)$ can be written as

$$\varphi(\gamma, \beta^a) = \mathbf{G}(\gamma)\beta^a + \mathbf{b}(\gamma), \quad (67)$$

where $\mathbf{G}(\gamma) \triangleq -\text{diag}^{-1}(\mathbf{A}\gamma + \mathbf{u})\text{diag}(\mathbf{p})\mathbf{A}$ and $\mathbf{b}(\gamma) \triangleq -\text{diag}^{-1}(\mathbf{A}\gamma + \mathbf{u})\text{diag}(\mathbf{p})\boldsymbol{\xi}$. Then, define the following sets

$$\mathcal{B}(m) = \{\beta^a \mid \|\beta^a - \phi\|_\infty^\omega \leq \eta_m\}, \quad (68)$$

where $\|\cdot\|_\infty^\omega$ is the weighted maximum norm with weighting ω ; ϕ is defined in (31);

$$\eta_m \triangleq \max_{\gamma \in \mathcal{V}(m-1)} \left\{ \|\mathbf{G}(\gamma)\|_\infty^\omega \eta_{m-1} + \mathcal{L}(\gamma) \right\} \quad (69)$$

with $\mathcal{L}(\gamma) \triangleq \|(\mathbf{G}(\gamma) - \mathbf{G}^*)\phi + \mathbf{b}(\gamma) - \mathbf{b}^*\|_\infty^\omega$; and $\|\cdot\|_\infty^\omega$ is the induced matrix norm of $\|\cdot\|_\infty^\omega$. To define the set $\mathcal{B}(m)$, the initial η_0 could be any positive value such that the chosen initialization $\beta^a(0)$ lies within the set $\mathcal{B}(0)$.

Due to the maximum weighted norm $\|\cdot\|_\infty^\omega$ in (68), set $\mathcal{B}(m)$ can always be represented as the product of subsets of individual components, and thereby the box condition [30, p. 431] is satisfied. Thus, to prove the convergence of $\beta^a(t)$, we only need to prove that $\mathcal{B}(m)$ satisfies the following three conditions: 1) $\mathcal{B}(m+1) \subseteq \mathcal{B}(m)$ for all m larger than some positive integer; 2) $\varphi(\gamma, \beta^a) \in \mathcal{B}(m+1)$ for any $\gamma \times \beta^a \in \mathcal{V}(m) \times \mathcal{B}(m)$; 3) $\mathcal{B}(m)$ converges to ϕ [30, p. 431].

1) Due to $\rho(|\mathbf{G}^*|) < 1$, there exists some weighted maximum matrix norm $\|\cdot\|_\infty^\omega$ such that

$\|\mathbf{G}^*\|_\infty^\omega < 1$ [30]. Since $\mathcal{V}(m)$ converges to $\mathbf{w}^*$, there exists a m_0 such that $\|\mathbf{G}(\gamma)\|_\infty^\omega < 1$ for

all $\gamma \in \mathcal{V}(m_0)$. On the other hand, for any $\gamma \in \mathcal{V}(m_0)$, due to $\mathcal{V}(m_0) \subseteq \mathcal{V}(0)$ and $\mathcal{V}(0) \subseteq \mathcal{A}$, it is known that $\gamma \in \mathcal{A}$. From (37), we have $\mathbf{A}\gamma + \mathbf{u} > \mathbf{0}$, and $\text{diag}^{-1}(\mathbf{A}\gamma + \mathbf{u})$ exists. Then, elements in $\mathbf{G}(\gamma)$ and $\mathbf{b}(\gamma)$ are finite, and thus $\mathcal{L}(\gamma) = \|(\mathbf{G}(\gamma) - \mathbf{G}^*)\phi + \mathbf{b}(\gamma) - \mathbf{b}^*\|_\infty^\omega$ must be finite for all $\gamma \in \mathcal{V}(m_0)$. Combining with $\|\mathbf{G}(\gamma)\|_\infty^\omega < 1$ for all $\gamma \in \mathcal{V}(m_0)$, we can infer that $\frac{\max_{\gamma \in \mathcal{V}(m_0)} \mathcal{L}(\gamma)}{1 - \max_{\gamma \in \mathcal{V}(m_0)} \|\mathbf{G}(\gamma)\|_\infty^\omega}$ is a positive finite value. Thus, by choosing η_0 large enough, we can always ensure $\beta^a(0) \in \mathcal{B}(0)$ and $\eta_{m_0} \geq \frac{\max_{\gamma \in \mathcal{V}(m_0)} \mathcal{L}(\gamma)}{1 - \max_{\gamma \in \mathcal{V}(m_0)} \|\mathbf{G}(\gamma)\|_\infty^\omega}$, or equivalently $\max_{\gamma \in \mathcal{V}(m_0)} \|\mathbf{G}(\gamma)\|_\infty^\omega \eta_{m_0} + \max_{\gamma \in \mathcal{V}(m_0)} \mathcal{L}(\gamma) \leq \eta_{m_0}$. Combining with the fact $\max_{\gamma \in \mathcal{V}(m_0)} \{\|\mathbf{G}(\gamma)\|_\infty^\omega \eta_{m_0} + \mathcal{L}(\gamma)\} \leq \max_{\gamma \in \mathcal{V}(m_0)} \|\mathbf{G}(\gamma)\|_\infty^\omega \eta_{m_0} + \max_{\gamma \in \mathcal{V}(m_0)} \mathcal{L}(\gamma)$, we can infer from the definition of η_{m_0+1} in (69) that $\eta_{m_0+1} \leq \eta_{m_0}$. Now, suppose $\eta_{m+1} \leq \eta_m$ for some $m \geq m_0$. Applying this to $\eta_{m+2} = \max_{\gamma \in \mathcal{V}(m+1)} \{\|\mathbf{G}(\gamma)\|_\infty^\omega \eta_{m+1} + \mathcal{L}(\gamma)\}$ from (69) gives $\eta_{m+2} \leq \max_{\gamma \in \mathcal{V}(m+1)} \{\|\mathbf{G}(\gamma)\|_\infty^\omega \eta_m + \mathcal{L}(\gamma)\}$. Due to $\mathcal{V}(m+1) \subseteq \mathcal{V}(m)$, it can be further inferred that $\eta_{m+2} \leq \max_{\gamma \in \mathcal{V}(m)} \{\|\mathbf{G}(\gamma)\|_\infty^\omega \eta_m + \mathcal{L}(\gamma)\}$. According to the definition of η_{m+1} in (69), we obtain $\eta_{m+2} \leq \eta_{m+1}$. Hence, $\eta_{m+1} \leq \eta_m$ for all $m \geq m_0$. Thus, the first condition is proved.

- 2) From (31), it is known that $\phi = \mathbf{G}^*\phi + \mathbf{b}^*$. Subtracting it from (67) and rearranging the terms gives $\varphi(\gamma, \beta^a) - \phi = \mathbf{G}(\gamma)(\beta^a - \phi) + (\mathbf{G}(\gamma) - \mathbf{G}^*)\phi + \mathbf{b}(\gamma) - \mathbf{b}^*$. Taking the norm on both sides of this equality leads to

$$\|\varphi(\gamma, \beta^a) - \phi\|_\infty^\omega \leq \|\mathbf{G}(\gamma)\|_\infty^\omega \|\beta^a - \phi\|_\infty^\omega + \mathcal{L}(\gamma). \quad (70)$$

For any $\gamma \times \beta^a \in \mathcal{V}(m) \times \mathcal{B}(m)$, it is known from (68) that $\|\beta^a - \phi\|_\infty^\omega \leq \eta_m$. Applying it to (70) gives $\|\varphi(\gamma, \beta^a) - \phi\|_\infty^\omega \leq \|\mathbf{G}(\gamma)\|_\infty^\omega \eta_m + \mathcal{L}(\gamma)$, and hence $\|\varphi(\gamma, \beta^a) - \phi\|_\infty^\omega \leq \max_{\gamma \in \mathcal{V}(m)} \{\|\mathbf{G}(\gamma)\|_\infty^\omega \eta_m + \mathcal{L}(\gamma)\} = \eta_{m+1}$. Thus, the second condition is proved.

- 3) Since $\eta_{m+1} \leq \eta_m$ for $m \geq m_0$ and η_m is lower bounded by 0, then η_m converges to a value η^* . Since $\mathcal{V}(m)$ converges to $\mathbf{w}^*$, it can be inferred from (69) that $\eta^* = \|\mathbf{G}^*\|_\infty^\omega \eta^*$. Due to $\|\mathbf{G}^*\|_\infty^\omega < 1$, we have $\eta^* = 0$. From (68), $\eta^* = 0$ means that $\mathcal{B}(m)$ converges to ϕ . Hence, the third condition is proved

APPENDIX B

PROOF OF PROPOSITION 4

First, if $\mathcal{S}_1 \neq \emptyset$, for any $\mathbf{c} \in \mathcal{S}_1$, according to the definition of $\mathcal{S}_1$ in (17), it is known that $c_{ij} \leq -\frac{p_{ij}^2}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} c_{ki}}$ and $p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} c_{ki} > 0$ for all $(i, j) \in \mathcal{E}$, which are equivalent

to $\frac{c_{ij}}{p_{jj}} \leq -\frac{p_{ij}^2}{p_{ii}p_{jj}(1+\sum_{k \in \mathcal{N}(i) \setminus j} c_{ki}/p_{ii})}$ and $p_{ii}(1 + \sum_{k \in \mathcal{N}(i) \setminus j} c_{ki}/p_{ii}) > 0$. With the definition $\tilde{c}_{ij} \triangleq c_{ij}/p_{jj}$, the two relations can be further written as $\tilde{c}_{ij} \leq -\frac{\tilde{p}_{ij}^2}{1+\sum_{k \in \mathcal{N}(i) \setminus j} \tilde{c}_{ki}}$ and $1+\sum_{k \in \mathcal{N}(i) \setminus j} \tilde{c}_{ki} > 0$ for all $(i, j) \in \mathcal{E}$, where $\tilde{p}_{ij} = \frac{p_{ij}}{\sqrt{p_{ii}}\sqrt{p_{jj}}}$ due to $\tilde{\mathbf{P}} = \text{diag}^{-\frac{1}{2}}(\mathbf{P})\mathbf{P}\text{diag}^{-\frac{1}{2}}(\mathbf{P})$. From the definition of $\tilde{\mathcal{S}}_1$ in (17), it can be seen that $\tilde{\mathbf{c}} \in \tilde{\mathcal{S}}_1$, where $\tilde{\mathbf{c}}$ contains elements $\tilde{c}_{ij}$ for $(i, j) \in \mathcal{E}$ arranged in ascending order first on j and then on i . Thus, we have $\tilde{\mathcal{S}}_1 \neq \emptyset$. Similarly, we can also prove that if $\tilde{\mathcal{S}}_1 \neq \emptyset$, then $\mathcal{S}_1 \neq \emptyset$.

Second, due to $\mathcal{S}_1 \neq \emptyset$ and thereby $\tilde{\mathcal{S}}_1 \neq \emptyset$, the optimal solution $\mathbf{w}^*$ and $\tilde{\mathbf{w}}^*$ of (19) exist, and thereby $\mathbf{G}^*$ and $\tilde{\mathbf{G}}^*$ defined in (29) exist, too. Using $\lim_{t \rightarrow \infty} \mathbf{v}^a(t) = \mathbf{w}^*$ and $\lim_{t \rightarrow \infty} \tilde{\mathbf{v}}^a(t) = \tilde{\mathbf{w}}^*$, we can obtain from (11) that $w_{ij}^* = -\frac{p_{ij}^2}{p_{ii}+\sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*}$ and $\tilde{w}_{ij}^* = -\frac{\tilde{p}_{ij}^2}{1+\sum_{k \in \mathcal{N}(i) \setminus j} \tilde{w}_{ki}^*}$ for all $t \geq 0$. Dividing p_{jj} on both sides of $w_{ij}^* = -\frac{p_{ij}^2}{p_{ii}+\sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^*}$ gives $\frac{w_{ij}^*}{p_{jj}} = -\frac{p_{ij}^2}{p_{jj}p_{ii}(1+\sum_{k \in \mathcal{N}(i) \setminus j} \frac{w_{ki}^*}{p_{ii}})}$. Comparing this result to $\tilde{w}_{ij}^* = -\frac{\tilde{p}_{ij}^2}{1+\sum_{k \in \mathcal{N}(i) \setminus j} \tilde{w}_{ki}^*}$, we can infer that

$$\tilde{w}_{ij}^* = \frac{w_{ij}^*}{p_{jj}} \quad (71)$$

for all $(i, j) \in \mathcal{E}$. By using (71), we have $1 + \sum_{k \in \mathcal{N}(i) \setminus j} \tilde{w}_{ki}^* = \frac{1}{p_{ii}} \left(p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{ki}^* \right)$. Writing it into a vector form gives

$$\mathbf{A}\tilde{\mathbf{w}}^* + \mathbf{1} = \text{diag}^{-1}(\mathbf{u})(\mathbf{A}\mathbf{w}^* + \mathbf{u}), \quad (72)$$

where $\mathbf{u}$ is defined after (16). On the other hand, due to $\tilde{p}_{ij} = \frac{p_{ij}}{\sqrt{p_{ii}}\sqrt{p_{jj}}}$, writing it into a vector form, we obtain

$$\tilde{\mathbf{p}} = \text{diag}^{-\frac{1}{2}}(\mathbf{u})\text{diag}^{-\frac{1}{2}}(\bar{\mathbf{u}})\mathbf{p}, \quad (73)$$

where the vector $\tilde{\mathbf{p}}$ contains $\tilde{p}_{ij}$ with $(i, j) \in \mathcal{E}$ arranged in ascending order first on j and then on i ; and $\bar{\mathbf{u}} = [p_{11}\mathbf{1}_{|\mathcal{N}(1)|}^T, p_{22}\mathbf{1}_{|\mathcal{N}(2)|}^T, \dots, p_{NN}\mathbf{1}_{|\mathcal{N}(N)|}^T]^T$ with $\mathbf{1}_{|\mathcal{N}(i)|}$ denoting an all ones vector of length $|\mathcal{N}(i)|$. Substituting (72) and (73) into the definition of $\tilde{\mathbf{G}}^* = -\text{diag}^{-1}(\mathbf{A}\tilde{\mathbf{w}}^* + \mathbf{1})\text{diag}(\tilde{\mathbf{p}})\mathbf{A}$, we obtain

$$\begin{aligned} \tilde{\mathbf{G}}^* &= -\text{diag}(\mathbf{u})\text{diag}^{-1}(\mathbf{A}\mathbf{w}^* + \mathbf{u})\text{diag}^{-\frac{1}{2}}(\mathbf{u})\text{diag}^{-\frac{1}{2}}(\bar{\mathbf{u}})\text{diag}(\mathbf{p})\mathbf{A} \\ &= \text{diag}^{-\frac{1}{2}}(\bar{\mathbf{u}})\text{diag}^{\frac{1}{2}}(\mathbf{u})\mathbf{G}^*, \end{aligned} \quad (74)$$

where the second step is due to the fact that positions of diagonal matrices can be interchanged. It can be verified that $\text{diag}^{\frac{1}{2}}(\mathbf{u})\mathbf{G}^* = \mathbf{G}^*\text{diag}^{\frac{1}{2}}(\bar{\mathbf{u}})$, thus we further have

$$\tilde{\mathbf{G}}^* = \text{diag}^{-\frac{1}{2}}(\bar{\mathbf{u}})\mathbf{G}^*\text{diag}^{\frac{1}{2}}(\bar{\mathbf{u}}). \quad (75)$$

Obviously, $\mathbf{G}^*$ and $\tilde{\mathbf{G}}^*$ are similar matrices. From basic properties of similar matrices, we have $\rho(\tilde{\mathbf{G}}^*) = \rho(\mathbf{G}^*)$. On the other hand, taking absolute value of elements on both sides of (75) gives

$$|\tilde{\mathbf{G}}^*| = \text{diag}^{-\frac{1}{2}}(|\bar{\mathbf{u}}|)|\mathbf{G}^*|\text{diag}^{\frac{1}{2}}(|\bar{\mathbf{u}}|). \quad (76)$$

It is obvious that $|\mathbf{G}^*|$ and $|\tilde{\mathbf{G}}^*|$ are similar, and hence $\rho(|\tilde{\mathbf{G}}^*|) = \rho(|\mathbf{G}^*|)$.

Third, putting (71) into $1 + \sum_{k \in \mathcal{N}(i)} \tilde{w}_{ki}^*$ gives $1 + \sum_{k \in \mathcal{N}(i)} \tilde{w}_{ki}^* = \frac{1}{p_{ii}}(p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^*)$. Thus, $p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{ki}^* \neq 0$ if and only if $1 + \sum_{k \in \mathcal{N}(i)} \tilde{w}_{ki}^* \neq 0$.

REFERENCES

- [1] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, pp. 498-519, 2001.
- [2] H. A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, pp. 28-41, 2004.
- [3] H. A. Loeliger, J. Dauwels, H. Junli, S. Korl, P. Li, and F. R. Kschischang, "The Factor graph approach to model-based signal processing," *Proc. IEEE*, vol. 95, no. 6, pp. 1295-1322, Jun. 2007.
- [4] C.M. Bishop, *Pattern Recognition and Machine Learning*. Springer, Aug. 2006.
- [5] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. The MIT Press, 2012.
- [6] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Comput.*, vol. 13, pp. 2173-2200, 2001.
- [7] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multiuser detection," in *Proc. IEEE Information Theory Workshop*, Punta del Este, Uruguay, Mar. 2006.
- [8] Q. Guo and D. Huang, "EM-based joint channel estimation and detection for frequency selective channels using Gaussian message passing," *IEEE Trans. Signal Process.*, vol. 59, pp. 4030-4035, 2011.
- [9] Q. Guo and P. Li, "LMMSE turbo equalization based on factor graphs," *IEEE J. Sel. Areas Commun.*, vol. 26, pp. 311-319, 2008.
- [10] Y. El-Kurdi, W. J. Gross, and D. Giannacopoulos, "Efficient implementation of Gaussian belief propagation solver for large sparse diagonally dominant linear systems," *IEEE Trans. Magn.*, vol. 48, no. 2, pp. 471-474, Feb. 2012.
- [11] 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. Inform. Theory (ISIT)*, pp. 1863-1867, 2008.
- [12] M. W. Seeger and D. P. Wipf, "Variational Bayesian inference techniques," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 8191, Nov. 2010.
- [13] V. Chandrasekaran, J. K. Johnson, and A. S. Willsky, "Estimation in Gaussian graphical models using tractable subgraphs: a walk-sum analysis," *IEEE Trans. Signal Process.*, vol. 56, pp. 1916-1930, 2008.
- [14] Y. Liu, V. Chandrasekaran, A. Anandkumar, and A. S. Willsky, "Feedback message passing for inference in Gaussian graphical models," *IEEE Trans. Signal Process.*, vol. 60, no. 8, pp. 4135-4150, 2012.
- [15] N. Boon Loong, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5491-5499, 2008.

- [16] D. Dolev, A. Zymnis, S. P. Boyd, D. Bickson, and Y. Tock, "Distributed large scale network utility maximization," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 829-833, 2009.
- [17] M. Leng and Y. C. Wu, "Distributed clock synchronization for wireless sensor networks using belief propagation," *IEEE Trans. Signal Process.*, vol.59, no.11, pp. 5404-5414, Nov. 2011.
- [18] A. Ahmad, D. Zennaro, E. Serpedin, and L. Vangelista, "A factor graph approach to clock offset estimation in wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4244-4260, Jul. 2012.
- [19] M. Leng, T. Wee-Peng, and T. Q. S. Quek, "Cooperative and distributed localization for wireless sensor networks in multipath environments," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Process. (ICASSP)*, pp. 3125-3128, 2012.
- [20] Y. Song, B. Wang, Z. Shi, K. Pattipati, and S. Gupta, "Distributed algorithms for energy-efficient even self-deployment in mobile sensor networks," in *IEEE Trans. Mobile Compt.*, pp. 1-14, 2013.
- [21] G. Zhang, W. Xu, Yaming, Wang, "Fast distributed rate control algorithm with QoS support in Ad Hoc networks," in *Proc. IEEE Global Telecommunications Conf. (Globecom)*, 2010.
- [22] F. Lehmann, "Iterative mitigation of intercell interference in cellular networks based on Gaussian belief propagation," *IEEE Trans. Veh. Technol.*, vol. 61, pp. 2544-2558, 2012.
- [23] C. C. Moallemi and B. Van Roy, "Convergence of min-sum message passing for quadratic optimization," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2413-2423, 2009.
- [24] D. M. Malioutov, J. Johnson, and A. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *J. Mach. Learn. Res.*, vol. 7, no. 1, pp. 2031-2064, 2006.
- [25] K. P. Murphy, Y. Weiss, and M. I. Jordan, "Loopy belief propagation for approximate inference: an empirical study," in *Proc. 15th Conf. Uncertainty in Artificial Intelligence (UAI)*, Stockholm, Sweden, pp. 467-475, Jul. 1999.
- [26] Y. El-Kurdi, D. Giannacopoulos, and W. J. Gross, "Relaxed Gaussian belief propagation," *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, pp. 2002-2006, 2012.
- [27] M. Pretti, "A message-passing algorithm with damping," *J. Stat. Mech.*, Nov. 2005.
- [28] P. Som, T. datta, N. Srinidhi, A. Chockalingam, and B. S. Rajan, "Low-complexity detection in large-dimension MIMO-ISI channels using graphical models," *IEEE J. Sel. Topics Signal Process.*, vol. 5, no. 8, pp. 1497-1511, Dec. 2011.
- [29] Q. Su and Y. C. Wu, "Convergence analysis of the variance in Gaussian belief propagation," *IEEE Trans. Signal Process.*, vol. 62, no. 19, pp. 5119-5131, Oct. 2014.
- [30] D.P. Bertsekas and J.N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1989.
- [31] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49-95, Mar. 1996.
- [32] M. Grant, S. Boyd, and Y. Ye, cvx: Matlab software for disciplined convex programming. [Online]. Available: <http://www.stanford.edu/~boyd/cvx/>
- [33] J. F. Sturm, "Using SeDumi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Methods Software*, no. 11-22, pp.625-653, 1999.
- [34] R. Horn and C.R. Johnson, *Matrix Analysis*. New York: Cambridge Univ. Press, 1985.



Qinliang Su received the B.Eng degree from Chongqing University in 2007 and M.Eng degree from Zhejiang University in 2010 (both with honors), respectively. He obtained his Ph.D. degree in electrical and electronic engineering from the University of Hong Kong, Hong Kong, in December 2014. His current research interests include signal processing for wireless communications and networks; statistical inference and learning; graphical models; distributed and parallel computation; convex optimization theories.



Yik-Chung Wu received the B.Eng. (EEE) degree in 1998 and the M.Phil. degree in 2001 from the University of Hong Kong (HKU). He received the Croucher Foundation scholarship in 2002 to study Ph.D. degree at Texas A&M University, College Station, and graduated in 2005. From August 2005 to August 2006, he was with the Thomson Corporate Research, Princeton, NJ, as a Member of Technical Staff. Since September 2006, he has been with HKU, currently as an Associate Professor. He was a visiting scholar at Princeton University, in summer 2011. His research interests are in general area of signal processing and communication systems, and in particular distributed signal processing and communications; optimization theories for communication systems; estimation and detection theories in transceiver designs; and smart grid. Dr. Wu served as an Editor for IEEE Communications Letters, is currently an Editor for IEEE Transactions on Communications and Journal of Communications and Networks.