Title	Convergence Analysis of the Variance in Gaussian Belief Propagation
Author(s)	SU, Q; Wu, YC
Citation	IEEE Transactions on Signal Processing, 2014, v. 62, p. 5119-5131
Issued Date	2014
URL	http://hdl.handle.net/10722/214162
Rights	Creative Commons: Attribution 3.0 Hong Kong License

Convergence Analysis of the Variance in Gaussian Belief Propagation

Qinliang Su, *Member, IEEE*, and Yik-Chung Wu

***Abstract*—It is known that Gaussian belief propagation (BP) is a low-complexity algorithm for (approximately) computing the marginal distribution of a high dimensional Gaussian distribution. However, in loopy factor graph, it is important to determine whether Gaussian BP converges. In general, the convergence conditions for Gaussian BP variances and means are not necessarily the same, and this paper focuses on the convergence condition of Gaussian BP variances. In particular, by describing the message-passing process of Gaussian BP as a set of updating functions, the necessary and sufficient convergence conditions of Gaussian BP variances are derived under both synchronous and asynchronous schedulings, with the converged variances proved to be independent of the initialization as long as it is chosen from the proposed set. The necessary and sufficient convergence condition is further expressed in the form of a semi-definite programming (SDP) optimization problem, thus can be verified more efficiently compared to the existing convergence condition based on computation tree. The relationship between the proposed convergence condition and the existing one based on computation tree is also established analytically. Numerical examples are presented to corroborate the established theories.**

***Index Terms*—Convergence, Gaussian belief propagation, graphical model, factor graph, loopy belief propagation, message passing, sum-product algorithm.**

I. Introduction

In signal processing and machine learning, many problems eventually come to the issue of computing the marginal mean and variance of an individual random variable from a high dimensional joint Gaussian probability density function (PDF). A direct way of marginalization involves the computation of the inverse of the precision matrix in the joint Gaussian PDF. The inverse operation is known to be computationally expensive for a high dimensional matrix, and would introduce heavy communication overhead when carried out in distributed scenarios.

By representing the joint PDF with a factor graph [1], Gaussian BP provides an alternative to (approximately) calculate the marginal mean and variance for each individual random variable by passing messages between neighboring nodes in the factor graph [2]–[5]. It is known that if the factor graph is of tree structure, the belief means and variances calculated with Gaussian BP both converge to the true marginal means and variances, respectively [1]. For a factor graph with loops, if the belief means and variances in Gaussian BP converge, the true marginal means and approximate marginal variances are obtained [6]. Recently, a novel message-passing algorithm has been proposed in [7] for inference in Gaussian graphical model by choosing a special set of nodes to break loops in the graph. Although this algorithm computes the true marginal means and improves the accuracy of variances, it still needs to use Gaussian BP as an underlying inference algorithm. Thus, in this paper, we focus on the message-passing algorithm of Gaussian BP only.

With the ability to provide the true marginal means upon convergence, Gaussian BP has been successfully applied in low-complexity detection and estimation problems arising in communication systems [8]–[10], fast solver for large sparse linear systems [11], [12], sparse Bayesian learning [13], estimation in Gaussian graphical model [14], etc. In addition, the distributed property inherited from message passing algorithms is particularly attractive for applications requiring distributed implementation, such as distributed beamforming [15], inter-cell interference mitigation [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 distributed network utility maximization [22]. Moreover, Gaussian BP is also exploited to provide approximate marginal variances for large-scale sparse Bayesian learning in a computationally efficient way [23].

However, Gaussian BP only works under the prerequisite that the belief means and variances calculated from the updating messages do converge. So far, several sufficient convergence conditions have been proposed, which can guarantee the means and variances converge simultaneously [6], [24], [25]. However, in general, the belief means and variances do not necessarily converge under the same condition. It is reported in [24] that if the variances converge, the convergence of the means can always be observed when suitable damping is imposed. Thus, it is important to ensure the variances of Gaussian BP to converge in the first place. In the pioneering work [24], the message-passing procedure of Gaussian BP in a loopy factor graph is equivalently translated into that in a loop-free computation tree [26]. Based on the computation tree, an almost necessary and sufficient convergence condition of variances is derived. It is proved that if the spectral radius of the infinite dimensional precision matrix of the computation tree is smaller than one, the variances converge; and if this spectral radius is larger than one, the Gaussian BP becomes ill-posed. However, this condition is only 'almost necessary and sufficient' since it does not cover

Manuscript received January 16, 2014; revised April 30, 2014 and July 24, 2014; accepted July 26, 2014. Date of publication August 06, 2014; date of current version August 27, 2014. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Ami Wiesel. This work was supported in part by the HKU Seed Funding Programme under Project 201311159066.

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).

Digital Object Identifier 10.1109/TSP.2014.2345635

the scenario when the spectral radius is equal to one. More crucially, this convergence condition requires the evaluation of the spectral radius of an infinite dimensional matrix, which is almost impossible to be calculated in practice.

In this paper, with the fact that the messages in Gaussian BP can be represented by linear and quadratic parameters, the message-passing process of Gaussian BP is described as a set of updating functions of parameters. Based on the monotonically non-decreasing and concave properties of the updating functions, the necessary and sufficient convergence condition of messages' quadratic parameters is derived first. Then, with the relation between BP messages' quadratic parameters and belief variances, the necessary and sufficient convergence condition of belief variances is developed under both synchronous and asynchronous schedulings. The initialization set under which the variances are guaranteed to converge is proposed as well. The convergence condition derived in this paper is proved to be equivalent to a semi-definite programming (SDP) problem, thus can be easily verified. Furthermore, the relationship between the proposed convergence condition and the existing one based on computation tree is also established. Our result fills in the missing part of the convergence condition in [24] on the case when the spectral radius of the infinite dimensional matrix is equal to one.

The rest of this paper is organized as follows. Gaussian BP is reviewed in Section II. Section III analyzes the updating process of Gaussian BP messages, followed by the necessary and sufficient convergence condition of quadratic parameters in the messages in Section IV. The derivation of the necessary and sufficient convergence condition of belief variances is presented in Section V. Relationship between the condition proposed in this paper and the existing one based on computation tree is established in Section VI. Numerical examples are presented in Section VII, followed by conclusions in Section VIII.

The following notations are used throughout this paper. For two vectors, $\mathbf{x}_1 \geq \mathbf{x}_2$ and $\mathbf{x}_1 > \mathbf{x}_2$ mean the inequalities are held in all corresponding elements. Notations $\lambda(\mathbf{G})$ and $\lambda_{\max}(\mathbf{G})$ represent any eigenvalue and the maximal eigenvalue of matrix $\mathbf{G}$, respectively. Notation $\rho(\mathbf{G})$ means the spectral radius of $\mathbf{G}$. Notations $[\cdot]_i$ and $[\cdot]_{ij}$ represent the i-th and (i,j)-th element of a vector and matrix, respectively.

II. Gaussian Belief Propagation

In general, a Gaussian PDF can be written as

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

where $\mathbf{x} = [x_1, x_2, \ldots, x_N]^T$ with N being the number of random variables; $\mathbf{P} \succ 0$ is the precision matrix with p_{ij} being its (i,j)-th element; and $\mathbf{h} = [h_1, h_2, \ldots, h_N]^T$ is the linear coefficient vector. In many signal processing applications, we want to find the marginalized PDF $p(x_i) = \int f(\mathbf{x})dx_1 \cdots dx_{i-1}dx_{i+1} \cdots dx_N$. For a Gaussian $f(\mathbf{x})$, this can be done by completing the square inside the exponent of (1) as $f(\mathbf{x}) \propto \exp\{-\frac{1}{2}(\mathbf{x} - \mathbf{P}^{-1}\mathbf{h})^T\mathbf{P}(\mathbf{x} - \mathbf{P}^{-1}\mathbf{h})\}$, and then the marginalized PDF is obtained as

$$p(x_i) \propto \exp\left\{-\frac{(x_i - \mu_i)^2}{2\sigma_i^2}\right\}, \tag{2}$$

with the mean $\mu_i = [\mathbf{P}^{-1}\mathbf{h}]_i$ and the variance $\sigma_i^2 = [\mathbf{P}^{-1}]_{ii}$. However, the required matrix inverse is computationally expensive for a large dimensional $\mathbf{P}$, and introduces heavy communication overhead to the network when the information p_{ij} is located distributedly.

On the other hand, the Gaussian PDF in (1) can be expanded as $f(\mathbf{x}) \propto \prod_{i=1}^N f_i(x_i) \prod_{j=1}^N \prod_{k=j+1}^N f_{jk}(x_j, x_k)$, 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[1] can be constructed by connecting each variable x_i with its associated factors $f_i(x_i)$ and $f_{ij}(x_i, x_j)$. It is known that Gaussian BP can be applied on this factor graph by passing messages among neighboring nodes to obtain the true marginal mean and approximate marginal variance for each variable.

In Gaussian BP, the departing and arriving messages corresponding to variables i and j are updated as

$$m_{i\to j}^d(x_i, t) \propto f_i(x_i) \prod_{k\in\mathcal{N}(i)\backslash j} m_{k\to i}^a(x_i, t), \tag{3}$$

$$m_{i\to j}^a(x_j, t+1) \propto \int f_{ij}(x_i, x_j) m_{i\to j}^d(x_i, t) dx_i, \tag{4}$$

where $(i,j) \in \mathcal{E}$ with $\mathcal{E} \triangleq \{(i,j) \,|\, p_{ij} \neq 0, \text{for}\, i,j \in \mathcal{V}\}$ being the set of index pairs of any two connected variables and $\mathcal{V} \triangleq \{1, 2, \ldots, N\}$ being the set of indices of all variables; t is the time index; $\mathcal{N}(i)$ is the set of indices of neighboring variable nodes of node i, and $\mathcal{N}(i)\backslash j$ is the set $\mathcal{N}(i)$ but excluding the node j. After obtaining the messages $m_{k\to i}^a(x_i, t)$, the belief at variable node i is equal to

$$b_i(x_i, t) \propto f_i(x_i) \prod_{k\in\mathcal{N}(i)} m_{k\to i}^a(x_i, t). \tag{5}$$

It has been proved that $b_i(x_i, t)$ always converges to $p(x_i)$ in (2) exactly if the factor graph is of tree structure [1], [6]. For loopy factor graph, if $b_i(x_i, t)$ converges, the mean of $b_i(x_i, t)$ will converge to the true mean μ_i, while the converged variance is an approximation of σ_i^2 [6].

Assume the arriving message is in Gaussian form of $m_{i\to j}^a(x_j, t) \propto \exp\{-\frac{v_{i\to j}^a(t)}{2}x_j^2 + \beta_{i\to j}^a(t)x_j\}$, where $v_{i\to j}^a(t)$ and $\beta_{i\to j}^a(t)$ are the arriving precision and arriving linear coefficient, respectively. Inserting it into (3), we obtain $m_{i\to j}^d(x_i, t) \propto \exp\{-\frac{v_{i\to j}^d(t)}{2}x_i^2 + \beta_{i\to j}^d(t)x_i\}$, where

$$v_{i\to j}^d(t) = p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} v_{k\to i}^a(t), \tag{6}$$

$$\beta_{i\to j}^d(t) = h_i + \sum_{k\in\mathcal{N}(i)\backslash j} \beta_{k\to i}^a(t) \tag{7}$$

are the departing precision and linear coefficient, respectively. Furthermore, substituting the departing message $m_{i\to j}^d(x_i, t)$ into (4), we obtain

$$\begin{aligned} m_{i\to j}^a(x_j, t+1) &\propto \exp\left\{\frac{p_{ij}^2}{2v_{i\to j}^d(t)}x_j^2 - \frac{p_{ij}\beta_{i\to j}^d(t)}{v_{i\to j}^d(t)}x_j\right\} \\ &\quad \times \int \exp\left\{-\frac{v_{i\to j}^d(t)}{2}\left(x_i - \frac{\beta_{i\to j}^d(t) - p_{ij}x_j}{v_{i\to j}^d(t)}\right)^2\right\} dx_i. \end{aligned} \tag{8}$$

[1]The factor graph is assumed to be connected in this paper.

If $v^d_{i\to j}(t) > 0$, the integration equals to a constant, and thus $m^a_{i\to j}(x_j, t+1) \propto \exp\{\frac{p^2_{ij}}{2v^d_{i\to j}(t)}x_j^2 - \frac{p_{ij}\beta^d_{i\to j}(t)}{v^d_{i\to j}(t)}x_j\}$. Therefore, $v^a_{i\to j}(t+1)$ and $\beta^a_{i\to j}(t+1)$ are updated as

$$v^a_{i\to j}(t+1) = \begin{cases} -\frac{p^2_{ij}}{v^d_{i\to j}(t)}, & \text{if } v^d_{i\to j}(t) > 0, \\ \text{not defined}, & \text{otherwise}, \end{cases} \tag{9}$$

$$\beta^a_{i\to j}(t+1) = \begin{cases} -\frac{p_{ij}\beta^d_{i\to j}(t)}{v^d_{i\to j}(t)}, & \text{if } v^d_{i\to j}(t) > 0, \\ \text{not defined}, & \text{otherwise}, \end{cases} \tag{10}$$

where 'not defined' arises since under the case of $v^d_{i\to j}(t) \le 0$, the integration in (8) becomes infinite and the message $m^a_{i\to j}(x_j, t+1)$ loses its Gaussian form. After obtaining $v^a_{k\to i}(t)$, the variance of belief at each iteration is computed as

$$\sigma_i^2(t) = \frac{1}{p_{ii} + \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)}. \tag{11}$$

From (9) and (10), it can be seen that the messages of Gaussian BP maintain the Gaussian form only when $v^d_{i\to j}(t) > 0$ for all $t \ge 0$. Therefore, we have the following lemma.

Lemma 1: The messages of Gaussian BP are always in Gaussian form if and only if $v^d_{i\to j}(t) > 0$ for all $t \ge 0$.

III. Analysis of the Message-Passing Process

In this section, we first analyze the updating process of $v^a_{i\to j}(t)$, and then derive the condition to guarantee the BP messages being maintained at Gaussian form.

A. Updating Function of $v^a_{i\to j}(t)$ and Its Properties

Under the Gaussian form of messages, substitutting (6) into (9) gives

$$v^a_{i\to j}(t+1) = -\frac{p^2_{ij}}{p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} v^a_{k\to i}(t)}. \tag{12}$$

By writing (12) into a vector form, we obtain

$$\mathbf{v}^a(t+1) = \mathbf{g}(\mathbf{v}^a(t)), \tag{13}$$

where $\mathbf{g}(\cdot)$ is a vector-valued function containing components $g_{ij}(\cdot)$ with $(i,j) \in \mathcal{E}$ arranged in ascending order first on j and then on i, and $g_{ij}(\cdot)$ is defined as

$$g_{ij}(\mathbf{w}) \triangleq -\frac{p^2_{ij}}{p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} w_{ki}}; \tag{14}$$

$\mathbf{v}^a(t)$ and $\mathbf{w}$ are vectors containing elements $v^a_{i\to j}(t)$ and w_{ij}, respectively, both with $(i,j) \in \mathcal{E}$ arranged in ascending order first on j and then on i. Moreover, define the set

$$\mathcal{W} \triangleq \left\{ \mathbf{w} \,|\, p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} w_{ki} > 0, \forall (i,j) \in \mathcal{E} \right\}. \tag{15}$$

Now, we have the following proposition about $\mathbf{g}(\cdot)$ and $\mathcal{W}$.

Proposition 1: The following claims hold:

P1) For any $\mathbf{w}_1 \in \mathcal{W}$, if $\mathbf{w}_1 \le \mathbf{w}_2$, then $\mathbf{w}_2 \in \mathcal{W}$;

P2) For any $\mathbf{w}_1 \in \mathcal{W}$, if $\mathbf{w}_1 \le \mathbf{w}_2$, then $\mathbf{g}(\mathbf{w}_1) \le \mathbf{g}(\mathbf{w}_2)$;

P3) $g_{ij}(\mathbf{w})$ is a concave function with respect to $\mathbf{w} \in \mathcal{W}$.

Proof: Consider two vectors $\bar{\mathbf{w}}$ and $\hat{\mathbf{w}}$ in $\mathcal{W}$, which contain elements $\bar{w}_{ki}$ and $\hat{w}_{ki}$ with $(k,i) \in \mathcal{E}$ arranged in ascending order first on k and then on i. For any $\bar{\mathbf{w}} \in \mathcal{W}$, according to the definition of $\mathcal{W}$ in (15), we have $p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} \bar{w}_{ki} > 0$. Then, if $\bar{\mathbf{w}} \le \hat{\mathbf{w}}$, it can be easily seen that $p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} \hat{w}_{ki} > 0$ as well. Thus, we have $\hat{\mathbf{w}} \in \mathcal{W}$.

The first-order derivative of $g_{ij}(\mathbf{w})$ with respect to w_{ki} for $k \in \mathcal{N}(i)\backslash j$ is computed to be

$$\frac{\partial g_{ij}}{\partial w_{ki}} = \frac{p^2_{ij}}{\left(p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} w_{ki}\right)^2} > 0. \tag{16}$$

Thus, $g_{ij}(\mathbf{w})$ is a continuous and strictly increasing function with respect to the components w_{ki} for $k \in \mathcal{N}(i)\backslash j$ with $\mathbf{w} \in \mathcal{W}$. Hence, we have if $\mathbf{w}_1 \le \mathbf{w}_2$, then $\mathbf{g}(\mathbf{w}_1) \le \mathbf{g}(\mathbf{w}_2)$.

To prove the third property, consider the simple function $-\frac{p^2_{ij}}{p_{ii}+r}$ under the domain of $r > -p_{ii}$ first. It can be easily derived that the second order derivative of $-\frac{p^2_{ij}}{p_{ii}+r}$ is $-\frac{2p^2_{ij}}{(p_{ii}+r)^3} < 0$, where the inequality holds due to $p_{ii} + r > 0$. Thus, $-\frac{p^2_{ij}}{p_{ii}+r}$ is a concave function with respect to $r > -p_{ii}$. Since $g_{ij}(\mathbf{w}) = -\frac{p^2_{ij}}{p_{ii}+\sum_{k\in\mathcal{N}(i)\backslash j} w_{ki}}$ is the composition of the concave function $-\frac{p^2_{ij}}{p_{ii}+r}$ and the affine mapping $\sum_{k\in\mathcal{N}(i)\backslash j} w_{ki}$, the composition function $g_{ij}(\mathbf{w})$ is also concave with respect to $\sum_{k\in\mathcal{N}(i)\backslash j} w_{ki} > -p_{ii}$ ([28], p. 79). Due to $\mathcal{W} \triangleq \{\mathbf{w} \,|\, p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} w_{ki} > 0, \forall (i,j) \in \mathcal{E}\}$ as defined in (15), then $\mathbf{w} \in \mathcal{W}$ implies that $\sum_{k\in\mathcal{N}(i)\backslash j} w_{ki} > -p_{ii}$. Thus, we can infer that $g_{ij}(\mathbf{w})$ is a concave function with respect to $\mathbf{w} \in \mathcal{W}$. ■

Notice that convexity is also exploited in [29] to derive the convergence condition for a general min-sum message-passing algorithm, which reduces to Gaussian BP when it is applied to the particular quadratic function $\frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} - \mathbf{h}^T\mathbf{x}$. The convexity in [29] means that the quadratic function can be decomposed into summation of convex functions, which is different from the convexity in the updating function $g_{ij}(\mathbf{w})$ here.

B. Condition to Maintain the Gaussian Form of Messages

First, define the following set[2]

$$\mathcal{S}_1 \triangleq \{\mathbf{w} \in \mathcal{W} \,|\, \mathbf{w} \le \mathbf{g}(\mathbf{w})\}. \tag{17}$$

With notations $\mathbf{g}^{(t)}(\mathbf{w}) \triangleq \mathbf{g}(\mathbf{g}^{(t-1)}(\mathbf{w}))$ and $\mathbf{g}^{(0)}(\mathbf{w}) \triangleq \mathbf{w}$, the following proposition can be established.

Proposition 2: The set $\mathcal{S}_1$ has the following properties:

P4) $\mathcal{S}_1$ is a convex set;

P5) If $\mathbf{s} \in \mathcal{S}_1$, then $\mathbf{s} < \mathbf{0}$;

P6) If $\mathbf{s} \in \mathcal{S}_1$, then $\mathbf{g}^{(t)}(\mathbf{s}) \in \mathcal{S}_1$ and $\mathbf{g}^{(t)}(\mathbf{s}) \le \mathbf{g}^{(t+1)}(\mathbf{s})$ for all $t \ge 0$.

Proof: As $g_{ij}(\mathbf{w})$ is a concave function for $\mathbf{w} \in \mathcal{W}$ according to P3), thus $g_{ij}(\mathbf{w}) - w_{ij}$ is also concave for

[2]The simpler symbol $\mathcal{S}$ is reserved for later use.

$\mathbf{w} \in \mathcal{W}$. The set of $\mathbf{w} \in \mathcal{W}$ which satisfies the condition of $g_{ij}(\mathbf{w}) - w_{ij} \geq 0$ is a convex set [28]. Thus, $\mathcal{S}_1 = \{\mathbf{w} \,|\, \mathbf{w} \leq \mathbf{g}(\mathbf{w}) \text{ with } \mathbf{w} \in \mathcal{W}\} = \{g_{ij}(\mathbf{w}) - w_{ij} \geq 0 \text{ for all } (i,j) \in \mathcal{E} \text{ with } \mathbf{w} \in \mathcal{W}\}$ is also a convex set.

If $\mathbf{s} \in \mathcal{S}_1$, we have $\mathbf{s} \leq \mathbf{g}(\mathbf{s})$ and $\mathbf{s} \in \mathcal{W}$. According to the definition of $\mathcal{W}$ in (15), $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} s_{ki} > 0$. Putting this fact into the definition of $g_{ij}(\cdot)$ in (14), it is obvious that $\mathbf{g}(\mathbf{s}) < \mathbf{0}$. Therefore, we have the relation that $\mathbf{s} \leq \mathbf{g}(\mathbf{s}) < \mathbf{0}$ for all $\mathbf{s} \in \mathcal{S}_1$.

Finally, if $\mathbf{s} \in \mathcal{S}_1$, we have $\mathbf{s} \leq \mathbf{g}(\mathbf{s})$ and $\mathbf{s} \in \mathcal{W}$. Hence, $\mathbf{g}(\mathbf{s}) \in \mathcal{W}$ according to the P1). Applying $\mathbf{g}(\cdot)$ on both sides of $\mathbf{s} \leq \mathbf{g}(\mathbf{s})$ and using P2), we obtain $\mathbf{g}(\mathbf{s}) \leq \mathbf{g}^{(2)}(\mathbf{s})$. Furthermore, since $\mathbf{g}(\mathbf{s}) \in \mathcal{W}$ and $\mathbf{g}(\mathbf{s}) \leq \mathbf{g}^{(2)}(\mathbf{s})$, we also have $\mathbf{g}(\mathbf{s}) \in \mathcal{S}_1$. By induction, we can prove in general that $\mathbf{g}^{(t)}(\mathbf{s}) \in \mathcal{S}_1$ and $\mathbf{g}^{(t)}(\mathbf{s}) \leq \mathbf{g}^{(t+1)}(\mathbf{s})$ for all $t \geq 0$. ■

Now, we give the following theorem.

Theorem 1: There exists at least one initialization $\mathbf{v}^a(0)$ such that the messages of Gaussian BP are maintained at Gaussian form for all $t \geq 0$ if and only if $\mathcal{S}_1 \neq \emptyset$.

Proof:

Sufficient Condition:

If $\mathcal{S}_1 \neq \emptyset$, by choosing an initialization $\mathbf{v}^a(0) \in \mathcal{S}_1$ and applying P6), it can be inferred that $\mathbf{v}^a(t) \in \mathcal{S}_1$ for all $t \geq 0$. Moreover, according to the definitions of $\mathcal{W}$ in (15) and $\mathcal{S}_1$ in (17), we have $\mathcal{S}_1 \subseteq \mathcal{W}$, and thereby $\mathbf{v}^a(t) \in \mathcal{W}$ for all $t \geq 0$, or equivalently $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} v^a_{k\to i}(t) > 0$ for all $(i,j) \in \mathcal{E}$ according to the definition of set $\mathcal{W}$ in (15). Due to $v^d_{i\to j}(t) = p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} v^a_{k\to i}(t)$ as given in (6), we obtain $v^d_{i\to j}(t) > 0$ for all $(i,j) \in \mathcal{E}$ and $t \geq 0$. According to Lemma 1, it can be inferred that the messages are maintained at Gaussian form for all $t \geq 0$.

Necessary Condition:

We prove this necessary condition by contradiction. When $\mathcal{S}_1 = \emptyset$, suppose that there exists a $\bar{\mathbf{v}}^a(0)$ such that the messages are maintained at Gaussian form for all $t \geq 0$. According to Lemma 1, we can infer that $\bar{v}^d_{i\to j}(t) = p_{ii} + \sum_{k \to \mathcal{N}(i)\backslash j} \bar{v}^a_{k\to i}(t) > 0$ for all $(i,j) \in \mathcal{E}$, or equivalently $\bar{\mathbf{v}}^a(t) \triangleq \mathbf{g}^{(t)}(\bar{\mathbf{v}}^a(0)) \in \mathcal{W}$ for all $t \geq 0$ from the definition of $\mathcal{W}$ in (15).

Now, choose another initialization $\mathbf{v}^a(0)$ satisfying both $\bar{\mathbf{v}}^a(0) \leq \mathbf{v}^a(0)$ and $\mathbf{v}^a(0) \geq \mathbf{0}$. Due to $\bar{\mathbf{v}}^a(0) \in \mathcal{W}$ and $\bar{\mathbf{v}}^a(0) \leq \mathbf{v}^a(0)$, by using P2), we have $\mathbf{g}(\bar{\mathbf{v}}^a(0)) \leq \mathbf{g}(\mathbf{v}^a(0))$, that is, $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1)$. Furthermore, substituting $\mathbf{v}^a(0) \geq \mathbf{0}$ into (14) gives $\mathbf{v}^a(1) \leq \mathbf{0}$, and thereby $\mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Combining with $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1)$ leads to $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Due to the assumption $\bar{\mathbf{v}}^a(1) \in \mathcal{W}$, by applying P2) to $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1) \leq \mathbf{v}^a(0)$, we obtain $\bar{\mathbf{v}}^a(2) \leq \mathbf{v}^a(2) \leq \mathbf{v}^a(1)$. Combining with the fact $\mathbf{v}^a(1) \leq \mathbf{0}$ as proved above, we have $\bar{\mathbf{v}}^a(2) \leq \mathbf{v}^a(2) \leq \mathbf{v}^a(1) \leq \mathbf{0}$. By induction, it can be derived that

$$\bar{\mathbf{v}}^a(t+1) \leq \mathbf{v}^a(t+1) \leq \mathbf{v}^a(t) \leq \mathbf{0}, \quad \forall\, t \geq 1. \tag{18}$$

Then, from the definition of $\mathcal{W}$ in (15), the assumption $\bar{\mathbf{v}}^a(t) \in \mathcal{W}$ is equivalent to $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} \bar{v}^a_{k\to i}(t) > 0$ for all $(i,j) \in \mathcal{E}$, where $\bar{v}^a_{k\to i}(t)$ are the elements of $\bar{\mathbf{v}}^a(t)$ with $(k,i) \in \mathcal{E}$ arranged in the same order as $\mathbf{v}^a(t)$. By rearranging the terms in $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} \bar{v}^a_{k\to i}(t) > 0$, we obtain $\bar{v}^a_{\gamma\to i}(t) > -p_{ii} - \sum_{k \in \mathcal{N}(i)\backslash j,\gamma} \bar{v}^a_{k\to i}(t)$. Applying $\bar{\mathbf{v}}^a(t) \leq \mathbf{0}$ shown in (18) into this inequality, it can be inferred that $\bar{v}^a_{\gamma\to i}(t) > -p_{ii} - \sum_{k \in \mathcal{N}(i)\backslash j,\gamma} \bar{v}^a_{k\to i}(t) \geq -p_{ii}$, that is, $\bar{v}^a_{\gamma\to i}(t) > -p_{ii}$ for all $(\gamma, i) \in \mathcal{E}$. Combining with $\mathbf{v}^a(t) \geq \bar{\mathbf{v}}^a(t)$ shown in (18), for all $(i,j) \in \mathcal{E}$, we have

$$v^a_{i\to j}(t) > -p_{jj}. \tag{19}$$

It can be seen from (18) and (19) that $\mathbf{v}^a(t)$ is a monotonically non-increasing and lower bounded sequence. Thus, $\mathbf{v}^a(t)$ must converge to a vector $\mathbf{v}^{a*}$, that is, $\mathbf{v}^{a*} = \mathbf{g}(\mathbf{v}^{a*})$.

On the other hand, substituting (12) into (19) gives $-\frac{p^2_{ij}}{p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} v^a_{k\to i}(t)} > -p_{jj}$. Due to $\bar{\mathbf{v}}^a(t) \in \mathcal{W}$, it can be inferred from P1) and (18) that $\mathbf{v}^a(t) \in \mathcal{W}$, or equivalently $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} v^a_{k\to i}(t) > 0$. Together with the fact $p_{jj} > 0$, we obtain $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} v^a_{k\to i}(t) > \frac{p^2_{ij}}{p_{jj}}$. Since $\mathbf{v}^a(t)$ converges to $\mathbf{v}^{a*}$, taking the limit on both sides of the inequality gives $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} v^{a*}_{k\to i} \geq \frac{p^2_{ij}}{p_{jj}}$. Hence, $p_{ii} + \sum_{k \in \mathcal{N}(i)\backslash j} v^{a*}_{k\to i} > 0$. From the definition of $\mathcal{W}$ in (15), we have $\mathbf{v}^{a*} \in \mathcal{W}$. Combining with $\mathbf{v}^{a*} = \mathbf{g}(\mathbf{v}^{a*})$, according to the definition of $\mathcal{S}_1$ in (17), it is clear that $\mathbf{v}^{a*} \in \mathcal{S}_1$. This contradicts with the prerequisite $\mathcal{S}_1 = \emptyset$. Therefore, $\mathcal{S}_1$ cannot be $\emptyset$. ■

IV. Necessary and Sufficient Convergence Condition of Message Parameter $\mathbf{v}^a(t)$

According to Theorem 1, Gaussian form of messages may only be maintained for some $\mathbf{v}^a(0)$. Thus, the choice of initialization $\mathbf{v}^a(0)$ also plays an important role in the convergence of $\mathbf{v}^a(t)$.

A. Synchronous Scheduling

In this section, we will investigate the convergence condition of $\mathbf{v}^a(t)$ under synchronous scheduling, which means that $\mathbf{v}^a(t)$ is updated as $\mathbf{v}^a(t+1) = \mathbf{g}(\mathbf{v}^a(t))$. The necessary and sufficient convergence condition under nonnegative initialization set is derived first. Then, the convergence condition is proved to hold under a more general initialization set.

Lemma 2: $\mathbf{v}^a(t)$ converges to the same point $\mathbf{v}^{a*} \in \mathcal{S}_1$ for all $\mathbf{v}^a(0) \geq \mathbf{0}$ if and only if $\mathcal{S}_1 \neq \emptyset$.

Proof: See Appendix A. ■

Obviously, the initialization $\mathbf{v}^a(0) = \mathbf{0}$ used by the convergence condition in [24] is implied by the proposed initialization set $\mathbf{v}^a(0) \geq \mathbf{0}$ in Lemma 2. In the following, we show that the initialization set could be further expanded to the set

$$\begin{aligned}\mathcal{A} \triangleq\ & \{\mathbf{w} \geq \mathbf{0}\} \cup \{\mathbf{w} \geq \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\} \\ & \times \cup \left\{\mathbf{w} \geq \mathbf{w}_0 \,\middle|\, \mathbf{w}_0 \in \mathcal{S}_1 \text{ and } \mathbf{w}_0 = \lim_{t\to\infty} \mathbf{g}^{(t)}(\mathbf{0})\right\},\end{aligned} \tag{20}$$

where $\mathrm{int}(\mathcal{S}_1)$ means the interior of $\mathcal{S}_1$. Notice that the set $\mathcal{A}$ consists of the union of three parts.

- The first part $\{\mathbf{w} \geq \mathbf{0}\}$ corresponds to the initialization set given in Lemma 2.
- For the case $\mathcal{S}_1 \neq \emptyset$ but $\mathrm{int}(\mathcal{S}_1) = \emptyset$, the second part is empty. Also, according to Lemma 2, we have $\lim_{t\to\infty} \mathbf{g}^{(t)}(\mathbf{0}) = \mathbf{v}^{a*} \in \mathcal{S}_1$. Thus, the third part of set $\mathcal{A}$ reduces to $\{\mathbf{w} \geq \mathbf{v}^{a*}\}$. Due to the fact that $\mathbf{v}^{a*} \in \mathcal{S}_1$ implies $\mathbf{v}^{a*} < \mathbf{0}$ given by P5), we have $\{\mathbf{w} \geq \mathbf{0}\} \subsetneq \{\mathbf{w} \geq \mathbf{v}^{a*}\}$.

- For the case $\mathrm{int}(\mathcal{S}_1) \neq \emptyset$, according to (63), we have $\mathbf{w}_0 \leq \mathbf{v}^{a*}$ for any $\mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)$. This implies $\{\mathbf{w} \geq \mathbf{v}^{a*}\} \subseteq \{\mathbf{w} \geq \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\}$. Obviously, in this case, we also have $\{\mathbf{w} \geq \mathbf{0}\} \subsetneq \{\mathbf{w} \geq \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\}$.

Therefore, if $\mathcal{S}_1 \neq \emptyset$, the set $\mathcal{A}$ is always larger than the set $\{\mathbf{w} \geq \mathbf{0}\}$.

Theorem 2: $\mathbf{v}^a(t)$ converges to the same point $\mathbf{v}^{a*}$ for all $\mathbf{v}^a(0) \in \mathcal{A}$ if and only if $\mathcal{S}_1 \neq \emptyset$.

Proof: The first part $\{\mathbf{w} \geq \mathbf{0}\}$ of $\mathcal{A}$ corresponds to the case covered in Lemma 2, thus has already been established. To prove the sufficiency, we consider the case $\mathrm{int}(\mathcal{S}_1) \neq \emptyset$ and the case $\mathrm{int}(\mathcal{S}_1) = \emptyset$ but $\mathcal{S}_1 \neq \emptyset$, respectively.

For the case $\mathrm{int}(\mathcal{S}_1) \neq \emptyset$, it is known from (20) that $\mathcal{A} = \{\mathbf{w} \geq \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\}$ in this case. To prove the theorem, we will first prove that $\bar{\mathbf{v}}^a(t)$ converges to the same point $\mathbf{v}^{a*}$ for all $\bar{\mathbf{v}}^a(0) \in \mathrm{int}(\mathcal{S}_1)$. Then, notice that for any $\mathbf{v}^a(0) \in \{\mathbf{w} \geq \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\}$, it can be upper bounded by a point from $\{\mathbf{w} \geq \mathbf{0}\}$ and lower bounded by a point from $\mathrm{int}(\mathcal{S}_1)$. Because both the upper bound and lower bound converge to $\mathbf{v}^{a*}$, we can easily extend the convergence to the much larger set $\mathbf{v}^a(0) \in \{\mathbf{w} \geq \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\}$.

First, we prove that $\bar{\mathbf{v}}^a(t)$ converges to $\mathbf{v}^{a*}$ for all $\bar{\mathbf{v}}^a(0) \in \mathrm{int}(\mathcal{S}_1)$. To do this, we establish the following three facts first.

1) For any $\bar{\mathbf{v}}^a(0) \in \mathrm{int}(\mathcal{S}_1)$, according to P6), we have $\bar{\mathbf{v}}^a(t) \leq \bar{\mathbf{v}}^a(t+1)$ and $\bar{\mathbf{v}}^a(t) \in \mathcal{S}_1$. With P5), we also have $\bar{\mathbf{v}}^a(t) < \mathbf{0}$. It can be seen that $\bar{\mathbf{v}}^a(t)$ is a monotonically non-decreasing and upper bounded sequence, thus $\bar{\mathbf{v}}^a(t)$ converges to a fixed point, which is denoted as $\bar{\mathbf{v}}^{a*}$. On the other hand, let $\mathbf{v}^a(0) \geq \mathbf{0}$ correspond to a point in the first part of $\mathcal{A}$, due to $\bar{\mathbf{v}}^a(0) < \mathbf{0}$, we have $\bar{\mathbf{v}}^a(0) < \mathbf{v}^a(0)$. By using P2), applying $\mathbf{g}(\cdot)$ to $\bar{\mathbf{v}}^a(0) < \mathbf{v}^a(0)$ for t times gives $\bar{\mathbf{v}}^a(t) \leq \mathbf{v}^a(t)$. Since $\bar{\mathbf{v}}^a(t)$ and $\mathbf{v}^a(t)$ converge to $\bar{\mathbf{v}}^{a*}$ and $\mathbf{v}^{a*}$, respectively, taking the limit on both sides of $\bar{\mathbf{v}}^a(t) \leq \mathbf{v}^a(t)$ gives $\bar{\mathbf{v}}^{a*} \leq \mathbf{v}^{a*}$. Due to $\bar{\mathbf{v}}^a(0) \in \mathrm{int}(\mathcal{S}_1)$, the definition of $\mathcal{S}_1$ in (17) implies that $\bar{\mathbf{v}}^a(0) < \mathbf{g}(\bar{\mathbf{v}}^a(0))$, that is, $\bar{\mathbf{v}}^a(0) < \bar{\mathbf{v}}^a(1)$. Combining with $\bar{\mathbf{v}}^a(t) \leq \bar{\mathbf{v}}^a(t+1)$ established above, we obtain $\bar{\mathbf{v}}^a(0) < \bar{\mathbf{v}}^{a*}$. Together with the fact $\bar{\mathbf{v}}^{a*} \leq \mathbf{v}^{a*}$, we obtain

$$\bar{\mathbf{v}}^a(0) < \bar{\mathbf{v}}^{a*} \leq \mathbf{v}^{a*}. \tag{21}$$

2) For a $\bar{\mathbf{v}}^a(0) \in \mathrm{int}(\mathcal{S}_1)$, construct a vector

$$\hat{\mathbf{v}}^a(0) = \lambda \bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*}, \tag{22}$$

where $\lambda \in (0,1)$. Since $g_{ij}(\mathbf{w})$ is a concave function over $\mathbf{w} \in \mathcal{S}_1$ and $\{\bar{\mathbf{v}}^a(0), \mathbf{v}^{a*}\} \in \mathcal{S}_1$, we have $\mathbf{g}(\lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*}) \geq \lambda\mathbf{g}(\bar{\mathbf{v}}^a(0)) + (1-\lambda)\mathbf{g}(\mathbf{v}^{a*})$, or equivalently $\mathbf{g}(\lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*}) \geq \lambda\bar{\mathbf{v}}^a(1) + (1-\lambda)\mathbf{v}^{a*}$. According to P2), applying $\mathbf{g}(\cdot)$ to this inequality gives $\mathbf{g}^{(2)}(\lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*}) \geq \mathbf{g}(\lambda\bar{\mathbf{v}}^a(1) + (1-\lambda)\mathbf{v}^{a*})$. Because of the concave property of $g_{ij}(\mathbf{w})$, it is known that $\mathbf{g}(\lambda\bar{\mathbf{v}}^a(1) + (1-\lambda)\mathbf{v}^{a*}) \geq \lambda\bar{\mathbf{v}}^a(2) + (1-\lambda)\mathbf{v}^{a*}$. Thus, we have $\mathbf{g}^{(2)}(\lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*}) \geq \lambda\bar{\mathbf{v}}^a(2) + (1-\lambda)\mathbf{v}^{a*}$. By induction, it can be inferred that $\mathbf{g}^{(t)}(\lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*}) \geq \lambda\bar{\mathbf{v}}^a(t) + (1-\lambda)\mathbf{v}^{a*}$ for all $t \geq 0$. With $\hat{\mathbf{v}}^a(0) = \lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*}$ given by (22), we obtain

$$\hat{\mathbf{v}}^a(t) \geq \lambda\bar{\mathbf{v}}^a(t) + (1-\lambda)\mathbf{v}^{a*}, \tag{23}$$

where $\hat{\mathbf{v}}^a(t) \triangleq \mathbf{g}^{(t)}(\hat{\mathbf{v}}^a(0))$. Since $\{\bar{\mathbf{v}}^a(0), \mathbf{v}^{a*}\} \in \mathcal{S}_1$ and $\mathcal{S}_1$ is a convex set as indicated in P4), we have $\hat{\mathbf{v}}^a(0) = \lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*} \in \mathcal{S}_1$. According to P6), $\hat{\mathbf{v}}^a(t) \triangleq \mathbf{g}^{(t)}(\hat{\mathbf{v}}^a(0))$ is a monotonically non-decreasing sequence. Furthermore, according to P5), $\hat{\mathbf{v}}^a(t)$ is also upper bounded by $\mathbf{0}$. Thus, $\hat{\mathbf{v}}^a(t)$ converges to a fixed point, which is denoted as $\hat{\mathbf{v}}^{a*}$. Taking limits on both sides of (23) gives

$$\begin{aligned}\hat{\mathbf{v}}^{a*} &\geq \lambda\bar{\mathbf{v}}^{a*} + (1-\lambda)\mathbf{v}^{a*} \\ &= \bar{\mathbf{v}}^{a*} + (1-\lambda)(\mathbf{v}^{a*} - \bar{\mathbf{v}}^{a*}).\end{aligned} \tag{24}$$

3) From $\bar{\mathbf{v}}^a(0) < \bar{\mathbf{v}}^{a*} \leq \mathbf{v}^{a*}$ given by (21), we have $\bar{\mathbf{v}}^{a*} - \bar{\mathbf{v}}^a(0) > \mathbf{0}$ and $\mathbf{v}^{a*} - \bar{\mathbf{v}}^a(0) > \mathbf{0}$. If $\lambda \in (0,1)$ and is close to 1, the relation $(1-\lambda)(\mathbf{v}^{a*} - \bar{\mathbf{v}}^a(0)) < \bar{\mathbf{v}}^{a*} - \bar{\mathbf{v}}^a(0)$ always holds, that is,

$$\exists\, \lambda \in (0,1), \text{ s.t. } (1-\lambda)(\mathbf{v}^{a*} - \bar{\mathbf{v}}^a(0)) < \bar{\mathbf{v}}^{a*} - \bar{\mathbf{v}}^a(0). \tag{25}$$

Rearranging the terms in (25) gives $\exists\, \lambda \in (0,1)$, s.t. $\lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*} < \bar{\mathbf{v}}^{a*}$. Due to $\lambda\bar{\mathbf{v}}^a(0) + (1-\lambda)\mathbf{v}^{a*} = \hat{\mathbf{v}}^a(0)$ given in (22), thus (25) implies that there always exists a $\lambda \in (0,1)$ such that $\hat{\mathbf{v}}^a(0) < \bar{\mathbf{v}}^{a*}$. Now, applying $\mathbf{g}(\cdot)$ to $\hat{\mathbf{v}}^a(0) < \bar{\mathbf{v}}^{a*}$ for t times, from P2), we obtain $\mathbf{g}^{(t)}(\hat{\mathbf{v}}^a(0)) \leq \mathbf{g}^{(t)}(\bar{\mathbf{v}}^{a*})$, or equivalently $\hat{\mathbf{v}}^a(t) \leq \bar{\mathbf{v}}^{a*}$. Taking the limit on both sides of $\hat{\mathbf{v}}^a(t) \leq \bar{\mathbf{v}}^{a*}$, it can be inferred that

$$\exists\, \lambda \in (0,1), \text{ s.t. } \hat{\mathbf{v}}^{a*} \leq \bar{\mathbf{v}}^{a*}. \tag{26}$$

Now, we make use of (21), (24) and (26) to prove that the convergence limit $\bar{\mathbf{v}}^{a*}$ is identical to the convergence limit $\mathbf{v}^{a*}$. Combining (21) and (26) gives

$$\exists\, \lambda \in (0,1), \text{ s.t. } \hat{\mathbf{v}}^{a*} \leq \bar{\mathbf{v}}^{a*} \leq \mathbf{v}^{a*}. \tag{27}$$

From the second inequality of (27), we have $(1-\lambda)(\mathbf{v}^{a*} - \bar{\mathbf{v}}^{a*}) \geq \mathbf{0}$. Substituting this relation into (24), we can infer that $\hat{\mathbf{v}}^{a*} \geq \bar{\mathbf{v}}^{a*}$. Comparing this result to the first inequality of (27), we obtain $\hat{\mathbf{v}}^{a*} = \bar{\mathbf{v}}^{a*}$. Due to $\hat{\mathbf{v}}^{a*} = \bar{\mathbf{v}}^{a*}$, (24) becomes $(1-\lambda)(\mathbf{v}^{a*} - \bar{\mathbf{v}}^{a*}) \leq \mathbf{0}$. For $\lambda \in (0,1)$, the relation is equivalent to $\mathbf{v}^{a*} \leq \bar{\mathbf{v}}^{a*}$. Combining it with the second inequality in (27), we obtain

$$\bar{\mathbf{v}}^{a*} = \mathbf{v}^{a*}. \tag{28}$$

Next, we will extend the initialization set to the whole set $\mathcal{A} = \{\mathbf{w} \geq \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\}$. For any $\bar{\mathbf{v}}^a(0) \in \mathcal{A}$, we can always find a $\mathbf{v}_l^a(0) \in \mathrm{int}(\mathcal{S}_1)$ and a $\mathbf{v}_u^a(0) \geq \mathbf{0}$ such that $\mathbf{v}_l^a(0) \leq \bar{\mathbf{v}}^a(0) \leq \mathbf{v}_u^a(0)$. According to P2), applying $\mathbf{g}(\cdot)$ to this inequality repeatedly gives $\mathbf{g}^{(t)}(\mathbf{v}_l^a(0)) \leq \mathbf{g}^{(t)}(\bar{\mathbf{v}}^a(0)) \leq \mathbf{g}^{(t)}(\mathbf{v}_u^a(0))$. Since $\mathbf{v}_l^a(t) = \mathbf{g}^{(t)}(\mathbf{v}_l^a(0))$ and $\mathbf{v}_u^a(t) = \mathbf{g}^{(t)}(\mathbf{v}_u^a(0))$ both converge to $\mathbf{v}^{a*}$, thus $\bar{\mathbf{v}}^a(t)$ converges to $\mathbf{v}^{a*}$, too.

Now, consider the case $\mathrm{int}(\mathcal{S}_1) = \emptyset$ but $\mathcal{S}_1 \neq \emptyset$. In this case, according to the discussion after (20), $\mathcal{A} = \{\mathbf{w} \geq \mathbf{v}^{a*}\}$. For any $\bar{\mathbf{v}}^a(0) \in \mathcal{A}$, we can always find a $\mathbf{v}_u^a(0) \in \{\mathbf{w} \geq \mathbf{0}\}$ such that $\mathbf{v}^{a*} \leq \bar{\mathbf{v}}^a(0) \leq \mathbf{v}_u^a(0)$. Applying $\mathbf{g}(\cdot)$ to this inequality for t times, from P2), we obtain $\mathbf{v}^{a*} \leq \mathbf{g}^{(t)}(\bar{\mathbf{v}}^a(0)) \leq \mathbf{g}^{(t)}(\mathbf{v}_u^a(0))$. Since $\mathbf{v}_u^a(t)$ converges to $\mathbf{v}^{a*}$, thus $\bar{\mathbf{v}}^a(t)$ also converges to $\mathbf{v}^{a*}$.

On the other hand, if $\mathcal{S}_1 = \emptyset$, according to Theorem 1, the messages of Gaussian BP passed in factor graph lose the Gaussian form, thus the parameters of messages $\mathbf{v}^a(t)$ cannot converge in this case. ■

B. Asynchronous Scheduling

In this section, convergence conditions under asynchronous scheduling schemes will be investigated. Modified from the synchronous updating equation in (12), arriving precision in asynchronous scheduling is updated as

$$v^a_{i\to j}(t+1) = -\frac{p_{ij}^2}{p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} v^a_{k\to i}(\tau^a_{k\to i}(t))}, \quad \forall\, t \in \mathcal{T}_{i\to j} \tag{29}$$

where $\mathcal{T}_{i\to j}$ is the set of time instants at which $v^a_{i\to j}(t)$ are updated; $0 \le \tau^a_{k\to i}(t) \le t$ is the last updated time instant of $v^a_{k\to i}(t)$ at time t. In this paper, we only consider the totally asynchronous scheduling defined as follows.

Definition 1. (Totally Asynchronous Scheduling) [27]: The sets $\mathcal{T}_{i\to j}$ are infinite, and $\tau^a_{i\to j}(t)$ satisfies $0 \le \tau^a_{i\to j}(t) \le t$ and $\lim_{t\to\infty} \tau^a_{i\to j}(t) = \infty$ for all $(i,j) \in \mathcal{E}$.

In order to prove the convergence of $\mathbf{v}^a(t)$ under asynchronous scheduling given by Definition 1, we need to find a sequence of nonempty sets $\{\mathcal{U}(m)\}$ for $m \ge 0$ such that the following two sufficient conditions are satisfied [27]:

- Synchronous Convergence Condition: the sequence of sets satisfy 1) $\mathcal{U}(m+1) \subseteq \mathcal{U}(m)$; 2) $\mathbf{g}(\mathbf{w}) \in \mathcal{U}(m+1)$ for $\forall\, \mathbf{w} \in \mathcal{U}(m)$; 3) $\mathcal{U}(m)$ converges to a fixed point;
- Box Condition: $\mathcal{U}(m)$ can be written as product of subsets of individual components $\mathcal{U}_{i\to j}(m)$ for all $(i,j) \in \mathcal{E}$.

While the formal proof of asynchronous convergence under these conditions are provided in [27], we give a brief intuitive explanation below. Suppose when $t \ge t_m$, $v^a_{i\to j}(t) \in \mathcal{U}_{i\to j}(m)$ for all $(i,j) \in \mathcal{E}$, or expressed using the Box Condition $\mathbf{v}^a(t) \in \mathcal{U}(m)$ for $t \ge t_m$. Now, consider the updating of an individual component $v^a_{i\to j}(t)$. Due to $\lim_{t\to\infty} \tau^a_{k\to i}(t) = \infty$, we can always find a $t^{(i\to j)}_{m+1} > t_m$ such that when $t \ge t^{(i\to j)}_{m+1}$, we have $\tau^a_{k\to i}(t) \ge t_m$ and thereby $v^a_{k\to i}(\tau^a_{k\to i}(t)) \in \mathcal{U}_{k\to i}(m)$ for all $k \in \mathcal{N}(i)\backslash j$. Putting $v^a_{k\to i}(\tau^a_{k\to i}(t))$ into function $g_{ij}(\cdot)$ in (29) and applying condition 2) in Synchronous Convergence Condition, we have $v^a_{i\to j}(t+1) \in \mathcal{U}_{i\to j}(m+1)$ for $t \ge t^{(i\to j)}_{m+1}$. Thus, by choosing $t_{m+1} = \max_{(i,j)\in\mathcal{E}} t^{(i\to j)}_{m+1}$, we have when $t \ge t_{m+1}$, $v^a_{i\to j}(t) \in \mathcal{U}_{i\to j}(m+1)$ for all $(i,j) \in \mathcal{E}$. Further due to conditions 1) and 3) in Synchronous Convergence Condition, we know that $\mathbf{v}^a(t)$ will gradually converge to a fixed point.

Theorem 3: $\mathbf{v}^a(t)$ converges for any $\mathbf{v}^a(0) \in \mathcal{A}$ and all choices of schedulings if and only if $\mathcal{S}_1 \ne \emptyset$.

Proof: We first prove the sufficient condition. Since $\mathcal{S}_1 \ne \emptyset$, as given in the proof of Theorem 2, for any $\mathbf{v}^a(0) \in \mathcal{A}$, there always exist an element $\bar{\mathbf{v}}^a(0) \in \mathrm{int}(\mathcal{S}_1) \cup \{\mathbf{v}^{a*}\}$ and $\hat{\mathbf{v}}^a(0) \ge \mathbf{0}$ such that $\bar{\mathbf{v}}^a(0) \le \mathbf{v}^a(0) \le \hat{\mathbf{v}}^a(0)$. Construct a sequence of sets as

$$\mathcal{U}(m) = \{\mathbf{w} \,|\, \bar{\mathbf{v}}^a(m) \le \mathbf{w} \le \hat{\mathbf{v}}^a(m)\}. \tag{30}$$

Since $\bar{\mathbf{v}}^a(0) \in \mathrm{int}(\mathcal{S}_1) \cup \{\mathbf{v}^{a*}\}$, we have $\bar{\mathbf{v}}^a(m) \le \bar{\mathbf{v}}^a(m+1)$ according to P6). Further, for any $\hat{\mathbf{v}}^a(0) \ge \mathbf{0}$, putting $\hat{\mathbf{v}}^a(0)$ into (14), we have $\hat{\mathbf{v}}^a(1) < \mathbf{0}$, thereby $\hat{\mathbf{v}}^a(1) < \hat{\mathbf{v}}^a(0)$. Since $\hat{\mathbf{v}}^a(m)$ converges according to Theorem 2, then $\hat{\mathbf{v}}^a(m) \in \mathcal{W}$ for all $m \ge 0$. Thus, according to P2), by applying $\mathbf{g}(\cdot)$ to $\hat{\mathbf{v}}^a(1) < \hat{\mathbf{v}}^a(0)$ and by induction, we have $\hat{\mathbf{v}}^a(m+1) \le \hat{\mathbf{v}}^a(m)$. By exploiting $\bar{\mathbf{v}}^a(m) \le \bar{\mathbf{v}}^a(m+1)$, $\hat{\mathbf{v}}^a(m+1) \le \hat{\mathbf{v}}^a(m)$ and the definition of $\mathcal{U}(m)$ in (30), we have

$$\mathcal{U}(m+1) \subseteq \mathcal{U}(m). \tag{31}$$

Now, for any $\mathbf{w} \in \mathcal{U}(m)$, by applying $\mathbf{g}(\cdot)$ to $\bar{\mathbf{v}}^a(m) \le \mathbf{w} \le \hat{\mathbf{v}}^a(m)$, we obtain $\bar{\mathbf{v}}^a(m+1) \le \mathbf{g}(\mathbf{w}) \le \hat{\mathbf{v}}^a(m+1)$. Thus, from the definition of $\mathcal{U}(m)$ in (30), we have the relation

$$\mathbf{g}(\mathbf{w}) \in \mathcal{U}(m+1), \quad \forall\, \mathbf{w} \in \mathcal{U}(m). \tag{32}$$

Furthermore, according to Theorem 2, we have $\lim_{m\to\infty} \bar{\mathbf{v}}^a(m) = \lim_{m\to\infty} \hat{\mathbf{v}}^a(m) = \mathbf{v}^{a*}$, hence $\mathcal{U}(m)$ will converge to the fixed point $\mathbf{v}^{a*}$. Thus, the set $\mathcal{U}(m)$ satisfies the Synchronous Convergence Condition [27]. Furthermore, according to the definition of $\mathcal{U}(m)$ in (30), $\mathbf{v}^a(m) \in \mathcal{U}(m)$ means $\bar{v}^a_{i\to j}(m) \le v^a_{i\to j}(m) \le \hat{v}^a_{i\to j}(m)$ for all $(i,j) \in \mathcal{E}$, that is, $v^a_{i\to j}(m) \in \mathcal{U}_{i\to j}(m)$ with $\mathcal{U}_{i\to j}(m) = \{w_{ij} \,|\, \bar{v}^a_{i\to j}(m) \le w_{ij} \le \hat{v}^a_{i\to j}(m)\}$. Thus, $\mathcal{U}(m)$ can be represented in form of product of subsets of individual components $\mathcal{U}_{i\to j}(m)$ for all $(i,j) \in \mathcal{E}$, and thereby the Box Condition is satisfied. Hence, the sufficient condition is proved.

On the other hand, if $\mathcal{S}_1 = \emptyset$, according to Theorem 1, the messages of Gaussian BP passed in factor graph lose the Gaussian form, thus the parameters of messages $\mathbf{v}^a(t)$ cannot converge in this case. ■

From Theorems 2 and 3, it can be seen that the convergence condition of $\mathbf{v}^a(t)$ under asynchronous scheduling is the same as that under synchronous scheduling.

V. Necessary and Sufficient Convergence Condition of Belief Variance $\sigma_i^2(t)$

Although the convergence conditions for $\mathbf{v}^a(t)$ are derived in the last section, what we are really interested in is the convergence condition for the belief variance $\sigma_i^2(t)$. According to (11), the variance $\sigma_i^2(t)$ and $\mathbf{v}^a(t)$ are related by $\sigma_i^2(t) = \frac{1}{p_{ii}+\sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)}$, or equivalently $\frac{1}{\sigma_i^2(t)} = p_{ii} + \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)$.

A. Convergence Condition

In order to investigate the convergence condition of $\sigma_i^2(t)$, we present the following lemma first.

Lemma 3: If $\mathcal{S}_1 \ne \emptyset$ and $\mathbf{v}^a(0) \in \mathcal{A}$, then $\lim_{t\to\infty} \frac{1}{\sigma_i^2(t)} > 0$ for all $i \in \mathcal{V}$ or $\lim_{t\to\infty} \frac{1}{\sigma_i^2(t)} = 0$ for all $i \in \mathcal{V}$.

Proof: In the proof, we first consider the special case under tree-structured factor graph as well as the impact of initializations. Then, for the case of loopy factor graph, we prove that there exists at least one node $\gamma \in \mathcal{V}$ such that $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} \ge 0$. At last, we demonstrate that either $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} > 0$ for all $\gamma \in \mathcal{V}$ or $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} = 0$ for all $\gamma \in \mathcal{V}$ will hold.

First, if the factor graph is of tree structure, according to properties of BP in a tree structured factor graph, it is known that $\lim_{t\to\infty} \sigma_\gamma^2(t) = [\mathbf{P}^{-1}]_{\gamma\gamma}$ [1]. Due to $\mathbf{P}^{-1} \succ 0$, we have $\lim_{t\to\infty} \sigma_\gamma^2(t) = [\mathbf{P}^{-1}]_{\gamma\gamma} > 0$, and thereby $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} > 0$ for

all $\gamma \in \mathcal{V}$. In the following, we focus on the factor graph containing loops.

If $\mathcal{S}_1 \neq \emptyset$, according to Theorem 2, $\mathbf{v}^a(t)$ converges to the same point for any initialization $\mathbf{v}^a(0) \in \mathcal{A}$. Due to $\frac{1}{\sigma_i^2(t)} = p_{ii} + \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)$, then $\lim_{t\to\infty} \frac{1}{\sigma_i^2(t)}$ always exists and is identical for all $\mathbf{v}^a(0) \in \mathcal{A}$. Therefore, we only need to consider a specific initialization, e.g., $\mathbf{v}^a(0) = \mathbf{0}$.

Second, we prove $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} \geq 0$ for at least one node $\gamma \in \mathcal{V}$. Due to the factor graph containing loops, we can always find a node γ such that $|\mathcal{N}(\gamma)| \geq 2$, where $|\cdot|$ means the cardinality of a set. Denote nodes i, j as two neighbors of node γ, that is, $i, j \in \mathcal{N}(\gamma)$. Since $\mathbf{v}^a(0) = \mathbf{0}$, from (63), we have $\mathbf{v}^{a*} \leq \mathbf{v}^a(t)$. Applying this relation to $\frac{1}{\sigma_\gamma^2(t)} = p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus i} v^a_{k\to\gamma}(t) + v^a_{i\to\gamma}(t)$ gives

$$\frac{1}{\sigma_\gamma^2(t)} \geq p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus i} v^a_{k\to\gamma}(t) + v^{a*}_{i\to\gamma}. \tag{33}$$

By further using $\mathbf{v}^{a*} \leq \mathbf{v}^a(t)$, we can easily obtain

$$\begin{aligned} &p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus i} v^a_{k\to\gamma}(t) + v^{a*}_{i\to\gamma} \\ &= p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus i,j} v^a_{k\to\gamma}(t) + v^{a*}_{i\to\gamma} + v^a_{j\to\gamma}(t) \\ &\geq p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma} + v^a_{j\to\gamma}(t). \end{aligned} \tag{34}$$

Notice that $p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma} + v^a_{j\to\gamma}(t)$ in (34) can be written as (35),

$$\begin{aligned} &p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma} + v^a_{j\to\gamma}(t) \\ &= \frac{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma}}{p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1)} \\ &\quad \times \Bigg(p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1) \\ &\quad + \frac{p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1)}{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma}} v^a_{j\to\gamma}(t) \Bigg) \\ &\overset{a}{=} \frac{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma}}{p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1)} \\ &\quad \times \Bigg(p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1) \\ &\quad - \frac{p_{\gamma j}^2}{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma}} \Bigg) \\ &\overset{b}{=} \frac{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma}}{p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1)} \\ &\quad \times \left(p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1) + v^{a*}_{\gamma\to j} \right), \end{aligned} \tag{35}$$

where the equality $\overset{a}{=}$ holds because $(p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1)) v^a_{j\to\gamma}(t) = -p_{\gamma j}^2$ from (12); and $\overset{b}{=}$ is obtained by using the limiting form of (12) $v^{a*}_{\gamma\to j} = -\frac{p_{\gamma j}^2}{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma}}$. Combining (34) and (35) gives

$$\begin{aligned} &p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus i} v^a_{k\to\gamma}(t) + v^{a*}_{i\to\gamma} \\ &\geq \frac{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v^{a*}_{k\to\gamma}}{p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1)} \\ &\quad \times \left(p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v^a_{k\to j}(t-1) + v^{a*}_{\gamma\to j} \right). \end{aligned} \tag{36}$$

Due to $\mathbf{v}^{a*} \in \mathcal{S}_1$ from Lemma 2, then $\mathbf{v}^{a*} \in \mathcal{W}$. From the definition of $\mathcal{W}$ in (15), for all $(\nu_0, \nu_1) \in \mathcal{E}$, we have

$$p_{\nu_0\nu_0} + \sum_{k\in\mathcal{N}(\nu_0)\setminus\nu_1} v^{a*}_{k\to\nu_0} > 0. \tag{37}$$

By using (37) and $\mathbf{v}^a(t) \geq \mathbf{v}^{a*}$ implied by (63), we can infer $p_{\nu_0\nu_0} + \sum_{k\in\mathcal{N}(\nu_0)\setminus\nu_1} v^a_{k\to\nu_0}(t) > 0$ for all $(\nu_0, \nu_1) \in \mathcal{E}$, and thereby

$$\frac{p_{\nu_0\nu_0} + \sum_{k\in\mathcal{N}(\nu_0)\setminus\nu_1} v^{a*}_{k\to\nu_0}}{p_{\nu_1\nu_1} + \sum_{k\in\mathcal{N}(\nu_1)\setminus\nu_0} v^a_{k\to\nu_1}(t)} > 0. \tag{38}$$

Since the factor graph contains loops, there always exists a walk $(j_0, j_1, \ldots, j_{t-1}, j_t, i)$ with $j_t = \gamma$ such that all $(j_{\ell-1}, j_\ell) \in \mathcal{E}$ and $|\mathcal{N}(j_\ell)| \geq 2$ for $\ell = 0, 1, \ldots, t$. Then, using (36) repeatedly on such a walk gives

$$\begin{aligned} &p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus i} v^a_{k\to\gamma}(t) + v^{a*}_{i\to\gamma} \\ &\geq \prod_{\ell=0}^{t-1} \frac{p_{j_{\ell+1}j_{\ell+1}} + \sum_{k\in\mathcal{N}(j_{\ell+1})\setminus j_\ell} v^{a*}_{k\to j_{\ell+1}}}{p_{j_\ell j_\ell} + \sum_{k\in\mathcal{N}(j_\ell)\setminus j_{\ell+1}} v^a_{k\to j_\ell}(\ell)} \\ &\quad \times \left(p_{j_0 j_0} + \sum_{k\in\mathcal{N}(j_0)\setminus j_1} v^a_{k\to j_0}(0) + v^{a*}_{j_1\to j_0} \right) \\ &= \left(p_{j_0 j_0} + v^{a*}_{j_1\to j_0} \right) \\ &\quad \times \prod_{\ell=0}^{t-1} \frac{p_{j_{\ell+1}j_{\ell+1}} + \sum_{k\in\mathcal{N}(j_{\ell+1})\setminus j_\ell} v^{a*}_{k\to j_{\ell+1}}}{p_{j_\ell j_\ell} + \sum_{k\in\mathcal{N}(j_\ell)\setminus j_{\ell+1}} v^a_{k\to j_\ell}(\ell)}, \end{aligned} \tag{39}$$

where the last equality holds due to $\mathbf{v}^a(0) = \mathbf{0}$. Due to $|\mathcal{N}(j_0)| \geq 2$, there exists a node ν such that $j_1 \in \mathcal{N}(j_0)\setminus\nu$. By exploiting the fact that $\mathbf{v}^{a*} < \mathbf{0}$ due to $\mathbf{v}^{a*} \in \mathcal{S}_1$, we obtain $p_{j_0 j_0} + v^{a*}_{j_1\to j_0} \geq p_{j_0 j_0} + \sum_{k\in\mathcal{N}(j_0)\setminus\nu} v^{a*}_{k\to j_0}$. Combining with (37), we can infer that

$$p_{j_0 j_0} + v^{a*}_{j_1\to j_0} > 0. \tag{40}$$

Putting (38) and (40) into (39), we obtain $p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus i} v^a_{k\to\gamma}(t) + v^{a*}_{i\to\gamma} > 0$. Applying this result to (33) gives $\frac{1}{\sigma_\gamma^2(t)} > 0$ for all $t \geq 0$, and thereby $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} \geq 0$.

At last, we will prove that $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} > 0$ for all $\gamma \in \mathcal{V}$ or $\lim_{t\to\infty} \frac{1}{\sigma_\gamma^2(t)} = 0$ for all $\gamma \in \mathcal{V}$. Notice that the relation in (35) holds for any $(\gamma, j) \in \mathcal{E}$. Thus, taking the limit on both sides of

(35) and recognizing that $\lim_{t\to\infty}\frac{1}{\sigma_\gamma^2(t)} = p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} v_{k\to\gamma}^{a*}$ and $\lim_{t\to\infty}\frac{1}{\sigma_j^2(t)} = p_{jj} + \sum_{k\in\mathcal{N}(j)} v_{k\to j}^{a*}$, we obtain

$$\lim_{t\to\infty}\frac{1}{\sigma_\gamma^2(t)} = \frac{p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v_{k\to\gamma}^{a*}}{p_{jj} + \sum_{k\in\mathcal{N}(j)\setminus\gamma} v_{k\to j}^{a*}} \cdot \lim_{t\to\infty}\frac{1}{\sigma_j^2(t)}. \quad (41)$$

Since $p_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus j} v_{k\to\gamma}^{a*} > 0$ for all $(\gamma, j) \in \mathcal{E}$ given in (37), according to (41), if $\lim_{t\to\infty}\frac{1}{\sigma_\gamma^2(t)} > 0$, then $\lim_{t\to\infty}\frac{1}{\sigma_j^2(t)} > 0$ for all nodes $j \in \mathcal{N}(\gamma)$, and vice versa. Similarly, if $\lim_{t\to\infty}\frac{1}{\sigma_\gamma^2(t)} = 0$, then $\lim_{t\to\infty}\frac{1}{\sigma_j^2(t)} = 0$ for all nodes $j \in \mathcal{N}(\gamma)$, and vice versa. Recall that we have proved there exists a node γ with $\lim_{t\to\infty}\frac{1}{\sigma_\gamma^2(t)} \geq 0$. Therefore, for a fully connected factor graph, either $\lim_{t\to\infty}\frac{1}{\sigma_i^2(t)} > 0$ for all $i \in \mathcal{V}$ or $\lim_{t\to\infty}\frac{1}{\sigma_i^2(t)} = 0$ for all $i \in \mathcal{V}$. ■

Obviously, if $\lim_{t\to\infty}\frac{1}{\sigma_i^2(t)} \neq 0$, then $\lim_{t\to\infty}\sigma_i^2(t)$ exists, and hence the convergence condition of $\sigma_i^2(t)$ is the same as that of $\frac{1}{\sigma_i^2(t)}$. However, Lemma 3 reveals that the scenario $\lim_{t\to\infty}\frac{1}{\sigma_i^2(t)} = 0$ cannot be excluded. Thus, we have the following theorem.

Theorem 4: $\sigma_i^2(t)$ with $i \in \mathcal{V}$ converges to the same positive value for all $\mathbf{v}^a(0) \in \mathcal{A}$ under synchronous and asynchronous schedulings if and only if $\mathcal{S} \neq \emptyset$, where

$$\mathcal{S} \triangleq \left\{ \mathbf{w} \,\middle|\, \mathbf{w} \in \mathcal{S}_1 \text{ and } p_{ii} + \sum_{k\in\mathcal{N}(i)} w_{ki} > 0, \forall\, i \in \mathcal{V} \right\}. \quad (42)$$

Proof: First, we prove if $\mathcal{S} \neq \emptyset$, then $\sigma_i^2(t)$ converges for any $\mathbf{v}^a(0) \in \mathcal{A}$. From the definition of $\mathcal{S}$ in (42), if $\mathcal{S} \neq \emptyset$, then $\mathcal{S}_1 \neq \emptyset$. According to Theorems 2 and 3, $\mathbf{v}^a(t)$ converges to $\mathbf{v}^{a*}$ for all $\mathbf{v}^a(0) \in \mathcal{A}$ under both synchronous and asynchronous schedulings. Thus, $\lim_{t\to\infty}\frac{1}{\sigma_i^2(t)} = p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t)$ will converge to $p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*}$. Due to $\mathcal{S} \neq \emptyset$, there exists a $\mathbf{w}$ such that $\mathbf{w} \in \mathcal{S}_1$ and $p_{ii} + \sum_{k\in\mathcal{N}(i)} w_{ki} > 0$ for all $i \in \mathcal{V}$. Due to $\mathbf{w} \in \mathcal{S}_1$, it can be inferred from (63) that $\mathbf{w} \leq \mathbf{v}^{a*}$. Putting this result into $p_{ii} + \sum_{k\in\mathcal{N}(i)} w_{ki} > 0$, we obtain $p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*} > 0$ and its inverse exists. Thus, if $\mathcal{S} \neq \emptyset$, the variance $\sigma_i^2(t) = \frac{1}{p_{ii}+\sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t)}$ converges to $\frac{1}{p_{ii}+\sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*}} > 0$.

Next, we prove by contradiction that if $\sigma_i^2(t)$ converges for $\mathbf{v}^a(0) \in \mathcal{A}$, then $\mathcal{S} \neq \emptyset$. Suppose that $\mathcal{S} = \emptyset$. This assumption contains two cases: 1) $\mathcal{S}_1 = \emptyset$; 2) $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \sum_{k\in\mathcal{N}(i)} w_{ki} \leq 0$ for some $i \in \mathcal{V}$. First, consider the case $\mathcal{S}_1 = \emptyset$. Due to $\mathcal{S}_1 = \emptyset$, according to Theorem 1, the messages of Gaussian BP passed in factor graph cannot maintain Gaussian form, hence $\mathbf{v}^a(t)$ becomes undefined. Therefore, the variance $\sigma_i^2(t)$ cannot converge in this case, which contradicts with the prerequisite that $\sigma_i^2(t)$ converges. Second, consider the case $\mathcal{S}_1 \neq \emptyset$. Due to $\mathcal{S}_1 \neq \emptyset$, according to Theorems 2 and 3, $\mathbf{v}^a(t)$ converges to $\mathbf{v}^{a*} \in \mathcal{S}_1$ for any $\mathbf{v}^a(0) \in \mathcal{A}$. According to Lemma 3, we further have $\lim_{t\to\infty}\frac{1}{\sigma^2(t)} = p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*} > 0$ for all $i \in \mathcal{V}$ or $\lim_{t\to\infty}\frac{1}{\sigma^2(t)} = p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*} = 0$ for all $i \in \mathcal{V}$. If $p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*} > 0$ for all $i \in \mathcal{V}$, combining with $\mathbf{v}^{a*} \in \mathcal{S}_1$, we can infer from the definition of $\mathcal{S}$ in (42) that $\mathbf{v}^{a*} \in \mathcal{S}$, which contradicts with the assumption $\mathcal{S} = \emptyset$. On the other hand, if $p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*} = 0$ for all $i \in \mathcal{V}$, obviously, the variance $\sigma_i^2(t) = \frac{1}{p_{ii}+\sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t)}$ cannot converge, which contradicts with the prerequisite that $\sigma_i^2(t)$ converges. In summary, we have proved that if $\sigma_i^2(t)$ converges for all $\mathbf{v}^a(0) \in \mathcal{A}$, then $\mathcal{S} = \emptyset$ cannot hold, or equivalently we must have $\mathcal{S} \neq \emptyset$. ■

Remark 1: Due to the belief mean $\mu_i(t) = \frac{h_i+\sum_{k\in\mathcal{N}(i)} \beta_{k\to i}^a(t)}{\sigma_i^2(t)}$, if the convergence condition of belief variance $\sigma_i^2(t)$ is already known, the convergence of $\mu_i(t)$ is determined by that of message parameters $\beta_{k\to i}^a(t)$. As pointed out in [24], $\beta_{k\to j}^a(t)$ is updated in accordance to a set of linear equations upon the convergence of belief variances. Thus, after deriving the convergence condition of $\sigma_i^2(t)$, we can then apply the theories from linear equations to analyze the convergence condition of belief mean $\mu_i(t)$. However, the convergence condition of belief mean $\mu_i(t)$ would be lengthy and will be investigated in the future.

B. Verification Using Semi-Definite Programming

Next, we show that whether $\mathcal{S}$ is $\emptyset$ can be determined by solving the following SDP problem [30]

$$\begin{aligned} \min_{\mathbf{w},\alpha} \quad & \alpha \\ \text{s.t.} \quad & \begin{bmatrix} p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki} & p_{ij} \\ p_{ij} & -w_{ij} \end{bmatrix} \succeq 0, \forall\,(i,j) \in \mathcal{E}; \\ & \alpha + p_{ii} + \sum_{k\in\mathcal{N}(i)} w_{ki} \geq 0, \qquad \forall\, i \in \mathcal{V}. \end{aligned} \quad (43)$$

The SDP problem (43) can be solved efficiently by existing softwares, such as CVX [31] and SeDuMi [32], etc.

Theorem 5: $\mathcal{S} \neq \emptyset$ if and only if the optimal solution α^* of (43) satisfies $\alpha^* < 0$.

Proof: First, notice that the SDP problem in (43) is equivalent to the following optimization problem

$$\begin{aligned} \min_{\mathbf{w},\alpha} \quad & \alpha \\ \text{s.t.} \quad & w_{ij} - g_{ij}(\mathbf{w}) \leq 0, \qquad \forall\,(i,j) \in \mathcal{E}; \\ & -p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki} \leq 0, \quad \forall\,(i,j) \in \mathcal{E}; \\ & -p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki} \leq \alpha, \qquad \forall\, i \in \mathcal{V}. \end{aligned} \quad (44)$$

If $\mathcal{S} \neq \emptyset$, according to definition of $\mathcal{S}$ in (42), there must exist a $\mathbf{w}$ such that $\mathbf{w} \leq \mathbf{g}(\mathbf{w})$, $\mathbf{w} \in \mathcal{W}$, and $p_{ii} + \sum_{k\in\mathcal{N}(i)} w_{ki} > 0$ for all $i \in \mathcal{V}$. Obviously, these three conditions are equivalent to $w_{ij} - g_{ij}(\mathbf{w}) \leq 0$, $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki} < 0$ for all $(i,j) \in \mathcal{E}$ and $-p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki} < 0$ for all $i \in \mathcal{V}$. Thus, by defining $\alpha = -p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki}$, it can be seen that $(\mathbf{w}, \alpha)$ satisfies the constraints in (44). Due to $-p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki} < 0$, thus we have $\alpha < 0$. Since α is a feasible solution of the minimization problem in (44), the optimal solution of (44) cannot be greater than α, hence it can be inferred that $\alpha^* < 0$.

Next, we prove the reverse also holds. If $(\mathbf{w}^*, \alpha^*)$ is the optimal solution of (44) with $\alpha^* < 0$, $(\mathbf{w}^*, \alpha^*)$ must satisfy the constraints in (44), thus the following three conditions hold: 1) $w_{ij}^* - g_{ij}(\mathbf{w}^*) \leq 0$; 2) $-p_{ii} -$

$\sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* \le 0$; 3) $-p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki}^* \le \alpha^*$. For the second constraint, 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 $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* = 0$ will never happen. Hence, we always have $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* < 0$. For the third constraint, due to $\alpha^* < 0$, it can be inferred that $-p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki}^* < 0$. Now, comparing $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* \le 0$, $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* < 0$ and $-p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki}^* < 0$ with the definition of set $\mathcal{S}$ in (42), we have $\mathbf{w}^* \in \mathcal{S}$, and hence $\mathcal{S} \neq \emptyset$. ■

Remark 2: Using the alternating direction method of multipliers (ADMM) [27], the SDP problem in (43) can be reformulated into N locally connected low-dimensional SDP sub-problems, thus can be solved distributively. Not only this avoids the gathering of information at a central processing unit, the complexity is also reduced from $\mathcal{O}((\sum_{i=1}^N |\mathcal{N}(i)|)^4)$ of directly solving SDP to $\mathcal{O}(\sum_{i=1}^N |\mathcal{N}(i)|^4)$ per iteration of using ADMM technique. Furthermore, since what we are really interested in is to know whether $\mathcal{S} \neq \emptyset$, ADMM can stop its updating immediately if an intermediary vector is found to be within $\mathcal{S}$. Thus, the required number of iterations can be reduced significantly. Because the derivation of ADMM is well-documented in [27], [33], we do not give the details here. It should be emphasized that despite of the low complexity of ADMM at each iteration, due to the unknown number of required iterations, it cannot be proved that the overall complexity is always lower than that of direct matrix inverse $\mathcal{O}(N^3)$.

Remark 3: As discussed after (20), we know that $\mathcal{A} = \{\mathbf{w} \ge \mathbf{w}_0 \,|\, \mathbf{w}_0 \in \mathrm{int}(\mathcal{S}_1)\}$ if $\mathrm{int}(\mathcal{S}_1) \neq \emptyset$. In particular, with a proof similar to that of Theorem 5, it can be shown that if the SDP

$$\begin{aligned} \min_{\mathbf{w},\alpha} \quad & \alpha \\ \text{s.t.} \quad & \begin{bmatrix} p_{ii} + \sum\limits_{k\in\mathcal{N}(i)\setminus j} w_{ki} & p_{ij} \\ p_{ij} & \alpha - w_{ij} \end{bmatrix} \succeq 0, \quad \forall (i,j) \in \mathcal{E} \end{aligned} \tag{45}$$

has a solution $\alpha^* < 0$, then the optimal $\mathbf{w}^*$ must be a point in $\mathrm{int}(\mathcal{S}_1)$.

VI. Relationship With the Condition Based on Computation Tree

In this section, we will establish the relation between the proposed convergence condition and the one proposed in [24], which is the best currently available. For the convergence condition in [24], the original PDF in (1) is first transformed into the normalized form

$$\tilde{f}(\tilde{\mathbf{x}}) \propto \exp\left\{-\frac{1}{2}\tilde{\mathbf{x}}^T\tilde{\mathbf{P}}\tilde{\mathbf{x}} + \tilde{\mathbf{h}}^T\tilde{\mathbf{x}}\right\}, \tag{46}$$

where $\tilde{\mathbf{x}} = \mathrm{diag}^{\frac{1}{2}}(\mathbf{P})\mathbf{x}$, $\tilde{\mathbf{P}} = \mathrm{diag}^{-\frac{1}{2}}(\mathbf{P})\mathbf{P}\mathrm{diag}^{-\frac{1}{2}}(\mathbf{P})$ and $\tilde{\mathbf{h}} = \mathrm{diag}^{-\frac{1}{2}}(\mathbf{P})\mathbf{h}$ with $\mathrm{diag}(\mathbf{P})$ denoting a diagonal matrix by only retaining the diagonal elements of $\mathbf{P}$. Then, Gaussian BP is carried out with $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{h}}$ using the updating equations in (6), (7), (9) and (10). Under the normalized $\tilde{\mathbf{P}}$ and $\tilde{\mathbf{h}}$, we denote the corresponding message parameters as $\tilde{\mathbf{v}}^a(t)$ and $\tilde{\boldsymbol{\beta}}^a(t)$. With $\tilde{\mathbf{v}}^a(t)$, the corresponding belief variance can be calculated using (11) as $\tilde{\sigma}_i^2(t) = \frac{1}{\tilde{p}_{ii}+\sum_{k\in\mathcal{N}(i)} \tilde{v}_{k\to i}^a(t)}$. The variance of the original PDF $f(\mathbf{x})$ can be recovered from $\frac{1}{p_{ii}}\tilde{\sigma}_i^2(t)$. Moreover, under the normalized $\tilde{\mathbf{P}}$, the sets $\tilde{\mathcal{W}}$, $\tilde{\mathcal{S}}_1$, $\tilde{\mathcal{A}}$ and $\tilde{\mathcal{S}}$ can also be defined similar to (15), (17), (20) and (42), respectively.

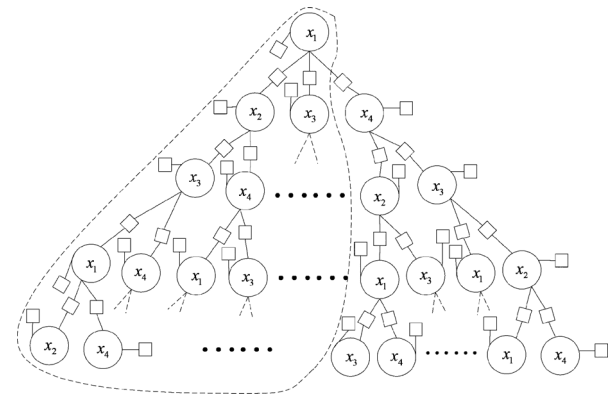

Fig. 1. Illustration of computation tree $\mathbb{T}(1,4)$ and computation sub-tree $\mathbb{T}_s(1,4,4)$ within the dashed line.

By noticing that the message-passing process of Gaussian BP in the factor graph can be translated exactly into the message passing in a computation tree, [24] proposed a convergence condition for belief variance $\tilde{\sigma}_i^2(t)$ in terms of computation tree. A computation tree $\mathbb{T}_{\gamma,\Delta}$ is constructed by choosing one of the variable nodes γ as root node and $\mathcal{N}(\gamma)$ as its direct descendants, connected through factor node $\tilde{f}_{\gamma,\nu}(\tilde{x}_\gamma, \tilde{x}_\nu) = \exp\{-\tilde{p}_{\gamma\nu}\tilde{x}_\gamma\tilde{x}_\nu\}$, where ν is the index of any descendant. For each descendant ν, the neighbors of ν excluding its parent node are connected as next level descendants, also connected through the corresponding factor nodes. The process is repeated until the computation tree $\mathbb{T}_{\gamma,\Delta}$ has $\Delta+1$ layers of variables nodes (including the root node). The computation tree is completed by further connecting factor node $\tilde{f}_i(\tilde{x}_i) = \exp\{-\frac{1}{2}\tilde{x}_i^2 + \tilde{h}_i\tilde{x}_i\}$ to variable node $\tilde{x}_i$ for $i \in \mathcal{V}$. Furthermore, a computation subtree $\mathbb{T}_s(\gamma,\nu,\Delta)$ can be obtained by cutting the branch of x_ν in the first layer from $\mathbb{T}(\gamma,\Delta)$. An example of computation tree $\mathbb{T}(1,4)$ as well as computation subtree $\mathbb{T}_s(1,4,4)$ corresponding to a fully connected factor graph with four variables are illustrated in Fig. 1.

By assigning to each variable node with a new index $i = 1,2,\ldots,|\mathbb{T}_{\gamma,\Delta}|$, $\mathbb{T}_{\gamma,\Delta}$ can be viewed as a factor graph for a new PDF, where $|\mathbb{T}_{\gamma,\Delta}|$ denotes the number of variable nodes in $\mathbb{T}_{\gamma,\Delta}$. The new PDF represented by $\mathbb{T}_{\gamma,\Delta}$ is

$$\tilde{f}_{\gamma,\Delta}(\tilde{\mathbf{x}}_{\gamma,\Delta}) \propto \exp\left\{-\frac{1}{2}\tilde{\mathbf{x}}_{\gamma,\Delta}^T\tilde{\mathbf{P}}_{\gamma,\Delta}\tilde{\mathbf{x}}_{\gamma,\Delta} + \tilde{\mathbf{h}}_{\gamma,\Delta}^T\tilde{\mathbf{x}}_{\gamma,\Delta}\right\}, \tag{47}$$

where $\tilde{\mathbf{x}}_{\gamma,\Delta}$ is a $|\mathbb{T}_{\gamma,\Delta}| \times 1$ vector; $\tilde{\mathbf{P}}_{\gamma,\Delta}$ and $\tilde{\mathbf{h}}_{\gamma,\Delta}$ are the corresponding precision matrix and linear coefficient vector, respectively. According to the equivalence between the message-passing process in original factor graph and that in computation tree [6], [24], under the prerequisite of $\tilde{\mathbf{P}}_{\gamma,\Delta} \succ 0$ and $\tilde{\mathbf{v}}^a(0) = \mathbf{0}$, it is known that

$$\tilde{\sigma}_\gamma^2(\Delta) = \left[\tilde{\mathbf{P}}_{\gamma,\Delta}^{-1}\right]_{11}. \tag{48}$$

Similarly, the computation sub-tree $\mathbb{T}_s(\gamma,\nu,\Delta)$ can also be viewed as a factor graph with $\tilde{\mathbf{P}}_{\gamma\setminus\nu,\Delta}$ being the corresponding

symmetric precision matrix in the PDF. Furthermore, if (48) holds, we also have

$$\tilde{\sigma}^2_{\gamma\backslash\nu}(\Delta) = \left[\tilde{\mathbf{P}}^{-1}_{\gamma\backslash\nu,\Delta}\right]_{11}, \tag{49}$$

where $\tilde{\sigma}^2_{\gamma\backslash\nu}(\Delta) \triangleq \frac{1}{\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)\backslash\nu}\tilde{v}^a_{k\to\gamma}(\Delta)}$. In ([24], Proposition 25), it is proved that if $\lim_{\Delta\to\infty}\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I}) < 1$, the variance $\tilde{\sigma}^2_\gamma(t)$ converges for $\tilde{\mathbf{v}}^a(0) = \mathbf{0}$; but if $\lim_{\Delta\to\infty}\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I}) > 1$, Gaussian BP becomes ill-posed. Notice that the limit $\lim_{\Delta\to\infty}\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I})$ always exist and are identical for all $\gamma\in\mathcal{V}$ ([24], Lemma 24). The following lemma and theorem reveal the relation between the proposed convergence condition and that based on computation tree.

Lemma 4: If $\tilde{\mathbf{P}}_{\gamma,\Delta}\succ 0$ for all $\gamma\in\mathcal{V}$ and $\Delta\geq 0$, then $\tilde{\mathbf{v}}^a(t)$ converges to $\tilde{\mathbf{v}}^{a*}\in\tilde{\mathcal{S}}_1$ for $\tilde{\mathbf{v}}^a(0)=\mathbf{0}$.

Proof: See Appendix B. ■

Now, we give the following theorem.

Theorem 6: $\tilde{\mathcal{S}}\neq\emptyset$ if and only if either of the following two conditions holds:

1) $\lim_{\Delta\to\infty}\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I}) < 1$;
2) $\lim_{\Delta\to\infty}\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I}) = 1$, $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}_{k\to\gamma}\neq 0$ with $\tilde{\mathbf{v}}^a(0)=\mathbf{0}$, and $\tilde{\mathbf{P}}_{\nu,\Delta}\succ 0$ for all $\nu\in\mathcal{V}$ and $\Delta\geq 0$.

Proof:

Sufficient Condition:

First, we prove that if $\lim_{\Delta\to\infty}\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I}) < 1$, then $\tilde{\mathcal{S}}\neq\emptyset$. With the monotonically increasing property of $\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I})$ ([24], Lemma 23) and the assumption that its limit is smaller than 1, we have $\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I})\leq\rho_u$ with $\rho_u\triangleq\lim_{\Delta\to\infty}\rho(\tilde{\mathbf{P}}_{\gamma,\Delta}-\mathbf{I})\in(0,1)$. Expressing in terms of eigenvalues, we have $-\rho_u\leq\lambda(\tilde{\mathbf{P}}_{\gamma,\Delta})-1\leq\rho_u$, or equivalently

$$1-\rho_u\leq\lambda(\tilde{\mathbf{P}}_{\gamma,\Delta})\leq 1+\rho_u. \tag{50}$$

For a symmetric $\tilde{\mathbf{P}}_{\gamma,\Delta}$, the eigenvalues of $\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}$ are $\{\frac{1}{\lambda(\tilde{\mathbf{P}}_{\gamma,\Delta})}\}$, and (50) is equivalent to

$$\frac{1}{1+\rho_u}\leq\lambda\left(\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}\right)\leq\frac{1}{1-\rho_u}. \tag{51}$$

Due to $0<\rho_u<1$, it can be obtained from (51) that $\lambda(\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta})>0$, that is,

$$\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}\succ 0,\ \forall\gamma\in\mathcal{V}\text{ and }\Delta\geq 0. \tag{52}$$

Notice that the diagonal element of a symmetric matrix is always smaller or equal to the matrix's maximum eigenvalue [34], that is, $[\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}]_{ii}\leq\lambda_{\max}(\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta})$. Combining with the fact that for a positive definite matrix, its diagonal elements are positive, we obtain $0<[\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}]_{ii}\leq\lambda_{\max}(\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta})$. Since the upper bound of (51) applies to all eigenvalues, we obtain

$$0<\left[\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}\right]_{ii}\leq\frac{1}{1-\rho_u}. \tag{53}$$

Due to $[\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}]_{11} = \tilde{\sigma}^2_\gamma(\Delta)$ from (48) and $\tilde{\sigma}^2_\gamma(\Delta) = \frac{1}{\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^a_{k\to\gamma}(\Delta)}$, it can be inferred from (53) that $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^a_{k\to\gamma}(\Delta)\geq 1-\rho_u$ for all $\Delta\geq 0$. Due to $\tilde{\mathbf{P}}^{-1}_{\gamma,\Delta}\succ 0$ from (52), we have $\tilde{\mathbf{P}}_{\gamma,\Delta}\succ 0$, and according to Lemma 4, $\tilde{\mathbf{v}}^a(t)$ always converges to $\tilde{\mathbf{v}}^{a*}\in\tilde{\mathcal{S}}_1$. Thus, it can be inferred that $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}_{k\to\gamma}\geq 1-\rho_u$. With $\rho_u<1$, we obtain

$$\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}_{k\to\gamma}>0,\quad\forall\gamma\in\mathcal{V}. \tag{54}$$

Combining (54) with $\tilde{\mathbf{v}}^{a*}\in\tilde{\mathcal{S}}_1$, it can be inferred that $\tilde{\mathbf{v}}^{a*}\in\tilde{\mathcal{S}}$ by the definition, thus $\tilde{\mathcal{S}}\neq\emptyset$.

Next, we prove the second scenario. Due to $\tilde{\mathbf{P}}_{\gamma,\Delta}\succ 0$, according to Lemma 4, $\tilde{\mathbf{v}}^a(t)$ converges to $\tilde{\mathbf{v}}^{a*}\in\tilde{\mathcal{S}}_1$, and thereby $\tilde{\mathcal{S}}_1\neq\emptyset$ and $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}$ exists. Due to $\tilde{\mathcal{S}}_1\neq\emptyset$, according to Lemma 3, it is known that $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}>0$ for all $\gamma\in\mathcal{V}$ or $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}=0$ for all $\gamma\in\mathcal{V}$. From the prerequisite $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}\neq 0$, we can infer that $\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)}\tilde{v}^{a*}>0$ holds for all $\gamma\in\mathcal{V}$. Combining with $\tilde{\mathbf{v}}^{a*}\in\tilde{\mathcal{S}}_1$, we can see that $\tilde{\mathbf{v}}^{a*}\in\tilde{\mathcal{S}}$, thus $\tilde{\mathcal{S}}\neq\emptyset$.

Necessary Condition:

In the following, we will first prove that the precision matrix of a computation tree can always be written into a special structure by re-indexing the nodes in the tree. Based on the special structure, the necessity is proved by the method of induction.

First, notice that changing indexing schemes in the computation tree does not affect the positive definiteness of the corresponding precision matrix $\tilde{\mathbf{P}}_{\gamma,\Delta}$. So, we consider an indexing scheme such that the precision matrix $\tilde{\mathbf{P}}_{\gamma,\Delta+1}$ for $\Delta\geq 0$ can be represented in the form

$$\tilde{\mathbf{P}}_{\gamma,\Delta+1} = \begin{bmatrix} \tilde{p}_{\gamma\gamma} & \tilde{\mathbf{a}}^T_{k_1\backslash\gamma,\Delta} & \tilde{\mathbf{a}}^T_{k_2\backslash\gamma,\Delta} & \cdots & \tilde{\mathbf{a}}^T_{k_{|\mathcal{N}(\gamma)|}\backslash\gamma,\Delta} \\ \tilde{\mathbf{a}}_{k_1\backslash\gamma,\Delta} & \tilde{\mathbf{P}}_{k_1\backslash\gamma,\Delta} & \mathbf{0} & \cdots & \mathbf{0} \\ \tilde{\mathbf{a}}_{k_2\backslash\gamma,\Delta} & \mathbf{0} & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \mathbf{0} \\ \tilde{\mathbf{a}}_{k_{|\mathcal{N}(\gamma)|}\backslash\gamma,\Delta} & \mathbf{0} & \cdots & \mathbf{0} & \tilde{\mathbf{P}}_{k_{|\mathcal{N}(\gamma)|}\backslash\gamma,\Delta} \end{bmatrix}, \tag{55}$$

where $k_i\in\mathcal{N}(\gamma)$ for $i=1,2,\ldots,|\mathcal{N}(\gamma)|$; $\tilde{\mathbf{a}}_{k_i\backslash\gamma,\Delta} = [\tilde{p}_{k_i\gamma},0,\ldots,0]^T$ is a vector with length $|\mathbb{T}_s(k_i,\gamma,\Delta)|$. Notice that the $(\Delta+1)$-order computation tree $\mathbb{T}(\gamma,\Delta+1)$ consists of the root node $[\tilde{\mathbf{x}}]_\gamma$ and a set of Δ-order computation sub-trees $\mathbb{T}_s(k_i,\gamma,\Delta)$ with the corresponding root node k_i. The lower-right block diagonal structure of (55) can be easily obtained by assigning consecutive indices to nodes inside each sub-tree $\mathbb{T}_s(k_i,\gamma,\Delta)$. Moreover, since there is only one connection from node γ to the root node k_i of each sub-tree $\mathbb{T}_s(k_i,\gamma,\Delta)$, $\tilde{\mathbf{a}}_{k_i\backslash\gamma,\Delta}$ contains only one nonzero element $\tilde{p}_{k_i\gamma}$. Then, by assigning the smallest index to the root node in each $\mathbb{T}_s(k_i,\gamma,\Delta)$, the only nonzero element in $\tilde{\mathbf{a}}_{k_i\backslash\gamma,\Delta}$ must locate at the first position. Therefore, the precision matrix $\tilde{\mathbf{P}}_{\gamma,\Delta+1}$ can be represented in the form of (55).

When $\Delta=0$, obviously, $\tilde{\mathbf{P}}_{\gamma,\Delta}=\tilde{p}_{\gamma\gamma}>0$. Suppose $\tilde{\mathbf{P}}_{\gamma,\Delta}\succ 0$ for some $\Delta\geq 0$, we need to prove $\tilde{\mathbf{P}}_{\gamma,\Delta+1}\succ 0$. From (55), it is clear that $\tilde{\mathbf{P}}_{\gamma,\Delta+1}\succ 0$ if and only if the following two conditions are satisfied ([34], p. 472):

$$\mathrm{Bdiag}\left(\tilde{\mathbf{P}}_{k_1\backslash\gamma,\Delta},\ldots,\tilde{\mathbf{P}}_{k_{|\mathcal{N}(\gamma)|}\backslash\gamma,\Delta}\right)\succ 0, \tag{56}$$

$$\tilde{p}_{\gamma\gamma}-\sum_{k\in\mathcal{N}(\gamma)}\tilde{\mathbf{a}}^T_{k\backslash\gamma,\Delta}\tilde{\mathbf{P}}^{-1}_{k\backslash\gamma,\Delta}\tilde{\mathbf{a}}_{k\backslash\gamma,\Delta}>0, \tag{57}$$

where $\mathrm{Bdiag}(\cdot)$ denotes block diagonal matrix with the elements located along the main diagonal. Due to $\tilde{\mathbf{P}}_{k,\Delta} \succ 0$ for all $k \in \mathcal{V}$ by assumption, then its sub-matrices $\tilde{\mathbf{P}}_{k\backslash\gamma,\Delta} \succ 0$ for all $(k,\gamma) \in \mathcal{E}$, thus the first condition (56) holds. On the other hand, for the second condition (57), we write

$$\begin{aligned}\tilde{p}_{\gamma\gamma} &- \sum_{k\in\mathcal{N}(\gamma)} \tilde{\mathbf{a}}_{k\backslash\gamma,\Delta}^T \mathbf{P}_{k\backslash\gamma,\Delta}^{-1} \tilde{\mathbf{a}}_{k\backslash\gamma,\Delta} \\ &\overset{a}{=} \tilde{p}_{\gamma\gamma} - \sum_{k\in\mathcal{N}(\gamma)} p_{k\gamma}^2 \cdot \tilde{\sigma}_{k\backslash\gamma}^2(\Delta) \\ &\overset{b}{=} \tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} \tilde{v}_{k\to\gamma}^a(\Delta+1), \end{aligned} \tag{58}$$

where the equality $\overset{a}{=}$ holds since $\tilde{\mathbf{a}}_{k\backslash\gamma,\Delta}$ has only one nonzero $\tilde{p}_{k\gamma}$ in the first element, and $[\mathbf{P}_{k\backslash\gamma,\Delta}^{-1}]_{11} = \tilde{\sigma}_{k\backslash\gamma}^2(\Delta)$ from (49); and $\overset{b}{=}$ holds due to $-p_{k\gamma}^2 \cdot \tilde{\sigma}_{k\backslash\gamma}^2(\Delta) = -\frac{\tilde{p}_{k\gamma}^2}{\tilde{p}_{\gamma\gamma}+\sum_{k\in\mathcal{N}(\gamma)\backslash\nu} \tilde{v}_{k\to\gamma}^a(\Delta)} = \tilde{v}_{k\to\gamma}^a(\Delta+1)$ given in (12). If $\tilde{\mathcal{S}} \neq \emptyset$, according to Theorem 4, we have

$$\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} \tilde{v}_{k\to\gamma}^{a*} > 0. \tag{59}$$

Due to $\tilde{\mathcal{S}} \subseteq \tilde{\mathcal{S}}_1$, then $\tilde{\mathcal{S}} \neq \emptyset$ also implies $\tilde{\mathcal{S}}_1 \neq \emptyset$. According to (63), it can be inferred that $\tilde{\mathbf{v}}^a(t) \geq \tilde{\mathbf{v}}^{a*}$. Combining with (59), we obtain $\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} \tilde{v}_{k\to\gamma}^a(t) > 0$. Substituting the result into (58), it can be inferred that the second condition (57) holds as well. Thus, we have $\tilde{\mathbf{P}}_{\gamma,\Delta} \succ 0$ for all $\gamma \in \mathcal{V}$ and $\Delta > 0$. Furthermore, from (59), it is obvious that $\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} \tilde{v}_{k\to\gamma}^{a*} \neq 0$.

From $\tilde{\mathbf{P}}_{\gamma,\Delta} \succ 0$, we obtain $\mathbf{I} - (\mathbf{I} - \tilde{\mathbf{P}}_{\gamma,\Delta}) \succ 0$, and hence $1 - \lambda_{\max}(\mathbf{I} - \tilde{\mathbf{P}}_{\gamma,\Delta}) > 0$. Since $\tilde{\mathbf{P}}_{\gamma,\Delta}$ represents a tree-structured factor graph, it is proved in ([24], Proposition 15) that $\lambda_{\max}(\mathbf{I} - \tilde{\mathbf{P}}_{\gamma,\Delta}) = \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I})$. Therefore, it can be obtained that $\rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) < 1$, and thereby

$$\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) \leq 1. \tag{60}$$

Finally, if $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) < 1$, due to $\tilde{\mathbf{P}}_{\gamma,\Delta}^{-1} \succ 0$ from (52), we have $\tilde{\mathbf{P}}_{\gamma,\Delta} \succ 0$. Together with (54), it can be inferred that under the prerequisite of $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) < 1$, the conditions $\tilde{\mathbf{P}}_{\gamma,\Delta} \succ 0$ and $\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} \tilde{v}_{k\to\gamma}^{a*} \neq 0$ are automatically satisfied. Therefore, if $\tilde{\mathcal{S}} \neq \emptyset$, we have either $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) < 1$ or $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) = 1$, $\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} \tilde{v}_{k\to\gamma}^{a*} \neq 0$ and $\tilde{\mathbf{P}}_{\gamma,\Delta} \succ 0$. ■

From Theorems 4 and 6, it can be obtained that the variance $\sigma_\gamma^2(t)$ converges as $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) < 1$, and diverges as $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) > 1$, which are consistent with the results proposed in [24]. Moreover, it can be seen from Theorem 6 that it is not sufficient to determine the convergence of variance $\tilde{\sigma}_\gamma^2(t)$ by using $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) = 1$ only. This fills in the gap of [24] in the scenario of $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) = 1$. Albeit with similar conclusions to [24], we need to emphasize that the criterion $\lim_{\Delta\to\infty} \rho(\tilde{\mathbf{P}}_{\gamma,\Delta} - \mathbf{I}) < 1$ proposed in [24] is not easy to check in practice due to the infinite dimension, while our condition $\tilde{\mathcal{S}} \neq \emptyset$ can be verified by solving an SDP problem given in Theorem 5. Moreover, the initialization is expanded from the a single choice $\tilde{\mathbf{v}}^a(0) = \mathbf{0}$ in [24] to a much larger set $\tilde{\mathbf{v}}^a(0) \in \tilde{\mathcal{A}}$ in this paper. The flexibility on the choice of initialization is useful to accelerate the convergence of variance $\sigma_i^2(t)$ if the initialization is chosen close to the convergent point.

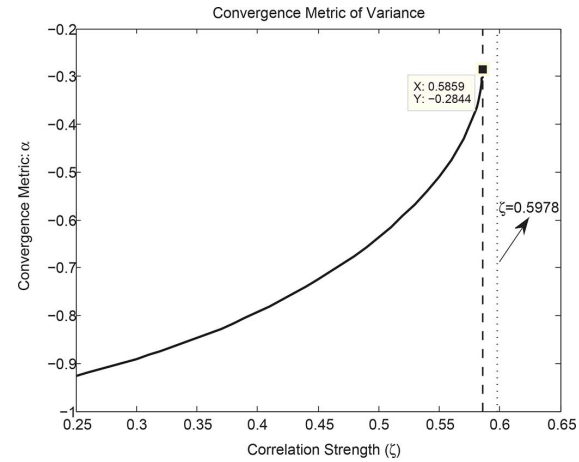

Fig. 2. The value of α^* under different correlation strength ζ.

VII. Numerical Examples

In this section, numerical experiments are presented to corroborate the theories in this paper. The example is based on the 20×20 precision matrices $\mathbf{P}$ constructed as

$$p_{ij} = \begin{cases} 1, & \text{if } i = j, \\ \zeta \cdot \theta_{\mathrm{mod}(i+j,10)+1}, & \text{if } i \neq j, \end{cases} \tag{61}$$

where ζ is a coefficient indicating the correlation strength among variables; and θ_k is the k-th element of the vector $\boldsymbol{\theta} = [0.13, 0.10, 0.71, -0.05, 0, 0.12, 0.07, 0.11, -0.02, -0.03]^T$. The varying of correlation strength ζ induces a series of matrices, and the positive definite constraint $\mathbf{P} \succ \mathbf{0}$ required by a valid PDF is guaranteed when $\zeta < 0.5978$.

Fig. 2 illustrates how the optimal solution α^* of (43) varies with the correlation strength ζ. It can be seen that the optimal solution α^* always exists and the condition $\alpha^* < 0$ holds for all $\zeta \leq 0.5859$, while no feasible solution exists in the SDP problem (43) when $\zeta > 0.5859$. According to Theorem 4, this means that if $\zeta \leq 0.5859$, the variance $\sigma_i^2(t)$ with $i \in \mathcal{V}$ converges to the same point for all initializations $\mathbf{v}^a(0) \in \mathcal{A}$ under both synchronous and asynchronous schedulings. On the other hand, if $\zeta > 0.5859$, the variance $\sigma_i^2(t)$ cannot converge.

To verify the convergence of belief variances under $\zeta \leq 0.5859$, Fig. 3 shows how the variance $\sigma_1^2(t)$ of the 1-st variable evolves as a function of t when $\zeta = 0.5858$, which is slightly smaller than 0.5859. It can be observed that the variance $\sigma_1^2(t)$ converges to the same value under both synchronous and asynchronous schedulings and different initializations of $\mathbf{v}^a(0) = \mathbf{0}$ and $\mathbf{v}^a(0) = \mathbf{w}^*$, where $\mathbf{w}^*$ is the optimal solution of (45). For the asynchronous case, a scheduling with 30% chance of not updating the messages at each iteration is considered. On the other hand, Fig. 4 verifies the divergence of variance $\sigma_1^2(t)$ when $\zeta = 0.5860$, which is slightly larger than 0.5859. In this figure, synchronous scheduling and the same initializations as

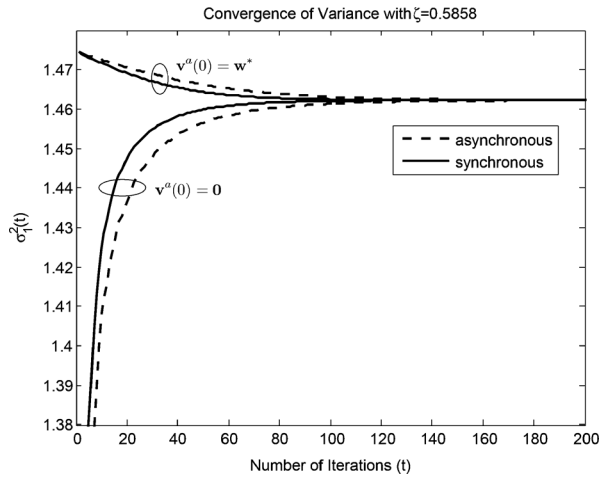

Fig. 3. Illustration for the convergence of variance $\sigma_1^2(t)$ under different schedulings and initializations with $\mathbf{w}^*$ being the optimal solution of (45).

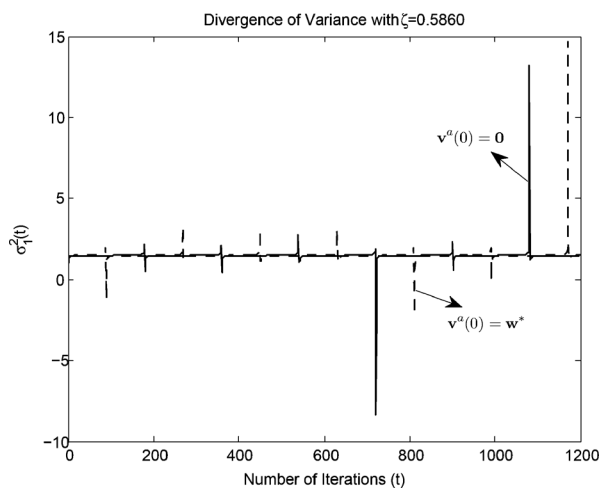

Fig. 4. Illustration for the divergence of variance $\sigma_1^2(t)$ with $\mathbf{w}^*$ being the optimal solution of (45).

that of Fig. 3 are used. It can be seen that $\sigma_1^2(t)$ fluctuates as iterations proceed, and does not show sign of convergence.

VIII. Conclusion

In this paper, the necessary and sufficient convergence condition for the variances of Gaussian BP was developed for synchronous and asynchronous schedulings. The initialization set of the proposed condition is much larger than the usual choice of a single point of zero. It is proved that the convergence condition can be verified efficiently by solving an SDP problem. Furthermore, the relationship between the convergence condition proposed in this paper and the one based on computation tree was established. The relationship fills in a missing piece of the result in [24] where the spectral radius of computation tree is equal to one. Numerical examples were further proposed to verify the proposed convergence conditions.

Appendix A
Proof of Lemma 2

First, we prove that $\mathbf{v}^a(t)$ converges for any $\mathbf{v}^a(0) \geq \mathbf{0}$ given $\mathcal{S}_1 \neq \emptyset$. For any $\mathbf{w} \in \mathcal{S}_1$, according to P5), we have $\mathbf{w} < \mathbf{0}$. Thus, for any $\mathbf{v}^a(0) \geq \mathbf{0}$, the relation $\mathbf{w} \leq \mathbf{v}^a(0)$ always holds. Notice that $\mathbf{w} \in \mathcal{W}$ due to $\mathbf{w} \in \mathcal{S}_1$ and $\mathcal{S}_1 \subseteq \mathcal{W}$. Applying P2) to $\mathbf{w} \leq \mathbf{v}^a(0)$, we obtain $\mathbf{g}(\mathbf{w}) \leq \mathbf{v}^a(1)$. Combining it with $\mathbf{w} \leq \mathbf{g}(\mathbf{w})$ from P6) gives $\mathbf{w} \leq \mathbf{v}^a(1)$. On the other hand, substituting $\mathbf{v}^a(0) \geq \mathbf{0}$ into (14) gives

$$\mathbf{v}^a(1) < \mathbf{0}. \tag{62}$$

Due to $\mathbf{v}^a(0) \geq \mathbf{0}$, thus $\mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Combining it with $\mathbf{w} \leq \mathbf{v}^a(1)$ gives $\mathbf{w} \leq \mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Applying $\mathbf{g}(\cdot)$ to $\mathbf{w} \leq \mathbf{v}^a(1) < \mathbf{v}^a(0)$, it can be inferred from P2) that $\mathbf{g}(\mathbf{w}) \leq \mathbf{v}^a(2) < \mathbf{v}^a(1)$. Together with $\mathbf{w} \leq \mathbf{g}(\mathbf{w})$ as claimed by P6), we obtain $\mathbf{w} \leq \mathbf{v}^a(2) < \mathbf{v}^a(1)$. By induction, we can infer that

$$\mathbf{w} \leq \mathbf{v}^a(t+1) \leq \mathbf{v}^a(t). \tag{63}$$

It can be seen from (63) that $\mathbf{v}^a(t)$ is a monotonically non-increasing but lower bounded sequence, thus it converges.

Next, we prove that $\mathbf{v}^a(t)$ converges to the same point for all $\mathbf{v}^a(0) \geq \mathbf{0}$. For any $\mathbf{v}^a(0) \geq \mathbf{0}$, according to P2), applying $\mathbf{g}(\cdot)$ on both sides of $\mathbf{0} \leq \mathbf{v}^a(0)$ gives $\mathbf{g}(\mathbf{0}) \leq \mathbf{g}(\mathbf{v}^a(0)) = \mathbf{v}^a(1)$. Combining this relation with (62) leads to $\mathbf{g}(\mathbf{0}) \leq \mathbf{v}^a(1) \leq \mathbf{0}$. Applying $\mathbf{g}(\cdot)$ on this inequality for t more times, it can be obtained from P2) that $\mathbf{g}^{(t+1)}(\mathbf{0}) \leq \mathbf{v}^a(t+1) \leq \mathbf{g}^{(t)}(\mathbf{0})$ for all $t \geq 0$. By denoting $\lim_{t\to\infty} \mathbf{g}^{(t)}(\mathbf{0}) = \mathbf{v}^{a*}$, it is obvious that $\mathbf{v}^a(t)$ also converges to the $\mathbf{v}^{a*}$.

Finally, since $\mathbf{v}^{a*}$ is a fixed point of $\mathbf{g}(\cdot)$, hence $\mathbf{v}^{a*} = \mathbf{g}(\mathbf{v}^{a*})$. From (63), we obtain $\mathbf{w} \leq \mathbf{v}^{a*}$. Due to $\mathbf{w} \in \mathcal{S}_1$, thus $\mathbf{w} \in \mathcal{W}$ by definition. Applying P1), it can be inferred that $\mathbf{v}^{a*} \in \mathcal{W}$. Combining with the fact $\mathbf{v}^{a*} = \mathbf{g}(\mathbf{v}^{a*})$, according to the definition of $\mathcal{S}_1$ in (17), we obtain $\mathbf{v}^{a*} \in \mathcal{S}_1$.

On the other hand, if $\mathcal{S}_1 = \emptyset$, according to Theorem 1, the messages of Gaussian BP passed in factor graph cannot be maintained at the Gaussian form, thus the parameters of messages $\mathbf{v}^a(t)$ cannot converge in this case.

Appendix B
Proof of Lemma 4

Due to $\tilde{\mathbf{P}}_{\gamma,\Delta} \succ 0$, then $\tilde{\mathbf{P}}_{\gamma,\Delta}^{-1} \succ 0$, hence the diagonal elements $[\tilde{\mathbf{P}}_{\gamma,\Delta}^{-1}]_{ii} > 0$ [34]. With $\tilde{\sigma}_\gamma^2(\Delta) = [\tilde{\mathbf{P}}_{\gamma,\Delta}^{-1}]_{11}$ and $\frac{1}{\tilde{\sigma}_\gamma^2(t)} = \tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)} \tilde{v}_{k\to\gamma}^a(t)$, we have

$$\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus\nu} \tilde{v}_{k\to\gamma}^a(t) > -\tilde{v}_{\nu\to\gamma}^a(t). \tag{64}$$

Next, we prove $\tilde{\mathbf{v}}^a(t) \in \tilde{\mathcal{W}}$ for $t \geq 0$. Obviously, $\tilde{\mathbf{v}}^a(0) = \mathbf{0} \in \tilde{\mathcal{W}}$. Suppose $\tilde{\mathbf{v}}^a(t) \in \tilde{\mathcal{W}}$ holds for some $t \geq 0$, and according to the definition of $\tilde{\mathcal{W}}$ in (15), this is equivalent to $\tilde{p}_{\nu\nu} + \sum_{k\in\mathcal{N}(\nu)\setminus\gamma} \tilde{v}_{k\to\nu}^a(t) > 0$. Substituting this inequality into (14) gives $\tilde{v}_{\nu\to\gamma}^a(t+1) < 0$. Then applying $\tilde{v}_{\nu\to\gamma}^a(t+1) < 0$ into (64) gives $\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus\nu} \tilde{v}_{k\to\gamma}^a(t+1) > 0$ for all $(\nu,\gamma) \in \mathcal{E}$, or equivalently $\tilde{\mathbf{v}}^a(t+1) \in \tilde{\mathcal{W}}$. Thus, we have $\tilde{\mathbf{v}}^a(t) \in \tilde{\mathcal{W}}$ for all $t \geq 0$.

Finally, substituting $\tilde{\mathbf{v}}^a(0) = \mathbf{0}$ into (13) gives $\tilde{\mathbf{v}}^a(1) < \mathbf{0}$, and thereby $\tilde{\mathbf{v}}^a(1) < \tilde{\mathbf{v}}^a(0)$. From $\tilde{\mathbf{v}}^a(t) \in \tilde{\mathcal{W}}$ and P2), applying $\mathbf{g}(\cdot)$ on both sides of $\tilde{\mathbf{v}}^a(1) < \tilde{\mathbf{v}}^a(0)$ for t times gives

$$\tilde{\mathbf{v}}^a(t+1) \leq \tilde{\mathbf{v}}^a(t). \tag{65}$$

From $\tilde{\mathbf{v}}^a(0) = \mathbf{0}$ and (65), we have $\tilde{\mathbf{v}}^a(t) \leq \mathbf{0}$. Putting the result into (64) gives $\tilde{v}_{\nu\to\gamma}^a(t) > -\tilde{p}_{\gamma\gamma}$ for all $(\nu,\gamma) \in \mathcal{E}$. Together with (65), it can be seen that $\tilde{\mathbf{v}}^a(t)$ is a monotonically

non-increasing and lower bounded sequence. Thus, $\tilde{\mathbf{v}}^a(t)$ converges to a fixed point $\tilde{\mathbf{v}}^{a*}$ with $\tilde{\mathbf{v}}^{a*} = \mathbf{g}(\tilde{\mathbf{v}}^{a*})$. Moreover, from $\tilde{\mathbf{v}}^a(1) < \mathbf{0}$ and (65), we have $\tilde{\mathbf{v}}^{a*} < \mathbf{0}$. Putting the result into the limiting form of (64) gives $\tilde{p}_{\gamma\gamma} + \sum_{k\in\mathcal{N}(\gamma)\setminus\nu} \tilde{v}^{a*}_{k\to\gamma} > 0$ for all $(\nu,\gamma) \in \mathcal{E}$, or equivalently $\tilde{\mathbf{v}}^{a*} \in \tilde{\mathcal{W}}$. Combining with $\tilde{\mathbf{v}}^{a*} = \mathbf{g}(\tilde{\mathbf{v}}^{a*})$, we obtain $\tilde{\mathbf{v}}^{a*} \in \tilde{\mathcal{S}}_1$ by the definition of $\mathcal{S}_1$.

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, no. 2, pp. 498–519, Feb. 2001.

[2] H. A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 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*. Berlin, Germany: Springer-Verlag, Aug. 2006.

[5] D. Koller and N. Friedman, *Probabilistic Graphical Models Principles and Techniques*. Cambridge, MA, USA: MIT Press, 2009.

[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] 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, Aug. 2012.

[8] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multiuser detection," presented at the IEEE Inf. Theory Workshop, Punta del Este, Uruguay, Mar. 2006.

[9] 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, no. 8, pp. 4030–4035, Aug. 2011.

[10] Q. Guo and P. Li, "LMMSE turbo equalization based on factor graphs," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 311–319, Feb. 2008.

[11] 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.

[12] 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. Inf. Theory (ISIT)*, 2008, pp. 1863–1867.

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

[14] 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, no. 5, pp. 1916–1930, May 2008.

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

[16] F. Lehmann, "Iterative mitigation of intercell interference in cellular networks based on Gaussian belief propagation," *IEEE Trans. Veh. Technol.*, vol. 61, no. 6, pp. 2544–2558, Jul. 2012.

[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. Acoustics, Speech, Signal Process. (ICASSP)*, 2012, pp. 3125–3128.

[20] Y. Song, B. Wang, Z. Shi, K. Pattipati, and S. Gupta, "Distributed algorithms for energy-efficient even self-deployment in mobile sensor networks," *IEEE Trans. Mobile Comput.*, vol. 13, no. 5, pp. 1–14, May 2013.

[21] G. Zhang, W. Xu, and Y. Wang, "Fast distributed rate control algorithm with QoS support in Ad Hoc networks," presented at the IEEE Global Telecommun. Conf. (Globecom), 2010.

[22] D. Dolev, A. Zymnis, S. P. Boyd, D. Bickson, and Y. Tock, "Distributed large scale network utility maximization," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, 2009, pp. 829–833.

[23] M. W. Seeger and D. P. Wipf, "Variational Bayesian inference techniques," *IEEE Signal Process. Mag.*, vol. 27, no. 6, pp. 81–91, Nov. 2010.

[24] D. M. Malioutov, J. K. Johnson, and A. S. Willsky, "Walk-sums and belief propagation in Gaussian graphical models," *J. Mach. Learn. Res.*, vol. 7, no. 1, pp. 2031–2064, 2006.

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

[26] S. Tatikonda and M. I. Jordan, "Loopy belief propagation and Gibbs measures," in *Proc. 18th Annu. Conf. Uncertainty Artif. Intell. (UAI-02)*, San Francisco, CA, USA, Aug. 2002, pp. 493–500.

[27] D. P. Bertsekas and J. N. Tsitsiklis, *Parallel and Distributed Computation: Numerical Methods*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1989.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

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

[30] L. Vandenberghe and S. Boyd, "Semidefinite programming," *SIAM Rev.*, vol. 38, no. 1, pp. 49–95, Mar. 1996.

[31] M. Grant, S. Boyd, and Y. Ye, "CVX: Matlab Software for Disciplined Convex Programming," [Online]. Available: http://www.stanford.edu/boyd/cvx/

[32] 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.

[33] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, 2011.

[34] R. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.

Qinliang Su (S'09–M'10) received the B.Eng. degree from Chongqing University in 2007 and M.Eng. degree from Zhejiang University in 2010 (both with honors), respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Electronic Engineering, the University of Hong Kong, Hong Kong. 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. (E.E.E.) 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 the IEEE COMMUNICATIONS LETTERS, is currently an Editor for the IEEE TRANSACTIONS ON COMMUNICATIONS and the *Journal of Communications and Networks*.