<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>On Convergence Conditions of Gaussian Belief Propagation</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>SU, Q; Wu, YC</td>
</tr>
<tr>
<td><strong>Citation</strong></td>
<td>IEEE Transactions on Signal Processing, 2015, v. 63, p. 1144-1155</td>
</tr>
<tr>
<td><strong>Issued Date</strong></td>
<td>2015</td>
</tr>
<tr>
<td><strong>URL</strong></td>
<td><a href="http://hdl.handle.net/10722/214163">http://hdl.handle.net/10722/214163</a></td>
</tr>
<tr>
<td><strong>Rights</strong></td>
<td>This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.</td>
</tr>
</tbody>
</table>
On Convergence Conditions of Gaussian Belief Propagation
Qinliang Su and Yik-Chung Wu

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

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

I. INTRODUCTION

MANY problems in signal processing and machine learning eventually come to the issue of computing marginal probability density function (PDF) from a high dimensional joint PDF. In general, the complexity of directly computing the marginal PDF could be very high. By passing messages among the neighboring nodes in factor graph, belief propagation (BP) provides an efficient way to compute the approximate marginal PDFs upon convergence [1]–[5].

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

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

The contributions of this paper are summarized as follow.
1) For synchronous Gaussian BP, the necessary and sufficient convergence condition of beliefs under a newly proposed initialization set is derived. The proposed initialization set is proved to be the largest among all currently known results.
2) For damped Gaussian BP, the necessary and sufficient convergence condition of beliefs under the proposed initialization set is also derived. It is proved that this convergence condition is more relaxed than that of Gaussian BP. To the best of our knowledge, this is the first convergence condition of damped Gaussian BP, and theoretically confirms the extensively reported conjecture that damping is helpful to improve the convergence of Gaussian BP [24]–[28]. A method on how to choose an appropriate damping factor is also proposed.
3) For asynchronous Gaussian BP, a sufficient condition is derived to guarantee the convergence of beliefs under all totally asynchronous schedulings.
4) Relationships between the proposed convergence conditions and existing ones are established analytically. The results demonstrate that the existing conditions are implied by the proposed ones.

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

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

II. GAUSSIAN BELIEF PROPAGATION

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

$$m_{i\to j}^a(x_i, t) \propto m_{i\to j}^b(x_i, t) f_i(x_i),$$

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

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

$$b_i(x_i, t) \propto \prod_{k \in \mathcal{N}(i) \setminus j} m_{i\to k}^c(x_i, t) f_i(x_i).$$

Without loss of generality, let the arriving message of node $j$ at time $t$ is in form of

$$m_{i\to j}^c(x_j, t) \propto \exp \left\{ -\frac{1}{2} \frac{v_{i\to j}^c(t)}{2} x_j^2 + \beta_{i\to j}^c(t) x_j \right\},$$

where $v_{i\to j}^c(t)$ and $\beta_{i\to j}^c(t)$ are the arriving precision and arriving linear coefficient, respectively. Inserting $m_{i\to j}^c(x_j, t)$ into (1), we obtain

$$m_{i\to j}^c(x_j, t) \propto \exp \left\{ -\frac{1}{2} \frac{v_{i\to j}^c(t)}{2} x_j^2 + \beta_{i\to j}^c(t) x_j \right\},$$

where

$$v_{i\to j}^c(t) = \mu_i + \sum_{k \in \mathcal{N}(i) \setminus j} v_{k\to i}^c(t),$$

$$\beta_{i\to j}^c(t) = h_i + \sum_{k \in \mathcal{N}(i) \setminus j} \beta_{k\to i}^c(t),$$

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

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

$$m_{i\to j}^d(x_j, t + 1) \propto \int \exp \left\{ -\frac{1}{2} \frac{v_{i\to j}^d(t)}{2} x_j^2 + \beta_{i\to j}^d(t) x_j \right\} dx_j.$$

If $v_{i\to j}^d(t) > 0$, the integration equals to a constant, and

$$m_{i\to j}^d(x_j, t + 1) \propto \exp \left\{ -\frac{1}{2} \frac{v_{i\to j}^d(t)}{2} x_j^2 + \beta_{i\to j}^d(t) x_j \right\}.$$
Fig. 1. Illustration of the updating process of $v_{i,j}(t+1)$, where $g_{ij}(v^\ast(t)) = -\frac{r_{ij} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{k,i}}{p_{ij} - w_{ij}}$.

where $\mathbf{b}'(t)$ is a vector containing $b_{i,j}'(t)$ with $(i,j) \in \mathcal{E}$ arranged in ascending order first on $j$ then on $i$; $\mathbf{G}(t)$ and $\mathbf{b}(t)$ are defined as

\[
\mathbf{G}(t) \triangleq - \text{diag}^{-1}(\mathbf{A} \mathbf{v}^\ast(t) + \mathbf{u}) \text{diag}(\mathbf{p}) \mathbf{A},
\]

\[
\mathbf{b}(t) \triangleq - \text{diag}^{-1}(\mathbf{A} \mathbf{v}^\ast(t) + \mathbf{u}) \text{diag}(\mathbf{p}) \mathbf{\xi}_i.
\]

$\mathbf{A}$ is a $|\mathcal{E}| \times |\mathcal{E}|$ matrix defined such that $\mathbf{A} \mathbf{v}^\ast(t)$ is a column vector containing elements $\sum_{k \in \mathcal{N}(i) \setminus j} v_{k,i}^\ast(t)$ with $(i,j) \in \mathcal{E}$ arranged first on $j$ then on $i$; $\mathbf{p}$ contains elements $p_{ij}$ with $(i,j) \in \mathcal{E}$ arranged in ascending order first on $j$ then on $i$; $\mathbf{u} = [\mathbf{u}_1^T, \mathbf{u}_2^T, \ldots, \mathbf{u}_N^T]^T$ with $\mathbf{u}_i$ being a column vector containing elements $p_{ij}$ for all $j \in \mathcal{N}(i)$ arranged in ascending order; $\mathbf{\xi}_i = [\mathbf{\xi}_i^1, \mathbf{\xi}_i^2, \ldots, \mathbf{\xi}_i^N]^T$ with $\mathbf{\xi}_i$ being a column vector containing elements $h_{ij}$ for all $j \in \mathcal{N}(i)$ arranged in ascending order.

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

III. CONVERGENCE OF MESSAGES IN GAUSSIAN BP

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

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

\[
\mathcal{S}_1 \triangleq \{ \mathbf{w} | |\mathbf{w}| \leq \mathbf{g}(\mathbf{w}) \text{ and } \mathbf{w} \in \mathcal{W} \},
\]

$\mathcal{A} \triangleq \{ \mathbf{w} \geq 0 \} \cup \{ \mathbf{w} \geq 0 \} \cap \text{int}(\mathcal{S}_1)$

\[
\cap \{ \mathbf{w} \geq \mathbf{w}_0 | \mathbf{w}_0 \in \mathcal{S}_1 \text{ and } \mathbf{w}_0 = \lim_{t \to \infty} \mathbf{g}(t)(\mathbf{0}) \}
\]

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

\[\min_{\mathbf{w}} \| \mathbf{w} \|^2\]

s.t. \[p_{ij} + \sum_{k \in \mathcal{N}(i) \setminus j} w_{k,i} \geq 0, \forall (i,j) \in \mathcal{E}. \]

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

Proposition 1: If $\mathcal{S}_1 \neq \emptyset$ then there always exists a $\mathbf{w} \in \mathcal{S}_1$. From the definition of $\mathbf{S}_1$ in (17), we have $\mathbf{w} \leq \mathbf{g}(\mathbf{w})$ and $\mathbf{w} \in \mathcal{W}$. That is, $w_{ij} - g_{ij}(\mathbf{w}) \leq 0, \forall (i,j) \in \mathcal{E}$. It can be seen that the vector $\mathbf{w}$ satisfies the constraints (20) and (21), thus it is a feasible solution to (19).

On the other hand, if $\mathcal{S}_1$ has a feasible solution $\mathbf{w}$, then $\mathbf{w}$ must satisfy (20) and (21), that is, $w_{ij} - g_{ij}(\mathbf{w}) \leq 0$ and $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{k,i} \leq 0$. Now, suppose $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{k,i} < 0$. Obviously, in this case, the determinant of the matrix in the constraint of (19) equals to $-p_{i,i}^2 < 0$. Hence, the matrix in the constraint of (19) is not positive semi-definite, which contradicts to the prerequisite that $\mathbf{w}$ is a feasible solution of (19). Thus, we must have $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{k,i} \geq 0$. Combining with the established result $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{k,i} \leq 0$, we obtain $-p_{ii} - \sum_{k \in \mathcal{N}(i) \setminus j} w_{k,i} = 0$. Together with the established result $w_{ij} - g_{ij}(\mathbf{w}) \leq 0$, from the definition of $\mathcal{S}_1$, we can infer that $\mathbf{w} \in \mathcal{S}_1$, and thus $\mathcal{S}_1 \neq \emptyset$.

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

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

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

\[\mathbf{w}^* \leq \mathbf{g}(\mathbf{w}^*).\]
Next, the first-order derivative of $g_{ij}(w)$ with respect to $w_{k_i}$ for $k \in N(i) \setminus j$ is equal to
\[
\frac{\partial g_{ij}(w)}{\partial w_{k_i}} = \frac{p_{ij}}{(p_{ii} + \sum_{l \in N(i) \setminus j} w_{l})^2}.
\] (24)

Obviously, it can be seen that $\frac{\partial g_{ij}(w)}{\partial w_{k_i}} > 0$. Together with the fact $g_{ij}(w)$ is continuous for $w \in \mathbb{W}$, we can infer that $g_{ij}(w)$ is a monotonically increasing function for $w \in \mathbb{W}$. Thus, by applying $g(\cdot)$ on $w^* < 0$ in (23), we obtain $g(w^*) \leq g(0)$. Combining with the constraint $w^* \leq g(w^*)$ in (22) gives $w^* \leq g(0)$. Then, applying $g(\cdot)$ to $w^* \leq g(0)$ leads to $g(w^*) \leq g(0)$. Due to $w^* < g(w^*)$, we further have $w^* < g(2)(0)$. By induction, it can be inferred that $w^* \leq g(t)(0)$ for all $t \geq 0$. By taking the limit on both sides of $w^* \leq g(t)(0)$, and due to $g(t)(0) = v^*(t)$ with $v^*(0) = 0$, we obtain $w^* \leq \lim_{t \to \infty} v^*(t)$ under the initialization of $v^*(0) = 0$. Since according to Theorem 1, $\lim_{t \to \infty} v^*(t) = 0$ is the same as $\lim_{t \to \infty} v^*(t)$ with $v^*(0) \in \mathcal{A}$, we further have
\[
w^* \leq \lim_{t \to \infty} v^*(t)
\] (25)
for any $v^*(0) \in \mathcal{A}$.

Finally, since $v^*(t)$ converges to $\lim_{t \to \infty} v^*(t)$ for $v^*(0) \in \mathcal{A}$, according to (11), this means that $\lim_{t \to \infty} v^*(t) = g\left(\lim_{t \to \infty} v^*(t)\right)$ and $\lim_{t \to \infty} v^*(t) \in \mathcal{W}$. Writing the two conditions into scalar form, we obtain $\lim_{t \to \infty} v^*(t) = g\left(\lim_{t \to \infty} v^*(t)\right)$ and $\lim_{t \to \infty} v^*(t) \in \mathcal{W}$. Comparing with the constraints (20) and (21), it can be easily seen that $\lim_{t \to \infty} v^*(t)$ is a feasible solution of (19). Since $w^*$ is the optimal solution of the minimization problem (19), we must have
\[
|w^*|^2 \leq \left|\lim_{t \to \infty} v^*(t)\right|^2.
\] (26)

On the other hand, substituting $p_{ii} + \sum_{k \in N(i) \setminus j} \lim_{t \to \infty} v^*(t) > 0$ into (12), we obtain $g_{ij}\left(\lim_{t \to \infty} v^*(t)\right) < 0$ for all $(i, j) \in \mathcal{E}$, or equivalently $g\left(\lim_{t \to \infty} v^*(t)\right) < 0$. Due to $\lim_{t \to \infty} v^*(t) = g\left(\lim_{t \to \infty} v^*(t)\right)$, we have $\lim_{t \to \infty} v^*(t) < 0$. Combining with (25), it can be inferred that
\[
|w^*|^2 \geq \left|\lim_{t \to \infty} v^*(t)\right|^2.
\] (27)
Comparing (26) with (27), it can be seen that $|w^*|^2 = \left|\lim_{t \to \infty} v^*(t)\right|^2$. Suppose $w^* \neq \lim_{t \to \infty} v^*(t)$. According to (25), it can be inferred that $|w^*|^2 < \left|\lim_{t \to \infty} v^*(t)\right|^2$, which contradicts with the established result $|w^*|^2 = \left|\lim_{t \to \infty} v^*(t)\right|^2$. Therefore, we must have $w^* = \lim_{t \to \infty} v^*(t)$.

Now, we derive the convergence condition for the message parameters $(v^a(t), \beta^a(t))$. If $v^a(t)$ has already converged to $w^*$, then $\beta^a(t)$ is updated as
\[
\beta^a(t + 1) = G^* \beta^a(t) + b^*,
\] (28)

where
\[
G^* \triangleq -\text{diag}^{-1}(A w^* + u)\text{diag}(p|A),
\]
\[
b^* \triangleq -\text{diag}^{-1}(A w^* + u)\text{diag}(p)|\mathcal{E}.
\]

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

However, in Gaussian BP, $v^a(t)$ and $\beta^a(t)$ are updated alternatively, rather than waiting for one quantity to converge before starting iterations of another. The following theorem shows that even with simultaneous updating, the message parameters $(v^a(t), \beta^a(t))$ converge under the same condition as if we wait for $v^a(t)$ to converge first.

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

**Proof:**

**Sufficient Condition:**

If $S_1 \neq \emptyset$, according to Theorem 1, it is known that $v^a(t)$ converges to the same point for all $v^a(0) \in \mathcal{A}$. Thus, in order to prove the sufficiency, we only need to prove $\beta^a(t)$ also converges to the same point for all $v^a(0) \in \mathcal{A}$ and $\beta^a(0) \in \mathbb{R}^{|\mathcal{E}|}$.

Due to $\rho(G^*) < 1$, we can infer that $I - G^*$ is invertible. Now, define the following vector
\[
\phi \triangleq (I - G^*)^{-1}b^*,
\] (31)
which can be equivalently written as $\phi = G^*\phi + b^*$. Subtracting it from (14) gives
\[
\beta^a(t + 1) - \phi = g(t)\beta^a(t) - G^*\phi + b(t) - b^*
\]
\[
= (G^* + (G(t) - G^*))\beta(t) - \phi
\]
\[
+ (G(t) - G^*)\phi + b(t) - b^*.
\] (32)

According to (32), for any matrix norm $\| \cdot \|$, we have
\[
\|\beta^a(t + 1) - \phi\| \leq (\|G^*\| + \|G(t) - G^*\|)\|\beta^a(t) - \phi\|
\]
\[
+ \|G(t) - G^*\|\phi + b(t) - b^*.
\] (33)

where $\| \cdot \|$ is the compatible vector norm associated with the matrix norm $\| \cdot \|$ [34, p. 297]. Denote $\rho(G^*)$ as $\rho$. It is known that there always exists a specific matrix norm $\| \cdot \|_s$ such that $|G^*|_s \leq \rho + \epsilon_s$ for any $\epsilon_s > 0$ [34]. Furthermore, due to $\|G(t) - G^*\| = G^*$ from Proposition 2, it can be seen that $\lim_{t \to \infty} G(t) = G^*$, meaning that there exists an integer $t_0$ such that $|G(t) - G^*|_s \leq \epsilon_G$ for all $t \geq t_0$ and any $\epsilon_G > 0$. Thus, for $t \geq t_0$, we can derive from (33) that
\[
\|\beta^a(t + 1) - \phi\| \leq (\rho + \epsilon_s + \epsilon_G)\|\beta^a(t) - \phi\|_s
\]
\[
+ \|G(t) - G^*\|\phi + b(t) - b^*\|_s,
\] (34)

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

Now, we define another sequence $z(t)$ for $t \geq t_0$ as
\[
z(t + 1) = (\rho + \epsilon_s + \epsilon_G)z(t)
\]
\[
+ (|G(t) - G^*|\phi + b(t) - b^*|)_s
\]
\[
= (\rho + \epsilon_s + \epsilon_G)^{t-t_0+1}z(t_0) + \sum_{t=t_0}^{t} (\rho + \epsilon_s + \epsilon_G)^{t-t_0} \times |G(t) - G^*|\phi + b(t) - b^*|_s.
\] (35)
where $z(t_0) \triangleq \| \beta^*(t_0) - \phi \|_s$. From (35), we further have
\[
z(2t_1 + 1) \leq (\rho + \varepsilon_s + \varepsilon_G) z(t_0) + \sum_{i=1}^{2t_1} (\rho + \varepsilon_s + \varepsilon_G)^{-t-1} z(2t_1 + 1) - b(t) + b^* \|_s.
\]
(36)
Since $\rho + \varepsilon_s + \varepsilon_G$ converges for any $\varepsilon_s$, according to (11), we must have $b(t)$ for all $t$, or equivalently in vector form
\[
(\rho + \varepsilon_s + \varepsilon_G)^{-t} b(t) + b^* \|_s.
\]
(37)
meaning that the entries in $\mathbf{A} v^*(t) + \mathbf{u} > 0$, all $t \geq 0$.

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

Subtracting (34) from (35) gives
\[
(\rho + \varepsilon_s + \varepsilon_G) (z(t) - \| \beta^*(t) - \phi \|_s) \leq z(t+1) - \| \beta^*(t+1) - \phi \|_s.
\]
(39)
Putting $t = t_1$ into (39), and since by definition $z(t_0) = \| \beta^*(t_0) - \phi \|_s$, it can be derived that $\| \beta^*(t_0+1) - \phi \|_s \leq z(t_0+1)$. In general, it can be recursively derived that $\| \beta^*(t) - \phi \|_s \leq z(t)$ for all $t \geq t_0$. Since $\lim z(t) = 0$, we have $\lim \| \beta^*(t) - \phi \|_s$ and $\beta^*(t)$ converges to the same point.

 Necessary Condition: Since $v^*(t)$ converges for all $v^*(0) \in \mathcal{A}$, according to Theorem 1, we have $S_1 \neq \emptyset$. Thus, to prove the necessity, we only need to prove $\rho (\mathbf{G}^*) < 1$ under the prerequisite of $S_1 \neq \emptyset$.

First, from Proposition 2 and 6, we know that $\lim_{i \to \infty} g_i(0) = w^*$. Together with $w^* < 0$ from (23), set $\mathcal{A}$ defined in (18) reduces to $\mathcal{A} = \{ w \geq w_0 | w \in \text{int}(S_1) \} \cup \{ w \geq w^* \}$. Thus, for any $v^*(0) \in \mathcal{A}$, we can always find a $\gamma \in \text{int}(S_1) \cup \{ w^* \}$ such that $\gamma \leq v^*(0)$. By applying $g(\cdot)$ to $\gamma \leq v^*(0)$, from the monotonically increasing property of $g(\cdot)$, we obtain
\[
g(\gamma) \leq v^*(1).
\]
(40)
If $\gamma \in \text{int}(S_1)$, it is seen from (17) that $\gamma < g(\gamma)$. On the other hand, if $\gamma = w^*$, then $\gamma = g(\gamma)$ since $w^*$ is the converged point. Thus, with $\gamma \in \text{int}(S_1) \cup \{ w^* \}$, we have $\gamma < g(\gamma)$. Combining with (40) gives $\gamma \leq v^*(1)$. By induction, it can be inferred that $\gamma \leq v^*(t)$ for $t \geq 0$. Due to $\gamma \in \text{int}(S_1) \cup \{ w^* \}$, from the definition of set $\mathcal{A}$ in (18), we have $v^*(t) \in \mathcal{A}$ for all $t \geq 0$.

Next, consider two initializations $(v^*(0), \beta^*(0))$ and $(v^*(0), \beta^*_0(0))$ with $v^*(0) \in \mathcal{A}$ and $\beta^*(0), \beta^*_0(0) \in \mathbb{R}^{|e|}$. Due to $\beta_i(t+1) = \mathbf{G}(t) \beta_i(t) + b(t)$ and $\beta_0^*_0(t+1) = \mathbf{G}(t) \beta^*_0(t) + b(t)$ as given in (14), subtracting the two equations gives
\[
\Delta \beta_i(t+1) = \mathbf{G}(t) \Delta \beta_i(t)
\]
(41)
where $\Delta \beta_i(t) = \beta_i(t) - \beta_0^*_0(t)$. Since $\beta_i(t)$ and $\beta_0^*_0(t)$ converge to the same point, then $\Delta \beta_i(t)$ converges to the zero vector for any $v^*(0) \in \mathcal{A}$ and $\Delta \beta_i(t) \in \mathbb{R}^{|e|}$. Thus, according to (41), we can infer $\lim_{t \to \infty} \prod_{k=0}^{t} \mathbf{G}(k) \Delta \beta_i(t) = 0$ for any $v^*(0) \in \mathcal{A}$ and $\Delta \beta_i(t) \in \mathbb{R}^{|e|}$. Similarly, for another $v^*(0) \in \mathcal{A}$, we also have $\lim_{t \to \infty} \prod_{k=0}^{t} \mathbf{G}(k) = 0$, where $\mathbf{G}(k) \triangleq \Delta \mathbf{A} \mathbf{v}^*(k) + \mathbf{u} \Delta \mathbf{A} \mathbf{v}^*(k) \mathbf{A}$ as defined in (15). Due to $v^*(t) \in \mathcal{A}$ for any $t \geq 0$ as proved above, we can choose $v^*(0) = v^*(t_1)$, which means that $\mathbf{G}(k) = \mathbf{G}(k + t_1)$. Substituting this result into $\lim_{t \to \infty} \prod_{k=0}^{t} \mathbf{G}(k) = 0$ for all $t \geq 0$. As $t_1 \to \infty$, using the fact $\lim_{t \to \infty} \mathbf{G}(k) = \mathbf{G}^*$, we obtain $\lim_{t \to \infty} \prod_{k=0}^{t} \mathbf{G}(k) = 0$, which holds if and only if $\rho (\mathbf{G}^*) < 1$.

IV. CONVERGENCE OF BELIEFS IN GAUSSIAN BP

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

A. Synchronous Gaussian BP

Theorem 3: In synchronous Gaussian BP, belief parameters $(\sigma^2(t), \mu(t))$ converge to the same point for all choices of $v^*(0) \in \mathcal{A}$ and $\beta^*(0) \in \mathbb{R}^{|e|}$ if and only if $S_1 \neq \emptyset$, $\rho (\mathbf{G}^*) < 1$ and $\mu_k + \sum_{k \in \mathcal{N}(i)} \omega_k^{(i)} \neq 0$.

Proof:

Sufficient Condition:
If $S_1 \neq \emptyset$, according to Proposition 2, it is known that $\mu_k + \lim_{t \to \infty} \sum_{k \in \mathcal{N}(i)} \omega_k^{(i)} \neq 0$ for all $v^*(0) \in \mathcal{A}$. Due to $\mu_k + \sum_{k \in \mathcal{N}(i)} \omega_k^{(i)} \neq 0$, we can infer that belief variance $\sigma^2(t) = \frac{1}{\mu_k + \sum_{k \in \mathcal{N}(i)} \omega_k^{(i)}}$ converges to $\frac{1}{\mu_k + \sum_{k \in \mathcal{N}(i)} \omega_k^{(i)}}$ for all $v^*(0) \in \mathcal{A}$.

Next, if $S_1 \neq \emptyset$ and $\rho (\mathbf{G}^*) < 1$, from Theorem 3, it can be inferred that $h_i + \lim_{t \to \infty} \sum_{k \in \mathcal{N}(i)} \beta_k^{(i)} \neq 0$ exists and is unique for all $v^*(0) \in \mathcal{A}$ and $\beta^*(0) \in \mathbb{R}^{|e|}$. Then, using the fact

\[
\sum_{t=0}^{\infty} \mathbf{G}(t) \Delta \beta_i(t)
\]
we can infer that belief mean $\mu(t)$ also converges to the same point for all $v^a(0) \in A$ and $\beta^a(0) \in \mathbb{R}_F$. Using the same arguments after (41), it can be inferred that

$$\lim_{t \to \infty} E \cdot \mathbf{G}^* t = 0.$$  

(44)

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

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

(45)

Writing (45) into a scalar form gives that

$$\sum_{k \in N(i)} e_{ki}^T \text{eig}_{\text{max}}(G^*) = 0,$$

or equivalently

$$\sum_{k \in N(i)} e_{ki}^T \text{eig}_{\text{max}}(G^*) = -e_j^T \cdot \text{eig}_{\text{max}}(G^*)$$  

(46)

Moreover, since $\sigma^T \sigma$ converges for any $v^a(0) \in A$, according to (11), we have $v^a(t) \in \mathcal{W}$ for $t \geq 0$. From the definition of $\mathcal{W}$ in (13), this means $p_i + \sum_{k \in N(i) \setminus \beta} v_{ki}^T(t) > 0$ for all $i, j \in E$, or equivalently $v_{ki}^T(t) > -p_{ij} - \sum_{k \in N(i) \setminus \beta} v_{ki}^T(t)$. Using $v^a(t) \leq v^a_0(t)$ for $t \geq 2$ from (42), we can infer that $v_{ki}^T(t) > p_{ij}$ for all $i, j \in E$. Combining with $v^a(t) \leq v^a(t)$ from (42), we can see that $v^a(t)$ is lower bounded. Together with the monotonically decreasing property of $v^a(t)$ from (42), we can infer that $v^a(t)$ converges, and thereby $g \left( \lim_{t \to \infty} v^a(t) \right) = \lim_{t \to \infty} v^a(t)$. Furthermore, since $v^a(t)$ converges from (42) and $v^a(t)$ is in $\mathcal{W}$, we can infer from the definition of $\mathcal{W}$ in (13) that $v^a(t)$ is in $\mathcal{W}$ for all $t \geq 0$, and thereby $\lim_{t \to \infty} v^a(t) \in \mathcal{W}$. Together with

$$g \left( \lim_{t \to \infty} v^a(t) \right) = \lim_{t \to \infty} v^a(t),$$

it can be seen from the definition of $S_1$ in (17) that $\lim_{t \to \infty} v^a(t) \in S_1$, and hence $S_1 \neq \emptyset$.

Second, due to $S_1 \neq \emptyset$, according to Proposition 2, it is known that $\lim_{t \to \infty} v^a(t) = \text{eig}_{\text{max}}(G^*)$ for all $v^a(0) \in A$, and thus we have $p_i + \sum_{k \in N(i) \setminus \beta} v_{ki}^T(t) = p_i + \sum_{k \in N(i) \setminus \beta} w_{ki}^T$. Since $\sigma^2(T) = \frac{1}{p_i + \sum_{k \in N(i) \setminus \beta} v_{ki}^T(t)}$ converges for $v^a(0) \in A$, we can infer that $p_i + \sum_{k \in N(i) \setminus \beta} w_{ki}^T \neq 0$.

Finally, we prove $\rho(G^*) < 1$. Due to $S_1 \neq \emptyset$, then $p_{ij} + \lim_{t \to \infty} \sum_{k \in N(i) \setminus \beta} v_{ki}^T(t) = p_{ij} + \sum_{k \in N(i) \setminus \beta} w_{ki}^T$, for all $v^a(0) \in A$. Thus, if $\mu(t)$ converges to the same point, we can infer that $\lim_{t \to \infty} \sum_{k \in N(i) \setminus \beta} \beta_{ki}(t)$, or equivalently $\lim_{t \to \infty} E \cdot \beta^a(t)$ is unique for all $v^a(0) \in A$ and $\beta^a(0) \in \mathbb{R}_F$, where $E$ is an $N \times E$ matrix such that the $i$-th column of $E \cdot \beta^a(t)$ is equal to $\sum_{k \in N(i) \setminus \beta} \beta_{ki}(t)$.

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

$$\lim_{t \to \infty} E \cdot \Delta \beta^a(t) = 0$$  

(43)

for all $v^a(0) \in A$ and $\Delta \beta^a(0) \in \mathbb{R}_F$, where $\Delta \beta^a(t) = \beta^a_{ki}(t) - \beta^a_{ki}(t)$. Substituting $\Delta \beta^a(t) = \sum_{k \in N(i) \setminus \beta} G(k) \cdot \Delta \beta^a(0)$ from (41) into (43) gives $\lim_{t \to \infty} E \cdot \prod_{k=0}^{t-1} G(k) \cdot \Delta \beta^a(0) = 0$ for all $v^a(0) \in A$ and $\Delta \beta^a(0) \in \mathbb{R}_F$. The same arguments after (41), it can be inferred that

$$\lim_{t \to \infty} E \cdot \mathbf{G}^* t = 0.$$  

(44)

It is known that if the variance $\sigma^2(t)$ converges, the converged variance $\lim_{t \to \infty} \sigma^2(t) = \frac{1}{p_i + \sum_{k \in N(i) \setminus \beta} w_{ki}} > 0$. Thus, we have $p_i + \sum_{k \in N(i) \setminus \beta} w_{ki} > 0$, or equivalently $p_i + \sum_{k \in N(i) \setminus \beta} w_{ki} - w_{ki} > 0$. Together with (49), it can be obtained that $\lambda_{\text{max}}(G^*) = -\frac{w_{ki}}{p_i + \sum_{k \in N(i) \setminus \beta} w_{ki}} < 1$, which
contradicts with the assumption that \( \rho(G^*) \geq 1 \). Thus, we have \( \rho(G^*) < 1 \).

B. Damped Gaussian BP

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

\[
\beta^\theta(t + 1) = (1 - d) \cdot \beta^\theta(t) + d \cdot \{G(t) \cdot \beta^\theta(t) + b(t)\},
\]

(50)

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

**Theorem 4:** In damped Gaussian BP, belief parameters \( \{\sigma^2_i(t), \mu_i(t)\} \) converge to the same point for all choices of \( \nu^\theta(0) \in A \) and \( \beta^\theta(0) \in \mathbb{R}^E \) under a nonzero damping factor \( d \) if and only if the three conditions hold: 1) \( S_1 \neq \emptyset \); 2) \( \max_{\lambda(G^*)} \mathbb{R}(\lambda(G^*)) < 1 \) or \( \min_{\lambda(G^*)} \mathbb{R}(\lambda(G^*)) > 1 \); 3) \( p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{i_k} \neq 0 \). Furthermore, the damping factor \( d \) should be chosen as

\[
d \in \left\{ \begin{array}{ll}
0, & \text{if } \max_{\lambda(G^*)} \mathbb{R}(\lambda(G^*)) < 1 \\
\frac{1 - \mathbb{R}(\lambda(G^*))}{1 - \lambda(G^*)}, & \text{if } \min_{\lambda(G^*)} \mathbb{R}(\lambda(G^*)) > 1.
\end{array} \right.
\]

(51)

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

\[
\beta^\theta(t + 1) = (1 - d) \cdot \mathbf{I} + d \cdot G(t) \cdot \beta^\theta(t) + d \cdot b(t),
\]

(52)

which is in the same form as (14) in Gaussian BP. By applying Theorem 2, it can be inferred that belief parameters \( \{\sigma^2_i(t), \mu_i(t)\} \) of damped Gaussian BP converge for all choices of \( \nu^\theta(0) \in A \) and \( \beta^\theta(0) \in \mathbb{R}^E \) if and only if \( S_1 \neq \emptyset \); \( \rho(1 - d) \cdot \mathbf{I} + d \cdot G^* < 1 \); and \( p_{ii} + \sum_{k \in \mathcal{N}(i)} w_{i_k} \neq 0 \).

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

\[
\max_{\lambda(G^*)} \left\{ 1 - \lambda(G^*) \cdot d^2 - 2(1 - \mathbb{R}(\lambda(G^*))) \cdot d + 1 \right\} < 1.
\]

(53)

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

\[
0 < d < 2 \cdot \frac{1 - \mathbb{R}(\lambda(G^*))}{|1 - \lambda(G^*)|},
\]

(54)

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

\[
2 \cdot \frac{1 - \mathbb{R}(\lambda(G^*))}{|1 - \lambda(G^*)|} < d < 0.
\]

(55)

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

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

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

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

**Proof:** Due to \( \rho(G^*) = \max_{\lambda(G^*)} \mathbb{R}(\lambda(G^*)) \), if the convergence condition of beliefs in Gaussian BP \( \rho(G^*) < 1 \) holds, we have \( \max_{\lambda(G^*)} \mathbb{R}(\lambda(G^*)) < 1 \), and thereby \( \rho(G^*; d = 1) = \max_{\lambda(G^*)} \mathbb{R}(\lambda(G^*)) < 1 \). That is, the convergence condition of beliefs in damped Gaussian BP holds.

Furthermore, from (56), we know that \( \lambda(G^*) \) is not empty for all \( \lambda(G^*) \). Adding \( 1 - 2 \cdot \mathbb{R}(\lambda(G^*)) \) on both sides of this equation gives \( 1 - 2 \cdot \mathbb{R}(\lambda(G^*)) + |\lambda(G^*)|^2 < 2 - 2 \cdot \mathbb{R}(\lambda(G^*)) \), or equivalently \( \frac{1 - \mathbb{R}(\lambda(G^*))}{1 - |\lambda(G^*)|^2} > 1 \). Since this holds for all \( \lambda(G^*) \), we can infer that \( \min_{\lambda(G^*)} 2 \cdot \frac{1 - \mathbb{R}(\lambda(G^*))}{1 - \lambda(G^*)} > 1 \). Comparing this with the first case of (51), it is obvious that \( d = 1 \) is in the feasible range.

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

C. Asynchronous Gaussian BP

Modified from the synchronous updating of \( \nu^\theta(t) \) in (11) and \( \beta^\theta(t) \) in (14), under asynchronous scheduling, \( \nu^\theta_{i,j}(t) \) and \( \beta^\theta_{i,j}(t) \) for \( t \in T_{i,j} \) are updated as

\[
\nu^\theta_{i,j}(t + 1) = -\frac{p_{ij}}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} \nu^\theta_{k \rightarrow i}(\tau^\theta_{k \rightarrow i}(t))},
\]

(57)

\[
\beta^\theta_{i,j}(t + 1) = -\frac{p_{ij}}{p_{ii} + \sum_{k \in \mathcal{N}(i) \setminus j} \beta^\theta_{k \rightarrow i}(\tau^\theta_{k \rightarrow i}(t))},
\]

(58)

where \( T_{i,j} \) is the set of time instants at which \( \nu^\theta_{i,j}(t) \) and \( \beta^\theta_{i,j}(t) \) are updated and \( \tau^\theta_{i,j}(t) \) satisfies the totally asynchronous scheduling defined below.
Definition 1. (Totally Asynchronous Scheduling) [30]: The sets $\mathcal{T}_{i,j}$ are infinite, and $\tau_{i,j}(t)$ satisfies $0 \leq \tau_{i,j}(t) \leq t$ and $\lim_{t \to \infty} \tau_{i,j}(t) = -\infty$ for all $(i,j) \in \mathcal{E}$. From the above definition, the updating of each message is executed independently at time instants $t \in \mathcal{T}_{i,j}$ with $(i,j) \in \mathcal{E}$. Moreover, at each updating time instant, only the previously available information $\mathbf{v}^{k-1}_{i,j}(t)$ is needed. Now, the following theorem can be presented.

Theorem 5: In asynchronous Gaussian BP, if $\rho(\mathbf{G}^*) < 1$ and $\rho(\mathbf{G}^*) < 1$, belief parameters converge to the same point for all choices of $\rho(\mathbf{G}^*)$ and $\rho(\mathbf{G}^*)$. Theorem 6: If $\rho(\mathbf{P} - 1) < 1$, then $\mathbf{S}_1 \neq \emptyset$, and $\rho(\mathbf{G}^*) < 1$. A relaxed initialization set is also proposed in [23] under the convergence condition of convex decomposition, which is proved to be equivalent to walk-summarizability [24]. The initialization set in [23] is defined as $D = \{ \mathbf{v}^a(0) \mid -\text{diag}^{-1}(\mathbf{P})\mathbf{v}^a(0) \geq -\mathbf{A} \mathbf{v}^a(0) \}$ for any $\mathbf{A} \in \mathbb{R}$. Convex decomposition requires the existence of $\theta_{ij}$ for all $(i,j) \in \mathcal{E}$ satisfying the two conditions:

\[
\begin{align*}
\theta_{ij} \mathbf{p}_{ij}^2 + \mathbf{p}_{ij} \mathbf{p}_{ij}^2 &> 0, \forall (i,j) \in \mathcal{E}, \quad (60) \\
1 - \sum_{k \in \mathcal{N}(i) \cup \mathcal{N}(j)} \mathbf{p}_{ik}^2 \mathbf{p}_{kj} &> 0, \forall i \in \mathcal{V}. \quad (61)
\end{align*}
\]

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

\[
\theta_{ij} > 0, \quad (62)
\]

Further, applying $\mathbf{p}_{ij}^2 \theta_{ij} \geq \frac{1}{\theta_{ij}}$ to (62), we also obtain

\[
-\mathbf{p}_{ij}^2 \theta_{ij} < -\frac{1}{1 + \sum_{k \in \mathcal{N}(i) \cup \mathcal{N}(j)} \mathbf{p}_{ik}^2 \mathbf{p}_{kj}}, \quad (65)
\]

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

Furthermore, it is proved in [23, Lemma 3] that if $\rho(\mathbf{P} - 1) < 1$, we must have $\rho(\mathbf{G}^*) < 1$. At last, since $\rho(\mathbf{P} - 1) < 1$ guarantees belief variance converge to $\mathbf{I}$ in $[24]$, we must have $1 + \sum_{k \in \mathcal{N}(i) \cup \mathcal{N}(j)} \mathbf{p}_{ik}^2 \mathbf{p}_{kj} \neq 0$.
the definition of $\mathcal{A} = \{ w \geq 0 \} \cup \{ w \geq w_0 | w_0 \in \text{int}(\mathcal{S}_I) \} \cup \{ w \geq w_0, w_0 \in \mathcal{S}_I \text{ and } w_0 = \lim_{t \to \infty} g^{(t)}(0) \}$ in (18), it is obvious that $\mathcal{D}$ is a subset of $\mathcal{A}$.

VI. NUMERICAL EXAMPLES

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

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

(66)

where $\zeta$ is a coefficient indicating the correlation strength among variables; and $a_k$ is the $k$th element of the vector $a = [0.13, 0.10, 0.71, -0.05, 0, 0.12, 0.07, 0.11, -0.02, -0.03]^{T}$. The varying of correlation strength $\zeta$ induces a series of matrices, and the positive definite constraint $P > 0$ can be guaranteed when $\zeta < 0.5978$. Furthermore, numerical calculation shows that when $\zeta < 0.5859$, we have $S_1 \neq \emptyset$ and $p_{ii} + \sum_{k \in N(i)} w_{ki} \neq 0$ by solving the SDP problem in (19). Thus, when $\zeta \leq 0.5859$, to determine the convergence of beliefs, we only need to check whether $\rho(G^{*}), \max_{\lambda(G^{*})} \Re[\lambda(G^{*})]$ and $\rho(G^{*})$ are satisfied for synchronous Gaussian BP, damped Gaussian BP and asynchronous Gaussian BP, respectively.

Fig. 2 illustrates how the spectral radii $\rho(\{P\} - 1)$, $\rho([G^{*}]), \rho(G^{*})$ and the maximum real part of eigenvalues $\max_{\lambda(G^{*})} \Re[\lambda(G^{*})]$ vary as a function of $\zeta$. The curves are plotted for $\zeta \leq 0.5859$ which guarantees $S_1 \neq \emptyset$ and $p_{ii} + \sum_{k \in N(i)} w_{ki} \neq 0$. In Fig. 2, three critical values can be observed: $\zeta_a = 0.3945, \zeta_b = 0.4621$ and $\zeta_c = 0.5859$, which are labelled as A, B and C, respectively. It can be seen that $\zeta < \zeta_a < \zeta_b < \zeta_c$ are the necessary and sufficient conditions for $\rho(G^{*}) < 1$, $\rho(G^{*}) < 1$ and $S_1 \neq \emptyset$, respectively. Then, according to Theorem 5, if $\zeta < \zeta_a$, beliefs converge for all initialization $w^{*}(0) \in \mathcal{A}$ and $\beta^{*}(0) \in \mathbb{R}$. In Fig. 2, the curves reveal the necessary and sufficient conditions of beliefs in synchronous and damped Gaussian BP, respectively. Lastly, from Fig. 2, it is obvious that $\rho(\{P\} - 1) < 1$, then $\rho([G^{*}]) < 1$. This numerically corroborates the claim in Theorem 6 that walk-summability is implied by the convergence conditions proposed in this paper. Next, we inspect the convergence behaviors around the three special points A, B and C under two different initializations: $w^{*}(0) = 0, \beta^{*}(0) = 0$ and $w^{*}(0) = w^{*}, \beta^{*}(0) = 1 \times 0.01$, where $w^{*}$ is the optimal solution of (19) with $\zeta = 0.5858$.

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

Second, consider $\zeta = 0.5858$ and 0.5860, which are chosen to be slightly smaller and larger than $\zeta_c = 0.5859$, respectively.

When $\zeta = 0.5858$ is slightly smaller than $\zeta_c = 0.5859$, Fig. 4(a) illustrates the convergence of $\mu_1(t)$ in damped Gaussian BP with damping factor $d = 0.61$, which locates within the required range $d \in (0, 0.6201)$ computed from (51). On the other hand, when the damping factor $d = 0.63$, which is outside of the range (0, 0.6201), Fig. 4(b) shows that belief mean $\mu_1(t)$ does not converge. Moreover, when $\zeta = 0.5860$ which is slightly larger than $\zeta_c = 0.5859$, Fig. 4(b) illustrates that belief mean $\mu_1(t)$ cannot converge neither, even damping is applied. Obviously, Fig. 4 corroborates the necessary and sufficient convergence condition of damped Gaussian BP as well as the proposed range of damping factor in Theorem 4.

Finally, consider $\zeta = 0.3900$, which is slightly smaller than the critical value $\zeta_a = 0.3945$. To demonstrate the effects of asynchronous update, two schedulings with 2% and 10% chances of missing the exchanged messages are considered. When missing rate is 2%, Fig. 5(a) shows that the belief mean
Fig. 4. Illustration of convergence and divergence of belief mean \( \mu_t \) in damped Gaussian BP under different initializations with \( \zeta \) chosen around the critical value \( \zeta^* = 0.5855 \). (a) Convergence of \( \mu_t \) of damped Gaussian BP with \( \zeta = 0.5855 \). (b) Divergence of \( \mu_t \) of damped Gaussian BP.

Fig. 5. Illustration of convergence of belief mean \( \mu_t \) under different initializations and message missing rates with \( \zeta \) chosen smaller than the critical value \( \zeta^* \). (a) Convergence of \( \mu_t \) of asynchronous Gaussian BP with \( \zeta = 0.3594 \) and message missing rate equal to 10\%. (b) Convergence of \( \mu_t \) of asynchronous Gaussian BP with \( \zeta = 0.3594 \) and message missing rate equal to 20\%.

\[ \mu_1(t) \] converges to the same value under the two different initializations. Moreover, the convergence of \( \mu_1(t) \) under the asynchronous scheduling with message missing rate equal to 10\% is also observed in Fig. 5(b). Obviously, Fig. 5 verifies the results in Theorem 5.

VII. CONCLUSIONS

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

APPENDIX A

PROOF OF THEOREM 5

Under totally asynchronous scheduling, due to \( S_1 \neq \emptyset \), it is known from Theorem 1 that \( \psi^a(t) \) converges to the same point for all \( t \). Thus, if \( S_1 \neq \emptyset \) and \( \mu^a + \sum_{k \in N(i)} \psi_{k}^a(\theta) \neq 0 \), we can infer that belief variance \( \sigma^2 = \frac{1}{p_t + \sum_{k \in N(i)} \psi_{k}^a(\theta)} \) converges to a unique value, to prove that the belief mean \( \mu^a(t) \) converges to a unique value under totally asynchronous scheduling for all \( \beta^a(0) \). First, define a sequence of sets \( \mathcal{V}(m) = \{ w | \mathbf{v}_i^a(m) \leq w \leq \mathbf{v}_i^a(m) \} \) with \( \mathbf{v}_i^a(0) \) and \( \mathbf{v}_i^a(0) \geq 0 \). Due to \( \mathbf{v}_i^a(0) \in \text{int}(S_1) \), from the definition of \( S_1 \) in (17), we have \( \mathbf{v}_i^a(0) \leq g(\mathbf{v}_i^a(0)) \), or equivalently \( \mathbf{v}_i^a(0) \leq \mathbf{v}_i^a(1) \). From the monotonically increasing property of \( g_i^a(\cdot) \) in (24), we can infer that \( g(\mathbf{v}_i^a(0)) \leq g(\mathbf{v}_i^a(1)) \), or equivalently \( \mathbf{v}_i^a(1) \leq \mathbf{v}_i^a(2) \). By induction, we have \( \mathbf{v}_i^a(m) \leq \mathbf{v}_i^a(m+1) \) for all \( m \geq 0 \). On the other hand, due to \( \theta = 0 \), substituting it into (12) gives \( \mathbf{v}(1) < 0 \), and thereby \( \mathbf{v}(1) < \mathbf{v}(0) \). From the monotonically increasing property of \( g_i^a(\cdot) \) in (24), we can infer that \( g(\mathbf{v}(1)) \leq g(\mathbf{v}(0)) \), or equivalently \( \mathbf{v}(2) \leq \mathbf{v}(1) \). By induction, we have \( \mathbf{v}(m+1) \leq \mathbf{v}(m) \) for all \( m \geq 0 \). Combining with \( \mathbf{v}(m) \leq \mathbf{v}(m+1) \), we can infer that \( \mathcal{V}(m+1) \subseteq \mathcal{V}(m) \). It is known from Theorem 1 that \( \mathbf{v}_i^a(m) \) and \( \mathbf{v}_i^a(m) \) both converge to \( \mathbf{w}^* \), thus we have the set \( \mathcal{V}(m) \) also converges to a single point \( \mathbf{w}^* \).

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

\[ \varphi(\gamma, \beta^a) = G(\gamma) \beta^a + b(\gamma), \]

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

\[ B(m) = \{ \beta^a | \| \beta^a - \phi \|_\infty \leq \eta_m \} \]

where \( \| \cdot \|_\infty \) is the weighted maximum norm with weighting \( \phi \). Define in (31):

\[ \eta_m \triangleq \max_{\gamma \in \mathcal{V}(m-1)} \left\{ \| G(\gamma) \|_\infty \eta_{m-1} + \varphi(\gamma) \right\} \]

with \( \varphi(\gamma) \triangleq \| G(\gamma) - G^* \|_\infty \phi + b(\gamma) - b^* \|_\infty \), and \( \| \cdot \|_\infty \) is the induced matrix norm of \( \| \cdot \|_\infty \). To define the set \( B(m) \), the initial \( \eta_0 \) could be any positive value such that the chosen initialization \( \beta^a(0) \) lies within the set \( B(0) \).
Due to the maximum weighted norm $\| \cdot \|_\infty$ in (68), set $B(m)$ can always be represented as the product of subsets of individual components, and thereby the box condition [30, p. 431] is satisfied. Thus, to prove the convergence of $\beta^*(t)$, we only need to prove that $B(m)$ satisfies the following three conditions: 1) $B(m + 1) \subseteq B(m)$ for all $m$ larger than some positive integer; 2) $\psi(\gamma, \beta^*) \in B(m + 1)$ for any $\gamma \times \beta^* \in V(m) \times B(m)$; 3) $B(m)$ converges to $\phi$ [30, p. 431].

1) Due to $\rho(1) \phi^* - \psi < 1$, there exists some weighted maximum matrix norm $\| \cdot \|_{\infty}$ such that $\| G(\gamma) \|_\infty < 1$ [30]. Since $V(m)$ converges to $\omega^*$, there exists a $m_0$ such that $\| G(\gamma) \|_\infty < 1$ for all $\gamma \in V(m_0)$. On the other hand, for any $\gamma \in V(m)$, due to $V(0) \subseteq V(0)$ and $V(0) \subseteq A$, it is known that $\gamma \in A$. From (37), we have $A \gamma + u > 0$, and $d^{-1}(A \gamma + u)$ exists. Then, elements in $G(\gamma)$ and $b(\gamma)$ are finite, and thus $\gamma = \{(G(\gamma) - \gamma \phi + b(\gamma) - \gamma b^* \|_\infty \}$ must be finite for all $\gamma \in V(m_0)$. Combining with $\| G(\gamma) \|_\infty < 1$ for all $\gamma \in V(m)$, we can infer that $\| G(\gamma) \|_\infty < 1$ for all finite $\gamma$. Thus, by choosing $\eta_0$ large enough, we can always ensure $\beta^*(0) \in B(t)$ and $\eta_m \geq \max_{\gamma \in V(m)} \| G(\gamma) \|_\infty \eta_m + \max_{\gamma \in V(m)} \| \gamma \|_\infty \leq \eta_m$. Combining with the fact $\gamma(\gamma) \|_\infty \leq \max_{\gamma \in V(m)} \| G(\gamma) \|_\infty \eta_m + \max_{\gamma \in V(m)} \| \gamma \|_\infty \| \gamma \|_\infty \| \gamma \|_\infty$, we can infer from the definition of $\eta_{m+1}$ in (69) that $\eta_{m+1} \leq \eta_m$. Now, suppose $\eta_{m+1} \leq \eta_m$ for some $m \geq m_0$. Applying this to $\eta_{m+2} = \max_{\gamma \in V(m+1)} \{\| G(\gamma) \|_\infty \eta_m + \| \gamma \|_\infty \eta_m + \| \gamma \|_\infty \}$. Due to $V(m + 1) \subseteq V(m)$, it can be further inferred that $\eta_{m+2} \leq \max_{\gamma \in V(m+1)} \{\| G(\gamma) \|_\infty \eta_m + \| \gamma \|_\infty \eta_m + \| \gamma \|_\infty \}. According to the definition of $\eta_{m+1}$ in (69), we obtain $\eta_{m+2} \leq \eta_{m+1}$. Hence, $\eta_{m+1} \leq \eta_m$ for all $m \geq m_0$. Thus, the first condition is proved.

2) From (31), it is known that $\phi = G(\gamma) \mathbf{1} + b^*$. Subtracting it from (67) and rearranging the terms gives $\phi(\gamma, \beta^*) - \phi = G(\gamma) (\beta^* - \phi(\gamma, \beta^*) - \mathbf{1} (G(\gamma) - \phi(\gamma, \beta^*) - \mathbf{1} b^*)$. Taking the norm on both sides of this equality leads to $\| \phi(\gamma, \beta^*) - \phi \|_\infty \leq \| G(\gamma) \|_\infty \| \beta^* - \phi \|_\infty + \| \gamma \|_\infty (\gamma)$. (70)

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

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

**APPENDIX B PROOF OF PROPOSITION 4**

First, if $S_1 \neq \emptyset$, for any $c \in S_1$, according to the definition of $S_1$ in (17), it is known that $c_{ij} \leq -p_{ij}^* + \frac{\sum_{k \in E(i) \setminus j} c_{ik}}{p_{ij}^*}$ and $p_{ii} + \sum_{k \in E(i) \setminus j} c_{ik} > 0$ for all $(i, j) \in E$, which are equivalent to $\frac{p_{ij}^*}{p_{ij}^*} \leq -\frac{p_{ij}^* + (1 + \sum_{k \in E(i) \setminus j} c_{ik})}{p_{ij}^*}$ and $p_{ii} > 0$ for all $(i, j) \in E$, where $\bar{p}_{ij} = \frac{p_{ij}^*}{\sqrt{p_{ii}^*}}$ due to $\bar{P} = diag^{-\frac{1}{2}}(P)Pdiag^{-\frac{1}{2}}(P)$. From (17), it can be seen that $c \in S_1$, where $c$ contains elements $c_{ij}$ for $(i, j) \in E$ arranged in ascending order first on $j$ and then on $i$. Thus, we have $S_1 \neq \emptyset$. Similarly, we can also prove that if $S_1 \neq \emptyset$, then $S_1 \neq \emptyset$.

Second, due to $S_1 \neq \emptyset$ and thereby $S_1 \neq \emptyset$, the optimal solution $w^*$ and $\hat{w}^*$ of (19) exist, and thereby $G(\gamma)$ and $\gamma$ defined in (29) exist. For any $\mu \rightarrow \infty$, we can obtain from (11) that $w_{ij}^* = -\frac{p_{ij}^*}{p_{ij}^* + \sum_{k \in E(i) \setminus j} c_{ik}}$ gives $\hat{w}_{ij}^* = -\frac{p_{ij}^*}{p_{ij}^* + \sum_{k \in E(i) \setminus j} c_{ik}}$ for all $t \geq 0$. Dividing $p_{ij}$ on both sides of $w_{ij}^* = -\frac{p_{ij}^*}{p_{ij}^* + \sum_{k \in E(i) \setminus j} c_{ik}}$, we can infer that $\hat{w}_{ij}^* = -\frac{p_{ij}^*}{p_{ij}^* + \sum_{k \in E(i) \setminus j} c_{ik}}$. Comparing this result to $\hat{w}_{ij}^* = -\frac{p_{ij}^*}{p_{ij}^* + \sum_{k \in E(i) \setminus j} c_{ik}}$, we can infer that $\hat{w}_{ij}^* = -\frac{p_{ij}^*}{p_{ij}^* + \sum_{k \in E(i) \setminus j} c_{ik}}$ for all $(i, j) \in E$. By using (71), we have $1 + \sum_{k \in E(i) \setminus j} \hat{w}_{ik}^* - \frac{1}{p_{ii}^*} (p_{ii} + \sum_{k \in E(i) \setminus j} \hat{w}_{ki}^*)$. Writing it into a vector form gives $A \hat{w}^* + 1 = diag^{-1}(u)(A \hat{w}^* + u)$. (72)

where $u$ is defined after (16). On the other hand, due to $\bar{p}_{ij} = \frac{p_{ij}^*}{\sqrt{p_{ii}^*}}$, writing it into a vector form, we obtain $\bar{p} = diag^{-\frac{1}{2}}(u)diag^{-\frac{1}{2}}(u)p$. (73)

where the vector $\bar{p}$ contains $\bar{p}_{ij}$ with $(i, j) \in E$ arranged in ascending order first on $j$ and then on $i$; and $\bar{u} = [p_{11}^*1_{N(1)} ; p_{22}^*1_{N(2)} ; \ldots ; p_{NN}^*1_{N(N)}]^T$ denotes the all ones vector of length $|V(i)|$. Substituting (72) and (73) into the definition of $G(\gamma)$, we obtain $G(\gamma) = -diag^{-1}(A \hat{w}^* + 1)diag^{-1}(p)A$. (74)

where the second step is due to the fact that positions of diagonal matrices can be interchanged. It can be verified that $diag^{-\frac{1}{2}}(u)G(\gamma) = G(\gamma)diag^{-\frac{1}{2}}(u)$, thus we further have $\hat{\gamma}^* = diag^{-\frac{1}{2}}(u)G^*diag^{-\frac{1}{2}}(u)$. (75)
Obviously, $\mathbf{G}^*$ and $\tilde{\mathbf{G}}^*$ are similar matrices. From basic properties of similar matrices, we have $\rho(\mathbf{G}^*) = \rho(\tilde{\mathbf{G}}^*)$. On the other hand, taking absolute value of elements on both sides of (75) gives

$$
|\tilde{\mathbf{G}}^*| = \text{diag} \left( \frac{1}{2} (|\mathbf{u}|) (|\mathbf{G}^*| \text{diag} \tau (|\mathbf{u}|)) \right).
$$

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

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

**REFERENCES**


