Journal of Machine Learning Research 20 (2019) 1-30 Submitted 3/18; Revised 9/19; Published 10/19

Convergence of Gaussian Belief Propagation Under General Pairwise Factorization: Connecting Gaussian MRF with Pairwise Linear Gaussian Model

Bin Li binli@eee.hku.hk

Yik-Chung Wu ycwu@eee.hku.hk

Department of Electrical and Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong

Editor: Gal Elidan

Abstract

Gaussian belief propagation (BP) is a low-complexity and distributed method for computing the marginal distributions of a high-dimensional joint Gaussian distribution. However, Gaussian BP is only guaranteed to converge in singly connected graphs and may fail to converge in loopy graphs. Therefore, convergence analysis is a core topic in Gaussian BP. Existing conditions for verifying the convergence of Gaussian BP are all tailored for one particular pairwise factorization of the distribution in Gaussian Markov random field (MRF) and may not be valid for another pairwise factorization. On the other hand, convergence conditions of Gaussian BP in pairwise linear Gaussian model are developed independently from those in Gaussian MRF, making the convergence results highly scattered with diverse settings. In this paper, the convergence condition of Gaussian BP is investigated under a general pairwise factorization, which includes Gaussian MRF and pairwise linear Gaussian model as special cases. Upon this, existing convergence conditions in Gaussian MRF are extended to any pairwise factorization. Moreover, the newly established link between Gaussian MRF and pairwise linear Gaussian model reveals an easily verifiable sufficient convergence condition in pairwise linear Gaussian model, which provides a unified criterion for assessing the convergence of Gaussian BP in multiple applications. Numerical examples are presented to corroborate the theoretical results of this paper.

Keywords: Gaussian belief propagation, convergence analysis, Gaussian Markov random field, pairwise linear Gaussian model, pairwise factorization

1. Introduction

Due to the distributed and parallel computation mechanisms, Gaussian belief propagation (BP) (Weiss and Freeman, 2000) is a low-complexity method for computing the marginal distributions of a high-dimensional joint Gaussian distribution, especially for those with a sparse information matrix. It is known that Gaussian BP is guaranteed to obtain the exact marginal distributions in singly connected or acyclic graphs. For loopy graphs (Kschischang et al., 2001), if Gaussian BP converges, the exact means of the marginal distributions can be obtained (Weiss and Freeman, 2000). These properties make Gaussian BP popular in many applications, ranging from consensus propagation (Moallemi and Roy, 2006), peer-ro-peer rating (Bickson et al., 2007), multi-user data detection (Bickson et al., 2008), solving systems of linear equations (Shental et al., 2008), quadratic minimization (Moallemi and Roy, 2009; Ruozzi and Tatikonda, 2013), to network synchronization (Leng and Wu, 2011; Du and Wu, 2013).

Despite the apparent advantages and popularity, a major hurdle in applying Gaussian BP is that it may not converge in loopy graphs. Consequently, deriving Gaussian BP convergence conditions is one of the major themes in Gaussian BP research. For Gaussian BP running in Gaussian Markov random field (MRF) (Rue and Held, 2005), diagonal dominance (Weiss and Freeman, 2000), walk-summability (Malioutov et al., 2006), pairwise-normalizability (Malioutov et al., 2006) and convex decomposition (Moallemi and Roy, 2009) are well-known sufficient convergence conditions under both synchronous and totally asynchronous schedulings (Bertsekas and Tsitsiklis, 1989). For these sufficient conditions, belief means and variances converge upon the same condition, which has been proved to be unnecessary recently (Su and Wu, 2014, 2015a,b). In particular, the convergence condition of belief variances was proved to be looser than that of belief means (Su and Wu, 2015a). Due to the separate treatment of belief means and variances, the necessary and sufficient convergence condition of Gaussian BP under synchronous scheduling was proposed in Su and Wu (2015a).

On the other hand, under pairwise linear Gaussian model, the convergence of Gaussian BP has been investigated in a case-by-case basis (Moallemi and Roy, 2006; Bickson et al., 2007, 2008; Leng and Wu, 2011; Du and Wu, 2013). In Moallemi and Roy (2006), consensus propagation for distributed averaging was viewed as an asynchronous implementation of Gaussian BP, where the convergence was shown to be guaranteed. In Bickson et al. (2007), Gaussian BP is used to solve the minimization of a quadratic cost function in peer-to-peer rating, where the convergence of Gaussian BP is shown by numerical results. In Bickson et al. (2008), multiuser detection problem is solved by Gaussian BP, where the convergence is checked by the diagonal dominance in the equivalent information matrix. Furthermore, in distributed synchronization applications (Leng and Wu, 2011; Du and Wu, 2013), Gaussian BP is guaranteed to converge under a reference node with perfect prior information.

Obviously, the understanding on the convergence conditions of Gaussian BP is far from complete, as the current results are scattered and derived under different assumptions. For example, existing convergence conditions for Gaussian MRF are derived under different pairwise factorizations of the joint Gaussian distribution. Since different factorizations lead to different Gaussian BP messages, it is not clear if the convergence condition obtained under one pairwise factorization is valid under another pairwise factorization. Another missing link is between the convergence results in Gaussian MRF and that of pairwise linear Gaussian model. On one hand, Bickson et al. (2008) advocate transforming the linear Gaussian model into Gaussian MRF, and then leveraging on the existing convergence condition for determining the convergence of Gaussian BP. On the other hand, in Moallemi and Roy (2006), Leng and Wu (2011), and Du and Wu (2013), Gaussian BP was proved to be convergent without using existing sufficient convergence conditions from Gaussian MRF. One may wonder if there is any special characteristic of pairwise linear Gaussian model that facilitates the Gaussian BP convergence.

To provide a more complete picture, the convergence of Gaussian BP is revisited but starting from a general pairwise factorization of the joint Gaussian distribution. This general pairwise factorization covers all possible pairwise factorizations with the exponents of factors being quadratic polynomial, and it includes all the existing pairwise factorizations used in Gaussian MRF and pairwise linear Gaussian model as special cases. By establishing the convergence conditions under this general pairwise factorization, existing convergence conditions can be compared and contrasted in a unified way. It is found that existing sufficient convergence conditions derived for Gaussian MRF (including walk-summability, pairwise-normalizability, convex decomposition, and diagonal dominance) are valid under any factorization within the general pairwise factorization model, and the converged beliefs are independent of the choice of pairwise factorization. Moreover, due to the newly established bridge between Gaussian MRF and pairwise linear Gaussian model under the generalized setting, the pairwise-normalizability from Gaussian MRF is further applied to pairwise linear Gaussian model, revealing an easily verifiable sufficient convergence condition in pairwise linear Gaussian model. This sufficient convergence condition is more general than the convergence conditions in Moallemi and Roy (2006), Leng and Wu (2011), and Du and Wu (2013), and further explains the empirical convergence behavior of peer-to-peer rating application (Bickson et al., 2007). Numerical results and applications are presented to corroborate the theoretical convergence results.

Notations: Scalars, vectors, matrices and sets are denoted by lower-case letters, bold lower-case letters, bold upper-case letters and calligraphic upper-case letters, respectively. Notations $\mathbf{a} > \mathbf{b}$, $\mathbf{a} \geq \mathbf{b}$, $\mathbf{a} < \mathbf{b}$, and $\mathbf{a} \leq \mathbf{b}$ indicate $a_i > b_i$, $a_i \geq b_i$, $a_i < b_i$, and $a_i \leq b_i$ for all i, respectively, where a_i and b_i are the i-th element of $\mathbf{a}$ and $\mathbf{b}$. For a matrix, $\mathbf{A} \succ \mathbf{0}$ indicates that $\mathbf{A}$ is positive definite, $\mathbf{A}^T$ denotes the transpose of $\mathbf{A}$, and $\mathbf{A}^{-1}$ denotes the inverse of $\mathbf{A}$. The notation $|\mathbf{A}|$ is a matrix with element-wise absolute value of $\mathbf{A}$, and $\rho(\mathbf{A})$ is the spectral radius of $\mathbf{A}$. Moreover, $\text{diag}(\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_n)$ denotes a block diagonal matrix with the diagonal blocks being $\mathbf{D}_1, \mathbf{D}_2, \cdots, \mathbf{D}_n$, starting from the upper left corner. For the notation $\mathcal{B}$, it denotes a set, and $|\mathcal{B}|$ denotes the number of elements in $\mathcal{B}$. The notation $\mathcal{B} \setminus i$ denotes all the elements in set $\mathcal{B}$ except i, and the notation $\mathcal{B} \setminus \mathcal{C}$ denotes all the elements in set $\mathcal{B}$ but not in $\mathcal{C}$. The notation $a \propto b$ denotes that a is proportional to b. For a Gaussian distributed random variable x with mean m and variance v, we denote it as $\mathcal{N}(x; m, v)$. Similarly, for a multivariate Gaussian distributed random vector $\mathbf{x}$ with mean vector $\mathbf{m}$ and covariance matrix $\mathbf{V}$, we denote it as $\mathcal{N}(\mathbf{x}; \mathbf{m}, \mathbf{V})$.

2. Gaussian Model and Belief Propagation

In this section, we first discuss the pairwise factorizations of Gaussian model (Section 2.1). Then, BP algorithm under a general pairwise factorization of Gaussian model is presented (Section 2.2).

2.1. Gaussian Model and Its Pairwise Factorizations

Consider the joint Gaussian probability density function (pdf) $p(\mathbf{x})$ of a random vector $\mathbf{x} \triangleq [x_1, x_2, \cdots, x_n]^T$ written in a pairwise factorization form

$$\boldsymbol{\mathcal{PFF}}: \; p(\mathbf{x}) \propto \prod_i f_i(x_i) \prod_{i,j>i} f_{ij}(x_i, x_j), \tag{1}$$

where $f_i(x_i)$ is a local function of x_i while $f_{ij}(x_i, x_j)$ is a local function modeling the interaction between x_i and x_j. In general, given a joint Gaussian distribution[1] $\mathcal{N}(\mathbf{x}; \mathbf{J}^{-1}\mathbf{h}, \mathbf{J}^{-1})$ in Gaussian MRF, the $\mathcal{PFF}$ is not unique. For example, one of the earliest pairwise factorizations in Gaussian MRF is given in Weiss and Freeman (2000) with

$$\mathbb{F}_1: \; f_i(x_i) \propto \exp\left\{-\frac{1}{2}p_i(x_i - \frac{h_i}{p_i})^2\right\}, \tag{2a}$$

$$f_{ij}(x_i, x_j) \propto \exp\left\{-\frac{1}{2}[x_i \; x_j]\begin{bmatrix} a_{ij} & J_{ij} \\ J_{ij} & a_{ji} \end{bmatrix}[x_i \; x_j]^T\right\}, \tag{2b}$$

$$\text{s.t. } \mathbf{J} \succ \mathbf{0}, \; p_i + \sum_{k \in \mathcal{B}_i} a_{ik} = J_{ii}, \; \mathcal{B}_i \triangleq \{k \mid J_{ik} \neq 0, k = 1, 2, \cdots, n\}, \tag{2c}$$

where J_{ik} is the (i, k)-th element of $\mathbf{J}$ and h_i is the i-th element of $\mathbf{h}$. On the other hand, the convex decomposition in Moallemi and Roy (2009) requires

$$\mathbb{F}_2: \; f_i(x_i) \propto \exp\left\{-\frac{1}{2}(J_{ii} - \sum_{k \in \mathcal{B}_i} \xi_{ik} J_{ik}^2)x_i^2 + (h_i - \sum_{k \in \mathcal{B}_i} \varsigma_{ik})x_i\right\}, \tag{3a}$$

$$f_{ij}(x_i, x_j) \propto \exp\left\{-\frac{1}{2}\xi_{ij} J_{ij}^2 x_i^2 - J_{ij} x_i x_j - \frac{1}{2}\xi_{ji} J_{ij}^2 x_j^2 + \varsigma_{ij} x_i + \varsigma_{ji} x_j\right\}, \tag{3b}$$

$$\text{s.t. } \mathbf{J} \succ \mathbf{0}, \; \xi_{ij}, \xi_{ji} > 0, \; J_{ii} - \sum_{k \in \mathcal{B}_i} \xi_{ik} J_{ik}^2 > 0, \; \xi_{ij}\xi_{ji} J_{ij}^2 \geq 1. \tag{3c}$$

However, given an information matrix $\mathbf{J}$, the conditions in $\mathbb{F}_2$ might not be satisfied, thus this decomposition may not exist. Consequently, a more popular factorization is given by

$$\mathbb{F}_3: \; f_i(x_i) \propto \exp\left\{-\frac{1}{2}J_{ii} x_i^2 + h_i x_i\right\}, \tag{4a}$$

$$f_{ij}(x_i, x_j) \propto \exp\left\{-J_{ij} x_i x_j\right\}, \tag{4b}$$

$$\text{s.t. } \mathbf{J} \succ \mathbf{0}. \tag{4c}$$

This factorization directly connects the joint distribution without any constraints and is used in Malioutov et al. (2006), and Su and Wu (2014, 2015a,b).

On the other hand, the $\mathcal{PFF}$ also covers the pairwise linear Gaussian model that frequently appears in distributed inference. In particular, assume there exists a local relationship $z_{ij} = c_{ij} x_i + c_{ji} x_j + n_{ij}$ with known coefficients c_{ij}, c_{ji}. If $n_{ij} \sim \mathcal{N}(n_{ij}; 0, 1/\eta_{ij})$ and the prior distribution of x_i is $p(x_i) = \mathcal{N}(x_i; y_i, 1/\zeta_i)$, then we have the joint posterior distribution given by the $\mathcal{PFF}$ with

$$\mathbb{F}_4: \; f_i(x_i) \propto \exp\left\{-\frac{1}{2}\zeta_i (x_i - y_i)^2\right\}, \tag{5a}$$

$$f_{ij}(x_i, x_j) \propto \exp\left\{-\frac{1}{2}\eta_{ij}\left[z_{ij} - (c_{ij} x_i + c_{ji} x_j)\right]^2\right\}, \tag{5b}$$

$$\text{s.t. } \mathrm{diag}(\zeta_1, \zeta_2, \cdots, \zeta_n) + \mathbf{C} \succ \mathbf{0}, \; \zeta_i \geq 0, \; \eta_{ij} > 0, \; \mathbf{C} \in \mathbb{R}^{n \times n}$$
$$\text{with } C_{ii} = \sum\nolimits_{k \in \mathcal{B}_i} \eta_{ik} c_{ik}^2 \text{ and } C_{ij} = \eta_{ij} c_{ij} c_{ji}, \tag{5c}$$

1. The Gaussian distribution is expressed in an information form, where $\mathbf{J}$ is the information matrix and $\mathbf{h}$ is the potential vector.

where $\text{diag}(\zeta_1, \zeta_2, \cdots, \zeta_n) + \mathbf{C} \succ \mathbf{0}$ is a condition guaranteeing the $\mathcal{PFF}$ with factors in $\mathbb{F}_4$ represents a valid Gaussian pdf (in fact, $\text{diag}(\zeta_1, \zeta_2, \cdots, \zeta_n) + \mathbf{C}$ is the information matrix of such Gaussian pdf). This model handles a broad class of distributed estimation applications. In particular, $\mathbb{F}_4$ is exactly the model in distributed clock synchronization (Leng and Wu, 2011) and distributed carrier frequency estimation (Du and Wu, 2013). On the other hand, if $z_{ij} = 0$, $c_{ij} = -c_{ji} = 1$, and y_i is obtained from local observation, $\mathbb{F}_4$ reduces to the model used in consensus propagation (Moallemi and Roy, 2006) and peer-to-peer rating (Bickson et al., 2007).

Currently, different convergence conditions of Gaussian BP are developed under different factorizations. For example, diagonal dominance (Weiss and Freeman, 2000) is developed from $\mathbb{F}_1$, convex decomposition (Moallemi and Roy, 2009) is developed from $\mathbb{F}_2$, walk-summability (Malioutov et al., 2006) and the recent necessary and sufficient condition (Su and Wu, 2014, 2015a) are developed from $\mathbb{F}_3$. One may wonder if the convergence condition of Gaussian BP developed from one factorization is applicable to another factorization. More importantly, for different pairwise factorizations, can we develop a unified treatment in the convergence analysis? In order to answer these questions, we write the general forms of factors $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ in the $\mathcal{PFF}$ as

$$\mathbb{F}_5: \ f_i(x_i) \propto \exp\left\{-\frac{1}{2}\phi_i x_i^2 + \psi_i x_i\right\}, \tag{6a}$$

$$f_{ij}(x_i, x_j) \propto \exp\left\{ -\frac{1}{2}\gamma_{ij} x_i^2 - \frac{1}{2}\gamma_{ji} x_j^2 - \tau_{ij} x_i x_j + \kappa_{ij} x_i + \kappa_{ji} x_j \right\}, \tag{6b}$$

$$\text{s.t. } \tau_{ij} = \tau_{ji} \text{ and } \begin{bmatrix} \phi_1 + \sum_{k\in\mathcal{B}_1} \gamma_{1k} & \tau_{12} & \cdots & \tau_{1n} \\ \tau_{21} & \phi_2 + \sum_{k\in\mathcal{B}_2} \gamma_{2k} & \cdots & \tau_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \tau_{n1} & \tau_{n2} & \cdots & \phi_n + \sum_{k\in\mathcal{B}_n} \gamma_{nk} \end{bmatrix} \succ \mathbf{0}, \tag{6c}$$

where the constraints in (6c) guarantee that the $\mathcal{PFF}$ with factors in $\mathbb{F}_5$ is a valid Gaussian pdf. Notice that $\mathbb{F}_5$ covers $\mathbb{F}_1$, $\mathbb{F}_2$, $\mathbb{F}_3$ and $\mathbb{F}_4$ as special cases. For example, when $\phi_i = J_{ii}$, $\psi_i = h_i$, $\gamma_{ij} = \gamma_{ji} = \kappa_{ij} = \kappa_{ji} = 0$, $\tau_{ij} = J_{ij}$, $\mathbb{F}_5$ reduces to $\mathbb{F}_3$. Furthermore, when $\phi_i = \zeta_i$, $\psi_i = \zeta_i y_i$, $\gamma_{ij} = \eta_{ij} c_{ij}^2$, $\gamma_{ji} = \eta_{ij} c_{ji}^2$, $\tau_{ij} = \eta_{ij} c_{ij} c_{ji}$, $\kappa_{ij} = \eta_{ij} z_{ij} c_{ij}$ and $\kappa_{ji} = \eta_{ij} z_{ij} c_{ji}$, $\mathbb{F}_5$ reduces to $\mathbb{F}_4$. In fact, $\mathbb{F}_5$ covers pairwise factorizations beyond that in $\mathbb{F}_1$, $\mathbb{F}_2$, $\mathbb{F}_3$ and $\mathbb{F}_4$, as long as the exponents of $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ are quadratic polynomials. For example, when $f_i(x_i) \propto 1$ and $f_{ij}(x_i, x_j) \propto \exp\left\{-\frac{1}{2}x_i^2 - \frac{1}{2}x_i x_j - \frac{1}{2}x_j^2 + x_i + x_j\right\}$ for all $i \neq j \in \{1, 2, \cdots, n\}$ in $\mathbb{F}_5$, it leads to a valid Gaussian pdf, but $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ cannot be represented by $\mathbb{F}_1$—$\mathbb{F}_4$. The relationships among $\mathbb{F}_1$—$\mathbb{F}_5$ are shown in Figure 1 at top of next page (relationships among $\mathbb{F}_1$—$\mathbb{F}_4$ are proved in Appendix A).

2.2. Gaussian Belief Propagation Under $\mathbb{F}_5$

In Gaussian BP, messages are updated and passed among variables. In particular, under synchronous scheduling[2], the message $m_{j\to i}(x_i)$ to be passed from variable x_j to variable

2. Synchronous scheduling requires that all messages at each iteration are updated before starting a new one.

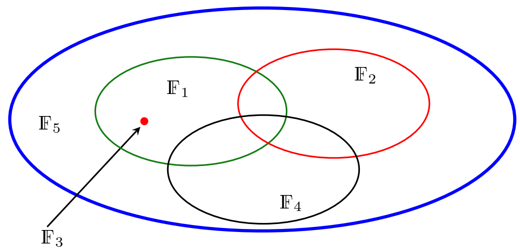


Figure 1: Relationships among different factorization sets.

x_i at the l-th iteration is computed as (Malioutov et al., 2006)

$$m_{j\to i}^{(l)}(x_i) \propto \int_{-\infty}^{\infty} f_{ij}(x_i,x_j) f_j(x_j) \prod_{k\in\mathcal{B}_j\setminus i} m_{k\to j}^{(l-1)}(x_j) dx_j, \tag{7}$$

where $m_{k\to j}^{(l-1)}(x_j)$ denotes the message $m_{k\to j}(x_i)$ passed from variable x_k to variable x_j at the $(l-1)$-th iteration.

Based on the expressions of factors in $\mathbb{F}_5$ and message update rule in (7), we can derive a general formula of messages $m_{j\to i}^{(l)}(x_i)$ for all $(i,j)\in\mathcal{E}$, where $\mathcal{E} \triangleq \{(i,j) \mid i=1,2,\cdots,n, j\in\mathcal{B}_i\}$ denotes the set of index pair (i,j) with $\tau_{ij}\neq 0$. Without loss of generality, it is assumed that the message $m_{j\to i}^{(l-1)}(x_i)$ with $(i,j)\in\mathcal{E}$ at the $(l-1)$-th iteration takes the expression

$$m_{j\to i}^{(l-1)}(x_i) \propto \exp\left\{-\frac{1}{2}\alpha_{j\to i}^{(l-1)} x_i^2 + \beta_{j\to i}^{(l-1)} x_i\right\}, \tag{8}$$

where the parameters $\alpha_{j\to i}^{(l-1)}$ and $\beta_{j\to i}^{(l-1)}$ are the precision and linear coefficient, respectively. Substituting (8) into (7) and if $\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(l-1)} > 0$, the message $m_{j\to i}(x_i)$ at the l-th iteration is well-defined and is shown in Appendix B to be in the form $m_{j\to i}^{(l)}(x_i) \propto \exp\{-\frac{1}{2}\alpha_{j\to i}^{(l)} x_i^2 + \beta_{j\to i}^{(l)} x_i\}$, with

$$\alpha_{j\to i}^{(l)} = \gamma_{ij} - \frac{\tau_{ij}^2}{\phi_j + \gamma_{ji} + \sum\limits_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(l-1)}}, \tag{9}$$

$$\beta_{j\to i}^{(l)} = \kappa_{ij} - \frac{\tau_{ij}\psi_j + \tau_{ij}\kappa_{ji} + \sum\limits_{k\in\mathcal{B}_j\setminus i} \tau_{ij}\beta_{k\to j}^{(l-1)}}{\phi_j + \gamma_{ji} + \sum\limits_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(l-1)}}. \tag{10}$$

After updating the messages $m_{j\to i}^{(l)}(x_i)$ by that of the parameters $\alpha_{j\to i}^{(l)}$ and $\beta_{j\to i}^{(l)}$ for all $(i,j)\in\mathcal{E}$, we can compute the belief of x_i at the l-th iteration as

$$b^{(l)}(x_i) \propto f_i(x_i) \prod_{j\in\mathcal{B}_i} m_{j\to i}^{(l)}(x_i). \tag{11}$$

By substituting the expression of $f_i(x_i)$ in $\mathbb{F}_5$ and $m_{j\to i}^{(l)}(x_i) \propto \exp\{-\frac{1}{2}\alpha_{j\to i}^{(l)}x_i^2 + \beta_{j\to i}^{(l)}x_i\}$ into (11), it becomes

$$b^{(l)}(x_i) \propto \exp\left\{-\frac{1}{2}\left(\phi_i + \sum_{j\in\mathcal{B}_i}\alpha_{j\to i}^{(l)}\right)x_i^2 + \left(\psi_i + \sum_{j\in\mathcal{B}_i}\beta_{j\to i}^{(l)}\right)x_i\right\}, \tag{12}$$

where the precision $\nu_{x_i}^{(l)}$ and linear coefficient $\varphi_{x_i}^{(l)}$ of belief $b^{(l)}(x_i)$ can be denoted as

$$\nu_{x_i}^{(l)} = \phi_i + \sum_{j\in\mathcal{B}_i}\alpha_{j\to i}^{(l)}, \tag{13}$$

$$\varphi_{x_i}^{(l)} = \psi_i + \sum_{j\in\mathcal{B}_i}\beta_{j\to i}^{(l)}. \tag{14}$$

For the belief $b^{(l)}(x_i)$, it represents a valid Gaussian pdf if and only if $\nu_{x_i}^{(l)} > 0$. Under $\nu_{x_i}^{(l)} > 0$, we can explicitly write $b^{(l)}(x_i) = \mathcal{N}(x_i; \varphi_{x_i}^{(l)}/\nu_{x_i}^{(l)}, 1/\nu_{x_i}^{(l)})$, where $\varphi_{x_i}^{(l)}/\nu_{x_i}^{(l)}$ and $1/\nu_{x_i}^{(l)}$ are the belief mean and belief variance of x_i, respectively. It has been proved (Pearl, 1988) that beliefs $b^{(l)}(x_i)$ are guaranteed to converge to the exact marginal distributions in singly connected graphs, such as tree-structured and chain-structured graphs. But for loopy graphs, it is known that beliefs may not converge. Fortunately, if beliefs do converge in Gaussian BP, belief means converge to the exact means of the marginal distributions and belief variances are approximate variances of the marginal distributions (Weiss and Freeman, 2000). This makes convergence analysis in loopy graphs a core research topic of Gaussian BP.

3. Convergence Analysis Under General Pairwise Factorization $\mathbb{F}_5$

In this section, we analyze the convergence of Gaussian BP under the $\mathfrak{PFF}$ with general factors in $\mathbb{F}_5$. For notational convenience, we stack $\alpha_{j\to i}^{(l)}$, $\beta_{j\to i}^{(l)}$, γ_{ij} and κ_{ij} for all $(i,j)\in\mathcal{E}$ into vectors $\boldsymbol{\alpha}^{(l)}$, $\boldsymbol{\beta}^{(l)}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\kappa}$, respectively, where the order of $(i,j)\in\mathcal{E}$ is ascending first on i and then on j. For the order of $(i,j)\in\mathcal{E}$ ascending first on i and then on j, there exists a one-to-one mapping function $order(\cdot,\cdot)$ from all $(i,j)\in\mathcal{E}$ to $1,2,\cdots,|\mathcal{E}|$. That is, for any $k\in\{1,2,\cdots,|\mathcal{E}|\}$, there exists only one $(i,j)\in\mathcal{E}$ satisfying $k = order(i,j)$. Moreover, we define a set

$$\begin{aligned}\mathcal{Q} \triangleq \Big\{\mathbf{q}\in\mathbb{R}^{|\mathcal{E}|} \mid \phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i} > 0, \frac{-\tau_{ij}^2}{\phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i}} - q_{i\to j} \geq 0,\\ \forall (i,j)\in\mathcal{E},\ \text{and}\ \phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i} q_{k\to i} > 0 \text{ for } i=1\Big\},\end{aligned} \tag{15}$$

where $\mathbf{q}\in\mathbb{R}^{|\mathcal{E}|}$ is a vector obtained by stacking $q_{j\to i}$ for all $(i,j)\in\mathcal{E}$ ordered in the same way as that in $\boldsymbol{\alpha}^{(l)}$. For the set $\mathcal{Q}$, it is non-empty if and only if there exists at least one vector $\mathbf{q}\in\mathbb{R}^{|\mathcal{E}|}$ satisfying the inequalities in (15). As shown later, the set $\mathcal{Q}$ would be used to verify the convergence of $\boldsymbol{\alpha}^{(l)}$.

First, we consider the convergence of BP beliefs under synchronous scheduling. For synchronous scheduling, if $\boldsymbol{\alpha}^{(l)}$ converge to $\boldsymbol{\alpha}^*$ with elements $\alpha_{j\to i}^*$ for all $(i,j)\in\mathcal{E}$ ordered in the same way as that in $\boldsymbol{\alpha}^{(l)}$, the update of $\beta_{j\to i}^{(l)}$ in (10) for all $(i,j) \in \mathcal{E}$ can be written in a vector form as

$$\boldsymbol{\beta}^{(l)} = \mathbf{A}\boldsymbol{\beta}^{(l-1)} + \mathbf{d}, \tag{16}$$

where $\mathbf{A}$ is a $|\mathcal{E}| \times |\mathcal{E}|$ matrix with the element

$$A_{i'j'} = \begin{cases} -\frac{\tau_{i_1 i_2}}{\phi_{i_1} + \gamma_{i_1 i_2} + \sum_{k \in \mathcal{B}_{i_1} \setminus i_2} \alpha^*_{k\to i_1}} & \text{if } j_1 \in \mathcal{B}_{i_1} \setminus i_2 \text{ and } j_2 = i_1, \\ 0 & \text{otherwise,} \end{cases} \tag{17}$$

for $i' = order(i_1, i_2)$ and $j' = order(j_1, j_2)$, and $\mathbf{d}$ is a column vector containing elements $\kappa_{ij} - \frac{\tau_{ij}\psi_j + \tau_{ij}\kappa_{ji}}{\phi_j + \gamma_{ji} + \sum_{k \in \mathcal{B}_j \setminus i} \alpha^*_{k\to j}}$ for all $(i,j) \in \mathcal{E}$ ordered in the same way as that in $\boldsymbol{\alpha}^{(l)}$. Conditioned on the convergence of $\boldsymbol{\alpha}^{(l)}$, the convergence of $\boldsymbol{\beta}^{(l)}$ can be easily analyzed with the linear update equation in (16). In the following, we present the convergence conditions of Gaussian BP under general pairwise factorization and synchornous scheduling.

Theorem 1 ***(Convergence conditions under synchronous scheduling)*** *Under synchronous scheduling and any initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, belief variances converge to the same positive values if and only if the set $\mathcal{Q}$ is non-empty. Furthermore, if $\mathcal{Q}$ is non-empty, belief means converge to unique values if and only if $\rho(\mathbf{A}) < 1$ under synchronous scheduling.*

Proof See Appendix C. ∎

Theorem 1 states that after we fix the factorization parameters in $\mathbb{F}_5$, Gaussian BP converges if and only if the set $\mathcal{Q}$ is non-empty and $\rho(\mathbf{A}) < 1$. Furthermore, the converged values of BP beliefs are independent of the choice of initialization as long as $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$. Notice that the spectral radius $\rho(\mathbf{A})$ depends on the converged value of $\boldsymbol{\alpha}^{(l)}$, which can be obtained by running the update of $\boldsymbol{\alpha}^{(l)}$ until convergence. On the other hand, the condition $\rho(\mathbf{A}) < 1$ will be useful in analyzing the invariant convergence properties of Gaussian BP in Section 4 without the need to know the converged value $\boldsymbol{\alpha}^*$.

Synchronous scheduling requires each variable receiving updated messages from all neighbors before computing the next round messages. In contrast, asynchronous scheduling allows distributed processors to operate more efficiently and flexibly without waiting for all its neighbors' messages. For totally asynchronous scheduling (Bertsekas and Tsitsiklis, 1989), the convergence condition is given by the following Theorem.

Theorem 2 ***(Convergence conditions under totally asynchronous scheduling)*** *Under totally asynchronous scheduling and any initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, belief variances converge to the same positive values if and only if the set $\mathcal{Q}$ is non-empty. Moreover, if $\mathcal{Q}$ is non-empty and $\rho(|\mathbf{A}|) < 1$, belief means converge to unique values under totally asynchronous scheduling.*

Proof See Appendix D. ∎

From Theorems 1 and 2, it is noticed that the convergence condition of belief variances under synchronous and totally asynchronous schedulings stays the same, i.e., $\mathcal{Q} \neq \emptyset$. Furthermore, since totally asynchronous scheduling includes synchronous scheduling as a special case, the converged value of BP message precision $\boldsymbol{\alpha}^*$ stays the same for both synchronous and totally asynchronous schedulings. Consequently, the matrix $\mathbf{A}$ in both Theorems 1 and 2 are the same. On the other hand, for the convergence condition of belief means, the condition in Theorem 2 is stricter than that in Theorem 1 since $\rho(\mathbf{A}) \leq \rho(|\mathbf{A}|)$. This implies that if the convergence condition in Theorem 2 for totally asynchronous scheduling is satisfied, the convergence condition in Theorem 1 for synchronous scheduling would automatically be satisfied. But Theorem 1 has its own value since it is the necessary and sufficient condition, while Theorem 2 is only a sufficient condition. Notice that the initialization $\boldsymbol{\alpha}^{(0)}$ also plays a role in the convergence of BP beliefs, but $\boldsymbol{\alpha}^{(0)}$ is a parameter controllable by the user, therefore the requirement $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$ can easily be satisfied.

For the verification of the non-emptiness of the set $\mathcal{Q}$, one may directly solve the following semidefinite programming (SDP) (Vandenberghe and Boyd, 1996) problem:

$$\min_{\mathbf{q}\in\mathbb{R}^{|\mathcal{E}|},a} \quad a, \tag{18a}$$

$$\text{s.t.} \quad \begin{bmatrix} \phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i} & \tau_{ij} \\ \tau_{ij} & -q_{i\to j} \end{bmatrix} \succeq \mathbf{0}, \forall (i,j)\in\mathcal{E}, \tag{18b}$$

$$a + \phi_1 + \sum_{k\in\mathcal{B}_1}\gamma_{1k} + \sum_{k\in\mathcal{B}_1} q_{k\to 1} \geq 0, \tag{18c}$$

where the constraints in (18b) are equivalent to the first two constraints in $\mathcal{Q}$, and the constraint in (18c) corresponds to the third constraint in $\mathcal{Q}$ after introducing a slack variable a. If the SDP problem in (18a)—(18c) is feasible with the solution $(\mathbf{q}^\dagger, a^\dagger)$ and $a^\dagger < 0$, then $\mathbf{q}^\dagger$ satisfies the constraints in (18b) and (18c) with $\phi_1 + \sum_{k\in\mathcal{B}_1}\gamma_{1k} + \sum_{k\in\mathcal{B}_1} q^\dagger_{k\to 1} \geq -a^\dagger > 0$, which implies $\mathcal{Q}$ is non-empty. Otherwise, $\mathcal{Q}$ would be empty. For solving the SDP problem in (18a)—(18c), interior point methods, with available solvers such as SeDuMi, SDPT3, and MOSEK, can be used.

On the other hand, the optimization problem in (18a)—(18c) can be reformulated in the form:

$$\min_{\mathbf{S}\in\mathcal{S}^{2|\mathcal{E}|+1},\mathbf{q}\in\mathbb{R}^{|\mathcal{E}|},a} \quad a, \tag{19a}$$

$$\text{s.t.} \quad \mathbf{S} \succeq \mathbf{0}, \tag{19b}$$

$$a\mathbf{M}_a + \sum_{(i,j)\in\mathcal{E}} q_{j\to i}\mathbf{M}_{j\to i} + \mathbf{S} = \mathbf{B}, \tag{19c}$$

where $\mathcal{S}^m$ denotes the set of $m \times m$ symmetric matrix, $\mathbf{M}_a \triangleq \text{diag}(-1, \mathbf{0}) \in \mathcal{S}^{2|\mathcal{E}|+1}$, $\mathbf{B} \triangleq \text{diag}(\phi_1 + \sum_{k\in\mathcal{B}_1}\gamma_{1k}, \mathbf{D}^{(B)}_{1,k_1}, \cdots, \mathbf{D}^{(B)}_{i,k_i}, \cdots, \mathbf{D}^{(B)}_{n,k_n}) \in \mathcal{S}^{2|\mathcal{E}|+1}$ with

$$\mathbf{D}^{(B)}_{i,k_i} = \begin{bmatrix} \phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} & \tau_{ik_i} \\ \tau_{ik_i} & 0 \end{bmatrix} \text{ for all } i = 1, 2, \cdots, n \text{ and } k_i \in \mathcal{B}_i,$$

and $\mathbf{M}_{j\to i} \triangleq \text{diag}(d^{(M)}, \mathbf{D}^{(M)}_{1,k_1}, \cdots, \mathbf{D}^{(M)}_{i,k_i}, \cdots, \mathbf{D}^{(M)}_{n,k_n}) \in \mathcal{S}^{2|\mathcal{E}|+1}$ with

$$d^{(M)} = \begin{cases} -1 & \text{if } i = 1 \text{ and } j \in \mathcal{B}_1, \\ 0 & \text{otherwise,} \end{cases}$$

$$\mathbf{D}^{(M)}_{i_1,j_1} = \begin{cases} \begin{bmatrix} -1 & 0 \\ 0 & 0 \end{bmatrix} & \text{if } i_1 = i \text{ and } j_1 \in \mathcal{B}_i \setminus j, \\ \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} & \text{else if } i_1 = j \text{ and } j_1 = i, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

The form of the SDP problem in (19a)—(19c) is exactly the same as that of the SDP problem in equation (3) of Wen et al. (2010). Therefore, the alternating direction augmented Lagrangian methods in Wen et al. (2010) can also be used to solve the SDP problem in (19a)—(19c).

Furthermore, we can interpret the physical meaning of the convergence condition in Theorems 1 and 2 as follows. Under initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, the first two constraints in the set $\mathcal{Q}$ correspond to the condition for guaranteeing the validity of integration in (32). Together with the third constraint in the set $\mathcal{Q}$, they correspond to the condition for guaranteeing the BP beliefs in (12) remain as valid Gaussian pdfs at each iteration. This is summarized and formally proved in the following Corollary.

Corollary 3 ***(Physical meaning of non-emptiness of*** $\mathcal{Q}$***)*** *Under any initialization* $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$*, BP beliefs at each iteration are valid Gaussian pdfs if and only if the set* $\mathcal{Q}$ *is non-empty.*

Proof See Appendix E. ∎

Note that being a valid Gaussian pdf in $\mathbb{F}_5$ does not necessarily imply that the BP beliefs are valid Gaussian pdfs at all iterations. Rather, being a valid Gaussian pdf in $\mathbb{F}_5$ is a prerequisite for Theorems 1, 2, and Corollary 3.

4. Invariant Properties of Convergence Conditions Under $\mathbb{F}_5$

For the $\mathfrak{PFF}$ with factors in $\mathbb{F}_5$, it can be equivalently written as a Gaussian pdf $\mathcal{N}(\mathbf{x}; \mathbf{J}^{-1}\mathbf{h}, \mathbf{J}^{-1})$ with the information matrix

$$\mathbf{J} \triangleq \begin{bmatrix} \phi_1 + \sum_{k\in\mathcal{B}_1} \gamma_{1k} & \tau_{12} & \cdots & \tau_{1n} \\ \tau_{21} & \phi_2 + \sum_{k\in\mathcal{B}_2} \gamma_{2k} & \cdots & \tau_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \tau_{n1} & \tau_{n2} & \cdots & \phi_n + \sum_{k\in\mathcal{B}_n} \gamma_{nk} \end{bmatrix}, \tag{20}$$

and the potential vector

$$\mathbf{h} \triangleq \left[\psi_1 + \sum_{k\in\mathcal{B}_1} \kappa_{1k}, \psi_2 + \sum_{k\in\mathcal{B}_2} \kappa_{2k}, \cdots, \psi_n + \sum_{k\in\mathcal{B}_n} \kappa_{nk} \right]. \tag{21}$$

Conversely, for a given Gaussian pdf $\mathcal{N}(\mathbf{x}; \mathbf{J}^{-1}\mathbf{h}, \mathbf{J}^{-1})$, there are many possible pairwise factorizations with each obtained by choosing different ϕ_i, ψ_i, γ_{ij}, τ_{ij} and κ_{ij} while satisfying the constraints in $\mathbb{F}_5$. Since the convergence conditions in Theorems 1 and 2 seem to depend on the factorization parameters. One may wonder would different factorizations demonstrate different convergence behaviors? We will give the answer by the following discussions.

From (20) and (21), it is obvious that for a Gaussian pdf with information matrix $\mathbf{J}$ and potential vector $\mathbf{h}$, the factorization parameters of any pairwise factorization with factors in the form of $\mathbb{F}_5$ should satisfy $\phi_i + \sum_{k\in\mathcal{B}_i} \gamma_{ik} = J_{ii}$, $\tau_{ij} = J_{ij}$ and $\psi_i + \sum_{k\in\mathcal{B}_i} \kappa_{ik} = h_i$. Using these relationships, the definition of the set $\mathcal{Q}$ can be rewritten as

$$\mathcal{Q} \triangleq \Big\{\mathbf{q} \in \mathbb{R}^{|\mathcal{E}|} \mid J_{ii} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i} > 0, \frac{-J_{ij}^2}{J_{ii} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i}} - q_{i\to j} \geq 0, \forall (i,j) \in \mathcal{E},$$
$$\text{and } J_{ii} + \sum_{k\in\mathcal{B}_i} q_{k\to i} > 0 \text{ for } i = 1\Big\}. \tag{22}$$

From (22), one can obtain that the emptiness or non-emptiness of the set $\mathcal{Q}$ only depends on the information matrix $\mathbf{J}$ and is independent of ϕ_i, ψ_i, γ_{ij}, τ_{ij} and κ_{ij} as long as they obey $\mathbb{F}_5$. The above discussion is summarized in the following Lemma.

Lemma 4 *For a Gaussian pdf with information matrix* $\mathbf{J}$*, the emptiness or non-emptiness of the set* $\mathcal{Q}$ *is independent of factorization parameters in* $\mathbb{F}_5$*.*

Furthermore, when belief variances converge, it can be proved that the matrix $\mathbf{A}$ is independent of the choice of factorization parameters ϕ_i, ψ_i, γ_{ij}, τ_{ij} and κ_{ij}. This is formalized in the following Lemma.

Lemma 5 *For a Gaussian pdf with information matrix* $\mathbf{J}$*, if the set* $\mathcal{Q}$ *is non-empty, matrix* $\mathbf{A}$ *is independent of factorization parameters in* $\mathbb{F}_5$ *for any initialization* $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$*.*

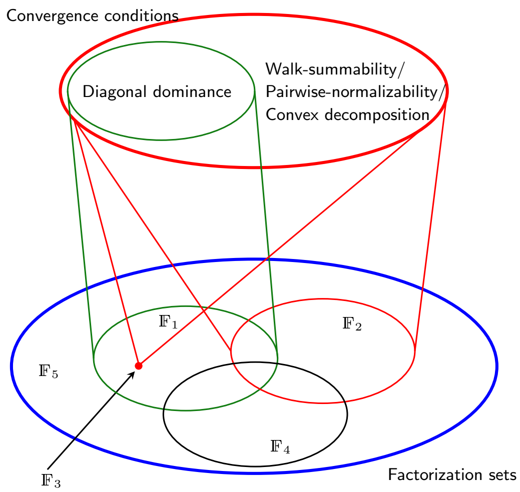
Proof See Appendix F. ■

Since $\mathcal{Q}$ and $\mathbf{A}$ are independent of the specific pairwise factorization, and the convergence conditions in Theorems 1 and 2 only depend on $\mathcal{Q}$ and $\mathbf{A}$, we obtain the invariant properties of Gaussian BP convergence in the following two Corollaries.

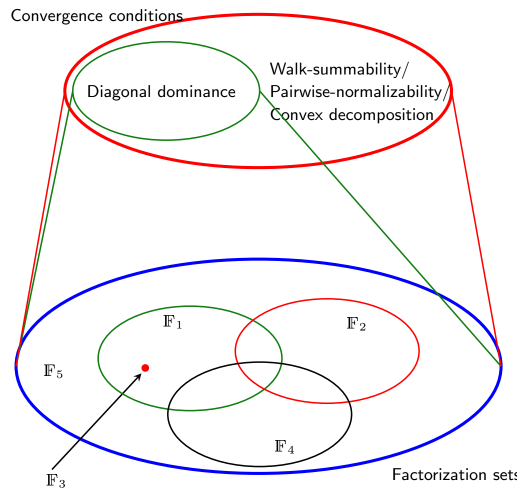
Corollary 6 ***(Convergence or divergence being independent of factorization)*** *For a Gaussian pdf with information matrix* $\mathbf{J}$*, under synchronous scheduling and any initialization* $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$*, the convergence or divergence of Gaussian BP is independent of factorization parameters in* $\mathbb{F}_5$*.*

Proof See Appendix G. ■

Corollary 6 is a very strong result as convergence (or divergence) in one factorization automatically implies convergence (or divergence) in all other factorizations. On the other



(a) Previous results.



(b) New relationships revealed by Corollary 7.

Figure 2: Relationships of classical convergence conditions and factorization sets.

hand, for totally asynchronous scheduling, due to the fact that Theorem 2 is only a sufficient condition under $\mathbb{F}_5$, we cannot have a similar conclusion to that of Corollary 6.

Previously, walk-summability was proved to be equivalent to pairwise-normalizability (Malioutov et al., 2006) and convex decomposition (Moallemi and Roy, 2009), thus they are all valid within factorizations $\mathbb{F}_2$ and $\mathbb{F}_3$. On the other hand, diagonal dominance is valid within $\mathbb{F}_1$ (Weiss and Freeman, 2000), and was proved to be a special case of walk-summability (Malioutov et al., 2006). These relationships are illustrated in Figure 2(a). But whether these classical conditions are valid for other pairwise factorizations is still unknown. To this end, we obtain a result that is practically very useful in the following Corollary.

Corollary 7 ***(Validity of existing sufficient conditions under any pairwise factorization)*** *For a Gaussian pdf with information matrix* $\mathbf{J}$*, if* $\mathbf{J}$ *is diagonally dominant, walk-summable, pairwise-normalizable or convex decomposable, Gaussian BP converges under any pairwise factorization within* $\mathbb{F}_5$ *in both synchronous and totally asynchronous schedulings as long as initialization* $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$.

Proof See Appendix H. ■

With Corollary 7, it gives a more complete picture: diagonal dominance, walk-summability, pairwise-normalizability, and convex decompositions are all valid sufficient conditions under $\mathbb{F}_5$, which includes $\mathbb{F}_1$, $\mathbb{F}_2$ and $\mathbb{F}_3$ as special cases. This extension is illustrated in Figure 2(b). Moreover, the initialization set of $\boldsymbol{\alpha}^{(0)}$ in Corollary 7 is larger than that in previous conditions (Weiss and Freeman, 2000; Malioutov et al., 2006; Moallemi and Roy, 2009). Notice that from Corollary 7, the initialization set is related to the parameter $\boldsymbol{\gamma}$, which means that typical initialization $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$ will not always work. This point will be further illustrated in Section 6.1. For comparison with the results in Su and Wu (2015a), it is noticed that Theorems 1 and 2 can be considered as extensions of the convergence conditions in Su and Wu (2015a) from a single factorization $\mathbb{F}_3$ to a much larger set $\mathbb{F}_5$.

Corollaries 6 and 7 only give us the information about the convergence or divergence of Gaussian BP under different pairwise factorizations. Another important question is if BP beliefs converge, would the converged values of beliefs be the same under different pairwise factorizations? The answer is given in the following Theorem.

Theorem 8 *(**Invariant converged beliefs**) For a Gaussian pdf with information matrix* $\mathbf{J}$ *and potential vector* $\mathbf{h}$*, under any initialization* $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$*, if Gaussian BP converges in synchronous or totally asynchronous scheduling, the converged beliefs are independent of factorization parameters in* $\mathbb{F}_5$*.*

Proof See Appendix I. ■

5. A Simple Convergence Condition in Pairwise Linear Gaussian Model

In the pairwise linear Gaussian model $\mathbb{F}_4$, we have the parameters of $\mathbb{F}_5$ as $\phi_i = \zeta_i$, $\psi_i = \zeta_i y_i$, $\gamma_{ij} = \eta_{ij} c_{ij}^2$, $\gamma_{ji} = \eta_{ij} c_{ji}^2$, $\tau_{ij} = \eta_{ij} c_{ij} c_{ji}$, $\kappa_{ij} = \eta_{ij} z_{ij} c_{ij}$ and $\kappa_{ji} = \eta_{ij} z_{ij} c_{ji}$. Further setting different values of ζ_i, y_i, η_{ij}, z_{ij}, c_{ij} and c_{ji}, the model covers various applications (Leng and Wu, 2011; Du and Wu, 2013; Moallemi and Roy, 2006; Bickson et al., 2007), where the convergence of Gaussian BP is studied separately. While we can use the general results in Theorems 1 and 2 to verify whether Gaussian BP converges in a particular pairwise linear Gaussian model, this approach requires us to verify the non-emptiness of the set $\mathcal{Q}$ and find the spectral radius of matrix $\mathbf{A}$. In the following, by further making use of the special structure of $\mathbb{F}_4$, we will provide an easily verifiable sufficient convergence condition for the pairwise linear Gaussian model. Before we state the result, we need the following Lemma.

Lemma 9 *A Gaussian model obeying* $\mathbb{F}_5$ *is pairwise-normalizable if there exist* ω_{ij} *for all* $(i,j) \in \mathcal{E}$ *such that*

$$\omega_{ij} \geq 0 \text{ for all } (i,j) \in \mathcal{E}, \tag{23a}$$

$$\omega_{ij}\omega_{ji} - \tau_{ij}^2 \geq 0 \text{ for all } (i, j > i) \in \mathcal{E}, \tag{23b}$$

$$\phi_i + \sum_{j \in \mathcal{B}_i} \gamma_{ij} - \sum_{j \in \mathcal{B}_i} \omega_{ij} > 0 \text{ for all } i = 1, 2, \cdots, n. \tag{23c}$$

Proof See Appendix J. ■

Making use of Lemma 9, we can prove that the pairwise linear Gaussian model $\mathbb{F}_4$ with at least one $\zeta_i > 0$ is pairwise-normalizable, therefore by Theorem 8, Gaussian BP is convergent under synchronous and totally asynchronous schedulings. The detail is given in the following Theorem.

Theorem 10 *(**At least one** $\zeta_i > 0$ **guarantees convergence**) For pairwise linear Gaussian model* $\mathbb{F}_4$*, under any initialization* $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$*, if* $\zeta_i > 0$ *for at least one* i*, Gaussian BP converges in both synchronous and totally asynchronous schedulings.*

Proof See Appendix K. ■

From Theorem 10, the convergence of Gaussian BP in pairwise linear Gaussian model can be guaranteed if there exists one $\zeta_i > 0$, which is much easier to check than directly verifying the conditions in Theorems 1 and 2. For example, in clock synchronization, ζ_i has the physical interpretation of precision of prior distribution of unknown clock offset. Theorem 10 implies that as long as there is at least one node having an informative prior on the clock parameter, the distributed clock synchronization algorithm (Leng and Wu, 2011) would converge. This is a much stronger result than that in Leng and Wu (2011) since Leng and Wu (2011) proves the convergence under one reference node with perfect prior information (i.e., $p(x_i) = \delta(x_i)$). In contrast, Theorem 10 allows the reference node to have imperfect prior, and more importantly, Theorem 10 allows more than one node having prior information, possibly with different precisions.

On the other hand, for peer-to-peer rating applications, ζ_i represents the confidence of node i's initial rating y_i. Theorem 10 shows that Gaussian BP converges with a minimum of one initial rating with corresponding confidence $\zeta_i > 0$. This is obviously true for a practical peer-to-peer rating system. Notice that while Bickson et al. (2007) requires all ζ_i to be the same for different nodes, our model $\mathbb{F}_4$ can accommodate different confidences at different nodes and Gaussian BP still converges due to Theorem 10. This is the first time that the convergence of Gaussian BP is proved in such an application. Furthermore, if all the observations $\{y_i\}$ exist with the same confidence $\zeta_i = 1$, the model $\mathbb{F}_4$ becomes the consensus propagation model in Moallemi and Roy (2006). Due to $\zeta_i > 0$ for all i, Theorem 10 guarantees the convergence of Gaussian BP in consensus propagation.

6. Numerical Results and Applications

In this section, numerical results are presented to corroborate the newly established theoretical results. In the following simulations, we consider two message passing schedules: synchronous scheduling, and an asynchronous schedule with each variable updating its message with probability 0.7 at each iteration.

6.1. Impact of Different Pairwise Factorizations

Consider a joint Gaussian distribution of 6 random variables, where the information matrix $\mathbf{J}$ and potential vector $\mathbf{h}$ are

$$\mathbf{J} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 6 & -1 & 2 & 0 & 0 \\ 0 & -1 & 5 & -6 & 0 & 0 \\ 0 & 2 & -6 & 36 & 3 & -10 \\ 0 & 0 & 0 & 3 & 9 & 0 \\ 0 & 0 & 0 & -10 & 0 & 4 \end{bmatrix}, \ \mathbf{h} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}.$$

We consider two pairwise factorizations with different parameters in $\mathbb{F}_5$. In pairwise factorization I, we set $\phi_1 = 1$, $\phi_2 = 6$, $\phi_3 = 5$, $\phi_4 = 36$, $\phi_5 = 9$, $\phi_6 = 4$, $\psi_i = 1$ for all $i = 1, 2, \cdots, 6$, $\tau_{12} = 1$, $\tau_{23} = -1$, $\tau_{24} = 2$, $\tau_{34} = -6$, $\tau_{45} = 3$, $\tau_{46} = -10$, and $\gamma_{ij} = \kappa_{ij} = 0$

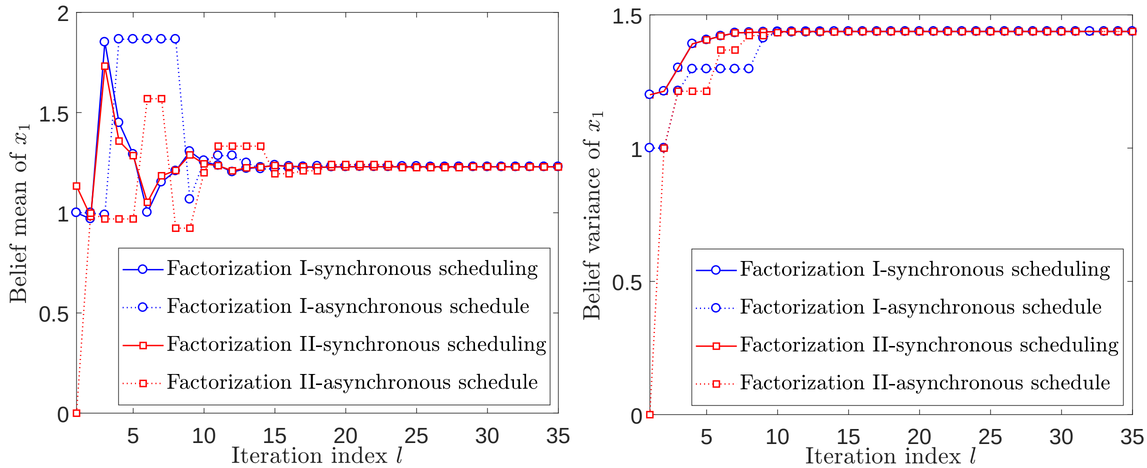


Figure 3: Convergence of beliefs under different pairwise factorizations and message schedulings.

for all $(i, j) \in \mathcal{E}$. The factors in this factorization are used in Malioutov et al. (2006), and Su and Wu (2014, 2015a). In pairwise factorization II, we set $\phi_i = 0$ and $\psi_i = 0$ for all $i = 1, 2, \cdots, 6$, $\gamma_{12} = \gamma_{21} = \gamma_{23} = \gamma_{32} = \gamma_{42} = \gamma_{45} = 1$, $\gamma_{24} = \gamma_{34} = \gamma_{64} = 4$, $\gamma_{43} = \gamma_{54} = 9$, $\gamma_{46} = 25$, $\tau_{12} = 1$, $\tau_{23} = -1$, $\tau_{24} = 2$, $\tau_{34} = -6$, $\tau_{45} = 3$, $\tau_{46} = -10$, $\kappa_{12} = \kappa_{54} = \kappa_{64} = 1$, $\kappa_{21} = \kappa_{23} = \kappa_{24} = 1/3$, $\kappa_{32} = \kappa_{34} = 1/2$, and $\kappa_{42} = \kappa_{43} = \kappa_{45} = \kappa_{46} = 1/4$.

It can be verified that the set $\mathcal{Q}$ is non-empty for the considered Gaussian distribution by solving the SDP problem in (18a)—(18c) (as discussed in Lemma 4, the set $\mathcal{Q}$ is independent of the chosen factorization). Furthermore, under initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, numerical computation shows that $\rho(\mathbf{A}) = 0.6934 < 1$ and $\rho(|\mathbf{A}|) = 0.6934 < 1$ for both pairwise factorizations. This corroborates Lemma 5 that $\mathbf{A}$ is independent of factorizations. By Theorems 1 and 2, under initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, Gaussian BP beliefs converge in both synchronous and totally asynchronous schedulings. Figure 3 shows belief means and variances of variable x_1 during the iterations, where $\boldsymbol{\alpha}^{(0)}$ is set to $\boldsymbol{\gamma}$ and $\boldsymbol{\beta}^{(0)}$ is set to $\mathbf{0}$ in both pairwise factorizations (i.e., $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$ for factorization I and $\boldsymbol{\alpha}^{(0)} = [1\ 1\ 1\ 4\ 1\ 4\ 1\ 9\ 1\ 25\ 9\ 4]^T$ for factorization II). It can be seen that belief means (variances) converge to the same values for both factorizations under synchronous scheduling and asynchronous scheduling, which corroborates Theorem 8.

Nevertheless, if we set the initialization $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$ in factorization II, by the update equation in (9), $\boldsymbol{\alpha}^{(l)}$ can be easily shown to maintain at $\mathbf{0}$ for all $l \geq 0$. Then belief variance $1/(\phi_i + \sum_{j \in \mathcal{B}_i} \alpha_{j \to i}^{(l)}) = 1/0$ is not defined. But for factorization I, BP beliefs do converge when $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$. Therefore, the initialization $\boldsymbol{\alpha}^{(0)} = \mathbf{0}$ will not always work for different pairwise factorizations. To guarantee the convergence of Gaussian BP, we need to choose the initialization $\boldsymbol{\alpha}^{(0)}$ according to the factorization parameter $\boldsymbol{\gamma}$.

6.2. Distributed Clock Synchronization

In clock synchronization of wireless sensor networks (Leng and Wu, 2011), x_i represents the unknown clock offset at node i, and the likelihood function can be regarded as $f_{ij}(x_i, x_j)$

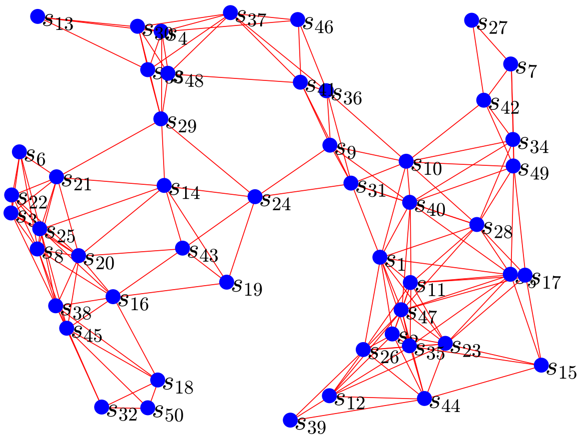


Figure 4: Network topology in clock synchronization.

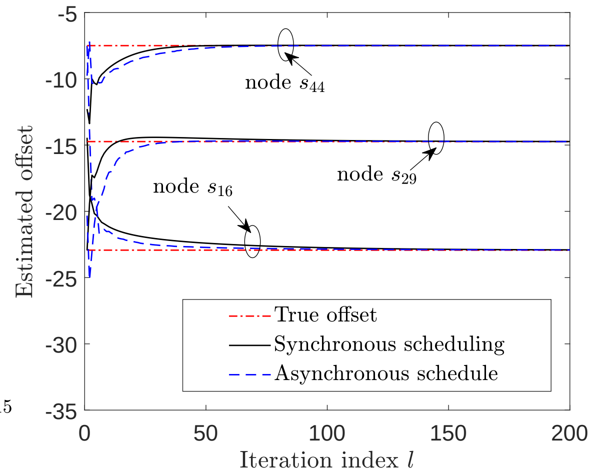


Figure 5: Estimated offsets.

with the expression

$$f_{ij}(x_i, x_j) \propto \exp\left\{-\frac{N}{2\sigma^2}\left[2\left(x_i - x_j\right) + \frac{\mathbf{1}^T \mathbf{t}_{ij}}{N}\right]^2\right\}, \tag{24}$$

where $\mathbf{1}$ is the all-ones vector of length N, $\mathbf{t}_{ij}$ is a column vector collecting N observations, and the observation errors are Gaussian distributed with zero mean and variance σ^2. Moreover, the prior of x_i is Gaussian distributed, and it can be regarded as $f_i(x_i)$ with the expression

$$f_i(x_i) \propto \exp\left\{-\frac{1}{2\bar{\sigma}_i^2}\left(x_i - \bar{\mu}_i\right)^2\right\}, \tag{25}$$

where $\bar{\mu}_i$ and $\bar{\sigma}_i^2$ are the mean and variance, respectively. In case there is no prior information for x_i, we can set $\bar{\sigma}_i^2 = \infty$ and $f_i(x_i) \propto 1$. On the other hand, if the prior of x_i is perfect, then $\bar{\sigma}_i^2 = 0$ and $f_i(x_i) \propto \delta(x_i - \bar{\mu}_i)$.

We consider a network with topology shown in Figure 4, where there are 50 nodes. Among these nodes, the clock offsets of nodes with no prior information are drawn uniformly from $[-30, 30]$. Moreover, we set $N = 4$, $\sigma^2 = 1$ and two reference nodes s_1 and s_4 with their offsets drawn from $\mathcal{N}(x_1; 0, 10^{-1})$ and $\mathcal{N}(x_4; 0, 10^{-2})$, respectively. By Theorem 10, if there exists at least one node with prior information in the network, BP beliefs always converge under both synchronous and totally asynchronous schedulings. Figure 5 shows the estimated offsets of nodes s_{16}, s_{29}, s_{44} under synchronous scheduling and asynchronous scheduling. It can be seen that the estimated offsets of each node converge to the same values under both synchronous and asynchronous schedulings, corroborating the results in Theorem 8.

6.3. Peer-to-Peer Rating

The ratings of items, such as movies, doctors and vendors, play an important role in the social networks and can affect the decisions of people in some degree. Define x_i as the rating

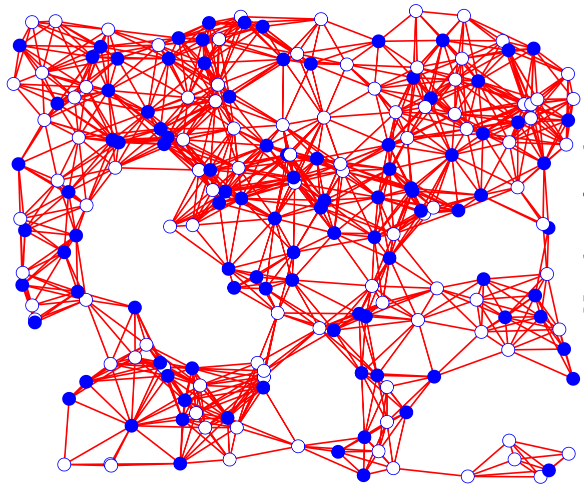


Figure 6: Network topology of peer-to-peer rating.

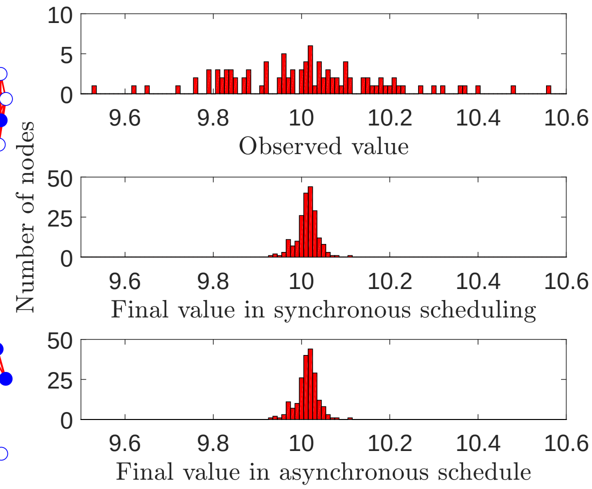


Figure 7: Histograms of consensus values.

at node i, and the peer-to-peer rating problem can be formulated as

$$\min_{x_1, x_2, \cdots, x_n} \sum_i \frac{1}{2}\varphi_{ii}(x_i - z_i)^2 + \sum_{i \neq j} \frac{1}{2}\varphi_{ij}(x_i - x_j)^2, \tag{26}$$

where z_i is the initial rating of node i, φ_{ii} and φ_{ij} denote the self confidence of node i and mutual trust between node i and node j, respectively. The term in the first summation of (26) denotes the rating difference between the final result and its initialization, and the term in the second summation of (26) denotes the rating difference between a pair of neighboring nodes. In particular, if we set

$$f_i(x_i) \propto \exp\left\{-\frac{1}{2}\varphi_{ii}(x_i - z_i)^2\right\}, \tag{27}$$

$$f_{ij}(x_i, x_j) \propto \exp\left\{-\frac{1}{2}\varphi_{ij}(x_i - x_j)^2\right\}, \tag{28}$$

then Gaussian BP can be performed to estimate $\{x_i\}$.

We consider a network with topology in Figure 6, where there are 200 nodes. Out of the 200 nodes, 100 randomly selected nodes (denoted as solid nodes in Figure 6) have observations z_i on a common object or variable with true value 10. The observations are disturbed by zero-mean Gaussian noise with variance φ_{ii} uniformly sampled from $[10, 10^2]$, and we set φ_{ij} to $10^{1.5}$. Comparing $f_i(x_i)$ in (27) with that in $\mathbb{F}_4$, it is noticed that $\xi_i = \varphi_{ii} > 0$ for the nodes with observations. By Theorem 10, Gaussian BP beliefs converge under both synchronous and totally asynchronous schedulings. This can be seen from Figure 7 that the final values at all nodes concentrate much more closely to the true common value compared to the initial observed values. Furthermore, the final consensus values are consistent under synchronous scheduling and asynchronous scheduling.

Figure 8: (a) Original image, (b) Corrupted image, (c) Recovered image by Gaussian BP, (d) Recovered image by median filter (5×5) for all pixels, (e) Recovered image by median filter (5×5) only for pixels with values 0 or 255.

6.4. Image Denoising

Consider a $M \times N$ gray level image (with level $[0, 255]$) corrupted by salt-and-pepper noise. With the true gray level of pixel at location (i, j) in the original image denoted by $x_{i,j}$, we can formulate the image restoration problem as

$$\min_{\{x_{i,j}\}} \quad \sum_{1\leq i\leq M, 1\leq j\leq N} \frac{1}{2} w_{i,j}(x_{i,j} - z_{i,j})^2 + \sum_{(i,j)\in\mathcal{D}} \frac{1}{2}(x_{i,j} - x_{i,j-1})^2 + \frac{1}{2}(x_{i,j} - x_{i,j+1})^2 + \frac{1}{2}(x_{i,j} - x_{i-1,j})^2 + \frac{1}{2}(x_{i,j} - x_{i+1,j})^2, \tag{29}$$

where $w_{i,j}$ is the confidence of the observed pixel value $z_{i,j}$ at the location (i, j), and $\mathcal{D}$ denotes the set of the locations of pixels with values 0 or 255. From (29), if we set the factors

$$f(x_{i,j}) \propto \exp\left\{-\frac{1}{2} w_{i,j}(x_{i,j} - z_{i,j})^2\right\}, \tag{30}$$

$$f(x_{i,j}, \widetilde{x}) \propto \exp\left\{-\frac{1}{2}(x_{i,j} - \widetilde{x})^2\right\}, \tag{31}$$

where $\widetilde{x}$ denotes the neighboring pixel of $x_{i,j}$ (i.e., $x_{i,j-1}$ or $x_{i,j+1}$ or $x_{i-1,j}$ or $x_{i+1,j}$), by treating all $x_{i,j}$ as continuous variables, we can perform Gaussian BP to recover the image.

As a demonstration, we choose a 256×256 gray-scale Lena image. The original image (shown in Figure 8(a)) is corrupted by 50% salt-and-pepper noise on the face and 5% salt-and-pepper noise on other parts, as shown in Figure 8(b). Figure 8(c) shows the recovered image by Gaussian BP under synchronous scheduling, where we set $w_{i,j} = 10^{-6}$ for $(i, j) \in \mathcal{D}$ and $w_{i,j} = 1$ for $(i, j) \notin \mathcal{D}$. Since $w_{i,j} > 0$, by Theorem 10, we know that Gaussian BP converges in solving the problem in (29) under both synchronous and totally asynchronous schedulings. Moreover, for comparison, Figure 8(d) and (e) show the images recovered by the median filter (5×5) for all pixels and only for pixels with values 0 or 255, respectively. It can be seen that the image recovered by Gaussian BP has a much better visual quality than those by the median filters.

7. Conclusions

In this paper, a unified convergence analysis of Gaussian BP in Gaussian MRF and pairwise linear Gaussian model was presented. By using a general pairwise factorization of the joint Gaussian distribution, general convergence conditions of the Gaussian BP beliefs were derived for both synchronous and totally asynchronous schedulings. With the general convergence conditions, existing convergence conditions such as walk-summability, pairwise-normalizability, convex decomposition, and diagonal dominance, were extended from their original considered pairwise factorizations to the proposed general pairwise factorization. Moreover, by further linking the pairwise-normalizability in Gaussian MRF to pairwise linear Gaussian model, an easily verifiable sufficient convergence condition was proposed. Numerical examples and applications were presented to corroborate the newly established convergence results.

Appendix A. Relationships Among Factorizations $\mathbb{F}_1$—$\mathbb{F}_4$

In $\mathbb{F}_1$, if we set $a_{ij} = a_{ji} = 0$, $p_i = J_{ii}$, then $\mathbb{F}_1$ is reduced to $\mathbb{F}_3$, which implies that $\mathbb{F}_3$ is a specific setting of $\mathbb{F}_1$. On the other hand, from the constraint of $\mathbb{F}_2$, ξ_{ij}, $\xi_{ji} > 0$, so the coefficients of the terms x_i^2 and x_j^2 in $f_{ij}(x_i, x_j)$ cannot be zero. Therefore, $\mathbb{F}_2$ cannot reduce to $\mathbb{F}_3$, and vice versa. Furthermore, in the constraint of $\mathbb{F}_4$, due to $\eta_{ij} > 0$, the coefficients of the terms x_i^2 and x_j^2 in $f_{ij}(x_i, x_j)$ cannot be zero. Therefore, $\mathbb{F}_4$ cannot reduce to $\mathbb{F}_3$, and vice versa.

In $\mathbb{F}_2$, if $\varsigma_{ij} = \varsigma_{ji} = 0$, by comparing $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ in $\mathbb{F}_2$ with those in $\mathbb{F}_1$, we have $\xi_{ij} J_{ij}^2 = a_{ij}$, $\xi_{ji} J_{ij}^2 = a_{ji}$, $J_{ii} - \sum_{k \in \mathcal{B}_i} \xi_{ik} J_{ik}^2 = p_i$, therefore $\mathbb{F}_2$ becomes the form as $\mathbb{F}_1$. However, if we set $\varsigma_{ij} \neq 0$ or $\varsigma_{ji} \neq 0$, $f_{ij}(x_i, x_j)$ in $\mathbb{F}_2$ will never become the form as that in $\mathbb{F}_1$ since $f_{ij}(x_i, x_j)$ in $\mathbb{F}_1$ does not contain the first-order terms of variable x_i and x_j. On the other hand, in $\mathbb{F}_1$, if $a_{ij} a_{ji} - J_{ij}^2 < 0$, $\mathbb{F}_1$ will not satisfy the convex decomposition in $\mathbb{F}_2$, which implies $\mathbb{F}_1$ cannot be included in $\mathbb{F}_2$ for these cases. Therefore, $\mathbb{F}_1$ and $\mathbb{F}_2$ are overlapping, but one cannot include the other.

In $\mathbb{F}_1$, if $p_i, a_{ij}, a_{ji} > 0$ and $a_{ij} a_{ji} = J_{ij}^2$, $\mathbb{F}_1$ can be converted into $\mathbb{F}_4$ with $\zeta_i = p_i$, $y_i = h_i / p_i$, $\eta_{ij} = 1$, $z_{ij} = 0$, $c_{ij} = \sqrt{a_{ij}}$ and $c_{ji} = \sqrt{a_{ji}}$. But if $p_i < 0$ in $\mathbb{F}_1$, due to $\zeta_i \geq 0$ in $\mathbb{F}_4$, $f_i(x_i)$ in $\mathbb{F}_1$ cannot be converted into that in $\mathbb{F}_4$, which implies $\mathbb{F}_1$ is not included in $\mathbb{F}_4$. On the other hand, if $z_{ij} \neq 0$ in $\mathbb{F}_4$, $f_{ij}(x_i, x_j)$ in $\mathbb{F}_4$ contains the first-order terms of variable x_i and x_j while $f_{ij}(x_i, x_j)$ in $\mathbb{F}_1$ does not contain the first-order terms of x_i and x_j, which implies $\mathbb{F}_4$ cannot be included in $\mathbb{F}_1$. Therefore, $\mathbb{F}_1$ and $\mathbb{F}_4$ are overlapping, but one cannot include the other.

In $\mathbb{F}_4$, if $\zeta_i = 0$, $f_i(x_i)$ in $\mathbb{F}_4$ does not satisfy the condition $J_{ii} - \sum_{k \in \mathcal{B}_i} \xi_{ik} J_{ik}^2 > 0$ of $f_i(x_i)$ in $\mathbb{F}_2$, which implies $\mathbb{F}_4$ is not included in $\mathbb{F}_2$ for this case. But if $\zeta_i > 0$, then $\mathbb{F}_4$ can be converted into $\mathbb{F}_2$ with $J_{ii} - \sum_{k \in \mathcal{B}_i} \xi_{ik} = \zeta_i > 0$, $h_i - \sum_{k \in \mathcal{B}_i} \varsigma_{ik} = \zeta_i y_i$, $\xi_{ij} J_{ij}^2 = \eta_{ij} c_{ij}^2 > 0$, $\xi_{ji} \eta_{ij}^2 = \eta_{ij} c_{ji}^2 > 0$, $J_{ij} = \eta_{ij} c_{ij} c_{ji}$, $\varsigma_{ij} = \eta_{ij} z_{ij} c_{ij}$ and $\varsigma_{ji} = \eta_{ij} z_{ij} c_{ji}$. On the other hand, comparing $\mathbb{F}_2$ with $\mathbb{F}_4$, if $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ in $\mathbb{F}_2$ can be expressed by those in $\mathbb{F}_4$, we must have $\xi_{ij} J_{ij}^2 = \eta_{ij} c_{ij}^2$, $\xi_{ji} J_{ij}^2 = \eta_{ij} c_{ji}^2$, and $J_{ij} = \eta_{ij} c_{ij} c_{ji}$. From these conditions, we obtain that $\xi_{ij} \xi_{ji} J_{ij}^4 = \eta_{ij}^2 c_{ij}^2 c_{ji}^2 = J_{ij}^2$. Due to $\eta_{ij}, c_{ij}, c_{ji} \neq 0$, $J_{ij} \neq 0$ and the above result reduces to $\xi_{ij} \xi_{ji} J_{ij}^2 = 1$ if $\mathbb{F}_2$ can be expressed by $\mathbb{F}_4$. Hence, when $\xi_{ij} \xi_{ji} J_{ij}^2 > 1$ in $\mathbb{F}_2$, it implies that $\mathbb{F}_2$ cannot be expressed by $\mathbb{F}_4$. Therefore, $\mathbb{F}_2$ and $\mathbb{F}_4$ are overlapping, but one cannot include the other.

Finally, in $\mathbb{F}_4$, if $\zeta_i > 0$ for all $i = 1, 2, \cdots, n$ and $z_{ij} = 0$ for all $(i,j) \in \mathcal{E}$, then $f_{ij}(x_i, x_j)$ reduces to $f_{ij}(x_i, x_j) \propto \exp\left\{-\frac{1}{2}\eta_{ij}\left(c_{ij}x_i + c_{ji}x_j\right)^2\right\}$. Then, comparing $\mathbb{F}_4$ with $\mathbb{F}_1$, if we put $\zeta_i = p_i$, $y_i = h_i/p_i$, $\eta_{ij}c_{ij}^2 = a_{ij}$, $\eta_{ij}c_{ji}^2 = a_{ji}$ and $\eta_{ij}c_{ij}c_{ji} = J_{ij}$, then $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ in $\mathbb{F}_4$ become those in $\mathbb{F}_1$. On the other hand, comparing $\mathbb{F}_4$ with $\mathbb{F}_2$, if we put $\zeta_i = J_{ii} - \sum_{k\in\mathcal{B}_i} \xi_{ik} J_{ik}^2$, $\zeta_i y_i = h_i - \sum_{k\in\mathcal{B}_i} \varsigma_{ik}$, $\eta_{ij}c_{ij}^2 = \xi_{ij}J_{ij}^2$, $\eta_{ij}c_{ji}^2 = \xi_{ji}J_{ij}^2$, $\eta_{ij}c_{ij}c_{ji} = J_{ij}$ and $\varsigma_{ij} = \varsigma_{ji} = 0$, then $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ in $\mathbb{F}_4$ reduce to those in $\mathbb{F}_2$ and satisfy the convex decomposition. Hence, we found a model that is a special case of $\mathbb{F}_4$, but at the same time can be expressed in the form of $\mathbb{F}_1$ and $\mathbb{F}_2$. Therefore, there exists an intersection of $\mathbb{F}_1$, $\mathbb{F}_2$ and $\mathbb{F}_4$.

Appendix B. Derivation of Gaussian BP Messages

By substituting the factors of $\mathbb{F}_5$ and (8) into (7), we obtain the expression

$$
\begin{aligned}
&m_{j\to i}^{(l)}(x_i) \propto \exp\left\{-\frac{1}{2}\gamma_{ij}x_i^2 + \kappa_{ij}x_i\right\} \times \\
&\int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\left(\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\backslash i} \alpha_{k\to j}^{(l-1)}\right)x_j^2 + \left(\psi_j + \kappa_{ji} - \tau_{ij}x_i + \sum_{k\in\mathcal{B}_j\backslash i} \beta_{k\to j}^{(l-1)}\right)x_j\right\} dx_j.
\end{aligned}
\tag{32}
$$

For the integration in (32), it remains finite if and only if $\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\backslash i} \alpha_{k\to j}^{(l-1)} > 0$. If the condition $\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\backslash i} \alpha_{k\to j}^{(l-1)} > 0$ is satisfied, by performing the integration in (32), we have

$$
\begin{aligned}
&m_{j\to i}^{(l)}(x_i) \\
&\propto \exp\left\{-\frac{1}{2}\underbrace{\left(\gamma_{ij} - \frac{\tau_{ij}^2}{\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\backslash i} \alpha_{k\to j}^{(l-1)}}\right)}_{\alpha_{j\to i}^{(l)}} x_i^2 + \underbrace{\left(\kappa_{ij} - \frac{\tau_{ij}\psi_j + \tau_{ij}\kappa_{ji} + \sum_{k\in\mathcal{B}_j\backslash i} \tau_{ij}\beta_{k\to j}^{(l-1)}}{\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\backslash i} \alpha_{k\to j}^{(l-1)}}\right)}_{\beta_{j\to i}^{(l)}} x_i\right\},
\end{aligned}
\tag{33}
$$

where the message $m_{j\to i}^{(l)}(x_i)$ is maintained the same form as in (8) and parameterized by a pair of parameters $\alpha_{j\to i}^{(l)}$ and $\beta_{j\to i}^{(l)}$. If $\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\backslash i} \alpha_{k\to j}^{(l-1)} \le 0$, the message $m_{j\to i}^{(l)}(x_i)$ is not defined.

Appendix C. Proof of Theorem 1

From (9) and (10), the message parameters update can be rewritten as

$$\alpha_{j\to i}^{(l)} - \gamma_{ij} = -\frac{\tau_{ij}^2}{\phi_j + \sum\limits_{k\in\mathcal{B}_j} \gamma_{jk} + \sum\limits_{k\in\mathcal{B}_j\setminus i} (\alpha_{k\to j}^{(l-1)} - \gamma_{jk})}, \tag{34}$$

$$\beta_{j\to i}^{(l)} - \kappa_{ij} = -\frac{\tau_{ij}\psi_j + \tau_{ij}\sum\limits_{k\in\mathcal{B}_j} \kappa_{jk} + \sum\limits_{k\in\mathcal{B}_j\setminus i} \tau_{ij}(\beta_{k\to j}^{(l-1)} - \kappa_{jk})}{\phi_j + \sum\limits_{k\in\mathcal{B}_j} \gamma_{jk} + \sum\limits_{k\in\mathcal{B}_j\setminus i} (\alpha_{k\to j}^{(l-1)} - \gamma_{jk})}. \tag{35}$$

Furthermore, from (13) and (14), the belief parameters update can be rewritten as $\nu_{x_i}^{(l)} = \phi_i + \sum_{j\in\mathcal{B}_i} \gamma_{ij} + \sum_{j\in\mathcal{B}_i} (\alpha_{j\to i}^{(l)} - \gamma_{ij})$ and $\varphi_{x_i}^{(l)} = \psi_i + \sum_{j\in\mathcal{B}_i} \kappa_{ij} + \sum_{j\in\mathcal{B}_i} (\beta_{j\to i}^{(l)} - \kappa_{ij})$. By taking $\alpha_{j\to i}^{(l)} - \gamma_{ij}$ and $\beta_{j\to i}^{(l)} - \kappa_{ij}$ as two new sequences in the updates of messages and beliefs, it is observed that the updates of messages and beliefs are in the same form as those in Su and Wu (2014) if we set $\phi_i + \sum_{k\in\mathcal{B}_i} \gamma_{ik} = J_{ii}$, $\tau_{ij} = J_{ij}$ for $i \neq j$ and $\psi_i + \sum_{k\in\mathcal{B}_i} \kappa_{ik} = h_i$. By using Su and Wu (2014, Theorem 4), under any initialization satisfying $\boldsymbol{\alpha}^{(0)} - \boldsymbol{\gamma} \geq \mathbf{0}$, we obtain that belief variances $\{1/\nu_{x_i}^{(l)}\}$ converge to the same positive values if and only if the set $\widetilde{\mathcal{Q}} = \{\mathbf{q} \in \mathbb{R}^{|\mathcal{E}|} \mid J_{ii} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i} > 0, -\frac{J_{ij}^2}{J_{ii} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i}} \geq q_{i\to j}, \forall (i,j) \in \mathcal{E},$ and $J_{ii} + \sum_{k\in\mathcal{B}_i} q_{k\to i} > 0$ for $i = 1, 2, \cdots, n\}$ is non-empty. Furthermore, by Su and Wu (2014, Lemma 3), the constraint $J_{ii} + \sum_{k\in\mathcal{B}_i} q_{k\to i} > 0$ will be either satisfied for all $i = 1, 2, \cdots, n$ or dissatisfied for all $i = 1, 2, \cdots, n$. Without loss of generality, checking $i = 1$ is sufficient. By putting $\phi_i + \sum_{k\in\mathcal{B}_i} \gamma_{ik} = J_{ii}$, $\tau_{ij} = J_{ij}$ for $i \neq j$ into $\widetilde{\mathcal{Q}}$, it leads to the set $\mathcal{Q} \triangleq \{\mathbf{q} \in \mathbb{R}^{|\mathcal{E}|} \mid \phi_i + \sum_{k\in\mathcal{B}_i} \gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i} > 0, -\frac{\tau_{ij}^2}{\phi_i + \sum_{k\in\mathcal{B}_i} \gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i}} \geq q_{i\to j}, \forall (i,j) \in \mathcal{E},$ and $\phi_i + \sum_{k\in\mathcal{B}_i} \gamma_{ik} + \sum_{k\in\mathcal{B}_i} q_{k\to i} > 0$ for $i = 1\}$. Furthermore, if the set $\mathcal{Q}$ is non-empty, $\boldsymbol{\alpha}^{(l)}$ converges, and the update of $\boldsymbol{\beta}^{(l)}$ becomes the linear update equation in (16). With Bertsekas and Tsitsiklis (1989, Proposition 2.6.1), $\boldsymbol{\beta}^{(l)}$ converges to a unique value if and only if $\rho(\mathbf{A}) < 1$ under synchronous scheduling. With a similar proof to Su and Wu (2015a, Theorem 3), belief means converge to unique values if and only if $\boldsymbol{\beta}^{(l)}$ converges to a unique value. Therefore, belief means converge to unique values if and only if $\rho(\mathbf{A}) < 1$ under synchronous scheduling.

Appendix D. Proof of Theorem 2

By the proof in Theorem 1, it is shown that the update of beliefs is in the same form as that in Su and Wu (2014) if we set $f_i(x_i) \propto \exp\{-\frac{1}{2}(\phi_i + \sum_{k\in\mathcal{B}_i} \gamma_{ik})x_i^2 + (\psi_i + \sum_{k\in\mathcal{B}_i} \kappa_{ik})x_i\}$ and $f_{ij}(x_i, x_j) \propto \exp\{-\tau_{ij}x_i x_j\}$. By Su and Wu (2014, Theorem 4), under any initialization $\boldsymbol{\alpha}^{(0)} - \boldsymbol{\gamma} \geq \mathbf{0}$, belief variances converge to the same positive values under totally asynchronous scheduling if and only if $\mathcal{Q}$ is non-empty. Moreover, if $\mathcal{Q}$ is non-empty, $\boldsymbol{\alpha}^{(l)}$ converges and thus the update of $\boldsymbol{\beta}^{(l)}$ becomes a linear equation as shown in (16). According to the asynchronous convergence theorem in Bertsekas and Tsitsiklis (1989, Proposition 6.2.1), $\boldsymbol{\beta}^{(l)}$ converge to a unique value if $\rho(|\mathbf{A}|) < 1$ under totally asynchronous scheduling. Furthermore, with the converged value of x_i's belief variance being $1/\nu_{x_i}^*$, the x_i's belief mean at the l-th iteration is computed as $\varphi_{x_i}^{(l)}/\nu_{x_i}^* = (\psi_i + \sum_{j\in\mathcal{B}_i}\beta_{j\to i}^{(l)})/\nu_{x_i}^*$. If $\boldsymbol{\beta}^{(l)}$ converges, belief means would also converge. Therefore, belief means converge to unique values if $\rho(|\mathbf{A}|) < 1$ under totally asynchronous scheduling.

Appendix E. Proof of Corollary 3

Since synchronous scheduling is a special case of totally asynchronous scheduling, we only need to prove the result under totally asynchronous scheduling. We define a positive integer set $\mathcal{T}_{j\to i}$ denoting the iteration indices when the message $m_{j\to i}^{(l)}(x_i)$ is updated under totally asynchronous scheduling. If $l \in \mathcal{T}_{j\to i}$, we have the update of $\alpha_{j\to i}^{(l)}$ as

$$\alpha_{j\to i}^{(l)} = \gamma_{ij} - \frac{\tau_{ij}^2}{\phi_j + \gamma_{ji} + \sum\limits_{k\in\mathcal{B}_j\setminus i}\alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l-1))}}, \tag{36}$$

where $0 \le T_{k\to j}^{j\to i}(l-1) \le l-1$ and $\alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l-1))}$ denotes the most recently received value of $\alpha_{k\to j}$ for the update of $\alpha_{j\to i}$ at the l-th iteration. Otherwise, $\alpha_{j\to i}^{(l)}$ will be maintained as the previous value at the $(l-1)$-th iteration.

Necessary condition:

We need to prove that if the set $\mathcal{Q}$ is non-empty, $\nu_{x_i}^{(l)} > 0$ for all $l \ge 0$. Firstly, we can rewrite the update of $\nu_{x_i}^{(l)}$ in (25) as $\nu_{x_i}^{(l)} = \phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i}(\alpha_{k\to i}^{(l)} - \gamma_{ik})$. If we can prove that $\alpha_{k\to i}^{(l)} - \gamma_{ik} > q_{k\to i}$, where $q_{k\to i}$ is the component of any $\mathbf{q} \in \mathcal{Q}$, then $\nu_{x_i}^{(l)} > \phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i} q_{k\to i} > 0$, where the last inequality is due to the third constraint of $\mathcal{Q}$. Hence, our task is to prove $\alpha_{k\to i}^{(l)} - \gamma_{ik} > q_{k\to i}$ for all $l \ge 0$. This is done by induction. First, consider $l = 0$. By putting the first constraint of $\mathcal{Q}$ into its second constraint, we must have $q_{i\to j} \le -\frac{\tau_{ij}^2}{\phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i}} < 0$. On the other hand, under any initialization $\boldsymbol{\alpha}^{(0)} \ge \boldsymbol{\gamma}$, we have $\alpha_{i\to j}^{(0)} - \gamma_{ji} \ge 0$. Combining with $q_{i\to j} < 0$, we obtain $\alpha_{i\to j}^{(0)} - \gamma_{ji} > q_{i\to j}$. Therefore, the statement is true for $l = 0$. Now, suppose that $\alpha_{i\to j}^{(l-1)} - \gamma_{ji} > q_{i\to j}$ is true. On one hand, if $l \notin \mathcal{T}_{i\to j}$, then $\alpha_{i\to j}^{(l)}$ is not updated at the l-th iteration and $\alpha_{i\to j}^{(l)} = \alpha_{i\to j}^{(l-1)}$. Hence, we can directly obtain $\alpha_{i\to j}^{(l)} - \gamma_{ji} > q_{i\to j}$. On the other hand, if $l \in \mathcal{T}_{i\to j}$, then $\alpha_{i\to j}^{(l)}$ is updated at the l-th iteration. Due to the assumption $\alpha_{i\to j}^{(l-1)} - \gamma_{ji} > q_{i\to j}$ and $T_{k\to i}^{i\to j}(l-1) \le l-1$, we have $\alpha_{k\to i}^{(T_{k\to i}^{i\to j}(l-1))} - \gamma_{ik} > q_{k\to i}$ and then $\phi_i + \gamma_{ij} + \sum_{k\in\mathcal{B}_i\setminus j}\alpha_{k\to i}^{(T_{k\to i}^{i\to j}(l-1))} > \phi_i + \gamma_{ij} + \sum_{k\in\mathcal{B}_i\setminus j}(q_{k\to i} + \gamma_{ik})$. Furthermore, due to the third constraint of $\mathcal{Q}$, we can obtain $\phi_i + \gamma_{ij} + \sum_{k\in\mathcal{B}_i\setminus j}\alpha_{k\to i}^{(T_{k\to i}^{i\to j}(l-1))} > \phi_i + \gamma_{ij} + \sum_{k\in\mathcal{B}_i\setminus j}(q_{k\to i} + \gamma_{ik}) > 0$, which implies $-\frac{\tau_{ij}^2}{\phi_i + \gamma_{ij} + \sum_{k\in\mathcal{B}_i\setminus j}\alpha_{k\to i}^{(T_{k\to i}^{i\to j}(l-1))}} > -\frac{\tau_{ij}^2}{\phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i}}$. Since we have $-\frac{\tau_{ij}^2}{\phi_i + \sum_{k\in\mathcal{B}_i}\gamma_{ik} + \sum_{k\in\mathcal{B}_i\setminus j} q_{k\to i}} \ge q_{i\to j}$ due to the second constraint of $\mathcal{Q}$, we can obtain $-\frac{\tau_{ij}^2}{\phi_i + \gamma_{ij} + \sum_{k\in\mathcal{B}_i\setminus j}\alpha_{k\to i}^{(T_{k\to i}^{i\to j}(l-1))}} > q_{i\to j}$. Noticing that the left hand side of this equation is simply $\alpha_{i\to j}^{(l)} - \gamma_{ji}$ (see (36)) under totally asynchronous scheduling, we obtain $\alpha_{i\to j}^{(l)} - \gamma_{ji} > q_{i\to j}$. Therefore, we have proved $\alpha_{k\to i}^{(l)} - \gamma_{ik} > q_{k\to i}$ for all $l \geq 0$.

Sufficient condition:

We need to prove that if BP beliefs stay as valid Gaussian pdfs at all iterations, the set $\mathcal{Q}$ will be non-empty. When BP beliefs stay as valid Gaussian pdfs at all iterations under totally asynchronous scheduling, we can obtain $\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l))} > 0$ and $\phi_i + \sum_{k\in\mathcal{B}_i} \alpha_{k\to i}^{(l)} > 0$ for all $l \geq 0$. When $l = 1$, if $1 \in \mathcal{T}_{j\to i}$, then $\alpha_{j\to i}^{(l)}$ is updated. Due to $\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(0))} > 0$, further by the update equation (36), we can obtain $\alpha_{j\to i}^{(1)} = \gamma_{ij} - \frac{\tau_{ij}^2}{\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(0))}} < \gamma_{ij}$. Due to $\alpha_{j\to i}^{(0)} \geq \gamma_{ij}$, we have $\alpha_{j\to i}^{(1)} < \alpha_{j\to i}^{(0)}$. If $1 \notin \mathcal{T}_{j\to i}$, we have $\alpha_{j\to i}^{(1)} = \alpha_{j\to i}^{(0)}$. Therefore, $\alpha_{j\to i}^{(l)} \leq \alpha_{j\to i}^{(l-1)}$ is true for $l = 1$. Suppose that $\alpha_{j\to i}^{(l)} \leq \alpha_{j\to i}^{(l-1)}$ is true. If $l+1 \notin \mathcal{T}_{j\to i}$, $\alpha_{j\to i}^{(l+1)}$ is not updated and $\alpha_{j\to i}^{(l+1)} = \alpha_{j\to i}^{(l)}$. If $l+1 \in \mathcal{T}_{j\to i}$, then $\alpha_{j\to i}^{(l+1)}$ is updated. We divide the discussion into two cases:

1) If $\alpha_{j\to i}^{(l)}$ is updated for the first time at the $(l+1)$-th iteration, we have $\alpha_{j\to i}^{(l)} = \alpha_{j\to i}^{(0)}$. From (36), we can obtain $\alpha_{j\to i}^{(l+1)} = \gamma_{ij} - \frac{\tau_{ij}^2}{\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{i\to j}^{j\to i}(l))}} < \gamma_{ij}$. Due to $\alpha_{j\to i}^{(0)} > \gamma_{ij}$, we can obtain $\alpha_{j\to i}^{(l+1)} < \alpha_{j\to i}^{(l)}$.

2) If $\alpha_{j\to i}^{(l)}$ has been updated before the $(l+1)$-th iteration, we suppose that l_1 is the most recent iteration for the update of $\alpha_{j\to i}^{(l)}$, where $l_1 \leq l$. Since $T_{k\to j}^{j\to i}(l) \geq T_{k\to j}^{j\to i}(l_1 - 1)$ and due to the assumption $\alpha_{k\to j}^{(l)} \leq \alpha_{k\to j}^{(l-1)}$, we can obtain $\alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l))} \leq \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l_1-1))}$. Further due to $\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l))} > 0$ and $\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l_1-1))} > 0$, we can obtain $\gamma_{ij} - \frac{\tau_{ij}^2}{\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l))}} \leq \gamma_{ij} - \frac{\tau_{ij}^2}{\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i} \alpha_{k\to j}^{(T_{k\to j}^{j\to i}(l_1-1))}}$, i.e., $\alpha_{j\to i}^{(l+1)} \leq \alpha_{j\to i}^{(l_1)}$. Moreover, since $\alpha_{j\to i}^{(l)}$ is not updated after the l_1-th iteration and before the $(l+1)$-th iteration, we have $\alpha_{j\to i}^{(l)} = \alpha_{j\to i}^{(l_1)}$. Hence, we can obtain $\alpha_{j\to i}^{(l+1)} \leq \alpha_{j\to i}^{(l)}$.

Therefore, $\alpha_{j\to i}^{(l)}$ is a monotonically decreasing sequence under totally asynchronous scheduling.

Furthermore, if $\alpha_{j\to i}^{(l)}$ is not bounded below, then it goes to $-\infty$ when the iterations goes to infinity, which leads to $\phi_i + \sum_{k\in\mathcal{B}_i} \alpha_{k\to i}^{(l)} < 0$ when l goes to infinity. This leads to a contradiction with the condition $\phi_i + \sum_{k\in\mathcal{B}_i} \alpha_{k\to i}^{(l)} > 0$ for all $l \geq 0$. Therefore, $\alpha_{j\to i}^{(l)}$ must be bounded below. Together with the monotonically decreasing property of $\alpha_{j\to i}^{(l)}$, $\alpha_{j\to i}^{(l)}$ converges. Here, we suppose that $\boldsymbol{\alpha}^*$ is the converged value of $\boldsymbol{\alpha}^{(l)}$. Defining $\alpha_{j\to i}^* = q_{j\to i} + \gamma_{ij}$, and putting it into the condition guaranteeing the validity of integration in (32) and $\nu_{x_i}^{(l)} > 0$, we can obtain $\phi_i+\gamma_{ij}+\sum_{k\in\mathcal{B}_i\setminus j}(q_{k\to i}+\gamma_{ik}) > 0$ and $\phi_i+\sum_{k\in\mathcal{B}_i}(q_{k\to i}+\gamma_{ik}) > 0$ for $l \geq 0$. Moreover, putting $\alpha_{j\to i}^* = q_{j\to i} + \gamma_{ij}$ into the update equation (36), we have

$q_{j\to i} = -\frac{\tau_{ij}^2}{\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i}(q_{k\to j}+\gamma_{jk})}$. But these are in fact the three constraints of the set $\mathcal{Q}$. Therefore, we have found a $\mathbf{q}\in\mathcal{Q}$ with the element $q_{j\to i}=\alpha^*_{j\to i}-\gamma_{ij}$, and the set $\mathcal{Q}$ is non-empty.

Appendix F. Proof of Lemma 5

For synchronous scheduling, by Corollary 3, under any initialization $\boldsymbol{\alpha}^{(0)}\geq\boldsymbol{\gamma}$, if the set $\mathcal{Q}$ is non-empty, BP beliefs stay being valid Gaussian pdfs. Furthermore, from the proof of sufficient condition in Corollary 3, it is proved that BP beliefs stay in valid Gaussian form leads to $\boldsymbol{\alpha}^{(l)}$ converges. Then, the matrix $\mathbf{A}$ consists of elements $\frac{\tau_{ij}}{\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i}\alpha^*_{k\to j}}$ for all $(i,j)\in\mathcal{E}$, where $\boldsymbol{\alpha}^*$ is the converged value of $\boldsymbol{\alpha}^{(l)}$. Since $\tau_{ij}=J_{ij}$ and $\phi_j+\gamma_{ji}+\sum_{k\in\mathcal{B}_j\setminus i}\alpha^*_{k\to j}=\phi_j+\sum_{k\in\mathcal{B}_j}\gamma_{jk}+\sum_{k\in\mathcal{B}_j\setminus i}(\alpha^*_{k\to j}-\gamma_{jk})=J_{jj}+\sum_{k\in\mathcal{B}_j\setminus i}(\alpha^*_{k\to j}-\gamma_{jk})$ for a Gaussian pdf with information matrix $\mathbf{J}$, the elements of $\mathbf{A}$ become $\frac{J_{ij}}{J_{jj}+\sum_{k\in\mathcal{B}_j\setminus i}(\alpha^*_{k\to j}-\gamma_{jk})}$ for all $(i,j)\in\mathcal{E}$. On the other hand, from (34), if we treat $\alpha^{(l)}_{j\to i}-\gamma_{ij}\triangleq\widetilde{\alpha}^{(l)}_{j\to i}$ as a whole and putting $\phi_i+\sum_{k\in\mathcal{B}_i}\gamma_{ik}=J_{ii}$ and $\tau_{ij}=J_{ij}$, the resultant equation becomes (12) in Su and Wu (2014). Furthermore, putting $\phi_i+\sum_{k\in\mathcal{B}_i}\gamma_{ik}=J_{ii}$ and $\tau_{ij}=J_{ij}$, the set $\mathcal{Q}$ reduces to $\widetilde{\mathcal{Q}}$ defined in the proof of Theorem 1. Therefore, non-empty $\mathcal{Q}$ implies that $\widetilde{\mathcal{Q}}$ is non-empty. By Su and Wu (2014, Theorem 2), under any initialization $\widetilde{\boldsymbol{\alpha}}^{(0)}\geq\mathbf{0}$ and non-empty $\widetilde{\mathcal{Q}}$, $\widetilde{\boldsymbol{\alpha}}^{(l)}$ converges to the same value $\widetilde{\boldsymbol{\alpha}}^*$. Recognizing that $\widetilde{\boldsymbol{\alpha}}^*=\boldsymbol{\alpha}^*-\boldsymbol{\gamma}$ is the same for any $\boldsymbol{\gamma}$, we can conclude that under synchronous scheduling, the elements of matrix $\mathbf{A}$ are independent of factorization parameters for any initialization $\boldsymbol{\alpha}^{(0)}\geq\boldsymbol{\gamma}$. Since the matrix $\mathbf{A}$ in totally asynchronous scheduling is the same as that in synchronous scheduling, Lemma 5 also holds for totally asynchronous scheduling.

Appendix G. Proof of Corollary 6

For a Gaussian pdf with information matrix $\mathbf{J}$, under any initialization $\boldsymbol{\alpha}^{(0)}\geq\boldsymbol{\gamma}$, by Lemmas 4 and 5, $\mathcal{Q}$ and $\mathbf{A}$ are independent of factorization parameters in $\mathbb{F}_5$. Furthermore, by Theorem 1, belief variances converge if and only if the set $\mathcal{Q}$ is non-empty and belief means converge if and only if $\rho(\mathbf{A})<1$. Therefore, the convergence or divergence of Gaussian BP under synchronous scheduling is independent of factorization parameters in $\mathbb{F}_5$.

Appendix H. Proof of Corollary 7

Since synchronous scheduling is a special case of totally asynchronous scheduling, proving the result under totally asynchronous scheduling is sufficient.

Under totally asynchronous scheduling, walk-summability guarantees that the set $\mathcal{Q}$ is non-empty and $\rho(|\mathbf{A}|)<1$ under pairwise factorization $\mathbb{F}_3$ and initialization $\boldsymbol{\alpha}^{(0)}\geq\mathbf{0}$ (Su and Wu, 2015a, Theorem 6). Using Lemma 4, $\mathcal{Q}$ is non-empty under any pairwise factorization within $\mathbb{F}_5$. Further by Lemma 5, we have $\rho(|\mathbf{A}|)<1$ under any pairwise factorization within $\mathbb{F}_5$ and any initialization $\boldsymbol{\alpha}^{(0)}\geq\boldsymbol{\gamma}$. Then applying Theorem 2, we obtain walk-summability leads to Gaussian BP convergence under any pairwise factorization within $\mathbb{F}_5$ and any initialization $\boldsymbol{\alpha}^{(0)}\geq\boldsymbol{\gamma}$. Since walk-summability is equivalent to convex decomposition (Moallemi and Roy, 2009) or pairwise-normalizability (Malioutov et al., 2006), and includes diagonal dominance as a special case (Malioutov et al., 2006), Gaussian BP converges under any one of these conditions, and the convergence result is valid under any pairwise factorization within $\mathbb{F}_5$ and any initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$.

Appendix I. Proof of Theorem 8

As given in the proof of Lemma 5, for a Gaussian pdf with information matrix $\mathbf{J}$, under any initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, if Gaussian beliefs converge, then $\widetilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha}^* - \boldsymbol{\gamma}$ is the same for any $\boldsymbol{\gamma}$. On the other hand, when belief variances converge under synchronous scheduling, we have $\boldsymbol{\beta}^{(l)} = \mathbf{A}\boldsymbol{\beta}^{(l-1)} + \mathbf{d}$. By taking $\beta_{j\to i}^{(l)} - \kappa_{ij} \triangleq \widetilde{\beta}_{j\to i}^{(l)}$, we have the update equation $\widetilde{\boldsymbol{\beta}}^{(l)} = \mathbf{A}\widetilde{\boldsymbol{\beta}}^{(l-1)} + \widetilde{\mathbf{d}}$, where $\widetilde{\mathbf{d}}$ contains the elements $-\frac{\tau_{ij}(\psi_j + \sum_{k\in\mathcal{B}_j}\kappa_{jk})}{\phi_j + \gamma_{ji} + \sum_{k\in\mathcal{B}_j\setminus i}\alpha_{k\to j}^*}$ for all $(i,j) \in \mathcal{E}$ ordered in the same way as that in $\boldsymbol{\alpha}^{(l)}$. Since $J_{jj} = \phi_j + \sum_{k\in\mathcal{B}_j}\gamma_{jk}$, $J_{ij} = \tau_{ij}$ and $\psi_j + \sum_{k\in\mathcal{B}_j}\kappa_{jk} = h_j$ for a Gaussian pdf with information matrix $\mathbf{J}$ and potential vector $\mathbf{h}$ (see (20) and (21)), the elements of $\widetilde{\mathbf{d}}$ become $-\frac{J_{ij}h_j}{J_{jj} + \sum_{k\in\mathcal{B}_j\setminus i}(\alpha_{k\to j}^* - \gamma_{jk})}$ for all $(i,j) \in \mathcal{E}$. Due to J_{jj}, J_{ij}, h_j are fixed and $\boldsymbol{\alpha}^* - \boldsymbol{\gamma}$ being the same under initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, the elements of $\widetilde{\mathbf{d}}$ are independent of factorization parameters in $\mathbb{F}_5$. Furthermore, from Theorem 1, when Gaussian BP beliefs converge under synchronous scheduling, we have $\rho(\mathbf{A}) < 1$. This leads to $\widetilde{\boldsymbol{\beta}}^{(l)}$ converge to $\widetilde{\boldsymbol{\beta}}^* = (\mathbf{I} - \mathbf{A})^{-1}\widetilde{\mathbf{d}}$ (Bertsekas and Tsitsiklis, 1989, Proposition 6.1). Since $\mathbf{A}$ and $\widetilde{\mathbf{d}}$ are independent of factorization parameters, we can obtain $\widetilde{\boldsymbol{\beta}}^* = \boldsymbol{\beta}^* - \boldsymbol{\kappa}$ is unique for any $\boldsymbol{\kappa}$. Moreover, Su and Wu (2015a, Theorem 5) has proved that the converged $\widetilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha}^* - \boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\beta}}^* = \boldsymbol{\beta}^* - \boldsymbol{\kappa}$ under totally asynchronous scheduling and any initialization $\boldsymbol{\alpha}^{(0)} - \boldsymbol{\gamma} \geq \mathbf{0}$ are the same values as those under synchronous scheduling. Therefore, under any initialization $\boldsymbol{\alpha}^{(0)} \geq \boldsymbol{\gamma}$, the converged $\widetilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha}^* - \boldsymbol{\gamma}$ and $\widetilde{\boldsymbol{\beta}}^* = \boldsymbol{\beta}^* - \boldsymbol{\kappa}$ are the same for different pairwise factorizations under both synchronous and totally asynchronous schedulings. Since $\phi_i + \sum_{j\in\mathcal{B}_i}\gamma_{ij} = J_{ii}$ and $\psi_i + \sum_{j\in\mathcal{B}_i}\kappa_{ij} = h_i$ are fixed for a Gaussian pdf with information matrix $\mathbf{J}$ and potential vector $\mathbf{h}$, and $\widetilde{\boldsymbol{\alpha}}^* = \boldsymbol{\alpha}^* - \boldsymbol{\gamma}$, $\widetilde{\boldsymbol{\beta}}^* = \boldsymbol{\beta}^* - \boldsymbol{\kappa}$ are unique when Gaussian BP converges, x_i's converged belief mean $\frac{\psi_i + \sum_{j\in\mathcal{B}_i}\beta_{j\to i}^*}{\phi_i + \sum_{j\in\mathcal{B}_i}\alpha_{j\to i}^*} = \frac{\psi_i + \sum_{j\in\mathcal{B}_i}\kappa_{ij} + \sum_{j\in\mathcal{B}_i}(\beta_{j\to i}^* - \kappa_{ij})}{\phi_i + \sum_{j\in\mathcal{B}_i}\gamma_{ij} + \sum_{j\in\mathcal{B}_i}(\alpha_{j\to i}^* - \gamma_{ij})} = \frac{h_i + \sum_{j\in\mathcal{B}_i}\widetilde{\beta}_{j\to i}^*}{J_{ii} + \sum_{j\in\mathcal{B}_i}\widetilde{\alpha}_{j\to i}^*}$ and belief variance $\frac{1}{\phi_i + \sum_{j\in\mathcal{B}_i}\alpha_{j\to i}^*} = \frac{1}{\phi_i + \sum_{j\in\mathcal{B}_i}\gamma_{ij} + \sum_{j\in\mathcal{B}_i}(\alpha_{j\to i}^* - \gamma_{ij})} = \frac{1}{J_{ii} + \sum_{j\in\mathcal{B}_i}\widetilde{\alpha}_{j\to i}^*}$ are unique and independent of factorization parameters under both synchronous and totally asynchronous schedulings.

Appendix J. Proof of Lemma 9

For any valid Gaussian pdf, $\mathbf{x}^T\mathbf{J}\mathbf{x}$ with $\mathbf{J}$ defined in (20) can be written as

$$\mathbf{x}^T\mathbf{J}\mathbf{x} = \sum_{i=1}^{n}(J_{ii} - \sum_{j\in\mathcal{B}_i}\omega_{ij})x_i^2 + \sum_{(i,j>i)\in\mathcal{E}}[x_i\ x_j]\begin{bmatrix}\omega_{ij} & J_{ij}\\ J_{ij} & \omega_{ji}\end{bmatrix}\begin{bmatrix}x_i\\ x_j\end{bmatrix} \text{ for all } \omega_{ij}. \tag{37}$$

While there are many possible ways to write (37) for an information matrix $\mathbf{J}$, the Gaussian model is pairwise-normalizable if there exists at least one decomposition (37) with $\omega_{ij} \geq 0$ for all $(i,j) \in \mathcal{E}$, $\omega_{ij}\omega_{ji} - J_{ij}^2 \geq 0$ for all $(i, j > i) \in \mathcal{E}$ and $J_{ii} - \sum_{j\in\mathcal{B}_i}\omega_{ij} > 0$ for all

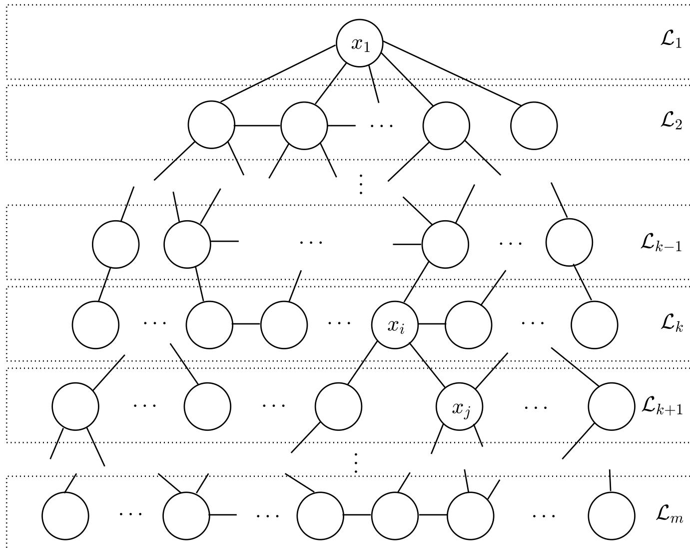


Figure 9: Clustering variables into multiple layers.

$i = 1, 2, \cdots, n$ (Koller and Friedman, 2009). Putting $J_{ii} = \phi_i + \sum_{j \in \mathcal{B}_i} \gamma_{ij}$ and $J_{ij} = \tau_{ij}$ into the above conditions, we have the constraints as in (23a)—(23c) if a Gaussian model is pairwise-normalizable.

Appendix K. Proof of Theorem 10

We put $\phi_i = \zeta_i$, $\gamma_{ij} = \eta_{ij} c_{ij}^2$, $\tau_{ij} = \eta_{ij} c_{ij} c_{ji}$ into (23a)—(23c) to reduce the conditions for the special case of pairwise linear Gaussian model $\mathbb{F}_4$:

$$\omega_{ij} \geq 0 \text{ for all } (i, j) \in \mathcal{E}, \tag{38a}$$

$$\omega_{ij}\omega_{ji} - \eta_{ij}^2 c_{ij}^2 c_{ji}^2 \geq 0 \text{ for all } (i, j > i) \in \mathcal{E}, \tag{38b}$$

$$\zeta_i + \sum_{j \in \mathcal{B}_i} \eta_{ij} c_{ij}^2 - \sum_{j \in \mathcal{B}_i} \omega_{ij} > 0, \text{ for all } i = 1, 2, \cdots, n. \tag{38c}$$

Below we will find ω_{ij} such that (38a)—(38c) are satisfied.

To find the ω_{ij}, we cluster all the variables into multiple layers as shown in Figure 9. Without loss of generality, we assume $\zeta_1 > 0$ and put x_1 into the first layer. Then we define a variable index set $\mathcal{L}_1 = \{1\}$ for variable x_1 in the first layer and the collection of variables indices in the second layer as a set $\mathcal{L}_2 = \{i \mid i \in \mathcal{B}_1\}$. In general, the set of variables' indices in the k-th layer is defined as $\mathcal{L}_k \triangleq \{i | i \in \mathcal{B}_t \setminus \{\mathcal{L}_{k-1} \cup \mathcal{L}_{k-2}\}$, for all $t \in \mathcal{L}_{k-1}\}$ with $2 \leq k \leq m$, where m is the number of total layers and $\mathcal{L}_1 \cup \mathcal{L}_2 \cup \cdots \cup \mathcal{L}_m = \{1, 2, \cdots, n\}$. Furthermore, based on the network in Figure 9, we can decompose the set $\mathcal{E} = \bigcup_{k=2}^{m} (\mathcal{E}_{k,k} \cup \mathcal{E}_{k-1,k}^{+} \cup \mathcal{E}_{k-1,k}^{-})$, where $\mathcal{E}_{k,k} \triangleq \{(i,j) \in \mathcal{E} \mid i, j \in \mathcal{L}_k\}$, $\mathcal{E}_{k-1,k}^{+} = \{(i,j) \in \mathcal{E} \mid i \in \mathcal{L}_{k-1}, j \in \mathcal{L}_k\}$, and $\mathcal{E}_{k-1,k}^{-} = \{(i,j) \in \mathcal{E} \mid i \in \mathcal{L}_k, j \in \mathcal{L}_{k-1}\}$. For the preparation of finding ω_{ij}, we first define ϵ_{ij} for all $(i,j) \in \mathcal{E}$ as follows.

For $(i,j) \in \mathcal{E}^-_{k-1,k}$ with $k \geq 2$, we define

$$\epsilon_{ij} = \begin{cases} -\frac{\eta_{i1}c_{i1}^2\zeta_1/(|\mathcal{B}_1|+1)}{\eta_{i1}c_{1i}^2+\zeta_1/(|\mathcal{B}_1|+1)} & \text{if } (i,j) \in \mathcal{E}^-_{1,2}, \\ \frac{\eta_{ij}c_{ij}^2\sum_{(j,t)\in\mathcal{E}^-_{k-2,k-1}}\epsilon_{jt}/(|\mathcal{B}_j\setminus\mathcal{L}_{k-2}|+1)}{\eta_{ij}c_{ji}^2-\sum_{(j,t)\in\mathcal{E}^-_{k-2,k-1}}\epsilon_{jt}/(|\mathcal{B}_j\setminus\mathcal{L}_{k-2}|+1)} & \text{otherwise.} \end{cases} \tag{39}$$

We can prove that $\epsilon_{ij} < 0$ for all $(i,j) \in \mathcal{E}^-_{k-1,k}$ with $k \geq 2$ by induction. When $k = 2$, we have $j = 1$. Due to $\eta_{i1} > 0$ and $\zeta_1 > 0$, we can obtain $\epsilon_{i1} = -\frac{\eta_{i1}c_{i1}^2\zeta_1/(|\mathcal{B}_1|+1)}{\eta_{i1}c_{1i}^2+\zeta_1/(|\mathcal{B}_1|+1)} < 0$. Therefore, $\epsilon_{i1} < 0$ is true for $k = 2$. Suppose $\epsilon_{ij} < 0$ is true for all $(i,j) \in \mathcal{E}^-_{k-1,k}$ for some $k \geq 2$. Then, for $(i,j) \in \mathcal{E}^-_{k,k+1}$, we have $\epsilon_{ij} = \frac{\eta_{ij}c_{ij}^2\sum_{(j,t)\in\mathcal{E}^-_{k-1,k}}\epsilon_{jt}/(|\mathcal{B}_j\setminus\mathcal{L}_{k-1}|+1)}{\eta_{ij}c_{ji}^2-\sum_{(j,t)\in\mathcal{E}^-_{k-1,k}}\epsilon_{jt}/(|\mathcal{B}_j\setminus\mathcal{L}_{k-1}|+1)}$. Since on the right hand side of this equation, $(j,t) \in \mathcal{E}^-_{k-1,k}$, by assumption, $\epsilon_{jt} < 0$. Furthermore, since $\eta_{ij} > 0$, we can obtain $\epsilon_{ij} < 0$. Therefore, we have proved $\epsilon_{ij} < 0$ for all $(i,j) \in \mathcal{E}^-_{k-1,k}$ with $k \geq 2$.

Based on ϵ_{ij} with $(i,j) \in \mathcal{E}^-_{k-1,k}$, we can define ϵ_{ij} for $(i,j) \in \mathcal{E}^+_{k-1,k}$ with $k \geq 2$ as

$$\epsilon_{ij} = \begin{cases} 0 & \text{if } (i,j) \in \mathcal{E}^+_{1,2}, \\ -\frac{\sum_{(i,t)\in\mathcal{E}^-_{k-2,k-1}}\epsilon_{it}}{|\mathcal{B}_i\setminus\mathcal{L}_{k-2}|+1} & \text{otherwise.} \end{cases} \tag{40}$$

Furthermore, we define $\epsilon_{ij} = -\frac{\sum_{(i,t)\in\mathcal{E}^-_{k-1,k}}\epsilon_{it}}{|\mathcal{B}_i\setminus\mathcal{L}_{k-1}|+1}$ for $(i,j) \in \mathcal{E}_{k,k}$ with $k \geq 2$.

Based on ϵ_{ij} for all $(i,j) \in \mathcal{E}$, we define $\omega_{ij} = \eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} + \epsilon_{ij}$ for all $(i,j) \in \mathcal{E}$. Then we will prove such defined ω_{ij} satisfies the conditions in (38a)—(38c).

Condition in (38a):

For $(i,j) \in \mathcal{E}^+_{k-1,k}$ with $k = 2$, we have $i = 1$. Due $\eta_{1j}, \zeta_1 > 0$ and $\epsilon_{1j} = 0$, we can obtain $\omega_{1j} = \eta_{1j}c_{1j}^2 + \frac{\zeta_1}{|\mathcal{B}_1|+1} + \epsilon_{1j} = \eta_{1j}c_{1j}^2 + \frac{\zeta_1}{|\mathcal{B}_1|+1} > 0$. For $(i,j) \in \mathcal{E}^+_{k-1,k}$ with $k > 2$, we have $\omega_{ij} = \eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} + \epsilon_{ij} = \eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}^-_{k-2,k-1}}\epsilon_{it}}{|\mathcal{B}_i\setminus\mathcal{L}_{k-2}|+1}$. Due to $\epsilon_{it} < 0$ for $(i,t) \in \mathcal{E}^-_{k-2,k-1}$, and $\eta_{ij} > 0$, $\zeta_i \geq 0$, we have $\omega_{ij} > 0$.

For $(i,j) \in \mathcal{E}^-_{k-1,k}$ with $k = 2$, we have $j = 1$. We can obtain $\omega_{i1} = \eta_{i1}c_{i1}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} + \epsilon_{i1} = \eta_{i1}c_{i1}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\eta_{i1}c_{i1}^2\zeta_1/(|\mathcal{B}_1|+1)}{\eta_{i1}c_{1i}^2+\zeta_1/(|\mathcal{B}_1|+1)} = \frac{\zeta_i}{|\mathcal{B}_i|+1} + \frac{\eta_{i1}^2c_{i1}^2c_{1i}^2}{\eta_{i1}c_{1i}^2+\zeta_1/(|\mathcal{B}_1|+1)}$. Due to $\zeta_i \geq 0$ and $\eta_{i1} > 0$, we have $\omega_{i1} > 0$. For $(i,j) \in \mathcal{E}^-_{k-1,k}$ with $k > 2$, we have

$$\begin{aligned} \omega_{ij} &= \eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} + \epsilon_{ij} \\ &= \eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} + \frac{\eta_{ij}c_{ij}^2\sum_{(j,t)\in\mathcal{E}^-_{k-2,k-1}}\epsilon_{jt}/(|\mathcal{B}_j\setminus\mathcal{L}_{k-2}|+1)}{\eta_{ij}c_{ji}^2-\sum_{(j,t)\in\mathcal{E}^-_{k-2,k-1}}\epsilon_{jt}/(|\mathcal{B}_j\setminus\mathcal{L}_{k-2}|+1)} \\ &= \frac{\zeta_i}{|\mathcal{B}_i|+1} + \frac{\eta_{ij}^2c_{ij}^2c_{ji}^2}{\eta_{ij}c_{ji}^2-\sum_{(j,t)\in\mathcal{E}^-_{k-2,k-1}}\epsilon_{jt}/(|\mathcal{B}_j\setminus\mathcal{L}_{k-2}|+1)} > 0, \end{aligned} \tag{41}$$

due to $\zeta_i \geq 0$, $\eta_{ij} > 0$, and $\epsilon_{jt} < 0$ with $(j,t) \in \mathcal{E}^-_{k-2,k-1}$.

Furthermore, for $(i,j) \in \mathcal{E}_{k,k}$ with $k \geq 2$, we have $\omega_{ij} = \eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} + \epsilon_{ij} = \eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}^-_{k-1,k}} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-1}|+1}$. Due to $\eta_{ij} > 0$, $\zeta_i \geq 0$ and $\epsilon_{it} < 0$ with $(i,t) \in \mathcal{E}^-_{k-1,k}$, we can obtain $\omega_{ij} > 0$.

<u>Condition in (38b):</u>

First notice that $(i, j > i) \in \mathcal{E}$ is equivalent to $(i,j) \in \bigcup_{k=2}^{m} \mathcal{E}^+_{k-1,k}$ or $(i, j > i) \in \bigcup_{k=2}^{m} \mathcal{E}_{k,k}$. Therefore, we only need to examine (38b) for the sets $(i,j) \in \mathcal{E}^+_{k-1,k}$ and $(i, j > i) \in \mathcal{E}_{k,k}$ for $k \geq 2$. For $(i,j) \in \mathcal{E}^+_{k-1,k}$ with $k = 2$, since $\eta_{1j}, \zeta_1 > 0$ and $\zeta_j \geq 0$, we can obtain $\omega_{1j}\omega_{j1} - \eta_{1j}^2 c_{1j}^2 c_{j1}^2 = (\eta_{1j}c_{1j}^2 + \frac{\zeta_1}{|\mathcal{B}_1|+1})(\frac{\zeta_j}{|\mathcal{B}_j|+1} + \frac{\eta_{j1}^2 c_{j1}^2 c_{1j}^2}{\eta_{j1}c_{1j}^2 + \zeta_1/(|\mathcal{B}_1|+1)}) - \eta_{1j}^2 c_{1j}^2 c_{j1}^2 = (\eta_{1j}c_{1j}^2 + \frac{\zeta_1}{|\mathcal{B}_1|+1})\frac{\zeta_j}{|\mathcal{B}_j|+1} \geq 0$. For $(i,j) \in \mathcal{E}^+_{k-1,k}$ with $k > 2$, we have

$$
\begin{aligned}
\omega_{ij}\omega_{ji} - \eta_{ij}^2 c_{ij}^2 c_{ji}^2 = &\left(\eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}^-_{k-2,k-1}} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-2}|+1}\right) \\
&\times \left(\frac{\zeta_j}{|\mathcal{B}_j|+1} + \frac{\eta_{ij}^2 c_{ij}^2 c_{ji}^2}{\eta_{ij}c_{ij}^2 - \sum_{(i,t)\in\mathcal{E}^-_{k-2,k-1}} \epsilon_{it}/(|\mathcal{B}_i \setminus \mathcal{L}_{k-2}|+1)}\right) - \eta_{ij}^2 c_{ij}^2 c_{ji}^2 \\
= &\left(\eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}^-_{k-2,k-1}} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-2}|+1}\right)\frac{\zeta_j}{|\mathcal{B}_j|+1} \\
&+ \frac{\zeta_i}{|\mathcal{B}_i|+1}\frac{\eta_{ij}^2 c_{ij}^2 c_{ji}^2}{\eta_{ij}c_{ij}^2 - \sum_{(i,t)\in\mathcal{E}^-_{k-2,k-1}} \epsilon_{it}/(|\mathcal{B}_i \setminus \mathcal{L}_{k-2}|+1)} \geq 0,
\end{aligned} \tag{42}
$$

due to $\eta_{ij} > 0$, $\zeta_i, \zeta_j \geq 0$, and $\epsilon_{it} < 0$ with $(i,t) \in \mathcal{E}^-_{k-2,k-1}$.

Furthermore, for $(i, j > i) \in \mathcal{E}_{k,k}$, we have

$$
\begin{aligned}
\omega_{ij}\omega_{ji} - \eta_{ij}^2 c_{ij}^2 c_{ji}^2 = &\left(\eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}^-_{k-1,k}} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-1}|+1}\right)\left(\eta_{ij}c_{ji}^2 + \frac{\zeta_j}{|\mathcal{B}_j|+1} - \frac{\sum_{(j,t)\in\mathcal{E}^-_{k-1,k}} \epsilon_{jt}}{|\mathcal{B}_j \setminus \mathcal{L}_{k-1}|+1}\right) \\
&- \eta_{ij}^2 c_{ij}^2 c_{ji}^2 \\
= &\left(\frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}^-_{k-1,k}} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-1}|+1}\right)\eta_{ij}c_{ji}^2 \\
&+ \left(\eta_{ij}c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}^-_{k-1,k}} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-1}|+1}\right)\left(\frac{\zeta_j}{|\mathcal{B}_j|+1} - \frac{\sum_{(j,t)\in\mathcal{E}^-_{k-1,k}} \epsilon_{jt}}{|\mathcal{B}_j \setminus \mathcal{L}_{k-1}|+1}\right).
\end{aligned} \tag{43}
$$

Due to $\eta_{ij} > 0$, $\zeta_i, \zeta_j \geq 0$ and $\epsilon_{it}, \epsilon_{jt} < 0$ with $(i,t), (j,t) \in \mathcal{E}^-_{k-1,k}$, we can obtain $\omega_{ij}\omega_{ji} - \eta_{ij}^2 c_{ij}^2 c_{ji}^2 > 0$.

<u>Condition in (38c):</u>

For $i=1$, we have $\zeta_1 + \sum_{j\in\mathcal{B}_1} \eta_{1j} c_{1j}^2 - \sum_{j\in\mathcal{B}_1} \omega_{1j} = \zeta_1 + \sum_{j\in\mathcal{B}_1} \eta_{1j} c_{1j}^2 - \sum_{j\in\mathcal{B}_1} (\eta_{1j} c_{1j}^2 + \frac{\zeta_1}{|\mathcal{B}_1|+1}) = \frac{\zeta_1}{|\mathcal{B}_1|+1} > 0$. For $i \in \mathcal{L}_k$, we have

$$\begin{aligned}
\zeta_i + \sum_{j\in\mathcal{B}_i} \eta_{ij} c_{ij}^2 - \sum_{j\in\mathcal{B}_i} \omega_{ij} =& \zeta_i + \sum_{j\in\mathcal{B}_i} \eta_{ij} c_{ij}^2 - \sum_{j\in\mathcal{B}_i\cap\mathcal{L}_{k-1}} \omega_{ij} - \sum_{j\in\mathcal{B}_i\setminus\mathcal{L}_{k-1}} \omega_{ij} \\
=& \zeta_i + \sum_{j\in\mathcal{B}_i} \eta_{ij} c_{ij}^2 - \sum_{j\in\mathcal{B}_i\cap\mathcal{L}_{k-1}} (\eta_{ij} c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} + \epsilon_{ij}) \\
& - \sum_{j\in\mathcal{B}_i\setminus\mathcal{L}_{k-1}} (\eta_{ij} c_{ij}^2 + \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}_{k-1,k}^-} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-1}|+1}) \\
=& \frac{\zeta_i}{|\mathcal{B}_i|+1} - \frac{\sum_{(i,t)\in\mathcal{E}_{k-1,k}^-} \epsilon_{it}}{|\mathcal{B}_i \setminus \mathcal{L}_{k-1}|+1}. \qquad (44)
\end{aligned}$$

Due to $\zeta_i \geq 0$, $\epsilon_{it} < 0$ with $(i,t) \in \mathcal{E}_{k-1,k}^-$, and $\sum_{j\in\mathcal{B}_i\cap\mathcal{L}_{k-1}} \epsilon_{ij} = \sum_{(i,t)\in\mathcal{E}_{k-1,k}^-} \epsilon_{it}$, we obtain $\zeta_i + \sum_{j\in\mathcal{B}_i} \eta_{ij} c_{ij}^2 - \sum_{j\in\mathcal{B}_i} \omega_{ij} > 0$.

Hence, we have found ω_{ij} satisfying (38a)—(38c). Therefore, pairwise factorization $\mathbb{F}_4$ with at least one $\zeta_i > 0$ is pairwise-normalizable. By Corollary 7, Gaussian BP converges in both synchronous and totally asynchronous schedulings.

References

D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and distributed computation: numerical methods*, volume 23. Prentice Hall, 1989.

D. Bickson, D. Malkhi, and L. Zhou. Peer-to-peer rating. In *IEEE International Conference on Peer-to-Peer Computing*, pages 211–218. IEEE, 2007.

D. Bickson, D. Dolev, O. Shental, P. H. Siegel, and J. K. Wolf. Gaussian belief propagation based multiuser detection. In *IEEE International Symposium on Information Theory*, pages 1878–1882. IEEE, 2008.

J. Du and Y. C. Wu. Network-wide distributed carrier frequency offsets estimation and compensation via belief propagation. *IEEE Transactions on Signal Processing*, 61(23): 5868–5877, 2013.

D. Koller and N. Friedman. *Probabilistic graphical models: principles and techniques.* MIT Press, 2009.

F. R. Kschischang, B. J. Frey, and H. A. Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on Information Theory*, 47(2):498–519, 2001.

M. Leng and Y. C. Wu. Distributed clock synchronization for wireless sensor networks using belief propagation. *IEEE Transactions on Signal Processing*, 59(11):5404–5414, 2011.

D. M. Malioutov, J. K. Johnson, and A. S. Willsky. Walk-sums and belief propagation in gaussian graphical models. *Journal of Machine Learning Research*, 7(1):2031–2064, 2006.

C. C. Moallemi and B. V. Roy. Consensus propagation. *IEEE Transactions on Information Theory*, 52(11):4753–4766, 2006.

C. C. Moallemi and B. V. Roy. Convergence of min-sum message passing for quadratic optimization. *IEEE Transactions on Information Theory*, 55(5):2413–2423, 2009.

J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference.* Morgan Kaufmann, 1988.

H. Rue and L. Held. *Gaussian Markov random fields: theory and applications.* CRC Press, 2005.

N. Ruozzi and S. Tatikonda. Message-passing algorithms for quadratic minimization. *Journal of Machine Learning Research*, 14(1):2287–2314, 2013.

O. Shental, P. H. Siegel, J. K. Wolf, D. Bickson, and D. Dolev. Gaussian belief propagation solver for systems of linear equations. In *IEEE International Symposium on Information Theory*, pages 1863–1867. IEEE, 2008.

Q. Su and Y. C. Wu. Convergence analysis of the variance in Gaussian belief propagation. *IEEE Transactions on Signal Processing*, 62(19):5119–5131, 2014.

Q. Su and Y. C. Wu. On convergence conditions of Gaussian belief propagation. *IEEE Transactions Signal Processing*, 63(5):1144–1155, 2015a.

Q. Su and Y. C. Wu. Distributed estimation of variance in Gaussian graphical model via belief propagation: Accuracy analysis and improvement. *IEEE Transactions on Signal Processing*, 63(23):6258–6271, 2015b.

L. Vandenberghe and S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

Y. Weiss and W. T. Freeman. Correctness of belief propagation in Gaussian graphical models of arbitrary topology. In *Advances in Neural Information Processing Systems*, pages 673–679, 2000.

Z. Wen, D. Goldfarb, and W. Yin. Alternating direction augmented lagrangian methods for semidefinite programming. *Mathematical Programming Computation*, 2(3-4):203–230, 2010.