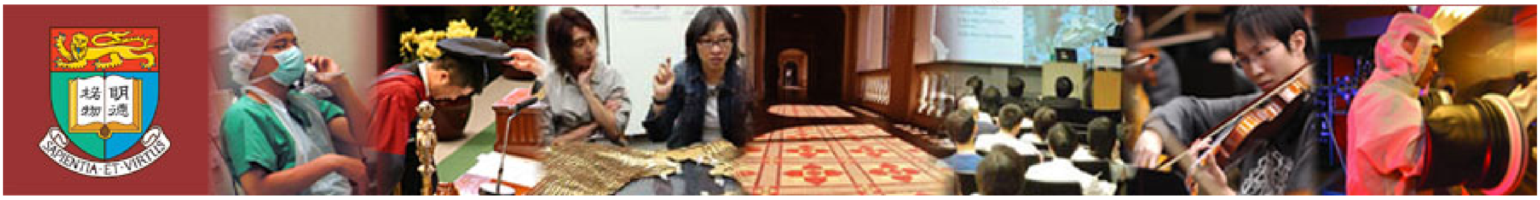

Title	Determining the convergence of variance in Gaussian belief propagation via semi-definite programming
Author(s)	Su, Q; Wu, YC
Citation	The 2014 IEEE International Symposium on Information Theory (ISIT), Honolulu, HI., 29 June-4 July 2014. In IEEE International Symposium on Information Theory Proceedings, 2014, p. 2614-2618
Issued Date	2014
URL	http://hdl.handle.net/10722/199396
Rights	IEEE International Symposium on Information Theory Proceedings. Copyright © IEEE.

Determining the Convergence of Variance in Gaussian Belief Propagation via Semi-definite Programming

Qinliang Su and Yik-Chung Wu
Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong
E-mail: {qlsu, ycwu}@eee.hku.hk

***Abstract*—In order to compute the marginal distribution from a high dimensional distribution with loopy Gaussian belief propagation (BP), it is important to determine whether Gaussian BP would converge. In general, the convergence condition for Gaussian BP variance and mean are not necessarily the same, and this paper focuses on the convergence condition of Gaussian BP variance. In particular, by describing the message-passing process of Gaussian BP as a set of updating functions, the necessary and sufficient convergence condition of Gaussian BP variance is derived, with the converged variance proved to be independent of the initialization as long as it is greater or equal to zero. It is further proved that the convergence condition can be verified efficiently by solving a semi-definite programming (SDP) optimization problem. Numerical examples are presented to corroborate the established theories.**

I. Introduction

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

By representing the joint PDF with a factor graph , Gaussian BP provides an alternative to calculate the marginal mean and variance for each individual random variable by passing messages between neighboring nodes in the factor graph [1]. It is known that for a factor graph with loops, if the marginal mean and variance in Gaussian BP converge, the true marginal mean and an approximate marginal variance are obtained [2].

With the ability to provide the true marginal mean upon convergence, Gaussian BP has been successfully applied in communication systems [3], fast solver for large sparse linear systems [4], etc. In addition, the distributed property inherited from message passing algorithms is particularly attractive to the applications requiring distributed implementation, such as distributed beamforming [5], synchronization in wireless sensor networks [6], [7], etc.

However, Gaussian BP only works under the prerequisite that the belief mean and variance calculated from the updating messages do converge. So far, several sufficient convergence conditions have been proposed, which can guarantee the mean and variance converge simultaneously [2], [8], [9]. However, in general, the belief mean and variance do not necessarily converge under the same condition. It is reported in [8] that if the variance converges, the convergence of mean can always be observed when suitable damping is imposed. Thus, it is important to ensure the variance of Gaussian BP to converge in the first place. In the pioneering work [8], based on the concept of computation tree [2], a general convergence condition of variance is derived. However, this convergence condition requires the evaluation of the spectral radius of an infinite dimensional matrix, which is almost impossible to be calculated in practice.

In this paper, we first describe the message passing of Gaussian BP as a set of updating functions, and analyze the properties of updating functions. Then, based on the relation between BP messages and variance, the necessary and sufficient convergence condition of variance are developed, with the converged variance is proved to be independent of the initialization as long as it is greater or equal to zero. Furthermore, it is proved that the convergence condition derived in this paper can be efficiently verified by solving a semi-definite programming (SDP) problem.

II. Gaussian Belief Propagation

A Gaussian PDF can be written as $f(\mathbf{x}) \propto \exp\left\{-\frac{1}{2}\mathbf{x}^T\mathbf{P}\mathbf{x} + \mathbf{h}^T\mathbf{x}\right\}$, where $\mathbf{x} = [x_1, x_2, \cdots, x_N]^T$; $\mathbf{P} \succ 0$ is the precision matrix with p_{ij} being its (i,j)-th element; and $\mathbf{h} = [h_1, h_2, \cdots, h_N]^T$. We expand $f(\mathbf{x})$ as $f(\mathbf{x}) \propto \prod_{i=1}^{N} f_i(x_i) \prod_{j=1}^{N} \prod_{k=j+1}^{N} f_{jk}(x_j, x_k)$, where $f_i(x_i) = \exp\left\{-\frac{p_{ii}}{2}x_i^2 + h_i x_i\right\}$ and $f_{jk}(x_j, x_k) = \exp\left\{-p_{jk}x_j x_k\right\}$. Based on this expansion, a factor graph $\mathbb{G}(\mathcal{V}, \mathcal{E})$ can be constructed by connecting factors $f_i(x_i)$ and $f_{jk}(x_j, x_k)$ with their associated variables, where $\mathcal{V} = \{1, 2, \cdots, N\}$ is the set of indices of variable nodes; and $\mathcal{E} = \{(i,j) | p_{ij} \neq 0, \text{ for } i,j \in \mathcal{V}\}$ is the set of index pairs of any two connected nodes.

In Gaussian BP, the departing and arriving messages of any two neighboring variable nodes i and j are updated as

$$m^d_{i\to j}(x_i, t) \propto \prod_{k\in\mathcal{N}(i)\setminus j} m^a_{k\to i}(x_i, t) f_i(x_i), \tag{1}$$

$$m^a_{i\to j}(x_j, t+1) \propto \int m^d_{i\to j}(x_i, t)\, f_{ij}(x_i, x_j)\, dx_i, \tag{2}$$

where t is the time index; $\mathcal{N}(i)$ is the set of indices of neighboring variable nodes of node i, and $\mathcal{N}(i)\backslash j$ is the set $\mathcal{N}(i)$ but excluding the node j. After obtaining the messages $m^a_{k\to i}(x_i,t)$, the belief at variable node i is equal to

$$b_i\left(x_i,t\right) \propto \prod_{k\in\mathcal{N}(i)} m^a_{k\to i}\left(x_i,t\right) f_i(x_i). \tag{3}$$

Assume the arriving message is in Gaussian form of $m^a_{i\to j}(x_j,t) \propto \exp\{-\frac{v^a_{i\to j}(t)}{2}x_j^2 + \beta^a_{i\to j}(t)x_j\}$, where $v^a_{i\to j}(t)$ and $\beta^a_{i\to j}(t)$ are the arriving precision and arriving linear coefficient, respectively. Inserting it into (1), we obtain $m^d_{i\to j}(x_i,t) \propto \exp\{-\frac{v^d_{i\to j}(t)}{2}x_i^2 + \beta^d_{i\to j}(t)x_i\}$, where

$$v^d_{i\to j}(t) = p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} v^a_{k\to i}(t), \tag{4}$$

$$\beta^d_{i\to j}(t) = h_i + \sum_{k\in\mathcal{N}(i)\backslash j} \beta^a_{k\to i}(t) \tag{5}$$

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

$$\begin{aligned} & m^a_{i\to j}(x_j,t+1) \\ & \propto \exp\left\{\frac{p_{ij}^2}{2v^d_{i\to j}(t)}x_j^2 - \frac{p_{ij}\beta^d_{i\to j}(t)}{v^d_{i\to j}(t)}x_j\right\} \\ & \times \int \exp\left\{-\frac{v^d_{i\to j}(t)}{2}\left(x_i - \frac{\beta^d_{i\to j}(t) - p_{ij}x_j}{v^d_{i\to j}(t)}\right)^2\right\} dx_i. \end{aligned} \tag{6}$$

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

$$v^a_{i\to j}(t+1) = -\frac{p_{ij}^2}{v^d_{i\to j}(t)}, \tag{7}$$

$$\beta^a_{i\to j}(t+1) = -\frac{p_{ij}\beta^d_{i\to j}(t)}{v^d_{i\to j}(t)}. \tag{8}$$

After obtaining $v^a_{k\to i}(t)$, the variance of belief at each iteration is computed as

$$\sigma_i^2(t) = \frac{1}{p_{ii} + \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)}. \tag{9}$$

However, from (6), it should be noticed that if $v^d_{i\to j}(t) \le 0$, the integration in (6) as well as the resulting message $m^a_{i\to j}(t+1)$ become infinite. Therefore, under the assumption of Gaussian initialization, we have the following lemma.

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

III. Analysis of the Message-Passing Process

Substituting (4) into (7) gives

$$v^a_{i\to j}(t+1) = -\frac{p_{ij}^2}{p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} v^a_{k\to i}(t)}. \tag{10}$$

By writing (10) into a vector form, we obtain

$$\mathbf{v}^a(t+1) = \mathbf{g}(\mathbf{v}^a(t)), \tag{11}$$

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

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

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

$$\mathcal{W} \triangleq \left\{\mathbf{w}\middle| p_{ii} + \sum\nolimits_{k\in\mathcal{N}(i)\backslash j} w_{ki} > 0,\ \forall\,(i,j)\in\mathcal{E}\right\}. \tag{13}$$

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

Proposition 1 *The following claims hold:*

P1) $\mathbf{w}_1 \in \mathcal{W}$ *and* $\mathbf{w}_2 \ge \mathbf{w}_1$ *implies* $\mathbf{w}_2 \in \mathcal{W}$*;*

P2) $\mathbf{w}_1 \in \mathcal{W}$ *and* $\mathbf{w}_2 \ge \mathbf{w}_1$ *implies* $\mathbf{g}(\mathbf{w}_2) \ge \mathbf{g}(\mathbf{w}_1)$.

Proof: Consider two vectors $\bar{\mathbf{w}}$ and $\hat{\mathbf{w}}$, which contain elements $\bar{w}_{ki}$ and $\hat{w}_{ki}$ with $(k,i)\in\mathcal{E}$ arranged in ascending order first on j and then on i. For any $\bar{\mathbf{w}}\in\mathcal{W}$, according to the definition of $\mathcal{W}$ in (13), we have $p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j}\bar{w}_{ki} > 0$. Then, if $\bar{\mathbf{w}} \le \hat{\mathbf{w}}$, it can be easily seen that $p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j}\hat{w}_{ki} > 0$ as well. Thus, we have $\hat{\mathbf{w}}\in\mathcal{W}$.

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

$$\frac{\partial g_{ij}}{\partial w_{ki}} = \frac{p_{ij}^2}{\left(p_{ii} + \sum_{k\in\mathcal{N}(i)\backslash j} w_{ki}\right)^2} > 0. \tag{14}$$

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

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

A. Convergence Condition

To derive the convergence condition of variance $\sigma_i^2(t)$, we first define the following set

$$\mathcal{S}_1 \triangleq \left\{\mathbf{w}\,\middle|\mathbf{w} \le \mathbf{g}(\mathbf{w}) \text{ and } \mathbf{w}\in\mathcal{W}\right\}. \tag{15}$$

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

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

P3) If $\mathbf{s}\in\mathcal{S}_1$*, then* $\mathbf{s} < \mathbf{0}$*;*

P4) If $\mathbf{s}\in\mathcal{S}_1$*, then* $\mathbf{g}^{(t)}(\mathbf{s})\in\mathcal{S}_1$ *and* $\mathbf{g}^{(t)}(\mathbf{s}) \le \mathbf{g}^{(t+1)}(\mathbf{s})$ *for all* $t \ge 0$.

Proof: If $\mathbf{s} \in \mathcal{S}_1$, we have $\mathbf{s} \leq \mathbf{g}(\mathbf{s})$ and $\mathbf{s} \in \mathcal{W}$. According to the definition of $\mathcal{W}$ in (13), $p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} s_{ki} > 0$. Putting this fact into the definition of $g_{ij}(\cdot)$ in (12), it is obvious that $\mathbf{g}(\mathbf{s}) < \mathbf{0}$. Therefore, we have the relation that $\mathbf{s} \leq \mathbf{g}(\mathbf{s}) < \mathbf{0}$ for all $\mathbf{s} \in \mathcal{S}_1$.

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

Now, we have the following lemma.

Lemma 2 *If $\mathcal{S}_1 \neq \emptyset$, $\mathbf{v}^a(t)$ converges to the same point $\lim_{t\to\infty} \mathbf{v}^a(t)$ for all $\mathbf{v}^a(0) \geq \mathbf{0}$.*

Proof: First, we prove that $\mathbf{v}^a(t)$ converges for any $\mathbf{v}^a(0) \geq \mathbf{0}$ given $\mathcal{S}_1 \neq \emptyset$. Due to $\mathcal{S}_1 \neq \emptyset$, for any $\mathbf{w} \in \mathcal{S}_1$, it can be obtained from P3) that $\mathbf{w} < \mathbf{0}$. Thus, for any $\mathbf{v}^a(0) \geq \mathbf{0}$, the relation $\mathbf{w} \leq \mathbf{v}^a(0)$ always holds. Notice that $\mathbf{w} \in \mathcal{W}$ due to $\mathbf{w} \in \mathcal{S}_1$ and $\mathcal{S}_1 \subseteq \mathcal{W}$. Applying P2) to $\mathbf{w} \leq \mathbf{v}^a(0)$, we obtain $\mathbf{g}(\mathbf{w}) \leq \mathbf{v}^a(1)$. Combining it with $\mathbf{w} \leq \mathbf{g}(\mathbf{w})$ from P4) gives $\mathbf{w} \leq \mathbf{v}^a(1)$. On the other hand, substituting $\mathbf{v}^a(0) \geq \mathbf{0}$ into (12) gives

$$\mathbf{v}^a(1) < \mathbf{0}. \tag{16}$$

Due to $\mathbf{v}^a(0) \geq \mathbf{0}$, thus $\mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Combining it with $\mathbf{w} \leq \mathbf{v}^a(1)$ gives $\mathbf{w} \leq \mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Applying $\mathbf{g}(\cdot)$ to $\mathbf{w} \leq \mathbf{v}^a(1) < \mathbf{v}^a(0)$, it can be inferred from P2) that $\mathbf{g}(\mathbf{w}) \leq \mathbf{v}^a(2) < \mathbf{v}^a(1)$. Together with $\mathbf{w} \leq \mathbf{g}(\mathbf{w})$ as claimed by P4), we obtain $\mathbf{w} \leq \mathbf{v}^a(2) < \mathbf{v}^a(1)$. By induction, we can infer that

$$\mathbf{w} \leq \mathbf{v}^a(t+1) \leq \mathbf{v}^a(t). \tag{17}$$

It can be seen from (17) that $\mathbf{v}^a(t)$ is a monotonically decreasing but lower bounded sequence, thus it converges.

Next, we prove that $\mathbf{v}^a(t)$ converges to the same point for all $\mathbf{v}^a(0) \geq \mathbf{0}$. For any $\mathbf{v}^a(0) \geq \mathbf{0}$, according to P2), applying $\mathbf{g}(\cdot)$ on both sides of $\mathbf{0} \leq \mathbf{v}^a(0)$ gives $\mathbf{g}(\mathbf{0}) \leq \mathbf{g}(\mathbf{v}^a(0)) = \mathbf{v}^a(1)$. Combining this relation with (16) leads to $\mathbf{g}(\mathbf{0}) \leq \mathbf{v}^a(1) \leq \mathbf{0}$. Applying $\mathbf{g}(\cdot)$ on this inequality for t more times, it can be obtained from P2) that $\mathbf{g}^{(t+1)}(\mathbf{0}) \leq \mathbf{v}^a(t+1) \leq \mathbf{g}^{(t)}(\mathbf{0})$ for all $t \geq 0$. By denoting $\lim_{t\to\infty} \mathbf{g}^{(t)}(\mathbf{0}) = \mathbf{v}^{a*}$, it is obvious that $\mathbf{v}^a(t)$ also converges to the $\mathbf{v}^{a*}$. ■

Now, we present the following theorem.

Theorem 1 *For any $\mathbf{v}^a(0) \geq \mathbf{0}$, $\sigma_i^2(t)$ converges to the same value if and only if $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t) \neq 0$.*

Proof:

<u>Sufficient Condition:</u>

If $\mathcal{S}_1 \neq \emptyset$, the limit $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)$ always exists and is equal to $p_{ii} + \sum_{k\in\mathcal{N}(i)} v^{a*}_{k\to i}$. From $\sigma_i^2(t) = \frac{1}{p_{ii}+\sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)}$ in (9) as well as $p_{ii} + \sum_{k\in\mathcal{N}(i)} v^{a*}_{k\to i} \neq 0$, it can be inferred that $\sigma_i^2(t)$ converges to the same $\frac{1}{p_{ii}+\sum_{k\in\mathcal{N}(i)} v^{a*}_{k\to i}}$.

<u>Necessary Condition:</u>

We will prove the necessary part by contradiction. Under the condition that $\sigma_i^2(t)$ converges for any $\mathbf{v}^a(0) \geq \mathbf{0}$, to prove the theorem, we only need to prove the following two cases can never happen: 1) $\mathcal{S}_1 = \emptyset$; 2) $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t) = 0$.

First, suppose $\mathcal{S}_1 = \emptyset$. Since $\sigma_i^2(t)$ converges, then the messages should be maintained at Gaussian form for all $t \geq 0$. According to Lemma 1, it can be inferred that $\bar{v}^d_{i\to j}(t) = p_{ii} + \sum_{k\to\mathcal{N}(i)\setminus j} \bar{v}^a_{k\to i}(t) > 0$ for all $(i,j) \in \mathcal{E}$. From $\mathcal{W}$ in (13), we obtain

$$\bar{\mathbf{v}}^a(t) \in \mathcal{W}, \tag{18}$$

where $\bar{\mathbf{v}}^a(t) \triangleq \mathbf{g}^{(t)}(\bar{\mathbf{v}}^a(0))$; $\bar{v}^a_{k\to i}(t)$ are the elements of $\bar{\mathbf{v}}^a(t)$ with $(k,i) \in \mathcal{E}$ arranged in the same order as $\mathbf{v}^a(t)$.

Now, choose another initialization $\mathbf{v}^a(0)$ satisfying both $\bar{\mathbf{v}}^a(0) \leq \mathbf{v}^a(0)$ and $\mathbf{v}^a(0) \geq \mathbf{0}$. Due to $\bar{\mathbf{v}}^a(0) \in \mathcal{W}$ and $\bar{\mathbf{v}}^a(0) \leq \mathbf{v}^a(0)$, by using the P2), we have $\mathbf{g}(\bar{\mathbf{v}}^a(0)) \leq \mathbf{g}(\mathbf{v}^a(0))$, that is, $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1)$. Furthermore, substituting $\mathbf{v}^a(0) \geq \mathbf{0}$ into (12) gives $\mathbf{v}^a(1) \leq \mathbf{0}$, and thereby $\mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Combining $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1)$ and $\mathbf{v}^a(1) \leq \mathbf{v}^a(0)$ leads to $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1) \leq \mathbf{v}^a(0)$. Due to the assumption $\bar{\mathbf{v}}^a(1) \in \mathcal{W}$, by applying P2) to $\bar{\mathbf{v}}^a(1) \leq \mathbf{v}^a(1) \leq \mathbf{v}^a(0)$, we obtain $\bar{\mathbf{v}}^a(2) \leq \mathbf{v}^a(2) \leq \mathbf{v}^a(1)$. Combining with the fact $\mathbf{v}^a(1) \leq \mathbf{0}$ as proved above, we have $\bar{\mathbf{v}}^a(2) \leq \mathbf{v}^a(2) \leq \mathbf{v}^a(1) \leq \mathbf{0}$. By induction, it can be derived that

$$\bar{\mathbf{v}}^a(t+1) \leq \mathbf{v}^a(t+1) \leq \mathbf{v}^a(t) \leq \mathbf{0} \tag{19}$$

for all $t \geq 1$.

Then, from the definition of $\mathcal{W}$ in (13) and $\bar{\mathbf{v}}^a(t) \in \mathcal{W}$ in (18), we have $p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} \bar{v}^a_{k\to i}(t) > 0$ for all $(i,j) \in \mathcal{E}$. By rearranging the terms in $p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} \bar{v}^a_{k\to i}(t) > 0$, we obtain $\bar{v}^a_{\gamma\to i}(t) > -p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j,\gamma} \bar{v}^a_{k\to i}(t)$. Applying $\bar{\mathbf{v}}^a(t) \leq \mathbf{0}$ shown in (19) into this inequality, it can be inferred that $\bar{v}^a_{\gamma\to i}(t) > -p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j,\gamma} \bar{v}^a_{k\to i}(t) \geq -p_{ii}$, that is, $\bar{v}^a_{\gamma\to i}(t) > -p_{ii}$ for all $(\gamma, i) \in \mathcal{E}$. Combining with $\mathbf{v}^a(t) \geq \bar{\mathbf{v}}^a(t)$ shown in (19), for all $(i,j) \in \mathcal{E}$, we have

$$v^a_{i\to j}(t) > -p_{jj}. \tag{20}$$

It can be seen from (19) and (20) that $\mathbf{v}^a(t)$ is a monotonically decreasing and lower bounded sequence. Thus, $\mathbf{v}^a(t)$ must converge to a vector $\mathbf{v}^{a*}$, that is, $\mathbf{v}^{a*} = \mathbf{g}(\mathbf{v}^{a*})$.

On the other hand, substituting (10) into (20) gives $-\frac{p_{ij}^2}{p_{ii}+\sum_{k\in\mathcal{N}(i)\setminus j} v^a_{k\to i}(t)} > -p_{jj}$. Due to $\bar{\mathbf{v}}^a(t) \in \mathcal{W}$ in (18), it can be inferred from P1) and (19) that $\mathbf{v}^a(t) \in \mathcal{W}$, or equivalently $p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} v^a_{k\to i}(t) > 0$. Together with the fact $p_{jj} > 0$, we obtain $p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} v^a_{k\to i}(t) > \frac{p_{ij}^2}{p_{jj}}$. Since $\mathbf{v}^a(t)$ converges to $\mathbf{v}^{a*}$, taking the limit on both sides of the inequality gives $p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} v^{a*}_{k\to i} \geq \frac{p_{ij}^2}{p_{jj}}$. Hence, $p_{ii} + \sum_{k\in\mathcal{N}(i)\setminus j} v^{a*}_{k\to i} > 0$. From the definition of $\mathcal{W}$ in (13), we have $\mathbf{v}^{a*} \in \mathcal{W}$. Combining with $\mathbf{v}^{a*} = \mathbf{g}(\mathbf{v}^{a*})$, according to the definition of $\mathcal{S}_1$ in (15), it is clear that $\mathbf{v}^{a*} \in \mathcal{S}_1$. This contradicts with the prerequisite $\mathcal{S}_1 = \emptyset$. Thus, we have $\mathcal{S}_1 \neq \emptyset$.

Finally, consider the second scenario. Suppose $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t) = 0$. Due to $\mathcal{S}_1 \neq \emptyset$, it is known from Lemma 2 that the limit $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v^a_{k\to i}(t)$

always exists and is equal to $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*}$. Due to $\sigma_i^2(t) = \frac{1}{p_{ii}+\sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t)}$, in order to ensure $\sigma_i^2(t)$ converge, we must have $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t) \neq 0$. ■

B. Verification by Semi-definite Programming

In this section, we will establish the connections between the proposed convergence condition and the SDP optimization problem [10], which can be solved efficiently by existing softwares, such as CVX. First, define the following SDP problem

$$\begin{aligned} \min_{\mathbf{w},\,\eta} \quad & \eta \\ s.t. \quad & \begin{bmatrix} p_{ii} + \sum\limits_{k\in\mathcal{N}(i)\setminus j} w_{ki} & p_{ij} \\ p_{ij} & \eta - w_{ij} \end{bmatrix} \succeq 0,\ \forall\,(i,j)\in\mathcal{E}. \end{aligned} \tag{21}$$

Now, we have the following theorem.

Theorem 2 *$\mathcal{S}_1 \neq \emptyset$ if and only if the optimal solution of* (21) *$\eta^* \leq 0$.*

Proof: First, notice that the SDP problem in (21) is equivalent to the following optimization problem

$$\begin{aligned} \min_{\mathbf{w},\,\eta} \quad & \eta \\ s.t. \quad & w_{ij} - g_{ij}(\mathbf{w}) \leq \eta, & \forall\,(i,j)\in\mathcal{E}; \\ & -p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki} \leq 0, & \forall\,(i,j)\in\mathcal{E}. \end{aligned} \tag{22}$$

If $\mathcal{S}_1 \neq \emptyset$, according to definition of $\mathcal{S}_1$ in (15), there must exist a $\mathbf{w}$ such that $w_{ij} - g_{ij}(\mathbf{w}) \leq 0$ and $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki} < 0$ for all $(i,j)\in\mathcal{E}$. Thus, $(\mathbf{w}, 0)$ satisfies the constraints of convex optimization problem (22). Since $\eta = 0$ is a feasible solution and (22) is a minimization problem, the optimal solution of (22) cannot be greater than 0, that is, $\eta^* \leq 0$.

Next, if $(\mathbf{w}^*, \eta^*)$ is the optimal solution of (22), we have

$$w_{ij}^* - g_{ij}(\mathbf{w}^*) \leq \eta^* \text{ and } -p_{ii} - \sum\nolimits_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* \leq 0. \tag{23}$$

Due to $\eta^* \leq 0$, we have $w_{ij}^* - g_{ij}(\mathbf{w}^*) \leq 0$ for all $(i,j)\in\mathcal{E}$. This is exactly the first constraint in $\mathcal{S}_1$. For the second constraint in (23), because $g_{ij}(\mathbf{w}) = -\frac{p_{ij}^2}{p_{ii}+\sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}}$ is undefined when $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* = 0$, the scenarios $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* = 0$ will never happen. Thus, the second constraint in (23) is equivalent to the second constraint of $\mathcal{S}_1$. Hence, we have $\mathbf{w}^* \in \mathcal{S}_1$ and $\mathcal{S}_1 \neq \emptyset$. ■

If $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t) \neq 0$ is known to hold, according Theorems 1 and 2, $\eta^* \leq 0$ can serve as the necessary and sufficient convergence condition of variance $\sigma_i^2(t)$. Although $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t) \neq 0$ happens almost surely, the condition can be guaranteed theoretically by using the following modified SDP problem:

$$\begin{aligned} \min_{\mathbf{w},\,\alpha} \quad & \alpha \\ s.t. \quad & \begin{bmatrix} p_{ii} + \sum\limits_{k\in\mathcal{N}(i)\setminus j} w_{ki} & p_{ij} \\ p_{ij} & -w_{ij} \end{bmatrix} \succeq 0,\ \forall\,(i,j)\in\mathcal{E}; \\ & \alpha + p_{ii} + \sum_{k\in\mathcal{N}(i)} w_{ki} \geq 0, \qquad \forall\, i\in\mathcal{V}. \end{aligned} \tag{24}$$

Now, the following theorem can be presented.

Theorem 3 *If the optimal solution of* (24) *$\alpha^* < 0$, then $\mathcal{S}_1 \neq \emptyset$ and $p_{ii} + \lim_{t\to\infty} \sum_{k\in\mathcal{N}(i)} v_{k\to i}^a(t) \neq 0$ for all $i \in \mathcal{V}$.*

Proof: First, notice that the SDP problem in (24) is equivalent to the following optimization problem

$$\begin{aligned} \min_{\mathbf{w},\,\alpha} \quad & \alpha \\ s.t. \quad & w_{ij} - g_{ij}(\mathbf{w}) \leq 0, & \forall\,(i,j)\in\mathcal{E}; \\ & -p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki} \leq 0, & \forall\,(i,j)\in\mathcal{E}; \\ & -p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki} \leq \alpha, & \forall\, i\in\mathcal{V}. \end{aligned} \tag{25}$$

If $(\mathbf{w}^*, \alpha^*)$ is the optimal solution of (25) with $\alpha^* < 0$, $(\mathbf{w}^*, \alpha^*)$ must satisfy the constraints in (25), thus the following three conditions hold: 1) $w_{ij}^* - g_{ij}(\mathbf{w}^*) \leq 0$; 2) $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* \leq 0$; 3) $-p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki}^* \leq \alpha^*$. For the second constraint, if $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* = 0$, the function $g_{ij}(\mathbf{w}^*) = -\frac{p_{ij}^2}{p_{ii}+\sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^*}$ becomes undefined, thus $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* = 0$ will never happen. Hence, we always have $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* < 0$. For the third constraint, due to $\alpha^* < 0$, it can be inferred that $-p_{ii} - \sum_{k\in\mathcal{N}(i)} w_{ki}^* < 0$, or equivalently

$$p_{ii} + \sum\nolimits_{k\in\mathcal{N}(i)} w_{ki}^* > 0. \tag{26}$$

Next, from the first two conditions $w_{ij}^* - g_{ij}(\mathbf{w}^*) \leq 0$ and $-p_{ii} - \sum_{k\in\mathcal{N}(i)\setminus j} w_{ki}^* < 0$, it can be obtained from the definition of $\mathcal{S}_1$ in (15) that $\mathbf{w}^* \in \mathcal{S}_1$. Due to $\mathbf{w}^* \in \mathcal{S}_1$, it can be inferred from (17) that $\mathbf{v}^{a*} \geq \mathbf{w}^*$. Combining with (26), it can be inferred that $p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*} > 0$, and thereby $p_{ii} + \sum_{k\in\mathcal{N}(i)} v_{k\to i}^{a*} \neq 0$. ■

Using the alternating direction method of multipliers (ADMM) technique [11], [12], the SDP problem in (24) can be further reformulated into a problem consisting of N low-dimensional sub-problems, and thus solved distributely. Not only this avoids the gathering of information at a central processing unit, the complexity is also reduced from $O\left(\left(\sum_{i=1}^N |\mathcal{N}(i)|\right)^4\right)$ in SDP to $O\left(\sum_{i=1}^N |\mathcal{N}(i)|^4\right)$ in ADMM. Since the derivation of ADMM is well-documented in [11], [12], we do not give the details here.

V. Numerical Examples

In this section, numerical experiments are presented to illustrate the theories in this paper. The example is based on

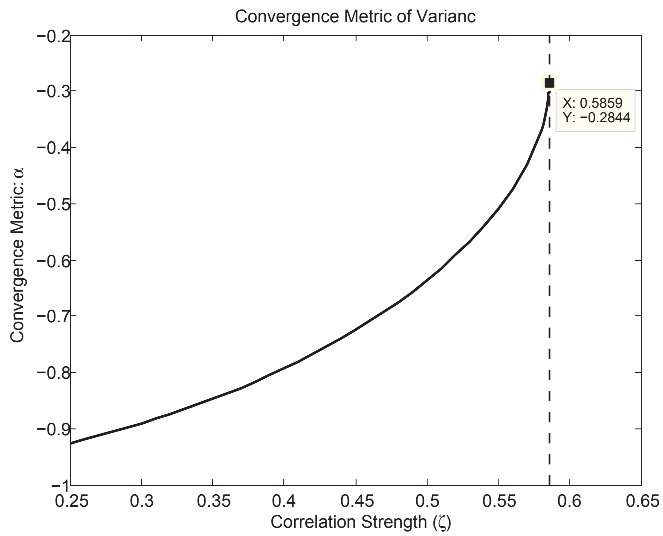

Fig. 1: The value of α^* under different correlation strength ζ.

the 20×20 precision matrices $\mathbf{P}$ constructed as

$$p_{ij} = \begin{cases} 1, & \text{if } i = j \\ \zeta \cdot \theta_{\text{mod}(i+j,10)+1}, & \text{if } i \neq j \end{cases}, \tag{27}$$

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

Fig. 1 illustrates how the optimal solution α^* of (24) varies with the correlation strength ζ. It can be seen that the optimal solution α^* always exists and the condition $\alpha^* < 0$ holds for all $\zeta \leq 0.5859$, while no feasible solution exists in the problem (24) when $\zeta > 0.5859$. According to Theorem 3, this means that if $\zeta \leq 0.5859$, the variance $\sigma_i^2(t)$ with $i \in \mathcal{V}$ converges to the same point for all initializations $\mathbf{v}^a(0) \geq \mathbf{0}$.

To verify the convergence of variance under $\zeta \leq 0.5859$, Fig. 2a shows how the variance $\sigma_1^2(t)$ of the 1-th variable evolves as a function of t when $\zeta = 0.5858$, which is slightly smaller than 0.5859. The convergence of variance $\sigma_1^2(t)$ can be observed under the initialization of $\mathbf{v}^a(0)$ being a vector with all elements equal to 10. On the other hand, Fig. 2b illustrates how the variance $\sigma_1^2(t)$ varies as iterations proceed when $\zeta = 0.5860$, which is slightly larger than 0.5859. It can be seen that $\sigma_1^2(t)$ fluctuates as iterations proceed, and does not show sign of convergence. Although Theorem 3 cannot tell what happens when $\alpha^* < 0$ is not satisfied, the numerical results demonstrate a very strong evidence that the convergence condition may also be necessary. Due to the limitation of space, the complete proof containing both sufficiency and necessity for the condition $\alpha^* < 0$ will be presented in a future paper.

VI. Conclusions

In this paper, the necessary and sufficient convergence condition for the variance of Gaussian BP was developed. And the converged variance is proved to be independent of the initialization as long as it is greater or equal to zero. Then, it is proved that the convergence condition can be verified efficiently by solving a SDP problem. Numerical examples are presented to corroborate the proposed theories.

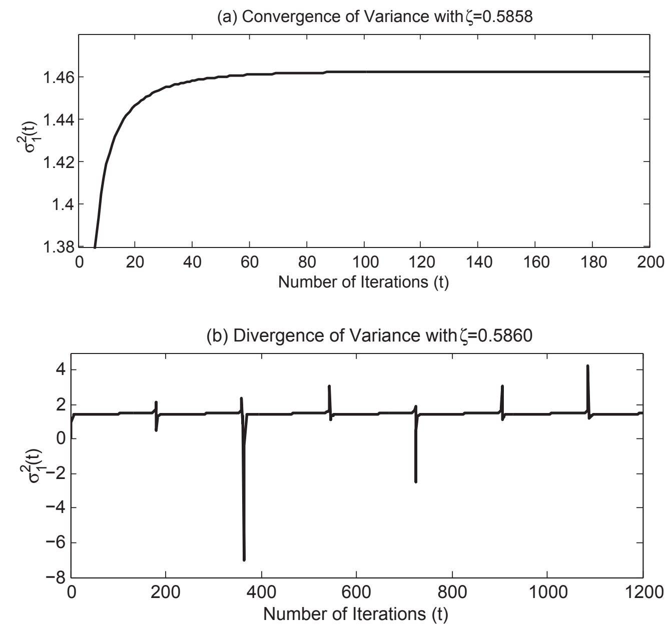

Fig. 2: Illustration for the convergence and divergence of variance $\sigma_1^2(t)$.

References

[1] F. R. Kschischang, B. J. Frey, and H. A. Loeliger, "Factor graphs and the sum-product algorithm," *IEEE Trans. Inf. Theory*, vol. 47, pp. 498-519, 2001.

[2] Y. Weiss and W. T. Freeman, "Correctness of belief propagation in Gaussian graphical models of arbitrary topology," *Neural Comput.*, vol. 13, pp. 2173-2200, 2001.

[3] A. Montanari, B. Prabhakar, and D. Tse, "Belief propagation based multiuser detection," in *Proc. IEEE Information Theory Workshop*, Punta del Este, Uruguay, Mar. 2006.

[4] O. Shental, P. H. Siegel, J. K. Wolf, D. Bickson, and D. Dolev, "Gaussian belief propagation solver for systems of linear equations," in *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, 2008, pp. 1863-1867.

[5] N. Boon Loong, J. S. Evans, S. V. Hanly, and D. Aktas, "Distributed downlink beamforming with cooperative base stations," *IEEE Trans. Inf. Theory*, vol. 54, pp. 5491-5499, 2008.

[6] M. Leng and Y. Wu, "Distributed clock synchronization for wireless sensor networks using belief propagation," *IEEE Trans. Signal Process.*, vol. PP, pp. 1-1, 2011.

[7] A. Ahmad, D. Zennaro, E. Serpedin, and L. Vangelista, "A factor graph approach to clock offset estimation in wireless sensor networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 7, pp. 4244-4260, July 2012.

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

[9] C. C. Moallemi and B. Van Roy, "Convergence of min-sum message passing for quadratic optimization," *IEEE Trans. Inf. Theory*, vol. 55, pp. 2413-2423, 2009.

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

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

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