JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 34, 1251-1272 (2018)
DOI: 10.6688/JISE.201809_34(5).0009

A Distributed Sparse Signal Reconstruction Algorithm in Wireless Sensor Network

ZHI ZHAO[1], JIU-CHAO FENG[1], WEI-YU YU[1], ZI-LIANG REN[1],
BAO PENG[2] AND ZHI-LI ZHOU[3]
[1]*School of Electronic and Information Engineering*
South China University of Technology
Guangzhou, 510640 P.R. China
[2]*School of Electronic and Communication*
Shenzhen Institute of Information Technology
Shenzhen, 518000 P.R. China
[3]*Nanjing University of Information Science and Technology*
Nanjing, 210000 P.R. China
E-mail: zhaozhi.perfect@163.com; {fengjc; yuweiyu}@scut.edu.cn;
ren_zl@126.com; {pengyu323pengbao; zhou_zhili}@163.com

We address the sparse signal reconstruction problem over networked sensing system. Signal acquisition is performed as in compressive sensing (CS), hence the number of measurements is reduced. Majority of existing algorithms are developed based on ℓ_p minimization in the framework of distributed convex optimization and thus whose performance is sensitive to the tuning of additional parameters. In this paper, we propose a distributed sparse signal reconstruction algorithm in the full Bayesian framework by using Variational Bayesian (VB) with embedded consensus filter. Specifically, each node executes one-step average-consensus with its neighbors per VB step and thus reaches a consensus on estimate of sparse signal finally. The proposed approach is ease of implementation and scalability to large networks. In addition, due to the observability of nodes can be enhanced by average-consensus, the number of measurements for each node can be further reduced and not necessary to satisfy lower bound required by CS. Simulation results demonstrate that the proposed distributed approach have good recovery performance and converge to their centralized counterpart.

Keywords: compressive sensing, sparse signal, variational Bayesian, consensus filter, wireless sensor networks

1. INTRODUCTION

The recently developed compressive sensing (CS) theory [1, 2] is a new sampling paradigm that can achieve acquisition of information contained in a large-scale data using only much fewer samples than that required by Nyquist sampling theorem. By exploiting sparsity, which is inherent characteristic of many natural signals, CS enables the signal to be stored in few samples and subsequently be recovered accurately. Indeed, advances in electronics and digital communications have made wireless sensor networks (WSNs) the predicted panacea for solving a variety of large-scale decision and information-processing tasks [3-6]. Hence, CS is a promising technique for WSNs in saving bandwidth and energy. However, most of CS reconstruction algorithms operate in centralized manner where all the measurements need to be concentrated for processing.

Received February 23, 2017; revised May 12 & June 14, 2017; accepted June 28, 2017.
Communicated by Wen-Liang Hwang.

From a practical perspective, these centralized approaches are only attractive to WSNs with a fusion center (FC). Moreover, the centralized approaches have the following drawbacks: (i) the pressure on storage and computation load at the FC tends to increase as the number of nodes grows; (ii) they are sensitive to FC failure. Due to the high fault tolerance and scalability, distributed processing is becoming increasingly popular in the WSNs applications. Different from the centralized approaches that rely on an FC, distributed processing requires no central coordinator and only single-hop communications among neighbors that aim to achieve consensus on local estimates. Recently, many researchers have attempted to discuss the distributed sparse signal recovery problems [7-9, 11] for WSN applications when a centralized approach is not possible or desirable. In [7], the distributed iterative hard thresholding (IHT) algorithm was proposed for tree networks, in which collaboration among the nodes is carried out via the broadcast of messages from the parent nodes and the converge cost of messages from the children nodes. In [8], João F. C. Mota *et al*. have developed a distributed basis pursuit (BP) algorithm for sparse signal reconstruction in noise-free scenario. In [9], the distributed least absolute shrinkage and selection operator algorithm (Lasso) reformulate the Lasso into a separable form which is iteratively minimized using the alternating-direction method of multipliers (ADMM) so as to achieve parallel optimization. In [13], authors proposed a distributed approach to Joint Sparse Model (JSM) based on Lasso and ADMM techniques. In [10], a reweighted ℓ_1 soft thresholding technique is leveraged in distributed recovery of jointly sparse signals. However, these aforementioned works based on convex optimization, such as BP, Lasso, IHT, *etc.*, usually require one or more practically unknown parameters, *e.g.*, the noise statistics, the regularization parameters, *etc.* In addition, a main issue of these deterministic methods is that the uncertainty of signal reconstruction is generally obscure.

On the one hand, an increasing attention has been recently paid to Bayesian algorithms since they generally achieve the better recovery performance and providing probabilistic estimates [14-17]. Moreover, the Bayesian CS algorithms naturally model the unknown signal along with the model parameters, which result in fully automated algorithms estimating all required parameters. As one important family of Bayesian algorithms, sparse Bayesian learning (SBL) was derived from the research area of machine learning [18] and then enriched by many researchers. In SBL, the sparse signal recovery problem is solved from a Bayesian perspective while the sparsity is modeled by a hierarchical prior which is called Automatic Relevance Determination (ARD). Considering JSM-1 model, a distributed Bayesian algorithm has been presented in [12] by using peripheral ADMM procedure. In [19], authors propose a distributed SBL algorithm for in-network recovery of joint sparse signal. These distributed reconstructions are developed by combining SBL with ADMM method. Nevertheless, the relevance vector machine (RVM) approach to SBL is known to converge rather slowly and the computational complexity scales $O(N^3)$. An alternative approach to SBL is the variational rendition of SBL (VSBL) with several advantages compared to SBL [20-22]. Moreover, both [12] and [19] apply the several ADMM procedure in each expectation maximization (EM).

On the other hand, average-consensus algorithms have lately issued as a family of low-complexity iterative distributed algorithms. The consensus formulation was early analyzed in [23] and deals with the problem of computing averages over graphs. Specifically, a group of networked nodes cooperate with each other to adjust their own state

with the goal to reach a consensus in a scalable and fault-tolerant manner [24, 25]. It has received considerable attention in different subjects [26-30]. As previously stated, the existing works, which apply the ADMM to the design of consensus-type sparse signal reconstruction algorithms, involve the implementations that necessitate the fine tuning of additional parameters. Compared with ADMM technique, average-consensus strategies naturally lead to an equalization effect across the sensors without complexity [31].

Motivated by all of the above, this paper aims to propose a distributed algorithm for sparse signal reconstruction in the full Bayesian framework by using variational approximation and average-consensus techniques. We start by deriving a centralized approach for reconstruction with VSBL, which facilitate our distributed algorithm design; afterwards, we move to the distributed scenario and employed average-consensus as a diffusion strategy for acquiring three global information quantities which extracted from centralized approach; particularly, the average-consensus step only needs to be performed once among neighbors in each fixed point iteration of variational SBL for saving communication resource and energy. Furthermore, both the convergence and scalability of algorithm are also discussed.

The remainder of this paper is structured as follows. The fundamental of compressive sampling is provided in Section 2, and introduces the sparse signal recovery using SBL. The system model is described in Section 3. In Section 4, we develop variational Bayesian inference for the system model. In the sequel, the proposed distributed sparse signal reconstruction is provided in Section 5. Numerical results are presented in Section 6, followed by conclusions in Section 7.

Notation: Throughout this paper, we use b, $\mathbf{B}$ and $\mathbf{b}$ for scalars, matrices and column vectors, respectively. The superscripts $(\cdot)^T$ and $(\cdot)^{-1}$ denote the transpose and the inverse of a matrix, respectively. T_r denotes the trace of matrix. $E_{p(x)}(\cdot)$ denotes expectation with respect to $p(x)$. $U_i[a, b]$ and $N(\mu, \Gamma)$ denote integer uniform distribution in the interval $[a, b]$ and multivariate Gaussian distribution with mean vector μ and covariance Γ, respectively. I_N denotes the $M\times N$ identity matrix. $\|\cdot\|_0$, $\|\cdot\|_p$ and $\|\cdot\|_2$ denote ℓ_0-norm, ℓ_p-norm and ℓ_2-norm, respectively.

2. BACKGROUND

In the following, we will briefly review the principle of compressive sensing and sparse Bayesian learning (SBL) [18, 32], which is a centralized CS algorithm.

As a framework for signal sensing and compression, CS theory claims that if a signal is sparse in one basis, it can be recovered from a small number of projections onto a second basis that is incoherent with the first one. If $s\in\Re^N$ represents the signal to be sensed, then noisy sensing process may be represented as

$$\mathbf{z} = \mathbf{\Psi s} + \mathbf{w}. \tag{1}$$

Where $z = \Re^M$ is the measurement vector, $\Psi\in\Re^{M\times N}$ denotes the sensing matrix, and $w \in \Re^M$ denotes the measurement noise. Signal s can be expressed in representation basis as $s = \Theta x$, where $x \in \Re^N$ is k-sparse vector, representing projection coefficients of s on Θ. The basis Θ can be a predefined one depending on the characteristics of signal, *e.g.*,

wavelet basis, Fourier basis, *etc.* While $||x||_0 = k \ll N$, the signal s is said to be sparse over the predefined basis. Thus, measurement vector z can be rewritten in terms of x as follows

$$\mathbf{z} = \mathbf{A}\mathbf{x} + \mathbf{w}. \tag{2}$$

Where $A = \Psi\Theta = \Re^{M\times N}$ is also referred as equivalent sensing matrix.

In order to recover x from the noisy measurements z, the solution of Eq. (2) formally be translated into the following optimization problem:

$$\min_{\mathbf{x}} \|\mathbf{x}\|_0, s.t. \|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2 \le \epsilon. \tag{3}$$

Where $\grave{o} > 0$ is an estimate of the measurement noise level? However, the problem (3) is NP-hard and impractical to solve in the general case. For this reason, conventional CS reconstruction approaches resort to solve the following unconstrained optimization problem by relaxing ℓ_0 to ℓ_p:

$$\min_{\mathbf{x}} \tfrac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{z}\|_2 + \lambda \|\mathbf{x}\|_p. \tag{4}$$

Where $0 < p \le 1$, and the Lagrange multiplier $\lambda > 0$ is a function of A, z and $\grave{o}$. For the case $p = 1$, the problem (4) is known as LASSO (Least Absolute Shrinkage and Selection Operator), the solution of which can be obtained with overwhelming probability. While $p < 1$, the problem (4) is non-convex, but the solution of which shows superior performance by using non-convex optimization techniques [33, 34], *e.g.*, SBL, Bayesian compressed sensing (BCS), *etc.*

From a Bayesian perspective, the CS problem can also be formulated by SBL whose close relationship to non-convex ℓ_p-norm minimization problem is revealed in [35, 36]. In SBL framework, sparse signal is assumed to be distributed according to a parameterized zero-mean Gaussian prior

$$p(\mathbf{x} \mid \mathbf{\Lambda}) = \prod_{i=1}^{N} \mathcal{N}(x_i \mid 0, \alpha_i). \tag{5}$$

Where $\Lambda \in \Re^{M\times N}$ is a diagonal matrix composed of N hyperparameters $\alpha_i (i = 1, \ldots, N)$, whose ith diagonal entry α_i models the variance of x_i for $1 < i \le N$. In [18, 32, 36], the rationale of using this prior have been elaborated. With uninformative hyperpriors $p(\alpha_i)$ and $p(\sigma^2)$, one can infer these hyperparameters by maximizing

$$\ln p(\mathbf{\Lambda}, \sigma^2 \mid \mathbf{z}) \propto \ln p(\mathbf{z} \mid \mathbf{\Lambda}, \sigma^2) = \ln \int p(\mathbf{z} \mid \mathbf{x}, \sigma^2) p(\mathbf{x} \mid \mathbf{\Lambda}) d\mathbf{x}. \tag{6}$$

In fact, the solution of Eq. (6) is equivalent to minimizing the cost function expressed as

$$\mathcal{L} = \ln |\mathbf{\Sigma}| + \mathbf{z}^T \mathbf{\Sigma}^{-1} \mathbf{z}. \tag{7}$$

Where $\sum = \sigma^2 I_M + A \wedge A^T$. In [18], authors apply the expectation-maximization (EM) algorithm to solve Eq. (7). Once the hyperparameters obtained, the MAP estimate of x can be inferred

$$\begin{aligned}\mathbf{x} &= \arg\max_{\mathbf{x}} p(\mathbf{x} \mid \mathbf{z}, \mathbf{\Lambda}, \sigma^2) \\ &= \arg\max_{\mathbf{x}} p(\mathbf{z} \mid \mathbf{x}, \sigma^2) p(\mathbf{x} \mid \mathbf{\Lambda}) \\ &= \mathbf{\Lambda}\mathbf{A}^T\mathbf{\Sigma}^{-1}\mathbf{z}.\end{aligned} \tag{8}$$

In addition, the effectiveness of SBL for sparse signal recovery are presented in [32, 36].

3. PROBLEM STATEMENT AND SYSTEM MODEL

Consider a network composed of K nodes whose connectivity is described by an undirected graph $G = (V, E, A)$ of the order K. Accordingly, node $K \in V$ represent a sensor and can communicate with node $l \in V$ if the edge (k, l) in the set $E \subseteq V \times V$, and the adjacency matrix $A = [\alpha_{kl}]_{K \times K}$ with nonnegative adjacency element α_{kl}, namely, $\alpha_{kl} > 0 \Leftrightarrow (k, l) \in E$. Node l is called a neighbor of node k if $(k, l) \in E$ and $l \neq k$. The neighborhood set of node k is denoted by N_k. The degree of vertex k is set as $d_i = |N_k|$, and the maximum degree is $d_{max} = \max_k d_k$. Each node is able to process the data and collaborate with its single-hop neighbors. See Fig. 1 for an example graph.

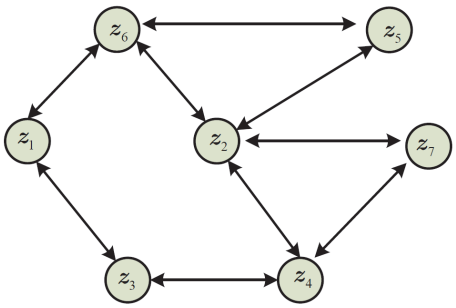

Fig. 1. An example of network structure.

Assume that each node k is interested in reconstructing an unknown sparse signal $x \in \Re^N$ from m_k noisy measurements $z_k \in \Re^{m_k}$. Thus, the sensing model at node k is

$$\mathbf{z}_k = \mathbf{A}_k\mathbf{x} + \mathbf{w}_k,\ 1 \le k \le K. \tag{9}$$

Where $A_k \in \Re^{m_{k \times N}}$ is the local sensing matrix for node k, and $w_k \in \Re^{m_k}$ is the zero-mean Gaussian noise with covariance $\beta^1 \mathrm{I}_{mk}$. Let M be the total number of measurements from all the nodes, *i.e.*, $M = \sum_{k=1}^{K} m_k$. We have the global measurement $Z \in \Re^M$, the global sensing matrix, $A \in \Re^{M \times N}$, and the global measurement noise $W \in \Re^M$ as follows

$$\mathbf{Z} = \begin{bmatrix} \mathbf{z}_1 \\ \vdots \\ \mathbf{z}_K \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \vdots \\ \mathbf{A}_K \end{bmatrix}, \mathbf{W} = \begin{bmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{bmatrix}. \tag{10}$$

Then the global sensing model is given by

$$\mathbf{Z} = \mathbf{A}\mathbf{x} + \mathbf{W}. \tag{11}$$

The construction of A satisfies so-called restricted isometry property (RIP) imposed in the design of compressive sensing schemes, *i.e.*, the elements of A are drawn from $N(0,1/M)$. From the sensing model (3) and noise statistics, the measurements likelihood function is given by

$$\begin{aligned} p(\mathbf{Z} \mid \mathbf{x}, \beta) &= \prod_{k=1}^{K} p(\mathbf{z}_k \mid \mathbf{x}, \beta) \\ &= \prod_{k=1}^{K} \frac{\beta^{m_k/2}}{(2\pi)^{m_k/2}} \exp\left[-\frac{\beta}{2} \left\| \mathbf{z}_k - \mathbf{A}_k \mathbf{x} \right\|^2 \right]. \end{aligned} \tag{12}$$

Moreover, as stated previously, the estimate of x is constrained to be sparse. To this end, the likelihood should be complemented by suitable conjugate priors over x and β. Specifically, a Gamma prior with parameters c and d is selected for precision β

$$p(\beta \mid c, d) = G(\beta \mid c, d) = \frac{d^c \beta^{c-1} \exp[-d\beta]}{\Gamma(c)}. \tag{13}$$

From a probabilistic point of view, a heavy-tailed distribution is well-suited to reflect prior knowledge about sparsity of signal. Here, two levels prior is set for our Bayesian model. For the first level, a Gaussian distribution is adopted for x similar to SBL, *i.e.*,

$$\begin{aligned} p(\mathbf{x} \mid \boldsymbol{\alpha}) &= \mathcal{N}(\mathbf{x} \mid 0, \boldsymbol{\Lambda}^{-1}) = \prod_{i=1}^{N} p(x_i \mid \alpha_i) \\ &= \prod_{i=1}^{N} (2\pi)^{-1/2} \alpha_i^{1/2} \exp\left[-\frac{1}{2} x_i^2 \alpha_i \right]. \end{aligned} \tag{14}$$

Where $\alpha = [\alpha_1, \alpha_2, \ldots, \alpha_N]^T$ and $\Lambda = \mathrm{diag}(\alpha)$. In the second level, a Gamma distribution is selected for the precision parameters $\alpha_i s$

$$p(\alpha_i \mid a_i, b_i) = G(\alpha_i \mid a_i, b_i) = \frac{b_i^{a_i} \alpha_i^{a_i - 1} \exp[-b_i \alpha_i]}{\Gamma(a_i)}. \tag{15}$$

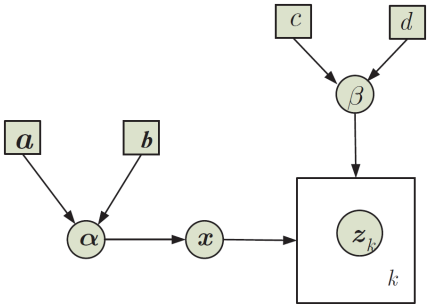

Fig. 2. DAG of the Bayesian model.

Heretofore, the Bayesian system model is developed for sparse signal reconstruction in WSN. Fig. 2 illustrates the directed acyclic graph (DAG) of the proposed Bayesian model. Comparing the proposed model with that in [18], the difference is the normalization of the variances of $x_i s$ by β in Eq. (6), which ensures the unimodality of posterior joint distribution. Based on the proposed Bayesian model, our goal is to recover the sparse signal x at each node using distributed processing. Next, we firstly derive centralized variational Bayesian method to facilitate the distributed algorithm design.

4. VARIATIONAL BAYESIAN APPROXIMATION FOR CENTRALIZED SPARSE SIGNAL RECONSTRUCTION

In this section, we briefly review the variational approximation technique and derive the centralized VSBL for aforementioned full Bayesian model. The obtained centralized algorithm facilitates our distributed algorithm design, which is presented in the next section. For ease of notation, we define $\psi = [x_1, \ldots, x_N, \beta, \alpha_1, \ldots, \alpha_N]^T$ as the unknown parameters and hidden variables of the model which are referred to as unknown variables, and $\theta = [\alpha_1, \ldots, \alpha_N, b_1, \ldots, b_N, c, d]^T$ as the hyperparameters of the imposed prior. Based on the Bayesian model previous, variational Bayesian is to approximate the posterior of ψ, $p(\psi|Z)$, by a more tractable distribution $Q(\psi)$. To this end, the hyperparameters θ is inferred by maximinzing the following log-likelihood

$$\ln p(\mathbf{Z}|\boldsymbol{\theta}) = F(Q(\boldsymbol{\psi})) + KL(Q(\boldsymbol{\psi}))||p(\boldsymbol{\psi}|\mathbf{Z}, \boldsymbol{\theta}). \tag{16}$$

Where F is the free energy

$$F(Q(\boldsymbol{\psi}), \boldsymbol{\theta}) = \int Q(\boldsymbol{\psi}) \ln\left(\frac{p(\mathbf{Z}\,|\,\boldsymbol{\psi})\, p(\boldsymbol{\psi}\,|\,\boldsymbol{\theta})}{Q(\boldsymbol{\psi})}\right). \tag{17}$$

and

$$KL(Q(\boldsymbol{\psi})\,||\,p(\boldsymbol{\psi}\,|\,\mathbf{Z}, \boldsymbol{\theta})) = \int Q(\boldsymbol{\psi}) \log\left(\frac{Q(\boldsymbol{\psi})}{p(\boldsymbol{\psi}\,|\,\mathbf{Z}, \boldsymbol{\theta})}\right) \tag{18}$$

is the Kullback-Leibler (KL) divergence between the posterior $p(\boldsymbol{\Psi}|\mathbf{Z}, \boldsymbol{\theta})$ and a tractable distribution $Q(\Psi)$. In Eq. (1), since the KL divergence is non-negative and the log-likelihood $\ln p(Z, \theta)$ is fixed with respect to $Q(\Psi)$, the variational free energy can be viewed as lower bound for $\ln p(Z, \theta)$. Therefore, minimizing the KL divergence is equivalent to maximizing the variational free energy. From an optimization point of view, the model parameters of $Q(\Psi)$ is well-suited selected so that the lower bound can be minimized. In order to make it tractable, we resort to a simpler variational free form $Q(\Psi)$ to approximate the posterior based on the mean-field theory from statistical physics. Specifically, $Q(\Psi)$ can be factorized into a family of q-distribution w.r.t. some partitions $\Psi = \{\Psi_1, \ldots, \Psi_L\}$ as follows

$$Q(\boldsymbol{\psi}) = \prod_{j=1}^{L} q_j(\boldsymbol{\psi}_j) \tag{19}$$

i.e., each partition ψ_j of the unknown variables is mutually independent given the measurements. In fact, if we let ψ_i denote the jth partition of the vector $\psi = [x_1, \ldots, x_N, \beta, \alpha_1, \ldots, \alpha_N]^T$ containing the parameters of the Bayesian hierarchical model, and ψ_{-j} refers to the other parameters after removing jth partition. Maximizing the free energy in Eq. (2) is realized by taking functional derivatives with respect to each of the $q(\cdot)$ distributions while fixing the other distributions and setting $\partial F(q)/\partial q(\cdot) = 0$ [37]. Furthermore, the computation of $\partial F(q)/\partial q(\cdot) = 0$ can be expressed as

$$\ln q^*(\boldsymbol{\psi}_j) \propto E_{q(\boldsymbol{\psi}_{-j})}[\ln(p(\mathbf{Z}|\boldsymbol{\psi})p(\boldsymbol{\psi}|\boldsymbol{\theta}))]. \tag{20}$$

Where $E_{q(\psi_{-j})}$ denotes the expectation w.r.t. $\prod_{i\neq j} q(\psi_i)$. From Eq. (5), it is noted that q^* don't represent explicit solution since it depends on the other factors $q(\psi_i)$ for $i \neq j$. Thus, the optimum solution $q^*(\psi_j)$ is sought by first initializing all the $q(\psi_j)$ appropriately and then cycling through the factors and replacing each in turn with updated estimate given by Eq. (5) using current estimates for all of the other factors. Since the lower bound is convex w.r.t. $q(\psi_j)$, the convergence is guaranteed.

By applying the variational bayesian techniques to our model, we first take the following logarithm of the joint distribution over Z and all unknown variables $\boldsymbol{\psi}$

$$\begin{aligned} \ln(p(\mathbf{Z}\,|\,\boldsymbol{\psi})p(\boldsymbol{\psi}\,|\,\boldsymbol{\theta})) = &\sum_{k=1}^{K}\ln p(\mathbf{z}_k\,|\,\mathbf{x},\beta) + \sum_{i=1}^{N}\ln p(x_i\,|\,\alpha_i) \\ &+ \sum_{i=1}^{N}\ln p(\alpha_i\,|\,a_i,b_i) + \ln\, p(\beta\,|\,c,d). \end{aligned} \tag{21}$$

For each factor, averaging w.r.t. those variables not in that factor by making use of Eq. (5), we can express the re-estimation equations for the factors analytically, namely

$$\begin{aligned} &\ln q(\mathbf{x}) \propto E_{q(\boldsymbol{\alpha})q(\beta)}\left[\ln(p(\mathbf{Z}\,|\,\boldsymbol{\psi})p(\boldsymbol{\psi}\,|\,\boldsymbol{\theta}))\right] \\ &\propto E_{q(\beta)}\left[\sum_{k=1}^{K}\ln p(\mathbf{z}_k\,|\,\mathbf{x},\beta)\right] + E_{q(\boldsymbol{\alpha})}\left[\sum_{i=1}^{N}\ln p(x_i\,|\,\alpha_i)\right] \\ &\propto -\sum_{k=1}^{K}\frac{E[\beta]}{2}(\mathbf{z}_k - \mathbf{A}_k\mathbf{x})^T(\mathbf{z}_k - \mathbf{A}_k\mathbf{x}) - \frac{1}{2}\mathrm{diag}(E[\alpha_i])\mathbf{x}^T\mathbf{x} \\ &\propto -\frac{1}{2}\mathbf{x}^T(\mathrm{diag}(E[\alpha_i]) + \sum_{k=1}^{K}E[\beta]\mathbf{A}_k^T\mathbf{A}_k)\mathbf{x} + \sum_{k=1}^{K}E[\beta]\mathbf{x}^T\mathbf{A}^T\mathbf{z}_k \\ &\propto \mathcal{N}(\mathbf{x}\,|\,\boldsymbol{\mu},\boldsymbol{\Gamma}). \end{aligned} \tag{22}$$

Where

$$\begin{aligned} \boldsymbol{\Gamma} &= \left(\mathrm{diag}(E[\alpha_i]) + \sum_{k=1}^{K}E[\beta]\mathbf{A}_k^T\mathbf{A}_k\right)^{-1} \\ \boldsymbol{\mu} &= \sum_{k=1}^{K}E[\beta]\boldsymbol{\Gamma}\mathbf{A}_k^T\mathbf{z}_k \end{aligned} \tag{23}$$

and

$$\begin{aligned}
&\ln q(\beta) \propto E_{q(\mathbf{x})q(\boldsymbol{\alpha})}[\ln(p(\mathbf{Z}\mid\boldsymbol{\psi})p(\boldsymbol{\psi}\mid\boldsymbol{\theta}))] \\
&\propto E_{q(\mathbf{x})}\left[\sum_{k=1}^{K}\ln p(\mathbf{z}_k\mid\mathbf{x},\beta)\right]+\ln p(\beta\mid c,d) \\
&\propto \sum_{k=1}^{K}-\frac{\beta}{2}(\mathbf{z}_k^T\mathbf{z}_k-\mathbf{z}_k^T\mathbf{A}_kE[\mathbf{x}]-E[\mathbf{x}]^T\mathbf{A}_k\mathbf{z}_k+E[\mathbf{x}^T\mathbf{A}_k^T\mathbf{A}_k\mathbf{x}]) \\
&\quad+\sum_{k=1}^{K}\frac{m_k}{2}\ln\beta+(c-1)\ln\beta-d\beta \\
&\propto \sum_{k=1}^{K}-\frac{\beta}{2}(\mathbf{z}_k^T\mathbf{z}_k-\mathbf{z}_k^T\mathbf{A}_kE[\mathbf{x}]-E[\mathbf{x}]^T\mathbf{A}_k\mathbf{z}_k+\mathrm{Tr}\left(\mathbf{A}_k^T\mathbf{A}_kE[\mathbf{x}\mathbf{x}^T]\right)) \\
&\quad+\frac{M}{2}\ln\beta+(c-1)\ln\beta-d\beta \\
&\propto G(\beta\mid\tilde{c},\tilde{d})
\end{aligned} \tag{24}$$

and

$$\begin{aligned}
&\ln q(\boldsymbol{\alpha}) \propto E_{q(\mathbf{x})q(\boldsymbol{\beta})}\left[\ln(p(\mathbf{Z}\mid\boldsymbol{\psi})p(\boldsymbol{\psi}\mid\boldsymbol{\theta}))\right] \\
&\propto E_{q(\mathbf{x})}\left[\sum_{i=1}^{N}\ln p(x_i\mid\alpha_i)\right]+\sum_{i=1}^{N}\ln p(\alpha_i\mid a_i,b_i) \\
&\propto \sum_{i=1}^{N}\left((a_i-1)\ln\alpha_i-b_i\alpha_i+\frac{1}{2}\ln\alpha_i-\alpha_iE[x_i^2]\right) \\
&\propto G(\alpha_i\mid\tilde{a}_i,\tilde{b}_i)=G(\boldsymbol{\alpha}\mid\tilde{a},\tilde{b}).
\end{aligned} \tag{25}$$

From Eqs. (22)-(25), we recognize that $q(x)$ is Gaussian distribution, $q(\beta)$ and $q(\alpha)$ are Gamma distributions, *i.e.*, $q(x) = N(x|\mu, \Gamma)$, $q(\beta) = G(\beta|\tilde{c}, \tilde{d})$, and $q(\alpha) = G(\alpha|\tilde{a}, \tilde{b})$. Now given $q(x)$, $q(\beta)$ and $q(\alpha)$, the hyperparameters can be updated by $\boldsymbol{\theta} = \arg\max_{\theta} F(Q(\boldsymbol{\psi}), \boldsymbol{\theta})$. Specifically, we have

$$\tilde{a}_i = a+\frac{1}{2}, \tag{26a}$$

$$\tilde{b}_i = b+\frac{E[x_i^2]}{2}, \tag{26b}$$

$$\tilde{c} = c+\frac{M}{2}, \tag{26c}$$

$$\begin{aligned}
\tilde{d} = d+&\frac{1}{2}\sum_{k=1}^{K}\mathbf{z}_k^T\mathbf{z}_k-\sum_{k=1}^{K}E[\mathbf{x}]^T\mathbf{A}_k^T\mathbf{z}_k+ \\
&\frac{1}{2}\mathrm{Tr}\left(\sum_{k=1}^{K}\mathbf{A}_k^T\mathbf{A}_k\cdot\left(\boldsymbol{\Gamma}+\boldsymbol{\mu}^T\boldsymbol{\mu}\right)\right).
\end{aligned} \tag{26d}$$

In the above, the required moments can be easily computed as follows

$$\begin{aligned} E[\mathbf{x}] &= \boldsymbol{\mu} \\ E[x_i^2] &= \boldsymbol{\Gamma}_{ii} + \boldsymbol{\mu}_i^2 \\ E[\alpha_i] &= \tilde{a}_i / \tilde{b}_i \\ E[\beta] &= \tilde{c} / \tilde{d} \end{aligned} \tag{27}$$

The variational optimization proceeds by iteratively updating Eqs. (7)-(12) until convergence to stable hyperparameters θ. From the aforementioned formulas, it is noted that a fusion center is required to gather all the measurements and then reconstruct the sparse signal. Thus, the centralized variational sparse bayesian algorithm for solving the joint sparse signal reconstruction can be summarized by the following steps:

(1) Input local measurements $\{z_k\}_{k\in V}$, local sensing matrices $\{A_k\}_{k\in V}$ and adjacency matrix A;
(2) Initialize the hyperparameters a_i, b_i, c, d with 10^{-6}, and μ by 0, Γ by I_N;
(3) All the nodes transmit the local measurements Z_K to fusion center;
(4) Fusion center receives the measurements from network-wide nodes and computes sufficient statistics $E[\alpha_i]$, $E[\beta]$, μ and Γ using Eq. (8), thus the variational distributions are calculated;
(5) Update the hyperparameters as in Eq. (12);
(6) Unless the hyperparameters converge to fixed values, Steps 3, 4 and 5 are iterated continually;
(7) Fusion center returns the estimate of sparse signal x to all the nodes.

However, there is no fusion center in the distributed scenario and formulas derived above can't be implemented directly. In order to develop the distributed counterpart, we attempt to reformulate the centralized formulas which will be presented in the following section.

5. DISTRIBUTED VARIATIONAL SPARSE BAYESIAN LEARNING ALGORITHM

In this section, we propose a distributed variational Bayesian algorithm for jointly reconstructing sparse signal, which is developed from the previous variational Bayesian inference. Hereinafter, the distributed algorithm will be referred to as DVSBL. From the aforementioned centralized algorithm, it is noted that both Steps 4, and 5 are performed at a fusion center, which gathers all the nodes' measurements and does the centralized variational inference. However, for a distributed scenario, it is assumed that each node performs the inference severally and has no knowledge of other nodes' measurement matrices and measurements. Therefore, by inspecting the former equations, we define the following global information quantities

$$\mathcal{I}^{(1)} = \sum_{k=1}^{K} \mathbf{A}_k^T \mathbf{A}_k$$

$$\begin{aligned}\mathcal{I}^{(2)} &= \sum_{k=1}^{K} \mathbf{A}_k^T \mathbf{z}_k \\ \mathcal{I}^{(3)} &= \sum_{k=1}^{K} \mathbf{z}_k^T \mathbf{z}_k\end{aligned} \tag{28}$$

and define the following local quantities

$$\begin{aligned}\mathcal{I}_k^{(1)} &= \mathbf{A}_k^T \mathbf{A}_k \\ \mathcal{I}_k^{(2)} &= \mathbf{A}_k^T \mathbf{z}_k \\ \mathcal{I}_k^{(3)} &= \mathbf{z}_k^T \mathbf{z}_k\end{aligned} \tag{29}$$

so the Eqs. (23) and (26d) can be reformulated as follows

$$\boldsymbol{\Gamma} = \left(\mathrm{diag}(E[\alpha_i]) + E[\beta]\mathcal{I}^{(1)}\right)^{-1}, \tag{30}$$

$$\boldsymbol{\mu} = E[\beta]\boldsymbol{\Gamma}\mathcal{I}^{(2)}, \tag{31}$$

$$\tilde{d} = d + \frac{1}{2}\mathcal{I}^{(3)} - E[\mathbf{x}]^T \mathcal{I}^{(2)} + \frac{1}{2}\mathrm{Tr}\left(\mathcal{I}^{(1)} \cdot \left(\boldsymbol{\Gamma} + \boldsymbol{\mu}^T \boldsymbol{\mu}\right)\right). \tag{32}$$

It is obvious that the calculation of required parameters involves three global quantities. However, each sensor only interacts solely with its neighbors in the distributed scenario, thus the global quantities $I^{(1)}$, $I^{(2)}$, $I^{(3)}$ can't be calculated locally. It is noted that these global information quantities can also be redetermined by averaging the local quantities from all nodes in Eq. (18).

$$\begin{aligned}\bar{\mathcal{I}}^{(1)} &= \frac{1}{K}\sum_{k=1}^{K} \mathcal{I}_k^{(1)} \\ \bar{\mathcal{I}}^{(2)} &= \frac{1}{K}\sum_{k=1}^{K} \mathcal{I}_k^{(2)} \\ \bar{\mathcal{I}}^{(3)} &= \frac{1}{K}\sum_{k=1}^{K} \mathcal{I}_k^{(3)}\end{aligned} \tag{33}$$

Note that the redefinition of global quantities has no impact on the parameter approximation in Eqs. (23) and (26). To obtain the global average at each node, an average consensus filter suggested in [38] can be employed to approximate the global information quantities defined in Eq. (33). Particularly, the local information quantities possessed by each node are interchanged with their neighbors, then the global average is approximated asymptotically at each node depending upon the local information quantities input from others by using consensus filer. Hence, the DVSBL algorithm can be developed by employing such average consensus filter.

According to [38], a consensus filter can be formulated by following continuous compact form

$$\dot{\xi}_k = \sum_{l \in \mathcal{N}_k} (\xi_l - \xi_k) + \sum_{l \in \mathcal{N}_k \cup k} (u_l - \xi_k). \tag{34}$$

Where x_k is the filter state of node k, which approximates the filter input u_l. The discrete-time form of consensus filter suggested in [38] is as follows:

$$\xi_k^{t+1} = \xi_k^t + \eta^t \left[\sum_{l \in \mathcal{N}_k} (\xi_l^t - \xi_k^t) + \sum_{l \in \mathcal{N}_k \cup k} (u_l^t - \xi_k^t) \right]. \tag{35}$$

Where the superscript t denotes the number of iteration; η is the updating rate and should be

$$\begin{aligned} &\eta^t \le \frac{1}{d_{max}}, \\ &\sum_{t=1}^{\infty} \eta^t = \infty, \\ &\sum_{t=1}^{\infty} (\eta)^2 < \infty. \end{aligned} \tag{36}$$

Eq. (36) is the stableness condition of the discrete consensus filter according to the Gershgorin theorem [39]. Thus, the filtering algorithm can be carried out in a distributed manner if the averages Eq. (33) can be obtained by every node. Before the distributed algorithm is presented, we use $u_k^t = \{\mathrm{I}_k^{(1)}, \mathrm{I}_k^{(2)}, \mathrm{I}_k^{(3)}\}$ to denote the local information quantities in the node k, and x_k^t denote the estimated global information quantities. Here, both u_k^t and ξ_k^t are referred as vector. In particular, the vector factors $x_k^t(1)$, $x_k^t(2)$, $x_k^t(3)$ are the approximations of $\mathrm{I}^{-(1)}$, $\mathrm{I}^{-(2)}$, $\mathrm{I}^{-(3)}$. The consensus filter in node k takes the local quantities u_k and neighbors' approximated global quantities x_l^t as inputs. The filter states asymptotically converge to

$$\begin{aligned} &\xi_k^t(1) \to \frac{1}{K} \sum_{k=1}^{K} \mathcal{I}_k^{(1)}, \\ &\xi_k^t(2) \to \frac{1}{K} \sum_{k=1}^{K} \mathcal{I}_k^{(2)}, \\ &\xi_k^t(3) \to \frac{1}{K} \sum_{k=1}^{K} \mathcal{I}_k^{(3)}. \end{aligned} \tag{37}$$

Simultaneously, the hyperparameters $\{\hat{a}_{k,i}, \hat{b}_{k,i}, \hat{c}_{k,i}, \hat{d}_{k,i}, \boldsymbol{\Gamma}_k, \boldsymbol{\mu}_k\}$ at each node will be approached by their estimate $\{\hat{a}_{k,i'}, \hat{b}_{k,i'}, \hat{c}_{k'}, \hat{d}_{k'}, \hat{\boldsymbol{\Gamma}}_k, \boldsymbol{\mu}_k\}$, as the exact global information quantity vector is approached by its estimated value. Thus

$$\hat{\boldsymbol{\Gamma}}_k^t = \left(\mathrm{diag}(E[\alpha_{k,i}^t]) + E[\beta_k^t] K \xi_k^t(1) \right)^{-1} \tag{38a}$$

$$\hat{\boldsymbol{\mu}}_k^t = E[\beta_k^t] \hat{\boldsymbol{\Gamma}}_k{}^t K \xi_k^t(2) \tag{38b}$$

$$\hat{a}_{k,i}^t = a + \frac{1}{2} \tag{38c}$$

$$\hat{b}_{k,i}^t = b + \frac{E[x_{k,i}^2]}{2} \tag{38d}$$

$$\hat{c}_k^t = c + \frac{M}{2} \tag{38e}$$

$$\hat{d}_k^t = d + \frac{1}{2} K\xi_k^t(3) - E[\mathbf{x}_k]^T K\xi_k^t(2) + \frac{1}{2}\mathrm{Tr}\left(K\xi_k^t(1)\cdot\left(\hat{\boldsymbol{\Gamma}}_k^t + \hat{\boldsymbol{\mu}}_k^{t^T}\hat{\boldsymbol{\mu}}_k^t\right)\right) \tag{38f}$$

Where the according moments are given by

$$\begin{aligned} E[\mathbf{x}_k] &= \hat{\boldsymbol{\mu}}_k^t \\ E[x_{k,i}^2] &= \hat{\boldsymbol{\Gamma}}_{ii}^t + \hat{\boldsymbol{\mu}}_k^t \\ E[\alpha_{k,i}^t] &= \hat{a}_{k,i}^t / \hat{b}_{k,i}^t \\ E[\beta_k^t] &= \hat{c}_k^t / \hat{d}_k^t \end{aligned} \tag{39}$$

Heretofore, the DVSBL algorithm has been developed above. The algorithm diagram is illustrated in Fig. 3. In the following, we will analyze the convergence property and scalability of the proposed distributed algorithm.

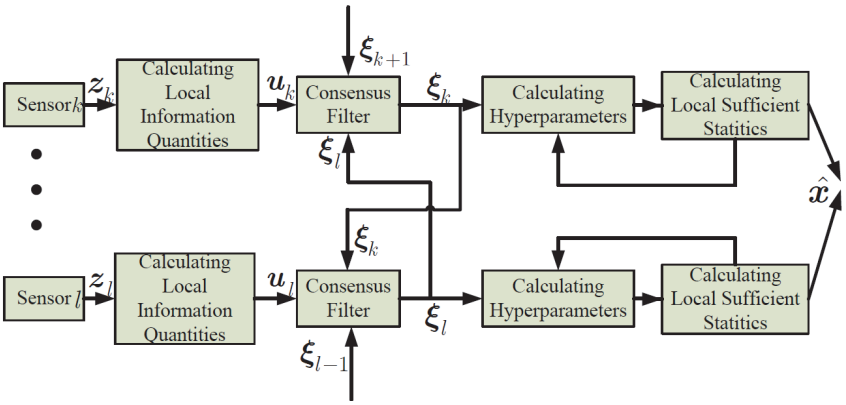

Fig. 3. Algorithm architecture.

Remark: As the variational approximation is essentially an EM-like scheme, the log-likelihood $P(Z|\theta)$ is guaranteed to be increased for each update iteration until a fixed point is reached. Hence, the proposed algorithm is guaranteed to converge [40]. Moreover, it is noted that the consensus update is independent of EM update in DVSBL, *i.e.*, the updated result of EM has no impact on consensus update. So, it is obvious that the output of consensus update finally converges to average of local information quantity as iteration

increasing. In the end, the updates of Eq. (2) will convergence to the true hyperparameters values as the global quantities obtained by consensus filter.

5.1 Scalability Analysis

Next, we investigate the scalability of the proposed algorithm by showing that the computational burden of each node will not be influenced by size of the network (K). For this purpose, a formula for the average computational burden will be given later. It will be shown that the computational burden of each node does not depend upon size of network but it depends only upon the average number of its neighboring nodes (N_{avg}).

Assume that T denotes total number of required iterations. From the distributed algorithm presented above, it is noted that the computational burden of each node can be given by

$$\mathcal{F} = \mathcal{T} \cdot (f_1(m_k) + f_2(N_{avg}) + f_3(N)).$$

Where f_1, f_2, f_3 are three functions, which represent the three different stages causing computational burden for each iteration. In the first stage, each node calculate the local information quantities. Thus, the computational burden of this stage is a function of the dimension of local measurement at each node, which is denoted by $f_1(m_k)$. The second stage involves updating the global information quantities using the consensus filter, and its computational load is dependent on number of neighboring nodes *i.e.* $f_2(N_{avg})$. In the last stage, the model parameters are updated; thus, its computational burden depends upon N, which is represented by $f_3(N)$. As a consequence, it is easy to see that the computational complexity of the proposed algorithm does not depend upon size of the network, namely, the algorithm is performed in a scalable manner.

Remark: In [11], it is hard to know the number of consensus iterations that are necessary in different scenarios. In [12], two thorough ADMM procedures are required in each EM iteration, which is inevitable to consume large time and energy. Compared with the works in [11, 12], our approach only need once average-consensus iteration in each VB step. It means the communication resource and energy can be saved to the great extent.

6. SIMULATIONS

In this section, simulation results are provided to demonstrate the performance of the proposed distributed algorithm by experiments with synthetic signals and real humidity signals. In our experiments, three approaches are compared:

(1) Distributed VSBL: each node reconstructs the signal severally in a distributed manner by the proposed algorithm;
(2) Independent VSBL: signal is reconstructed independently at each node by VSBL, which only exploits the local information;
(3) Centralized VSBL: signal is jointly reconstructed at a fusion center by VSBL, where FC can gather all the information from nodes.

For comparison purposes, the reconstruction performance is evaluated by normalized mean square error (NMSE) and averaged relative error (ARE), which are defined as follows:

$$\text{NMSE} = \frac{\left\|\mathbf{x} - \hat{\mathbf{x}}_k\right\|_2}{\left\|\mathbf{x}\right\|_2}, \tag{40}$$

$$\text{ARE} = \frac{\sum_{k=1}^{K} \left\|\hat{\mathbf{x}}_k - \mathbf{x}\right\|_2^2}{K \cdot \left\|\mathbf{x}\right\|_2^2}. \tag{41}$$

In addition, the computer simulations involve two different networks. The first one is an L-connected Harary graph model, where each node is only available to communicate with L adjacent neighbors to exchange information. The other one is Erdös-Rényi graph model, which is constructed by connecting nodes randomly, *i.e.*, each edge is included in the graph with probability p independent from every other edge. In the proposed distributed algorithm, the updating rate of the consensus filter is chosen as

$$\eta = \frac{1}{d_{max} + \tau t}, \text{ where } 0 < \tau < 1. \tag{42}$$

6.1 Experiments with Synthetic Data

Fig. 4 shows a simulated set of sensor nodes placement. The sensors are located in a 1km×1km square randomly. According to underlying scheduling scheme, a network with 6 nodes is employed to observe linear combinations of sparse signal, which is represented by an undirected graph $G = (V, E, A)$ with the set of nodes $A = (1,2,3,4,5,6)$, the set of edges $E = \{(1,2), (1,4), (1,6), (2,3), (2,5), (3,4), (3,6), (4,1), (4,5), (5,2), (5,6), (6,1), (6,3)\}$ and the adjacency matrix

$$\mathcal{A} = \begin{bmatrix} 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 \end{bmatrix}. \tag{43}$$

In this example, the signal $x = \Re^{256}$ is assumed to be sparse itself. There are altogether 10 non-zero elements $x\{i\} \neq 0$ in the sparse signal, where i is the index of support and $x\{i\}$ denote the value of support. Here, the index and value are unknown and sampled over $i \in U_i[1, 256]$ and $x\{i\} \in N(0, 5^2)$, respectively. We randomly generate a set of $m_k \times 256$ measurement matrices $\{A_k, k = 1, 2, \ldots, 6\}$ from the standard independent and identically distributed (i.i.d) Gaussian ensemble. Then, we generate the local measurements $\{z_k, k = 1, 2, \ldots, 6\}$ for each node. The received measurements are corrupted by additive zero-mean Gaussian noise to yield signal noise ratio (SNR), *i.e.* $\frac{\|\mathbf{A}_k \mathbf{x}_k\|_2^2}{\|\mathbf{w}_k\|_2^2}$, of 20dB.

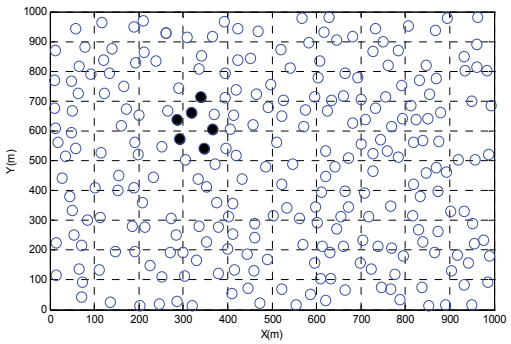

Fig. 4. Random placement.

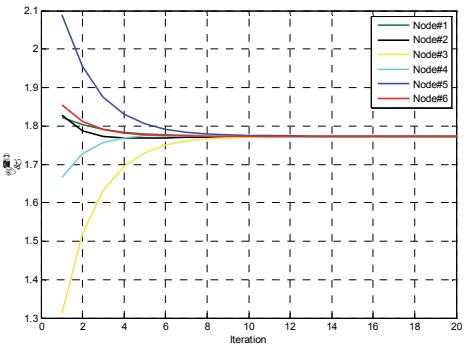

Fig. 5. Evolution of global information quantity $\mathbf{I}_k^{(3)}$.

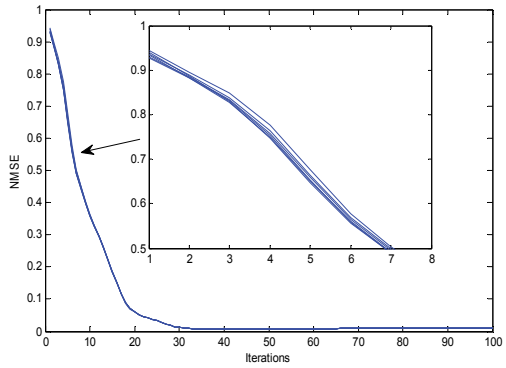

Fig. 6. Normalized MSE versus iteration ($K = 6$, $m_k = 12$).

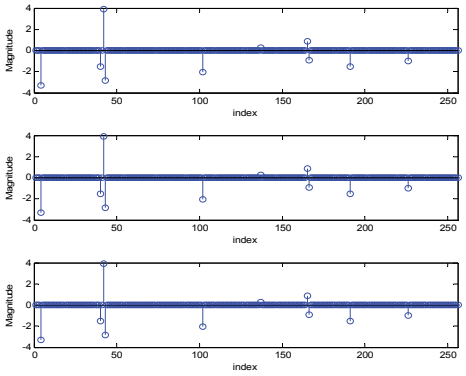

Fig. 7. Estimation of x ($K = 6$, $m_k = 12$).

The convergence rate of algorithm is given in Figs. 5 and 6, which show the local quantity $\mathrm{I}_k^{(3)}$ versus the number of iteration and Normalized MSE versus iteration, respectively. As expected, our numerical simulation results confirm that all the nodes reach a consensus on the local quantity and estimate of sparse signals as the iteration increasing. Moreover, it is easy to see that both the NMSE and local quantity have almost identical convergence rate.

In addition, Fig. 7 compares the estimates of x from three nodes using DVSBL algorithm. It is obvious that all nodes give satisfactory estimates of the actual sparse signal. Furthermore, it is noted that the performance is achieved with far fewer measurements than the unknowns (< 5%) for each node. For instance, in this example, only 20% of the lower bound $10\ln 256 \approx 56$ are required by each node.

In order to show scalability of the proposed algorithm, the size of network is increased to 24 nodes and the network is modeled as a 5-connected Harary graph which is illustrated in Fig. 8. Here, the performance of the centralized VSBL is also given as a benchmark, but the performance of aforementioned optimization-based algorithms is not provided due to their huge computational complexity. As a final illustration to highlight the consensus property of the algorithms, Fig. 9 shows the evolution of the signal supports estimated by each node as the iteration index grows. For each one of the $s = 10$ supports, $K = 24$ different curves are shown, one per node. As we can see, all local esti-

mates reach a common limiting value after several iterations. We also give the NMSE of 24 nodes shown in Fig. 10. From the figure, we can see that the convergence rate has no significant changes compared with Fig. 6. In Fig. 11, it is noted that the reconstruction quality is very close to that of the centralized VSBL and the difference can be ignored.

Fig. 12 shows the Average Relative Error of the proposed algorithm with different instances for a Erdös-Rényi model, which is illustrated in Fig. 13. In addition, the iteration steps while convergence attained for different network settings are provided in Table 1.

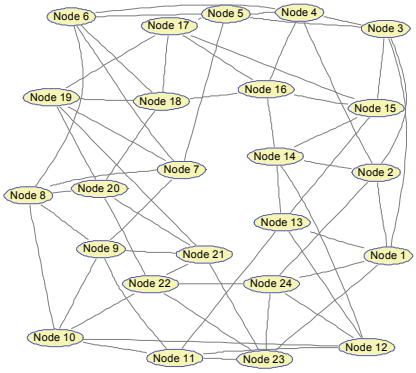

Fig. 8. Harary graph.

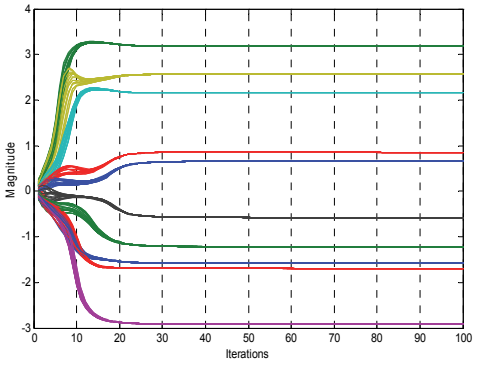

Fig. 9. Evolution of the per node estimates.

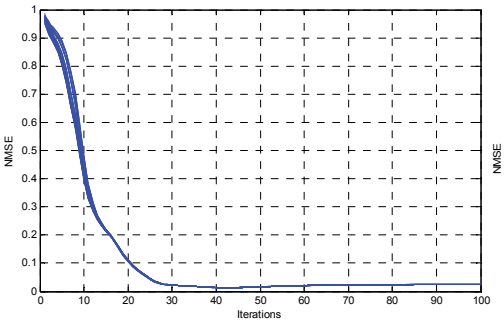

Fig. 10. Normalized MSE versus iteration (K = 24, m_k = 3).

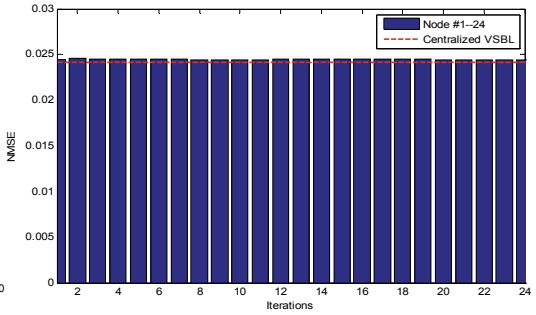

Fig. 11. Comparison of NMSE after t = 100.

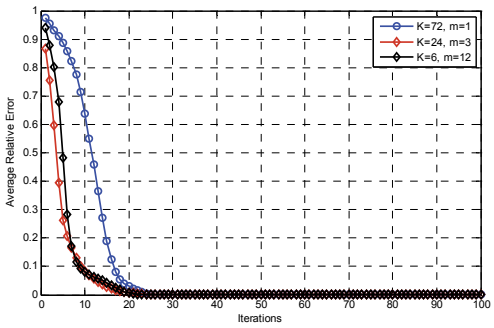

Fig. 12. ARE versus iterations for different instances.

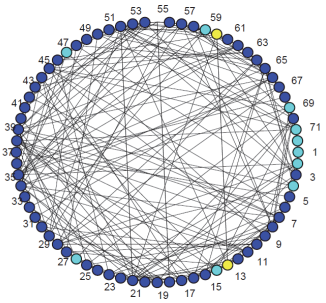

Fig. 13. Erdös-Rényi model.

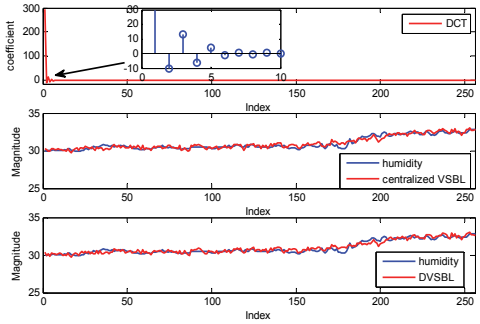

Fig. 14. Reconstructed results of environmental humidity signal ($K=24$, $m_k=3$ and Harary graph model with $L=3$).

Table 1. The iterations of convergence in the proposed algorithm.

	$K=24$ $m_k=3$	$K=24$ $m_k=6$	$K=24$ $m_k=9$	$K=72$ $m_k=1$	$K=72$ $m_k=3$
Harary graph model ($L=2$)	23	12	9	52	16
Harary graph model ($L=3$)	20	9	6	26	12
Erdös-Rényi model ($p=0.3$)	15	12	5	27	10
Erdös-Rényi model ($p=0.5$)	19	8	4	24	7

Table 2. The average relative error for different approach.

	$K=24$ $m_k=3$	$K=24$ $m_k=6$	$K=72$ $m_k=1$	$K=72$ $m_k=3$
Independent VSBL	0.9631	0.9611	0.9976	0.9856
Harary graph model ($L=3$)	3.3×10^{-4}	4.3×10^{-4}	4.4×10^{-4}	2.8×10^{-4}
Erdös-Rényi model ($p=0.5$)	6.3×10^{-4}	4.4×10^{-4}	7.8×10^{-4}	1.4×10^{-3}

6.2 Experiments with Real Data

Next, we verify the effectiveness of the proposed algorithm with real signal. Different from the synthetic signal previous, the real signal itself may not be sparse but sparse over some representation basis. To this end, we adopt the humidity signals obtained from the Intel Berkeley Research lab. Moreover, the discrete cosine transform (DCT) is employed as the sparse representation basis. In the experiments, the length of signal is also restricted to $N=256$. The samples of humidity signal are sensed by $K=24$ nodes severally. By exploiting the samples, both centralized and distributed reconstructed results are evaluated.

From Fig. 14, it is obvious that both the proposed distributed algorithm and centralized algorithm successfully reconstruct the humidity signal. For our proposed algorithm, the reconstructed result is only for one of the 24 nodes. In fact, the other nodes have a nearly consensus reconstructed results. For the centralized VSBL, all the nodes' samples need to be gathered at an FC for joint reconstruction, while the proposed distributed algorithm enables joint reconstruction without an FC. Table 2 gives more performance comparisons under different settings, and further evaluates the effectiveness of algorithm.

7. CONCLUSION

In this paper, we tackle the sparse signal reconstruction in WSNs, where all the nodes cooperate and interact with each other without centralized coordination. A distributed variational Bayesian algorithm is proposed for full Bayesian system model. First, we have addressed the centralized scenario, where all node measurements are available at fusion center. In this case, a centralized approach is derived to facilitate the design of distributed counterpart. Next, a distributed reconstruction algorithm is obtained by means of consensus filter due to its efficiency and low-complexity. In particular, there exists merely one-step average-consensus iteration in each VB update. To evaluate the effectiveness of the proposed algorithm, the numerical simulations on both synthetic and real data demonstrate that the proposed algorithm has comparable recovery performance and convergence properties.

REFERENCES

1. D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, Vol. 52, 2004, pp. 1289-1306.
2. R. G. Baraniuk, E. Candes, R. Nowak, and M. Vetterli, "Compressive sampling," *IEEE Signal Processing Magazine*, Vol. 25, 2008, pp. 12-13.
3. D. Brunelli and C. Caione, "Sparse recovery optimization in wireless sensor networks with a sub-nyquist sampling rate," *Sensors*, Vol. 15, 2015, pp. 16654-16673.
4. J. Shen, H. W. Tan, and J. W. Wang, *et al.*, "A novel routing protocol providing good transmission reliability in underwater sensor networks," *Journal of Internet Technology*, Vol. 16, 2015, pp. 171-178.
5. S. D. Xie and Y. X. Wang, "Construction of tree network with limited delivery latency in homogeneous wireless sensor networks," *Wireless Personal Communications*, Vol. 78, 2014, pp. 231-246.
6. Y. Weng, W. D. Xiao, and L. H. Xie, "Diffusion-based EM algorithm for distributed estimation of Gaussian mixtures in wireless sensor networks," *Sensors*, Vol. 11, 2011, pp. 6297-6316.
7. S. Patterson, Y. Eldar, and I. Keidar, "Distributed compressed sensing for static and time-varying networks," *IEEE Transactions on Signal Processing*, Vol. 62, 2014, pp. 4931-4946.
8. J. F. C. Mota, J. M. F. Xavier, and P. M. Q. Aguiar, *et al.*, "Distributed basis pursuit," *IEEE Transactions on Signal Processing*, Vol. 25, 2012, pp. 1942-1956.
9. G. Mateos, J. A. Bazerque, and G. B. Giannakis, "Distributed sparse linear regression," *IEEE Transactions on Signal Processing*, Vol. 10, 2010, pp. 5262-5276.
10. S. Fosson, J. Matamoros, and C. Antón-Haro, "Distributed recovery of jointly sparse signals under communication constraints," *IEEE Transactions on Signal Processing*, Vol. 64, 2016, pp. 3470-3482.
11. B. Peng, Z. Zhao and G. Han, *et al.*, "Consensus-based sparse signal reconstruction algorithm for wireless sensor networks," *International Journal of Distributed Sensor Networks*, Vol. 12, 2016, pp. 1-12.
12. W. Chen and I. Wassell, "A decentralized bayesian algorithm for distributed com pressive sensing in networked sensing systems," *IEEE Transactions on Wireless Communications*, Vol. 15, 2015, pp. 1282-1292.

13. J. Matamoros, S. M. Fosson and E. Magli, *et al.*, "Distributed admm for in-network reconstruction of sparse signals with innovations," *IEEE Transactions on Signal and Information Processing over Networks*, Vol. 1, 2015, pp. 225-234.
14. S. Ji, Y. Xue and L. Carin, "Bayesian compressive sensing," *IEEE Transactions on Signal Processing*, Vol. 56, 2008, pp. 2346-2356.
15. D. Baron, S. Sarvotham, and R. G. Baraniuk, "Bayesian compressive sensing via belief propagation," *IEEE Transactions on Signal Processing*, Vol. 58, 2010, pp. 269-280.
16. J. P. Vila and P. Schniter, "Expectation-maximization Gaussian-mixture approximate message passing," *IEEE Transactions on Signal Processing*, Vol. 61, 2014, pp. 4658-4672.
17. J. P. Vila and P. Schniter, "An empirical-Bayes approach to recovering linearly constrained non-negative sparse signals," *IEEE Transactions on Signal Processing*, Vol. 62, 2014, pp. 4689-4703.
18. M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The Journal of Machine Learning Research*, Vol. 1, 2001, pp. 211-244.
19. S. Khanna and C. R. Murthy, "Decentralized joint-sparse signal recovery: A sparse Bayesian learning approach," *IEEE Transactions on Signal and Information Processing over Networks*, Vol. 3, 2016, pp. 29-45.
20. M. E. Tipping, "Variantional relevance vector machines," *Eprint Arxiv*, Vol. 28, 2000, pp. 46-53.
21. D. Shutin and T. Buchgraber, "Fast variational sparse Bayesian learning with automatic relevance determination for superimposed signals," *IEEE Transactions on Signal Processing*, Vol. 59, 2011, pp. 6527-6261.
22. K. E. Themelis, A. A. Rontogiannis, and K. D. Koutroumbas, "A variational bayes framework for sparse adaptive estimation," *IEEE Transactions on Signal Processing*, Vol. 62, 2014, pp. 4723-4736.
23. J. Tsitsiklis, D. Bertsekas, and M. Athans, "Distributed asynchronous deterministic and stochastic gradient optimization algorithms," *IEEE Transactions on Automatic Control*, Vol. 31, 2003, pp. 803-812.
24. R. Olfati-Saber and J. S. Shamma, "Consensus filters for sensor networks and distributed sensor fusion," in *Proceedings of the 44th IEEE Conference on Decision and Control*, 2005, pp. 6698-6703.
25. R. Olfati-Saber and R. M. Murray, "Consensus problems in networks of agents with switching topology and time-delays," *IEEE Transactions on Automatic Control*, Vol. 49, 2004, pp. 1520-1533.
26. N. Amelina, A. Fradkov, and Y. Jiang, *et al.*, "Approximate consensus in stochastic networks with application to load balancing," *IEEE Transactions on Information Theory*, Vol. 61, 2015, pp. 1739-1752.
    Y. Wang, K. T. Ta, X. Y. Peng, and P. L. So, *et al.*, "Coordinated control of distributed energy storage systems for voltage regulation in distribution networks," *IEEE Transactions on Power Delivery*, Vol. 31, 2016, pp. 1132-1141.
27. Z. Feng, G. Wen, and G. Hu, "Distributed secure coordinated control for multiagent systems under strategic attacks," *IEEE Transactions on Cybernetics*, Vol. 47, 2017, pp. 1273-1284.
28. B. Safarinejadian and M. E. Estahbanati, "Consensus filter-based distributed variational bayesian algorithm for flow and speed density prediction with distributed traffic sensors," *IEEE Systems Journal*, 2015, pp. 1-10.

29. H. M. La and W. Sheng, "Distributed sensor fusion for scalar field mapping using mobile sensor networks," *IEEE Transactions on Cybernetics*, Vol. 43, 2013, pp. 766-778.
30. A. H. Sayed, "Adaptive networks," *Proceedings of the IEEE*, Vol. 102, 2014, pp. 460-497.
31. D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Transactions on Signal Processing*, Vol. 52, 2004, pp. 2153-2164.
32. M. E. Davies and R. Gribonval, "Restricted isometry constants where ℓ_p sparse recovery can fail for $0 \le p \le 1$," *IEEE Transactions on Information Theory*, Vol. 55, 2009, pp. 2203-2214.
33. R. Wu and D. R. Chen, "The improved bounds of restricted isometry constant for recovery via ℓ_p-minimization," *IEEE Transactions on Information Theory*, Vol. 59, 2013, pp. 6142-6147.
34. D. P. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *IEEE Journal of Selected Topics in Signal Processing*, Vol. 4, 2010, pp. 317-329.
35. D. P. Wipf, B. D. Rao and S. Nagarajan, "Latent variable bayesian models for promoting sparsity," *IEEE Transactions on Information Theory*, Vol. 57, 2010, pp. 6236-6255.
36. H. Zhu, H. Leung, and Z. He, "State estimation in unknown non-Gaussian measurement noise using variational Bayesian technique," *IEEE Transactions on Aerospace and Electronic System*, Vol. 49, 2013, pp. 2601-2614.
37. D. Kingston and R. Beard, "Discrete-time average-consensus under switching network topologies," in *Proceedings of American Control Conference*, 2006, pp. 3551-3556.
38. R. A. Horn and C. R. Johnson, *Matrix Analysis*, Publishing House, Cambridge, UK, 1986.
39. J. M. Winn and C. M. Bishop, "Variational message passing," *Journal of Machine Learning Research*, Vol. 6, 2005, pp. 661-694.

Zhi Zhao (赵智) received the B.Eng. degree in Electronic and Information Engineering from Xiangtan University, Xiangtan, China, in 2008; and the M.Eng. degree in Signal and Information Processing from Xiangtan University, Xiangtan, China, in 2011. From 2011 to 2013, he worked at Datang Telecom Technology and Industry Group. He is currently a Ph.D. candidate at the School of Electronic and Information, South China University of Technology, Guangzhou, China. His areas of interests include signal processing in wireless sensor network and nonlinear theories.

Jiu-Chao Feng (冯久超) received the B.Sc. degree in Physics from Southwest China Normal University in 1986, the M.Eng. degree in Communication and Electronic Systems from the South China University of Technology, Guangzhou, China, in 1997, the Ph.D. degree from The Hong Kong Polytechnic University, Hong Kong, in 2002. His areas of interests include digital signal processing, digital communications, nonlinear theories, and their applications.

Wei-Yu Yu (余卫宇) received the Ph.D. degree in Information and Communication Engineering from the Institute of Electronic Information Engineering, South China University of Technology, Guangzhou, 2005. His research interests are in the areas of sensor networks, image processing.

Zi-Liang Ren (任子良) is currently a Ph.D. candidate at the School of Electronic and Information, South China University of Technology, Guangzhou, China. His current research interests include blind signal processing and machine learning. E-mail: ren_zl @126.com

Bao Peng (彭保) received the Ph.D. degree in Information and Communication Engineering from the Institute of Electronic Information Engineering, Harbin Institute of Technology (HIT) in Harbin, Heilongjiang, 2009. His research interests are in the areas of sensor networks, distributed systems, and information management.

Zhi-Li Zhou (周志立) received the B.S. degree in Communication Engineering from Hubei University in 2007, and the M.S. and Ph.D. degrees in Computer Application from the School of Information Science and Engineering, Hunan University, in 2010 and 2014, respectively. He joined Nanjing University of Information Science and Technology, China, as an Assistant Professor in 2014. He is currently a Postdoctoral Fellow with the Department of Electrical and Computer Engineering, University of Windsor, Canada. His current research interests include near-duplicate image/video retrieval, image search, image/video copy detection, coverless information hiding, digital forensics, and image processing.