A Decentralized Bayesian Algorithm For Distributed Compressive Sensing in Networked Sensing Systems

Wei Chen, *Member, IEEE,* and Ian J. Wassell

***Abstract*—Compressive sensing (CS), as a new sensing/sampling paradigm, facilitates signal acquisition by reducing the number of samples required for reconstruction of the original signal, and thus appears to be a promising technique for applications where the sampling cost is high, e.g., the Nyquist rate exceeds the current capabilities of analog-to-digital converters (ADCs). Conventional CS, although effective for dealing with one signal, only leverages the intra-signal correlation for reconstruction. This paper develops a decentralized Bayesian reconstruction algorithm for networked sensing systems to jointly reconstruct multiple signals based on the distributed compressive sensing (DCS) model that exploits both intra- and inter-signal correlations. The proposed approach is able to address networked sensing system applications with privacy concerns and/or for a fusion-centre-free scenario, where centralized approaches fail. Simulation results demonstrate that the proposed decentralized approaches have good recovery performance and converge reasonably quickly.**

I. Introduction

RECENTLY developed compressed sensing (CS) theory and principles [2], [3] enable sampling and processing of analog signals at rates far below the Nyquist rate. Therefore, it has been proposed for applications where the sampling cost is high, e.g., wideband spectrum sensing for cognitive radio [4], [5], multipath channel identification with a high time resolution [6], super-resolution radar [7], imaging systems [8]–[10], and air quality monitoring [11].

CS exploits the intra-signal correlation, i.e., the sparse structure of a signal, to reconstruct the original signal from a few random measurements. In addition to the intra-signal correlation, signals in a network may have high spatial correlation. Such spatial correlation, which represents inter-signal correlation, has not been considered in the conventional CS framework. In [12], Quer et al. propose to adaptively update the sparsifying basis to capture the spatial and temporal characteristics of the network signal via principal component analysis (PCA), and reconstruct the signal by CS under this basis. Another widely used type of approach is to jointly reconstruct of a group of signals to leverage inter-signal correlation. For example, considering a cognitive radio sensor network application with CS measurements taken at different positions, one could improve the spectrum detection accuracy by jointly recovering all the nodes' spectra. The most straightforward way for modeling multiple measurement vectors (MMVs) [13], [14] is to assume that all signals share a common support, which is however too strict in many applications. In [15] a method to statistically characterize real world signals in space and time is provided. As an extension of CS, distributed compressive sensing (DCS) is proposed in [16], [17] to model the intra-signal and inter-signal correlations. In this paper, we consider a particular signal model, namely the Type-1 joint sparse model (JSM-1), which is one of the three generative models for joint sparse signals introduced in the context of DCS [16], [17]. This DCS model involves a common component and an innovation component for modelling the global factors and local factors corresponding to distinct signals, respectively.

For the centralized approach to joint signal reconstruction in DCS, all the data needs to be communicated to a fusion centre (FC) for processing. However, this scheme has the following drawbacks: i) the pressure on storage and computation load at the FC tends to increase as the number of nodes grows; ii) sensitive or private local sensor data is exposed to the FC; iii) it cannot be applied in a fusion-centre-free scenario. Decentralized processing in networked sensing systems avoids these drawbacks, and thus is attractive for applications involving sensitive data, those lacking a FC, or for a big data scenario. While most CS reconstruction algorithms operate in a centralized manner, some decentralized sparse reconstruction algorithms [18], [19] have been proposed for CS applications when a centralized approach is not possible or desirable. A survey of the state-of-the-art in CS for distributed systems is given in [20]. However, these decentralized algorithms are designed for the CS rather than the DCS setting that involves multiple distinct signals.

In this paper, a decentralized Bayesian algorithm is proposed for joint reconstruction of multiple sensor signals which follow the JSM-1 DCS model. In contrast to centralized algorithms where CS measurements are reported to a FC and the reconstruction is performed at the FC, the proposed algorithm is performed at each node with some inter-node communication. To achieve the goal of decentralized processing, we first decouple the common component from innovation components by applying variational Bayesian approximation. Then we cast the decoupled reconstruction problem as a set of decentralized problems with consensus constraints, where

Wei Chen is with the State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China, and also with the Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, United Kingdom (e-mail: wc253@cam.ac.uk).

Ian J. Wassell is with the Computer Laboratory, University of Cambridge, Cambridge CB3 0FD, United Kingdom (e-mail: ijw24@cam.ac.uk).

A Part of this work was presented at the 2015 IEEE International Conference on Communications [1].

each node exchanges limited non-sensitive information with its neighbors and recovers its own innovation component by using local data. Therefore, in the proposed approach, the innovation components, that reflect the intra-node correlations of different nodes, and which can be considered as sensitive data are not shared, and the common component, which reflects the inter-node correlation, is jointly reconstructed. For example consider a distributed air quality monitoring application that runs on a peer-to-peer network of smartphones. In this case the sensing cost is high, e.g., nondispersive infrared (NDIR) or metal oxide based gas sensors consume energy several orders of magnitude greater than that owing to inter-node communication, consequently the energy consumption per node can be reduced using the proposed distributed technique while also maintaining privacy for the sensitive local data at each device. Experimental results show that the proposed decentralized algorithm permits a good reconstruction quality in comparison to other existing approaches, and exhibits a good convergence rate.

The rest of the paper is organized as follows: Section II describes the background of CS and DCS. In Section III, we develop variational Bayesian inference for JSM-1 DCS. In the sequel, the proposed decentralized Bayesian DCS approach is provided in Section IV. Numerical results are presented in Section V, followed by conclusions in Section VI.

The following notation is used. Lower-case letters denote numbers, boldface upper-case letters denote matrices, and boldface lower-case letters denote column vectors. The superscripts $(\cdot)^T$, $(\cdot)^{-1}$ and $(\cdot)^\dagger$ denote the transpose, the inverse and the pseudoinverse of a matrix, respectively. $rank(\mathbf{X})$ and $|\mathbf{X}|$ denotes the rank and the determinant of matrix $\mathbf{X}$, respectively. x_i denotes the ith element of $\mathbf{x}$ and $X_{i,i}$ denotes the ith diagonal element of $\mathbf{X}$. $\texttt{diagm}(\mathbf{x})$ and $\texttt{diagv}(\mathbf{X})$ denote a diagonal matrix corresponding to the vector $\mathbf{x}$ and a vector corresponding to the diagonal matrix $\mathbf{X}$, respectively. $\mathbb{E}_{p(x)}(\cdot)$ denotes expectation with respect to $p(x)$, i.e., the distribution of x. $\mathcal{N}(\mathbf{x};\boldsymbol{\mu},\boldsymbol{\Sigma})$ denotes that $\mathbf{x}$ follows the multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$. $\mathbf{I}_n$ denotes the $n \times n$ identity matrix. The ℓ_0 norm, ℓ_p norm ($0 < p \leq 1$) and the ℓ_2 norm of vectors, are denoted by $\|\cdot\|_0$, $\|\cdot\|_p$ and $\|\cdot\|_2$, respectively. The Frobenius norm of a matrix $\mathbf{X}$ is denoted by $\|\mathbf{X}\|_F$.

II. Background

In this section, we first briefly introduce the background of CS and sparse Bayesian learning (SBL) [21], [22], which is a centralized CS algorithm. Then the JSM-1 DCS model is presented that is an extension of CS for joint reconstruction of multiple signals with both sparse structures and inter-signal correlation.

A. CS Model

For the CS data acquisition, a signal $\mathbf{f} \in \mathbb{R}^n$ is measured as

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{f} + \mathbf{e}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ is the measurement vector, $\boldsymbol{\Phi} \in \mathbb{R}^{m\times n}$ denotes the sensing matrix, and $\mathbf{e} \in \mathbb{R}^m$ denotes the noise term for the measuring process. It is assumed that $\mathbf{f}$ has a sparse representation $\mathbf{x} \in \mathbb{R}^n$ on some basis $\boldsymbol{\Psi} \in \mathbb{R}^{n\times n}$, so that $\mathbf{f} = \boldsymbol{\Psi}\mathbf{x}$. The basis can be a predefined one, e.g., a wavelet transform or a Fourier transform, depending upon the signal characteristics. The signal is said to be sparse over the basis when $\|\mathbf{x}\|_0 = s \ll n$. Therefore, we can rewrite (1) by

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{2}$$

where $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Psi} \in \mathbb{R}^{m\times n}$ denotes the equivalent sensing matrix.

The CS recovery procedure corresponds to the solution of the optimization problem given by:

$$\min_{\mathbf{x}} \quad \|\mathbf{x}\|_0, \qquad \text{s.t.} \quad \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon, \tag{3}$$

where $\epsilon_k > 0$ is an estimate of the measurement noise level. As solving (3) is NP-hard, the typical signal reconstruction process behind conventional CS approaches involves solving the following optimization problem:

$$\min_{\mathbf{x}} \quad \frac{1}{2}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 + \lambda\|\mathbf{x}\|_p, \tag{4}$$

where $0 < p \leq 1$, and $\lambda > 0$ is a penalty parameter. When $p = 1$, the problem in (4) becomes a convex problem, which is often referred to as the least absolute shrinkage and selection operator (LASSO). The ℓ_1-type regularizer often achieves suboptimal performance as it is a convex relaxation of the ℓ_0-type one, while the regularizer with $p < 1$, which is non-convex but a closer approximation of sparsity, shows superior performance [23], [24].

SBL formulates the CS problem from a Bayesian perspective, and its close relationship to a non-convex ℓ_p-norm ($p < 1$) minimization problem is unveiled in [25], [26]. The SBL framework considers a zero-mean Gaussian prior distribution

$$p(\mathbf{x};\boldsymbol{\Gamma}) = \mathcal{N}(\mathbf{x};\mathbf{0},\boldsymbol{\Gamma}) \tag{5}$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{n\times n}$ is a diagonal matrix composed of n hyperparameters γ_i ($i = 1,\ldots,n$). The rationale for the using this prior to model sparse signals is provided in [21], [22], [26]. With uniform hyperpriors $p(\gamma_i)$ and $p(\sigma^2)$, the value of these hyperparameters can be inferred by

$$\begin{aligned}\max_{\boldsymbol{\Gamma},\sigma^2} \log p(\boldsymbol{\Gamma},\sigma^2|\mathbf{y}) &\propto \max_{\boldsymbol{\Gamma},\sigma^2} \log p(\mathbf{y};\boldsymbol{\Gamma},\sigma^2) \\ &= \max_{\boldsymbol{\Gamma},\sigma^2} \log \int p(\mathbf{y}|\mathbf{x};\sigma^2)p(\mathbf{x};\boldsymbol{\Gamma})d\mathbf{x} \\ &\propto \min_{\boldsymbol{\Gamma},\sigma^2} \log|\boldsymbol{\Sigma}| + \mathbf{y}^T\boldsymbol{\Sigma}^{-1}\mathbf{y},\end{aligned} \tag{6}$$

where $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T$. In [21], the expectation-maximization (EM) algorithm is employed to solve (6). Given these hyperparameters, $\mathbf{x}$ can be inferred by maximizing the posterior distribution

$$\begin{aligned}\mathbf{x} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y};\boldsymbol{\Gamma},\sigma^2) \\ &= \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x};\sigma^2)p(\mathbf{x};\boldsymbol{\Gamma}) \\ &= \boldsymbol{\Gamma}\mathbf{A}^T\boldsymbol{\Sigma}^{-1}\mathbf{y}.\end{aligned} \tag{7}$$

The effectiveness of SBL for solving sparse reconstruction in comparison to many other algorithms has been demonstrated in [22], [26].

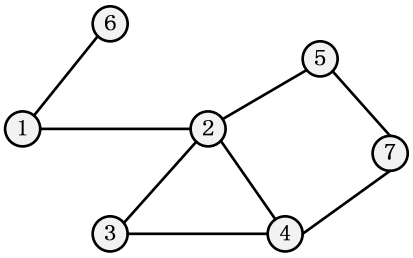

Fig. 1. A generic network structure.

B. DCS Model

We now consider a network with K nodes modeled by an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \ldots, K\}$ is the set of nodes and $\mathcal{E} \subset \mathcal{V} \times \mathcal{V}$ is the set of edges that describe the communication links among the nodes. Each node is able to process locally stored data and exchange messages with its neighbors. See Fig. 1 for an example graph.

Assume each node performs sampling based on the CS principle, and the samples are corrupted by some noise. Then we have

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{e}_k, \tag{8}$$

where $\mathbf{y}_k \in \mathbb{R}^{m_k}$, $\mathbf{A}_k \in \mathbb{R}^{m_k \times n}$, $\mathbf{x}_k \in \mathbb{R}^n$ and $\mathbf{e}_k \in \mathbb{R}^{m_k}$ denote the measurement vector, the sensing matrix, the sparse signal representation, and noise of node k, respectively.

Conventional CS only exploits the intra-signal correlation that is reflected in the sparse signal structure, while DCS provides a means to further leverage the inter-signal correlation and to jointly recover multiple signals. For the JSM-1 DCS setting, the sparse signal representation $\mathbf{x}_k$ $(k = 1, \ldots, K)$ can be decomposed as

$$\mathbf{x}_k = \mathbf{z}_c + \mathbf{z}_k, \tag{9}$$

where $\mathbf{z}_c \in \mathbb{R}^n$ with $\|\mathbf{z}_c\|_0 = s_c \ll n$ denotes the common component of the sparse representation $\mathbf{x}_k$, which captures the inter-signal correlation and is common to all signals, and $\mathbf{z}_k \in \mathbb{R}^n$ $(i = 1, \ldots, K)$ with $\|\mathbf{z}_k\|_0 = s_k \ll n$ denotes the innovations component of the sparse representation $\mathbf{x}_k$, which captures the intra-signal correlation and is specific to the signal k.

In [16], Baron et al. propose to jointly reconstruct multiple signals with the JSM-1 DCS model by solving the following optimization problem:

$$\min_{\tilde{\mathbf{z}}} \ \frac{1}{2}\|\mathbf{A}\tilde{\mathbf{z}} - \tilde{\mathbf{y}}\|_F^2 + \lambda \|\tilde{\mathbf{z}}\|_1 \tag{10}$$

where $\lambda > 0$, $\tilde{\mathbf{z}} = \begin{bmatrix}\mathbf{z}_c^T & \mathbf{z}_1^T & \ldots & \mathbf{z}_K^T\end{bmatrix}^T \in \mathbb{R}^{(K+1)n}$ is the extended signal vector, $\tilde{\mathbf{y}} = \begin{bmatrix}\mathbf{y}_1^T & \ldots & \mathbf{y}_K^T\end{bmatrix}^T \in \mathbb{R}^{\sum_{k=1}^K m_k}$ is the extended measurements vector and $\mathbf{A} \in \mathbb{R}^{\sum_{k=1}^K m_k \times (K+1)n}$ is the extended sensing matrix given by:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_1 & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & & & & \ddots & \vdots \\ \mathbf{A}_K & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{A}_K \end{bmatrix}.$$

In [27], a Fréchet mean approach is proposed for joint reconstruction of multiple correlated signals with a reduced computational complexity. Instead of solving (10) with concatenated measurements $\tilde{\mathbf{y}}$, a crude estimate of the common component is inferred directly from the measurements, and then those signals are recovered one by one with the use of the estimate of the common component. The Fréchet mean of K sparse signals, i.e., $\tilde{\mathbf{z}}_c \in \mathbb{R}^n$, can be obtained from the measurements as follows:

$$\tilde{\mathbf{z}}_c = \arg\min_{\tilde{\mathbf{z}}_c} \sum_{k=1}^{K} \lambda_k d^2(\mathbf{A}_k \tilde{\mathbf{z}}_c, \mathbf{y}_k), \tag{11}$$

where $\lambda_k > 0$ denotes the contribution weight of the kth signal and $d(\mathbf{A}_k\tilde{\mathbf{x}}, \mathbf{y}_k)$ denotes the distance function between the vector $\mathbf{A}_k\tilde{\mathbf{z}}_c$ and $\mathbf{y}_k$. By using the Euclidean distance function, the Fréchet mean is given by:

$$\tilde{\mathbf{z}}_c = (\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^T\hat{\mathbf{y}}, \tag{12}$$

where the extended sensing matrix $\hat{\mathbf{A}} \in \mathbb{R}^{(\sum_{k=1}^K m_k)\times n}$ and the extended measurement vector $\hat{\mathbf{y}} \in \mathbb{R}^{\sum_{k=1}^K m_k}$ are given by $\hat{\mathbf{A}} = \left[\sqrt{\lambda_1}\mathbf{A}_1^T, \cdots, \sqrt{\lambda_K}\mathbf{A}_K^T\right]^T$ and $\hat{\mathbf{y}} = \left[\sqrt{\lambda_1}\mathbf{y}_1^T, \cdots, \sqrt{\lambda_K}\mathbf{y}_K^T\right]^T$ respectively.

III. Variational Bayesian Inference for DCS

In this section, we develop variational Bayesian inference for solving the joint reconstruction problem for the JSM-1 DCS setting. This approach decouples the reconstruction of the common component, that characterizes inter-node correlation, from the innovation components, that represent intra-node correlation, and thus facilitate our decentralized algorithm design, which is presented in Section IV.

A. Variational Sparse Bayesian Inference

Akin to the SBL framework [21], we adopt zero-mean Gaussian prior distributions for the common component and innovation components, respectively, which are given as

$$p(\mathbf{z}_c; \mathbf{\Gamma}_c) = \mathcal{N}(\mathbf{z}_c; \mathbf{0}, \mathbf{\Gamma}_c) \tag{13}$$

and

$$p(\mathbf{z}_k; \mathbf{\Gamma}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{0}, \mathbf{\Gamma}_k), \tag{14}$$

where $\mathbf{\Gamma}_c \in \mathbb{R}^{n\times n}$ is a diagonal matrix with hyperparameters $\gamma_{c,i}$ $(i = 1, \ldots, n)$, and $\mathbf{\Gamma}_k \in \mathbb{R}^{n\times n}$ is a diagonal matrix with hyperparameters $\gamma_{k,i}$ $(k = 1, \ldots, K; i = 1, \ldots, n)$. Assuming elements of the measurement noise vector $\mathbf{e}_k$ are drawn from independent and identically distributed (i.i.d.) zero-mean Gaussian distributions with variance σ^2, we can write the likelihood function as

$$p(\mathbf{y}_k|\mathbf{z}_c, \mathbf{z}_k; \sigma^2) = \mathcal{N}(\mathbf{y}_k; \mathbf{A}_k(\mathbf{z}_c + \mathbf{z}_k), \sigma^2\mathbf{I}_{m_k}). \tag{15}$$

We now adopt the variational approximation in the Bayesian formulation of JSM-1 DCS to find separable functions that approximate the posterior of $\mathbf{z}_c$ and $\mathbf{z}_k$, which facilitates the development of a decentralized algorithm. The essence of variational inference is to find some distribution which usually has a factorized form and closely approximates the true posterior distribution. Variational approximation provides a method to bypass the requirement of exactly knowing the posterior.

To simplify the notation, here we define $\mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_K\}$, $\mathbf{Z} = \{\mathbf{z}_c, \mathbf{z}_1, \ldots, \mathbf{z}_K\}$ and $\boldsymbol{\theta} = \{\boldsymbol{\Gamma}_c, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_K, \sigma^2\}$. Our goal is to estimate the value of the hyperparameters, i.e., $\boldsymbol{\theta}$, which maximize the following log-likelihood

$$\log p(\mathbf{Y};\boldsymbol{\theta}) = F(q(\mathbf{Z}),\boldsymbol{\theta}) + \mathtt{KL}(q(\mathbf{Z})\|p(\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta})), \quad (16)$$

where

$$F(q(\mathbf{Z}),\boldsymbol{\theta}) = \int q(\mathbf{Z}) \log\left(\frac{p(\mathbf{Z},\mathbf{Y};\boldsymbol{\theta})}{q(\mathbf{Z})}\right) d\mathbf{Z}, \quad (17)$$

and

$$\mathtt{KL}(q(\mathbf{Z})\|p(\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta}) = -\int q(\mathbf{Z}) \log\left(\frac{p(\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta})}{q(\mathbf{Z})}\right) d\mathbf{Z} \quad (18)$$

is the Kullback-Leibler (KL) divergence between the true posterior $p(\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta})$ and a variational distribution $q(\mathbf{Z})$. The KL divergence $\mathtt{KL}(q(\mathbf{Z})\|p(\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta})) \geq 0$ and equality holds only when $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta})$. Therefore, $F(q(\mathbf{Z}),\boldsymbol{\theta})$ can be viewed as a lower bound of the log-likelihood $\log p(\mathbf{Y};\boldsymbol{\theta})$. The maximization of the log-likelihood involves iterations involving two steps: i) the maximization of the lower bound $F(q(\mathbf{Z}),\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$; ii) updating $q(\mathbf{Z})$ so that the approximation of $\log p(\mathbf{Y};\boldsymbol{\theta})$ by $F(q(\mathbf{Z}),\boldsymbol{\theta})$ is tight.

For the JSM-1 DCS setting, we can assume $q(\mathbf{Z})$ has a factorized form:

$$q(\mathbf{Z}) = q(\mathbf{z}_c)q(\mathbf{z}_1)\ldots q(\mathbf{z}_K), \quad (19)$$

which is a common assumption in variational approximation. By applying similar techniques to that used in [28], to yield a tight approximation of $\log p(\mathbf{Y};\boldsymbol{\theta})$ by $F(q(\mathbf{Z}),\boldsymbol{\theta})$, we derive the following variational distributions:

$$\begin{aligned} q(\mathbf{z}_c) &\propto \exp\big(\mathbb{E}_{q(\mathbf{z}_1),\ldots,q(\mathbf{z}_K)}\big[\ln p(\mathbf{y}_1,\ldots,\mathbf{y}_K, \\ &\qquad \mathbf{z}_c,\mathbf{z}_1,\ldots,\mathbf{z}_K;\boldsymbol{\Gamma}_c,\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_K,\sigma^2)\big]\big) \\ &\propto \exp\big(\mathbb{E}_{q(\mathbf{z}_1)}\big[\ln p(\mathbf{y}_1|\mathbf{z}_c,\mathbf{z}_1,\sigma^2)\big] + \ldots \\ &\qquad +\mathbb{E}_{q(\mathbf{z}_K)}\big[\ln p(\mathbf{y}_K|\mathbf{z}_c,\mathbf{z}_K,\sigma^2)\big] + \ln p(\mathbf{z}_c|\boldsymbol{\Gamma}_c)\big) \\ &\propto \mathcal{N}(\mathbf{z}_c;\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c), \end{aligned} \quad (20)$$

where $\boldsymbol{\mu}_c = \sigma^{-2}\boldsymbol{\Sigma}_c \sum_{k=1}^{K} \mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_k)$, $\boldsymbol{\Sigma}_c = \left(\sum_{k=1}^{K} \frac{\mathbf{A}_k^T\mathbf{A}_k}{\sigma^2} + \boldsymbol{\Gamma}_c^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_k = \mathbb{E}_{q(\mathbf{z}_1)}\big[\mathbf{z}_k\big]$, and

$$\begin{aligned} q(\mathbf{z}_k) &\propto \exp\big(\mathbb{E}_{q(\mathbf{z}_c),q(\mathbf{z}_j),j\neq k}\big[\ln p(\mathbf{y}_1,\ldots,\mathbf{y}_K, \\ &\qquad \mathbf{z}_c,\mathbf{z}_1,\ldots,\mathbf{z}_K;\boldsymbol{\Gamma}_c,\boldsymbol{\Gamma}_1,\ldots,\boldsymbol{\Gamma}_K,\sigma^2)\big]\big) \\ &\propto \exp\big(\mathbb{E}_{q(\mathbf{z}_c)}\big[\ln p(\mathbf{y}_k|\mathbf{z}_c,\mathbf{z}_k,\sigma^2)\big] + \ln p(\mathbf{z}_k|\boldsymbol{\Gamma}_k)\big) \\ &\propto \mathcal{N}(\mathbf{z}_k;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k), \end{aligned} \quad (21)$$

where $\boldsymbol{\mu}_k = \sigma^{-2}\boldsymbol{\Sigma}_k\mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_c)$ and $\boldsymbol{\Sigma}_k = \left(\frac{\mathbf{A}_k^T\mathbf{A}_k}{\sigma^2} + \boldsymbol{\Gamma}_k^{-1}\right)^{-1}$.

According to (20) and (21), it can be confirmed that $q(\mathbf{z}_c)$ and $q(\mathbf{z}_k)$ are Gaussian distributions, i.e., $q(\mathbf{z}_c) = \mathcal{N}(\mathbf{z}_c;\boldsymbol{\mu}_c,\boldsymbol{\Sigma}_c)$ and $q(\mathbf{z}_k) = \mathcal{N}(\mathbf{z}_k;\boldsymbol{\mu}_k,\boldsymbol{\Sigma}_k)$ $(k = 1,\ldots,K)$. Now given $q(\mathbf{z}_c)$ and $q(\mathbf{z}_k)$ $(k = 1,\ldots,K)$, the hyperparameters can be updated by $\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}} F(q(\mathbf{Z}),\boldsymbol{\theta})$. Specifically, we have

$$\gamma_{c,i}^{\text{new}} = (\Sigma_c)_{i,i} + \mu_{c,i}^2, \quad (22a)$$

$$\gamma_{k,i}^{\text{new}} = (\Sigma_k)_{i,i} + \mu_{k,i}^2, \quad (22b)$$

$$\begin{aligned} (\sigma^2)^{\text{new}} = \frac{1}{K\sum_{k=1}^{K} m_k}\Bigg(&\sum_{k=1}^{K}\|\mathbf{y}_k - \mathbf{A}_k(\boldsymbol{\mu}_c+\boldsymbol{\mu}_k)\|_2^2 + \\ &(\sigma^2)^{\text{old}}\sum_{k=1}^{K}\sum_{i=1}^{n}\big(1-(\gamma_{k,i}^{\text{old}})^{-1}(\Sigma_k)_{i,i}\big) + \\ &(\sigma^2)^{\text{old}}\sum_{i=1}^{n}\big(1-(\gamma_{c,i}^{\text{old}})^{-1}(\Sigma_c)_{i,i}\big)\Bigg). \end{aligned} \quad (22c)$$

The variational optimization proceeds by iteratively updating (20), (21) and (22) until convergence to stable hyperparameters $\boldsymbol{\theta}$. Finally, we can obtain the reconstructed signal by applying the maximum a posteriori estimation

$$\begin{aligned} \mathbf{x}_k &= \arg\max_{\mathbf{z}_c+\mathbf{z}_k} p(\mathbf{Z}|\mathbf{Y};\boldsymbol{\theta}) \\ &= \arg\max_{\mathbf{z}_c} q(\mathbf{z}_c) + \arg\max_{\mathbf{z}_k} q(\mathbf{z}_k) \\ &= \boldsymbol{\mu}_c + \boldsymbol{\mu}_k. \end{aligned} \quad (23)$$

The proposed variational Bayesian algorithm for solving the JSM-1 DCS problem can be summarized by the following steps:

1) Initialize $\boldsymbol{\Gamma}_c$ and $\boldsymbol{\Gamma}_k$ $(k = 1,\ldots,K)$ by identity matrices, and $\boldsymbol{\mu}_k$ $(k = 1,\ldots,K)$ by vectors composed of zero entries;
2) Compute $\boldsymbol{\Sigma}_c = \left(\sum_{k=1}^{K}\frac{\mathbf{A}_k^T\mathbf{A}_k}{\sigma^2} + \boldsymbol{\Gamma}_c^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_c = \sigma^{-2}\boldsymbol{\Sigma}_c\sum_{k=1}^{K}\mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_k)$, i.e., the variational distribution for the common component;
3) Compute $\boldsymbol{\Sigma}_k = \left(\frac{\mathbf{A}_k^T\mathbf{A}_k}{\sigma^2} + \boldsymbol{\Gamma}_k^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_k = \sigma^{-2}\boldsymbol{\Sigma}_k\mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_c)$ for $k = 1,\ldots,K$, i.e., the variational distributions for the innovation components;
4) Update the hyperparameters as in (22);
5) Iterate steps 2, 3 and 4 until convergence occurs to fixed hyperparameters;
6) Output $\mathbf{x}_k = \boldsymbol{\mu}_c + \boldsymbol{\mu}_k$ for $k = 1,\ldots,K$.

Note that although the proposed variational Bayesian algorithm operates in a centralized manner, it facilitates our design of a distributed algorithm that we will present in Section IV.

B. Analysis

1) Comparison with the Fréchet mean approach: The proposed variational Bayesian algorithm for JSM-1 DCS is derived directly from a Bayesian perspective, however it exhibits some similarities to the Fréchet mean approach [27] in the estimation of the common component. Specifically, in each iteration of the proposed algorithm, the mean of the common component is updated by

$$\boldsymbol{\mu}_c = \left(\sum_{k=1}^{K}\mathbf{A}_k^T\mathbf{A}_k + \sigma^2\boldsymbol{\Gamma}_c^{-1}\right)^{-1}\sum_{k=1}^{K}\mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_k), \quad (24)$$

while the Fréchet mean approach using equal weights and the Euclidean distance function gives a crude estimate of the common component as

$$\tilde{\mathbf{z}}_c = \left(\sum_{k=1}^{K} \mathbf{A}_k^T \mathbf{A}_k\right)^{-1} \sum_{k=1}^{K} \mathbf{A}_k^T \mathbf{y}_k. \tag{25}$$

Comparing (24) (25), we note that the Fréchet mean approach employs least squares estimation and ignores the impact of innovation components, while the proposed approach essentially applies minimum mean square error estimation with previous estimate of innovation components.

Given the estimated mean and covariance of the common component, the innovation components are updated separately in the proposed algorithm, which is similar to the process used by the sparse Bayesian learning and the Fréchet mean approach.

2) An Iterative Reweighting Algorithm for DCS: While replacing the ℓ_0 norm with the convex ℓ_1 norm is well justified for sparse signal recovery, recent research has been able to show great advantages from the use of the iterative reweighting algorithms for sparse signal processing [25], [29]. For conventional CS reconstruction of the kth SN's signal, the iterative reweighted ℓ_2 minimization algorithm computes

$$\begin{aligned} \mathbf{x}_k &= \arg\min_{\mathbf{x}_k} \|\mathbf{y}_k - \mathbf{A}_k\mathbf{x}_k\|_2^2 + \nu \mathbf{x}_k^T \mathbf{W}_k \mathbf{x}_k \\ &= \left(\mathbf{A}_k^T\mathbf{A}_k + \nu\mathbf{W}_k\right)^{-1} \mathbf{A}_k^T \mathbf{y}_k \end{aligned} \tag{26}$$

in each iteration, where $\mathbf{W}_k$ is a diagonal weighting matrix and ν is a tradeoff parameter balancing the two terms. Then the weighting matrix $\mathbf{W}_k$ is updated based on $\mathbf{x}_k$ in order to achieve a more accurate estimate of $\mathbf{x}_k$ in the next iteration. The motivation behind this approach relates to the fact that the term $\mathbf{x}_k^T\mathbf{W}_k\mathbf{x}_k$ in (26) is a better approximation to the ℓ_0 norm than the ℓ_1 norm in (4) because $\mathbf{x}_k^T\mathbf{W}_k\mathbf{x}_k \approx \|\mathbf{x}_k\|_0$ for a good weighting matrix $\mathbf{W}_k$.

In the previous subsection, the proposed algorithm emerges from a Bayesian model and a variational Bayesian Inference for JSM-1 DCS. However, based on its update procedure, it can also be seen as an extension of the iterative reweighted ℓ_2 minimization algorithm from the CS case to the JSM-1 DCS case. In each iteration of the proposed algorithm, we update the common component and innovation components separately, that are given by

$$\begin{aligned} \mathbf{z}_c &= \arg\min_{\mathbf{z}_c} \sum_{k=1}^{K} \|\mathbf{y}_k - \mathbf{A}_k\mathbf{z}_k - \mathbf{A}_k\mathbf{z}_c\|_2^2 + \sigma^2 \mathbf{z}_c^T \mathbf{\Gamma}_c^{-1} \mathbf{z}_c \\ &= \left(\sum_{k=1}^{K} \mathbf{A}_k^T\mathbf{A}_k + \sigma^2\mathbf{\Gamma}_c^{-1}\right)^{-1} \sum_{k=1}^{K} \mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{z}_k), \end{aligned} \tag{27}$$

and

$$\begin{aligned} \mathbf{z}_k &= \arg\min_{\mathbf{z}_k} \|\mathbf{y}_k - \mathbf{A}_k\mathbf{z}_c - \mathbf{A}_k\mathbf{z}_k\|_2^2 + \sigma^2 \mathbf{z}_k^T \mathbf{\Gamma}_k^{-1} \mathbf{z}_k \\ &= \left(\mathbf{A}_k^T\mathbf{A}_k + \sigma^2\mathbf{\Gamma}_k^{-1}\right)^{-1} \mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\mathbf{z}_c) \end{aligned} \tag{28}$$

for $k = 1, \ldots, K$. Here, $\mathbf{\Gamma}_c$ and $\mathbf{\Gamma}_k$ $(k = 1, \ldots, K)$, which can be seen as weighting matrices, are then updated based on the value of $\mathbf{z}_c$ and $\mathbf{z}_k$ $(k = 1, \ldots, K)$, as given in (22).

3) Convergence Analysis: As the variational Bayesian approach is essentially an EM update, the proposed algorithm is guaranteed to converge [28], i.e., each iteration is guaranteed to increase the log-likelihood $p(\mathbf{Y};\boldsymbol{\theta})$ until a fixed point is reached. By comparing the update rule of the proposed algorithm and the EM update rule for the SBL [22], it is observed that the solution $\mathbf{\Gamma}_k$ of the proposed algorithm is a minima of the following cost function:

$$\begin{aligned} L(\mathbf{\Gamma}_k) = &(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_c)^T \left(\sigma^2\mathbf{I}_{m_k} + \mathbf{A}_k\mathbf{\Gamma}_k\mathbf{A}_k^T\right)^{-1} (\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_c) \\ &+ \log|\sigma^2\mathbf{I}_{m_k} + \mathbf{A}_k\mathbf{\Gamma}_k\mathbf{A}_k^T|, \end{aligned} \tag{29}$$

and the solution $\mathbf{\Gamma}_c$ is a minima of the following cost function:

$$\begin{aligned} \tilde{L}(\mathbf{\Gamma}_c) = &\tilde{\mathbf{b}}^T \left(\sigma^2\mathbf{I}_{\sum_{k=1}^K m_k} + \tilde{\mathbf{A}}\mathbf{\Gamma}_c\tilde{\mathbf{A}}^T\right)^{-1} \tilde{\mathbf{b}} + \\ &\log|\sigma^2\mathbf{I}_{\sum_{k=1}^K m_k} + \tilde{\mathbf{A}}\mathbf{\Gamma}_c\tilde{\mathbf{A}}^T|, \end{aligned} \tag{30}$$

where $\tilde{\mathbf{A}} = \left[\mathbf{A}_1^T, \cdots, \mathbf{A}_K^T\right]^T$ and $\tilde{\mathbf{b}} = \left[\mathbf{y}_1^T - \boldsymbol{\mu}_1^T\mathbf{A}_1^T, \cdots, \mathbf{y}_K^T - \boldsymbol{\mu}_K^T\mathbf{A}_K^T\right]^T$. Therefore, the proposed framework involves a multi-objective optimization problem and the multiple cost functions are linked via constraints $\boldsymbol{\mu}_c = \sigma^{-2}\mathbf{\Sigma}_c \sum_{k=1}^{K} \mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_k)$ and $\boldsymbol{\mu}_k = \sigma^{-2}\mathbf{\Sigma}_k\mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_c)$ $(k = 1, \ldots, K)$.

We have the following result on the global minimum of the cost functions (29) and (30).

Theorem 1: In the limit as $\sigma^2 \to 0$, assuming $s_c + s_k$, i.e., the sparsity level of the maximally sparse solution $\hat{\mathbf{x}}_k = \hat{\mathbf{z}}_c + \hat{\mathbf{z}}_k$ to $\mathbf{y}_k = \mathbf{A}_k\mathbf{x}_k$, satisfies $s_c < \sum_{k=1}^K m_k$ and $s_k < m_k$ for $\forall k$, then $\{\hat{\mathbf{\Gamma}}_k\}$ and $\hat{\mathbf{\Gamma}}_c$, i.e., the global minima of (29) and (30), respectively, lead to a source estimate that equals $\{\hat{\mathbf{z}}_c, \hat{\mathbf{z}}_1, \ldots, \hat{\mathbf{z}}_K\}$.

Proof: See Appendix A. ∎

This theorem ensures the global minimum of our algorithm is achieved at the most sparse signal representation in the noiseless case. The global minimum property guarantees structural correctness, i.e., the proposed algorithm converges to a minimum (possibly global) with a cost function value no smaller than the value of the most sparse one, while the LASSO, a widely used convex optimization algorithm for sparse signal reconstruction, does not have a guarantee on the structural correctness.

Now we discuss the local minimum property. Similarly to the SBL, the cost functions of the proposed algorithm can potentially have many local minima. For the cost function in (29), we have the following result, which ensures all local minima $\mathbf{\Gamma}_k$ of our algorithm are sparse.

Theorem 2: Every local minimum of the cost function in (29) with respect to $\mathbf{\Gamma}_k$ is achieved at a solution with $\|\texttt{diagv}(\mathbf{\Gamma}_k)\|_0 \leq m_k$, regardless of the values of σ^2 and $\mathbf{\Gamma}_c$.

Proof: See Appendix B. ∎

For the cost function in (30), if $\sum_{k=1}^K m_k < n$, akin to Theorem 2 and using the same techniques in Appendix B, it can be proved that all local minima $\mathbf{\Gamma}_c$ are sparse. If $\sum_{k=1}^K m_k \geq n$, solving $\mathbf{\Gamma}_c$ by minimizing (30) is equivalent to obtaining sparse solutions in regression problems via the

relevance vector machine (RVM) [21] where the size of the training data is larger than the number of variables, and thus our algorithm has the same capability as the classic RVM to converge to a highly sparse $\mathbf{\Gamma}_c$.

We note that although the global minima of the cost functions in (29) and (30) is equivalent to the global minima of independently solving a sparsity maximization problem for each task, the entire cost function landscapes are not identical. We observe that our approach exploits the inter-signal structure, which could be advantageous in avoiding distracting local minima.

IV. A Decentralized Bayesian Algorithm for DCS

In this section, we propose a decentralized Bayesian algorithm for JSM-1 DCS, which exploits the variational Bayesian inference developed previously. By casting the decoupled reconstruction problem as a set of decentralized problems with consensus constraints, the variational Bayesian inference is carried out in a decentralized way without sharing sensitive information with respect to the innovation component.

For a centralized scenario, both steps 2) and 3) of the variational Bayesian inference is carried out at a FC, which collects all the nodes' measurements and performs the computation. However, for a decentralized scenario, we assume all the computation should be performed at the nodes and that each node has no knowledge of other nodes' sensing matrices and measurements. In view of the fact that the computation of innovation components (step 3) are decoupled from the common component in the variational SBL algorithm, the nodes can work in parallel to execute step 3) and update (22b). Therefore, we now only need to decentralize the computation of the common component in step 2) and the update of hyperparameters in (22a) and (22c).

According to the definition of $\mathbf{\Sigma}_c$ and $\boldsymbol{\mu}_c$, we have

$$\frac{\sigma^2}{K}\left(\mathbf{\Sigma}_c^{-1}-\mathbf{\Gamma}_c^{-1}\right)=\frac{1}{K}\sum_{k=1}^{K}\mathbf{A}_k^T\mathbf{A}_k \tag{31}$$

and

$$\frac{\sigma^2}{K}\mathbf{\Sigma}_c^{-1}\boldsymbol{\mu}_c=\frac{1}{K}\sum_{k=1}^{K}\mathbf{A}_k^T(\mathbf{y}_k-\mathbf{A}_k\boldsymbol{\mu}_k), \tag{32}$$

which are the average of $\mathbf{A}_k^T\mathbf{A}_k$ and the average of $\mathbf{A}_k^T(\mathbf{y}_k-\mathbf{A}_k\boldsymbol{\mu}_k)$ $(k=1,\ldots,K)$, respectively. The two averages can be obtained by solving the following couple of average consensus problems

$$\min_{\mathbf{W}}\quad\sum_{k=1}^{K}\left\|\mathbf{W}-\mathbf{A}_k^T\mathbf{A}_k\right\|_F^2, \tag{33}$$

and

$$\min_{\mathbf{r}}\quad\sum_{k=1}^{K}\left\|\mathbf{r}-\mathbf{A}_k^T(\mathbf{y}_k-\mathbf{A}_k\boldsymbol{\mu}_k)\right\|_2^2, \tag{34}$$

respectively.

The optimization problems in (33) can be reformulated into

$$\begin{aligned}&\min_{\mathbf{W}^1,\ldots,\mathbf{W}^K}\sum_{k=1}^{K}\left\|\mathbf{W}^k-\mathbf{A}_k^T\mathbf{A}_k\right\|_F^2\\&\text{s.t.}\quad\mathbf{W}^k=\mathbf{W}^{j_k},\ \forall j_k\in\mathcal{N}_k,\ \forall k\in\{1,\ldots,K\},\end{aligned} \tag{35}$$

where $\mathbf{W}^k$ denotes the local estimate of $\mathbf{W}=\frac{1}{K}\sum_{k=1}^{K}\mathbf{A}_k^T\mathbf{A}_k$ at node k, respectively, and $\mathcal{N}_k$ denotes the neighbors of node k. Two nodes are called as neighbors if they can communicate with each other to interchange information. Optimization problems (33) and (35) are equivalent if their neighborhood relationship can lead to a connected graph. Similarly, the optimization problems in (34) can be reformulated into

$$\begin{aligned}&\min_{\mathbf{r}^1,\ldots,\mathbf{r}^K}\sum_{k=1}^{K}\left\|\mathbf{r}^k-\mathbf{A}_k^T(\mathbf{y}_k-\mathbf{A}_k\boldsymbol{\mu}_k)\right\|_2^2\\&\text{s.t.}\quad\mathbf{r}^k=\mathbf{r}^{j_k},\ \forall j_k\in\mathcal{N}_k,\ \forall k\in\{1,\ldots,K\}.\end{aligned} \tag{36}$$

We employ the alternating direction method of multipliers (ADMM) [30] to solve (35) and (36) in a decentralized manner. Note that (35) only needs to be solved for consensus once, while (36) needs to be performed in each iteration of the variational SBL when $\boldsymbol{\mu}_k$ $(k=1,\ldots,K)$ are updated. According to [18], the simplified ADMM form of (35) consists of the following iterations

$$\begin{aligned}(\mathbf{P}_w^k)^{\text{new}}&=(\mathbf{P}_w^k)^{\text{old}}+\rho\sum_{j_k\in\mathcal{N}_k}\left((\mathbf{W}^k)^{\text{old}}-(\mathbf{W}^{j_k})^{\text{old}}\right),\\(\mathbf{W}^k)^{\text{new}}&=\frac{1}{2+2\rho|\mathcal{N}_k|}\Bigg(2\mathbf{A}_k^T\mathbf{A}_k-(\mathbf{P}_w^k)^{\text{new}}+\\&\qquad\rho\sum_{j_k\in\mathcal{N}_k}\left((\mathbf{W}^k)^{\text{old}}+(\mathbf{W}^{j_k})^{\text{old}}\right)\Bigg)\end{aligned} \tag{37}$$

for $\forall k\in\{1,\ldots,K\}$, where $\rho>0$ is a preselected penalty coefficient. Note that nodes can execute (37) in parallel with the information concerning $\mathbf{W}^{j_k}$ passed from their neighbors. In addition, it has been proved that iteratively executing the steps in (37) will converge to the global solution $\mathbf{W}$ for any $\rho>0$ [18]. Similarly, the ADMM form of (36) consists of the following iterations

$$\begin{aligned}(\mathbf{p}_r^k)^{\text{new}}&=(\mathbf{p}_r^k)^{\text{old}}+\rho\sum_{j_k\in\mathcal{N}_k}\left((\mathbf{r}^k)^{\text{old}}-(\mathbf{r}^{j_k})^{\text{old}}\right),\\(\mathbf{r}^k)^{\text{new}}&=\frac{1}{2+2\rho|\mathcal{N}_k|}\Bigg(2\mathbf{A}_k^T(\mathbf{y}_k-\mathbf{A}_k\boldsymbol{\mu}_k)-(\mathbf{p}_r^k)^{\text{new}}+\\&\qquad\rho\sum_{j_k\in\mathcal{N}_k}\left((\mathbf{r}^k)^{\text{old}}+(\mathbf{r}^{j_k})^{\text{old}}\right)\Bigg).\end{aligned} \tag{38}$$

Given $\mathbf{W}_k$ and $\mathbf{r}_k$, according to (31) and (32), each node can execute step 2 of the variational SBL by computing

$$\mathbf{\Sigma}_c=\left(\frac{K}{\sigma^2}\mathbf{W}_k+\mathbf{\Gamma}_c^{-1}\right)^{-1} \tag{39}$$

and

$$\boldsymbol{\mu}_c=\sigma^{-2}\mathbf{\Sigma}_c\mathbf{r}_k, \tag{40}$$

Algorithm 1 A Decentralized Bayesian Algorithm for DCS

Input: A set of signals $\{\mathbf{x}_i\}$ $(i = 1, \ldots, L)$, a dictionary $\mathbf{\Psi}$ and a positive value β.

Output: The common component $\mathbf{z}_c$ and the innovation components $\mathbf{z}_i$ $(i = 1, \ldots, L)$.

Process: Do

1) Initialize $\mathbf{\Gamma}_c$ and $\mathbf{\Gamma}_k$ $(k = 1, \ldots, K)$ by identity matrices, and $\boldsymbol{\mu}_k$ $(k = 1, \ldots, K)$ by $\mathbf{0}$;
2) Initialize $\mathbf{P}_w^k$, and $\mathbf{W}_k$ $(k = 1, \ldots, K)$ by $\mathbf{0}$, and iteratively compute (37) in parallel at each node until a predefined stopping criterion is satisfied;
3) Initialize $\mathbf{p}_r^k$ and $\mathbf{r}_k$ $(k = 1, \ldots, K)$ by $\mathbf{0}$, and iteratively compute (38) in parallel at each node until a predefined stopping criterion is satisfied;
4) Compute (39) and (40) at each node;
5) Compute $\mathbf{\Sigma}_k = \left(\frac{\mathbf{A}_k^T \mathbf{A}_k}{\sigma^2} + \mathbf{\Gamma}_k^{-1}\right)^{-1}$ and $\boldsymbol{\mu}_k = \sigma^{-2}\mathbf{\Sigma}_k \mathbf{A}_k^T(\mathbf{y}_k - \mathbf{A}_k \boldsymbol{\mu}_c)$ in parallel at each node;
6) Compute (22a) and (22b) in parallel at each node;
7) Compute (22c) in a distributed manner until a predefined stopping criterion is satisfied;
8) If halting condition is true, return $\mathbf{z}_c$ and $\mathbf{z}_i$ $(i = 1, \ldots, L)$; otherwise go to step 3;

which directly enable the update of the hyperparameter in (22a) locally at each node. The hyperparameter in (22c) can also be computed in a distributed manner using the same ADMM technique. The pseudo-code of the proposed decentralized variational SBL algorithm is given in Algorithm 1.

V. Numerical Simulations

In this section, we compare the performance of the proposed decentralized Baysian algorithm for DCS reconstruction with other existing approaches by experiments with synthetic signals and real temperature signals.

The following approaches are compared:

1) Decentralized Proposal: signals are reconstructed at each node in a decentralized manner by the proposed algorithm, which exploits both intra- and inter-signal correlations;
2) Independent SBL: signals are reconstructed independently at each node by SBL, which only exploits the intra-signal sparse structure;
3) Centralized SBL: signals are jointly reconstructed at a FC by SBL, and both intra- and inter-signal correlations are exploited;
4) Centralized Fréchet mean approach: Joint signal reconstruction by the Fréchet mean approach [27] at a FC, where both intra- and inter-signal correlations are exploited.

For the Fréchet mean approach, we use CVX, a package for specifying and solving convex programs [31]. Note that the independent SBL considers the CS setting and acts as a baseline, while the other three approaches consider the JSM-1 DCS setting.

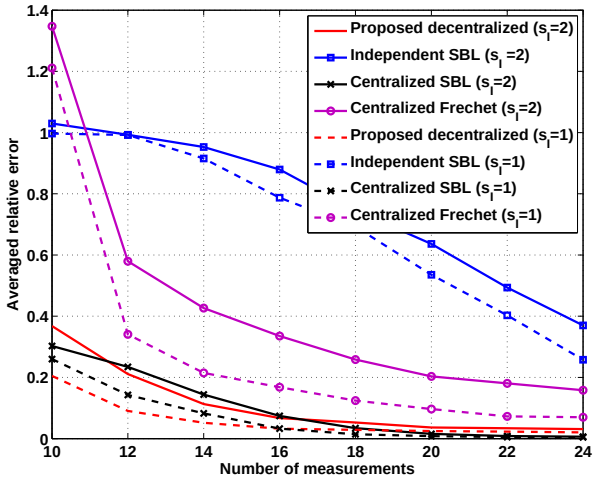

Fig. 2. Reconstruction quality vs. number of measurements ($n = 50$, $K = 10$, $s_c = 8$ and Harary graph model with $L = 3$).

In the comparison, the reconstruction quality is measured by averaged relative error, which is defined as the average of $\frac{\sum_{k=1}^{K} \|\hat{\mathbf{x}}_k - \mathbf{x}_k\|_2^2}{\sum_{k=1}^{K} \|\mathbf{x}_k\|_2^2}$, where $\hat{\mathbf{x}}_k$ denotes the reconstructed signal k. We conduct 100 trials for each experiment setting and provide the averaged result.

In addition, two different networks are considered. The first network is an L-connected Harary graph, where each node is only available to communicate with L adjacent neighbors to exchange information. For the other network, the Erdös-Rényi model [32] is applied to generate the neighborhood relationship, where the probability of any two nodes being connected is p. In the proposed decentralized algorithm, parameter ρ of the ADMM step is set to 0.3 in our simulations.

A. Experiments With Synthetic Data

We consider a set of K correlated signals following the JSM-1 DCS model. Without loss of generality, we let $m = m_k$ $(k = 1, \ldots, K)$, i.e., all signals have the same number of measurements, and $s_I = s_k$ $(k = 1, \ldots, K)$, i.e., the innovation components of different signals have the same sparsity level. We first generate the sparse common component $\mathbf{z}_c$ randomly for all signals and then generate the sparse innovation component $\mathbf{z}_k$ $(k = 1, \ldots, K)$ randomly for each of the signals independently, where the non-zero components of both $\mathbf{z}_c$ and $\mathbf{z}_k$ are drawn from i.i.d. Gaussian distributions $\mathcal{N}(0, 1)$. The sensing matrices $\mathbf{A}_k$ are generated randomly for different signals, where the elements are drawn from the i.i.d. Gaussian distribution $\mathcal{N}(0, 1)$, followed by a column normalization. The received measurements are corrupted by additive zero-mean Gaussian noise to yield signal noise ratio (SNR), i.e., $\frac{\|\mathbf{A}_k \mathbf{x}_k\|_2^2}{\|\mathbf{e}_k\|_2^2}$, of 20dB.

The reconstruction quality for different approaches is given in Fig. 2 and 3, where we have compared the averaged relative error against the number of measurements and the innovation component sparsity level, respectively. Our numerical simulation results confirm that joint signal reconstruction

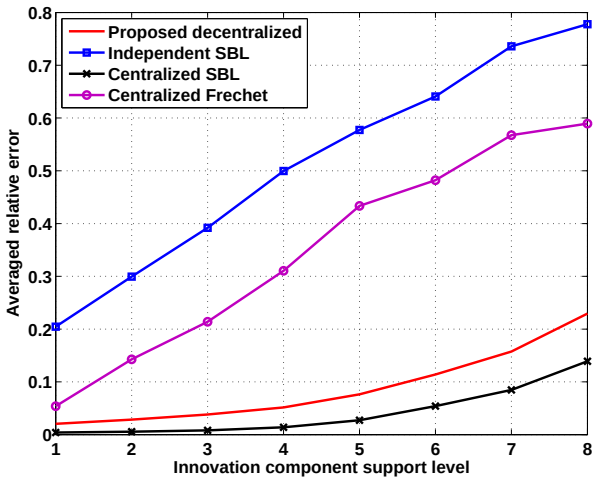

Fig. 3. Reconstruction quality vs. innovation component sparsity level ($n = 50$, $m = 25$, $K = 10$, $s_c = 8$ and Harary graph model with $L = 3$).

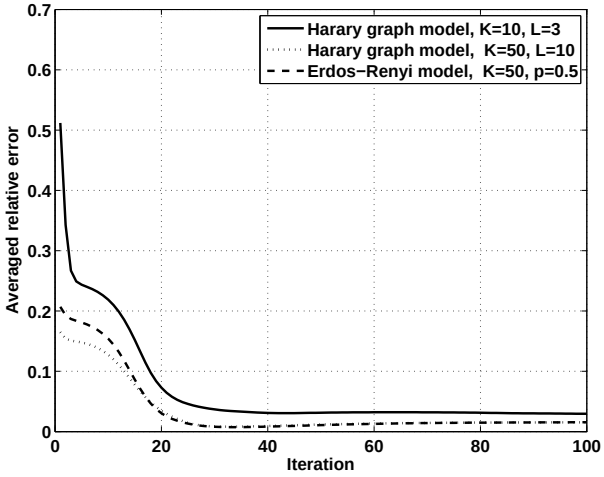

Fig. 4. Convergence rate of the outer loop (variational Bayesian inference loop) of proposed algorithm for a single instance. ($n = 50$, $m = 25$, $s_c = 8$ and $s_I = 2$).

algorithms have better performance than independent signal reconstruction, i.e., independent SBL that neglects inter-signal correlation. The proposed decentralized approach outperforms the centralized Fréchet mean approach, which can be explained by our analysis regarding the comparison between the two approaches, i.e., the Fréchet mean approach ignores the impact of innovation components in the estimation of common component. In Fig. 3, it is noted that the reconstruction quality of the proposed decentralized approach is slightly degraded in comparison with the centralized SBL. This behaviour is caused by the fact that the proposed decentralized algorithm employs variational Bayesian approximations to facilitate decentralized computation, while the centralized SBL employs EM updates. However, the decentralized approach avoids sharing sensitive or private local data and is more robust to attacks, since if the FC is compromised, the entire centralized system will fail. More performance comparisons with various settings are given in Table I, which further confirms our observations.

Fig. 4 illustrates the convergence rates of the proposed algorithm with different network settings for a single instance, and Fig. 5 shows the convergence of the inner consensus loop, i.e., iteratively solving optimization problem (35) by ADMM. It is observed that for all the three different network settings, both the outer loop and the inner loop of the algorithm exhibit reasonably fast convergence. In particular, the algorithm has converged after 30 outer loop iterations with 20 inner loop consensus iterations. In addition, the communication costs of the consensus loop in a single instance for different network settings are provided in Table II, where the numeric value indicates the number of transmissions of a single node. Considering the communication cost, the proposed approach potentially can be beneficial for the following two scenarios: i) the Nyquist rate of the signal exceeds the current capabilities of analog-to-digital converters (ADCs) so that the main bottleneck of the system is in the sampling rate and the communication cost introduced by the proposed distributed algorithm can be tolerated; ii) the cost of sampling is much higher than

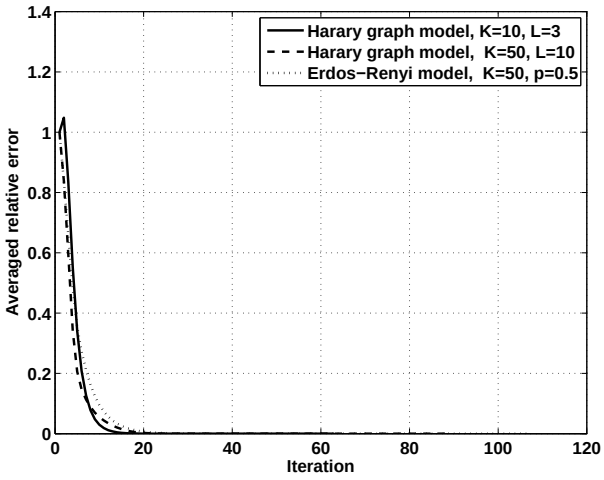

Fig. 5. Convergence rate of inner consensus loop (ADMM loop) in the proposed algorithm for a single instance.

communication. For example, vibrating wire strain-gauges that are used in civil engineering applications, and NDIR or metal oxide sensors that are used to measure the concentration of a gas or gasses of interest in the atmosphere, consume energy several orders of magnitude greater than that required for communication. For instance, it is indicated in [33] that metal oxide based sensors for measuring ozone concentration typically consume in excess of 90 mW.

B. Experiments With Real Data

We now investigate the effectiveness of the proposed decentralized algorithm with real signals which could be nearly sparse rather than exactly sparse as in the synthetic data experiment. To do this, we use the temperature signals obtained the Intel Berkeley Research lab [34]. In the following evaluations, we use the discrete cosine transform (DCT) as the sparsifying domain. Instead of uniform sampling, we assume each SN

TABLE I
THE COMPARISON OF AVERAGED RELATIVE ERROR WITH DIFFERENT SETTINGS. ($n = 100$, $s_c = 8$ AND $s_I = 2$)

	$K = 20$ $m = 10$	$K = 20$ $m = 15$	$K = 20$ $m = 20$	$K = 50$ $m = 10$	$K = 50$ $m = 15$	$K = 50$ $m = 20$
Independent SBL	0.8925	0.8046	0.6182	0.8746	0.8017	0.6006
Centralized SBL	0.1862	0.0556	0.0100	0.1921	0.0506	0.0072
Centralized Fréchet mean approach	0.5635	0.3385	0.2253	0.5242	0.3470	0.2078
Proposed decentralized algorithm, Harary graph model (L=5)	0.2185	0.0663	0.0212	0.2153	0.0517	0.0152
Proposed decentralized algorithm, Erdös-Rényi model (p=0.25)	0.1815	0.0467	0.0194	0.2037	0.0491	0.0148

TABLE II
THE COMMUNICATION COST OF THE CONSENSUS LOOP (ADMM LOOP) IN THE PROPOSED ALGORITHM. ($n = 50$, $s_c = 8$ AND $s_I = 2$)

	$K = 20$ $m = 15$	$K = 20$ $m = 20$	$K = 20$ $m = 25$	$K = 50$ $m = 15$	$K = 50$ $m = 25$
Harary graph model (L=2)	136	130	121	192	168
Harary graph model (L=4),	54	43	46	159	130
Erdös-Rényi model (p=0.3)	18	23	16	28	28
Erdös-Rényi model (p=0.5)	20	20	20	39	41

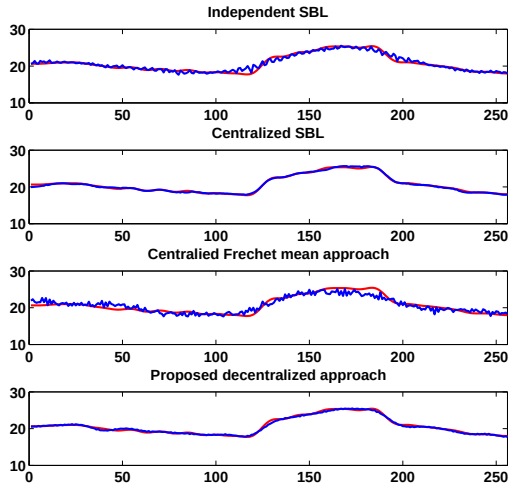

Fig. 6. Environmental temperature signal detected by one node and reconstruction results. ($m = 50$, $K = 10$ and Harary graph model with $L = 5$).

independently and randomly collects a small portion of the original samples. Therefore, the sensing matrices $\mathbf{A}_k$ become random partial DCT matrices. All the signals we employ in the following study have a length of $n = 256$.

Fig. 6 demonstrates the effectiveness of the proposed approach by comparing the original temperature signal with the reconstructed results for one of the nodes. In this experiment, the temperature signals from $K = 10$ nodes are reconstructed by different algorithms, and one of these signals and its reconstruction results are shown in the figure. It is observed that both the proposed decentralized Bayesian algorithm and the centralized SBL successfully recover the temperature signals with only 50 measurements, while the independent SBL and the centralized Fréchet mean approach have significant visible errors in the reconstructed results. We report that the reconstructed temperature signals of other nodes show a similar result. For the centralized SBL, all the nodes' samples needs to be gathered at a FC for joint reconstruction, while the proposed decentralized algorithm enables joint reconstruction without a FC and in addition the reconstructed signal at a particular node is only available to that node itself. Table III gives more performance comparisons with different settings, and similar trends are observed.

VI. CONCLUSION

In this paper, we propose a decentralized Bayesian DCS algorithm to efficiently reconstruct multiple signals in a networked sensing system. Both the intra- and inter-signal correlations are exploited by the proposed approach with the JSM-1 DCS model, and thus it possess advantages beyond conventional independent CS algorithms, that neglect inter-signal correlations. The decentralized characteristics of the proposed algorithm make it suitable for applications needing enhanced privacy and for those that require fusion-centre-free operation. Experimental results demonstrate good recovery performance and convergence properties of the proposed decentralized algorithm.

APPENDIX A
PROOF OF THE THEOREM 1

As the proposed variational sparse Bayesian framework extends the SBL framework to the JSM-1 DCS model, many of the following proofs are based on the theoretic work in [22]. However, some essential modifications are required.

According to the formulations of the cost functions in (29) and (30), the minimum occurs when

$$|\sigma^2\mathbf{I}_{m_k} + \mathbf{A}_k\mathbf{\Gamma}_k\mathbf{A}_k^T| = 0, \quad |\sigma^2\mathbf{I}_{\sum_{k=1}^{K} m_k} + \tilde{\mathbf{A}}\mathbf{\Gamma}_c\tilde{\mathbf{A}}^T| = 0, \tag{41}$$

TABLE III
THE AVERAGED RELATIVE ERROR OF THE RECONSTRUCTED ENVIRONMENTAL TEMPERATURE SIGNAL FOR DIFFERENT APPROACHES.

	$m=50$ $K=10$	$m=40$ $K=10$	$m=50$ $K=6$	$m=40$ $K=6$	$m=50$ $K=15$
Independent SBL	$1.4*10^{-3}$	$3.6*10^{-3}$	$6.2*10^{-4}$	$1.4*10^{-3}$	$2.3*10^{-3}$
Centralized SBL	$7.4*10^{-4}$	$9.8*10^{-4}$	$1.2*10^{-4}$	$1.9*10^{-4}$	$8.7*10^{-4}$
Centralized Fréchet mean approach	$1.2*10^{-3}$	$3.4*10^{-3}$	$1.1*10^{-3}$	$4.4*10^{-3}$	$1.1*10^{-3}$
Proposed decentralized algorithm, Harary graph model (L=5)	$1.1*10^{-3}$	$2.4*10^{-4}$	$4.6*10^{-4}$	$7.8*10^{-4}$	$1.9*10^{-3}$
Proposed decentralized algorithm, Erdös-Rényi model (p=0.5)	$1.0*10^{-3}$	$2.1*10^{-3}$	$5.1*10^{-4}$	$1.2*10^{-3}$	$1.6*10^{-3}$

and

$$\begin{aligned}&\left(\mathbf{y}_k-\mathbf{A}_k\mathbf{z}_c\right)^T\left(\sigma^2\mathbf{I}_{m_k}+\mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^T\right)^{-1}\left(\mathbf{y}_k-\mathbf{A}_k\mathbf{z}_c\right)\le\rho\\&\tilde{\mathbf{b}}^T\left(\sigma^2\mathbf{I}_{\sum_{k=1}^K m_k}+\tilde{\mathbf{A}}\boldsymbol{\Gamma}_c\tilde{\mathbf{A}}^T\right)^{-1}\tilde{\mathbf{b}}\le\rho,\end{aligned}\tag{42}$$

for $k=1,\ldots,K$, where $\rho>0$ denotes some finite bound. Now, all that is required is to prove that the solutions, which lead to accurate signal reconstruction, satisfy these conditions.

When $\sigma^2=0$, following results from linear algebra as given in [22], the solutions of the cost functions (29) and (30) lead to the estimate such that

$$\hat{\mathbf{z}}_k=\hat{\boldsymbol{\Gamma}}_k^{1/2}(\mathbf{A}\hat{\boldsymbol{\Gamma}}_k^{1/2})^{\dagger}(\mathbf{y}_k-\mathbf{A}\hat{\mathbf{z}}_c),\tag{43}$$

and

$$\hat{\mathbf{z}}_c=\frac{1}{K}\hat{\boldsymbol{\Gamma}}_c^{1/2}(\tilde{\mathbf{A}}\hat{\boldsymbol{\Gamma}}_c^{1/2})^{\dagger}\tilde{\mathbf{b}},\tag{44}$$

which suggests that the support of $\hat{\mathbf{z}}_c$ is the same as the support associated with $\hat{\boldsymbol{\Gamma}}_c$, and the support of $\hat{\mathbf{z}}_k$ is the same as the support associated with $\hat{\boldsymbol{\Gamma}}_k$. Since $s_c<\sum_{k=1}^K m_k$ and $s_k<m_k$, the conditions in (41) are satisfied. In addition, we have

$$\begin{aligned}&\lim_{\sigma^2\to0}\left(\mathbf{y}_k-\mathbf{A}\hat{\mathbf{z}}_c\right)^T\left(\sigma^2\mathbf{I}_m+\mathbf{A}\boldsymbol{\Gamma}_k\mathbf{A}^T\right)^{-1}\left(\mathbf{y}_k-\mathbf{A}\hat{\mathbf{z}}_c\right)\\&=\lim_{\sigma^2\to0}\hat{\mathbf{z}}_k^T\hat{\boldsymbol{\Gamma}}_k^{-1/2}\hat{\boldsymbol{\Gamma}}_k^{1/2}\mathbf{A}^T(\sigma^2\mathbf{I}_m+\mathbf{A}\boldsymbol{\Gamma}_k\mathbf{A}^T)^{-1}\mathbf{A}\hat{\boldsymbol{\Gamma}}_k^{1/2}\hat{\boldsymbol{\Gamma}}_k^{-1/2}\hat{\mathbf{z}}_k\\&=\hat{\mathbf{z}}_k^T\hat{\boldsymbol{\Gamma}}_k^{-1}\hat{\mathbf{z}}_k\le\frac{1}{\delta}\|\hat{\mathbf{z}}_k\|_2^2,\end{aligned}\tag{45}$$

where $\delta>0$ is the minimum nonzero entry of $\hat{\boldsymbol{\Gamma}}_k$. Using the same procedures, we can also prove that $\tilde{\mathbf{b}}^T\left(\sigma^2\mathbf{I}_{\sum_{k=1}^K m_k}+\tilde{\mathbf{A}}\boldsymbol{\Gamma}_c\tilde{\mathbf{A}}^T\right)^{-1}\tilde{\mathbf{b}}$ is bounded, which complete the proof.

APPENDIX B
PROOF OF THE THEOREM 2

Now before discussing the local minimum property, we provide two lemmas, which are given in [22] and are needed in proving our results.

Lemma 1: $\log|\sigma^2\mathbf{I}_{m_k}+\mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^T|$ is concave with respect to $\boldsymbol{\Gamma}_k$.

This lemma can be proved by the composition property of concave functions [35].

Lemma 2: Let $\mathbf{r}_k=\mathbf{y}_k-\mathbf{A}_k\boldsymbol{\mu}_c$. Then $\mathbf{r}_k^T\left(\sigma^2\mathbf{I}_{m_k}+\mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^T\right)^{-1}\mathbf{r}_k$ equals a constant c_k when $\boldsymbol{\Gamma}_k$ satisfies the linear constraints

$$\mathbf{b}=\mathbf{G}\mathtt{diagv}(\boldsymbol{\Gamma}_k),\tag{46}$$

with

$$\begin{aligned}\mathbf{b}&\triangleq\mathbf{r}_k-\sigma^2\mathbf{v}_k\\\mathbf{G}&\triangleq\mathbf{A}_k\mathtt{diagm}(\mathbf{A}_k^T\mathbf{v}_k),\end{aligned}\tag{47}$$

where $\mathbf{v}_k$ is any fixed vector such that $\left(\mathbf{y}_k-\mathbf{A}_k\boldsymbol{\mu}_c\right)^T\mathbf{v}_k=c_k$.

This lemma can be proved by rewriting the equation $\mathbf{r}_k^T\left(\sigma^2\mathbf{I}_{m_k}+\mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^T\right)^{-1}\mathbf{r}_k=c_k$ with $\mathbf{r}_k^T\mathbf{v}_k=c_k$, where $\mathbf{v}_k=\left(\sigma^2\mathbf{I}_{m_k}+\mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^T\right)^{-1}\mathbf{r}_k$. Then we have

$$\begin{aligned}\mathbf{r}_k-\sigma^2\mathbf{v}_k&=\mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^T\mathbf{v}_k\\&=\mathbf{A}_k\mathtt{diagm}(\mathbf{A}_k^T\mathbf{v}_k)\mathtt{diagv}(\boldsymbol{\Gamma}_k).\end{aligned}\tag{48}$$

The proof of the Theorem 2 follows along the line of Theorem 2 in [22] with the use of the above two lemmas. Consider the following optimization problem:

$$\begin{aligned}\min_{\boldsymbol{\Gamma}_k}\quad&\log|\sigma^2\mathbf{I}_{m_k}+\mathbf{A}_k\boldsymbol{\Gamma}_k\mathbf{A}_k^T|\\\text{s.t.}\quad&\mathbf{b}=\mathbf{G}\mathtt{diagv}(\boldsymbol{\Gamma}_k)\\&\mathtt{diagv}(\boldsymbol{\Gamma}_k)\ge0,\end{aligned}\tag{49}$$

where $\mathbf{b}$ and $\mathbf{G}$ are defined in Lemma 2. According to Lemma 1 and Lemma 2, the optimization problem (49) optimizes a concave function over a closed, bounded convex polytope. From the Theorem 7.5.3 in [36], the minimum of (49) is achieved at an extreme point. In addition, the Theorem 2.5 in [36] suggests that the extreme point is a basis feasible solution to $\mathbf{b}=\mathbf{G}\mathtt{diagv}(\boldsymbol{\Gamma}_k)$ and $\mathtt{diagv}(\boldsymbol{\Gamma}_k)\ge0$, which indicates $\|\mathtt{diagv}(\boldsymbol{\Gamma}_k)\|_0\le m_k$.

REFERENCES

[1] W. Chen and I. Wassell, "Variational Bayesian algorithm for distributed compressive sensing," in *2015 IEEE International Conference on Communications (ICC)*, accepted.

[2] E. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 489–509, 2006.

[3] D. Donoho, "Compressed sensing," *Information Theory, IEEE Transactions on*, vol. 52, no. 4, pp. 1289–1306, 2006.

[4] Y. Wang, Z. Tian, and C. Feng, "Sparsity order estimation and its application in compressive spectrum sensing for cognitive radios," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 6, pp. 2116–2125, June 2012.

[5] Z. Tian and G. Giannakis, "Compressed sensing for wideband cognitive radios," in *Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on*, vol. 4, April 2007, pp. IV–1357–IV–1360.

[6] K. Gedalyahu and Y. Eldar, "Time-delay estimation from low-rate samples: A union of subspaces approach," *Signal Processing, IEEE Transactions on*, vol. 58, no. 6, pp. 3017–3031, June 2010.

[7] O. Bar-Ilan and Y. Eldar, "Sub-Nyquist radar via doppler focusing," *Signal Processing, IEEE Transactions on*, vol. 62, no. 7, pp. 1796–1811, April 2014.

[8] M. Lustig, D. Donoho, J. Santos, and J. Pauly, "Compressed sensing MRI," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 72–82, March 2008.

[9] M. Mishali, Y. C. Eldar, O. Dounaevsky, and E. Shoshan, "Xampling: Analog to digital at sub-Nyquist rates," *IET circuits, devices & systems*, vol. 5, no. 1, pp. 8–20, 2011.

[10] M. Duarte, M. Davenport, D. Takhar, J. Laska, T. Sun, K. Kelly, and R. Baraniuk, "Single-pixel imaging via compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, no. 2, pp. 83–91, March 2008.

[11] L. Xu, X. Hao, N. D. Lane, X. Liu, and T. Moscibroda, "Cost-aware compressive sensing for networked sensing systems," in *Proceedings of the 14th International Conference on Information Processing in Sensor Networks*. ACM, 2015, pp. 130–141.

[12] G. Quer, R. Masiero, G. Pillonetto, M. Rossi, and M. Zorzi, "Sensing, compression, and recovery for wsns: Sparse signal modeling and monitoring framework," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 10, pp. 3447–3461, October 2012.

[13] M. Duarte and Y. Eldar, "Structured compressed sensing: From theory to applications," *Signal Processing, IEEE Transactions on*, vol. 59, no. 9, pp. 4053–4085, Sept 2011.

[14] M. Davies and Y. Eldar, "Rank awareness in joint sparse recovery," *Information Theory, IEEE Transactions on*, vol. 58, no. 2, pp. 1135–1146, Feb 2012.

[15] D. Zordan, G. Quer, M. Zorzi, and M. Rossi, "Modeling and generation of space-time correlated signals for sensor network fields," in *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*, Dec 2011, pp. 1–6.

[16] D. Baron, M. Wakin, M. Duarte, S. Sarvotham, and R. Baraniuk, "Distributed compressed sensing," *Technical Report ECE-0612, Electrical and Computer Engineering Department, Rice University*, Dec. 2006.

[17] M. Duarte, M. Wakin, D. Baron, S. Sarvotham, and R. Baraniuk, "Measurement bounds for sparse signal ensembles via graphical models," *Information Theory, IEEE Transactions on*, vol. 59, no. 7, pp. 4280–4289, 2013.

[18] G. Mateos, J. Bazerque, and G. Giannakis, "Distributed sparse linear regression," *Signal Processing, IEEE Transactions on*, vol. 58, no. 10, pp. 5262–5276, Oct 2010.

[19] J. Mota, J. Xavier, P. Aguiar, and M. Puschel, "Distributed basis pursuit," *Signal Processing, IEEE Transactions on*, vol. 60, no. 4, pp. 1942–1956, April 2012.

[20] G. Coluccia, C. Ravazzi, and E. Magli, *Compressed Sensing for Distributed Systems*, ser. SpringerBriefs in Signal Processing. Springer, 2015.

[21] M. E. Tipping, "Sparse Bayesian learning and the relevance vector machine," *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[22] D. Wipf and B. Rao, "Sparse Bayesian learning for basis selection," *Signal Processing, IEEE Transactions on*, vol. 52, no. 8, pp. 2153–2164, Aug 2004.

[23] M. Davies and R. Gribonval, "Restricted isometry constants where ℓ^p sparse recovery can fail for $0 < p \leq 1$," *Information Theory, IEEE Transactions on*, vol. 55, no. 5, pp. 2203–2214, May 2009.

[24] R. Wu and D.-R. Chen, "The improved bounds of restricted isometry constant for recovery via ℓ_p-minimization," *Information Theory, IEEE Transactions on*, vol. 59, no. 9, pp. 6142–6147, Sept 2013.

[25] D. Wipf and S. Nagarajan, "Iterative reweighted ℓ_1 and ℓ_2 methods for finding sparse solutions," *Selected Topics in Signal Processing, IEEE Journal of*, vol. 4, no. 2, pp. 317–329, April 2010.

[26] D. Wipf, B. Rao, and S. Nagarajan, "Latent variable Bayesian models for promoting sparsity," *Information Theory, IEEE Transactions on*, vol. 57, no. 9, pp. 6236–6255, Sept 2011.

[27] W. Chen, M. Rodrigues, and I. Wassell, "A Fréchet Mean Approach for Compressive Sensing Date Acquisition and Reconstruction in Wireless Sensor Networks," *Wireless Communications, IEEE Transactions on*, vol. 11, no. 10, pp. 3598 –3606, 2012.

[28] J. M. Winn and C. M. Bishop, "Variational message passing," in *Journal of Machine Learning Research*, 2005, pp. 661–694.

[29] I. Daubechies, R. DeVore, M. Fornasier, and C. S. Gntrk, "Iteratively reweighted least squares minimization for sparse recovery," *Communications on Pure and Applied Mathematics*, vol. 63, no. 1, pp. 1–38, 2010.

[30] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2011.

[31] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 2.0 beta," http://cvxr.com/cvx, Sep. 2013.

[32] P. Erdös and A. Rényi, "On random graphs," *Publications Math.*, vol. 6, pp. 290–297, 1959.

[33] R. Piedrahita, Y. Xiang, N. Masson, J. Ortega, A. Collier, Y. Jiang, K. Li, R. Dick, Q. Lv, M. Hannigan *et al.*, "The next generation of low-cost personal air quality sensors for quantitative exposure monitoring," *Atmospheric Measurement Techniques*, vol. 7, no. 10, pp. 3325–3336, 2014.

[34] P. Bodik, W. Hong, C. Guestrin, S. Madden, M. Paskin, and R. Thibaux. (2004, Feb.) Intel lab data. [Online]. Available: http://db.csail.mit.edu/labdata/labdata.html

[35] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge University Press, 2004.

[36] D. G. Luenberger and Y. Ye, *Linear and nonlinear programming*. Springer Science & Business Media, 2008, vol. 116.