CORE
Provided by Apollo
Metadata, citation and similar papers at core.ac.uk

Variational Bayesian Algorithm For Distributed Compressive Sensing

Wei Chen*† and Ian J. Wassell†
* State Key Laboratory of Rail Traffic Control and Safety, Beijing Jiaotong University, China
† Computer Laboratory, University of Cambridge, UK
wc253@cam.ac.uk, ijw24@cam.ac.uk

***Abstract*—Distributed compressive sensing (DCS) concerns the reconstruction of multiple sensor signals with reduced numbers of measurements, which exploits both intra- and inter-signal correlations. In this paper, we propose a novel Bayesian DCS algorithm based on variational Bayesian inference. The proposed algorithm decouples the common component, that characterizes inter-signal correlation, from innovation components, that represent intra-signal correlation. Such an operation results in a computational complexity of reconstruction which is linear with the number of signals. The superior performance of the algorithm, in terms of the computing time and reconstruction quality, is demonstrated by numerical simulations in comparison with other existing reconstruction methods.**

***Index Terms*—Distributed compressive sensing (DCS), Bayesian inference, signal reconstruction.**

I. Introduction

COMPRESSIVE sensing (CS) [1], [2] enables one to reconstruct compressible signals from a reduced number of linear measurements. Owning to its convenience for data acquisition, CS has been proposed to take the place of the traditional sampling-and-compression approach in applications such as wireless sensor networks (WSNs) where data acquisition is costly. In a WSN, the sensor nodes are embedded into the environment being sensed, which often places stringent constraints on power consumption, since it may be highly impractical to regularly replace or recharge embedded or implanted batteries, especially if there are many nodes forming a network. By applying CS, the number of samples required can be reduced and the compression operation is simpler than that for traditional compression methods. It has been shown that the limited energy supply in WSNs can be used more efficiently with CS, which leads to a longer network lifetime [3]–[5].

CS exploits the sparse structure of a signal and applies random measurements. However, in WSN applications, densely deployed sensors within the event area have high spatial correlations. Such spatial correlations, which represent inter-signal correlations, are not considered in conventional CS. To further exploit inter-signal correlations, the CS framework has been extended for multiple measurement vectors (MMVs) [6], [7] which, similarly to sparse signal reconstruction, jointly recover a set of signals. The MMVs model assumes that all signals share a common support, which may not be true in WSN applications. In [8], [9], distributed compressive sensing (DCS) [8], [9] is proposed to model the intra- and inter-signal correlations with a common component and an innovation component. This DCS model occurs for example in WSN applications that involve monitoring of various physical parameters, such as temperature, humidity, light intensity and air pressure. Global factors, e.g., the sun and prevailing winds, are common to all sensors and contribute to the common component, while local factors corresponding to distinct locations results in different innovation components.

The convenience of the compression operation in CS leads to the increased complexity of the decoding operation, and sophisticated algorithms are required to recover the original signal. Joint reconstruction of multiple signals in DCS has a much higher computational complexity than signal reconstruction in CS. The DCS reconstruction algorithm proposed in [8] concatenates measurements of each signal and performs a weighted ℓ_1-norm minimization to jointly recover multiple signals. This scheme has the following drawbacks: i) the computational complexity of the joint reconstruction is $\mathcal{O}(K^{3.5})$ (where K is the number of signals) times higher than conventional CS with basis pursuit (BP) [10], and thus the algorithm is impractical for a large number of signals with a high dimensionality; ii) one needs to choose weights for the common component and innovation components respectively, and improper selection of weights degrades the reconstruction performance. To reduce the computational complexity, Chen et al. propose a Fréchet mean approach in [4], which first estimates the common signal support from multiple correlated signals and then leverages the support estimate to enhance the reconstruction of each signal.

In this paper, we focus on the sparse common and innovations model of DCS, and propose a Bayesian DCS algorithm, which extends the sparse Bayesian learning framework [11] to the DCS scenario. By applying variational approximation, the new approach decouples the common component, that characterizes inter-signal correlation, from innovation components, that represent intra-signal correlation. Such an operation results in a computational complexity which is linear with the number of signals. The performance of the proposed algorithm

This work is supported by EPSRC Research Grant (EP/K033700/1); the Natural Science Foundation of China (61401018, U1334202); the State Key Laboratory of Rail Traffic Control and Safety (RCS2014ZT08), Beijing Jiaotong University; the Fundamental Research Funds for the Central Universities (2014JBM149); the Key Grant Project of Chinese Ministry of Education (313006); the Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry.

is studied by numerical simulations and compared with other existing approaches.

The rest of the paper is organized as follows: Section II describes in detail the background for CS and DCS. In Section III, we provide the Bayesian DCS framework and the proposed variational Bayesian algorithm. Numerical results are presented in Section IV, followed by conclusions in Section V.

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 $(\cdot)^T$ and $(\cdot)^{-1}$ denote the transpose and the inverse 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 $\mathbf{X}$. $\mathbb{E}_{p(x)}(\cdot)$ denotes expectation with respect to $p(x)$, i.e., the distribution of x. $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes 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, the ℓ_1 norm, and the ℓ_2 norm of vectors, are denoted by $\|\cdot\|_0$, $\|\cdot\|_1$, 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 provide a brief overview of CS under the sparse Bayesian learning framework, and then introduce DCS, i.e., an extension of CS for joint reconstruction of multiple signals with both sparse structures and inter-signal correlation.

A. Compressive Sensing and Sparse Bayesian Learning

The classical CS model is given by:

$$\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{e}, \tag{1}$$

where $\mathbf{y} \in \mathbb{R}^m$ $(m < n)$ denotes the vector of measurements, $\mathbf{A} \in \mathbb{R}^{m \times n}$ denotes the projection matrix, $\mathbf{e} \in \mathbb{R}^m$ denotes the noise term for the measuring process, and $\mathbf{x} \in \mathbb{R}^n$ denotes the s-sparse vector to be estimated. Here, s-sparse means that only s $(s \ll n)$ elements in vector $\mathbf{x}$ are non-zeros while all the other elements are zeros, i.e., $\|\mathbf{x}\|_0 = s$. In practice, one may obtain the measurements vector from the original signal using analogue CS encoders [12], whereby the measurements vector is obtained directly from the analogue continuous-time signal, or using digital CS encoders [13], whereby the measurements vector is obtained from the Nyquist sampled discrete-time signal. Recent studies suggest that digital CS encoders are more energy efficient than analogue CS encoders for WSNs [13].

The typical signal reconstruction process behind conventional CS approaches involves solving the following optimization problem to recover the original signal:

$$\min_{\mathbf{x}} \qquad \|\mathbf{x}\|_1 + \lambda \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2, \tag{2}$$

where λ is a parameter to trade-off sparsity level and distortion. It has been demonstrated that only $m = \mathcal{O}(s \log \frac{n}{s})$ measurements [14] are required for robust reconstruction in the CS framework.

The conventional CS problem can be formulated from a Bayesian perspective. Under the assumption of Gaussian measurement noise, i.e., $\mathbf{e} \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I}_m)$, where σ^2 denotes the noise variance, we have the following likelihood

$$p(\mathbf{y}|\mathbf{x}; \sigma^2) = \mathcal{N}(\mathbf{y}; \mathbf{A}\mathbf{x}, \sigma^2 \mathbf{I}_m). \tag{3}$$

To obtain a sparse solution, it is necessary to consider the use of a sparse-enforcing prior $p(\mathbf{x})$. For example, by using a Laplace prior, the maximum a posteriori estimate, i.e., $\arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y})$, is the solution of the optimization problem (2).

The sparse Bayesian learning framework [11] considers a zero-mean Gaussian prior distribution

$$p(\mathbf{x}; \boldsymbol{\Gamma}) = \mathcal{N}(\mathbf{x}; \mathbf{0}, \boldsymbol{\Gamma}) \tag{4}$$

where $\boldsymbol{\Gamma} \in \mathbb{R}^{n \times n}$ is a diagonal matrix composed of n hyperparameters γ_i $(i = 1, \ldots, n)$. With uniform hyperpriors, i.e., $p(\gamma_i)$ and $p(\sigma^2)$, the value of these hyperparameters can be inferred by

$$\begin{aligned}
\max_{\boldsymbol{\Gamma}, \sigma^2} \log p(\boldsymbol{\Gamma}, \sigma^2 | \mathbf{y}) &\propto \max_{\boldsymbol{\Gamma}, \sigma^2} \log p(\mathbf{y}; \boldsymbol{\Gamma}, \sigma^2) \\
&= \max_{\boldsymbol{\Gamma}, \sigma^2} \log \int p(\mathbf{y}|\mathbf{x}; \sigma^2) p(\mathbf{x}; \boldsymbol{\gamma}) d\mathbf{x} \\
&\propto \min_{\boldsymbol{\Gamma}, \sigma^2} \log |\boldsymbol{\Sigma}| + \mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y},
\end{aligned} \tag{5}$$

where $\boldsymbol{\Sigma} = \sigma^2 \mathbf{I}_m + \mathbf{A}\boldsymbol{\Gamma}\mathbf{A}^T$. In [11], the expectation-maximization (EM) algorithm is employed to solve (5). Given these hyperparameters, $\mathbf{x}$ can be inferred by maximizing the posterior distribution

$$\begin{aligned}
\hat{\mathbf{x}} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{y}; \boldsymbol{\gamma}, \sigma^2) \\
&= \arg\max_{\mathbf{x}} p(\mathbf{y}|\mathbf{x}; \sigma^2) p(\mathbf{x}; \boldsymbol{\gamma}) \\
&= \boldsymbol{\Gamma}\mathbf{A}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}.
\end{aligned} \tag{6}$$

It has been demonstrated in [15], [16] that the sparse Bayesian learning approach penalizes non-sparse solutions with a non-seperable cost function which is superior to solving the optimization problem (2).

B. Distributed Compressive Sensing

DCS extends CS to the application of joint reconstruction of multiple correlated signals. In the DCS setting, K signals are measured by

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{e}_k \quad (k = 1, \ldots, K), \tag{7}$$

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 vector of measurements, the projection matrix, the sparse signal to be estimated, and measurement noise for signal k, respectively. In the DCS model, the sparse signal $\mathbf{x}_k$ $(k = 1, \ldots, K)$ can be represented as

$$\mathbf{x}_k = \mathbf{z}_c + \mathbf{z}_k, \tag{8}$$

where $\mathbf{z}_c \in \mathbb{R}^n$ with $\|\mathbf{z}_c\|_0 = s_c \ll n$ denotes the common component of the sparse signal $\mathbf{x}_k$, 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 signal $\mathbf{x}_k$, which captures the intra-signal correlation and is specific to the signal k.

In [8], Baron et al. propose to jointly reconstruct multiple sparse signals by solving the following optimization problem:

$$\min_{\tilde{\mathbf{z}}} \ \|\tilde{\mathbf{z}}\|_1 \qquad \text{s.t. } \|\mathbf{A}\tilde{\mathbf{z}} - \tilde{\mathbf{y}}\|_2^2 \leq \epsilon, \tag{9}$$

where $\epsilon \geq 0$, $\tilde{\mathbf{z}} = \begin{bmatrix}\mathbf{z}_c^T & \mathbf{z}_1^T & \ldots & \mathbf{z}_K^T\end{bmatrix}^T \in \mathbb{R}^{(K+1)n}$ 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:

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

The optimization problem in (9) can be seen as recovering a $(K+1) \times n$ signal with $\sum_{k=1}^K m_k$ measurements, which in general requires more computing power and storage resources than does independent reconstruction of K signals. In [4], a Fréchet mean approach is proposed for joint reconstruction of multiple correlated signals with a reduced computational complexity. Instead of solving (9) 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:

$$\tilde{\mathbf{z}}_c = \arg\min_{\tilde{\mathbf{z}}_c} \sum_{k=1}^K \lambda_k d^2(\mathbf{A}_k \tilde{\mathbf{z}}_c, \mathbf{y}_k), \tag{10}$$

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:

$$\tilde{\mathbf{z}}_c = (\hat{\mathbf{A}}^T\hat{\mathbf{A}})^{-1}\hat{\mathbf{A}}^T\hat{\mathbf{y}}, \tag{11}$$

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 $\hat{\mathbf{A}} = \begin{bmatrix}\sqrt{\lambda_1}\mathbf{A}_1^T, \cdots, \sqrt{\lambda_K}\mathbf{A}_K^T\end{bmatrix}^T$ and $\hat{\mathbf{y}} = \begin{bmatrix}\sqrt{\lambda_1}\mathbf{y}_1^T, \cdots, \sqrt{\lambda_K}\mathbf{y}_K^T\end{bmatrix}^T$ respectively[1].

III. Variational Bayesian Learning for Distributed Compressive Sensing

In this section, we provide the Bayesian formulation for the DCS model and a variation inference approach for solving the joint reconstruction problem.

[1]Equation (11) requires $rank(\hat{\mathbf{A}}) = n$, which can be satisfied when $\sum_{k=1}^K m_k \geq n$ for randomly generated sensing matrices $\mathbf{A}_k$ $(k = 1, \ldots, K)$.

A. Bayesian Formulation for Distributed Compressive Sensing

Akin to the sparse Bayesian learning framework [11], we adopt zero-mean Gaussian prior distributions for the common component and innovation components, respectively, which are given as

$$p(\mathbf{z}_c; \mathbf{\Gamma}_c) = \mathcal{N}(\mathbf{z}_c; \mathbf{0}, \mathbf{\Gamma}_c) \tag{12}$$

and

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

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

$$p(\mathbf{y}_k|\mathbf{z}_c, \mathbf{z}_k; \sigma^2) = \mathcal{N}(\mathbf{y}_k; \mathbf{A}_k(\mathbf{z}_c + \mathbf{z}_k), \sigma^2\mathbf{I}_{m_k}). \tag{14}$$

Thus, the marginalized probability density function (PDF) is given by

$$\begin{aligned} & p(\mathbf{y}_1, \ldots, \mathbf{y}_K; \mathbf{\Gamma}_c, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K, \sigma^2) \\ &= \int p(\mathbf{z}_c; \mathbf{\Gamma}_c) \prod_{k=1}^K \int p(\mathbf{y}_k|\mathbf{z}_c, \mathbf{z}_k; \sigma^2) p(\mathbf{z}_k; \mathbf{\Gamma}_k) d\mathbf{z}_k \ d\mathbf{z}_c \\ &= \left|\mathbf{I}_n + \mathbf{\Gamma}_c \sum_{k=1}^K \mathbf{A}_k^T \mathbf{\Sigma}_k^{-1} \mathbf{A}_k\right|^{-\frac{1}{2}} \prod_{k=1}^K (2\pi)^{-\frac{m_k}{2}} |\mathbf{\Sigma}_k|^{-\frac{1}{2}} \\ & \exp\left(\frac{1}{2}\left(\sum_{k=1}^K \mathbf{y}_k^T \mathbf{\Sigma}_k^{-1} \mathbf{A}_k\right)\left(\mathbf{\Gamma}_c^{-1} + \sum_{k=1}^K \mathbf{A}_k^T \mathbf{\Sigma}_k^{-1} \mathbf{A}_k\right)^{-1} \right. \\ & \left. \left(\sum_{k=1}^K \mathbf{A}_k^T \mathbf{\Sigma}_k^{-1} \mathbf{y}_k\right) - \frac{1}{2}\left(\sum_{k=1}^K \mathbf{y}_k^T \mathbf{\Sigma}_k^{-1} \mathbf{y}_k\right)\right) \end{aligned} \tag{15}$$

where $\mathbf{\Sigma}_k = \sigma^2\mathbf{I}_{m_k} + \mathbf{A}_k\mathbf{\Gamma}_k\mathbf{A}_k^T$ $(k = 1, \ldots, K)$.

As with sparse Bayesian learning in [11], it is impossible to directly find the optimal hyperparameters that maximize the marginalized PDF (15). By employing Bayesian inference, we can express the posterior as

$$\begin{aligned} & p(\mathbf{z}_c, \mathbf{z}_1, \ldots, \mathbf{z}_K|\mathbf{y}_1, \ldots, \mathbf{y}_K; \mathbf{\Gamma}_c, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K, \sigma^2) \\ &= \frac{p(\mathbf{z}_c; \mathbf{\Gamma}_c) \prod_{k=1}^K p(\mathbf{y}_k|\mathbf{z}_c, \mathbf{z}_k; \sigma^2) p(\mathbf{z}_k; \mathbf{\Gamma}_k)}{p(\mathbf{y}_1, \ldots, \mathbf{y}_K; \mathbf{\Gamma}_c, \mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_K, \sigma^2)}. \end{aligned} \tag{16}$$

We note that the common component and the innovation components are coupled in (15) which makes the posterior (16) become non-separable for $\mathbf{z}_c$ and $\mathbf{z}_k$. Therefore, operations in sparse Bayesian learning [11] cannot be directly applied to solve this problem. In order to apply sparse Bayesian learning, one has to concatenate all sensing matrices and solve a sparse signal reconstruction problem, which leads to manipulations on an $(K+1)n \times (K+1)n$ covariance matrix.

B. Variational Bayesian Algorithm for Distributed Compressive Sensing

In order to reduce the computational complexity, we propose a variational Bayesian algorithm for DCS reconstruction. 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. We adopt the variational approximation in the Bayesian formulation of DCS to find separable functions that approximate the posterior of $\mathbf{z}_c$ and $\mathbf{z}_k$.

To simplify the notations, 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

$$\log p(\mathbf{Y}; \boldsymbol{\theta}) = F(q(\mathbf{Z}), \boldsymbol{\theta}) + \mathtt{KL}(q(\mathbf{Z}) \| p(\mathbf{Z}|\mathbf{Y}; \boldsymbol{\theta})), \quad (17)$$

where

$$F(q(\mathbf{Z}), \boldsymbol{\theta}) = \int q(\mathbf{Z}) \log \left(\frac{p(\mathbf{Z}, \mathbf{Y}; \boldsymbol{\theta})}{q(\mathbf{Z})} \right) d\mathbf{Z}, \quad (18)$$

and

$$\mathtt{KL}(q(\mathbf{Z}) \| p(\mathbf{Z}|\mathbf{Y}; \boldsymbol{\theta}) = - \int q(\mathbf{Z}) \log \left(\frac{p(\mathbf{Z}|\mathbf{Y}; \boldsymbol{\theta})}{q(\mathbf{Z})} \right) d\mathbf{Z} \quad (19)$$

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 $q(\mathbf{Z}) = p(\mathbf{Z}|\mathbf{Y}; \boldsymbol{\theta})$. We assume $q(\mathbf{Z})$ has a factorized form, which is given by

$$q(\mathbf{Z}) = q(\mathbf{z}_c) q(\mathbf{z}_1) \ldots q(\mathbf{z}_K). \quad (20)$$

According to [17], to maximize $F(q(\mathbf{Z}), \boldsymbol{\theta})$, the variational distributions satisfy

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

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

$$\begin{aligned} q(\mathbf{z}_k) &\propto \exp\big(\mathbb{E}_{q(\mathbf{z}_c), q(\mathbf{z}_j), j \neq k}\big[\ln p(\mathbf{y}_1, \ldots, \mathbf{y}_K, \\ &\qquad \mathbf{z}_c, \mathbf{z}_1, \ldots, \mathbf{z}_K; \boldsymbol{\Gamma}_c, \boldsymbol{\Gamma}_1, \ldots, \boldsymbol{\Gamma}_K, \sigma^2)\big]\big) \\ &\propto \exp\big(\mathbb{E}_{q(\mathbf{z}_c)}\big[\ln p(\mathbf{y}_k|\mathbf{z}_c, \mathbf{z}_k, \sigma^2)\big] + \ln p(\mathbf{z}_k|\boldsymbol{\Gamma}_k)\big) \\ &\propto \mathcal{N}(\mathbf{z}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k), \end{aligned} \quad (22)$$

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 (21) and (22), we conclude 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 $q(\mathbf{z}_k)$ $(k = 1, \ldots, K)$, hyperparameters can be updated by $\boldsymbol{\theta} = \arg\max_{\boldsymbol{\theta}} F(q(\mathbf{Z}), \boldsymbol{\theta})$. Specifically, we have

$$\begin{aligned} \gamma_{c,i}^{\text{new}} &= (\Sigma_c)_{i,i} + \mu_{c,i}^2, \\ \gamma_{k,i}^{\text{new}} &= (\Sigma_k)_{i,i} + \mu_{k,i}^2, \\ (\sigma^2)^{\text{new}} &= \frac{1}{K \sum_{k=1}^{K} m_k} \Bigg(\sum_{k=1}^{K} \|\mathbf{y}_k - \mathbf{A}_k(\boldsymbol{\mu}_c + \boldsymbol{\mu}_k)\|_2^2 + \\ &\quad (\sigma^2)^{\text{old}} \sum_{k=1}^{K} \sum_{i=1}^{n} \big(1 - (\gamma_{k,i}^{\text{old}})^{-1} (\Sigma_k)_{i,i}\big) + \\ &\quad (\sigma^2)^{\text{old}} \sum_{i=1}^{n} \big(1 - (\gamma_{c,i}^{\text{old}})^{-1} (\Sigma_c)_{i,i}\big) \Bigg). \end{aligned} \quad (23)$$

The variational optimization proceeds by iteratively updating (21), (22) and (23) until convergence to stable hyperparameters $\boldsymbol{\theta}$. In the end, we can obtain the reconstructed signal by applying the maximum a posteriori estimation

$$\begin{aligned} \hat{\mathbf{x}}_k &= \arg\max_{\mathbf{z}_c + \mathbf{z}_k} p(\mathbf{Z}|\mathbf{Y}; \boldsymbol{\theta}) \\ &= \arg\max_{\mathbf{z}_c} q(\mathbf{z}_c) + \arg\max_{\mathbf{z}_k} q(\mathbf{z}_k) \\ &= \boldsymbol{\mu}_c + \boldsymbol{\mu}_k. \end{aligned} \quad (24)$$

C. Comparison with the Fréchet mean approach

The proposed variational Bayesian algorithm for DCS is derived directly from a Bayesian perspective, while it exhibits some similarities to the Fréchet mean approach [4] in the estimation of the common component. In specific, in each iteration of the proposed algorithm, the mean of the common component is updated by

$$\boldsymbol{\mu}_c = \left(\sum_{k=1}^{K} \mathbf{A}_k^T \mathbf{A}_k + \sigma^2 \boldsymbol{\Gamma}_c^{-1} \right)^{-1} \sum_{k=1}^{K} \mathbf{A}_k^T (\mathbf{y}_k - \mathbf{A}_k \boldsymbol{\mu}_k), \quad (25)$$

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

$$\tilde{\mathbf{z}}_c = \left(\sum_{k=1}^{K} \mathbf{A}_k^T \mathbf{A}_k \right)^{-1} \sum_{k=1}^{K} \mathbf{A}_k^T \mathbf{y}_k. \quad (26)$$

Comparing (25) (26), 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.

IV. NUMERICAL SIMULATIONS

In this section, we compare the performance of the proposed variational Baysian algorithm for DCS reconstruction with other existing approaches by numerical simulations.

A. Experiment Setup

We consider a set of K correlated signals following the 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.

Two performance metrics including computing time and averaged relative error are considered in the comparison. The averaged relative error is defined as the average of $\frac{\sum_{k=1}^K \|\hat{\mathbf{x}}_k - \mathbf{x}_k\|_2^2}{\sum_{k=1}^K \|\mathbf{x}_k\|_2^2}$. We conduct 1000 trials for each experiment setting and provide the averaged result.

The following approaches are compared:

1) ℓ_1 minimization: Signals are reconstructed independently by basis pursuit denoising;
2) Joint ℓ_1 minimization: Joint signal reconstruction by the concatenated and weighted ℓ_1-norm minimization as (9);
3) Fréchet mean approach: Joint signal reconstruction by the Fréchet mean approach [4];
4) Proposed approach: Joint signal reconstruction by the proposed variational Baysian algorithm.

We use CVX, a package for specifying and solving convex programs [18], to solve inverse problems in ℓ_1 minimization, joint ℓ_1 minimization and the Fréchet mean approach.

B. Performance Comparison for DCS

In the first experiment, we compare the computing time consumed by the different approaches in the joint reconstruction of multiple correlated signals that satisfy the DCS model. Our simulations are performed in a MATLAB R2012b environment on a system with a quad-core 3.4GHz CPU and 32 GB RAM, running under the Microsoft Windows 7 operating system. As shown in Fig. 1, a significant improvement of the required computing time can be observed using the proposed approach. This simulation result agrees with the analysis that the computational complexity of joint ℓ_1 minimization is $\mathcal{O}\left((Km)^2 (Kn)^{1.5}\right)$, which is much higher than $\mathcal{O}\left(Km^2 n^{1.5}\right)$, i.e., the complexity of solving an ℓ_1 optimization problem [10].

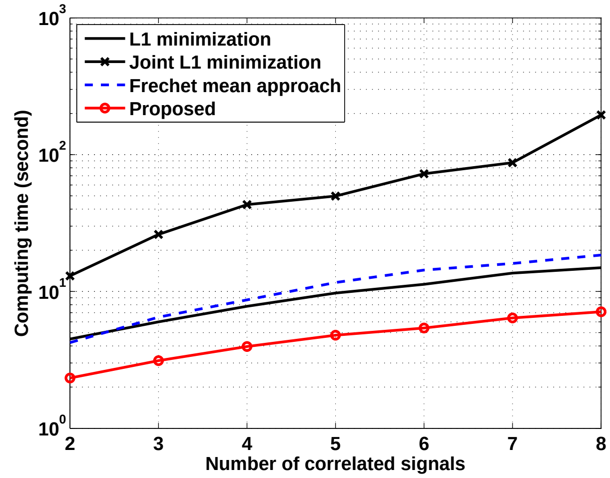

Fig. 1. Comparison of computing time consumed by different approaches ($m = 250$, $n = 500$, $s_c = 90$ and $s_I = 10$).

The reconstruction quality for different approaches is given in Fig. 2. In this experiment, we have compared the averaged relative error against number of measurements, number of signals and innovation component sparsity level. Comparing with conventional CS, i.e., performing independent ℓ_1 minimization, improved reconstruction quality is observed for the three approaches that exploit inter-signal correlations, and the proposed approach outperforms the other two joint reconstruction approaches. In addition, from Fig. 2 (c) we note that a high innovation component sparsity level results in a poor estimation quality of the common component by using the Fréchet mean, and thus degrades the performance of the Fréchet mean approach. However, the gain of the proposed approach is maintained for the case of high innovation component sparsity levels.

V. CONCLUSION

In this paper, we provide a Bayesian DCS framework for joint reconstruction of multiple correlated signals. An algorithm is proposed based on variational inference under the Bayesian DCS framework. The superiority of the proposed approach in relation to other existing approaches is revealed by our experimental study. Future work is to explore theoretical guarantees for the proposed approach.

ACKNOWLEDGMENT

The authors would like to thank Dr. David Wipf for generously lending his expertise in a series of insightful discussions.

REFERENCES

[1] 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.

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

[3] R. Xie and X. Jia, “Transmission-efficient clustering method for wireless sensor networks using compressive sensing,” *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 3, pp. 806–815, March 2014.

[4] 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.

[5] W. Chen and I. Wassell, “Energy-efficient signal acquisition in wireless sensor networks: a compressive sensing framework,” *Wireless Sensor Systems, IET*, vol. 2, no. 1, pp. 1–8, 2012.

[6] 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.

[7] 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.

[8] 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.

[9] 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.

[10] D. Needell and J. Tropp, “Cosamp: Iterative signal recovery from incomplete and inaccurate samples,” *Applied and Computational Harmonic Analysis*, vol. 26, no. 3, pp. 301 – 321, 2009.

[11] M. E. Tipping, “Sparse bayesian learning and the relevance vector machine,” *The journal of machine learning research*, vol. 1, pp. 211–244, 2001.

[12] F. Chen, A. Chandrakasan, and V. Stojanovic, “Design and analysis of a hardware-efficient compressed sensing architecture for data compression in wireless sensors,” *Solid-State Circuits, IEEE Journal of*, vol. 47, no. 3, pp. 744–756, March 2012.

[13] F. Chen, F. Lim, O. Abari, A. Chandrakasan, and V. Stojanovic, “Energy-aware design of compressed sensing systems for wireless sensors under performance and reliability constraints,” *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 60, no. 3, pp. 650–661, March 2013.

[14] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin, “A simple proof of the restricted isometry property for random matrices,” *Constructive Approximation*, vol. 28, no. 3, pp. 253–263, 2008.

[15] 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.

[16] 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.

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

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

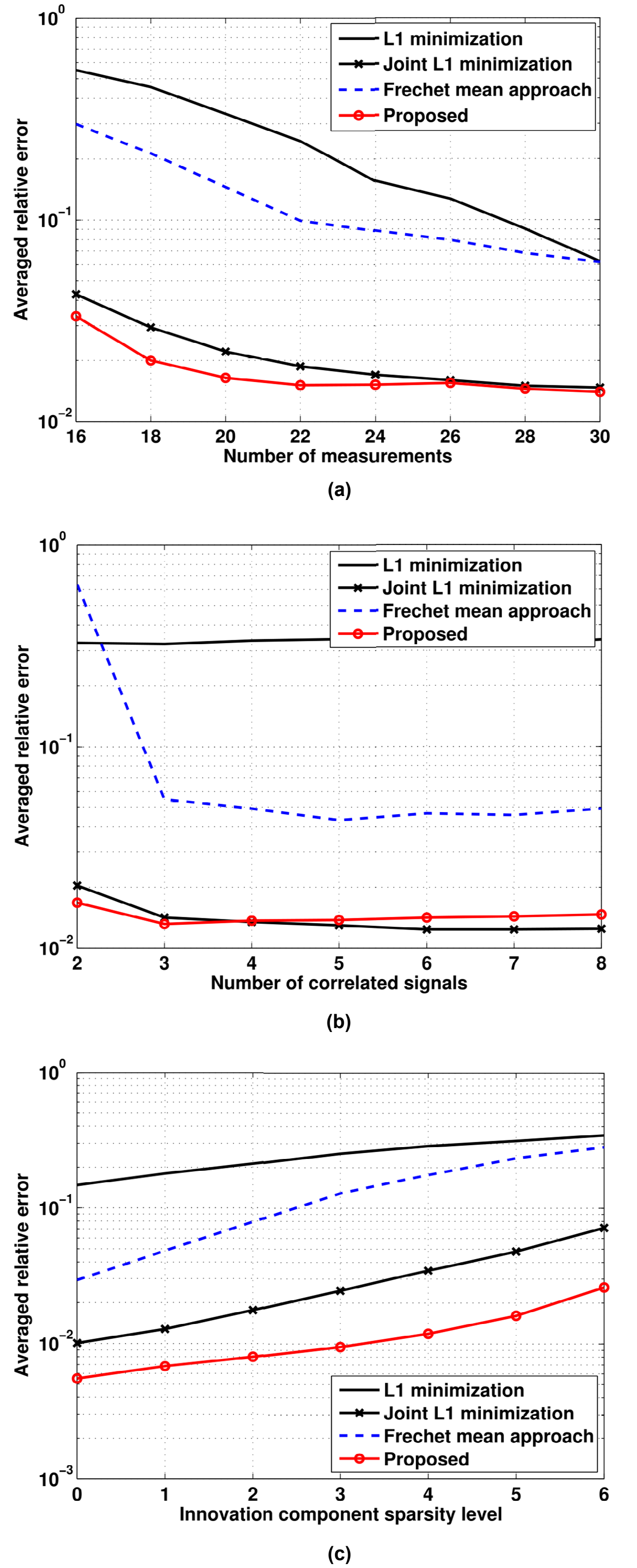

Fig. 2. Comparison of reconstruction performance for different approaches. (a) reconstruction quality vs. number of measurements ($n = 50$, $K = 4$, $s_c = 8$ and $s_I = 2$); (b) reconstruction quality vs. number of correlated signals ($n = 50$, $m = 25$, $s_c = 10$ and $s_I = 1$); (c) reconstruction quality vs. innovation component sparsity level ($n = 50$, $m = 25$, $K = 4$ and $s_c = 10$).