Bayesian Learning For The Type-3 Joint Sparse Signal Recovery

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, ijw24}@cam.ac.uk

Abstract—**Compressed sensing (CS) is a signal acquisition paradigm that utilises the finding that a small number of linear projections of a sparse signal have enough information for stable recovery. This paper develops a Bayesian CS algorithm to simultaneously recover multiple signals that follow the Type-3 joint sparse model [1], [2], where signals share a non-sparse common component and have distinct sparse innovation components. By employing the expectation-maximization (EM) algorithm, the proposed algorithm iteratively updates the estimates of the common component and innovation components. In particular, we find that the update rule for the non-sparse common component in the proposed algorithm, differs from all the other methods in the literature, and we provides an interpretation that gives a valuable insight into why the proposed algorithm is successful in estimating the non-sparse common component. The superior performance of the proposed algorithm is demonstrated by numerical simulation results.**

Index Terms—**compressive sensing (CS), distributed compressive sensing (DCS), Bayesian learning, signal reconstruction.**

I. INTRUDUCTION

COMPRESSED sensing (CS) [3], [4] enables one to reconstruct compressible signals from a reduced number of samples, and thus has been proposed for applications where data acquisition is costly. Typical applications that could benefit from CS include but are not limited to spectrum sensing for cognitive radio systems [5], sparse channel estimation [6], image acquisition by a mobile phone camera sensor [7] and wireless sensor networks [8].

CS exploits the sparse structure of naturally occurring signals under some basis, in order to reduce the degree of freedom in recovering the original signal from a reduced number of random linear measurements. The key problem in CS reconstruction is the search for the sparsest solution for a linear underdetermined system, which is unfortunately NP-hard. Iterative reweighted schemes [9], [10] have demonstrated success for finding sparse solutions, where in particular sparse Bayesian learning (SBL), i.e., an iterative reweighted scheme with non-separable penalties, has been shown to outperform many of the other schemes in terms of the recovery accuracy [11], [12].

The conventional CS framework mainly focuses on the case of a single signal, while for a sensor network with multiple nodes as an example, in addition to the sparse structure embedded in each sensor's signal, it is desired to exploit the inter-signal correlation to further improve the reconstruction performance. To get benefits from the inter-node correlation, the CS framework has been extended to joint signal reconstruction for multiple nodes. In [1], [2], the Type-3 joint sparse model (JSM-3), i.e., one of the available distributed compressive sensing (DCS) models, is proposed to model the intra-signal and inter-signal correlations with a non-sparse common component and a sparse innovation component. This JSM-3 model occurs for example in a verification system in a component production factory, where cameras acquire snapshots of each component to check for manufacturing defects with a common background that is not sparse in any basis.

In this paper, we focus on the development of a decentralized SBL algorithm for the JSM-3 [1], [2], where the innovation components, which reflect the intra-node correlations of different nodes, are sensitive data and are not shared, while the common component, which reflects the inter-node correlation, is jointly reconstructed. By applying variational approximation, the common component is decoupled from the innovation components in the SBL framework. Such an operation leads to iterative reconstruction of the common component and the innovation components. In the sequel, we cast the decoupled reconstruction problem as a set of decentralized problems with consensus constraints, where each node exchanges limited non-sensitive information with its neighbors and recovers its own innovation component by using its local data. Experimental results show that the proposed decentralized algorithm achieves a reconstruction accuracy close to the centralized SBL algorithm, and exhibits a good convergence rate.

The rest of the paper is organized as follows: Section II describes the background of CS and the JSM-3. In Section III, we provide details of the proposed Bayesian learning algorithm for JSM-3. Numerical results are presented in Section IV, followed by conclusions in Section V.

The following notation is used throughout. Boldface upper-case letters and boldface lower-case letters denotes matrices and column vectors, respectively. Calligraphic upper-case letters denote support sets. The superscripts $(\cdot)^T$, $(\cdot)^{-1}$ and $(\cdot)^\dagger$

denote the transpose, the inverse and the pseudoinverse of a matrix, respectively. $\mathbb{E}_x(\cdot)$ denotes expectation with respect to $p(x)$, i.e., the distribution of x. $\mathcal{N}(\mathbf{x}; \boldsymbol{\mu}, \boldsymbol{\Sigma})$ 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, 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

A. Compressive Sensing Model and Reconstruction

Given a signal $\mathbf{f} \in \mathbb{R}^n$, we consider a measurement system that acquires m $(m \leq n)$ linear measurements by projecting the signal with a sensing matrix $\boldsymbol{\Phi} \in \mathbb{R}^{m \times n}$. This sensing system can be presented as $\mathbf{y} = \boldsymbol{\Phi}\mathbf{f}$, where $\mathbf{y} \in \mathbb{R}^m$ denotes the measurement vector. The standard CS framework assumes that the sensing matrices are randomized and non-adaptive, which means each measurement is derived independently to previously acquired measurements.

As the signal $\mathbf{f}$ can be represented by an s-sparse vector $\mathbf{x} \in \mathbb{R}^n$ in some basis $\boldsymbol{\Psi} \in \mathbb{R}^{n \times n}$, i.e., $\mathbf{f} = \boldsymbol{\Psi}\mathbf{x}$, the sensing system can be rewritten as $\mathbf{y} = \boldsymbol{\Phi}\boldsymbol{\Psi}\mathbf{x} = \mathbf{A}\mathbf{x}$, where $\mathbf{A} = \boldsymbol{\Phi}\boldsymbol{\Psi}$ denotes an equivalent sensing matrix. More generally, measurements are considered to be contaminated by some noise term $\mathbf{n} \in \mathbb{R}^m$ owing to the sampling noise or the quantization process. Then the CS model can be described as

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

In generally, it is not possible to solve (1) even if the noise term is equal to zero, as there are an infinite number of solutions satisfying (1). However, a suitable sparsity constraint may rule out all the solutions except for the one that is expected. Therefore, the most natural strategy to recover the sparse representation from the measurements uses ℓ_0 norm minimization, which is, however, a combinatorial optimization problem and thus computationally intractable.

Consequently, as a convex relaxation of ℓ_0 norm minimization, ℓ_1 norm minimization is used instead to solve the sparse signal representation, which leads to a linear program (LP) and thus straight forward to solve. Therefore, the optimization problem becomes

$$\begin{aligned} \min_{\mathbf{x}} \quad & \|\mathbf{x}\|_1 \\ \text{s.t.} \quad & \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2 \leq \epsilon, \end{aligned} \tag{2}$$

where ϵ is an estimate of the noise level. This program is also known as the basis pursuit de-noising (BPDN), and it has been demonstrated that only $m = \mathcal{O}(s \log \frac{n}{s})$ measurements [13] are required for robust reconstruction of the original signal.

The ℓ_1 norm is a convex relaxation of the ℓ_0 norm, so that a globally optimal solution can be obtained. However, this convex optimization problem is known to be too loose to approximate the original ℓ_0 norm problem and thus it requires sufficiently strict conditions on $\mathbf{A}$. In addition, convergence to a global minimum with a cost function value smaller than the value of the most sparse one can occur in this convex

optimization problem. To break this limitation, optimization problems with a non-convex regularizer have been used instead, e.g., the ℓ_p norm minimization $(p < 1)$ [14], and Bayesian algorithms [15].

B. The Type-3 Joint Sparse Model

This standard CS framework only exploits the sparse characteristics of the signal to reduce the dimensionality required for sensing the signal. A recent growing trend relates to the use of joint sparse signal models to further enhance the performance of CS. For a group of K signals, which are acquired via CS measurements with some additive noise, we have

$$\mathbf{y}_k = \mathbf{A}_k \mathbf{x}_k + \mathbf{e}_k, \tag{3}$$

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 equivalent sensing matrix, the signal representation, and noise of node k, respectively. The JSM-3 [1], [2] models the inter-signal correlation by letting

$$\mathbf{x}_k = \mathbf{z}_c + \mathbf{z}_k, \tag{4}$$

where $\mathbf{z}_c \in \mathbb{R}^n$ denotes the common component which is the same for all nodes and is non-sparse, and $\mathbf{z}_k \in \mathbb{R}^n$ denotes the innovation component of node k and $\|\mathbf{z}_k\|_0 \leq s$. This model is suitable for applications where multiple signals share a common background which is too complicated to be represented by only a few atoms in any basis.

Under the JSM-3, no individual signal representation $\mathbf{x}_k$ is sparse, and thus recovery of each signal separately will fail when $m_k < n$. To achieve signal recovery using far fewer than n measurements per signal, joint reconstruction is necessary. In [1], an approach, namely transpose estimation of common component (TECC), is proposed to recover signals following the JSM-3 model, and is listed below:

1) Estimate common component $\hat{\mathbf{z}}_c$ given $\mathbf{y}_k$ and $\mathbf{A}_k$ $(k = 1, \ldots, K)$;
2) Estimate measurements generated by innovation components by $\hat{\mathbf{y}}_k = \mathbf{y}_k - \mathbf{A}_k \hat{\mathbf{z}}_c$ for all k;
3) Recover each innovation component $\hat{\mathbf{z}}_k$ from $\hat{\mathbf{y}}_k$ by a standard single-signal CS recovery algorithm;
4) Obtain each signal estimate $\hat{\mathbf{x}}_k = \hat{\mathbf{z}}_c + \hat{\mathbf{z}}_k$.

In TECC, the common component $\mathbf{z}_c$ is firstly estimated from the measurements of all signals without considering the effect of innovation components, and then the innovation components are estimated by standard CS recovery after removing the impact of the common component. It has been proved in [1] that the number of measurements of an individual signal can be significantly decreased for successful recovery by the TECC, as long as the total number of measurements is sufficiently large to capture enough information about the non-sparse common component.

In TECC, the estimated common component may not be sufficiently accurate to enable correct reconstruction of the sparse innovation components, which leads to incorrect estimate of signals. To remedy the drawback of TECC, another

approach, called alternating common and innovation estimation (ACIE), is proposed in [1], where the common component and innovation components are estimated iteratively. Suppose the support of the kth estimated innovation component is $\mathcal{J}_k$, and define $\mathbf{A}_{k,\mathcal{J}_k}$ as a submatrix of $\mathbf{A}_k$ corresponding to $\mathcal{J}_k$. A matrix $\mathbf{Q}_k$ with orthonormal columns can be found that spans the orthogonal complement of $\mathbf{A}_{k,\mathcal{J}_k}$. Then the ACIE algorithm can be described as follows:

1) Set $\mathcal{J}_k = \emptyset$ for each k;
2) Estimate common component $\hat{\mathbf{z}}_c = \tilde{\mathbf{A}}^\dagger \tilde{\mathbf{y}}$, where $\tilde{\mathbf{A}} = [\tilde{\mathbf{A}}_1^T \dots \tilde{\mathbf{A}}_K^T]^T$, $\tilde{\mathbf{A}}_k = \mathbf{Q}_k^T \mathbf{A}_k$, $\tilde{\mathbf{y}} = [\tilde{\mathbf{y}}_1^T \dots \tilde{\mathbf{y}}_K^T]^T$ and $\tilde{\mathbf{y}}_k = \mathbf{Q}_k^T \mathbf{y}_k$;
3) Recover each innovation component $\hat{\mathbf{z}}_k$ from $\mathbf{y}_k - \mathbf{A}_k \hat{\mathbf{z}}_c$ using a standard single-signal CS recovery algorithm, and find the support $\mathcal{J}_k$;
4) Iterate step 2 and 3 until some halting criterion is satisfied and then obtain each signal estimate $\hat{\mathbf{x}}_k = \hat{\mathbf{z}}_c + \hat{\mathbf{z}}_k$.

The construction of $\tilde{\mathbf{A}}$ and $\tilde{\mathbf{y}}$ in the step 2 allows one to project the measurements into the subspace that is orthogonal to the innovation component subspace, so that the estimated common component is obtained exclusively by vectors not in $\mathcal{J}_k$.

III. BAYESIAN LEARNING FOR JSM-3

In this section, we present Bayesian algorithms for solving the JSM-3 reconstruction problem. We adopt Gaussian prior distributions for the signal representation $\mathbf{x}_k$, that is given as

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

where $\boldsymbol{\mu} \in \mathbb{R}^n$ and $\boldsymbol{\Gamma}_k \in \mathbb{R}^{n \times n}$ is a diagonal matrix with hyperparameters $\gamma_{k,i}$ $(k = 1, \dots, K; i = 1, \dots, n)$. 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

$$p(\mathbf{y}_k | \mathbf{x}_k; \sigma^2) = \mathcal{N}(\mathbf{y}_k; \mathbf{A}_k \mathbf{x}_k, \sigma^2 \mathbf{I}_{m_k}). \tag{6}$$

With uniform hyperpriors $p(\gamma_{k,i})$, $p(\mu_i)$ and $p(\sigma^2)$, the value of these hyperparameters can be inferred by

$$\max_{\boldsymbol{\mu}, \{\boldsymbol{\Gamma}_k\}, \sigma^2} \log p(\boldsymbol{\mu}, \{\boldsymbol{\Gamma}_k\}, \sigma^2 | \{\mathbf{y}_k\})$$

$$\propto \max_{\boldsymbol{\mu}, \{\boldsymbol{\Gamma}_k\}, \sigma^2} \log p(\{\mathbf{y}_k\}; \boldsymbol{\mu}, \{\boldsymbol{\Gamma}_k\}, \sigma^2)$$

$$= \max_{\boldsymbol{\mu}, \{\boldsymbol{\Gamma}_k\}, \sigma^2} \sum_{k=1}^K \log \int p(\mathbf{y}_k | \mathbf{x}_k; \sigma^2) p(\mathbf{x}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k) d\mathbf{x}_k$$

$$\propto \min_{\boldsymbol{\mu}, \{\boldsymbol{\Gamma}_k\}, \sigma^2} \sum_{k=1}^K \log |\boldsymbol{\Sigma}_k| + \sum_{k=1}^K (\mathbf{y}_k - \mathbf{A}_k \boldsymbol{\mu})^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_k - \mathbf{A}_k \boldsymbol{\mu}),$$
$$\tag{7}$$

where $\boldsymbol{\Sigma}_k = \sigma^2 \mathbf{I}_{m_k} + \mathbf{A}_k \boldsymbol{\Gamma}_k \mathbf{A}_k^T$. Given these hyperparameters,

$\mathbf{x}_k$ can be inferred by maximizing the posterior distribution

$$\begin{aligned}
\mathbf{x}_k &= \arg \max_{\mathbf{x}_k} p(\mathbf{x}_k | \mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k, \sigma^2) \\
&= \arg \max_{\mathbf{x}_k} p(\mathbf{y}_k | \mathbf{x}_k; \sigma^2) p(\mathbf{x}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k) \\
&= (\sigma^{-2} \mathbf{A}_k^T \mathbf{A}_k + \boldsymbol{\Gamma}_k^{-1})^{-1} (\sigma^{-2} \mathbf{A}_k^T \mathbf{y}_k + \boldsymbol{\Gamma}_k^{-1} \boldsymbol{\mu}) \\
&= \boldsymbol{\Gamma}_k \mathbf{A}_k^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_k - \mathbf{A}_k \boldsymbol{\mu}) + \boldsymbol{\mu}.
\end{aligned} \tag{8}$$

Therefore, the learning problem amounts to the estimation of $\boldsymbol{\mu}$, $\{\boldsymbol{\Gamma}_k\}$ and σ^2 in (7).

A. Algorithm Derivation

For fixed hyperparameters, the posterior density of each signal representation follows a Gaussian distribution, i.e.,

$$p(\mathbf{x}_k | \mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k, \sigma^2) = \mathcal{N}(\mathbf{x}_k; \boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k^x), \tag{9}$$

where

$$\boldsymbol{\mu}_k = \boldsymbol{\Gamma}_k \mathbf{A}_k^T \boldsymbol{\Sigma}_k^{-1} (\mathbf{y}_k - \mathbf{A}_k \boldsymbol{\mu}) + \boldsymbol{\mu}, \tag{10}$$

and

$$\boldsymbol{\Sigma}_k^x = (\sigma^{-2} \mathbf{A}_k^T \mathbf{A}_k + \boldsymbol{\Gamma}_k^{-1})^{-1} = \boldsymbol{\Gamma}_k - \boldsymbol{\Gamma}_k \mathbf{A}_k^T \boldsymbol{\Sigma}_k^{-1} \mathbf{A}_k \boldsymbol{\Gamma}_k. \tag{11}$$

The second equality in (11) is derived by using the Woodbury matrix indentity. While the inversion of an $n \times n$ matrix requires $\mathcal{O}(n^2 \log n)$ operations [16], this matrix transform reduces the computational complexity as it only requires the inversion of an $m_k \times m_k$ matrix $\boldsymbol{\Sigma}_k$ where $m_k \ll n$.

We employ the expectation-maximization (EM) algorithm to find the maximum likelihood estimate (MLE) of the unknown hyperparameters in (7), where all the signal representations $\{\mathbf{x}_k\}$ are treated as hidden variables. In the E step we compute the posterior of the hidden variables, i.e., $p(\mathbf{x}_k | \mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k, \sigma^2)$, and find the hyperparameters that maximize the expected value of the log likelihood function $\sum_{k=1}^K \log p(\mathbf{y}_k, \mathbf{x}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k, \sigma^2)$ with respect to the conditional distribution of $\{\mathbf{x}_k\}$ given $\{\mathbf{y}_k\}$ in the M step. The update rules are given as follows:

E step:
$$\mathbb{E}_{\mathbf{x}_k | \mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k, \sigma^2}[\mathbf{x}_k] = \boldsymbol{\mu}_k, \tag{12a}$$

$$\mathbb{E}_{\mathbf{x}_k | \mathbf{y}_k; \boldsymbol{\mu}, \boldsymbol{\Gamma}_k, \sigma^2}\left[(x_{ki} - \mu_i)^2\right], = \Sigma_{k\,i,i}^x + \mu_i^2 \tag{12b}$$

M step:
$$\gamma_{k\,i}^{\text{new}} = \arg\max_{\gamma_{k\,i}} \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k;\boldsymbol{\mu},\boldsymbol{\Gamma}_k,\sigma^2}\left[\log p(\mathbf{x}_k;\boldsymbol{\mu},\boldsymbol{\Gamma}_k)\right]$$

$$= \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k;\boldsymbol{\mu},\boldsymbol{\Gamma}_k,\sigma^2}\left[(x_{k\,i}-\mu_i)^2\right] \tag{13a}$$

$$\boldsymbol{\mu}^{\text{new}} = \arg\max_{\boldsymbol{\mu}} \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k;\boldsymbol{\mu},\boldsymbol{\Gamma}_k,\sigma^2}\left[\sum_{k=1}^{K}\log p(\mathbf{x}_k;\boldsymbol{\mu},\boldsymbol{\Gamma}_k)\right]$$

$$= \sum_{k=1}^{K}\left(\mathbf{I}_n + \sum_{k'\neq k}\boldsymbol{\Gamma}_k\boldsymbol{\Gamma}_{k'}^{-1}\right)^{-1}\mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k;\boldsymbol{\mu},\boldsymbol{\Gamma}_k,\sigma^2}[\mathbf{x}_k]$$

$$= \sum_{k=1}^{K}\left(\mathbf{I}_n + \sum_{k'\neq k}\boldsymbol{\Gamma}_k\boldsymbol{\Gamma}_{k'}^{-1}\right)^{-1}\boldsymbol{\mu}_k \tag{13b}$$

$$(\sigma^2)^{\text{new}} = \arg\max_{\sigma^2} \mathbb{E}_{\mathbf{x}_k|\mathbf{y}_k;\boldsymbol{\mu},\boldsymbol{\Gamma}_k,\sigma^2}\left[\sum_{k=1}^{K}\log p(\mathbf{y}_k|\mathbf{x}_k;\sigma^2)\right]$$

$$= \frac{\sum_{k=1}^{K}\|\mathbf{y}_k - \mathbf{A}_k\boldsymbol{\mu}_k\|_2^2 + \sigma^2\sum_{i=1}^{N}\left(1 - \gamma_{k\,i}^{-1}\Sigma_{k\,i,i}^x\right)}{KN}$$
$$\tag{13c}$$

The E step and the M step are iteratively applied until convergence is achieved. Then the estimated reconstructed signal representations are computed by (8).

B. Analysis

According to the property of the EM algorithm, the proposed algorithm is guaranteed to converge to either a local maxima or a saddle point of the log-likelihood $\log p(\{\mathbf{y}_k\};\boldsymbol{\mu},\{\boldsymbol{\Gamma}_k\},\sigma^2)$. By comparing the update rule of the proposed algorithm and that of the SBL [15], we note that in the proposed algorithm there is an additional update for the common component $\boldsymbol{\mu}$ in (13b), which enables the proposed algorithm to find the sparse innovation components of different signals while also estimating the non-sparse common component.

Given the estimated innovation component $\mathbf{z}_k$, the common component can be estimated using various standard tools from linear algebra, e.g., performing least squares (LS) estimation or by the ones used in TECC and ACIE [1]. The update rule of the common component in (13b) which is different to these standard approaches, is simply the weighted sum of the estimate of all signals. Interestingly, upon convergence, we find that the proposed algorithm finds an accurate estimate of the common component and also good reconstruction accuracy of the original signal. We now provide some intuition behind the estimation step (13b) by investigating the value of the weight corresponding to signal k, i.e., $\mathbf{w}_k = \left(\mathbf{I}_n + \sum_{k'\neq k}\boldsymbol{\Gamma}_k\boldsymbol{\Gamma}_{k'}^{-1}\right)^{-1}$.

- Case I: If $\gamma_{k\,i} \to 0$, we have $w_{k\,i} > 0$.
- Case II: If $\gamma_{k\,i} > 0$ and $\gamma_{k'\,i} > 0$ $\forall k'$, we have $w_{k\,i} > 0$.
- Case III: If $\gamma_{k\,i} > 0$ and there is some other k' such that $\gamma_{k'\,i} \to 0$, we have $w_{k\,i} \to 0$.

For both case I and II, we have $w_{k\,i} > 0$, which means the estimate of the ith element of signal k has a non-zero weight and will contribute to the estimate of the common component in (13b). However, the reasons which lead to the same result in the two cases are significantly different. In case I, it can be

inferred from $\gamma_{k\,i} \to 0$ that the ith element of the innovation component k is likely to be zero, and thus the estimate of the ith element of signal k is actually the estimate of the common component. For case II where $\gamma_{k\,i} > 0$ and $\gamma_{k'\,i} > 0$ $\forall k'$, the ith elements of all the estimated innovation components are likely to be non-zeros and they are all used in calculating the common component in (13b). In contrast to case I and II, in case III we have $w_{k\,i} \to 0$. In this scenario, $\gamma_{k\,i} > 0$ infers that the ith element of the estimated innovation component k is non-zero, and $\gamma_{k'\,i} \to 0$ infers a zero-valued element of some innovation component k'. Thus the ith element of the estimated innovation component k will not contribute to the calculation of the common component in (13b).

The update rule of the common component in (13b) is significantly different to other JSM-3 recovery approaches proposed in the literature [1]. In view of the fact that, upon convergence, the hyperparameter $\gamma_{k\,i} = \frac{1}{K}\sum_{k=1}^{K}(x_{k\,i}-z_{c\,i})^2$ according to (13a), a generalized strategy for JSM-3 can be derived, which is given as follows:

1) Initialize $\hat{\mathbf{z}}_c = \mathbf{0}$ or use the result of any JSM-3 algorithm as a warm start;
2) Estimate innovation components: compute the estimate $\hat{\mathbf{z}}_k$ from $\mathbf{y}_k - \mathbf{A}_k\hat{\mathbf{z}}_c$ by some CS recovery algorithm;
3) Estimate common component: update the estimate[1] $\hat{z}_{c\,i} = \sum_{k=1}^{K}\left(1 + \sum_{k'\neq k}\hat{z}_{k\,i}^2/\hat{z}_{k'\,i}^2\right)^{-1}(\hat{z}_{k\,i} + \hat{z}_{c\,i})$;
4) Iterate Step 2 and 3 until convergence and compute the estimate of each signal $\mathbf{x}_k = \hat{\mathbf{z}}_k + \hat{\mathbf{z}}_c$.

This general recovery strategy for JSM-3 can exploit any CS technique to estimate the innovation components, and has the advantage of a low computational complexity in calculating the common component. Note that to update the common component, the ACIE algorithm [1] requires computation of the pseudoinverse of a matrix that is the concatenation of all equivalent sensing matrix $\mathbf{A}_k$ ($k = 1, \ldots, K$).

IV. NUMERICAL SIMULATIONS

This section presents the numerical results of solving the JSM-3 reconstruction problem by using the proposed approach. For comparison, we also present the performance of the TECC and the ACIE proposed in [1].

A. Simulation Settings

Our numerical simulations consider K signals that follow the JSM-3. Without loss of generality, we let all signals have the same number of measurements, i.e., $m = m_k$ ($k = 1, \ldots, K$), and the innovation components of different nodes have the same sparsity level, i.e., $s = \|\mathbf{z}_k\|_0$ ($k = 1, \ldots, K$). The innovation component supports are selected randomly, and the non-sparse common component and the non-zero elements in the innovation components are drawn from independent and identically distributed (i.i.d.) Gaussian distributions $\mathcal{N}(0, 1)$. The equivalent sensing matrix of each node is independently and randomly generated with i.i.d. Gaussian entries following

---

[1]To clarify the calculation of division by zero, we define $\frac{a}{0} = +\infty$ if $a > 0$ and $\frac{a}{0} = 0$ if $a = 0$.
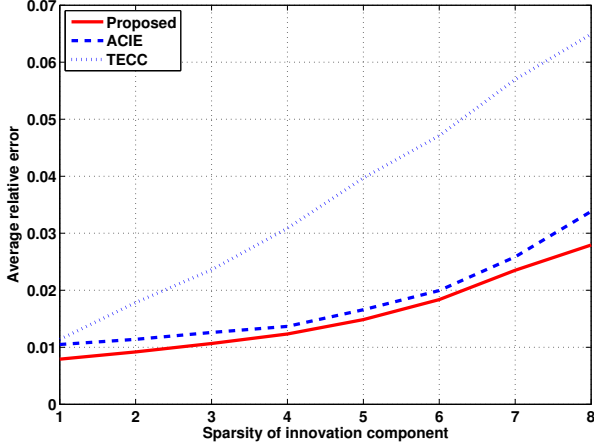
Fig. 1. Reconstruction accuracy vs. sparsity of innovation component. ($K = 10$, $n = 50$, $m = 30$ and SNR of 20dB).
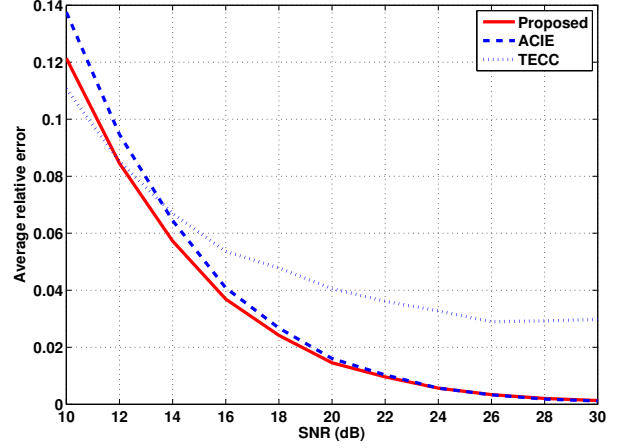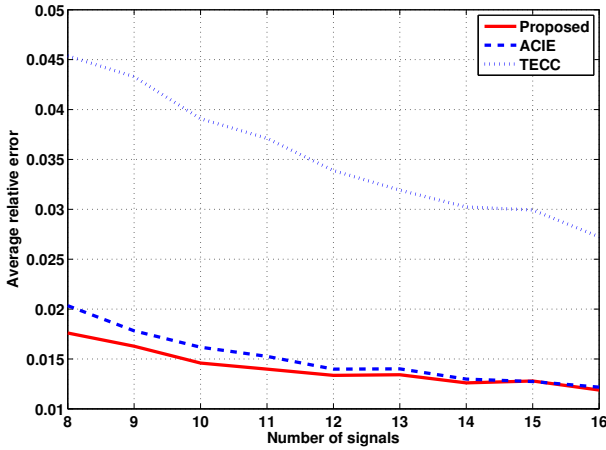

Fig. 2. Reconstruction accuracy vs. number of signals. ($s = 5$, $n = 50$, $m = 30$ and SNR of 20dB).

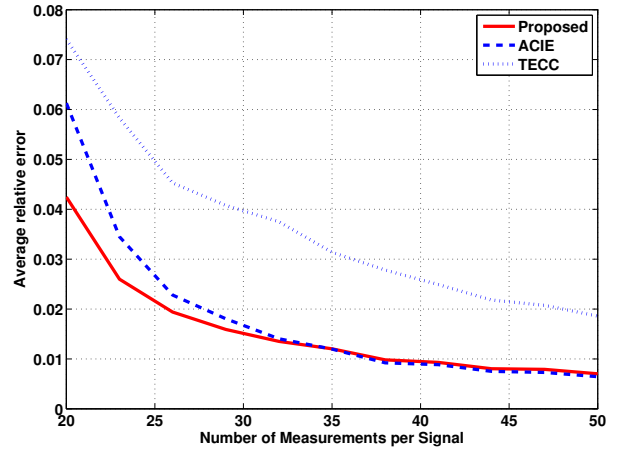Fig. 3. Reconstruction accuracy vs. SNR. ($K = 10$, $s = 5$, $n = 50$ and $m = 30$).

Fig. 4. Reconstruction accuracy vs. number of measurements per signal. ($K = 10$, $s = 5$, $n = 50$ and SNR of 20dB).

$\mathcal{N}(0, 1)$, followed by a column normalization. The received measurements are corrupted by additive zero-mean Gaussian noise to yield a signal noise ratio (SNR), i.e., $\frac{\|\mathbf{A}_k \mathbf{x}_k\|_2^2}{\|\mathbf{e}_k\|_2^2}$. In order to make a fair comparison, we use the SBL algorithm in the TECC and the ACIE to update the innovation components.

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

B. Recovery Performance

The reconstruction accuracy for different approaches is given in Fig. 1, 2, 3 and 4, where we have compared the averaged relative error against various factors including the sparsity of innovation component, the number of signals, SNR

and the number of measurements per signal. Our numerical simulation results confirm that the proposed Bayesian CS algorithm for JSM-3 has a good performance in comparison to both the TECC and the ACIE.

Both the proposed algorithm and the ACIE alternately and iteratively update the common component and innovation components, while the TECC doesn't update the estimate of the common component based on the innovation components. Although the TECC has a low computational complexity as it only computes the common component and each innovation component once, it fails to provide an accurate reconstruction result. It can be seen in Fig. 4 that the performance of the ACIE is similar to that of the proposed algorithm when a large number of measurements per signal is available. It should be noted that the ACIE requires the support of each innovation component to be determined in each iteration, which is selected according to the s largest elements in our

experiments. However, knowledge concerning the sparsity of innovation component is difficult to obtain in practice, and an inappropriate criterion for selecting the the support of innovation component will impair the reconstruction accuracy of the ACIE. However, the proposed algorithm does not need such empirical knowledge to remove the impact of innovation components, and the computational complexity in calculating the common component is also lower than the ACIE which involves the pseudoinverse of a matrix.

V. Conclusion

In this paper, we provide a Bayesian CS algorithm for jointly recovering multiple signals that follow the JSM-3. Such a joint sparse model occurs in applications where multiple signals have different sparse objects but a common background that is not sparse in any existing basis. We demonstrate the good performance of the proposed algorithm by numerical simulation results. In addition, the update rule for the common component, which is derived from the EM algorithm, is quite different to the other techniques used in existing reconstruction approaches for solving the JSM-3 recovery problem, and is the key to understanding the properties of the proposed algorithm.

References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] Y. Barbotin, A. Hormati, S. Rangan, and M. Vetterli, "Estimation of sparse mimo channels with common support," *Communications, IEEE Transactions on*, vol. 60, no. 12, pp. 3705–3716, 2012.

[7] D. Schneider, "New camera chip captures only what it needs," *Spectrum, IEEE*, vol. 50, no. 3, pp. 13–14, 2013.

[8] 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.

[9] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *Journal of Fourier analysis and applications*, vol. 14, no. 5-6, pp. 877–905, 2008.

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

[11] 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.

[12] 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.

[13] 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.

[14] R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," *Signal Processing Letters, IEEE*, vol. 14, no. 10, pp. 707–710, 2007.

[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] A. Tveit, "On the complexity of matrix inversion," *Mathematical Note*, p. 1, 2003.