Variational Bayesian Inference-Based Counting and Localization for Off-Grid Targets With Faulty Prior Information in Wireless Sensor Networks

Yan Guo, Baoming Sun, Ning Li, and Dagang Fang, *Life Fellow, IEEE*

***Abstract*— The benefits of compressive sensing (CS) on counting and localization in wireless sensor networks have already been demonstrated in previous works. However, most existing solutions usually become infeasible in the presence of off-grid targets, which is inevitable for the traditional CS-based localization methods. To mitigate the error caused by off-grid targets, in this paper, a more accurate sparse approximation model is applied where the true and unknown sparsifying dictionary is approximated with its first-order Taylor expansion around a known dictionary. Based on the model, the counting and localization problem is formulated as a sparse reconstruction problem that recovers two sparse signals with related supports. To solve the problem, we develop an iterative algorithm under a variational Bayesian inference framework. Moreover, in practice, some faulty prior information (e.g., coarse positions) is usually available. To incorporate the information, a three-level hierarchical prior model is introduced and imposed on the sparse signals to be recovered, where the supports of sparse signals are learned in the third level. Additionally, a grid pruning mechanism and matrix inversion lemma are leveraged to accelerate the inference process. Extensive simulation results demonstrate that, compared with state-of-the-art algorithms, the proposed algorithm can achieve counting and localization with higher accuracy as well as lower cost.**

***Index Terms*— Wireless sensor networks, counting and localization, faulty prior information, variational Bayesian inference.**

I. Introduction

WIRELESS sensor networks (WSNs) [1] have been widely deployed to acquire some physical data, such as temperature, humidity and acceleration, etc. Localization is an indispensable and fundamental component for WSNs since the data sensed by sensors will be meaningless without location information, attracting increasingly considerable attention in the past few years. As a popular system, Global Positioning System (GPS) [2] has been broadly used in daily life. However, it may be infeasible in WSNs as they are usually deployed in complicated environments (e.g., indoor or underground). What's more, it is extremely expensive to equip each target a GPS receiver in WSNs.

The limitations of GPS have promoted researchers to pursuit more feasible localization methods for WSNs. As a result, a large number of solutions have been developed, which can be divided into two classes: range-based and range-free. Range-based methods [3]–[6] first measure one or some (hybrid) physical features of the radio signal transmitted between sensors and targets, and then determine the positions of targets using trilateration, triangulation or maximum likelihood. Typical physical features include *Time of Arrival* (TOA), *Time Difference of Arrival* (TDOA), *Direction of Arrival* (DOA) and *Received Signal Strength* (RSS). These methods are accurate but susceptible to fading, noise and non-line-of-sight. In contrast, range-free methods [7]–[9] achieve localization by exploiting the connectivity or proximity relationship between targets and sensors. These methods are easy to implement but usually inaccurate and sensitive to the density of sensors. Moreover, there is a common weakness for range-based and range-free methods that they both require to collect mass data to obtain signal features or determine connectivity. Unfortunately, as we know, sensors are usually resource-constrained since they are powered by batteries, leading the above solutions unsuitable. Therefore, a meaningful and challenging problem arises that how to achieve accurate localization with a few data. It should be noted that though an adaptive RF fingerprint sampling scheme [10] is proposed to cut down the amount of samples, the method can not be applied in our context since we assume that the RSS readings from different targets superpose together and can not be distinguished by sensors.

As a new signal acquirement technique, compressive sensing (CS) [11], [12] shows that fewer samples will suffice for recovering sparse signals than the traditional Nyquist sampling theorem. In fact, the signal in localization problem is naturally sparse since the number of targets is usually limited. Therefore, CS attracts extensive attention for localization and many CS-based localization methods (see Section II) have been proposed. Among these methods, the continuous physical space is sampled into a discrete and fixed grid. A common assumption is made that all targets fall on the grid precisely. Based on the assumption, the RSS measurements can be sparsely represented in a sparsifying dictionary corresponding

Manuscript received February 4, 2017; revised June 27, 2017 and September 11, 2017; accepted October 26, 2017. Date of publication November 7, 2017; date of current version March 15, 2018. This work is supported by the National Natural Science Foundation of China (61571463, 61371124 and 61472445) and the Jiangsu Province Natural Science Foundation (BK20171401). The associate editor coordinating the review of this paper and approving it for publication was A. Nallanathan. *(Corresponding author: Baoming Sun.)*

Y. Guo, B. Sun, and N. Li are with the College of Communications Engineering, PLA University of Science and Technology, Nanjing 210007, China (e-mail: guoyan_1029@sina.com; baomings1988@sina.com).

D. Fang is with the School of Electronic and Optical Engineering, Nanjing University of Science and Technology, Nanjing 210094, China

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCOMM.2017.2770139

to the grid and the representation coefficient encodes the number and positions of multiple targets. As a result, counting and localization can be implemented by reconstructing the sparse signal and detecting its support. However, as a matter of fact, it is almost impossible to pre-define such a fixed grid so that all targets fall on the grid precisely, this is because we do not know the true number and positions of multiple targets in advance. The presence of off-grid targets will result in mismatch between the adopted and true sparsifying dictionaries. This is the so-called "dictionary mismatch" problem. Existing researches show that even a small dictionary mismatch may degrade the reconstruction performance of compressed sensing dramatically [27]–[30]. To handel this, some researchers tried to sample the physical space into a denser grid. However, these trials do not provide a thorough solution to the dictionary mismatch problem. In fact, a denser grid may result in higher correlation between the atoms of sparsifying dictionary. The high correlation may violate the restricted isometry property (RIP) [31] of sparsifying dictionary and lead to poor reconstruction performance. Furthermore, in [32], the authors solved the dictionary mismatch problem by leveraging the framework of atomic norm minimization. However, the algorithm is only suitable for reconstructing frequency-sparse signals but can not be applied in our context. Additionally, in [33], the authors studied dictionary mismatch problem in DOA estimation. To address the problem, the true array manifold matrix was approximated using its first order expansion at a fixed grid and an algorithm was proposed from the Bayesian perspective. However, the method assume that the model order is known in advance.

It should be noted that, in practice, some additional prior information is usually available. For example, in mobile networks, the positions of targets in the previous moment may provide coarse estimation for localization at present. Though the information may be partial and faulty, it can be utilized to improve localization accuracy. To the best knowledge of the authors, leveraging faulty prior information to achieve accurate counting and localization for off-grid targets has not been reported in literature so far. Motivated by this, we study the counting and localization problem for off-grid targets when faulty prior information is available. The main contributions of this paper are summarized as follows:

- To alleviate the error caused by off-grid targets, a more accurate sparse approximation model is applied. In the model, the true sparsifying dictionary is approximated by its first order Taylor expansion around the dictionary formed by a pre-defined grid. Based on the model, the counting and localization problem is formulated as a sparse recovery problem that recovers two sparse vectors. More importantly, the supports of two sparse vectors are related and encode the number and positions of multiple targets.
- To solve the above sparse recovery problem, an iterative algorithm is developed in a variational Bayesian inference framework. To integrate faulty prior information into the framework, a three-level hierarchical prior model is imposed on the sparse signals to be recovered. Based on the model, the supports of two sparse signals are leaned in the third level.
- To improve the convergence speed of the proposed algorithm, a grid pruning mechanism and matrix inversion lemma are introduced into the variational Bayesian inference framework. Based on these operations, the signal dimension and computational cost are reduced significantly.
- Extensive simulations are conducted to evaluate the counting and localization performance of the proposed algorithm. The superiority of our algorithm compared with the existing CS-based reconstruction algorithms is validated by the simulation results.

The remainder of this paper is organized as follows. Section 2 presents the signal model used in this paper and formulates the problem mathematically. Section 3 continues with a particular description of the proposed algorithm. In section 4, extensive numerical simulations are conducted to confirm the superiority of our algorithm. Finally, the conclusions are summarized in Section 5.

Notations used in this paper are as follows. Capital boldface letters and lowercase boldface letters are reserved for matrices and vectors respectively. $|\cdot|$ denotes the absolute value of a scalar or the cardinality of a set. $\langle\cdot\rangle$ denotes the expectation of a stochastic variable. $\|\cdot\|_p$ represents the ℓ_p norm of a vector. For a given matrix $\mathbf{A}$, $\mathbf{A}^{-1}$, $\mathbf{A}^{\mathrm{T}}$, $\mathrm{tr}(\mathbf{A})$ and $[\mathbf{A}]_{ij}$ denote the inverse, transpose, trace, and (i, j)-th entry of $\mathbf{A}$, respectively.

II. Related Work

Since CS was been proposed, the technology has been widely used to achieve counting and localization for targets. In the section, we will review the existing CS-based localization methods.

In [13], CS was utilized to localize multiple targets and a CS-based localization framework was developed for the first time. The unknown target positions on a discrete grid are modeled as a sparse signal, whose support encodes the positions of targets. Therefore, the localization problem is formulated as a sparse signal reconstruction problem. However, the weakness of the framework is that each sensor must maintain a dictionary and a large number of inter-sensor communications are required. In [14], a Bayesian framework was introduced for localization and the sparse approximation to its optimal solution was provided. Furthermore, 1-bit compressive sensing was used to reduce the amount of inter-sensor communications. Meanwhile, in [15], the positions of multiple targets were modeled as a sparse matrix and estimated by sequentially reconstructing the columns of the matrix. However, the data compression is not sufficient and the number of targets must be known in advance. Unlike [15], the authors in [16] modeled the positions of multiple targets as a sparse vector, which encodes the number and positions of these targets. Additionally, a greedy matching pursuit (GMP) algorithm was developed to reconstruct the sparse vector with a high probability. In [17]–[19], a CS-based indoor positioning system was developed. The system first obtains a coarse estimate using proximity constraint in the coarse step, and then gets a refined estimate using compressive sensing in the

fine step. Recently, a range-free CS-based localization method was proposed in [20]. Instead of measuring signal features directly, the method achieves localization with the proximation relationship between targets and sensors. In parallel, in [21], a CS-based localization method was developed where deterministic sensing matrices instead of random sensing matrices are designed to conduct RSS measurements. Wu *et al.* [22] developed a new counting and localization system using online compressive sensing. In the system, RSS values are recorded at runtime, then the number and positions of targets are recovered immediately based on few RSS readings. To circumvent channel mismatch problem, in [23], a multi-channel localization system was proposed where sparse Bayesian learning (SBL) is applied to improve localization accuracy. The conventional single measurement vector (SMV) model is infeasible in time-varying environments. To overcome this, Lagunas *et al.* [24] developed a counting and localization method by exploiting the joint sparsity of the multiple measurement vector (MMV) model. More recently, Xue *et al.* [25] for the first time applied SBL to improve localization accuracy by exploiting the joint posterior distribution of model parameters. Inspired by the sparse and compressible natures of signals in both spatial domain and temporal domain, Sun *et al.* [26] introduced a two-dimensional localization framework using compressive sensing, which consists of a spatial localization module and a temporal localization module.

III. Problem Formulation

A. Signal Model

Without loss of generality, we consider to count and localize multiple targets in a two-dimensional region. These targets are equipped with radio transmitting devices that broadcast beacons periodically. Simultaneously, a number of sensors are randomly deployed in the region to collect the beacons and calculate the received signal strength (RSS) values. As a matter of fact, the different signals from all targets always superpose together at sensors. To capture such a fact, we model each RSS value as the sum of the strengths of all received signals by a sensor. Specially, denote by $\mathbf{T} = \{\boldsymbol{t}_m\}_{m=1}^{M}$ and $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^{K}$ the positions of sensors and targets respectively, where M and K represent the numbers of sensors and targets respectively, $\boldsymbol{t}_m$ and $\boldsymbol{\theta}_k$ represent the positions of the m-th sensor and the k-th target respectively. Then, the RSS value of the m-th sensor can be represented as

$$\begin{aligned} y_m &= \sum_{k=1}^{K} a_k f\left(\boldsymbol{t}_m, \boldsymbol{\theta}_k\right) + \varepsilon_m \\ &= s_m + \varepsilon_m, \quad m = 1, \cdots, M \end{aligned} \tag{1}$$

where a_k and ε_m denote transmitted power and additive noise, $f\left(\boldsymbol{t}_m, \boldsymbol{\theta}_k\right)$ denotes an energy attenuation function which is determined by the path loss model presented in [34]:

$$f\left(\boldsymbol{t}_m, \boldsymbol{\theta}_k\right) = \begin{cases} 1, & \text{if } d < d_0 \\ \dfrac{1}{(d/d_0)^{\gamma}}, & \text{otherwise} \end{cases} \tag{2}$$

where $d = \|\boldsymbol{t}_m - \boldsymbol{\theta}_k\|_2$ denotes the distance between the m-th sensor and the k-th target; d_0 means the reference distance; and finally, γ represents the path loss coefficient determined by environments. Typically, the value of γ ranges from 2 to 5.

Define

$$\mathbf{d}\left(\boldsymbol{\theta}_k\right) = \left[f\left(\boldsymbol{t}_1, \boldsymbol{\theta}_k\right), \cdots, f\left(\boldsymbol{t}_M, \boldsymbol{\theta}_k\right)\right]^{\mathrm{T}}, \tag{3}$$

$$\mathbf{D}\left(\boldsymbol{\Theta}\right) = \left[\mathbf{d}\left(\boldsymbol{\theta}_1\right), \cdots, \mathbf{d}\left(\boldsymbol{\theta}_K\right)\right]. \tag{4}$$

Then, the signal model in (1) can be expressed in vector-matrix form as

$$\mathbf{y} = \mathbf{D}\left(\boldsymbol{\Theta}\right) \cdot \mathbf{a} + \boldsymbol{\varepsilon}, \tag{5}$$

with $\mathbf{y} = [y_1, \cdots, y_M]^{\mathrm{T}}$, $\mathbf{a} = [a_1, \cdots, a_K]^{\mathrm{T}}$, $\boldsymbol{\varepsilon} = [\varepsilon_1, \cdots, \varepsilon_M]^{\mathrm{T}}$.

For the sake of simplicity, we assume that the transmitted powers of all targets are the same and known in advance. Therefore, our goal is to estimate the number K and the positions $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_k\}_{k=1}^{K}$ of multiple targets from the noisy measurements $\mathbf{y}$. In the next subsection, we will describe the dictionary mismatch problem caused by off-grid targets and elaborate the sparse approximation model used to address the problem.

B. Sparse Approximation Model

As pointed out early, the traditional CS-based localization methods commonly assume that all targets fall exactly on a fixed grid $\tilde{\boldsymbol{\Theta}} = \left\{\tilde{\boldsymbol{\theta}}_n\right\}_{n=1}^{N}$, namely, $\boldsymbol{\Theta} \subset \tilde{\boldsymbol{\Theta}}$, where N denotes the number of grid points and $\tilde{\boldsymbol{\theta}}_n$ denotes the position of the i-th grid point. However, it is almost impossible to predefine such a grid since we do not know the true number and positions of multiple targets in advance. As a result, the performance of these methods deteriorates dramatically.

In fact, the model in the above methods can be considered as the zeroth order approximation of the true measurement model. In general cases, the approximation model is coarse so that the representation coefficient is not fully sparse or compressible. To address this problem, in this paper, we apply a more accurate sparse approximation model, which is the first order Taylor approximation of the true measurement model.

Denote by $\tilde{\boldsymbol{\theta}}_{I_k}$ the nearest grid point to the true target $\boldsymbol{\theta}_k$, then the first order Taylor approximation of $\mathbf{d}\left(\boldsymbol{\theta}_k\right)$ around $\tilde{\boldsymbol{\theta}}_{I_k}$ can be written as

$$\mathbf{d}\left(\boldsymbol{\theta}_k\right) \approx \mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_k}\right) + \mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_k}\right)\right)\left(\boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_{I_k}\right), \tag{6}$$

where $\mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_k}\right)\right)$ represents the Jacobian matrix of $\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_k}\right)$ with respect to $\tilde{\boldsymbol{\theta}}_{I_k}$. Then, the matrix $\mathbf{D}\left(\boldsymbol{\Theta}\right)$ and measurements $\mathbf{y}$ can be approximated as

$$\begin{aligned} \mathbf{D}(\boldsymbol{\Theta}) \approx & \left[\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_1}\right), \cdots, \mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_k}\right)\right] \\ & + \left[\mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_1}\right)\right), \cdots, \mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_K}\right)\right)\right] \\ & \cdot \operatorname{diag}\left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_{I_1}, \cdots, \boldsymbol{\theta}_K - \tilde{\boldsymbol{\theta}}_{I_K}\right), \end{aligned} \tag{7}$$

$$\begin{aligned} \mathbf{y} \approx & \left[\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_1}\right), \cdots, \mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_k}\right)\right] \cdot \mathbf{a} \\ & + \left[\mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_1}\right)\right), \cdots, \mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_{I_K}\right)\right)\right] \\ & \cdot \operatorname{diag}\left(\boldsymbol{\theta}_1 - \tilde{\boldsymbol{\theta}}_{I_1}, \cdots, \boldsymbol{\theta}_K - \tilde{\boldsymbol{\theta}}_{I_K}\right) \cdot \mathbf{a} + \boldsymbol{\varepsilon}. \end{aligned} \tag{8}$$

Define $\mathbf{J}\left(\tilde{\mathbf{\Theta}}\right) = \left[\mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_1\right)\right), \cdots, \mathbf{J}\left(\mathbf{d}\left(\tilde{\boldsymbol{\theta}}_N\right)\right)\right]$, $\mathbf{\Lambda} = \text{diag}\left(\boldsymbol{\lambda}_1, \boldsymbol{\lambda}_2, \cdots, \boldsymbol{\lambda}_N\right)$, and $\mathbf{w} = [w_1, \cdots, w_N]^{\mathrm{T}}$, where for $n = 1, \cdots, N$,

$$\boldsymbol{\lambda}_n = \begin{cases} \boldsymbol{\theta}_k - \tilde{\boldsymbol{\theta}}_{I_k}, & \text{if } n = I_k \quad \forall k \in \{1, \cdots, K\} \\ \mathbf{0}, & \text{otherwise.} \end{cases} \tag{9}$$

$$w_n = \begin{cases} a_k, & \text{if } n = I_k \quad \forall k \in \{1, \cdots, K\} \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

Then, the approximated signal model in (8) can be written as

$$\mathbf{y} \approx \left[\mathbf{D}\left(\tilde{\mathbf{\Theta}}\right) + \mathbf{J}\left(\tilde{\mathbf{\Theta}}\right)\mathbf{\Lambda}\right]\mathbf{w} + \boldsymbol{\varepsilon}. \tag{11}$$

Let $\mathbf{v} = \mathbf{\Lambda}\mathbf{w}$, then the sparse approximation model in (11) can be given as

$$\mathbf{y} \approx \mathbf{D}\left(\tilde{\mathbf{\Theta}}\right)\mathbf{w} + \mathbf{J}\left(\tilde{\mathbf{\Theta}}\right)\mathbf{v} + \boldsymbol{\varepsilon}. \tag{12}$$

The above sparse approximation model clearly shows that the measurements $\mathbf{y}$ can be sparsely approximated using two known dictionaries $\mathbf{D}\left(\tilde{\mathbf{\Theta}}\right)$ and $\mathbf{J}\left(\tilde{\mathbf{\Theta}}\right)$ with sparse representation coefficients $\mathbf{w}$ and $\mathbf{v}$. More importantly, according to definitions, the supports of $\mathbf{w}$ and $\mathbf{v}$ are related and encode the number and positions of multiple targets. In the next subsection, we will introduce the available and faulty prior information that can be used to learn the supports of $\mathbf{w}$ and $\mathbf{v}$ in WSNs.

C. Faulty Prior Information

In practice, some additional prior information is usually available and can be utilized to improve localization accuracy. For example, in mobile networks, once the target positions in the previous moment are available, we can obtain faulty prior information that which grid points are nearest to the true targets at present, namely the support of the sparse signal $\mathbf{w}$.

Mathematically speaking, we denote by $\mathcal{P}$ the faulty prior information on the support of the sparse signal $\mathbf{w}$. It should be pointed out that the prior information may be partial and erroneous. Therefore, we divide the prior information $\mathcal{P}$ into two parts: $\mathcal{P} = \mathcal{C} \cup \mathcal{E}$, where $\mathcal{C}$ denotes the subset containing correct information and $\mathcal{E}$ denotes the subset containing wrong information. Additionally, denote by $\mathcal{T}$ the true support of $\mathbf{w}$ and $\mathcal{T}^c$ the complement of the set $\mathcal{T}$, i.e. $\mathcal{T} \cup \mathcal{T}^c = \{1, 2, \cdots, N\}$, then it is easy to find that $\mathcal{C} \subset \mathcal{T}$, and $\mathcal{E} \subset \mathcal{T}^c$. It is noteworthy that the only prior information we have is $\mathcal{P}$, its division of $\mathcal{C}$ and $\mathcal{E}$ is unavailable for us. Hence, our goal is to simultaneously reconstruct sparse signals $\mathbf{w}$ and $\mathbf{v}$ with faulty prior information $\mathcal{P}$ from noisy measurements $\mathbf{y}$.

IV. Counting and Localization Via Variational Bayesian Inference With Support Learning

In this paper, we reconstruct the sparse representation coefficients $\mathbf{w}$ and $\mathbf{v}$ under the variational Bayesian inference framework. Before this, we first give a brief description on variational Bayesian inference and the proposed three-level hierarchical prior model.

A. Variational Bayesian Inference

As an efficient sparse reconstruction algorithm, the variational Bayesian inference [35] is gaining popularity within the signal processing field. In the methodology, the posteriors of hidden variables are inferred by maximizing a lower bound of the likelihood function which guarantees local convergence. This methodology allows inference in the case of complex signal models, that in certain cases provide significant improvements as compared to the other reconstruction algorithms.

Specifically, consider a model with observed variables $\mathbf{y}$, hidden variables $\mathbf{z}$ and unknown deterministic parameters $\boldsymbol{\theta}$. For any probability density function $q(\mathbf{z})$, the log-likelihood function can be written as

$$\ln p(\mathbf{y}; \boldsymbol{\theta}) = \int q(\mathbf{z}) \ln p(\mathbf{y}; \boldsymbol{\theta}) \, \mathrm{d}\mathbf{z} = \mathcal{F}(q, \boldsymbol{\theta}) + KL(q \parallel p), \tag{13}$$

with

$$\mathcal{F}(q, \boldsymbol{\theta}) = \int q(\mathbf{z}) \ln \left(\frac{p(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta})}{q(\mathbf{z})}\right) \mathrm{d}\mathbf{z}, \tag{14}$$

$$KL(q \parallel p) = -\int q(\mathbf{z}) \ln \left(\frac{p(\mathbf{z}|\mathbf{y}; \boldsymbol{\theta})}{q(\mathbf{z})}\right) \mathrm{d}\mathbf{z}, \tag{15}$$

where $KL(q \parallel p)$ is the Kullback-Leibler divergence between $p(\mathbf{z}|\mathbf{y}; \boldsymbol{\theta})$ and $q(\mathbf{z})$. Since $KL(q \parallel p) \geq 0$, it always holds that $\ln p(\mathbf{y}; \boldsymbol{\theta}) \geq \mathcal{F}(q, \boldsymbol{\theta})$, indicating that $\mathcal{F}(q, \boldsymbol{\theta})$ is a lower bound of $p(\mathbf{y}; \boldsymbol{\theta})$. Variational Bayesian inference is a two-step iterative algorithm where the lower bound $\mathcal{F}(q, \boldsymbol{\theta})$ is maximized with respect to $q(\mathbf{z})$ and $\boldsymbol{\theta}$ alternately. In the fist step, the lower bound $\mathcal{F}(q, \boldsymbol{\theta})$ is maximized with respect to $q(\mathbf{z})$. It is easy to see that this happens when $KL(q \parallel p) = 0$, which implies $q(\mathbf{z}) = p(\mathbf{z}|\mathbf{y}; \boldsymbol{\theta})$. In the second step, $q(\mathbf{z})$ is kept fixed and the lower bound $\mathcal{F}(q, \boldsymbol{\theta})$ is maximized with respect to $\boldsymbol{\theta}$.

However, in many complicated models, the posterior distributions of multiple interacting hidden variables are intractable to obtain. Fortunately, they are usually of a tractable and factorized form:

$$q(\mathbf{z}) = \prod_i q_i(z_i). \tag{16}$$

Denoting $q_j = q_j\left(z_j\right)$ and substituting (16) into (14), we have

$$\begin{aligned} \mathcal{F}(q, \boldsymbol{\theta}) &= \int q_j \ln \tilde{p}\left(\mathbf{y}, z_j; \boldsymbol{\theta}\right) dz_i - \int q_j \ln q_j dz_j \\ &\quad - \sum_{i \neq j} \int q_i \ln q_i dz_i \\ &= -KL\left(q_j \parallel \tilde{p}\right) - \sum_{i \neq j} \int q_i \ln q_i dz_i, \end{aligned} \tag{17}$$

where $\ln \tilde{p}\left(\mathbf{y}, z_j; \boldsymbol{\theta}\right) = \langle \ln p(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}) \rangle_{i \neq j}$ and $\langle \cdot \rangle_{i \neq j}$ denotes the expectation with respect to $q_i \, (i \neq j)$. Since $KL\left(q_j \parallel \tilde{p}\right) \geq 0$, it is easy to see that $\mathcal{F}(q, \boldsymbol{\theta})$ is maximized when $KL\left(q_j \parallel \tilde{p}\right) = 0$. Therefore, the optimal distribution for q_j (denoted by q_j^*) can be obtained as

$$\ln q_j^* = \langle \ln p(\mathbf{y}, \mathbf{z}; \boldsymbol{\theta}) \rangle_{i \neq j} + \text{const}. \tag{18}$$

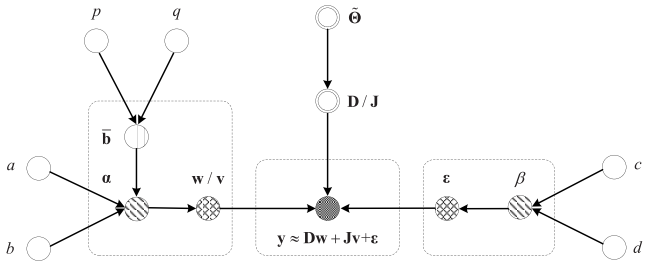


Fig. 1. Three-level hierarchical prior model for variational Bayesian inference.

To recover $\mathbf{w}$ and $\mathbf{v}$, we will model both of them as hidden variables. In fact, the dictionaries in model (12) are formed by a fixed grid. Consequently, (12) becomes a model containing only hidden variables and no parameters. In such a situation, variational Bayesian inference involves only the first step. It should be noted that the algorithm are still able to guarantee local convergence since the lower bound of log-likelihood function is improved in each iteration.

B. Hierarchical Prior Model

In the traditional variational Bayesian inference, a two-level hierarchical prior model is usually imposed on sparse coefficients to induce sparsity. Unlike the convention, to learn the supports of sparse coefficients $\mathbf{w}$ and $\mathbf{v}$, we expand the traditional two-level hierarchical prior model into a three-level hierarchical prior model [36]. Specially, the graphical model for variational Bayesian inference is shown in Fig. 1.

In the first level, we impose a Gaussian prior for $\mathbf{w}$ to induce its sparsity. Denote by α_i^{-1} the prior variance of w_i. Then, the prior distribution of $\mathbf{w}$ can be expressed as

$$p\left(\mathbf{w}|\boldsymbol{\alpha}\right) = \frac{1}{(2\pi)^{N/2}|\mathbf{A}|^{-1/2}} \exp\left(-\frac{1}{2}\mathbf{w}^{\mathrm{T}}\mathbf{A}\mathbf{w}\right), \tag{19}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_N]^{\mathrm{T}}$ and $\mathbf{A} = \mathrm{diag}\left(\alpha_1, \alpha_2, \cdots, \alpha_N.\right)$. For a uniform grid $\tilde{\boldsymbol{\Theta}}$, denote by Δ the spacing between two adjacent grid lines. According to (9), we have $\left[-\frac{1}{2}\Delta, -\frac{1}{2}\Delta\right]^{\mathrm{T}} \le \boldsymbol{\lambda}_n \le \left[\frac{1}{2}\Delta, \frac{1}{2}\Delta\right]^{\mathrm{T}}$. Then, by exploiting the relationship between the supports of $\mathbf{v}$ and $\mathbf{w}$, we set the prior variance of v_i to be $\frac{1}{4}\Delta^2\alpha_{\lceil i/2\rceil}^{-1}$ to accommodate the worst case where some target falls on the center of one lattice. Therefore, the prior distribution of $\mathbf{v}$ conditioned on $\boldsymbol{\alpha}$ can be written as

$$p\left(\mathbf{v}|\boldsymbol{\alpha}\right) = \frac{1}{(2\pi)^{N}|\mathbf{B}|^{-1/2}} \exp\left(-\frac{1}{2}\mathbf{v}^{\mathrm{T}}\mathbf{B}\mathbf{v}\right), \tag{20}$$

where $\mathbf{B} = 4\Delta^{-2}\mathrm{diag}\left(\boldsymbol{\alpha} \otimes \mathbf{1}_{2\times 1}\right)$. Similarly, the measurement noise $\boldsymbol{\varepsilon}$ is also assumed to be independent and Gaussian with zero-mean and variance β^{-1}, i.e.,

$$p\left(\boldsymbol{\varepsilon}|\beta\right) = \left(2\pi\beta^{-1}\right)^{-M/2} \exp\left(-\frac{\beta}{2}\left\|\boldsymbol{\varepsilon}\right\|_2^2\right). \tag{21}$$

In the second level, we impose a Gamma prior distribution to the random variables $\boldsymbol{\alpha}$ and β since Gamma distribution is conjugate to Gaussian distribution. Then, the prior distribution of $\boldsymbol{\alpha}$ and β can be given as

$$p(\boldsymbol{\alpha}; a, b_i) = \prod_{i=1}^{N} \mathrm{Gamma}\left(\alpha_i|a, b_i\right), \tag{22}$$

$$p\left(\beta; c, d\right) = \mathrm{Gamma}\left(\beta|c, d\right). \tag{23}$$

where $\mathrm{Gamma}\left(\cdot|a, b_i\right)$ denotes the Gamma distribution with parameters a and b_i.

In the third level, different from the traditional two-level hierarchical prior model where b_i is viewed as a deterministic variable, we divide the parameters $\{b_i\}$ into two subsets: $\{b_i\,|\forall i \in \mathcal{P}\}$ and $\{b_i\,|\forall i \in \mathcal{P}^c\}$, For the former, the parameters are still considered to be deterministic and assigned a very small value, i.e.

$$b_i = b, \quad \forall i \in \mathcal{P}^c. \tag{24}$$

For the latter, to learn the supports of sparse signals, we view them as random parameters and assign them Gamma distributions:

$$p\left(b_i; p, q\right) = \mathrm{Gamma}\left(b_i\,|p, q\right), \quad \forall i \in \mathcal{P}, \tag{25}$$

where p and q denote the parameters of Gamma distribution.

By doing so, the three-layer hierarchical prior model effectively integrates the faulty prior information into the variational Bayesian inference framework, and therefore has the ability to learn the true supports of sparse signals $\mathbf{w}$ and $\mathbf{v}$. In the next subsection, we will illustrate how to update $\mathbf{w}$ and $\mathbf{v}$ according to variational Bayesian inference.

C. Sparse Representation Coefficient Reconstruction

To reconstruct sparse representation coefficients $\mathbf{w}$ and $\mathbf{v}$, we model both of them as hidden variables. Based on (12) and (21), the likelihood function can be written as

$$p\left(\mathbf{y}|\mathbf{w}, \mathbf{v}, \beta\right) = \left(2\pi\beta^{-1}\right)^{-M/2} \cdot \exp\left(-\frac{\beta}{2}\left\|\mathbf{y} - \mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\mathbf{w} - \mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\mathbf{v}\right\|_2^2\right). \tag{26}$$

Define $\bar{\mathbf{b}} \triangleq \{b_i\,|i \in \mathcal{P}\}$. According to (18), the posterior distributions of $\mathbf{w}$, $\mathbf{v}$, $\boldsymbol{\alpha}$, β and $\bar{\mathbf{b}}$ can be evaluated as

$$\begin{aligned}\ln q\left(\mathbf{w}\right) &= \left\langle \ln p\left(\mathbf{y}, \mathbf{w}, \mathbf{v}, \boldsymbol{\alpha}, \beta, \bar{\mathbf{b}}\right)\right\rangle_{q(\mathbf{v})q(\boldsymbol{\alpha})q(\beta)q(\bar{\mathbf{b}})} + \mathrm{const} \\ &= \left\langle \ln p\left(\mathbf{y}|\mathbf{w}, \mathbf{v}, \beta\right) p\left(\mathbf{w}|\boldsymbol{\alpha}\right)\right\rangle_{q(\mathbf{v})q(\boldsymbol{\alpha})q(\beta)} + \mathrm{const}, \end{aligned} \tag{27}$$

$$\begin{aligned}\ln q\left(\mathbf{v}\right) &= \left\langle \ln p\left(\mathbf{y}, \mathbf{w}, \mathbf{v}, \boldsymbol{\alpha}, \beta, \bar{\mathbf{b}}\right)\right\rangle_{q(\mathbf{w})q(\boldsymbol{\alpha})q(\beta)q(\bar{\mathbf{b}})} + \mathrm{const} \\ &= \left\langle \ln p\left(\mathbf{y}|\mathbf{w}, \mathbf{v}, \beta\right) p\left(\mathbf{v}|\boldsymbol{\alpha}\right)\right\rangle_{q(\mathbf{w})q(\boldsymbol{\alpha})q(\beta)} + \mathrm{const}, \end{aligned} \tag{28}$$

$$\begin{aligned}\ln q\left(\boldsymbol{\alpha}\right) &= \left\langle \ln p\left(\mathbf{y}, \mathbf{w}, \mathbf{v}, \boldsymbol{\alpha}, \beta, \bar{\mathbf{b}}\right)\right\rangle_{q(\mathbf{w})q(\mathbf{v})q(\beta)q(\bar{\mathbf{b}})} + \mathrm{const} \\ &= \left\langle \ln p\left(\mathbf{w}|\boldsymbol{\alpha}\right) p\left(\mathbf{v}|\boldsymbol{\alpha}\right) p\left(\boldsymbol{\alpha}\,|a, \mathbf{b}\right)\right\rangle_{q(\mathbf{w})q(\mathbf{v})q(\bar{\mathbf{b}})} + \mathrm{const}, \end{aligned} \tag{29}$$

$$\begin{aligned}\ln q\left(\beta\right) &= \left\langle \ln p\left(\mathbf{y}, \mathbf{w}, \mathbf{v}, \boldsymbol{\alpha}, \beta, \bar{\mathbf{b}}\right)\right\rangle_{q(\mathbf{w})q(\mathbf{v})q(\boldsymbol{\alpha})q(\bar{\mathbf{b}})} + \mathrm{const} \\ &= \left\langle \ln p\left(\mathbf{y}|\mathbf{w}, \mathbf{v}, \beta\right) p\left(\beta\right)\right\rangle_{q(\mathbf{w})q(\mathbf{v})} + \mathrm{const}, \end{aligned} \tag{30}$$

$$\begin{aligned}\ln q\left(\bar{\mathbf{b}}\right) &= \left\langle \ln p\left(\mathbf{y}, \mathbf{w}, \mathbf{v}, \boldsymbol{\alpha}, \beta, \bar{\mathbf{b}}\right)\right\rangle_{q(\mathbf{w})q(\mathbf{v})q(\boldsymbol{\alpha})q(\beta)} + \mathrm{const} \\ &= \left\langle \ln p\left(\boldsymbol{\alpha}|a, \mathbf{b}\right) p\left(\bar{\mathbf{b}}\,|p, q\right)\right\rangle_{q(\boldsymbol{\alpha})} + \mathrm{const}. \end{aligned} \tag{31}$$

After some arrangement and simplification, we arrive at

$$q(\mathbf{w}) = \mathcal{N}(\mathbf{w}|\boldsymbol{\mu}_{\mathbf{w}}, \boldsymbol{\Sigma}_{\mathbf{w}}), \tag{32}$$

$$q(\mathbf{v}) = \mathcal{N}(\mathbf{v}|\boldsymbol{\mu}_{\mathbf{v}}, \boldsymbol{\Sigma}_{\mathbf{v}}), \tag{33}$$

$$q(\boldsymbol{\alpha}) = \prod_{i=1}^{N} \text{Gamma}\left(\alpha_i|\tilde{a}, \tilde{b}_i\right), \tag{34}$$

$$q(\beta) = \text{Gamma}\left(\beta|\tilde{c}, \tilde{d}\right), \tag{35}$$

$$q\left(\bar{\mathbf{b}}\right) = \prod_{i\in\mathcal{P}} \text{Gamma}\left(b_i|p, \tilde{q}_i\right), \tag{36}$$

with

$$\boldsymbol{\mu}_{\mathbf{w}} = \langle\beta\rangle \boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{D}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\left[\mathbf{y} - \mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\mu}_{\mathbf{v}}\right], \tag{37}$$

$$\boldsymbol{\Sigma}_{\mathbf{w}} = \left[\langle\beta\rangle \mathbf{D}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right) + \langle\mathbf{A}\rangle\right]^{-1}, \tag{38}$$

$$\boldsymbol{\mu}_{\mathbf{v}} = \langle\beta\rangle \boldsymbol{\Sigma}_{\mathbf{v}}\mathbf{J}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\left[\mathbf{y} - \mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\mu}_{\mathbf{w}}\right], \tag{39}$$

$$\boldsymbol{\Sigma}_{\mathbf{v}} = \left[\langle\beta\rangle \mathbf{J}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right) + \langle\mathbf{B}\rangle\right]^{-1}, \tag{40}$$

$$\tilde{a} = a + \frac{3}{2}, \tag{41}$$

$$\tilde{b}_i = \langle b_i\rangle + 2\Delta^{-2}\left(\left\langle v_{2i-1}^2\right\rangle + \left\langle v_{2i}^2\right\rangle\right) + \frac{1}{2}\left\langle w_i^2\right\rangle, \tag{42}$$

$$\tilde{c} = c + \frac{M}{2}, \tag{43}$$

$$\begin{aligned}\tilde{d} &= d + \frac{1}{2}\left\langle\left\|\mathbf{y} - \mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\mathbf{w} - \mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\mathbf{v}\right\|_2^2\right\rangle_{q(\mathbf{w})q(\mathbf{v})} \\ &= d + \frac{1}{2}\left\|\mathbf{y} - \mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\mu}_{\mathbf{w}} - \mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\mu}_{\mathbf{w}}\right\|_2^2 \\ &\quad + \frac{1}{2}\left\{\text{tr}\left(\mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\Sigma}_{\mathbf{w}}\mathbf{D}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\right) + \text{tr}\left(\mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\Sigma}_{\mathbf{v}}\mathbf{J}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\right)\right\},\end{aligned} \tag{44}$$

$$\tilde{q}_i = \langle\alpha_i\rangle + q. \tag{45}$$

In summary, the variational Bayesian inference obtains the posterior distributions of hidden variables $\mathbf{w}$, $\mathbf{v}$, $\boldsymbol{\alpha}$, β and $\bar{\mathbf{b}}$ by iteratively performing (32)−(36) until convergence. Some expectations used during the process are summarized as

$$\langle\alpha_i\rangle = \frac{\tilde{a}}{\tilde{b}_i}, \tag{46}$$

$$\langle\beta\rangle = \frac{\tilde{c}}{\tilde{d}}, \tag{47}$$

$$\langle b_i\rangle = \begin{cases}\dfrac{p}{\tilde{q}_i}, & i \in \mathcal{P} \\ b. & i \in \mathcal{P}^c\end{cases} \tag{48}$$

In the next subsection, we will demonstrate how to implement the above inference process in a faster way.

D. Fast Implementation

The above variational Bayesian inference is an iterative procedure. The computational burden in each iteration is dominated by the matrix inversions in (38) and (40) and the matrix-vector multiplications in (37) and (39). The complexities of these operations are $O\left(N^3\right)$ and $O\left(N^2\right)$, with N denoting the number of grid points. Therefore, the computational burden is significant when N is large. In fact, the number of grid points N is generally required to be large in order to reduce the error incurred by the sparse approximation model in (12).

We handle the above high computational burden problem from the following two aspects. On one hand, we utilize the matrix inversion lemma [37] to calculate the covariance matrices in (38) and (40), i.e.

$$\boldsymbol{\Sigma}_{\mathbf{w}} = \langle\mathbf{A}\rangle^{-1} - \langle\mathbf{A}\rangle^{-1}\mathbf{D}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\Sigma}_{\mathbf{D}}^{-1}\mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\langle\mathbf{A}\rangle^{-1}, \tag{49}$$

$$\boldsymbol{\Sigma}_{\mathbf{v}} = \langle\mathbf{B}\rangle^{-1} - \langle\mathbf{B}\rangle^{-1}\mathbf{J}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\Sigma}_{\mathbf{J}}^{-1}\mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\langle\mathbf{B}\rangle^{-1}, \tag{50}$$

where

$$\boldsymbol{\Sigma}_{\mathbf{D}} = \mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\langle\mathbf{A}\rangle^{-1}\mathbf{D}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right) + \langle\beta\rangle^{-1}\mathbf{I}, \tag{51}$$

$$\boldsymbol{\Sigma}_{\mathbf{J}} = \mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\langle\mathbf{B}\rangle^{-1}\mathbf{J}^{\mathrm{T}}\left(\tilde{\boldsymbol{\Theta}}\right) + \langle\beta\rangle^{-1}\mathbf{I}. \tag{52}$$

It should be pointed out that the inversions of $\langle\mathbf{A}\rangle$ and $\langle\mathbf{B}\rangle$ are easy to be obtained as they are diagonal matrices. By doing so, we must now only invert the $M \times M$ matrices $\boldsymbol{\Sigma}_{\mathbf{D}}$ and $\boldsymbol{\Sigma}_{\mathbf{J}}$, reducing the computational cost to $O\left(M^3\right)$, which is considerably desirable since the measurement number M is usually much smaller than the number of grid points N.

On the other hand, we apply the grid pruning mechanism presented in [38] and [39] to gradually reduce the number of grid points in the inference process. As a result, the computational burden will decrease with the reduction of the number of considered grid points. Specifically, we prune the current grid according to the element values in $\langle\boldsymbol{\alpha}\rangle$. From (37) and (38), it can be seen that when α_i is large enough, μ_{w_i} and $[\boldsymbol{\Sigma}_{\mathbf{w}}]_{ii}$ will approximate to zero, which indicates that the contribution of the i-th grid point to the measurements is negligible, and thus we remove this point from the current grid. The grid pruning results can be expressed as

$$\tilde{\boldsymbol{\Theta}}^{\text{NEW}} = \tilde{\boldsymbol{\Theta}}\left[\{i\,|\langle\alpha_i\rangle \leqslant \alpha_{th}\}\right], \tag{53}$$

where $\tilde{\boldsymbol{\Theta}}^{\text{NEW}}$ denotes the pruned grid and α_{th} denotes the threshold used to judge the elements in $\langle\boldsymbol{\alpha}\rangle$. It should also be pointed out that the corresponding posterior mean vectors and covariance matrices should also be pruned along with the grid pruning.

In the next subsection, we will exhibit the detailed steps of the proposed algorithm and how to achieve counting and localization according to the reconstructed $\mathbf{w}$ and $\mathbf{v}$.

E. Counting and Localization

Based on the above discussions, the sparse representation coefficients $\mathbf{w}$ and $\mathbf{v}$ can be reconstructed by performing the following steps:

1) *Initialization:* Set the iteration count $k = 0$ and divide the physical space into a uniform grid $\tilde{\boldsymbol{\Theta}}$. Set the initial values of deterministic variables a, b, c, d, p and q. Initialize the posterior distributions of $\mathbf{w}$ and $\mathbf{v}$.
2) *Posterior Updating:* Increase the iteration count k by 1. Update the posterior distributions of $\bar{\mathbf{b}}$ according to the current posterior distributions of $\boldsymbol{\alpha}$ via (45). Then, update the posterior distributions of $\boldsymbol{\alpha}$ and β according to the current posterior distributions of $\bar{\mathbf{b}}$, $\mathbf{w}$ and $\mathbf{v}$

via (41)-(44). At last, update the posterior distributions of $\mathbf{w}$ and $\mathbf{v}$ according to the current posterior distributions of $\boldsymbol{\alpha}$ and β via (37)-(40).

3) *Grid Pruning:* Prune the current grid according to $\langle\boldsymbol{\alpha}\rangle$. Meanwhile, prune the corresponding posterior mean vectors and covariance matrices.
4) *Termination Checking:* If the termination criterion is satisfied, perform Step 5; otherwise, go to Step 2.
5) *Output:* Output the sparse reconstruction results: $\hat{\mathbf{w}} \leftarrow \boldsymbol{\mu}_{\mathbf{w}}^{(k)}$, $\hat{\mathbf{v}} \leftarrow \boldsymbol{\mu}_{\mathbf{v}}^{(k)}$.

Remark 1: Note that we set a, b, c and d to very small values (e.g. $a = b = c = d = 10^{-6}$) to make the prior distributions uninformative and set p and q to relatively large values (e.g. $p = q = 1$) to impose a non-sparsity-encouraging prior on the corresponding coefficients. Initially, the prior distributions of $\boldsymbol{\alpha}$ and β are viewed as their posterior densities. The posterior distributions of $\mathbf{w}$ and $\mathbf{v}$ are initialized via (37)-(40).

Remark 2: The above algorithm terminates when the termination criterion is satisfied, for example when the iteration has reached the maximal allowable value k_{max} or the current residual reduction is smaller than some threshold η, namely $\left|r^{(k+1)} - r^{(k)}\right| < \eta$. The residual in the k-th iteration can be calculated as

$$r^{(k)} = \left\| \mathbf{y} - \mathbf{D}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\mu}_{\mathbf{w}}^{(k)} - \mathbf{J}\left(\tilde{\boldsymbol{\Theta}}\right)\boldsymbol{\mu}_{\mathbf{v}}^{(k)} \right\|_2^2. \tag{54}$$

As discussed early, the output results $\hat{\mathbf{w}}$ and $\hat{\mathbf{v}}$ encode the number and positions of multiple targets. According to (9) and (10), the number $\hat{K}$ and positions $\hat{\boldsymbol{\theta}}_k$ of the considered targets can be estimated as

$$\hat{K} = \#\left\{\hat{w}_i | \hat{w}_i > \delta, i = 1, \cdots, N\right\}, \tag{55}$$

$$\begin{aligned}\hat{\boldsymbol{\theta}}_k &= \tilde{\boldsymbol{\theta}}_{I_k} + \hat{\boldsymbol{\lambda}}_{I_k} \\ &= \tilde{\boldsymbol{\theta}}_{I_k} + \left[\hat{v}_{2I_k-1}, \hat{v}_{2I_k}\right] / \hat{w}_{I_k}.\end{aligned} \tag{56}$$

where δ denotes a sparsity threshold used to measure the values of $\hat{\mathbf{w}}$, and I_k denotes the index of the grid point nearest to the k-th target.

V. Simulations and Discussions

A. Simulation Setup

In this section, we conduct simulations with MATLAB R2015b 64-bit version running on Windows 10 64bit system. We randomly deploy $K = 9$ targets in a 2-dimensional square region with length of $19m$. The transmitted powers of all targets are assumed to be the same and set to $a_1 = a_2 = \cdots = a_9 = 100$ mW. Besides, M sensors are randomly deployed in the region to measure RSS values. Initially, we divide the region into a uniform grid with $N = 400$ grid points. As a result, the spacing between two adjacent grid lines is $\Delta = 1m$.

It is inevitable for measurements to be contaminated with additive noise. In order to check the robustness of the proposed algorithm, we intentionally add Gaussian white noise $\mathcal{N}\left(0, \sigma^2\right)$ to each measurement, where σ^2 denotes the variance of noise. We define the signal-to-noise ratio (SNR) as $10\log_{10}\left(\|\mathbf{s}\|_2^2 / \left(M\sigma^2\right)\right)$, where $\mathbf{s} = [s_1, s_2, \cdots, s_M]^{\mathrm{T}}$ denotes the RSS measurements of M sensors. The proposed algorithm is referred to as Variational Bayesian Inference with Support Learning (VBI-SL) and compared with the existing sparse reconstruction algorithms: (i) Basic Pursuit (BP) [40], (ii) Orthogonal Matching Pursuit (OMP) [41], (iii) Bayesian Compressive Sensing (BCS) [42], and (iv) Variational Bayesian Expectation Maximization (VBEM) [43]. The values of the parameters used in our simulations are summarized in Table I.

TABLE I
PARAMETER VALUES FOR SIMULATIONS

Parameters	Explains	Values
d_0	reference distance	1m
γ	path loss coefficient	2
α_{th}	pruning threshold	100
η	termination threshold	10^{-6}
k_{max}	maximum iteration number	10^3
δ	sparsity threshold	0.5

B. An Illustrative Example

In the first simulation, we apply different algorithms to a general case where $K = 9$ targets are randomly scattered in the region when SNR = 20 dB. We set $M = 100$, $|\mathcal{C}| = 6$ and $|\mathcal{E}| = 3$. We check the performance of different algorithms in terms of sparse reconstruction, counting and localization respectively.

The sparse reconstruction results by different algorithms are illustrated in Fig. 2 where blue circles and red asterisks respectively denote the true and reconstructed signals. As can be seen from Figs. 2(a)–(c), the reconstruction performance is quite poor when the traditional CS reconstruction algorithms (BP, OMP and BCS) are applied. This is because that the targets do not fall precisely on the pre-defined grids, incurring dictionary mismatch. Figs. 2(d)–(e) plot the sparse reconstruction results by VBEM and VBI-SL. As expected, their performance is significantly superior over the other three algorithms. The superiority can be attributed to the fact that they both apply a more accurate sparse approximation model than the others. Additionally, It can be seen that VBI-SL presents a light performance advantage over VBEM. The performance gain is primarily due to the fact that VBI-SL can learn the true supports of sparse signals from the faulty prior information while VBEM can not.

Correspondingly, Fig. 3 presents the counting and localization results by different algorithms. As for counting, it can be seen that false or missing targets are incurred when the traditional CS reconstruction algorithms are applied. As comparison, VBEM and VBI-SL are able to accurately estimate the number of targets in the region. From the perspective of localization, Fig. 3 clearly shows that the traditional CS reconstruction algorithms only localize some targets with significant errors. In contrast, the localization performance of VBEM is generally satisfactory with negligible errors. What's better, the positions estimated by VBI-SL are so accurate that they nearly superpose with the original positions. Based on these results, we can see that the proposed algorithm is more advantageous in counting and localizing multiple targets than the existing algorithms.

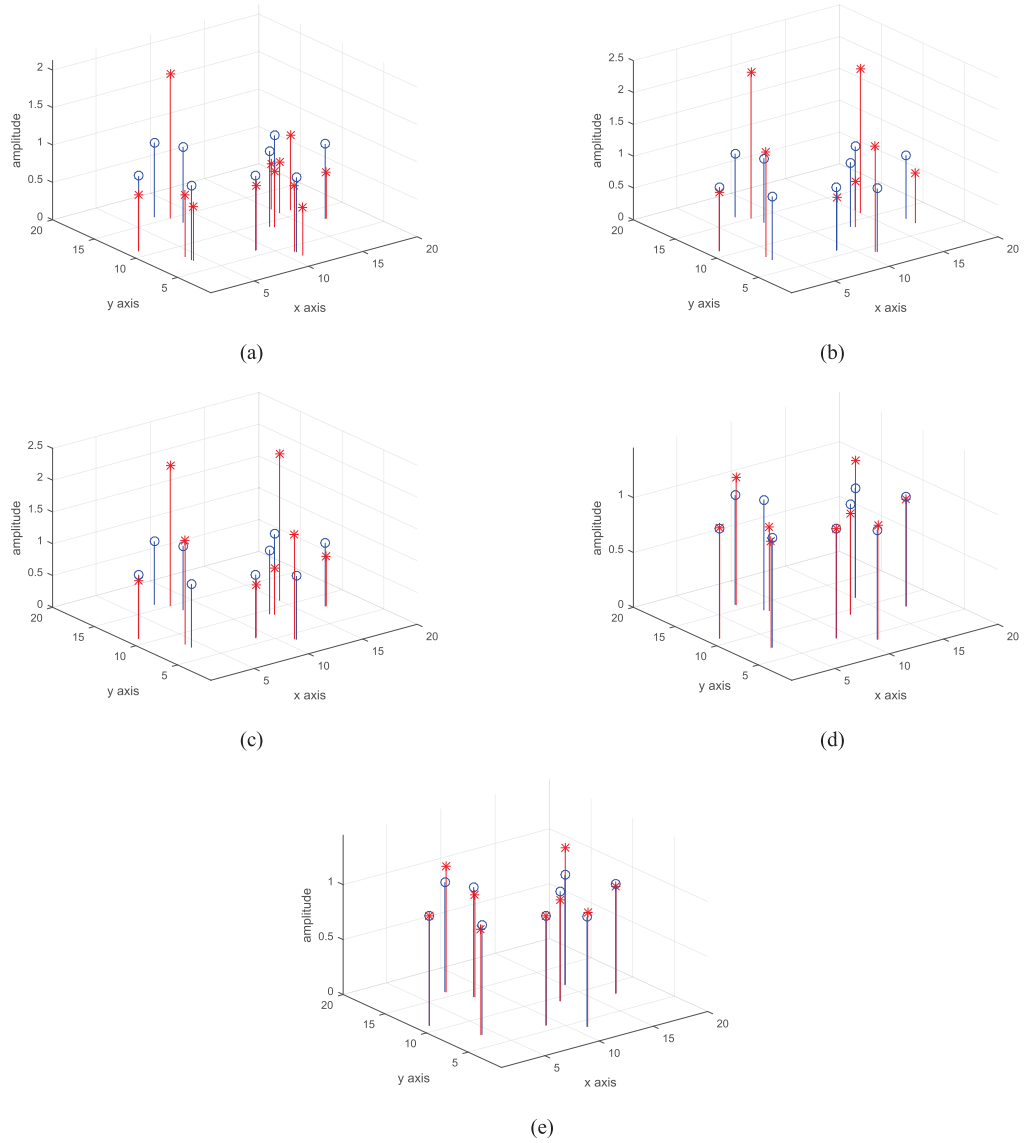


Fig. 2. Sparse reconstruction results by different algorithms when SNR = 20 dB. Blue circles and red asterisks denote the true and reconstructed signals.

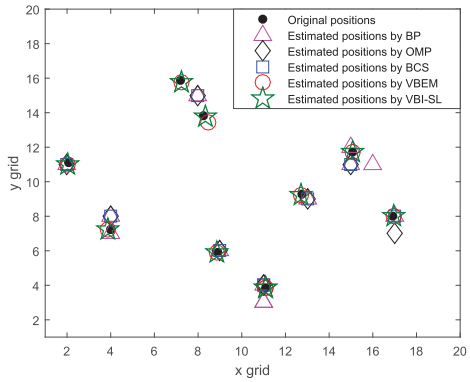


Fig. 3. Counting and localization results by different algorithms when SNR = 20 dB.

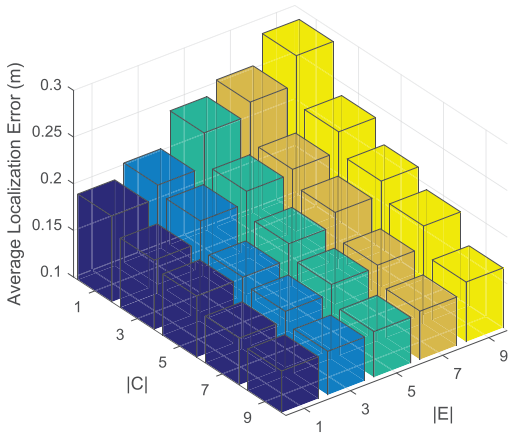


Fig. 4. Average localization error *vs.* amount of prior information when $M = 100$ and SNR=20 dB.

C. Effect of Amount of Information

Then, we check how the amount of prior information influences the reconstruction accuracy of the proposed algorithm. The amount of prior information includes the amount of correct prior information $|\mathcal{C}|$ and the amount of wrong prior information $|\mathcal{E}|$. We set $M = 100$ and SNR = 20 dB. The information amounts $|\mathcal{C}|$ and $|\mathcal{E}|$ vary from 1 to 9 at a

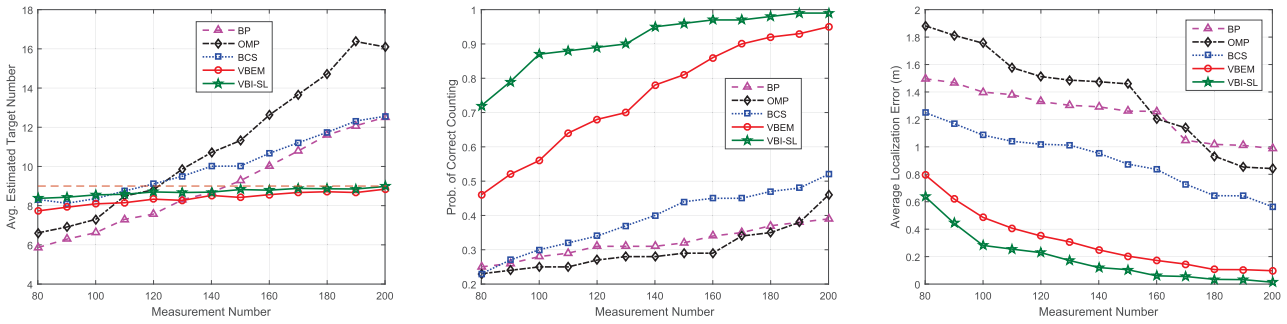


Fig. 5. Counting and localization performance *vs.* number of measurements M when SNR=20dB. (a) Average estimated target number; (b) Probability of correct counting; (c) Conditional average localization error.

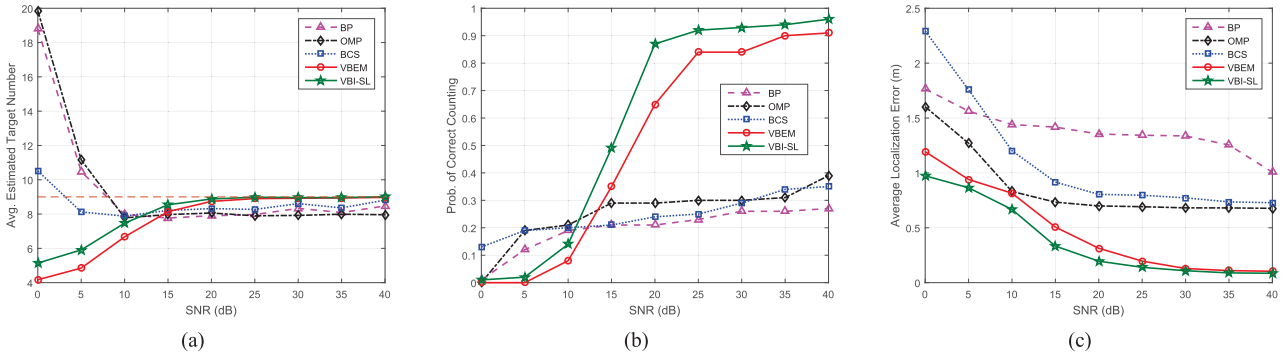


Fig. 6. Counting and localization performance *vs.* SNR when measurement number $M = 100$. (a) Average estimated target number; (b) Probability of correct counting; (c) Conditional average localization error.

step of 2. For each combine of $|\mathcal{C}|$ and $|\mathcal{E}|$, we conduct 100 independent trials. Fig. 4 plots the average localization errors conditioned on correct counting as a function of $|\mathcal{C}|$ and $|\mathcal{E}|$. As can be predicted, the conditional average localization error decreases with the increase of $|\mathcal{C}|$ and reduction of $|\mathcal{E}|$. It is rational because larger $|\mathcal{C}|$ (smaller $|\mathcal{E}|$) may provide more helpful information for localization. Specially, even in the worst case, namely when $|\mathcal{C}| = 1$ and $|\mathcal{E}| = 9$, the average localization error is no more than 0.3m. Based on these results, we can see that the proposed algorithm is robust to amount of wrong information.

D. Effect of Number of Measurement

We now focus on the impact of the number of measurements M on the counting and localization performance of the proposed algorithm. Figs. 5(a)–(c) plot the average estimated target number, the probability of correct counting and the average localization error conditioned on correct counting as a function of M. As can be seen, VBEM and VBI-SL perform significantly better than the traditional CS reconstruction algorithms. The counting and localization performance for the traditional CS reconstruction algorithms is considerably poor over the full M range while increasing M brings about obvious performance improvement for VBEM and VBI-SL. The reason accounting for the observation is the fact that the traditional CS reconstruction algorithms suffer from serious dictionary mismatch problem while VBEM and VBI-SL address the problem effectively by applying a more accurate sparse approximation model. Furthermore, it is obvious that VBI-SL is slightly inferior to VBEM for all M values. The performance gap is due to the fact that VBI-SL is able to learn the supports of sparse signals by utilizing the faulty prior information.

E. Robustness to Measurement Noise

Next, we check the robustness of the proposed algorithm to measurement noise. Figs. 6(a)–(c) illustrate the average estimated target number, the probability of correct counting and the average localization error conditioned on correct counting when SNR varies from 0 dB to 40 dB. As can be predicted, VBEM and VBI-SL extremely outperform the other algorithms in the entire SNR range, which can be explained by the fact that they both benefit from a more accurate sparse approximation model. In addition, compared with VBEM, the proposed algorithm VBI-SL achieves a small performance improvement as the results of support learning. Another key observation is that, when the traditional CS reconstruction algorithms are applied, the probabilities of correct counting keep below 40% no matter how high SNR is. However, increasing SNR results in larger probability of correct counting for VBEM and VBI-SL. This is because the traditional CS reconstruction algorithms suffer from dictionary mismatch while VBEM and VBI-SL do not. As can be seen, when SNR $=$ 20 dB, VBI-SL can achieve correct counting with a probability of 88% and localization with an average error of 0.2m. The above results distinctly show that the proposed algorithm VBI-SL can tolerate a certain level of measurement noise.

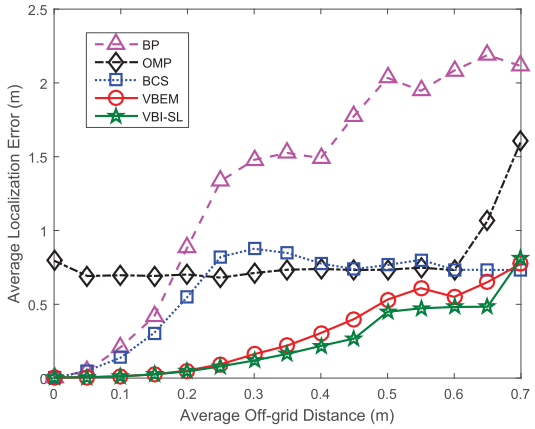


Fig. 7. Average localization error *vs.* grid mismatch level.

F. Sensitivity to Grid Mismatch

Then, we test the sensitivity of the VBI-SL algorithm to grid mismatch. We choose the average off-grid distance ΔR (the distance from the true target to the nearest grid point) to evaluate grid mismatch level. The larger the distance, the more serious the grid mismatch. We set simulation parameters to $M = 100$, $K = 9$ and SNR $= 20$ dB. In this simulation, we vary ΔR from 0 to 0.7; for each ΔR, 100 independent trials are carried out.

Fig. 7 plots the average localization errors as a function of grid mismatch level for different algorithms. It can be clearly seen that, compared to the other algorithms, the VBI-SL algorithm is much more advantageous throughout the grid mismatch level. Moreover, as the off-grid distance increases, the average localization errors of the other algorithms increase sharply or keep at a large level while the VBI-SL algorithm increases slowly. Furthermore, the average localization error of the VBI-SL algorithm is smaller than 0.5m even if the off-grid distance reaches up to 0.5m. The result indicates the robustness of the proposed algorithm to grid mismatch.

G. Complexity Analysis

Finally, we focus on the complexities of the proposed algorithm. For BP, OMP, BCS and VBEM, their computational complexities are $O\left(N^3\right)$, $O\left(NK^2\right)$, $O\left(NM^2\right)$ and $O\left(N^3\right)$, where K, M and N denote the numbers of targets, sensors and grid points respectively. As discussed above, the complexity of the proposed algorithm VBI-SL is reduced from $O\left(N^3\right)$ to $O\left(M^3\right)$ by using matrix inversion lemma. Moreover, Table II lists the conditioned average localization errors and average CPU running times (over 100 trials) of different algorithms when $M = 100$ and SNR $= 20$ dB. The results distinctly demonstrate that the VBEM and VBI-SL are much more accurate than the other algorithms with longer running time. As a matter of fact, the longer running time can be attributed to the fact that both VBEM and VBI-SL are iterative algorithms and they usually converge until numerous iterations. Meanwhile, the running efficiency of VBI-SL is remarkably advantageous over the VBEM algorithm. The efficiency improvement is due to the matrix inversion lemma and grid pruning mechanism.

TABLE II

AVERAGE LOCALIZATION ERRORS AND TOTAL CPU RUNNING TIMES (OVER 100 TRIALS) OF DIFFERENT ALGORITHMS

Algorithms	Localization error (m)	CPU running time (s)
BP	1.38	0.073
OMP	1.77	0.023
BCS	1.15	0.158
VBEM	0.48	76.868
VBI-SL	0.21	16.023

VI. CONCLUSIONS

In this paper, we addressed the counting and localization problem for off-grid targets with default prior information in WSNs. To alleviate the error caused by off-grid targets, a more accurate sparse approximation model was employed. Based on the model, the counting and localization problem was formulated as a sparse recovery problem that recovers two sparse support-related vectors. Then, we developed an iterative algorithm to solve the problem in a variational Bayesian inference framework. Additionally, a three-level hierarchical prior model was imposed on the sparse signals to learn their supports with default prior information. Experimental results demonstrated that the proposed algorithm can achieve accurate counting and localization with low computational cost. Meanwhile, it is robust to measurement noise and wrong prior information.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers for their constructive comments and Donoho *et al.* for sharing the source codes of related recovery algorithms online.

REFERENCES

[1] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "A survey on sensor networks," *IEEE Commun. Mag.*, vol. 40, no. 8, pp. 102–114, Aug. 2002.

[2] *Global Positioning System Standard Positioning Service Specification*, United States Coast Guard Navigation Center, Alexandria, VA, USA, Jun. 1995.

[3] G. Wang and K. Yang, "A new approach to sensor node localization using RSS measurements in wireless sensor networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 5, pp. 1389–1395, May 2011.

[4] I. Guvenc and C.-C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Commun. Surveys Tuts.*, vol. 11, no. 3, pp. 107–124, 3rd Quart., 2009.

[5] H. Jamali-Rad and G. Leus, "Sparsity-aware multi-source TDOA localization," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4874–4887, Oct. 2013.

[6] D. Liu, K. Liu, Y. Ma, and J. Yu, "Joint TOA and DOA localization in indoor environment using virtual stations," *IEEE Commun. Lett.*, vol. 18, no. 8, pp. 1423–1426, Aug. 2014.

[7] N. Bulusu, J. Heidemann, and D. Estrin, "GPS-less low-cost outdoor localization for very small devices," *IEEE Pers. Commun.*, vol. 7, no. 5, pp. 28–34, Oct. 2000.

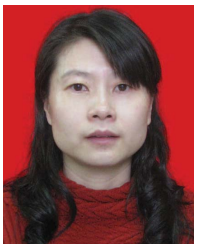
[8] T. He, C. Huang, B. Blum, J. Stankovic, and T. Abdelzaher, "Range-free localization schemes for large scale sensor networks," in *Proc. 9th Annu. Int. Conf. Mobile Comput. Netw.*, 2003, pp. 81–95.

[9] Y. Shang, W. Ruml, Y. Zhang, and M. Fromherz, "Localization from mere connectivity," in *Proc. 4th ACM Int. Symp. Mobile Ad Hoc Netw. Comput.*, 2003, pp. 201–212.

[10] X. Y. Liu, S. Aeron, V. Aggarwal, X. Wang, and M.-Y. Wu, "Adaptive sampling of RF fingerprints for fine-grained indoor localization," *IEEE Trans. Mobile Comput.*, vol. 15, no. 10, pp. 2411–2423, Oct. 2016.

[11] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.

[12] R. Baraniuk, "Compressive sensing," *IEEE Signal Process. Mag.*, vol. 24, no. 4, pp. 1–9, Apr. 2007.
[13] V. Cevher, M. F. Duarte, and R. G. Baraniuk, "Distributed target localization via spatial sparsity," in *Proc. 16th Eur. IEEE Signal Process. Conf.*, Aug. 2008, pp. 1–5.
[14] V. Cevher, P. Boufounos, R. G. Baraniuk, A. C. Gilbert, and M. J. Strauss, "Near-optimal Bayesian localization via incoherence and sparsity," in *Proc. IEEE Comput. Soc. Int. Conf. Inf. Process. Sensor Netw. (IPSN)*, 2009, pp. 205–216.
[15] C. Feng, S. Valaee, and Z. Tan, "Multiple target localization using compressive sensing," in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM)*, Dec. 2009, pp. 1–6.
[16] B. Zhang, X. Cheng, N. Zhang, Y. Cui, Y. Li, and Q. Liang, "Sparse target counting and localization in sensor networks based on compressive sensing," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 2255–2263.
[17] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Compressive sensing based positioning using RSS of WLAN access points," in *Proc. IEEE INFOCOM*, Mar. 2010, pp. 1–9.
[18] C. Feng, W. S. A. Au, S. Valaee, and Z. Tan, "Received-signal-strength-based indoor positioning using compressive sensing," *IEEE Trans. Mobile Comput.*, vol. 11, no. 2, pp. 1983–1993, Dec. 2012.
[19] A. W. S. Au *et al.*, "Indoor tracking and navigation using received signal strength and compressive sensing on a mobile device," *IEEE Trans. Mobile Comput.*, vol. 12, no. 10, pp. 2050–2062, Oct. 2013.
[20] L. Liu, T. Cui, and W. Lv, "A range-free multiple target localization algorithm using compressive sensing theory in wireless sensor networks," in *Proc. IEEE 11th Int. Conf. Mobile Ad Hoc Sensor Syst. (MASS)*, Oct. 2014, pp. 690–695.
[21] T. L. N. Nguyen and Y. Shin, "Multiple target localization in WSNs based on compressive sensing using deterministic sensing matrices," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 8, Aug. 2015, Article ID 947016.
[22] D. Wu, D. I. Arkhipov, Y. Zhang, C. H. Liu, and A. C. Regan, "Online war-driving by compressive sensing," *IEEE Trans. Mobile Comput.*, vol. 14, no. 11, pp. 2349–2362, Nov. 2015.
[23] S. Nikitaki, G. Tsagkatakis, and P. Tsakalides, "Efficient multi-channel signal strength based localization via matrix completion and Bayesian sparse learning," *IEEE Trans. Mobile Comput.*, vol. 14, no. 11, pp. 2244–2256, Nov. 2015.
[24] E. Lagunas, S. K. Sharma, B. Ottersten, and S. Chatzinotas, "Compressive sensing based target counting and localization exploiting joint sparsity," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2016, pp. 3231–3235.
[25] B. Xue, L. Zhang, and Y. Yu, "Multi-target localization based on sparse Bayesian learning in wireless sensor networks," *IEICE Trans. Commun.*, vol. 99, no. 5, pp. 1093–1100, May 2016.
[26] B. Sun, Y. Guo, N. Li, L. Peng, and D. Fang, "TDL: Two-dimensional localization for mobile targets using compressive sensing in wireless sensor networks," *Comput. Commun.*, vol. 78, pp. 45–55, Mar. 2016.
[27] M. Herman and T. Strohmer, "General deviants: An analysis of perturbations in compressed sensing," *IEEE J. Sel. Topics Signal Process.*, vol. 4, no. 2, pp. 342–349, Feb. 2010.
[28] Y. Chi, L. Scharf, and A. Pezeshki, "Sensitivity to basis mismatch in compressed sensing," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2182–2195, May 2011.
[29] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.
[30] Z. Yang, C. Zhang, and L. Xie, "Robustly stable signal recovery in compressed sensing with structured matrix perturbation," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4658–4671, Sep. 2012.
[31] E. J. Candès, "The restricted isometry property and its implications for compressed sensing," *Comput. Rendus Math.*, vol. 346, nos. 9–10, pp. 589–592, May 2008.
[32] G. Tang, B. N. Bhaskar, P. Shah, and B. Recht, "Compressed sensing off the grid," *IEEE Trans. Inf. Theory*, vol. 59, no. 11, pp. 7465–7490, Nov. 2013.
[33] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse Bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan. 2013.
[34] T. Clouqueur, K. K. Saluja, and P. Ramanathan, "Fault tolerance in collaborative sensor networks for target detection," *IEEE Trans. Comput.*, vol. 53, no. 3, pp. 320–333, Mar. 2004.
[35] D. G. Tzikas, A. C. Likas, and N. P. Galatsanos, "The variational approximation for Bayesian inference," *IEEE Signal Process. Mag.*, vol. 25, no. 6, pp. 131–146, Nov. 2008.
[36] J. Fang, Y. Shen, F. Li, H. Li, and Z. Chen, "Support knowledge-aided sparse Bayesian learning for compressed sensing," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 3786–3790.
[37] D. P. Wipf and B. D. Rao, "Sparse Bayesian learning for basis selection," *IEEE Trans. Signal Process.*, vol. 52, no. 8, pp. 2153–2164, Aug. 2004.
[38] L. Hu, Z. Shi, J. Zhou, and Q. Fu, "Compressed sensing of complex sinusoids: An approach based on dictionary refinement," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3809–3822, Jul. 2012.
[39] L. Hu, J. Zhou, Z. Shi, and Q. Fu, "A fast and accurate reconstruction algorithm for compressed sensing of complex sinusoids," *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5744–5754, Nov. 2013.
[40] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, Jan. 1999.
[41] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, Dec. 2007.
[42] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, Jun. 2008.
[43] B. Sun, Y. Guo, N. Li, and D. Fang, "An efficient counting and localization framework for off-grid targets in WSNs," *IEEE Commun. Lett.* vol. 21, no. 4, pp. 809–812, Apr. 2017, doi: 10.1109/LCOMM.2016.2644659.



Yan Guo received the B.S. and M.S. degrees from the PLA Institute of Information Engineering, Zhengzhou, China, in 1993 and 1996, respectively, and the Ph.D. degree from Xidian University, Xi'an, China, in 2002. She has been a Visiting Scholar with Chonbuk National University, South Korea. She is currently a Professor with the Institute of Communication and Engineering, PLA University of Science and Technology, Nanjing, China. Her main research interests include smart antenna, compressive sensing, MIMO, and cognitive radio.

Baoming Sun received the B.S. degree from the School of Mechanical Engineering and Electronic Information, China University of Geosciences, Wuhan, China, in 2010, and the M.S. degree from the Institute of Communications Engineering, PLA University of Science and Technology, Nanjing, China, in 2013, where he is currently pursuing the Ph.D. degree with the College of Communications Engineering. His research interests include wireless sensor networks, signal processing, and compressive sensing.

Ning Li received the B.S. and M.S. degrees from the PLA Institute of Information Engineering, Zhengzhou, China, in 1989 and 1996, respectively. He is currently an Associate Professor with the Institute of Communication and Engineering, PLA University of Science and Technology, Nanjing, China. His main research interests include Ad hoc networks, digital beamforming, and cognitive radio.

Dagang Fang (SM'90–F'03–LF'14) received the speciality degree in electromagnetic field and microwave technology from the Graduate School, Beijing Institute of Posts and Telecommunications, Beijing, China, in 1966. He is currently a Professor with the School of Electronic Engineering and Opto-electronic Technology, Nanjing University of Science and Technology, Nanjing, China. His current interests are computational electromagnetics, microwave integrated circuits, and antennas and EM scattering. He is a Fellow of the Chinese Institute of Electronics. He was a recipient of the National Outstanding Teacher Award and the People's Teacher Medal. He is an Associate Editor of *The Chinese Journal*. He is on the editorial or review board of several international and Chinese journals.