Multi-Task Bayesian Compressive Sensing Exploiting Intra-Task Dependency

Qisong Wu, *Member, IEEE*, Yimin D. Zhang, *Senior Member, IEEE*,
Moeness G. Amin, *Fellow, IEEE*, and Braham Himed, *Fellow, IEEE*

Abstract—In this letter, we propose a multi-task compressive sensing algorithm for the reconstruction of clustered sparse entries based on hierarchical Bayesian framework. The algorithm is based on a paired spike-and-slab prior, which is extended to a general multi-task model. The proposed algorithm has the capability of modeling both inter-task and intra-task dependencies of the observation data. The latter is achieved by imposing a clustered prior on non-zero entries and finds applications in radar where targets exhibit spatial extent. Simulation results demonstrate that the proposed algorithm outperforms existing state-of-the-art group sparse Bayesian learning algorithms.

I. INTRODUCTION

Sparse signal recovery and the related compressive sensing (CS) problems have attracted significant attention in recent years [1], [2]. CS techniques have the capability of recovering signals from a small number of measurement samples with a high probability, given that the signals are sparse or can be sparsely represented in some known domain.

A typical CS model with a single measurement vector (SMV) $\mathbf{y} \in \mathcal{R}^K$ is given by

$$\mathbf{y} = \mathbf{\Phi}\mathbf{w} + \boldsymbol{\varepsilon}, \tag{1}$$

where $\mathbf{\Phi} \in \mathcal{R}^{K \times M}$, $K \ll M$, is a known dictionary matrix, and $\boldsymbol{\varepsilon} \in \mathcal{R}^K$ is an unknown zero-mean Gaussian noise vector. The objective of sparse signal reconstruction is to estimate the sparse weight vector $\mathbf{w} \in \mathcal{R}^M$ from $\mathbf{y}$. Any K columns of $\mathbf{\Phi}$ are assumed to be linearly independent to satisfy the unique representation property which ensures sparse signal recovery with a high probability [2].

In many applications, the weight vector $\mathbf{w}$ exhibits group sparsity. Such problems are described by a multi-task CS model as [3],

$$\mathbf{y}_l = \mathbf{\Phi}_l\mathbf{w}_l + \boldsymbol{\epsilon}_l, \quad l \in [1, \cdots, L]. \tag{2}$$

In this model, $\mathbf{w}_l$ has the same sparsity support, i.e., the respective positions of the non-zero entries are the same for different l. Denote w_{ml} as the mth element of $\mathbf{w}_l$. Then, the above group sparsity implies that w_{ml} share the same sparsity for all $l \in [1, \cdots, L]$, whereas their values are generally different. For some applications, $\mathbf{\Phi}_l$ may take the same value, i.e., $\mathbf{\Phi}_l = \mathbf{\Phi}, \forall l$. In this case, the model is also referred to as multiple measurement vector (MMV) [4].

For example, in direction-of-arrival (DOA) estimation using different polarizations, L represents the number of available polarizations, and the weight vector $\mathbf{w}_l$ represents the source spatial entries which have group sparsity for these polarizations [4], [5]. On the other hand, in multi-static synthetic aperture radar (SAR) and inverse SAR (ISAR) imaging, L denotes the number of available illuminators, and

Q. Wu, Y. D. Zhang, and M. G. Amin are with the Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA.

B. Himed is with the RF Technology Branch, Air Force Research Laboratory (AFRL/RYMD), WPAFB, OH 45433, USA.

$\mathbf{w}_l$ represents the sparse scene that demonstrates group sparsity across all bistatic pairs corresponding to the illuminators. In this case, the dictionary matrices differ for each bistatic pair [6], [7].

A number of algorithms have been proposed to recover group sparse signals. These algorithms include greedy-based algorithms, such as Block-OMP (BOMP) [8], and basis pursuit-based ones, such as Group Basis Pursuit (GBP) [9] and Group Lasso [10]. All these algorithms require information about block partition either explicitly or through setting regularization parameters.

Group sparse Bayesian learning algorithms form a different class of sparse signal reconstruction algorithms, which generally yield improved performance in such situations [11]–[14]. The multi-task compressive sensing (mt-CS) algorithm [12] provides solutions to a large class of group sparse problems. This approach is extended to effectively process complex-valued problems [15].

There is a different class of group sparsity in which non-zero entries are grouped and the number of groups is difficult to specify. A representative example is the case where sparse targets, e.g., vehicles or aircrafts, have an extended spatial occupancy, forming a cluster. In this case, its non-zero entries are clustered in a spatial region, but the exact size and shape are difficult to specify in advance [16], [17].

Sparse Bayesian learning algorithms are suited to handle this type of clustering problems because they have the flexibility of exploiting the underlying signal structures. For example, the block sparse Bayesian learning algorithm (BSBL) uses the intra-block correlation to improve the signal reconstruction performance [18], and the spatio-temporal sparse Bayesian learning algorithm recovers multichannel signal by extending and exploiting inter-channel correlation [19]. In addition, Bayesian group-sparse modeling based on variational inference (GS-VB) [14] was developed based on the Laplace prior to recover group sparse signals, whereas the work in [20] uses the spike-and-slab prior to recover sparse signal with the group structure. Also, an overlapping group sparse algorithm is developed to reconstruct sparse signals with the cluster structure [21], and a Bayesian variant was considered in [14].

In this letter, we consider problems where the above two types of group sparsity coexist. Such problems characterize, for example, the abovementioned multi-static SAR/ISAR imaging where the targets have spatial extent. Similar issues arise in multipath through-the-wall radar imaging (TWRI) of spatially extended targets. To solve such problems, we propose a novel algorithm, termed clustered multi-task Bayesian compressive sensing (clustered MT-BCS). This algorithm exploits a multi-task model, where (a) all tasks share the same or similar sparse support, i.e., they assume inter-task dependency, and (b) the non-zero entries in each task are clustered, i.e., they exhibit intra-task dependency. In the proposed approach, we first extend a paired spike-and-slab priors to form a generalized multi-task model and induce the relationship between tasks based on a hierarchical model. Inspired by the cluster and graphical model in the Clustered Sparse Solver via Markov Chain Monte Carlo (CluSS-MCMC) algorithm [22], we place a cluster prior to capture the intra-task dependencies. A Gibbs sampler scheme is proposed to implement the posterior inference. Since the hierarchical Bayesian model allows the estimation of prior parameters in an unsupervised manner, the proposed algorithm does not require any information regarding either

sparse prior or cluster prior.

Notations: We use lower-case (upper-case) bold characters to denote vectors (matrices). $p(\cdot)$ denotes the probability density function (pdf), and $\mathcal{N}(x|a,b)$ denotes that random variable x follows a Gaussian distribution with mean a and variance b. In addition, $(\cdot)^T$ denotes transpose, $\mathbf{I}_N$ denotes the $N \times N$ identity matrix, and $\circ$ denotes element-wise (Hadamard) multiplication.

II. Cluster Multi-task Bayesian Compressive Sensing

A. Paired Spike-and-Slab Prior

Consider a clustered sparse reconstruction problem with L tasks, each consisting of M entries. To encourage the group sparsity described in (2), we place a spike-and-slab prior to $\mathbf{w}_l$, i.e., [23]–[26]

$$p(\mathbf{w}_l|\boldsymbol{\pi},\boldsymbol{\beta}) = \prod_{i=1}^{M}\left[(1-\pi_i)\delta(w_{il}) + \pi_i\mathcal{N}(w_{il}|0,\beta_i^{-1})\right], \quad (3)$$

where π_i is the prior probability of a non-zero element, i.e., a large weight π_i corresponds to a high probability that the entry takes a non-zero value, whereas a small π_i tends to generate a zero entry. In addition, β_i is the precision (reciprocal of the variance) of Gaussian distribution and $\delta(x)$ is the Dirac delta function of x.

The presence of the delta function in (3) makes the inference troublesome. However, there exists a simple reparameterization of the spike-and-slab prior [23]–[26]. Specially, assume a Gaussian random vector $\boldsymbol{\theta}_l = [\theta_{1l},...,\theta_{Ml}]^T$ with $p(\boldsymbol{\theta}_l) = \prod_{i=1}^{M}\mathcal{N}(\theta_{il}|0,\beta_i^{-1})$, $l = 1,...,L$, and a Bernoulli random vector $\mathbf{z} = [z_1,...,z_M]^T$ with $p(\mathbf{z}) = \prod_{i=1}^{M}\text{Bern}(z_i|\pi_i)$, where $z_i = 1$ corresponds to a non-zero entry in the ith position. The product of these latent vectors $\boldsymbol{\theta}_l \circ \mathbf{z}$ forms a new random vector that follows the pdf in (3), i.e.,

$$\mathbf{w}_l = \boldsymbol{\theta}_l \circ \mathbf{z}. \quad (4)$$

The group sparsity is characterized by the same z_i for the ith position across the L tasks. On the other hand, scattering coefficients in the ith block $\boldsymbol{\theta}_{i\cdot} = [\theta_{i1},\cdots,\theta_{iL}]$ have different values in general. In SAR/ISAR and TWRI applications, these values differ due to angle-dependent scattering coefficients and multipath returns [?], [6], [16].

Standard spike-and-slab prior independently places Gaussian distribution $p(\boldsymbol{\theta}_l)$ and Bernoulli distribution $p(\mathbf{z})$ on $\boldsymbol{\theta}_l$ and $\mathbf{z}$, respectively. However, $\boldsymbol{\theta}_l$ and $\mathbf{z}$ are strongly correlated since their product interacts with the data and a paired spike-and-slab prior was proposed to enhance the performance of sparse reconstruction [20]. The following paired spike-and-slab prior $p(\boldsymbol{\theta}_l,\mathbf{z})$ is introduced,

$$p(\boldsymbol{\theta}_l,\mathbf{z}) = \prod_{i=1}^{M}\left[\mathcal{N}(\theta_{il}|0,\beta_i^{-1})\right]^{z_i}\pi_i^{z_i}(1-\pi_i)^{1-z_i}. \quad (5)$$

To acquire the trackable posterior of β_i, we place a Gamma prior, which is the conjugate to the Gaussian distribution, on β_i, i.e., $\beta_i \sim \text{Gamma}(a,b), i \in [1,\cdots,M]$, where a and b are hyper-parameters. A Gaussian prior is placed on the additive noise as $\boldsymbol{\epsilon}_l \sim \mathcal{N}(\boldsymbol{\epsilon}_l|\mathbf{0},\alpha_l^{-1}\mathbf{I}_N)$. In a similar way, we place a Gamma prior on α_l to acquire an analytical posterior distribution, i.e., $\alpha_l \sim \text{Gamma}(c,d), l \in [1,\cdots,L]$, where c and d are hyper-parameters.

B. Cluster Prior on Intra-Task

Similar to [22], we define the neighborhood blocks of the ith block as $\eta_i = \{j|D(i,j) = D_{\max}, j \neq i\}$ with $D(i,j)$ denoting the distance between the ith and the jth pixel positions, and $D_{\max}$ denotes the maximum neighboring distance to be considered. Let $\boldsymbol{\theta}_{\eta_i}$ and $\mathbf{z}_{\eta_i}$ respectively denote collections of $\boldsymbol{\theta}$ and $\mathbf{z}$ located in the neighborhood of the ith block, and $\kappa_i = \sum \mathbf{z}_{\eta_i}$ be the number of non-zero neighboring blocks for the ith block. Consider the simple
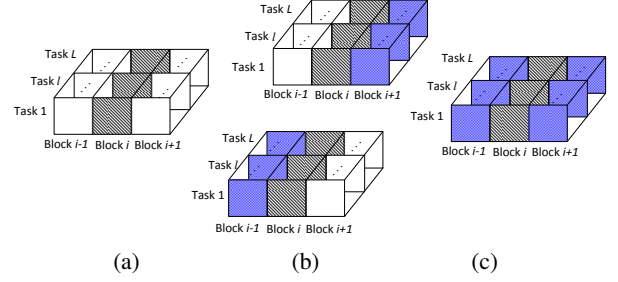

Fig. 1. Three cluster pattern for 1-D signal. (a) Pattern 0: strong rejection; (b) Pattern 1: weak rejection; (c) Pattern 2: strong acceptance.

example, depicted in Fig. 1, where the cluster is one-dimensional (1-D) with $D_{\max} = 1$. We categorize the relationship into three different cluster patterns. In Pattern 0 ("strong rejection"), all the neighboring blocks are zero valued, i.e., $\kappa_i = 0$, and the ith block would be zero valued with a high probability, since the targets of interest are unlikely to be isolated. In this case, $e_0 < f_0$ is assumed in the prior Beta(e_0,f_0) to encourage a small value of π_i^0 so as to reject this block. In Pattern 1 ("weak rejection"), one of the neighboring blocks is zero valued, i.e., $\kappa_i = 1$. In this case, the probability that the ith block takes zero values is fair, and $e_1 = f_1$ is used in the prior Beta(e_1,f_1) to exert non-informative prior on π_i^1. In Pattern 2 ("strong acceptance"), $\kappa_i = 2$, i.e., all the neighboring blocks have non-zero values. In this case, the ith block would take non-zero values with a high probability, and we use $e_2 > f_2$ in the prior Beta(e_2,f_2) to encourage a large value of π_i^2 to accept this block.

III. Bayesian inference

We adopt a Gibbs sampler to carry out the Bayesian inference of the proposed algorithm. For convenience, we define the collection of hyper-parameters as $\boldsymbol{\Xi} \triangleq \{a,b,c,d,\mathbf{e},\mathbf{f}\}$ with $\mathbf{e} \triangleq \{e_0,e_1,e_2\}$ and $\mathbf{f} \triangleq \{f_0,f_1,f_2\}$, and define the collection of random variables as $\Theta \triangleq \{\boldsymbol{\theta},\mathbf{z},\boldsymbol{\pi},\boldsymbol{\alpha},\boldsymbol{\beta}\}$. We also denote $\mathbf{Y} = \{\mathbf{y}_1,\cdots,\mathbf{y}_L\}$ and $\boldsymbol{\Phi} = \{\boldsymbol{\Phi}_1,\cdots,\boldsymbol{\Phi}_L\}$. The explicit form of the joint pdf is

$$\begin{aligned}p(\mathbf{Y},\boldsymbol{\Phi},\Theta|\boldsymbol{\Xi}) &= \prod_{l=1}^{L}\mathcal{N}(\mathbf{y}_l|\boldsymbol{\Phi}_l(\boldsymbol{\theta}_l \circ \mathbf{z}_l),\alpha_l^{-1}\mathbf{I})\text{Gamma}(\alpha_l|c,d)\\ &\times \prod_{l=1}^{L}\prod_{i=1}^{M}[\mathcal{N}(\theta_{il}|0,\beta_i^{-1})]^{z_i}\pi_i^{z_i}(1-\pi_i)^{1-z_i}\\ &\times \prod_{i=1}^{M}\text{Beta}(\pi_i|\mathbf{e},\mathbf{f},\mathbf{z}_{\eta_i})\text{Gamma}(\beta_i|a,b). \quad (6)\end{aligned}$$

The posterior distributions of each random variable will be analytically obtained based on the conjugate property.

A. Updating paired variables $\{\boldsymbol{\theta},\mathbf{z}\}$

Unlike the inference of standard spike-and-slab prior in [26], the paired Gibbs sampler iteratively samples from the following conditional pdf,

$$p(z_i,\theta_{il}|\boldsymbol{\theta}_{\setminus il},\mathbf{z}_{\setminus i},\mathbf{y}_l) = p(\theta_{il}|z_i,\boldsymbol{\theta}_{\setminus il},\mathbf{z}_{\setminus i},\mathbf{y}_l)p(z_i|\boldsymbol{\theta}_{\setminus il},\mathbf{z}_{\setminus i},\mathbf{y}_l), \quad (7)$$

where $\boldsymbol{\theta}_{\setminus il}$ and $\mathbf{z}_{\setminus i}$ respectively denote $\boldsymbol{\theta}_l$ except the variable θ_{il} and $\mathbf{z}$ except the variable z_i. Its marginal distribution is expressed as

$$p(z_i|\boldsymbol{\theta}_{\setminus il},\mathbf{z}_{\setminus i},\mathbf{y}_l) = \prod_{l=1}^{L}\int p(z_i,\theta_{il}|\boldsymbol{\theta}_{\setminus il},\mathbf{z}_{\setminus i},\mathbf{y}_l)d\theta_{il}, \quad (8)$$

where

$$p(z_i, \theta_{il}|\boldsymbol{\theta}_{\setminus il}, \mathbf{z}_{\setminus i}, \mathbf{y}_l) \propto \prod_{l=1}^{L} \mathcal{N}(\mathbf{y}_{\setminus il}|\boldsymbol{\phi}_{il} z_{il} \theta_{il}, \alpha_l^{-1}\mathbf{I})$$
$$\times [\mathcal{N}(\theta_{il}|0, \beta_i^{-1})]^{z_i} \pi_i^{z_i}(1-\pi_i)^{1-z_i}, \quad (9)$$

$\mathbf{y}_{\setminus il} = \mathbf{y}_l - \sum_{k \neq i} \boldsymbol{\phi}_{kl} z_k \theta_{kl}$, and $\boldsymbol{\phi}_{il}$ is the ith column of the measurement matrix $\boldsymbol{\Phi}_l$. The probability $p(z_i = 1|\boldsymbol{\theta}_{\setminus il}, \mathbf{z}_{\setminus i}, \mathbf{y}_l)$ is acquired analytically by utilizing the logistic function [26] as,

$$p(z_i = 1|\boldsymbol{\theta}_{\setminus il}, \mathbf{z}_{\setminus i}, \mathbf{y}_l) = (1 + e^{-u})^{-1}, \quad (10)$$

where u is derived as

$$u = \frac{1}{2}\sum_{l=1}^{L}\left(\log\beta_i - \log\sigma_{il} + \sigma_{il}\alpha_l^2 \mathbf{y}_{\setminus il}^H \boldsymbol{\phi}_{il}\boldsymbol{\phi}_{il}^H \mathbf{y}_{\setminus il}\right)$$
$$+ \log\pi_i - \log(1-\pi_i), \quad (11)$$
$$\sigma_{il} = (\alpha_l \boldsymbol{\phi}_{il}^H \boldsymbol{\phi}_{il} + \beta_i)^{-1}. \quad (12)$$

The conditional distribution of $p(\theta_{il}|z_i = 1, \boldsymbol{\theta}_{\setminus il}, \mathbf{z}_{\setminus i}, \mathbf{y}_l)$ can be expressed as,

$$p(\theta_{il}|z_i = 1, \boldsymbol{\theta}_{\setminus il}, \mathbf{z}_{\setminus i}, \mathbf{y}_l) = \mathcal{N}(\theta_{il}|\sigma_{il}\alpha_l\boldsymbol{\phi}_{il}^H\mathbf{y}_{\setminus il}, \sigma_{il}). \quad (13)$$

For $z_i = 0$, because the value of θ_{il} does not affect the result of w_{il}, we conveniently draw the value of variable θ_{il} from its prior.

B. Updating mixing weight $\boldsymbol{\pi}$

As depicted in Section II-B, each element π_i is chosen according to its corresponding sparsity pattern, i.e., π_i is assigned as $\pi_i^{(q)}$ for sparsity pattern $q \in \{0, 1, 2\}$. The Beta distribution on π_i leads to an analytical posterior distribution for a certain sparsity pattern q as [22],

$$p(\pi_i^{(q)}|\mathbf{e}, \mathbf{f}, \mathbf{z}) = \text{Beta}(e_q + \kappa_i + z_i, f_q + \nu_i + 1 - \kappa_i - z_i), \quad (14)$$

where $i \in \{1, \cdots, M\}$, and ν_i is the length of $\mathbf{z}_{\eta_i}$. In the 1-D signal example depicted in Fig. 1, $\nu_i = 2$, i.e., 2 neighboring blocks are considered for each block.

C. Updating signal precision β_i

By utilizing the conjugate property of the Gaussian and Gamma distributions, we analytically acquire the posterior distribution of the precision variable β_i as

$$p(\beta_i|a, b, \boldsymbol{\theta}) = \text{Gamma}(\tilde{a}, \tilde{b}), \quad (15)$$

where $\tilde{a} = a + (L/2)$, $\tilde{b} = b + \sum_l(\theta_{il}^2/2)$, and $i \in [1, \cdots, M]$.

D. Updating noise precision $\boldsymbol{\alpha}$

In a similar manner as β_i, we also obtain the posterior distribution of noise precision α_l

$$p(\alpha_l|c, d, \mathbf{y}_l, \boldsymbol{\Phi}_l, \boldsymbol{\theta}_l, \mathbf{z}) = \text{Gamma}(\tilde{c}, \tilde{d}), \quad (16)$$

where $\tilde{c} = c + (K/2)$, $\tilde{d} = d + (\|\mathbf{y}_l - \boldsymbol{\Phi}_l(\boldsymbol{\theta}_l \circ \mathbf{z})\|^2/2)$, and $l \in \{1, \cdots, L\}$.

The proposed algorithms is summarized in Table I. Since the inference of model parameters is implemented by the Gibbs sampler which inherently requires sequential sampling, it is straightforward to show from the procedure of the proposed algorithm, depicted in Table I, that its computational complexity is $\mathcal{O}(K \times L \times M \times N_{\text{maxiter}})$, where N_{maxiter} is the maximum number of iterations. It is higher than that of the other Bayesian based algorithms like mt-CS based on fast greedy algorithm, and the GS-VS and BSBL algorithms based on the dimension reduction operation, because it needs to go through all $L \times M$ elements in each MCMC iteration.

Table I. Clustered MT-BCS algorithm

Initialize the hyper-parameters $\boldsymbol{\Xi}$ and random variables $\boldsymbol{\Theta}$;
Assign N_{maxiter};
for $n = 1, \cdots, N_{\text{maxiter}}$, **do**
a. **for** $i = 1, \cdots, M$, **do**
    update $z_i \sim p(z_i|\boldsymbol{\theta}_{\setminus il}, \mathbf{z}_{\setminus i}, \mathbf{y}_l)$ [Eq. (9)];
    for $l = 1, \cdots, L$, **do**
        update $\theta_{il} \sim p(\theta_{il}|z_i, \boldsymbol{\theta}_{\setminus il}, \mathbf{z}_{\setminus i}, \mathbf{y}_l)$ [Eq. (14)];
    end for
    update $\pi_i \sim p(\pi_i^{(q)}|\mathbf{e}, \mathbf{f}, \mathbf{z})$ [Eq. (15)];
    update $\beta_i \sim p(\beta_i^{-1}|a, b, \boldsymbol{\theta})$ [Eq. (16)];
end for
b. **for** $l = 1, \cdots, L$, **do**
    update $\alpha_l \sim p(\alpha_l|c, d, \mathbf{y}_l, \boldsymbol{\Phi}_l, \boldsymbol{\theta}_l, \mathbf{z})$ [Eq. (17)];
end for
end for

IV. SIMULATION AND EXPERIMENT RESULTS

Two examples are considered in this section, one dealing with a 1-D simulated data set, and the other processing multi-static ISAR imaging. The following hyper-parameters are used: $a = b = c = d = 10^{-6}$, $(e_0, f_0) = (1/K, 1 - 1/K)$ for strong rejection, $(e_1, f_1) = (1/K, 1/K)$ for weak rejection, and $(e_1, f_1) = (1 - 1/K, 1/K)$ for strong acceptance. The maximum number of iterations in the Gibbs sampling is 200, and the sample with the maximum marginal likelihood in the last 20 samples is chosen as the estimate of $\mathbf{w}$ [22]. The normalized mean-square error (NMSE) $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_{\text{gen}}\|_2^2/\|\boldsymbol{\theta}_{\text{gen}}\|_2^2$ is used as the performance index, where $\hat{\boldsymbol{\theta}}$ is the estimate of the true signal $\boldsymbol{\theta}_{\text{gen}}$. For comparison, we also provide the performances of other state-of-the-art algorithms, including overlapping group sparse (GS) reconstruction [21], BOMP [8], mt-CS [12], BSBL [13], GS-VB [14], and CluSS-MCMC [22].

A. Simulated 1-D example

In the first example, we consider $L = 2$ signals, and the length of each 1-D task is $M = 256$. Each task contains a sparsity of $T = 40$ non-zero entries grouped into $B = 2$ clusters. The cardinality of non-zero entries in each cluster is randomly selected, as shown in Fig. 2(a). These two signals share the identical spike positions, but take random ± 1 values. Each measurement matrix $\boldsymbol{\Phi}_l \in R^{K \times M}$, $l = 1, ..., L$, is generated as a zero mean random Gaussian matrix whose columns are normalized to have a unit l_2 norm. Zero-mean Gaussian noise with standard deviation $\epsilon_0 = 0.005$ is added to each of K measurements that define the data $\{\mathbf{y}_l\}_{l=1}^{L}$, and the length of the measurements is $K = 70$.

The reconstructed result of signal 1 based on the proposed algorithm is shown in Fig. 2(b). It is evident that it successfully recovers the entire signal waveform by utilizing the cluster structure of neighborhood blocks. For comparison, the results obtained from several state-of-the-art algorithms are respectively shown in Figs. 2(c)-(h). In the first four algorithms, the entire tasks are partitioned into M blocks due to sharing the same non-zero support between those two tasks, i.e., $\{w_{m1}, w_{m2}\}$ in the mth block, $m = 1, ..., M$. Because of insufficient number of measurements K (K is smaller than the minimum quantity required for faithful reconstruction [2]), the reconstructed signals are highly noisy. For the overlapping GS algorithm, the block partitions are predefined such that $g_1 = \{w_{11}, w_{12}, w_{21}, w_{22}\}$, $g_2 = \{w_{21}, w_{22}, w_{31}, w_{32}\}$, ..., with each block overlapping half of the previous block, to encourage the continuity with nearby points. Two continuous clusters are reconstructed well with some distortion in the magnitudes, and a few noisy spikes appear due to overlapping partition, as shown in Fig. 2(g). For the CluSS-MCMC algorithm, the sparse signals are separately recovered with 200 Gibbs sampling
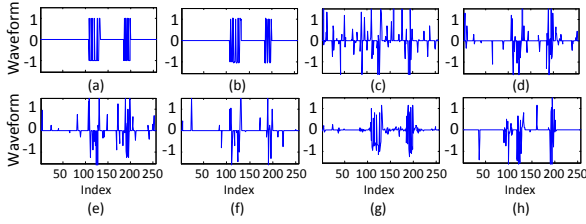
Fig. 2. Reconstructed results of the signal 1. (a) Original signal. (b) Result of the proposed algorithm. (c) Result of the BOMP algorithm using the true sparsity number. (d) Result of the BSBL algorithm. (e) Result of the GS-VB algorithm. (f) Result of the mt-CS algorithm. (g) Result of overlapping group sparse reconstruction. (h) Result of the CluSS-MCMC algorithm.
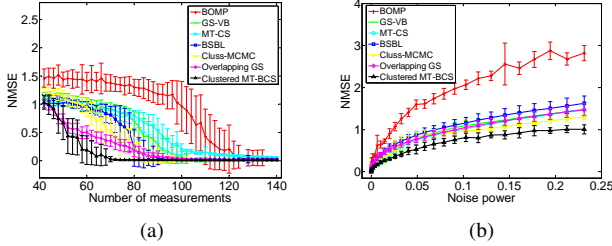

Fig. 3. (a) Performance comparison versus the number of measurements. (b) Performance comparison versus the noise power.

iterations. The signal in Fig. 2(h) is very noisy. As such, it is verified that the proposed algorithm enhances the reconstruction performance of sparse signals by transferring information between tasks based on a hierarchical Bayesian model when compared with the CluSS-MCMC algorithm.

1) Performance versus number of measurements: We take into account the oversampling rate K/T, which determines the number of measurements needed to achieve successful signal reconstructions. The number of measurements is varied from 42 to 140 with a step size of 2, resulting in an oversampling rate between 1.05 and 3.5. In each step, the measurement matrices $\{\boldsymbol{\Phi}_l\}_{l=1}^2$ are drawn from the Gaussian distribution $\mathcal{N}(0,1)$, and 100 independent trials are repeated for each number of measurements. Fig. 3(a) shows the errorbar plot which includes both average mean and standard deviation obtained from the 100 trials for each oversampling rate. It is evident that the proposed algorithm requires a much lower number of measurements for successful reconstruction. The sparse signals, which have 40 non-zero entries, can be reconstructed with NMSE<0.01 using only $K = 70$ measurements by the proposed algorithm, whereas the required number is $K = 92$ for BSBL, $K = 100$ for overlapping GS, and is much higher for the GS-VB, mt-CS, BOMP, and CluSS-MCMC algorithms.

2) Robustness in different noise level: To examine the effect of additive noise, we vary the noise power. To enable fair comparisons, we use $K = 128$ measurements to ensure that all the algorithms used can reconstruct the sparse signals according to Fig. 3(a). It is demonstrated in Fig. 3(b) that, as the noise power increases, the reconstruction error generally increases for all algorithms. Nevertheless, the proposed algorithm is more robust to noise than other algorithms and has the lowest reconstruction error. Regarding the performance comparison between BOMP and the Bayesian learning based algorithms, the superiority of the latter is evidently demonstrated by their lower reconstruction errors across the entire range of noise power being evaluated.

B. Application on multi-static ISAR imaging

In the second example, we perform ISAR imaging based on experimental data with a Yak-42 airplane recorded by a C-band (5.52
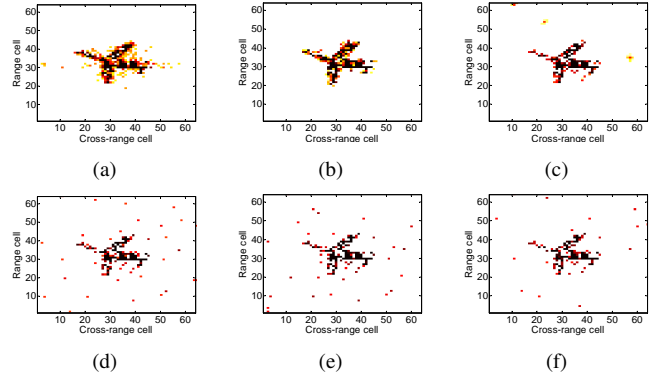

Fig. 4. ISAR imaging results obtained from the CS algorithms with half measurements. (a) Image obtained from the RDA algorithm with full measurements. (b) Result of the proposed algorithm. (c) Result of overlapping group sparse reconstruction. (d) Result of the mt-CS algorithm. (e) Result of the GS-VB algorithm. (f) Result of the BSBL algorithm.

GHz) ISAR system. The system transmits 400 MHz linear frequency modulated signals, and the pulse repetition frequency is 400 Hz. The number of points in the range domain is 64, and 64 pulses are used. The ISAR image with full measurements is shown in Fig. 4(a) based on the Range Doppler algorithm (RDA) [27]. It is observed that the target demonstrates strong clustered structure. To perform multi-task reconstruction, we generate another synthetic ISAR observation by randomly altering the phase and adding random perturbations on the measured radar scatterer coefficients. Similar to [17], additive complex white Gaussian noises are added to generate the experiment data with the equivalent signal-to-noise ratio (SNR) to be 10 dB.

To demonstrate CS technique applied to the ISAR imaging, we randomly select 50% measurements (32 pluses) for experiments. Among the aforementioned CS algorithms with group sparsity, the BOMP is excluded from comparison due to the unknown sparsity number in this data set. As seen in Fig. 4, all methods can properly recover most of the strong scatterers. However, comparing to those obtained by the BCS, GS-VB and BSBL algorithms, as respectively shown in Figs. 4(d)−4(f), the reconstructed results obtained by the proposed algorithm, shown in Fig. 4(b), are much more concentrated with relatively clear background. The overlapping GS algorithm yields similar but slightly inferior performance, as shown in Fig. 4(c). In these two approaches, a 2-D neighborhood clustered pattern is used in the proposed algorithm, whereas each pixel with all adjacent pixels is considered in the same group in the overlapping GS algorithm.

Due to the additional constraints representing the intra-task dependency, the proposed technique has a higher complexity. The CPU time of the proposed algorithm with a non-optimal Matlab code is about 2.85 times of that of the GB-VB algorithm and 3.71 times of that of the BSBL algorithm. The overlapping GS based on alternating direction optimization has the least CPU time, which is about 19% of that of the proposed method.

V. CONCLUSION

In this letter, we proposed a clustered multi-task compressive sensing algorithm for the recovery of group sparse signals whose non-zero entries are respectively clustered. The proposed technique is based on a hierarchical Bayesian framework, and takes both sparse prior and cluster prior into account. As such, it significantly improves the reconstruction performance of sparse signals by transferring information between tasks and exploiting the cluster structures of underlying targets. Simulation and experiment results demonstrated the superiority of the proposed algorithm over the state-of-the-art group sparse reconstruction algorithms.

References

[1] E. Candes and M. Wakin, "An introduction to compressive sampling," *IEEE Signal Processing Mag.*, vol. 25, no. 2, pp. 21–30, 2008.

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

[3] C. Rich, "Multitask learning," *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997.

[4] J. Yang, A. Bouzerdoum, F. H. C. Tivive, and M. G. Amin, "Multiple-measurement vector model and its application to through-the-wall radar imaging," in *Proc. IEEE ICASSP*, (Prague, Czech Republic), May 2011.

[5] Y. Zhang, B. Obeidat, and M. G. Amin, "Spatial polarimetric time-frequency distributions for direction-of-arrival estimations," *IEEE Trans. Signal Proc.*, vol. 54, no. 4, pp. 1327–1340, 2006.

[6] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Multi-static passive radar SAR imaging based on Bayesian compressive sensing," in *Proc. SPIE Defense, Security, and Sensing*, (Baltimore, MD), May 2014.

[7] G. Li, P. K. Varshney, and Y. D. Zhang, "Multistatic radar imaging via decentralized and collaborative subspace pursuit," in *Proc. Int. Conf. on Digital Signal Processing*, (Hong Kong, China), Aug. 2014.

[8] L. Jacob, G. Obozinski, and J. Vert, "Group Lasso with overlap and graph Lasso," in *Proc. Int. Conf. on Machine Learning*, (Montreal, Canada), Jun. 2009.

[9] E. V. D. Berg and M. Friedlander, "Probing the pareto frontier for basis pursuit solutions," *SIAM Scientific Computing*, vol. 31, no. 2, pp. 890–912, 2008.

[10] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Royal Statistical Society Series B*, vol. 68, no. 1, pp. 49–67, 2006.

[11] M. E. Tipping, "Sparse Bayesian shrinkage and selection learning and the relevance vector machine," *Machine Learning Research*, vol. 1, no. 9, pp. 211–244, 2001.

[12] S. Ji, D. Dunson, and L. Carin, "Multitask compressive sampling," *IEEE Trans. Signal Proc.*, vol. 57, no. 1, pp. 92–106, 2009.

[13] Z. Zhang and B. D. Rao, "Extension of SBL algorithm for the recovery of block sparse signals with inter-block correlation," *IEEE Trans. Signal Proc.*, vol. 92, no. 7, pp. 1580–1590, 2012.

[14] S. D. Babacan, S. Nakajima, and M. N. Do, "Bayesian group-sparse modeling and variational inference," *IEEE Trans. Signal Proc.*, vol. 62, no. 11, pp. 2906–2921, 2014.

[15] Q. Wu, Y. D. Zhang, M. G. Amin, and B. Himed, "Complex multitask Bayesian compressive sensing," in *Proc. IEEE ICASSP*, (Florence, Italy), May 2014.

[16] M. Leigsnering, F. Ahmad, M. G. Amin, and A. M. Zoubir, "Compressive sensing based specular multipath exploitation for through-the-wall radar imaging," in *Proc. IEEE ICASSP*, (Vancouver, Canada), May 2013.

[17] L. Wang, L. Zhao, G. Bi, C. Wan, and L. Yang, "Enhanced ISAR imaging by exploiting the continuity of the target scene," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 9, pp. 5736–5750, 2014.

[18] Z. Zhang and B. D. Rao, "Sparse signal recovery with temporally correlated source vector using sparse Bayesian learning," *IEEE Jounal of Selected Topics in Signal Processing*, vol. 5, no. 5, pp. 1–15, 2011.

[19] Z. Zhang, T. Jung, S. Makeig, Z. Pi, and B. D. Rao, "Spatiotemporal sparse Bayesian learning with applications to compressed sensing of multichannel physiological signals," *IEEE Trans. Neural Systems and Rehabilitation Engineering*, (in press).

[20] M. K. Titsias and M. Lázaro-Gredilla, "Spike and slab variational inference for multi-task and multiple kernel learning," in *Proc. Adv. Neural Info. Proc. Syst.*, (Granada, Spain), Dec. 2011.

[21] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," Department of Computation and Applied Mathematics, Rice University, Technical Report TR11-06,2011.

[22] L. Yu, H. Sun, J. P. Barbot, and G. Zheng, "Bayesian compressive sensing for cluster structured sparse signals," *Signal Processing*, vol. 92, no. 1, pp. 259–269, 2012.

[23] E. I. George and R. E. Mcculloch, "Variable selection via Gibbs sampling," *American Stat. Assoc.*, vol. 88, no. 423, pp. 881–889, 1993.

[24] T. J. Mitchell and J. J. Beauchamp, "Bayesian variable selection in linear regression," *American Stat. Assoc.*, vol. 83, no. 404, pp. 1023–1032, 1988.

[25] L. Yu, J. P. Barbot, G. Zheng, and H. Sun, "Compressive sensing for cluster structured sparse signals: Variational Bayes approach," Technical Report, 2011. Available at http://hal.archives-ouvertes.fr/docs/00/57/39/53/PDF/cluss_vb.pdf.

[26] M. Zhou, H. Chen, J. Paisley, L. Ren, L. Li, Z. Xing, D. Dunson, G. Sapiro, and L. Carin, "Nonparametric Bayesian dictionary learning for analysis of noisy and incomplete images," *IEEE Trans. Image Proc.*, vol. 21, no. 2, pp. 130–144, 2012.

[27] I. G. Cumming and F. H. Wong, *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Artech House, 2005.