Signal Processing 172 (2020) 107560

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Group adaptive matching pursuit with intra-group correlation learning for sparse signal recovery

Qisong Wu[a,c,1,*], Jiahao Liu[a,c,2], Moeness G. Amin[b,3]

[a] Key Laboratory of Underwater Acoustic Signal Processing of Ministry of Education, Southeast University, Nanjing, 210096, China
[b] Center for Advanced Communications, Villanova University, Villanova, PA 19085, USA
[c] Department of Electrical Engineering, Southeast University, and also with Purple Mountain Laboratories, Nanjing, 210096, China

ARTICLE INFO

Article history:
Received 23 September 2019
Revised 4 February 2020
Accepted 25 February 2020
Available online 28 February 2020

Keywords:
Compressive sensing
Bayesian inference
Spike-and-slab prior
Sparse signal recovery
Intra-group correlation learning

ABSTRACT

In this paper, a Bayesian model is adopted for sparse signal recovery where sparsity is enforced on the reconstructed coefficients via probabilistic priors. In particular, we focus on a group spike-and-slab prior and a kernel matrix which capture both the underlying group structure and the element correlation within groups. A novel greedy based group adaptive matching pursuit (GAMP) algorithm is introduced, which integrates both prior parameter learning and intra-group correlation parameter learning into one single problem. The proposed approach improves the reconstruction accuracy and offers strong robustness to signal-to-noise ratio. We consider a fast implementation method of GAMP which applies the preconditioned conjugate gradient method. Simulations, MNIST dataset based experiments and multi-static radar imaging application are used to verify the superior performance of the proposed method over existing techniques.

1. Introduction

SPARSE signal recovery and the related compressive sensing (CS) problems have attracted significant attention in recent years. The CS techniques are able to recover signals from a small number of measurements with a high probability if the signals are sparse or can be sparsely represented in some known domains [1–3].

A typical CS model aiming to recover a sparse signal $\boldsymbol{\omega} \in \mathbb{R}^n$ from a set of fewer measurements $\boldsymbol{y} \in \mathbb{R}^m$ ($m \ll n$) is given by:

$$\boldsymbol{y} = \mathbf{A}\boldsymbol{\omega} + \boldsymbol{n}, \tag{1}$$

where $\mathbf{A} \in \mathbb{R}^{m\times n}$ is a known sensing matrix, and $\boldsymbol{n} \in \mathbb{R}^m$ is an unknown Gaussian noise vector. The columns of $\mathbf{A}$ are assumed to be linearly independent to satisfy the unique representation property which ensures reliable sparse signal recovery [1].

In order to further enhance recoverability, recent studies have gone beyond sparse reconstruction and taken into account additional information about the underlying structure of the signal. In practice, a wide class of signals are known to have a "group-sparsity" structure. That is, the coefficient vector $\boldsymbol{\omega}$ has a natural grouping of its components, and the components within a group are likely to be either all zeros or all nonzeros. In addition, the elements within groups are often correlated [4]. With group-sparsity structure, $\boldsymbol{\omega}$ can be viewed as a concatenation of K groups, i.e.,

$$\boldsymbol{\omega} = \left[\underbrace{\omega_{11}, \cdots, \omega_{1L}}_{\boldsymbol{\omega}_1^T}, \cdots, \underbrace{\omega_{K1}, \cdots, \omega_{KL}}_{\boldsymbol{\omega}_K^T} \right]^T, \tag{2}$$

where $\boldsymbol{\omega}_k$ is the kth group of $\boldsymbol{\omega}$, ω_{kl} is the lth element in $\boldsymbol{\omega}_k$, K is the number of groups and L is the number of elements in each group. It should be pointed that the identical size assumption is not necessary, and the method is also applicable to different group sizes. A sparsity level p amounts to having p non-zero groups among K ($p \ll K$) groups, with unknown locations. A number of algorithms have been proposed to recover group-sparsity signals. Typical algorithms include block-OMP (BOMP) [5], group basis pursuit (GBP) [6], group Lasso (GLASSO) [7] and block sparse Bayesian learning (BSBL) [4]. In Bayesian sparse recovery, the choice of priors plays a key role in promoting sparsity and improving the recovery performance. For example, BSBL [4] places a Gaussian-Gamma distribution to model the group sparsity. In this paper, we focus on spike-and-slab prior [8–14], which is a well-suited sparsity promoting prior and is considered the golden standard for sparse inference in the Bayesian set-up [15]. In this prior, each individual

* Corresponding author.
E-mail address: qisong.wu@seu.edu.cn (Q. Wu).
[1] Member, IEEE.
[2] Student Member, IEEE.
[3] Fellow, EURASIP and Fellow, IEEE.

https://doi.org/10.1016/j.sigpro.2020.107560
0165-1684/

coefficient ω_{kl} of $\boldsymbol{\omega}$ is modeled as a mixture of two components,

$$\omega_{kl} \sim (1-\gamma_{kl})\delta_0 + \gamma_{kl} P_{kl}(\omega_{kl}), \tag{3}$$

where δ_0 is a point mass concentrated at zero (the "spike"), and P_{kl} is any suitable distribution on the non-zero coefficient (e.g. a Gaussian). The binary indicator variable γ_{kl} controls the sparsity level of the signal. When γ_{kl} is chosen to be close to zero, ω_{kl} tends to remain zero, otherwise P_{kl} will be the dominant distribution encouraging ω_{kl} to take a non-zero value. Although the spike-and-slab prior is suitable to encourage signal sparsity, it is unable to exploit the underlying structure of the sparse signal. A number of extended spike-and-slab priors were proposed to model the group structure, and the spatial-temporal dependence of sparse signals [16–20].

Motivated by these methods, the main contributions of our work are as follows: (1) A novel generative model in sparse signal recovery is provided, which takes into consideration the group structure and correlation learning within groups. A group spike-and-slab prior is applied to capture the group structure of the signals, whereas a kernel matrix is used to learn the correlations within groups. (2) A greedy based group adaptive matching pursuit (GAMP) algorithm is proposed to deal with the non-convex optimization problem resulting from the employed prior. This algorithm integrates the prior parameter learning and correlation parameter estimation into one single problem. (3) A fast implementation method of the GAMP is proposed which exploits the preconditioned conjugate gradient method.

The remainder of the paper is organized as follows. Section 2 describes the generative model and the optimization problem. Section 3 presents the solution of the non-convex optimization problem in Section 2. Section 4 provides a fast implementation of GAMP. The simulation and experimental results are shown in Section 5.

Notations: We use lower-case (upper-case) bold characters to denote vectors (matrices). $f(x|a,\ b)$ is the conditional probability distribution function (pdf) of the variable x given a and b. $\mathcal{N}(x|a,b)$ denotes that random variable x follows a Gaussian distribution with mean a and variance b, $\mathbf{I}_L$ denotes the $L \times L$ identity matrix, and $\mathbb{I}(x=0)$ stands for an indicator function, which equals to 1 if and only if $x=0$, and is 0 otherwise. In addition, $(\ \cdot\ )^T$ and $(\ \cdot\ )^H$ respectively denote transpose and conjugate transpose, and $\|\boldsymbol{x}\|_l$ represents the lth norm.

2. The proposed model

2.1. Group spike-and-slab prior

To encourage group sparsity, a group spike-and-slab prior is imposed on the vector $\boldsymbol{\omega}$, i.e.,

$$\boldsymbol{\omega}_k \sim \gamma_k \mathcal{N}(0, \sigma^2\lambda^{-1}\boldsymbol{B}_k) + (1-\gamma_k)\mathbb{I}(\boldsymbol{\omega}_k = \mathbf{0}), \tag{4}$$

where γ_k is a binary indicator variable, which follows the Bernoulli distribution. This prior implies that all the elements in the kth group share the identical γ_k, which is different from the standard spike-and-slab prior. When γ_k is 1, all the elements in the kth group are nonzero. σ^2 is the noise precision of the measurements, and λ is a learned parameter related to the variance of the estimated vector $\boldsymbol{\omega}$. It has been shown that, in many applications, the elements within the same group are often correlated [11,13,21], and sparse reconstruction can be significantly improved if such correlation is properly exploited. Motivated by the kernel techniques in correlation learning models [4,21], we define matrix $\boldsymbol{B}_k \in \mathbb{R}^{L\times L}$ as a positive definite kernel matrix to capture the correlation of the vector elements within the kth groups. To avoid overfitting in parameter estimation, we constrain $\boldsymbol{B}_k = \boldsymbol{B}\ \forall k$ [21]. That is, all K groups share the identical kernel matrix, given by

$$\boldsymbol{B} = \begin{bmatrix} 1 & e^{-c\|d_{21}\|_1} & \dots & e^{-c\|d_{L1}\|_1} \\ \vdots & & & \vdots \\ e^{-c\|d_{L1}\|_1} & e^{-c\|d_{L2}\|_1} & \dots & 1 \end{bmatrix},$$

where $\|d_{ij}\|_1 = \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_1$ is the absolute distance between the elements $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ within each group, and $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are, respectively, the spatial representations of elements i and j, with $i, j = 1, \cdots, L$. Notice that the above kernel matrix is real and symmetric, and all its entries take values within [0,1]. The diagonal entries take unit values, whereas off-diagonal element values exponentially decrease as a function of their respective distance from the main diagonal. It is noted that $c \geq 0$ is a key scaler parameter that leverages the correlation between elements within groups. When the scalar c approaches infinity, the kernel matrix reduces to the identity matrix, implying that all vector elements are independent. On the other hand, small values of c amount to strong correlations between the elements. In the proposed model, the intra-group correlation will be automatically learned by estimating the coefficient c.

Without loss of generality, the measurements are assumed to follow Gaussian distributions, expressed as,

$$\boldsymbol{y}|\boldsymbol{\omega}, \sigma^2 \sim \mathcal{N}\left(\boldsymbol{y}|\mathbf{A}\boldsymbol{\omega}, \sigma^2\mathbf{I}_m\right), \tag{5}$$

where $\boldsymbol{\omega}$ is a vector composed of $\boldsymbol{\omega}_k$. A latent variable based model for sparse signal recovery is employed and the Bernoulli distribution is imposed on vector $\boldsymbol{\gamma}$ as,

$$\boldsymbol{\gamma}|\boldsymbol{\kappa} \sim \prod_{k=1}^{K} \text{Bernoulli}(\gamma_k|\kappa_k), \quad k \in \{1, \cdots, K\}, \tag{6}$$

where $\boldsymbol{\kappa}$ is a hyperparameter vector.

2.2. Optimization problem

According to the generative model above, the posterior probability distribution of latent random variables $\boldsymbol{\omega}$, $\boldsymbol{\gamma}$, c, σ^2 and λ can be expressed by

$$f(\boldsymbol{\omega}, \boldsymbol{\gamma}, c, \lambda, \sigma^2|\boldsymbol{y}, \boldsymbol{\kappa}) \propto f(\boldsymbol{y}|\boldsymbol{\omega}, \sigma^2) f(\boldsymbol{\omega}|\boldsymbol{\gamma}, c, \lambda, \sigma^2) f(\boldsymbol{\gamma}|\boldsymbol{\kappa}).$$

A maximum a posteriori (MAP) estimation is performed [22], and the optimal values of $\boldsymbol{\omega}^*$, $\boldsymbol{\gamma}^*$, c^*, λ^* and σ^{2^*} are given by :

$$(\boldsymbol{\omega}^*,\ \boldsymbol{\gamma}^*,\ c^*, \lambda^*,\ \sigma^{2^*}) = \arg \max_{\boldsymbol{\omega}, \boldsymbol{\gamma}, c, \lambda, \sigma^2} \left\{ f(\boldsymbol{\omega}, \boldsymbol{\gamma}, c, \lambda, \sigma^2|\boldsymbol{y}, \boldsymbol{\kappa}) \right\}. \tag{7}$$

The optimization problem above is equivalent to the following minimization problem (see supplementary material in Appendix A):

$$(\boldsymbol{\omega}^*,\ \boldsymbol{\gamma}^*,\ c^*,\ \lambda^*,\ \sigma^{2^*}) = \arg \min_{\boldsymbol{\omega}, \boldsymbol{\gamma}, c, \lambda, \sigma^2} L(\boldsymbol{\omega},\ \boldsymbol{\gamma},\ c,\ \lambda,\ \sigma^2), \tag{8}$$

where

$$L(\boldsymbol{\omega}, \boldsymbol{\gamma}, c, \lambda, \sigma^2) = \frac{\|\ \boldsymbol{y} - \mathbf{A}\boldsymbol{\omega}\ \|_2^2}{\sigma^2} + \frac{\lambda}{\sigma^2} \sum_{k=1}^{K} \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k + \sum_{k=1}^{K} \rho_k \gamma_k + m \ln \sigma^2,$$

where $\rho_k = \ln((\frac{2\pi\sigma^2}{\lambda})^L |\boldsymbol{B}| \frac{(1-\kappa_k)^2}{\kappa_k^2})$ and $\boldsymbol{\rho}$ is a vector composed of ρ_k with $k = 1, \cdots, K$.

The cost function $L(\boldsymbol{\omega}, \boldsymbol{\gamma}, c, \lambda, \sigma^2)$ consists of three terms, namely, the data-fitting term, the correlation learning term and the hyperprior-based regularization term. Specifically,

1) The first term is related to data-fitting, and it is a well-known least-square problem. It aims to fit the recovered sparse signal into the measurement vector of $\boldsymbol{y}$.

2) The second term is for learning the intra-group correlation. It can also induce group sparsity together with the third term.
3) The third term is a hyperprior-based regularization term, that controls the activeness of each group in $\boldsymbol{\omega}$. The existence of this term results in a non-convex optimization problem which is addressed by applying a greedy-based algorithm, as discussed in Section 3.

It can be observed that signal recovery, sparsity prior learning, intra-correlation parameter learning and noise estimation are integrated into one unified model. The sparsity prior and intra-group correlation learning can effectively reduce the solution space of signal, thus improving the performance of sparse signal recovery. The recovered signal can further improve the accuracy of noise estimates which can be subsequently used to refine the optimization procedure.

To solve the above problem, the GAMP method is proposed where the alternative minimization scheme [23] is adopted to decompose (8) into four subproblems. The first subproblem, which includes signal recovery and sparsity prior learning, can be expressed as

$$(\boldsymbol{\omega}^*,\ \boldsymbol{\gamma}^*) = \arg\min_{\boldsymbol{\omega},\boldsymbol{\gamma}} \frac{\| \boldsymbol{y} - \mathbf{A}\boldsymbol{\omega} \|_2^2}{\sigma^2} + \frac{\lambda}{\sigma^2}\sum_{k=1}^{K} \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1}\boldsymbol{\omega}_k + \boldsymbol{\rho}^H\boldsymbol{\gamma}. \quad (9)$$

Given the updated $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$, the second subproblem is for estimating λ, and is written as

$$\lambda^* = \arg\min_{\lambda} \frac{\lambda}{\sigma^2}\sum_{k=1}^{K} \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1}\boldsymbol{\omega}_k + \boldsymbol{\rho}^H\boldsymbol{\gamma}. \quad (10)$$

The third subproblem optimizes for the noise precision σ^2, given the learned $\boldsymbol{\omega}$, $\boldsymbol{\gamma}$ and λ, and it is expressed as

$$\sigma^{2^*} = \arg\min_{\sigma^2} \frac{\| \boldsymbol{y} - \mathbf{A}\boldsymbol{\omega} \|_2^2}{\sigma^2} + \frac{\lambda}{\sigma^2}\sum_{k=1}^{K} \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1}\boldsymbol{\omega}_k + \boldsymbol{\rho}^H\boldsymbol{\gamma} + m\ln\sigma^2. \quad (11)$$

The last subproblem can be expressed as

$$c^* = \arg\min_{c} \frac{\lambda}{\sigma^2}\sum_{k=1}^{K} \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1}\boldsymbol{\omega}_k + \sum_{k=1}^{K}\rho_k\gamma_k. \quad (12)$$

The optimization procedure of these subproblems is provided in Section 3 in details.

3. Group adaptive matching pursuit

3.1. Signal recovery and sparsity prior learning: Optimization for $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$

In this subsection, we optimize $\boldsymbol{\omega}$ and $\boldsymbol{\gamma}$ based on Eq. (9) given σ^2 and λ. A greedy-based algorithm which solves the first subproblem by adding/removing elements into/from the group support set of $\boldsymbol{\omega}$ is proposed to handle this non-convex optimization problem. The subproblem can be rewritten as

$$(\boldsymbol{\omega}^*,\ \boldsymbol{\gamma}^*) = \arg\min_{\boldsymbol{\omega},\boldsymbol{\gamma}} \| \boldsymbol{y} - \mathbf{A}\boldsymbol{\omega} \|_2^2 + \lambda\sum_{k=1}^{K} \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1}\boldsymbol{\omega}_k + \sigma^2\boldsymbol{\rho}^H\boldsymbol{\gamma}. \quad (13)$$

Assuming that the group support of $\boldsymbol{\omega}$ is given, that is, the locations of the non-zero groups are known, we define the set of non-zero group index as $s = \{k : \gamma_k \neq 0\}$ with $k \in 1, \cdots, K$. Then, the optimization problem Eq. (13) reduces to

$$\boldsymbol{\omega}^s = \arg\min_{\boldsymbol{\omega}^s} \| \boldsymbol{y} - \mathbf{A}_s\boldsymbol{\omega}^s \|_2^2 + \lambda\sum_{k\in s} (\boldsymbol{\omega}_k^s)^H \boldsymbol{B}^{-1}\boldsymbol{\omega}_k^s, \quad (14)$$

where $\boldsymbol{\omega}^s$ represents the vector containing only active groups of $\boldsymbol{\omega}$ which are indexed by s. $\mathbf{A}_s$ is a sub-matrix of $\mathbf{A}$ composed by the group columns of $\mathbf{A}$ indexed by s. This problem is equivalent to the well-known ridge regression problem which can be easily solved by the conventional optimization algorithms.

For convenience, we define

$$\theta_s = \sigma^2\sum_{k\in s}\rho_k,\ \boldsymbol{r}_s = \boldsymbol{y} - \mathbf{A}_s\boldsymbol{\omega}^s \quad (15)$$

and

$$g(s) = \min_{\boldsymbol{\omega}^s} \| \boldsymbol{y} - \mathbf{A}_s\boldsymbol{\omega}^s \|_2^2 + \lambda\sum_{k\in s} (\boldsymbol{\omega}_k^s)^H \boldsymbol{B}^{-1}\boldsymbol{\omega}_k^s + \theta_s. \quad (16)$$

The one-to-one correspondence between s and $\boldsymbol{\omega}^s$ implies that solving Eq. (13) is equivalent to finding the group support set s. This prompts us to use a greedy-based method to find the set s by adding/removing a group index into/from s, whichever further reduces the cost function, and then solve problem Eq. (14). In particular, the choice of group index and the action of adding/removing are decided by calculating these two values at each iteration:

$$U_s = \min_{k\notin s} g(s\cup\{k\}) - g(s), \quad (17)$$

where U_s represents the decrease in the cost function if adding one of the unselected group indices into active set s. Similarly,

$$V_s = \min_{j\in s} g(s\backslash\{j\}) - g(s) \quad (18)$$

is the reduction of cost function if removing one of the existing indices in s. When both U_s and V_s are not less than 0, the iteration would be terminated, and it suggests the algorithm has converged. Otherwise, we compare U_s and V_s to update s by adding k (if $U_s < V_s$) or removing j (if $U_s > V_s$). However, the burden of calculating the precise U_s and V_s is very expensive. To address this problem, $\overline{U}_s$ and $\overline{V}_s$, which are respectively the upper bounds of U_s and V_s, will be calculated to reduce the computational cost instead of direct estimates of U_s and V_s. According to the estimated upper bounds $\overline{U}_s$ and $\overline{V}_s$, an updated group support s is acquired in each iteration. Given the updated s, $\boldsymbol{\omega}^s$ and $\boldsymbol{r}_s$ can be estimated from Eqs. (14) and (15), respectively.

The initialization of s may significantly influence the convergence of GAMP, and a beneficial initialization guidance will be given. The following three propositions support GAMP in initializing s and calculating $\overline{U}_s$ and $\overline{V}_s$, respectively.

Proposition 1. *If $\rho_k < 0$, then $k \in s_{opt}$, where s_{opt} is the optimal group support set.*

Proof. Assume that $k \notin s_{opt}$, then

$$\begin{aligned} g(s_{opt}\cup\{k\}) &\le \| \boldsymbol{r}_{s_{opt}} - \boldsymbol{a}_k\boldsymbol{\omega}_k \|_2^2 + \lambda\sum_{q\in s} \boldsymbol{\omega}_q^H\boldsymbol{B}^{-1}\boldsymbol{\omega}_q \\ &\quad + \sigma^2\rho_k + \theta_s + \lambda\boldsymbol{\omega}_k^H\boldsymbol{B}^{-1}\boldsymbol{\omega}_k \\ &= g(s_{opt}) + (\boldsymbol{a}_k\boldsymbol{\omega}_k)^H(\boldsymbol{a}_k\boldsymbol{\omega}_k) - 2\boldsymbol{r}_{s_{opt}}^H(\boldsymbol{a}_k\boldsymbol{\omega}_k) \\ &\quad + \lambda\boldsymbol{\omega}_k^H\boldsymbol{B}^{-1}\boldsymbol{\omega}_k + \sigma^2\rho_k, \end{aligned} \quad (19)$$

where $\boldsymbol{a}_k$ is a sub-matrix which contains the kth group column of $\boldsymbol{A}$. Let

$$p(\boldsymbol{\omega}_k) = (\boldsymbol{a}_k\boldsymbol{\omega}_k)^H(\boldsymbol{a}_k\boldsymbol{\omega}_k) - 2\boldsymbol{r}_{s_{opt}}^H\boldsymbol{a}_k\boldsymbol{\omega}_k + \lambda\boldsymbol{\omega}_k^H\boldsymbol{B}^{-1}\boldsymbol{\omega}_k + \sigma^2\rho_k, \quad (20)$$

where $p(\cdot) : \mathbb{R}^L \to \mathbb{R}$. It is obvious that as all the elements of $\boldsymbol{\omega}_k$ approaches to positive infinity, $p(\boldsymbol{\omega}_k)$ will equal to positive infinity. Furthermore, $p(\mathbf{0}) = \sigma^2\rho_k < 0$ and $p(\boldsymbol{\omega}_k)$ is continuous. Thus, there exist $\widehat{\boldsymbol{\omega}}_k$ whose all elements are nonzero such that $\sigma^2\rho_k < p(\widehat{\boldsymbol{\omega}}_k) < 0$. Then we can obtain

$$g(s_{opt}\cup\{k\}) \le g(s_{opt}) + p(\widehat{\boldsymbol{\omega}}_k) \le g(s_{opt}), \quad (21)$$

which suggests that adding k into s_{opt} can decrease the cost function. However, it is contradict to the assumption $k \notin s_{opt}$. This implies that if $\rho_k < 0$, then $k \in s_{opt}$. □

According to Proposition 1 we can initialize $s_0 = \{k : \rho_k < 0\}$.

Proposition 2.

$$U_s \le \min_{k \notin s} \left\{ \sigma^2 \rho_k - \boldsymbol{r}_s^H \boldsymbol{a}_k (\boldsymbol{a}_k^H \boldsymbol{a}_k + \lambda \boldsymbol{B})^{-1} \boldsymbol{a}_k^H \boldsymbol{r}_s \right\} \triangleq \overline{U}_s \tag{22}$$

Proof. Based on Eq. (19), we have

$$\begin{aligned} U_s &= \min_{k \notin s} g(s \cup \{k\}) - g(s) \\ &\le \min_{k \notin s} (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{a}_k \boldsymbol{\omega}_k) - 2 \boldsymbol{r}_s^H \boldsymbol{a}_k \boldsymbol{\omega}_k + \lambda \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k + \sigma^2 \rho_k \end{aligned}$$

Define

$$h(\boldsymbol{\omega}_k) = \min_{k \notin s} (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{a}_k \boldsymbol{\omega}_k) - 2 \boldsymbol{r}_s^H \boldsymbol{a}_k \boldsymbol{\omega}_k + \lambda \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k + \sigma^2 \rho_k. \tag{23}$$

Let

$$\frac{\partial h(\boldsymbol{\omega}_k)}{\partial \boldsymbol{\omega}_k} = 0, \tag{24}$$

then we have

$$\widehat{\boldsymbol{\omega}}_k = (\boldsymbol{a}_k^H \boldsymbol{a}_k + \lambda \boldsymbol{B})^{-1} \boldsymbol{a}_k^H \boldsymbol{r}_s. \tag{25}$$

We can obtain the minimum of $h(\boldsymbol{\omega}_k)$ as follows by substituting Eq. (25) into $h(\boldsymbol{\omega}_k)$

$$\min\{h(\boldsymbol{\omega}_k)\} = \sigma^2 \rho_k - \boldsymbol{r}_s^H \boldsymbol{a}_k (\boldsymbol{a}_k^H \boldsymbol{a}_k + \lambda \boldsymbol{B})^{-1} \boldsymbol{a}_k^H \boldsymbol{r}_s. \tag{26}$$

Since Eq. (26) holds for every $k \notin s$, it can be derived as

$$U_s \le \min_{k \notin s} \left\{ \sigma^2 \rho_k - \boldsymbol{r}_s^H \boldsymbol{a}_k (\boldsymbol{a}_k^H \boldsymbol{a}_k + \lambda \boldsymbol{B})^{-1} \boldsymbol{a}_k^H \boldsymbol{r}_s \right\}. \tag{27}$$

□

Proposition 3.

$$\begin{aligned} V_s \le \min_{k \in s} \{ & (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{a}_k \boldsymbol{\omega}_k) + 2 (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H \boldsymbol{r}_s \\ & - \lambda \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k - \sigma^2 \rho_k \} \triangleq \overline{V}_s \end{aligned} \tag{28}$$

Proof. For any $k \in \boldsymbol{S}$, we have

$$\begin{aligned} g(s) = & \min_{\boldsymbol{\omega}^s} \ \| \boldsymbol{r}_s + \boldsymbol{a}_k \boldsymbol{\omega}_k - \boldsymbol{a}_k \boldsymbol{\omega}_k \|_2^2 + \lambda \sum_{q \in s \setminus \{k\}} \boldsymbol{\omega}_q^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_q \\ & + \lambda \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k + \theta_{s \setminus \{k\}} + \sigma^2 \rho_k \\ = & \min_{\boldsymbol{\omega}^s} \ \| \boldsymbol{r}_s + \boldsymbol{a}_k \boldsymbol{\omega}_k \|_2^2 + \lambda \sum_{q \in s \setminus \{k\}} \boldsymbol{\omega}_q^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_q \\ & + \lambda \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k + \theta_{s \setminus \{k\}} + (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{a}_k \boldsymbol{\omega}_k) \\ & - 2 (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{r}_s + \boldsymbol{a}_k \boldsymbol{\omega}_k) + \sigma^2 \rho_k \\ \ge & g(s \setminus \{k\}) + (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{a}_k \boldsymbol{\omega}_k) - 2 (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{r}_s + \boldsymbol{a}_k \boldsymbol{\omega}_k) \\ & + \lambda \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k + \sigma^2 \rho_k. \end{aligned}$$

Since Eq. (29) holds for any $k \in s$, then we have

$$\begin{aligned} V_s = & \min_{k \in s} g(s \setminus \{k\}) - g(s) \\ \le & \min_{k \in s} \left\{ (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H (\boldsymbol{a}_k \boldsymbol{\omega}_k) + 2 (\boldsymbol{a}_k \boldsymbol{\omega}_k)^H \boldsymbol{r}_s - \lambda \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k - \sigma^2 \rho_k \right\}. \end{aligned} \tag{29}$$

□

According to these propositions above, the calculation of U_s and V_s is converted to approximate their upper bounds, thus significantly reducing the computation complexity.

3.2. Noise estimation: optimization for σ^2 and λ

Given $\boldsymbol{\omega}^s$ and s, the subproblem over λ is derived as

$$\lambda^* = \arg\min_{\lambda} \frac{\lambda}{\sigma^2} \sum_{q \in s} \boldsymbol{\omega}_q^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_q + \boldsymbol{\rho}^H \boldsymbol{\gamma}, \tag{30}$$

which results in the following closed-form solution

$$\lambda^* = \frac{L \sigma^2}{\sum_{q \in s} \boldsymbol{\omega}_q^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_q} \cdot \sum_{k=1}^{K} \gamma_k. \tag{31}$$

The subproblem over σ^2 is derived as

$$\sigma^{2^*} = \arg\min_{\sigma^2} \frac{\| \boldsymbol{y} - \boldsymbol{A}_s \boldsymbol{\omega}^s \|_2^2}{\sigma^2} + \frac{\lambda}{\sigma^2} \sum_{q \in \mathcal{S}} \boldsymbol{\omega}_q^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_q + \boldsymbol{\rho}^H \boldsymbol{\gamma} + m \ln \sigma^2. \tag{32}$$

Solving this problem, then we have

$$\sigma^{2^*} = \frac{\| \boldsymbol{y} - \boldsymbol{A}_s \boldsymbol{\omega}^s \|_2^2 + \lambda \sum_{q \in s} \boldsymbol{\omega}_q^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_q}{L \sum_{k=1}^{K} \gamma_k + m}. \tag{33}$$

3.3. Correlation learning: optimization for c

Given $\boldsymbol{\omega}^s$ and s, the updating step of c is performed at each iteration. We empirically calculate the value of c by $c \triangleq -\ln \frac{m_1}{m_0}$ [4], where m_0 is the average of elements along the main diagonal of matrix $\boldsymbol{B}$, whereas m_1 represents that along the main sub-diagonal. The learning rule of matrix $\boldsymbol{B}$ is obtained by setting $\frac{\partial L(\boldsymbol{\omega}, \boldsymbol{\gamma}, c, \lambda, \sigma^2)}{\partial \boldsymbol{B}} = 0$, which results in the following closed-form solution

$$\boldsymbol{B} = \frac{\lambda \sum_{k \in s} \boldsymbol{\omega}_k \boldsymbol{\omega}_k^H}{\sigma^2 \sum_{k \in s} \gamma_k}. \tag{34}$$

Then parameter c can be updated using the new matrix $\boldsymbol{B}$. Therefore, the intra-group correlation will be automatically learned and estimated.

3.4. Summary of the algorithm

The whole optimization procedure is summarized in Algorithm 1. Since the update step of $\overline{U}_s$ and $\overline{V}_s$ guarantees the decrease of cost function after each iteration, the algorithm can converge within finite steps.

Algorithm 1 Adaptive Matching Pursuit.

Input: $\boldsymbol{A}, \boldsymbol{y}$.

while true **do**
  $\boldsymbol{\omega}^s = \arg\min_{\boldsymbol{\omega}^s} \| \boldsymbol{y} - \boldsymbol{A}_s \boldsymbol{\omega}^s \|_2^2 + \lambda \sum_{k=1}^{K} \boldsymbol{\omega}_k^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k$
  Calculate: $[\overline{U}_s, k]$ and $[\overline{V}_s, j]$
  if $\min(\overline{U}_s, \overline{V}_s) > 0$ **then** break while
  else if $\overline{U}_s < \overline{V}_s$ **then** $s = s \cup \{k\}$
  else $s = s \setminus \{j\}$
  end if
  Update λ by Eq. (31)
  Update σ by Eq. (33)
  Update $\boldsymbol{B}$ by Eq. (34)
  Update $c : c = -\ln \frac{m_1}{m_0}$
end while

Output: $\boldsymbol{\omega}$, $\boldsymbol{\gamma}$, λ, c and σ

4. Fast implementation of GAMP

In this section, we recall a fast implementation ideal that uses the preconditioned conjugate gradient (PCG) to reduce the computation of the signal reconstruction process [24–26]. Then we incorporate it into our GAMP algorithm to acquire the fast GMAP (FGAMP). Looking back on the whole procedure of the GAMP method, we observe that the most time-consuming part is to calculate $\boldsymbol{\omega}^s$ in Eq. (14). The standard solution to Eq. (14) by solving the inverse matrix has a computational complexity of $\mathcal{O}(n^3)$, especially with a large n. To leverage the PCG, we rewrite Eq. (14) as

$$\begin{aligned}\boldsymbol{\omega}^s &= \arg\min_{\boldsymbol{\omega}^s} \| \boldsymbol{y} - \mathbf{A}_s\boldsymbol{\omega}^s \|_2^2 + \lambda \sum_{k\in s} (\boldsymbol{\omega}_k^s)^H \boldsymbol{B}^{-1} \boldsymbol{\omega}_k^s \\ &= \arg\min_{\boldsymbol{\omega}^s} \| \boldsymbol{y} - \mathbf{A}_s\boldsymbol{\omega}^s \|_2^2 + \lambda \| \boldsymbol{G}\boldsymbol{\omega}^s \|_2^2 \\ &= \arg\min_{\boldsymbol{\omega}^s} \| \bar{\boldsymbol{y}} - \boldsymbol{\phi}\boldsymbol{\omega}^s \|_2^2, \end{aligned} \tag{35}$$

where the matrix $\boldsymbol{L}$ is obtained by the Cholesky decomposition of matrix $\boldsymbol{B}^{-1}$ due to its positive definitive property and the matrix $\boldsymbol{G}$ is obtained by overlapping the matrix $\boldsymbol{L}$ to a block diagonal matrix

$$\boldsymbol{G} = \begin{bmatrix} \boldsymbol{L} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \boldsymbol{L} \end{bmatrix}. \tag{36}$$

The vector $\boldsymbol{s}$ and matrix $\boldsymbol{\phi}$ are obtained by

$$\boldsymbol{\phi} = \begin{bmatrix} \boldsymbol{A}_s \\ \sqrt{\lambda}\boldsymbol{G} \end{bmatrix}, \bar{\boldsymbol{y}} = \begin{bmatrix} \boldsymbol{y} \\ \boldsymbol{0} \end{bmatrix}. \tag{37}$$

The problem Eq. (35) is a least square problem which is equivalent to solve the following linear equations:

$$\boldsymbol{\phi}^H\boldsymbol{\phi}\boldsymbol{\omega}^s = \boldsymbol{\phi}^H\bar{\boldsymbol{y}}. \tag{38}$$

The preconditioned conjugate gradient (PCG) method and the Cholesky decomposition method are arguably the best known iterative and direct algorithms for solving this linear equation system. The Cholesky decomposition method has a complexity of $\mathcal{O}(n^2)$ and PCG has $\mathcal{O}(\eta(1+\log_2 n))$, where η denotes the number of iterations taken by the PCG. The PCG is faster than Cholesky decomposition when the size n is lager than 30 [25]. In PCG, the preconditioned matrix $\boldsymbol{D}$ is introduced. The idea is that matrix $\boldsymbol{\phi}^H\boldsymbol{\phi}$ itself maybe ill-conditioned, however, hopefully, by choosing some good preconditioned $\boldsymbol{D}$, the condition number of $\boldsymbol{D}\boldsymbol{\phi}^H\boldsymbol{\phi}$ will be much better. We can solve $\boldsymbol{D}\boldsymbol{\phi}^H\boldsymbol{\phi}\boldsymbol{\omega}^s = \boldsymbol{D}\boldsymbol{\phi}^H\bar{\boldsymbol{y}}$ by conjugate gradient method [26].

5. Simulations and experiments

In this section, simulations and MNIST image experiments are performed to show the effectiveness of the proposed GAMP. The mean square error (MSE) algorithm is used as performance index. Several state-of-the-art methods, including AMP [14], GLASSO [7] and BSBL [4] are considered for performance comparison. The hyperparameters used in the experiments are: $\lambda = 0.001, \sigma = 0.1$ and $\kappa_k = 0.48$ for $k \in \{1, \cdots, K\}$. All experimental results were obtained by averaging 100 trials and all experiments were conducted on a 2.80 GHz PC using Matlab R2018a.

5.1. 1-D simulated data

In this subsection, a Gaussian sensing matrix $\boldsymbol{A} \in \mathbb{R}^{256\times 512}$ is considered. The number of group K is 128 and the group size L is 4. The number of non-zero groups is 10, and the coefficients within groups are generated in the Toeplitz kernel matrix with $c = 0.5$. Without loss of generality, an additive noise is also considered with the signal-to-noise ratio (SNR) of 20 dB.

Table 1
Comparisons of runtime.

Method	AMP	GLASSO	BSBL	GAMP	FGAMP
Time(sec)	0.05	0.45	3.32	0.10	0.04
MSE	7.00e−2	7.94e−4	1.58e−4	7.90e−5	8.10e−5

Fig. 1 show the reconstruction results of the sparse signal. Owing to neglecting the group structure, it is observed that the AMP method has a poor reconstruction performance, shown in Fig. 1(a), with MSE of 0.07. On the other hand, the GLASSO and BSBL methods have improved performances, shown in Fig. 1(b) and (c), respectively. Fig. 1(d) and (e) show the reconstructed results based on the proposed GAMP and FGAMP methods, respectively. It is clear that the proposed methods accurately reconstruct the sparse signal with small MSE of 7.90×10^{-5} in GAMP and of 8.10×10^{-5} in FGAMP. This superior performance is due the exploitation of the group structure and the intra-group correlation learning in the proposed methods. The runtime and corresponding MSE of the above methods are also provided and compared in Table 1. It is observed that the greedy based methods, such as AMP, GAMP and FGAMP are slightly faster than GLASSO and BSBL methods, and the FGAMP method has the smallest runtime due to the use of the PCG.

Phase Transition: In this simulation, the MSE versus the number of measurements is firstly analyzed. All other parameters are the same as in the above simulation. The result is shown in Fig. 2(a). When the number of measurements m is from 80 to 230, the MSE generally decreases with increased m. It is observed that the MSEs for the proposed methods are generally lower than other methods and for $m = 94$, the MSE decreases to -32.4 dB. The corresponding MSE values are -25.48 dB, -15.18 dB and -12.28 dB for BSBL, GLASSP and AMP, respectively. Both GAMP and FGAMP methods have almost the same reconstruction performance, attributing to the fact that the only difference between them lies in the solution to the objective function in Eq. (14). In addition, F1 Score, which is an important and commonly used performance metric for binary prediction tasks, and is denoted as

$$\text{F1 Score} = 2\frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

is also used as an extra performance index for the non-zero support evaluation of sparse signal [27]. The “Precision” is the number of True Positives (TP) divided by sum of TP and the number of False Positives (FP), and it can be thought of as a measure metric of a sparse reconstruction exactness. The “Recall” is TP divided by TP and the number of False Negatives (FN), and it can be regarded as a measure metric of a sparse reconstruction completeness. It is clear that F1 Scores generally increase with increased measurement numbers among all methods, as shown in Fig. 2(b). It is observed that the F1 Scores in both BSBL and the proposed methods almost approach to 1, and it means that these approaches are capable of accurately recovering non-zero supports of sparse signals. However, the proposed methods has higher reconstruction accuracy and lower MSE than that in BSBL, shown in Fig. 2(a).

Reconstruction performances in group-sparsity levels: The MSE with respect to group-sparsity levels $p \ll K$, which is defined as the number of non-zero groups, is shown in Fig. 3(a). It is observed that the reconstruction performances generally decay with increased sparsity level. The methods with the group structure, such as GLASSO, BSBL, FAMP and GAMP methods, have lower MSE than that of the AMP method. But, the BSBL, GAMP and FGAMP methods have better performances compared to GLASSO, due to the intra-correlation learning. The proposed GAMP and FGAMP methods have MSEs of -20 dB across the sparsity levels from 2 to 16. We use success rate, which is defined as the ratio between the

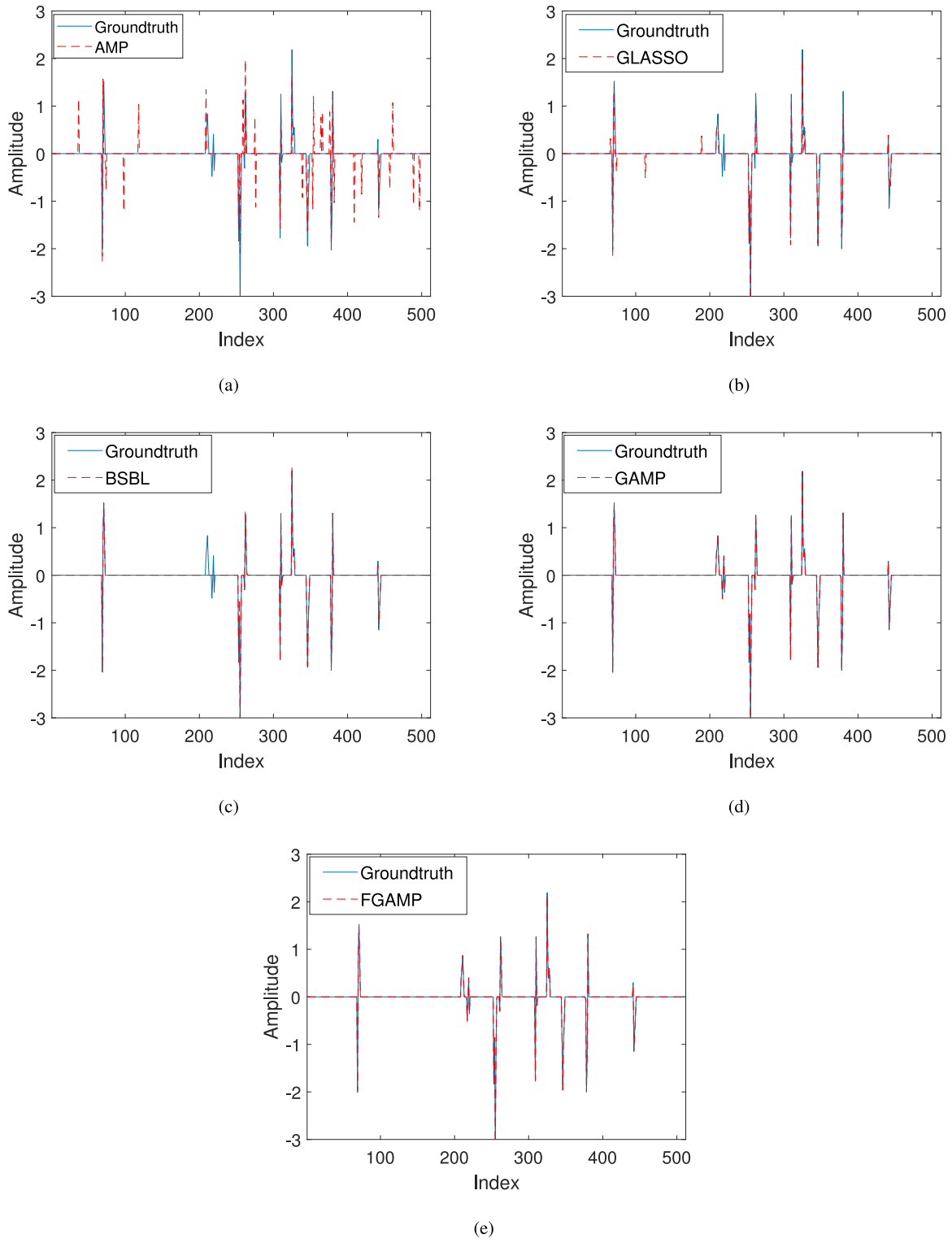
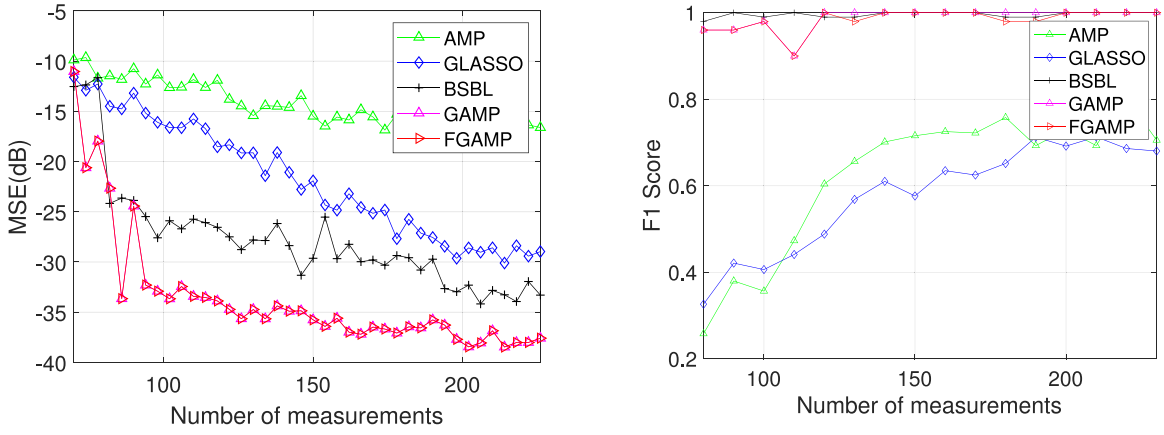


Fig. 1. Reconstruction performance comparisons in the 1-D sparse signal. (a) AMP, (b) GLASSO, (c) BSBL, (d) GAMP, (e) FGAMP.

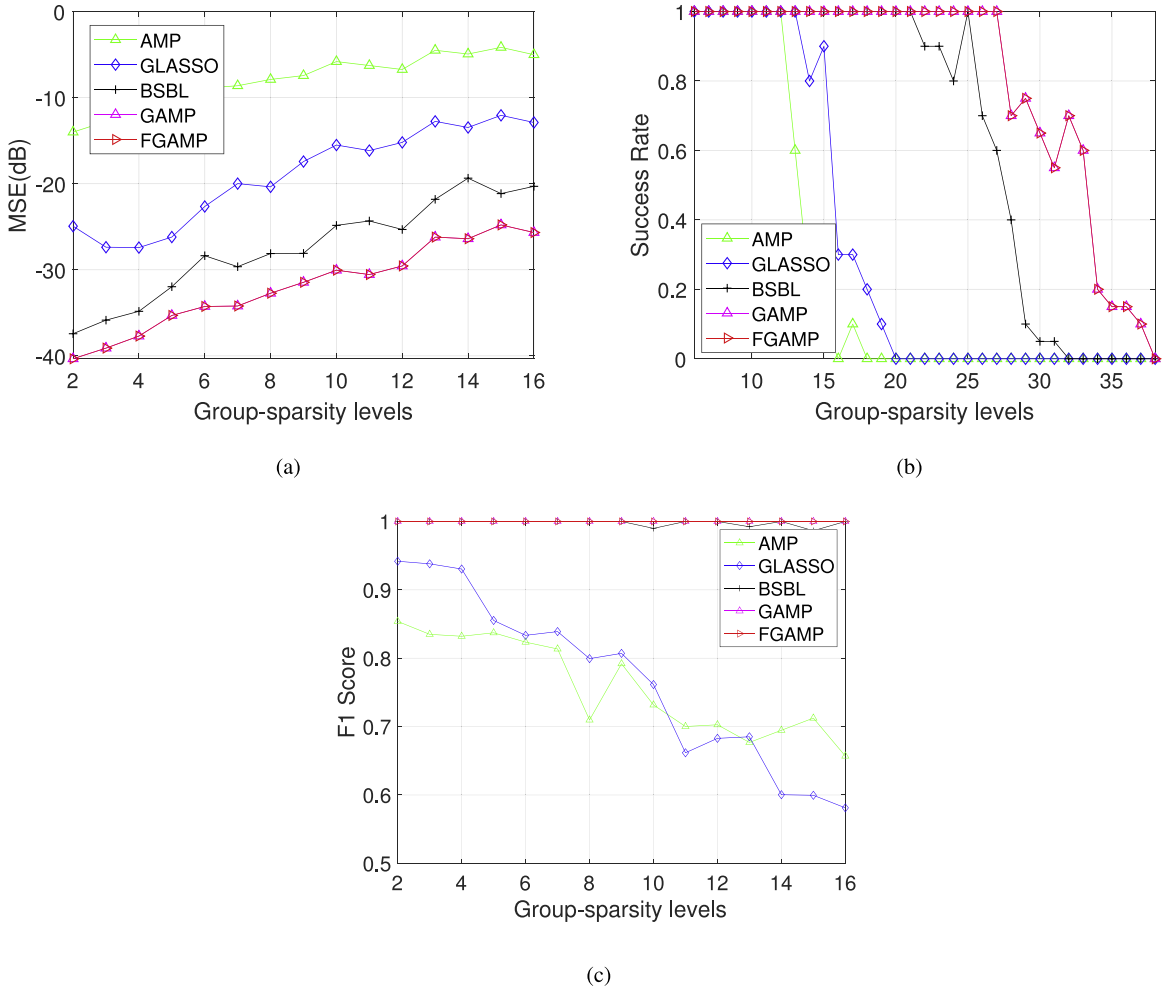
total number of success events whose MSE < 0.01 (corresponding to -20 dB) and the total number of trials, as an additional performance index. It is found that the success rate maintains unite value until the group-sparsity level reaches 27. The corresponding values in the BSBL, GLASSO and AMP methods are 21, 13 and 12, respectively. The reconstruction fails when the group-sparsity level is beyond 38 for the proposed methods. The corresponding value is 32 in the BSBL method, 20 in the GLASSO method, and 16 in the AMP method, as shown in Fig. 3(b). Furthermore, F1 Scores in the proposed method have constant values of 1 across sparsity levels, and show the superiorities of the proposed methods over other methods, shown in Fig. 3(c).

Intra-group correlation parameter estimation: The above results demonstrate that BSBL, FGAMP and FGAMP generally have better performances than GLASSO and AMP due to learning of the intra-group correlation. Fig. 4 shows the parameter estimation results of



(a) (b)

Fig. 2. MSE and F1 score versus the number of measurements. (a) MSE versus the number of measurements, (b) F1 score versus the number of measurements.



(a) (b)

(c)

Fig. 3. Reconstruction performance comparisons. (a) MSE versus group-sparsity levels, (b) Comparison of the reconstructed success rate, (c) F1 Score versus group-sparsity levels.

correlation c. It is noticed that the estimated values of c through GAMP and FGAMP are the closest to the true value. One possible reason is that the proposed methods have smaller MSE, and they are capable of more accurately modelling and learning the dependency of elements within groups, compared to the BSBL method.

Reconstruction performances in SNR: Fig. 5(a) shows reconstructed MSE versus SNR. It is observed that MSE generally decreases with increased SNR for all methods, with the proposed methods showing the lowest values. The MSE in the AMP method appears to almost flatten for SNR $>$ 10 dB. The reason could be attributed to the effect of the relatively limited number of observations in absence of group structure consideration. There exists about 4 dB difference between the proposed methods and other methods, when SNR is below 0. The difference between the BSLB

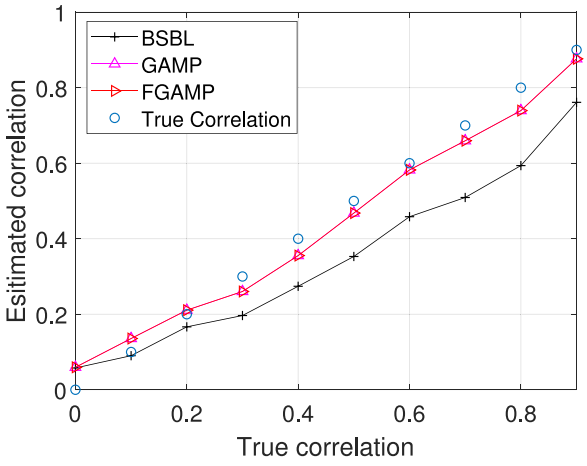


Fig. 4. Correlation parameter estimation.

method and the proposed methods becomes slightly smaller when SNR is above 10, while the difference above this SNR becomes larger between the proposed methods and the others. Fig. 5(b) shows that F1 Score versus SNR. It is clear that the proposed methods generally have the highest values among all methods. When the SNR increases up to 5 dB, the values of F1 Score in the proposed methods approach to 1. However, it is at least 10 dB in the BSBL method.

5.2. MNIST image recovery

Reconstructions of handwritten digital images from the MNIST dataset [28] are also considered for performance comparison. The dataset contains 60000 digital images of size 28 × 28 pixels. Most of the pixels in these images are inactive and the nonzero pixels are grouped together. That is, if one pixel is nonzero, the pixels adjacent to that pixel are typically nonzero. Therefore, these images are naturally group sparse and can be effectively recovered by the proposed methods. The measurements $\boldsymbol{y}$ are obtained by projecting the aligned vector $\boldsymbol{\omega} \in \mathbb{R}^{784\times 1}$ on a randomly generated Gaussian matrix $\boldsymbol{A} \in \mathbb{R}^{300\times 784}$. Additive noise with SNR of 5 dB is also considered. In this experiment, the group size is $L = 4$. The recovered images and their corresponding MSE are shown in Fig. 6. It is evident that the images recovered by the proposed methods are closest to the original images and have MSEs of 1.05×10^{-3} and 1.09×10^{-3}.

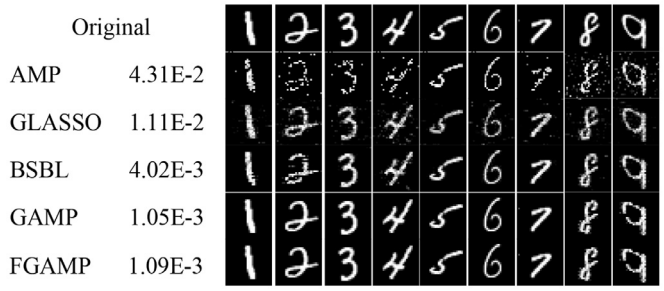


Fig. 6. Reconstructed images from MNIST dataset. The numbers appeared next to each method is the average MSE.

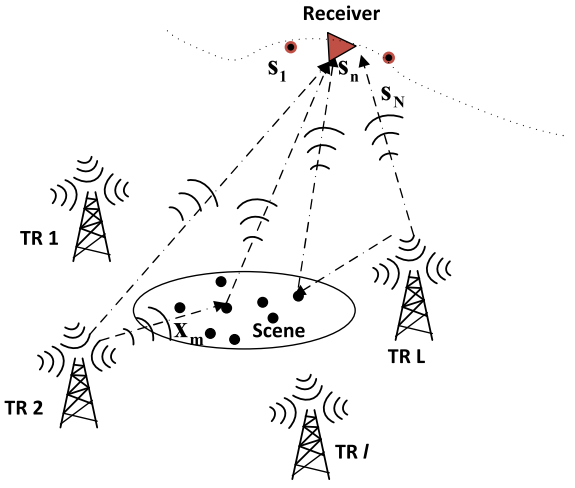
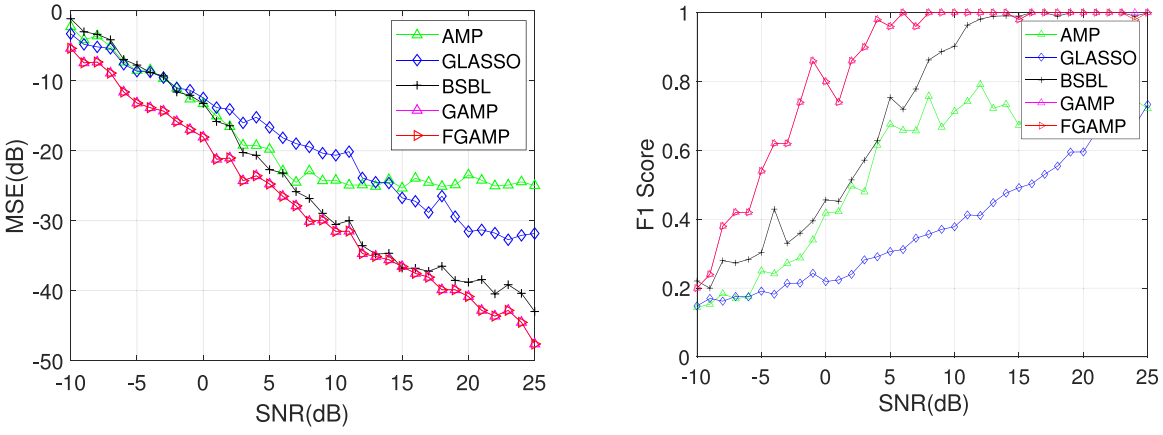


Fig. 7. Geometry of a multi-static passive SAR system.

5.3. Multi-static passive radar imaging

Passive radar systems, which utilize broadcast and navigation signals as sources of opportunity, have attracted significant interests in recent years due to their low cost, covertness, and availability of rich illuminator sources. Consider a multi-static passive synthetic aperture radar (SAR) system consisting of L stationary transmitters and a moving receiver, as depicted in Fig. 7. The receiver, which is mounted on an airborne platform, has the capability of receiving reflected signals from the targets and direct-path signals



(a) (b)

Fig. 5. Reconstruction MSE and F1 Scores versus SNR. (a) MSE versus SNR, (b) F1 Score versus SNR.

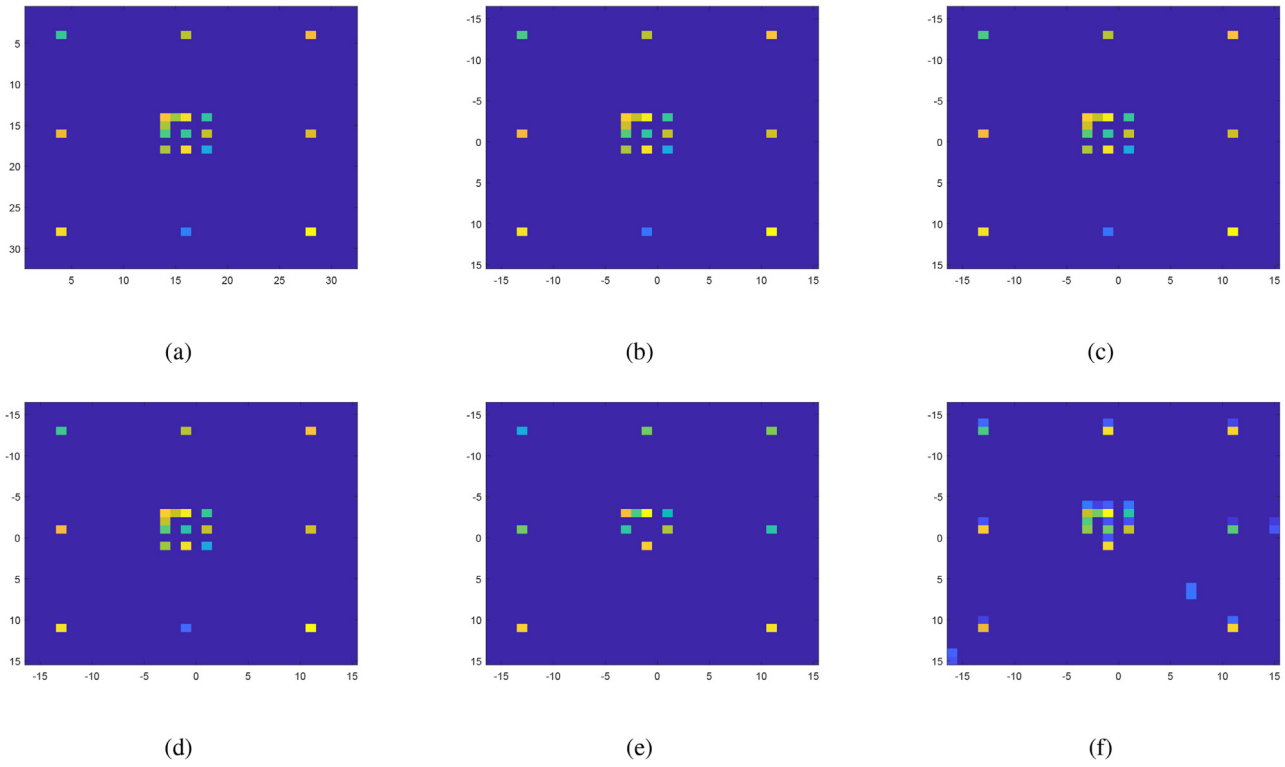


Fig. 8. Reconstruction results in the multi-static passive SAR system. (a) Normalized magnitude of the scattering coefficients of the original scene, (b) Reconstructed result based on GAMP, (c) Reconstructed result based on FGAMP, (d) Reconstructed result based on BSBL, (e) Reconstructed result based on GLASSO, (f) Reconstructed result based on BOMP.

from the transmitters, and forms the synthetic aperture for high azimuth resolution. According to [11,29], multi-static SAR imaging configuration can be equal to multiple bistatic cases. Furthermore, the multi-static passive SAR imaging can be regarded as a typical sparse reconstruction problem based on the forward model. By stacking the echoes with respect to the observation aperture positions in the lth bistatic pair, we obtain the lth bistatic SAR imaging

$$\boldsymbol{y}_l = \boldsymbol{\Phi}_l \boldsymbol{\omega}_l + \boldsymbol{\epsilon}_l, \quad l \in [1, \cdots, L], \tag{39}$$

where $\boldsymbol{y}_l$ is the observed data in the lth bistatic SAR, $\boldsymbol{\Phi}_l$ is the sensing matrix based on the SAR forward model, $\boldsymbol{\omega}_l$ is the reconstructed sparse scattering coefficients vector, and $\boldsymbol{\epsilon}_l$ is the additive noise. It should be noted that the supports of the sparse targets are approximately identical across the L bistatic geometries, (i.e., the non-zero entries of $\boldsymbol{\omega}_l$ lie in the same positions across different values of l).

Stacking all L bistatic observed data into a vector, and we have

$$\boldsymbol{y} = \boldsymbol{\Phi}\boldsymbol{\omega} + \boldsymbol{\epsilon}, \tag{40}$$

where $\boldsymbol{y} = [\boldsymbol{y}_1, \cdots, \boldsymbol{y}_L]^T$, $\boldsymbol{\Phi} = [\phi_{11}, \cdots, \phi_{1L}, \cdots, \cdots, \phi_{KL}]$, and

$$\boldsymbol{\omega} = \left[\underbrace{\omega_{11}, \cdots, \omega_{1L}}_{\boldsymbol{\omega}_1^T}, \cdots, \underbrace{\omega_{K1} \cdots \omega_{KL}}_{\boldsymbol{\omega}_K^T} \right]^T. \tag{41}$$

Notice that all scattering coefficients in each group should share identical sparse patterns, that is, all L elements from L bistatic SAR geometries within a group are nonzeros, when a target exists in the corresponding spatial location; otherwise, all the elements within the group are zeros. Further, the scattering coefficients within the group should be generally correlated, according to the electromagnetic backscattering theory, and the correlation of elements would be dependent on the observed aspects among multi-static geometries. When two transmitters are spatially close, the corresponding scattering coefficients should be highly correlated. Re-examining Eq. (1), the multi-static passive SAR imaging can be exactly regarded as sparse reconstruction problem, and can be solved by the proposed GAMP method.

Following the simulations in [29], DVB-T signals with a carrier frequency of 850 MHz and a bandwidth of 7.8 MHz, which corresponds to around 20 m range resolution, is used. The four illuminators are located 10 km away from the scene center with their respective aspect angles of -45°, -15° and 15°. These illuminators emit their individual DVB-T waveforms with different frequencies. The correlation of scattering coefficients among 3 bistatic geometries is assumed 0.9 due to similar spatial bistatic aspects. 48 synthetic aperture positions are acquired by uniformly dividing synthetic aperture width.

As shown in Fig. 8(a), the sparse scene consists of 32×32 pixels. With the consideration of the sparsity measure that most signals can be recovered by the l_1 norm minimization, we choose the total number of sparse targets to be $Q = 19$. The inter-pixel spacings are 1m in both range and azimuth.

The performance comparisons of the proposed method with other state-of-the-art algorithms are as follows. Fig. 8(b) to (f) represent the recovered images of GAMP, FGAMP, BSBL, GLASSO and BOMP, respectively. As shown in Fig. 8(e) and (f), the GLASSO and BOMP algorithms yield very poor image qualities. The GLASSO algorithm is highly dependent on the hyper-parameter to determine the sparsity level. The BOMP algorithm is suited to recover sparse signals when the measurement matrix has low coherence. The proposed GAMP has the best performance, as shown in Fig. 8(b), because of the employed golden standard prior - "spike-and-slab" and the correlation learning strategy. The MSEs are shown in

Table 2
Comparison of reconstruction MSE.

Methods	SNR=-5dB	SNR=0dB	SNR=5dB	SNR=10dB
GAMP	**5.92e−4**	**1.25e−4**	**3.99e−5**	**1.26e−5**
FGAMP	5.95e−4	1.26e−4	4.12e−5	1.33e−5
BSBL	7.62e−4	1.99e−4	1.15e−4	9.05e−5
GLASSO	4.20e−3	2.2e−3	2.00e−3	1.90e−3
BOMP	1.78e−2	7.2e−3	3.10e−3	3.01e−3

Table 2. In essence, the proposed GAMP achieves superior performance in reconstructing super-resolution imaging in a multi-static scenario characterized as a group sparsity model.

6. Conclusion

Compressive sensing techniques are capable of reconstructing sparse signals from a small number of measurements. In addition to the sparse property, the coefficients of real-world sparse signals often exhibit special structures, i.e., group structure. In this paper, a novel greedy-based group adaptive matching pursuit algorithm is proposed for the recovery of group sparse signals. A generative model in the sparse signal recovery with exploiting the group structure and correlation learning within groups is provided. We apply a group spike-and-slab prior to capture the sparse signal with group structure, and introduce a kernel matrix to learn and model the correlation of elements within groups. A greedy based group adaptive matching pursuit approach is proposed to address the non-convex optimization problem resulting from the prior. A fast implementation approach is also provided and examined based on the preconditioned conjugate gradient method technique. The proposed method significantly improves the reconstruction performance of the sparse signals by jointly exploiting sparse reconstruction methods.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Qisong Wu: Conceptualization, Methodology, Writing - review & editing, Funding acquisition, Supervision. **Jiahao Liu:** Data curation, Writing - original draft, Software, Validation. **Moeness G. Amin:** Writing - review & editing.

Acknowledgements

The work was supported in part National Natural Science Foundation under Grants no. 61701109, 11704069, and 11574048, in part supported by National Natural Science Foundation of Jiangsu Province under Grant no. BK20160701, in part by Science and Technology on Sonar Laboratory under Grant no. 6142109180202, and in part by National Defense Basis Scientific Research program of China under Grant no. JCKY2019110.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.sigpro.2020.107560.

References

[1] D.L. Donoho, Compressed sensing, IEEE Trans. Inf. Theory 52 (4) (2006) 1289–1306.
[2] R.G. Baraniuk, Compressive sensing [lecture notes], IEEE Signal Process. Mag. 24 (4) (2007) 118–121.
[3] E.J. Candes, M.B. Wakin, An introduction to compressive sampling, IEEE Signal Process. Mag. 25 (2) (2008) 21–30.
[4] Z. Zhang, B.D. Rao, Extension of SBL algorithms for the recovery of block sparse signals with intra-block correlation, IEEE Trans. Signal Process. 61 (8) (2013) 2009–2015.
[5] L. Jacob, G. Obozinski, J.P. Vert, Group lasso with overlap and graph Lasso, in: Proc. International Conference on Machine Learning, 2009, pp. 433–440. Montreal, Quebec, Canada
[6] V.D.B. Ewout, M.P. Friedlander, Probing the pareto frontier for basis pursuit solutions, SIAM J. Sci. Comput. 31 (2) (2008) 890–912.
[7] M. Yuan, Y. Lin, Model selection and estimation in regression with grouped variables, J. R. Stat. Soc. 68 (1) (2006) 49–67.
[8] T.J. Mitchell, J.J. Beauchamp, Bayesian variable selection in linear regression, Publ. Am. Stat.Assoc. 83 (404) (1988) 1023–1032.
[9] Q. Wu, Y.D. Zhang, M.G. Amin, B. Himed, Complex multitask Bayesian compressive sensing, in: Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2014. Florence, Italy
[10] Q. Wu, Y.D. Zhang, M.G. Amin, B. Himed, Mutli-task Bayesian compressive sensing exploiting intra-task dependency, IEEE Signal Proc. Lett. 22 (4) (2015) 430–434.
[11] Q. Wu, Y.D. Zhang, M.G. Amin, B. Himed, High-resolution passive SAR imaging exploiting structured Bayesian compressive sensing, IEEE J. Sel. Top. Signal Process. 9 (8) (2015) 1484–1497.
[12] H.S. Mousavi, V. Monga, T.D. Tran, Iterative convex refinement for sparse recovery, IEEE Signal Process. Lett. 22 (11) (2015) 1903–1907.
[13] Q. Wu, Y.D. Zhang, M.G. Amin, B. Himed, Space-time adaptive processing and motion parameter estimation in multistatic passive radar using sparse Bayesian learning, IEEE Trans. Geosci. Remote Sens. 54 (2) (2016) 944–957.
[14] T.H. Vu, H.S. Mousavi, V. Monga, Adaptive matching pursuit for sparse signal recovery, in: IEEE International Conference on Acoustics, Speech and Signal Processing, 2017, pp. 4331–4335.
[15] M.K. Titsias, Spike and slab variational inference for multi-task and multiple kernel learning, Adv. Neural Inf. Process. Syst. (2011).
[16] L. Yu, H. Sun, J.P. Barbot, Bayesian compressive sensing for cluster structured sparse signals, Signal Process. 92 (1) (2012) 259–269.
[17] H. Daniel, J. Hernndez-Lobato, P. Dupont, Generalized spike-and-slab priors for Bayesian group feature selection using expectation propagation, J. Mach. Learn. Res. 14 (1) (2013) 1891–1945.
[18] M.R. Andersen, O. Winther, L.K. Hansen, Bayesian inference for structured spike and slab priors, in: Proc. Adv. Neural Info. Proc. Syst., 2014. Montreal, Canada
[19] M.R. Andersen, A. Vehtari, O. Winther, Bayesian inference for spatio-temporal spike-and-slab priors, J. Mach. Learn. Res. 18 (139) (2017) 1–58.
[20] Q. Wu, S. Fang, Structured Bayesian compressive sensing with spatial location dependence via variational Bayesian inference, Digit. Signal Process. 71 (2017) 95–107.
[21] Z. Zhang, B.D. Rao, Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning, IEEE J. Sel. Top. Signal Process. 5 (5) (2011) 912–926.
[22] C.M. Bishop, Pattern Recognition and Machine Learning, Springer, New York, NY, 2006.
[23] I. Csiszar, Information geometry and alternating minimization procedures, Stat. Decis. 1 (1984) 205–237.
[24] J.R. Shewchuk, An Introduction to the Conjugate Gradient Method Without the Agonizing Pain, Carnegie Mellon University, 1994.
[25] A. Hasan, E.C. Kerrigan, G.A. Constantinides, Solving a positive definite system of linear equations via the matrix exponential, in: Decision and Control and European Control Conference, 2011, pp. 2299–2304.
[26] G.H. Golub, C.F.V. Loan, Matrix Computations, (third ed.), JHU Press, 1996.
[27] V. Rijsbergen, C. Joost, The Geometry of Information Retrieval, Cambridge University Press, 2004.
[28] Y. Lecun, The MNIST database of handwritten digits, http://yann.lecun.com/exdb/mnist/ (1998).
[29] Q. Wu, Y.D. Zhang, M.G. Amin, B. Himed, Multi-static passive sar imaging based on bayesian compressive sensing, in: Proc. International Society for Optics and Photonics, 2014. Baltimore, USA.