Research Article

A Greedy Multistage Convex Relaxation Algorithm Applied to Structured Group Sparse Reconstruction Problems Based on Iterative Support Detection

Liangtian He and Yilun Wang

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu, Sichuan 611731, China

Correspondence should be addressed to Yilun Wang; yilun.wang@uestc.edu.cn

Received 27 April 2014; Revised 27 August 2014; Accepted 29 August 2014; Published 21 October 2014

Academic Editor: Yi-Kuei Lin

Copyright © 2014 L. He and Y. Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

We propose a new effective algorithm for recovering a group sparse signal from very limited observations or measured data. As we know that a better reconstruction quality can be achieved when encoding more structural information besides sparsity, the commonly employed $\ell_{2,1}$ -regularization incorporating the prior grouping information has a better performance than the plain ℓ_1 -regularized models as expected. In this paper we make a further use of the prior grouping information as well as possibly other prior information by considering a weighted $\ell_{2,1}$ model. Specifically, we propose a multistage convex relaxation procedure to alternatively estimate weights and solve the resulted weighted problem. The procedure of estimating weights makes better use of the prior grouping information and is implemented based on the iterative support detection (Wang and Yin, 2010). Comprehensive numerical experiments show that our approach brings significant recovery enhancements compared with the plain $\ell_{2,1}$ model, solved via the alternating direction method (ADM) (Deng et al., 2013), either in noiseless or in noisy environments.

1. Introduction

1.1. Group Sparse Reconstruction and Related Work. It is becoming a hot research topic to find sparse solutions of underdetermined linear systems in the last few years in various fields, for example, compressive sensing (CS), signal processing, statistics, and machine learning [1], for example, multiple kernel learning [2], microarray data analysis [3], and channel estimations in doubly dispersive multicarrier systems [4]. For example, in machine learning, the high dimensionality poses significant challenges for us to build interpretable models with high prediction accuracy, and many sparsity related regularization techniques have been commonly utilized to obtain more stable and interpretable models. In compressive sensing, the sparsity regularization allows us to reconstruct high dimensional data with only a small number of samples. However, recent studies encourage us to go beyond sparsity to further enhance the recoverability, that is, taking into account additional information about the underlying structure of the solutions. As an important

case, lots of solutions are known to have certain groups sparsity structure, and more precisely, not only do they have a natural grouping division of its components, but also the components within a group are likely to be all nonzeros or all zeros. Thus encoding the group sparsity structure will reduce the degrees of freedom in the solution, resulting in better recovery performance.

In this paper, we focus on the above group sparsity and corresponding commonly used $\ell_{2,1}$ regularized model and try to extend it for better recovery performance. Assume that $x \in \mathbb{R}^n$ denotes an unknown group sparse solution. Let $\{x_{g_i} \in \mathbb{R}^{n_i} : i = 1, 2, \dots, s\}$ be the grouping division of x , where $g_i \subseteq \{1, 2, \dots, n\}$ is an index set corresponding to the i th group, and x_{g_i} denotes the subvector of x indexed by g_i , which is predefined based on the prior information of the underlying solution generally. The mixed $\ell_{2,1}$ norm is defined as follows:

$$\|x\|_{2,1} = \sum_{i=1}^s \|x_{g_i}\|_2. \quad (1)$$

Compared with the classical using of ℓ_1 -regularization for sparse reconstruction, $\ell_{2,1}$ -regularization will take the grouping information into consideration and facilitate group sparsity. Notice that the resulted problem is convex, and this mixed $\ell_{2,1}$ optimization problem (1) is commonly solved through several efficient first-order algorithms proposed in the literature, for example, spectral projected gradient method (SPGL1) [5], accelerated gradient method (SLEP) [6], block-coordinated descent algorithms [7], and SpaRSA [8]. Among them, the authors proposed the alternating direction method (ADM) to solve the primal and dual formulation of $\ell_{2,1}$ regularized optimization problems in [9]. Their preliminary numerical results have shown that the ADM algorithms are fast, stable, and robust, outperforming the previously known state-of-art algorithms.

1.2. Weighted $\ell_{2,1}$ -Norm Regularized Group Sparse Reconstruction Problem. In this paper, for better reconstruction quality, instead of considering the $\ell_{2,1}$ -norm, we consider a more general formulation, that is, the weighted $\ell_{2,1}$ (or $\ell_{w,2,1}$)-norm [9] defined as follows:

$$\|x\|_{w,2,1} = \sum_{i=1}^s w_i \|x_{g_i}\|_2, \quad (2)$$

where $w_i > 0$ ($i = 1, 2, \dots, s$) are the corresponding weights associated with each group, consisting of a weighting vector $\mathbf{w} = [w_1, w_2, \dots, w_s]$. We assume that the groups $\{x_{g_i} : i = 1, 2, \dots, s\}$ form a division of x unless otherwise specified. We emphasize that the weighted $\ell_{2,1}$ -norm can also be extended to the more general group configurations, for example, overlapping and/or incomplete cover group sparsity.

As a nonconvex model, (2) is well known to behavior better than the unweighted counterpart with appropriate settings of weights (1). The key is how to determine the weights and it constitutes the main contribution as summarized in Section 1.3.

We will consider several weighted models in this paper. The following basis pursuit (BP) model is considered when the measurement vector b do not contain noise:

$$\min_x \|x\|_{w,2,1} \quad \text{s.t. } Ax = b, \quad (3)$$

where $A \in \mathbb{R}^{m \times n}$ ($m < n$) and $b \in \mathbb{R}^m$. Without loss of generality, we suppose A has full rank. The basis pursuit denoising (BPDN) models are commonly employed when the measurement vector b contains noise, including the following constrained form:

$$\min_x \|x\|_{w,2,1} \quad \text{s.t. } \|Ax - b\|_2 \leq \sigma, \quad (4)$$

and the unconstrained form

$$\min_x \|x\|_{w,2,1} + \frac{1}{2\mu} \|Ax - b\|_2, \quad (5)$$

where $\sigma \geq 0$ and $\mu > 0$ are penalty parameters, respectively. It should be noted that the constrained form is equivalent to unconstrained form from the viewpoint of optimization theory when the parameters σ and μ are properly selected.

In this paper, we will focus on the basis pursuit model, and the extension of our proposed algorithms for the basis pursuit denoising models (4) and (5) follows similarly. Furthermore, it should be pointed out that the basis pursuit model (3) can also be good for reconstructing noisy data if the iterations are stopped properly prior to convergence based on the noise level.

1.3. Contributions. In this paper, we propose an effective way to determine the weights of the above weighed models, by extending the iterative support detection (ISD) proposed in [10], from sparsity to group sparsity, which is a special case of structure sparsity [11–13]. In other words, for the weighted $\ell_{2,1}$ reconstruction problem in this paper, based on ISD, we obtain the final result via multistage convex relaxation process [14], which consists of solving a series of weighted $\ell_{2,1}$ (or referred to as $\ell_{w,2,1}$) convex optimization problems, where weights have been estimated via the support detection applied to the reconstructed signal of the previous stage. The solution of multistage weighted $\ell_{2,1}$ model is usually better than the traditional $\ell_{1,2,1}$ model in terms of the relative error and reconstruction quality, from both the theoretical and practical points of view, and numerical results demonstrate that our proposed algorithm can recover a satisfying result from a failed reconstruction of $\ell_{1,2,1}$ model in our cases.

In addition, we empirically demonstrate that, in the cases of grouping sparsity, the previous requirements of the fast decaying property in the common plain sparse signal recovery are not necessary any longer, for the *threshold*-ISD proposed in [10], and we will give some intuitive explanation.

1.4. Organization. The rest of this paper is organized as follows. In Section 2, we present our algorithmic framework. In Section 3, we provide comprehensive numerical experiments to evaluate the performance of our proposed algorithm for group sparse signal reconstruction and compare it with the ADM approach proposed in [9]. We end this paper with some conclusions and discussions on some possible future work.

2. Algorithmic Framework

In this section, without loss of generality, we suppose that the grouping is a partition of the solution, which is predefined as a prior knowledge, and it can be easily extended to the general group configurations, for example, overlapping and incomplete cover grouping cases.

The main difficulty of setting the weights is that we do not usually know the true solution and that even when we have some knowledge about the true solution, we still need to find a proper way to make use of it to help obtain a better solution. The iterative support detection is an effective way to deal with the above difficulty and we will generalize this idea to the cases of group sparsity. So we first review the iterative support detection.

2.1. Revisiting and Some New Thoughts of Iterative Support Detection. Iterative support detection was first proposed in

our early work [10], and the idea of exploiting partially support detection arises in several subsequent literatures, for example, [15–20]. We first briefly review the iterative support detection in compressive sensing in terms of the single sparse signal reconstruction [10]. In addition, we give some novel analysis of the advantages of 0-1 weighting scheme adopted in ISD compared with other weighting alternatives [21, 22]. Compressive sensing (CS) [23, 24] reconstructs a sparse signal from a small set of linear projections. Let $\bar{x}$ denote a k -sparse signal, let $A \in \mathbb{R}^{m \times n}$ be the measurement matrix, and $b = A\bar{x}$ represent the linear projections of $\bar{x}$. The general optimization model is the basis pursuit (BP) problem:

$$(\text{BP}) \min_x \|x\|_1 \quad \text{s.t. } Ax = b. \quad (6)$$

ISD alternatively calls its two components: support detection and signal reconstruction. Support detection identifies an index set I from an incorrect reconstruction, which contains some elements of $\text{supp}(\bar{x}) = \{i : x_i \neq 0\}$. After acquiring the support detection, a resulting truncated BP problem is considered:

$$(\text{Truncated BP}) \min_x \|x_T\|_1 \quad \text{s.t. } Ax = b, \quad (7)$$

where $T = I^C$ and $\|x_T\|_1 = \sum_{i \notin I} |x_i|$. Let $\bar{x}$ be the true sparse signal. If the support detection $I = \text{supp}(\bar{x})$, then the solution of (7) is, of course, equal to true $\bar{x}$. But we should point out that even if I contains enough, not necessarily all, entries of $\text{supp}(\bar{x})$, a better solution can still be expected. When I does have enough $\text{supp}(\bar{x})$, those entries of $\text{supp}(\bar{x})$ in I will help (7) return a better solution in comparison to (6), and from this better solution, support detection will be able to identify more entries in $\text{supp}(\bar{x})$ and then yield an even better I . In this way, the two components of ISD work together to gradually recover $\text{supp}(\bar{x})$ and improve the reconstruction performance. It is clear that ISD is a multistage procedure.

ISD requires reliable true support detection from inexact reconstruction, which can be obtained by taking advantages of the features and prior information about the original true signal $\bar{x}$ [10, 25]. For example, for the sparse or compressible signals with components having a fast decaying distribution of nonzeros in [10]. One can perform the support detection by thresholding the solution of (7), and the corresponding ISD implementation is denoted as *threshold-ISD*.

In this paper, we would like to present some further analysis for ISD. ISD adopts a specific 0-1 weighting scheme and it has several advantages. First, its performance over the single stage pure ℓ_1 model has been proved rigorously in [10] once we can detect correct partial support information (by thresholding, e.g., in this paper), while most of the related alternatives such as the reweighted ℓ_1 algorithm [21] do not have such rigorous theoretical guarantees. Secondly, for the reweighted ℓ_1 algorithm, the weights are usually set like $\theta_i = 1/(|x_i| + \epsilon)$. In [10], we have pointed out that the tuning parameter $\epsilon > 0$ is the key parameter and should be determined carefully. Roughly, ϵ should not be a fixed value and should decrease from a large value to a small value. In an extreme case, if ϵ is always fixed to 0, then we would not get a better solution for the next stage, even if there is no

numerical trouble of dividing 0, because we just passively make use of all the information of the current solution without any filtering out for the inaccurate information. From our analysis, the choice of ϵ is like controlling the extraction of the useful information and suppressing the distortion of the recovery noise. Our 0-1 weights of ISD are a more explicit and straightforward way to imply the idea of making use of the correct information (mainly about the locations of components of large magnitude and setting the corresponding weights as 0) and give up the rest of the too noisy information (for those components of small magnitudes, they are mostly overwhelmed by the recovery noise and therefore there is very little meaning to set different weights according to their magnitudes; so it is more reasonable to set the same weights as 1).

We need to point out that the estimation of w based on the support detection is often advantageous. The implementations of support detection can be more flexible and different for specific signals, in order to make use of their different underlying structure, for example, grouping structure, tree structure, or even graph structure.

2.2. ISD Extended to the Weighted $\ell_{2,1}$ Model. The main effort of this paper is to extend the idea of ISD from the plain sparse vector recovery to the group sparsity cases and demonstrate the extraordinary advantages of thresholding support detection in these cases, compared to the plain sparsity cases.

As the original ISD does, our extension is also in general an alternating optimization procedure to decouple the non-linear combination of w and x , by repeatedly applying the following two steps.

Step 1. First we optimize x with w fixed (initially $\bar{1}$): this is a convex problem in x .

Step 2. Second we determine the value of w according to the currently reconstructed x . The value of w will be used in Step 1 of the next iteration.

The plain $\ell_{1,2,1}$ model, as a single stage process, is solved once generally and treats the solution as the final result. For the weighted $\ell_{2,1}$ model, we will obtain the final result from a multistage process by solving a series of weighted $\ell_{2,1}$ problems. At each stage, the adaptive weights will change according to the newly reconstructed signal. The full procedure and the details of Steps 1 and 2 will be presented in the following sections.

2.3. Step 1: Solving the Weighted $\ell_{2,1}$ Model Given Weights. Assume the weights are given; [9] proposed the approach for solving the weighted $\ell_{2,1}$ -problem (3), based on the variable splitting technique and the alternating direction method (ADM, for short) [26–32]. However, how to select proper weights was not presented and they used the uniform weights in their numerical experiments, that is, plain $\ell_{1,2,1}$ model. Here, we briefly review their approach, using the nonoverlapping case. But in the numerical experiments, the overlapping group sparsity [33] will also be considered.

2.3.1. Applying ADM to the Primal Weighted $\ell_{2,1}$ Model. For the primal $\ell_{w,2,1}$ -problem (3), we first introduce an auxiliary variable z and transform (3) as an equivalent constrained optimization problem as follows:

$$\min_{x,z} \left\{ \|z\|_{w,2,1} = \sum_{i=1}^s w_i \|z_{g_i}\|_2 : z = x, Ax = b \right\}. \quad (8)$$

The corresponding augmented Lagrangian function of (8) is defined as

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(x, z, \lambda_1, \lambda_2) \triangleq & \|z\|_{w,2,1} - \lambda_1^T (z - x) + \frac{\beta_1}{2} \|z - x\|_2^2 \\ & - \lambda_2^T (Ax - b) + \frac{\beta_2}{2} \|Ax - b\|_2^2, \end{aligned} \quad (9)$$

where $\lambda_1 \in \mathbb{R}^n, \lambda_2 \in \mathbb{R}^m$ are multipliers and $\beta_1, \beta_2 > 0$ are penalty parameters, respectively. If we start at $x = x^k$ and $(\lambda_1^k, \lambda_2^k)$, the iterative framework of the augmented Lagrangian problem has the following form:

$$\begin{aligned} z^{k+1} &\leftarrow \arg \min_z \mathcal{L}_{\mathcal{A}}(x^k, z, \lambda_1^k, \lambda_2^k), \\ x^{k+1} &\leftarrow \arg \min_x \mathcal{L}_{\mathcal{A}}(x, z^{k+1}, \lambda_1^k, \lambda_2^k), \\ \lambda_1^{k+1} &\leftarrow \lambda_1^k - \beta_1 (z^{k+1} - x^{k+1}), \\ \lambda_2^{k+1} &\leftarrow \lambda_2^k - \beta_2 (Ax^{k+1} - b). \end{aligned} \quad (10)$$

The x -subproblem of this iterative framework, namely, the minimization of $\mathcal{L}_{\mathcal{A}}(x, z^{k+1}, \lambda_1^k, \lambda_2^k)$ with respect to x , is

$$\min_x (\lambda_1^k)^T x + \frac{\beta_1}{2} \|z^{k+1} - x\|_2^2 - (\lambda_2^k)^T Ax + \frac{\beta_2}{2} \|Ax - b\|_2^2, \quad (11)$$

and it can be transformed as an equivalent convex quadratic problem:

$$\begin{aligned} \min_x & \frac{1}{2} x^T (\beta_1 I + \beta_2 A^T A) x \\ & - (\beta_1 z^{k+1} - \lambda_1^k + \beta_2 A^T b + A^T \lambda_2^k)^T x. \end{aligned} \quad (12)$$

Note that it can be reduced to solving the linear system based on the optimality condition as follows:

$$(\beta_1 I + \beta_2 A^T A) x = \beta_1 z^{k+1} - \lambda_1^k + \beta_2 A^T b + A^T \lambda_2^k. \quad (13)$$

Similarly, minimizing with respect to z in the iterative framework has the following formula:

$$\min_x \|z\|_{w,2,1} - (\lambda_1^k)^T z + \frac{\beta_1}{2} \|z - x^k\|_2^2. \quad (14)$$

Set $k = 0$. Initialize $x^0 \in \mathbb{R}^n, \lambda_1^0 \in \mathbb{R}^n, \lambda_2^0 \in \mathbb{R}^m$, and $\beta_1, \beta_2 > 0, \gamma_1 > 0, \gamma_2 > 0$.

While stopping criterion is not met **Do**

(1) Update z :

 Compute z^{k+1} according to (14) for given $(x^k, \lambda_1^k, \lambda_2^k)$.

(2) Update x :

 Compute x^{k+1} by solving (13).

(3) Update λ_1 and λ_2 :

 Update λ_1^{k+1} and λ_2^{k+1} via (17).

(4) $k = k + 1$.

end while

ALGORITHM 1: Primal-based ADM for group sparsity [9].

After simple manipulations, it is easy to see that (14) is equivalent to

$$\min_z \sum_{i=1}^s \left\{ w_i \|z_{g_i}\|_2 + \frac{\beta_1}{2} \|z_{g_i} - x_{g_i}^k - \frac{1}{\beta_1} (\lambda_1^k)_{g_i}\|_2^2 \right\}. \quad (15)$$

Notice that the solution of (15) has a closed form according to the one-dimensional shrinkage (or referred to as soft-thresholding) formula [34]:

$$z_{g_i} = \max \left\{ \|s_i\|_2 - \frac{w_i}{\beta_1}, 0 \right\} \frac{s_i}{\|s_i\|_2}, \quad (16)$$

where $s_i = x_{g_i}^k + (1/\beta_1)(\lambda_1^k)_{g_i}$, and the convention $0 \cdot (0/0) = 0$ is assumed. For shortness, we denote the above group-wise shrinkage operation as $z = \text{shrink}(x^k + (\lambda_1^k/\beta_1), w/\beta_1)$.

Finally, we update the multipliers $(\lambda_1^{k+1}, \lambda_2^{k+1})$. Note that step lengths can be incorporated to the update of $(\lambda_1^{k+1}, \lambda_2^{k+1})$; that is,

$$\begin{aligned} \lambda_1^{k+1} &\leftarrow \gamma_1 \lambda_1^k - \beta_1 (z^{k+1} - x^{k+1}), \\ \lambda_2^{k+1} &\leftarrow \gamma_2 \lambda_2^k - \beta_2 (Ax^{k+1} - b), \end{aligned} \quad (17)$$

where $\gamma_1, \gamma_2 > 0$ are step lengths. Under certain assumptions, convergence of the ADM framework with step lengths $\gamma_1, \gamma_2 \in (0, (\sqrt{5} + 1)/2)$ was demonstrated in [31, 32] in the context of variational inequality. The above procedure can be summarized in Algorithm 1.

2.3.2. Applying ADM to the Dual Weighted $\ell_{2,1}$ Model. In this part, we briefly review the ADM approach to the dual form of the weighted $\ell_{2,1}$ model and derive an equally simple yet more efficient algorithm. The dual form of (3) is given as follows:

$$\max_y \{b^T y : \|A^T y\|_2 \leq w_i, i = 1, 2, \dots, s\}. \quad (18)$$

Similarly, we introduce an auxiliary variable z and transform (18) as an equivalent constraint optimization problem:

$$\min_{y,z} \{-b^T y : z = A^T y, \|z_{g_i}\|_2 \leq w_i, i = 1, 2, \dots, s\}. \quad (19)$$

The Lagrangian function of (19) is given by

$$\begin{aligned} \mathcal{L}_{\mathcal{A}}(y, z, x) &\triangleq -b^T y - x^T (z - A^T y) + \frac{\beta}{2} \|z - A^T y\|_2^2 \\ \text{s.t. } \|z_{g_i}\|_2 &\leq w_i, \quad i = 1, 2, \dots, s, \end{aligned} \quad (20)$$

where $\beta > 0$ is a penalty parameter, and note that $x \in \mathbb{R}^n$ is a multiplier and essentially the primal variable. If we start at $x = x^k$ and $z = z^k$, the iteration framework of the augmented Lagrangian problem is given by

$$\begin{aligned} y^{k+1} &\leftarrow \arg \min_y \mathcal{L}_{\mathcal{A}}(x^k, y, z^k), \\ z^{k+1} &\leftarrow \arg \min_z \mathcal{L}_{\mathcal{A}}(x^k, y^{k+1}, z), \\ x^{k+1} &= x^k - \gamma \beta (z^{k+1} - A^T y^{k+1}). \end{aligned} \quad (21)$$

The y subproblem of this iterative framework

$$\min_y -b^T y + (Ax^k)^T y + \frac{\beta}{2} \|z - A^T y\|_2^2 \quad (22)$$

can be solved as a linear system according to its optimality condition

$$\beta A A^T y = b - Ax^k + \beta A z^k. \quad (23)$$

The z subproblem has the form

$$\begin{aligned} \min_z & -(x^k)^T z + \frac{\beta}{2} \|z - A^T y^{k+1}\|_2^2 \\ \text{s.t. } & \|z_{g_i}\|_2 \leq w_i, \quad i = 1, 2, \dots, s. \end{aligned} \quad (24)$$

It is easy to see that (24) has a closed-form solution:

$$z_{g_i} = \mathcal{P}_{\mathbf{B}_2^i} \left(A^T y^{k+1} + \frac{1}{\beta} x_{g_i}^k \right), \quad i = 1, 2, \dots, s, \quad (25)$$

where $\mathcal{P}$ represents a projection (in Euclidean norm) onto a convex set denoted by a subscript and $\mathbf{B}_2^i \triangleq \{z_{g_i} \in \mathbb{R}_{g_i}^n : \|z_{g_i}\|_2 \leq w_i\}$. Finally, we update the multiplier x , that is, the primal variable essentially:

$$x^{k+1} = x^k - \gamma \beta (z^{k+1} - A^T y^{k+1}), \quad (26)$$

where $\gamma \in (0, (\sqrt{5} + 1)/2)$ is a step length.

Now, we rewrite the ADM iteration scheme for (18) as follows.

2.4. Step 2: Adaptive Weights Determined Based on Iterative Support Detection. In this section, we will present the way to determine the weights $\mathbf{w}$ in Step 2, extending the idea

of iterative support detection (ISD) [10] from single sparse vector cases to group sparsity cases.

The support detection based on thresholding in terms of group sparsity is as follows:

$$I^{(L+1)} = \{i : \|x_{g_i}^{(L)}\|_2^2 > \epsilon^{(L)}\}, \quad (27)$$

where $L = 1, 2, \dots$ denotes the stage number. The elements of the weighting vector $\mathbf{w}^{(L+1)}$ are equal to 0 if the corresponding positions belong to the support detection $I^{(L+1)}$, or 1 otherwise. Before discussing the choice of $\epsilon^{(L)}$, it should be pointed out that support detection sets $I^{(L)}$ are not necessarily increased and nested; that is, $I^{(L)} \subset I^{(L+1)}$ may not hold for all L , because the support detection we get from the current solution may contain wrong detections by $\epsilon^{(L)}$ thresholding, and not requiring $I^{(L)}$ to be monotonic leaves the chance for support detection to remove previous wrong detections. This makes $I^{(L)}$ less sensitive to $\epsilon^{(L)}$, thus making the threshold value $\epsilon^{(L)}$ easier to choose. The tuning parameter $\epsilon^{(L)}$ is a key parameter, and it is not a fixed value but is preferred to decrease from a large value to a small value, which can extract more correct nonzero information from the intermediate reconstruction results as the ISD iteration proceeds. In addition, we have proved that ISD can tolerate certain ratio of wrong support detections and still achieve a better reconstruction in [10].

We set the threshold value as follows:

$$\epsilon^{(L)} := \frac{\max \{\|x_{g_1}^{(L)}\|_2^2, \|x_{g_2}^{(L)}\|_2^2, \dots, \|x_{g_s}^{(L)}\|_2^2\}}{\rho^{(L+1)}}, \quad (28)$$

with $\rho > 0$. An excessively large ρ will result in too many false detections and lead to low solution quality, while an excessively small ρ tends to need a large number of iterations. This rule will be quite effective with an appropriate ρ , and the proper range of ρ is case-dependent [35]. Empirically, the performance of our algorithm is not very sensitive to the choice of ρ in our cases.

Here we would like to point out the difference of the situations between this paper and our pervious work in [10]. While we still use the *threshold-ISD* [10] in our cases, the fast-decaying property of the unknown signal is not required any more, because the prior information of grouping improves the performance of threshold based support detection. Some simple intuitive explanation is given here. In [10], the components of a sparse signal are considered separately while the known grouping information of the sparse signal is considered in this paper. Therefore, when performing the thresholding via (27), the prior grouping information reduces the freedom of the unknown signal and provides us with better robustness to the recovery errors of the intermediate results than the plain sparse recovery, which is either the arbitrary grouping or only component-wise.

2.5. Our Algorithm Framework and Some Further Analysis. Now, we summarize the algorithm framework of the multi-stage convex relaxation for the weighted $\ell_{2,1}$ model based on

```

Set  $k = 0$ . Initialize  $x^0 \in \mathbb{R}^n$ ,  $z^0 \in \mathbb{R}^n$ ,  $\beta > 0$  and  $\gamma > 0$ .
While stopping criterion is not met Do
(1) Update  $y$ :
    Compute  $y^{k+1}$  according to (23) for given  $(x^k, z^k)$ .
(2) Update  $z$ :
    Compute  $z^{k+1}$  by solving (25).
(3) Update  $x$ :
    Update  $x^{k+1}$  via (26).
(4)  $k = k + 1$ .
end while

```

ALGORITHM 2: Dual-based ADM for group sparsity [9].

```

Given observation  $b$  and the measurement matrix  $A$ .
(1) Set the iteration number  $L \leftarrow 0$  and initialize the set of
detected entries  $I^{(s)} \leftarrow \emptyset$ .
(2) while the stopping condition is not met Do
(a)  $w_i^{(L)} \leftarrow 0$  if  $i \in I^{(L)}$ ; otherwise,  $w_i^{(L)} \leftarrow 1$ .
(b)  $x^{(L)} \leftarrow$  solving weighted  $\ell_{2,1}$  model (3) for  $w = w^{(L)}$ .
(c)  $I^{(L+1)} \leftarrow$  support detection using  $x^{(L)}$  as the reference.
(d)  $L \leftarrow L + 1$ .
end while

```

ALGORITHM 3: The GM-ADM algorithm.

ISD. The algorithm repeatedly performs the two steps mentioned above: support detection to determine the 0-1 weights and solving the resulted weighted $\ell_{2,1}$ utilizing the ADM algorithms. Moreover, we present some new viewpoints of ISD.

Notice that since $w^{(0)} = \vec{1}$, the weighted model (3) in Step 2(b) is nothing but the plain $\ell_{1,2,1}$ model in iteration 0. The weighted $\ell_{2,1}$ model in 2(b) of Algorithm 3 can be solved by Algorithm 1 or Algorithm 2. The support detection in 2(c) has been introduced in part C of this section. In each iteration, ISD estimates the indices of the nonzero groups using thresholding. Since ISD belongs to the greedy methods and it is also a multistage procedure, the new method is denoted by GM-ADM. Each iteration GM-ADM needs to solve a weighted $\ell_{2,1}$ problem and hence GM-ADM is computationally more demanding, compared to ADM. However, its costing time is not necessarily several times of that of original $\ell_{1,2,1}$ model. The reason is due to the adopted warm-starting; that is, the output of the current stage (outside iteration) is employed as the input of the next stage, and then in the next stage, we often just need to run a few inner iterations to obtain a better updated solution. In addition, usually the number of the stages is not necessarily large (no more than 9 empirically). Notice that, for our problems, we are mainly focusing on the reconstruction quality and GM-ADMs are performing much better than ADMs in this aspect, and it is worth the extra computing cost.

ISD can be also considered as a procedure based on mixed soft-thresholding and hard-thresholding. Specifically, The processing of obtaining the solution z of the subproblem (15) is actually a soft-thresholding procedure, and we will

obtain a group sparse solution via this group-wise shrinkage operation if the weights are all equal to 1. However, this kind of shrinkage has a fatal disadvantage; that is, it shrinks the components of the true nonzero groups as well and reduces the groups' sharpness of the solution. ISD aims not to use the uniform weights in the weighted $\ell_{2,1}$ ($\ell_{w,2,1}$) model but uses the greedy 0-1 weights. It is easy to see that some components of certain groups will not be shrunk if we believe they are unlikely to be zero groups. In such cases, the weights of these certain groups are set as 0, which corresponds to the hard-thresholding. Thus the process of solving z becomes a selective shrinkage procedure, that is, mixed soft-thresholding and hard-thresholding when ISD is applied.

3. Numerical Experiments

In this section, we show numerical results to evaluate the performance of our proposed GM-ADM approach in comparison with the ADM approach [9] in the case of group sparsity because [9] has made comprehensive comparisons with some other group sparsity methods, and their proposed algorithms are outperforming the previously known state-of-the-art algorithms mostly. The code of ADM (referred to as YALL1-group package) can be downloaded in the website [36]. All the experiments were performed under Windows 7 and MATLAB v7.10.0 (R2010a) running on a desktop with an Intel(R) pentium(R) CPU G640 (2.80 GHz) and 2 GB of memory.

3.1. Synthetic Nonoverlapping Group Sparsity Experiment. We generate the nonoverlapping group sparse solutions as follows: we first randomly divide an n -vector x into s groups and then randomly pick k of them as active (nonzero) groups, whose entries are iid random Gaussian or ± 1 (Bernoulli signals; instead of randomly dividing n -vector x into s groups, we fix the number of components in each group to be n/s and randomly pick k out of s groups as active groups). The purpose of testing the sparse Bernoulli signal is to demonstrate that in, the cases of grouping sparsity, the fast decaying property of the magnitudes of the nonzero components is not required any longer, even in terms of the magnitudes of groups of variables. We use the standard iid Gaussian sensing matrices A generated by $A = \text{randn}(m,n)$ in MATLAB. We add the Gaussian noise into the observation by $\text{noise} = \text{randn}(m,1)$, $b = b + \text{sigma} * \text{norm}(b) / \text{norm}(\text{noise}) * \text{noise}$ in MATLAB. The test sets are summarized in Table 1. It needs to be pointed out that the YALL1-group package not only includes solvers for the constrained model (3) but also includes 6 different approaches for solving the corresponding nonoverlapping group sparse problems, as summarized in Table 2. The continuation technique takes the rules as follows: for the primal-based ADM, that is, PADM3, if

$$\frac{\|z^{k+1} - x^{k+1}\|_2}{\|z^k - x^k\|_2} > \tau, \quad (29)$$

FIGURE 1: The sorted groups' magnitude of our tested nonoverlapping group sparse Gaussian signal has fast decaying property.

TABLE 1: Summary of test sets.

Test number #	Noise level σ	Dimension n	Groups number s	Active number k	Measurements m	Repetitions
1	0	1024	126	12	180 : 20 : 280	100
2	$5 \times 10^{-3}, 5 \times 10^{-2}$	1024	126	12	180 : 20 : 280	100

TABLE 2: The 6 different primal-based and dual-based ADM approaches.

Primal-based or dual-based ADM	If exactly solving the linear system	If adopting continuation technique	Algorithm name
Primal-based	Yes	No	PADM1
	No	No	PADM2
	No	yes	PADM3
Drimal-based	Yes	No	DADM1
	No	No	DADM2
	No	Yes	DADM3

TABLE 3: The optimal parameter setting of 6 different primal-based and dual-based ADM approaches that achieve the best reconstruction quality.

Algorithm name	Parameter setting
PADM1, PADM2	$\beta_1 = \beta_2 = 1/\text{mean}(\text{abs}(b))$
PADM3	$\beta_1 = \beta_2 = 0.01/\text{mean}(\text{abs}(b))$, $\tau = 0.9, \eta = 1.2$
DADM1, DADM2	$\beta = 1/\text{mean}(\text{abs}(b))$
DADM3	$\beta = 0.5/\text{mean}(\text{abs}(b))$, $\tau = 0.9$, $\eta = 1.2$

we update $\beta_1 = \eta \times \beta_1, \beta_2 = \eta \times \beta_2$. Similarly, for the dual-based ADM, that is, DADM3, if

$$\frac{\|z^{k+1} - A^T y^{k+1}\|_2}{\|z^k - A^T y^k\|_2} > \tau, \quad (30)$$

we update $\beta = \eta \times \beta$. The ADMs are terminated when one of the following situations are met:

$$\frac{\|x^{k+1} - x^k\|_2}{\|x^k\|_2} < \text{tol}. \quad (31)$$

FIGURE 2: Comparison results of different primal-based ADMs, dual-based ADMs, and the corresponding GM-ADMs (nonoverlapping group sparse Gaussian cases). Noise level $\sigma = 0$. The x-axe represents number of measurements, and the y-axe represents relative error, respectively. The maximum number of stages of the top row: $L_{\max} = 3$; middle row: $L_{\max} = 5$; bottom row: $L_{\max} = 8$, respectively.

FIGURE 3: Comparison results of different primal-based ADMs, dual-based ADMs, and the corresponding GM-ADMs (nonoverlapping group sparse Gaussian cases). Top row: noise level $\sigma = 5 \times 10^{-3}$. Bottom row: noise level $\sigma = 5 \times 10^{-2}$. The x-axis represents number of measurements, and the y-axis represents relative error, respectively. The maximum number of stages $L_{\max} = 3$.

That is, the relative change of two consecutive iterations becomes smaller than the tolerance. Or the iteration number reaches the prescribed maximal number, for example, 1000. We use the same inner iteration terminal condition for GM-ADMs, that is, the terminal condition of 2(b) in Algorithm 3. We set $tol = \max(0.1 * \sigma, 1e - 16)$ in all of the numerical experiments.

We set the optimal parameters for the 6 different ADMs in Table 3, which can achieve the best reconstruction quality in terms of relative error. These parameter values are mostly borrowed from [9], and the parameters for our proposed 6 different GM-ADMs are set the same as the corresponding ADMs and one can refer to [9] for some guidance. Empirically, these parameters are not very sensitive, and we believe the comparisons are fair under this parameter setting. Here

we use the MATLAB-type notation $mean(abs(b))$ to denote the arithmetic average of the absolute value b . For the primal-based ADMs and dual-based ADMs, we set $\gamma_1 = \gamma_2 = 1.618$ and $\gamma = 1.618$, respectively. With regard to our proposed GM-ADM, considering that the 2(b) in Algorithm 3 can be solved by the 6 different above-mentioned ADMs, we call them GM-PADM1, GM-PADM2, GM-PADM3, GM-DADM1, GM-DADM2, and GM-DADM3, respectively. The parameters for our proposed 6 different GM-ADMs are set same as the corresponding ADMs. In addition, we set $\rho = 5$ in (28), and empirically the parameter ρ is not very sensitive, and thus we choose to fix it in our experiments.

As mentioned before, in the cases of grouping sparsity, the fast decaying property of the magnitudes of the nonzero components is not required any longer for the effectiveness

FIGURE 4: Visual comparison of the performance between PADM3 and GM-PADM3, DADM3 and GM-DADM3 (nonoverlapping group sparse Gaussian cases). Noise level $\sigma = 0$. The maximum number of stages $L_{\max} = 3$.

of ISD in the case of group sparsity, though fast decaying property might be able to further enhance the performance of ISD. Our experiments will include the results of both fast decaying signals and non-fast decaying signals.

Figure 1 shows that the sorted groups' magnitude ($\{\|x_{g_1}^{(L)}\|_2^2, \|x_{g_2}^{(L)}\|_2^2, \dots, \|x_{g_s}^{(L)}\|_2^2\}$) of our test nonoverlapping group sparse Gaussian signal has fast decaying property, and this is good for our *threshold*-ISD, though not necessarily. In Figure 2, we present the relative error between recovered signal and original true signal of the test 1, and we can see that our algorithm GM-ADMs brings significant enhancements compared with the corresponding ADMs, for either primal-based ADMs or dual-based ADMs. From Figure 2, we can also see that GM-ADMs can obtain satisfactory promotion compared with ADMs, even if the maximum stage number

$L_{\max}$ is not very large. The key factor to our algorithm is that ISD requires reliable true support detection from inexact reconstruction, if the output of the first iteration (the result of ADMs) is rather unsatisfactory, for example, suffering from insufficient measurement number and/or considerable quantity of noise: it also becomes a hard work for GM-ADMs to achieve very huge promotion because of the inexact support detection. Thus the improvement by ISD when the measurements number $m = 180$ is less obvious than other cases. In Figure 3, we show the comparison results between ADMs and GM-ADMs in noise environments, and ISD still brings better results.

In order to better illustrate the practical advantage of our algorithm, we also give visual comparisons of the reconstruction. Figures 4, 5, and 6 plot the reconstructed signals in

FIGURE 5: Visual comparison of the performance between PADM3 and GM-PADM3, DADM3 and GM-DADM3 (nonoverlapping group sparse Gaussian cases). Noise level $\sigma = 5 \times 10^{-3}$. The maximum number of stages $L_{\max} = 3$.

the case of $m = 200$, $L_{\max} = 3$, and the noise levels are $\sigma = 0$, $\sigma = 5 \times 10^{-3}$, $\sigma = 5 \times 10^{-2}$, respectively. For paper conciseness, here we just give the comparisons between PADM3 and GM-PADM3, DADM3 and GM-DADM3, since the other cases have the similar conclusions. From Figures 4, 5, and 6, we can see that the results of GM-PADM3 and GM-DADM3 are much better than those of the correspondent PADM3 and DADM3, in either noiseless environment or noise environment.

In Figure 7, we show the sorted groups' magnitude ($\{\|x_{g_1}^{(L)}\|_2^2, \|x_{g_2}^{(L)}\|_2^2, \dots, \|x_{g_s}^{(L)}\|_2^2\}$) of nonoverlapping group sparse Bernoulli signal whose nonzero components are generated by randomly either 1 or -1 and do not have fast decaying property. In Figure 8, we present the comparison

results between ADMs and corresponding GM-ADMs, and we can see that GM-ADMs can achieve considerable promotion either in noiseless environment or in noise environment. We also give the visual comparison between PADM3 and GM-ADM3 in Figure 9. From Figures 8 and 9, for nonoverlapping group sparse Bernoulli signals, we can observe the similar conclusions as nonoverlapping group sparse Gaussian signals.

3.2. Synthetic Overlapping Group Sparsity Experiment. To assess the performance of our algorithm when overlapping groups are given a priori, we generate the simulation data with $n = 1024$ variables, covered by 126 groups of 10 variables with 2 variables of overlap between two consecutive groups:

FIGURE 6: Visual comparison of the performance between PADM3 and GM-PADM3, DADM3 and GM-DADM3 (nonoverlapping group sparse Gaussian cases). Noise level $\sigma = 5 \times 10^{-2}$. The maximum number of stages $L_{\max} = 3$.

$\{1, 2, \dots, 10\}, \{9, 10, \dots, 18\}, \dots, \{1015, \dots, 1024\}$. Then we randomly pick k of them as active (nonzero) groups, whose entries are either iid random Gaussian or ± 1 , while the remaining groups are all zeros. We use the same test sets as shown in Table 1, and the only difference is that the signal has overlapping group sparse structure here. In addition, the optimal parameters and terminal condition in Part A are still applicable here. Considering that the YALL1-group package only includes primal-based ADMs for overlapping group sparse problems, here we just compare the results between PADM3 and GM-PADM3, for fair comparison and paper conciseness. Empirically, $L_{\max} = 2$ can already achieve satisfactory reconstruction quality in terms of relative error, and thereby we set $L_{\max} = 2$ for all of the experiments.

In Figure 10, we show the sorted groups' magnitude $(\{\|x_{g_1}^{(L)}\|_2^2, \|x_{g_2}^{(L)}\|_2^2, \dots, \|x_{g_s}^{(L)}\|_2^2\})$ of our tested overlapping group sparse signals whose nonzero components are generated by Gaussian distribution also has fast decaying property. In Figure 11, we present the comparison between PADM3 and GM-PADM3, and we can see that GM-ADM3 achieve considerable promotion in either noiseless environment or noisy environment.

In order to better illustrate the practical advantage of our algorithm for overlapping group sparse reconstruction, we also give visual comparisons in Figure 12, and we can see that the accuracy rate of the elements' value of PADM3 is fairly low, while GM-ADM3 can achieve a rather high accuracy rate.

FIGURE 7: The sorted groups' magnitude of our tested nonoverlapping group sparse Bernoulli signal does not have fast decaying property.

In Figure 13, we show that the sorted groups' magnitude $(\{\|x_{g_1}^{(L)}\|_2^2, \|x_{g_2}^{(L)}\|_2^2, \dots, \|x_{g_s}^{(L)}\|_2^2\})$ of overlapping group sparse signals whose nonzeros are generated by randomly either 1 or -1 does not have fast decaying property. In Figure 14, we present the comparison between PADM3 and GM-PADM3, where GM-ADM3 achieves considerable promotion no matter in noiseless environment or noise environment. We also give the visual comparison between PADM3 and GM-ADM3 in Figure 15. From Figures 14 and 15, for overlapping group sparse Bernoulli signals, we can observe the similar conclusions as overlapping group sparse Gaussian signals.

3.3. A Nonoverlapping Group Sparsity Simulation Example from Collaborative Spectrum Sensing. In this part, we study an interesting special case of the nonoverlapping group sparsity structure called joint sparsity; namely, a set of sparse solutions share a common nonzero support. This example that comes from collaborative spectrum sensing [37], which aims at detecting spectrum holes (i.e., channels not used by any primal users), is the precondition for the implementation of Cognitive Radio (CR). The Cognitive Radio (CR) nodes must constantly sense the spectrum in order to detect the presence of the Primary Radio (PR) nodes and use the spectrum holes without causing harmful interference to the PRs. Hence, sensing the spectrum in a reliable manner is of critical importance and constitutes a major challenge in CR networks. Collaborative spectrum sensing is expected to improve the ability of checking complete spectrum usage. We consider a cognitive radio network with m CR nodes that locally monitor a subset of n channels. A channel is either occupied by a PR or unoccupied, corresponding to the states 1 and 0, respectively. It is assumed that the number s of occupied channels is much smaller than n . The goal is to recover the occupied channels from the CR nodes' observations. Via frequency-selective filters, a CR takes a

small number of measurements that are linear combinations of multiple channels. In order to mix the different channel sensing information, the filter coefficients are designed to be random numbers. Then, the filter outputs are sent to the fusion center. Assume that there are p frequency selective filters in each CR node sending out p reports regarding the n channels. The sensing process at each CR can be represented by $p \times n$ filter coefficients matrix A . Let an $n \times n$ diagonal matrix R represent the states of all the channel sources using 0 and 1 as diagonal entries, indicating the unoccupied or occupied states, respectively. There are s nonzero entries in the diagonal matrix R . In addition, channel gains between the CRs and channels are described in an $m \times n$ channel gain matrix G given by [38]. Then, the measurements reports sent to the fusion center can be written as a $p \times m$ matrix as follows:

$$B_{p \times m} = A_{p \times n} R_{n \times n} (G_{m \times n})^T. \quad (32)$$

Now, we need a highly effective method for recovering

$$X_{n \times m} = R_{n \times n} \times (G_{m \times n})^T. \quad (33)$$

In X , each column (denoted by $X_{:,j}$) corresponds to the channel occupancy status received by CR j , and each row (denoted by $X_{i,:}$) corresponds to the status of channel i . Since there are only a small number of channels which are used, X is sparse in terms of the number of rows containing nonzero. In each nonzero row $X_{i,:}$, if $X_{i,j} \neq 0$, other entries in the same row are likely nonzeros. Therefore, X is a joint sparse matrix.

The weighted $l_{2,1}$ model of the joint sparsity problem is

$$\min_X \|X\|_{w,2,1} = \sum_{i=1}^n w_i \|x^i\|_2, \quad \text{s.t. } AX = B, \quad (34)$$

FIGURE 8: Comparison results of different primal-based ADMs, dual-based ADMs, and the corresponding GM-ADMs (nonoverlapping group sparse Bernoulli cases). The x -axis represents number of measurements, and the y -axis represents relative error, respectively. The maximum number of stages: $L_{\max} = 3$. Noise level of top row: $\sigma = 0$. Middle row: $\sigma = 5 \times 10^{-3}$. Bottom row: $\sigma = 5 \times 10^{-2}$.

FIGURE 9: Visual comparison of the performance between PADM3 and GM-ADM3 (nonoverlapping group sparse Bernoulli cases). Noise level from top row to bottom row: $\sigma = 0$, $\sigma = 5 \times 10^{-3}$, and $\sigma = 5 \times 10^{-2}$, respectively. The maximum number of stages $L_{\max} = 3$.

FIGURE 10: The sorted groups' magnitude of our tested overlapping group sparse Gaussian signal has fast decaying property.

where $X = [x_1, x_2, \dots, x_m] \in \mathbb{R}^{n \times m}$ denotes the collection of m jointly sparse solutions, and x^i and x_j denote the i th row and j th column of X , respectively.

Let us define

$$\tilde{A} \triangleq I_m \otimes A = \begin{pmatrix} A & & \\ & A & \\ & & \ddots \\ & & & A \end{pmatrix}, \quad (35)$$

$$\tilde{x} \triangleq \text{vec}(X) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix}, \quad \tilde{b} \triangleq \text{vec}(B) = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix}, \quad (36)$$

where $I_m \in \mathbb{R}^{m \times m}$ is the identity matrix, and $\text{vec}(\cdot)$ and $\otimes$ are standard notations for the vectorization of a matrix and the Kronecker product, respectively. We partition $\tilde{x}$ into n groups $\{x_{g_1}, x_{g_2}, \dots, x_{g_n}\}$, where $x_{g_i} \in \mathbb{R}^m$ ($i = 1, 2, \dots, n$) corresponds to the i th row of matrix X . Thus, we can obtain an equivalent group $l_{w,2,1}$ problem to (34) as follows:

$$\min_{\tilde{x}} \|\tilde{x}\|_{w,2,1}, \quad \text{s.t. } \tilde{A}\tilde{x} = \tilde{b}. \quad (37)$$

The main advantage of joint sparsity reconstruction is its applications to large scale problems. Therefore, the following simulations is carried out for a relatively large dimensional applications with the following settings: we consider a 16-node cognitive radio network (i.e., $m = 16$), the number of channels is 1024 (i.e., $n = 1024$), the number of active PR nodes ranges from 100 to 120 on the given set of 1024 channels, the measurement matrix A is Gaussian random matrix, and the size is fixed as $p = 256$.

Here, we adopt the support detection strategy originally used in our previous work [10]. In order to better illustrate

this strategy, we introduce a vector gx , and the elements of gx are the magnitudes of these groups (i.e., $\|x_{g_i}\|_2, i = 1, 2, \dots, n$). The rule of our choice is based on locating the “first significant jump in the increasingly sorted sequence $gx^{(L)}$ ($gx_{[t]}^{(L)}$ denotes the t th largest component of gx by magnitude), and the rule looks for the smallest t such that

$$gx_{[t+1]}^{(L+1)} - gx_{[t]}^{(L)} > \tau^{(L)}. \quad (38)$$

Then, we set $\epsilon^L = gx_{[t]}^{(L)}$ in (27).

For paper conciseness, here we just give the comparison between the ADM1 algorithm and GM-ADM1 algorithm, to demonstrate the superiority of our method. Namely, the multistage $l_{w,2,1}$ process can achieve impressive improvement compared to the original single stage $l_{2,1}$ process. The parameters and tolerance of ADM1 method are the same as the nonoverlapping group sparse experiment. In addition, for GM-ADM1 algorithm, empirically, we fix the maximum stage number $L_{\max} = 3$, $\tau^{(L)} = 0.01 * \text{mean}(\text{diff}(\text{sort}(gx^{(L)})))$ in MATLAB.

In Figure 16, we present the comparison results between the ADM1 algorithm and the corresponding GM-ADM1 algorithm, in terms of the relative error, both in noiseless environment and noisy environment (noise level $\sigma = 5 \times 10^{-3}$). We can draw similar conclusions as those in parts A and B. Moreover, we also give the ideal results of our algorithm; that is, the support detection is based on the underlying true solution. While we usually do not know the true solution in practice, here we just use it as a reference and name it as ideal GM-PADM1 method (shortened as IGM-PADM1), as an ideal golden upper bound of the performance of iterative support detection based multistage methods. It fully demonstrates the superiority of our new idea; the iterative support detection based multistage process can bring significantly enhancement compared to the standard single stage process. In addition, it makes us believe that the GM-ADMs

FIGURE 11: Comparison of the performance between PADM3 and GM-PADM3 (overlapping group sparse Gaussian cases). Noise level from left to right: $\sigma = 0$, $\sigma = 5 \times 10^{-3}$, and $\sigma = 5 \times 10^{-2}$, respectively. The maximum number of stages $L_{\max} = 2$.

can bring dramatically better reconstructions compared to corresponding ADMs, as long as we can acquire enough reliable support detections. Even though the ideal case is not possible in practice, however, it serves as a benchmark and chalks out a path for us to explore. In addition, to make a further comparison between the ADM1 algorithm and GM-ADM1 algorithm, in Figure 17, we show the recoverability results of the two algorithms. It is worth emphasizing that we believe that the reconstruction is successful if the relative error is below the given threshold. Here, we just consider the noiseless case, and we set the threshold $\eta = 1 \times 10^{-6}$ and $\eta = 1 \times 10^{-4}$, respectively. Not surprisingly, the recoverability of the GM-ADM1 is better than ADM1.

4. Conclusions and Possible Future Work

In this paper, we propose a novel GM-ADM approach for group sparse reconstruction problems. The final result is obtained from multistage process consisting of solving a series of weighted $\ell_{2,1}$ model and determining the adaptive weights via ISD. The numerical results demonstrate the extendability of iterative support detection from sparsity to group sparsity.

Our previous work [10] has demonstrated the dependence of *threshold*-ISD on the fast decaying property in the case of the plain sparse signal recovery for effectiveness. However, in this paper, the *threshold*-ISD can still

FIGURE 12: Visual comparison of the performance between PADM3 and GM-ADM3 (overlapping group sparse Gaussian cases). Noise level from top row to bottom row: $\sigma = 0$, $\sigma = 5 \times 10^{-3}$, and $\sigma = 5 \times 10^{-2}$, respectively. The maximum number of stages $L_{\max} = 2$.

FIGURE 13: The sorted groups' magnitude of our tested overlapping group sparse Bernoulli signal does not have fast decaying property.

FIGURE 14: Comparison of the performance between PADM3 and GM-PADM3 (overlapping group sparse Bernoulli cases). Noise level from left to right: $\sigma = 0$, $\sigma = 5 \times 10^{-3}$, and $\sigma = 5 \times 10^{-2}$, respectively. The maximum number of stages $L_{\max} = 2$.

FIGURE 15: Visual comparison of the performance between PADM3 and GM-ADM3 (overlapping group sparse Bernoulli cases). Noise level from top row to bottom row: $\sigma = 0$, $\sigma = 5 \times 10^{-3}$, and $\sigma = 5 \times 10^{-2}$, respectively. The maximum number of stages $L_{\max} = 2$.

FIGURE 16: Comparison results of ADM1, GM-ADM1, and IGM-ADM1 on joint sparsity problem ($n = 1024$, $p = 256$, and $m = 16$) from Cognitive Radio. The x -axes represent the number of nonzero rows, and the y -axes represent the relative error, respectively. The first column corresponds to the noiseless data and the second corresponds to the data with Gaussian noise (noise level $\sigma = 5 \times 10^{-3}$). The results are average of 100 runs.

FIGURE 17: Comparison results of ADM1, and GM-ADM1 on the joint sparsity problem ($n = 1024$, $p = 256$, and $m = 16$) from Cognitive Radio. The x -axes represent the number of nonzero rows, and the y -axes represent the recoverability, respectively. The success thresholds of the first column and the second column are $\eta = 1 \times 10^{-6}$ and $\eta = 1 \times 10^{-4}$, respectively. The results are average of 100 runs.

achieve impressive performance without the dependence on the fast decaying property, in the presence of the prior grouping information. Considering that support detection is not limited to thresholding and reliable support detection guarantees better performance, we would like to study other specific signals, for example, structure sparsity signals in the future, and design more effective support

detection methods based on the particular property of these signals.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by the Natural Science Foundation of China, Grant nos. 11201054 and 91330201, and by the Fundamental Research Funds for the Central Universities ZYGX2012J118, ZYGX2013Z005.

References

- [1] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, vol. 68, no. 1, pp. 49–67, 2006.
- [2] F. R. Bach, "Consistency of the group lasso and multiple kernel learning," *Journal of Machine Learning Research*, vol. 9, pp. 1179–1225, 2008.
- [3] S. Ma, X. Song, and J. Huang, "Supervised group Lasso with applications to microarray data analysis," *BMC Bioinformatics*, vol. 8, no. 1, article 60, 2007.
- [4] D. Eiwen, G. Taubock, F. Hlawatsch, and H. Feichtinger, *Group Sparsity Methods for Compressive Channel Estimation in Doubly Dispersive Multicarrier Systems*, 2010.
- [5] E. van den Berg, M. Schmidt, M. Friedlander, and K. Murphy, "Group sparsity via linear-time projection," Tech. Rep., Department of Computer Science, University of British Columbia, Vancouver, Canada, 2008.
- [6] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, 2009.
- [7] Z. Qin, K. Scheinberg, and D. Goldfarb, "Efficient block-coordinate descent algorithms for the Group Lasso," *Mathematical Programming Computation*, vol. 5, no. 2, pp. 143–169, 2013.
- [8] S. J. Wright, R. D. Nowak, and M. A. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2479–2493, 2009.
- [9] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," in *Wavelets and Sparsity XV*, vol. 8858 of *Proceedings of SPIE*, 2013.
- [10] Y. Wang and W. Yin, "Sparse signal reconstruction via iterative support detection," *SIAM Journal on Imaging Sciences*, vol. 3, no. 3, pp. 462–491, 2010.
- [11] J. Huang, T. Zhang, and D. Metaxas, "Learning with structured sparsity," *The Journal of Machine Learning Research*, vol. 12, pp. 3371–3412, 2011.
- [12] S. Kim and E. P. Xing, "Tree-guided group lasso for multi-task regression with structured sparsity," in *Proceedings of the 27th International Conference on Machine Learning (ICML '10)*, pp. 543–550, June 2010.
- [13] F. Bach, R. Jenatton, J. Mairal, and G. Obozinski, "Structured sparsity through convex optimization," *Statistical Science*, vol. 27, no. 4, pp. 450–468, 2012.
- [14] T. Zhang, "Analysis of multi-stage convex relaxation for sparse regularization," *The Journal of Machine Learning Research*, vol. 11, pp. 1081–1107, 2010.
- [15] N. Vaswani and W. Lu, "Modified-CS: modifying compressive sensing for problems with partially known support," *IEEE Transactions on Signal Processing*, vol. 58, no. 9, pp. 4595–4607, 2010.
- [16] W. Lu and N. Vaswani, "Regularized modified BPDN for noisy sparse reconstruction with partial erroneous support and signal value knowledge," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 182–196, 2012.
- [17] L. Jacques, "A short note on compressed sensing with partially known signal support," *Signal Processing*, vol. 90, no. 12, pp. 3308–3312, 2010.
- [18] M. P. Friedlander, H. Mansour, R. Saab, and O. Yilmaz, "Recovering compressively sampled signals using partial support information," *IEEE Transactions on Information Theory*, vol. 58, no. 2, pp. 1122–1134, 2012.
- [19] R. E. Carrillo, L. F. Polania, and K. E. Barner, "Iterative algorithms for compressed sensing with partially known support," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '10)*, pp. 3654–3657, March 2010.
- [20] R. E. Carrillo, L. F. Polania, and K. E. Barner, "Iterative hard thresholding for compressed sensing with partially known support," in *Proceedings of the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '11)*, pp. 4028–4031, May 2011.
- [21] E. J. Candes, M. B. Wakin, and S. P. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *The Journal of Fourier Analysis and Applications*, vol. 14, no. 5–6, pp. 877–905, 2008.
- [22] R. Chartrand and W. Yin, "Iteratively reweighted algorithms for compressive sensing," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '08)*, pp. 3869–3872, April 2008.
- [23] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Communications on Pure and Applied Mathematics*, vol. 59, no. 8, pp. 1207–1223, 2006.
- [24] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.
- [25] W. Guo and W. Yin, "Edge guided reconstruction for compressive imaging," *SIAM Journal on Imaging Sciences*, vol. 5, no. 3, pp. 809–834, 2012.
- [26] Y. Zhang, *An Alternating Direction Algorithm for Nonnegative Matrix Factorization*, TR10-03, Rice University, 2010.
- [27] Z. Wen, W. Yin, and Y. Zhang, "Solving a low-rank factorization model for matrix completion by a nonlinear successive over-relaxation algorithm," *Mathematical Programming Computation*, vol. 4, no. 4, pp. 333–361, 2012.
- [28] Y. Shen, Z. Wen, and Y. Zhang, "Augmented Lagrangian alternating direction method for matrix separation based on low-rank factorization," *Optimization Methods & Software*, vol. 29, no. 2, pp. 239–263, 2014.
- [29] Y. Xu, W. Yin, Z. Wen, and Y. Zhang, "An alternating direction algorithm for matrix completion with nonnegative factors," *Frontiers of Mathematics in China*, vol. 7, no. 2, pp. 365–384, 2011.
- [30] M. J. D. Powell, "A method for nonlinear constraints in minimization problems," in *Optimization*, R. Fletcher, Ed., pp. 283–298, Academic Press, New York, NY, USA, 1969.
- [31] R. Glowinski and P. Le Tallec, *Augmented Lagrangian and Operatorsplitting Methods in Nonlinear Mechanics*, Society for Industrial Mathematics, 1989.
- [32] R. Glowinski, *Numerical Methods for Nonlinear Variational Problems*, Springer, 2008.
- [33] L. Jacob, G. Obozinski, and J.-P. Vert, "Group lasso with overlap and graph lasso," in *Proceedings of the 26th International Conference On Machine Learning (ICML '09)*, pp. 433–440, ACM, June 2009.
- [34] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Transactions on Information Theory*, vol. 41, no. 3, pp. 613–627, 1995.

- [35] D. H. Wolpert and W. G. Macready, "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, vol. 1, no. 1, pp. 67–82, 1997.
- [36] YALL1-Group: A solver for group/joint sparse reconstruction, <http://www.convexoptimization.com/wikimization/index.php>.
- [37] J. Meng, W. Yin, H. Li, E. Hossain, and Z. Han, "Collaborative spectrum sensing from sparse observations in cognitive radio networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 2, pp. 327–337, 2011.
- [38] T. S. Rappaport, *Wireless Communications: Principles and Practice*, Prentice Hall, 2nd edition, 2002.