DOA Estimation Using a Greedy Block Coordinate Descent Algorithm

Xiaohan Wei, Yabo Yuan, and Qing Ling

***Abstract*—This paper presents a novel jointly sparse signal reconstruction algorithm for the DOA estimation problem, aiming to achieve faster convergence rate and better estimation accuracy compared to existing $\ell_{2,1}$-norm minimization approaches. The proposed greedy block coordinate descent (GBCD) algorithm shares similarity with the standard block coordinate descent method for $\ell_{2,1}$-norm minimization, but adopts a greedy block selection rule which gives preference to sparsity. Although greedy, the proposed algorithm is proved to also have global convergence in this paper. Through theoretical analysis we demonstrate its stability in the sense that all nonzero supports found by the proposed algorithm are the actual ones under certain conditions. Last, we move forward to propose a weighted form of the block selection rule based on the MUSIC prior. The refinement greatly improves the estimation accuracy especially when two point sources are closely spaced. Numerical experiments show that the proposed GBCD algorithm has several notable advantages over the existing DOA estimation methods, such as fast convergence rate, accurate reconstruction, and noise resistance.**

***Index Terms*—Block coordinate descent, DOA estimation, joint sparsity.**

I. Introduction

DIRECTION of arrival (DOA) estimation of multiple narrowband signals plays an important role in array signal processing. The most well-known classical DOA estimation methods include Capon's method [1], MUSIC [2], ESPRIT [3], etc. However, their performances are generally not satisfactory when the point sources are closely spaced, or the product of the SNR and the number of snapshots is small.

In order to address these issues, [4] and [5] introduce the concept of sparse optimization in DOA estimation, given that the number of point sources is limited. Specifically, [4] exploits the beamspace domain sparsity and reconstructs a sparse vector[1] from an ℓ_1-norm regularized least squares problem, which is solved with a quadratic programming (QP) routine. While [5] utilizes the property that the spatial spectra of the point sources over time are jointly sparse[2]. Hence [5] proposes an $\ell_{2,1}$-norm minimization formulation which penalizes the joint sparsity of the spatial spectra; this problem is then solved in a second-order cone programming (SOCP) framework. Recent work along this line includes [6], where a weighted $\ell_{2,1}$-norm minimization formulation replaces the standard $\ell_{2,1}$-norm minimization formulation in [5]. Another notable work is [7], which deals with perturbations and inaccuracies in the basis matrix using the branch-and-bound technique. However, the above approaches are all quite time-consuming especially when the problem dimension is large, e.g., the sparse vector or the jointly sparse vectors are with large size, the number of snapshots is large, etc. To overcome this computational difficulty, [8] proposes a cyclic iterative method and solves a semi-definite programming (SDP) problem by exploiting covariance-based domain sparsity.

This paper adopts the $\ell_{2,1}$-norm minimization formulation for DOA estimation, and develops an efficient algorithm to solve it. The $\ell_{2,1}$-norm minimization problem, also known as group LASSO (least absolute soft shrinkage operator) [9], has attracted much research interest in recent years. Existing algorithms include a spectral projected gradient method (SPGL1) [10], an accelerated gradient method (SLEP) [11], a variant of the matching pursuit algorithm [12], SpaRSA [13], YALL1-Group [14], and block coordinate descent (BCD) algorithms [15]. However, none of them is specifically designed for the DOA estimation problem. In light of this issue, we utilize the prior knowledge as well as the problem structure in DOA estimation, and propose an algorithm with fast convergence rate and excellent recovery accuracy.

Our contributions in this paper are three-fold. First, we develop a novel greedy block coordinate descent (GBCD) algorithm which improves the BCD algorithms in [15] by using a greedy strategy for choosing descent directions. At each iteration, the algorithm updates the block which yields the greatest descent distance, and hence achieves faster convergence rate. Different from most greedy methods without convergence guarantee, global convergence of the proposed algorithm can be proved. Second, we prove that under certain conditions, all nonzero supports found by the proposed GBCD algorithm are the actual ones. Finally, we refine the algorithm by proposing a weighted form of the block selection rule based on the MUSIC prior, aiming to overcome the difficulty in resolving closely spaced point sources. Numerical experiments demonstrate that the proposed GBCD algorithm and its refined

Manuscript received May 15, 2012; revised July 30, 2012 and September 01, 2012; accepted September 03, 2012. Date of publication September 13, 2012; date of current version November 20, 2012. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jean Pierre Delmas. The work of Q. Ling is supported in part by NSFC Grant 61004137 and Fundamental Research Funds for the Central Universities.

X. Wei and Y. Yuan are with the Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei, Anhui 230026, China (e-mail: weixh@mail.ustc.edu.cn; yyb@mail.ustc.edu.cn).

Q. Ling is with the Department of Automation, University of Science and Technology of China, Hefei, Anhui 230026, China (e-mail: qingling@mail.ustc.edu.cn).

Digital Object Identifier 10.1109/TSP.2012.2218812

[1]A vector is sparse when only few elements are nonzero.

[2]Multiple vectors are jointly sparse when all of them are sparse and their nonzero supports are the same.

version are much faster than several popular DOA estimation methods and simultaneously yield quite satisfactory recovery performance.

We start with an introduction on formulating DOA estimation as an $\ell_{2,1}$-norm minimization problem in Section II. We propose the GBCD algorithm in Section III.A and prove its global convergence in Section III.B. Section IV proves the conditions which guarantee that all nonzero supports found by the proposed GBCD algorithm are the actual ones. In Section V, we describe a weighted form of the block selection rule based on the MUSIC prior. Choice of the regularization parameter is discussed in Section VI. Numerical experiments are shown in Section VII. Finally Section VIII concludes the paper.

In our notation, $\mathbf{I}_M$ represents an $M \times M$ identity matrix. The symbol $vec(\cdot)$ denotes the vectorization operator by stacking the columns of a matrix one underneath the other. The operators $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^\dagger$ denote transpose, conjugate transpose and Moore-Penrose inverse, respectively. Also, $\mathcal{C}^M$ represents the space of M-dimensional complex column vectors. For any $\mathbf{x} \in \mathcal{C}^M$ and a nonempty index set $\mathcal{J} \subseteq \mathcal{M} \triangleq \{1, 2, \ldots, M\}$, x_j denotes the j-th component of vector $\mathbf{x}$ and $\mathbf{x}_\mathcal{J}$ denotes the subvector of $\mathbf{x}$ comprising $\{x_j | j \in \mathcal{J}\}$. Moreover, for any matrix $\mathbf{A}$, we use $\mathbf{a}^i$ to represent its i-th row and $\mathbf{a}_i$ to represent its i-th column. $\mathbf{A}_{\mathcal{J}\mathcal{J}} = [\mathbf{A}_{ij}]_{i,j\in\mathcal{J}}$ denotes the principal submatrix of $\mathbf{A}$ indexed by $\mathcal{J}$.

II. Problem Formulation

Consider R narrowband far-field point source signals impinging on an N-element uniform linear array (ULA). For the r-th point source, its signal is $\bar{s}_r(t)$ with t being the time index. The r-th point source comes from direction θ_r with power σ_r^2. Hence, the $N \times 1$ array output vector $\bar{\mathbf{x}}(t)$ is given by

$$\bar{\mathbf{x}}(t) = \sum_{r=1}^{R} \mathbf{a}(\theta_r)\bar{s}_r(t) + \bar{\mathbf{n}}(t), \quad t = t_1, \ldots, t_L \tag{1}$$

where $\mathbf{a}(\theta_r) = [1, e^{-j\pi\cos\theta_r}, \ldots, e^{-j(N-1)\pi\cos\theta_r}]^T$ is the steering vector, and $\bar{\mathbf{n}}(t)$ is the $N \times 1$ noise vector with the power of each entry equal to σ^2. By assumption, the entries of $\{\bar{s}_r(t)\}_{r=1}^R$ and $\bar{\mathbf{n}}(t)$ are zero mean wide-sense stationary random processes, and the entries in $\bar{\mathbf{n}}(t)$ are uncorrelated with each other and the signals $\bar{s}_r(t)$. The above parameter estimation problem (i.e., estimating $\{\theta_r\}_{r=1}^R$) can be solved by many existing approaches, such as Capon's method [1], MUSIC [2], ESPRIT [3], etc. However, their performances are often unsatisfactory when the point sources are closely spaced or the SNR is low.

A popular sparse optimization approach in [5] rewrites (1) as

$$\bar{\mathbf{X}} = \mathbf{B}\bar{\mathbf{S}} + \bar{\mathbf{N}} \tag{2}$$

where $\bar{\mathbf{X}} = [\bar{\mathbf{x}}(t_1), \ldots, \bar{\mathbf{x}}(t_L)]$ and $\bar{\mathbf{N}} = [\bar{\mathbf{n}}(t_1), \ldots, \bar{\mathbf{n}}(t_L)]$. $\mathbf{B} = [\mathbf{b}_1, \ldots, \mathbf{b}_M]$ is an overcomplete basis matrix, in which $\mathbf{b}_i$ is the steering vector corresponding to an angle $\overline{\theta_i}$, and the set $\{\overline{\theta_i}\}_{i=1}^M$ denotes a grid that covers the correct directions. It is obvious that when the sampling grid is fine enough (i.e., $M \gg R$), only a few rows of the signal matrix

$$\bar{\mathbf{S}} = \begin{pmatrix} \bar{s}_1(1) & \cdots & \bar{s}_1(L) \\ \vdots & \ddots & \vdots \\ \bar{s}_M(1) & \ldots & \bar{s}_M(L) \end{pmatrix} \tag{3}$$

are different from zero. The DOA estimation problem is hence to find out which rows of the above signal matrix are nonzero (i.e., to find out the nonzero supports from the jointly sparse columns of $\bar{\mathbf{S}}$), and then to estimate θ_r, $r = 1, \ldots, R$. However, the number of snapshots L is usually very large, which, in most cases, prohibits us from recovering $\bar{\mathbf{S}}$ directly. Thus, according to [5], we can take the singular value decomposition (SVD) of $\bar{\mathbf{X}}$ as

$$\bar{\mathbf{X}} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H. \tag{4}$$

Define $\mathbf{X} = \bar{\mathbf{X}}\mathbf{V}\mathbf{W}$, $\mathbf{S} = \bar{\mathbf{S}}\mathbf{V}\mathbf{W}$, $\mathbf{N} = \bar{\mathbf{N}}\mathbf{V}\mathbf{W}$, where $\mathbf{W} = [\mathbf{I}_R, \mathbf{0}_{R\times(L-R)}]$ and $\mathbf{0}_{R\times(L-R)}$ is an $R \times (L-R)$ zero matrix. We get

$$\mathbf{X} = \mathbf{B}\mathbf{S} + \mathbf{N}. \tag{5}$$

Note that with the SVD, $\mathbf{S}$ is with the same sparsity pattern as $\bar{\mathbf{S}}$, and is hence jointly sparse too. This fact motivates the following $\ell_{2,1}$-norm minimization formulation for jointly sparse signal recovery [5]:

$$\min \ \frac{1}{2}\|\mathbf{X} - \mathbf{B}\mathbf{S}\|_F^2 + \lambda\|\mathbf{S}\|_{2,1} \tag{6}$$

where $\|\cdot\|_F$ denotes the matrix Frobenius norm, $\|\mathbf{S}\|_{2,1} = \sum_{i=1}^M \|\mathbf{s}^i\|_2$ is the $\ell_{2,1}$-norm of $\mathbf{S}$ (recall that $\mathbf{s}^i$ is the i-th row of $\mathbf{S}$), and λ is a regularization parameter.

In [5], (6) is solved with the optimization software package SeDuMi [16], which is designed for general convex optimization problems using the interior point method. Meanwhile, although there are a bunch of $\ell_{2,1}$-norm minimization algorithms, none of them utilizes the prior knowledge of DOA estimation and the specific problem structure (i.e., the structure of the overcomplete basis matrix $\mathbf{B}$). This paper aims at addressing these issues and developing an efficient $\ell_{2,1}$-norm minimization algorithm for DOA estimation.

III. Greedy Block Coordinate Descent (GBCD) Algorithm

This section elaborates on a greedy block coordinate descent (GBCD) algorithm, which exploits the advantages of the block coordinate descent (BCD) strategy, but uses a greedy block selection rule which explicitly gives preference to sparsity. Global convergence of the GBCD algorithm to the optimal solution is also proved in this section.

A. Greedy Block Coordinate Descent

The objective function in (6) can be written as

$$F(\mathbf{S}) = G(\mathbf{S}) + H(\mathbf{S}) \tag{7}$$

where $G(\mathbf{S}) = \frac{1}{2}\|\mathbf{X} - \mathbf{B}\mathbf{S}\|_F^2 = \frac{1}{2}\|vec(\mathbf{X}) - \mathbf{I}_R \otimes \mathbf{B}vec(\mathbf{S})\|_2^2$ with $\otimes$ denoting the Kronecker product, and $H(\mathbf{S}) =$

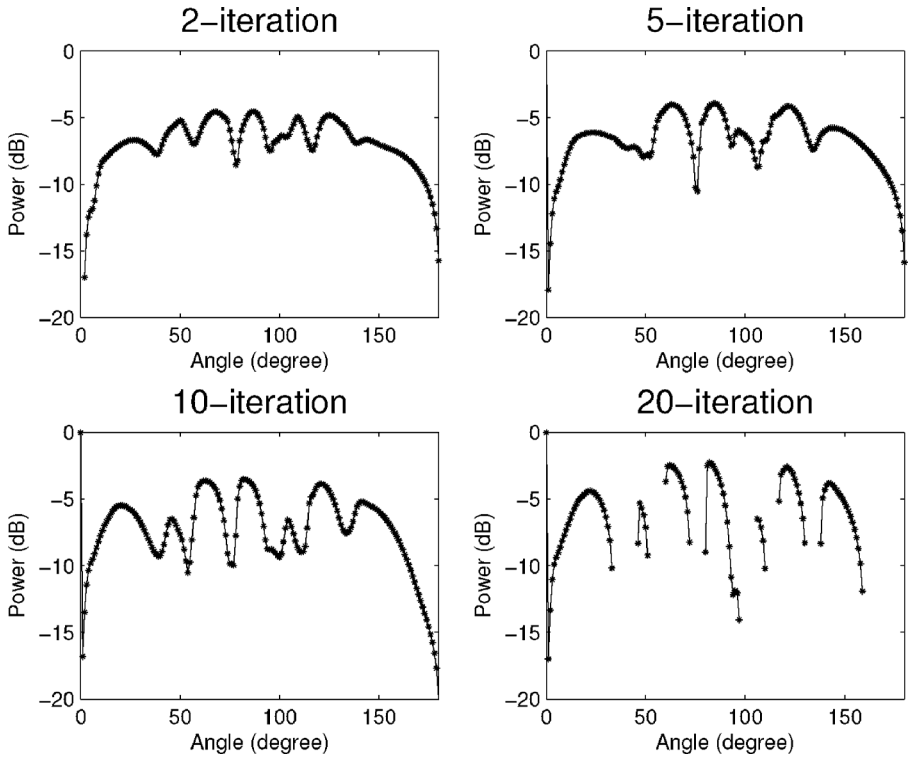

Fig. 1. Solutions of the standard BCD algorithm after 2, 5, 10, 20 major iterations. Here each iteration contains M parallel optimization steps.

$\lambda\|\mathbf{S}\|_{2,1} = \lambda\sum_{i=1}^{M}\|\mathbf{s}^i\|_2$. At the k-th major iteration, we first build a quadratic approximation for $G(\mathbf{S})$ at the previous solution $\mathbf{S}(k)$, such that the approximation of $F(\mathbf{S})$ is

$$\begin{aligned} F_a(\mathbf{S}) &= G(\mathbf{S}(k)) + vec(\nabla G(\mathbf{S}(k)))^H vec(\mathbf{S}-\mathbf{S}(k)) \\ &\quad + \frac{1}{2\beta}\|\mathbf{S}-\mathbf{S}(k)\|_F^2 + H(\mathbf{S}) \\ &= \sum_{i=1}^{M}\left\{\frac{1}{2\beta}\|\mathbf{s}^i - \mathbf{p}^i(k)\|_2^2 + \lambda\|\mathbf{s}^i\|_2\right\} + c(k). \end{aligned} \quad (8)$$

Here, $\mathbf{P}(k) = \mathbf{S}(k) - \beta\nabla G(\mathbf{S}(k)) = \mathbf{S}(k) - \beta\mathbf{B}^H(\mathbf{B}\mathbf{S}(k) - \mathbf{X})$; $\mathbf{s}^i$ and $\mathbf{p}^i(k)$ are the i-th rows of $\mathbf{S}$ and $\mathbf{P}(k)$, respectively; $c(k) = G(\mathbf{S}(k)) - \frac{\beta}{2}\|\nabla G(\mathbf{S}(k))\|_F^2$; β is chosen to be $\beta = \frac{1}{\|\mathbf{b}_i\|_2^2}$ where $\mathbf{b}_i$ is any i-th column of $\mathbf{B}$. Since all columns of the basis matrix $\mathbf{B}$ have the same norm in our DOA estimation problem, $\frac{1}{\|\mathbf{b}_i\|_2^2}$ is identical for $i = 1, \ldots, M$. The reason for choosing such a value will be given in Section III.B and Appendix B.

The standard BCD algorithm directly minimizes (8) to update $\mathbf{S}(k+1)$ [15]. Notice that (8) is separable; therefore, it can be minimized in a parallel manner. The solution to the i-th subproblem is given by a soft-thresholding operator (see [17])

$$\mathbf{s}^i(k+1) = \frac{\mathbf{p}^i(k)}{\|\mathbf{p}^i(k)\|_2}\max(0, \|\mathbf{p}^i(k)\|_2 - \lambda\beta). \quad (9)$$

Therefore, each major iteration consists of M parallel optimization steps. Notice that for our DOA estimation problem, we ultimately seek an extremely sparse signal. But the standard BCD algorithm does not give preference to sparsity explicitly. More precisely, each parallel optimization step may change a block (i.e., a row of $\mathbf{S}$) which should be zero during the sweep (i.e., the major iteration). Fig. 1 depicts the spatial spectra solved by the standard BCD algorithm after 2, 5, 10, 20 major iterations[3]. Obviously, the solved spatial spectra are far from sparse.

In order to maintain sparsity during the major iterations, we propose a greedy block selection rule which leads to a greedy block coordinate descent (GBCD) algorithm. Instead of sweeping through all the blocks in parallel, we only update the block that yields the greatest descent distance, namely

$$\begin{aligned} i_0 &= \arg\max_i \triangle d_i \\ &= \arg\max_i \|\mathbf{s}^i(k+1) - \mathbf{s}^i(k)\|_2,\ i = 1, \ldots, M. \end{aligned} \quad (10)$$

Here $\triangle d_i = \|\mathbf{s}^i(k+1) - \mathbf{s}^i(k)\|_2$ is the descent distance for the i-th block and $\mathbf{s}^i(k+1)$ is solved as (9). We choose to only update the block i_0; namely, $\mathbf{s}^{i_0}(k+1)$ is updated as (9) while $\mathbf{s}^i(k+1) = \mathbf{s}^i(k)$, $i \neq i_0$. This way, we avoid update of unnecessary blocks in each major iteration and improve the efficiency of the algorithm. On the other hand, the block selection rule brings little computational load over the standard BCD algorithm. The resulting GBCD algorithm is outlined in Table I.

The spatial spectra solved after 2, 5, 10, 20 iterations by the GBCD algorithm are demonstrated in Fig. 2, where the experimental settings are the same as those in Fig. 1. We can see that the GBCD algorithm achieves much sparser solutions than the standard BCD algorithm, and reaches near the true signals much faster. Now a natural question arises: *though the greedy block*

[3]In this experiment, we set the number of array elements to be $N = 11$. The total number of snapshots is $L = 200$ and the grid spacing is 1° from 0° to 180°. Five narrowband far-field signals with equal power impinge on the array from different angles [20° 65° 85° 123° 150°]. The SNR is set to be 0 dB; λ is chosen according to Section VI. The starting point is set to be $\mathbf{S}(0) = \mathbf{0}_{M\times R}$ which is an $M \times R$ zero matrix.

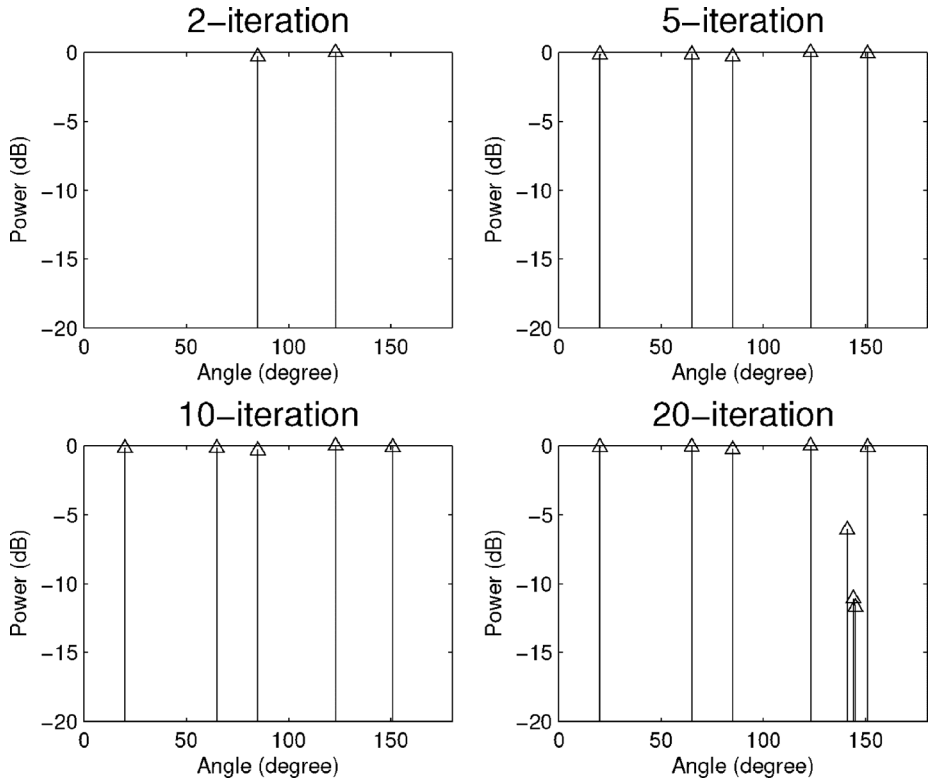

Fig. 2. Solutions of the GBCD algorithm after 2, 5, 10, 20 major iterations. Here each iteration updates the block that yields the greatest descent distance, as shown in (10).

TABLE I
GBCD: GREEDY BLOCK COORDINATE DESCENT ALGORITHM

Given: $\mathbf{S}(0) = \mathbf{0}_{M\times R}$, $\mathbf{B} \in \mathcal{C}^{N\times M}$, $\mathbf{X} \in \mathcal{C}^{N\times R}$, $\beta = \frac{1}{\|\mathbf{b}_i\|_2^2}$, $i = 1, \cdots, M$
repeat
  $\hat{\mathbf{S}} \leftarrow \mathbf{S}(k-1)$
  for $i = 1 : M$
    $\mathbf{p}^i = \hat{\mathbf{s}}^i - \beta \mathbf{b}_i^H(\mathbf{B}\hat{\mathbf{S}} - \mathbf{X})$ where $\hat{\mathbf{s}}^i$ is the i-th row of $\hat{\mathbf{S}}$
    $\mathbf{s}^i(k) = \frac{\mathbf{p}^i}{\|\mathbf{p}^i\|_2} \max(0, \| \mathbf{p}^i \|_2 - \lambda\beta)$
    $comp(i) = \| \mathbf{s}^i(k) - \hat{\mathbf{s}}^i \|_2$
  end for
  Choose the index i_0 such that $comp(i_0) = \max(comp)$
  $\mathbf{S}(k) \leftarrow [\hat{\mathbf{s}}^1; \cdots; \hat{\mathbf{s}}^{i_0-1}; \mathbf{s}^{i_0}(k); \hat{\mathbf{s}}^{i_0+1}; \hat{\mathbf{s}}^M]$
until satisfy certain optimality condition or reach certain number of iterations

selection rule gives preference to sparsity and has faster convergence, does it guarantee global convergence? In the following subsection, we will prove that the GBCD algorithm still guarantees convergence to the global optimal solutions of (6) with the greedy block selection rule.

B. Global Convergence of the GBCD Algorithm

We prove the global convergence of the proposed GBCD algorithm by showing that it indeed belongs to the general block coordinate gradient descent framework in [18]. The existing convergence proof of the general block coordinate gradient descent framework, combining with our convergence analysis on the solution sequence, leads to the global convergence of the GBCD algorithm.

We first introduce the general block coordinate gradient descent framework and its existing convergence proof. Consider minimizing a function $f(\tilde{\mathbf{s}}) = g(\tilde{\mathbf{s}}) + h(\tilde{\mathbf{s}})$ where $\tilde{\mathbf{s}} \in \mathcal{C}^{M'}$, $g(\tilde{\mathbf{s}})$ is a smooth and often inseparable convex function, and $h(\tilde{\mathbf{s}})$ is a separable and often non-smooth convex function. In our settings, $\tilde{\mathbf{s}} = vec(\mathbf{S})$, and $M' = RM$. At the k-th iteration, the general block coordinate gradient descent framework in [18] replaces $g(\tilde{\mathbf{s}})$ by a quadratic approximation at its previous solution $\tilde{\mathbf{s}}(k)$ and then inexactly solves a problem with the form

$$f_a(\tilde{\mathbf{s}}) \triangleq g(\tilde{\mathbf{s}}(k)) + \nabla g(\tilde{\mathbf{s}}(k))^H(\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(k)) + \frac{1}{2}(\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(k))^H \mathbf{H}(\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(k)) + h(\tilde{\mathbf{s}}) \quad (11)$$

where $\mathbf{H}$ is a given positive semidefinite matrix to approximate the Hessian $\nabla^2 g(\tilde{\mathbf{s}})$.

Inexactly solving (11) involves finding a descent direction $\tilde{\mathbf{d}}(k)$ and a stepsize $\alpha^{(k)}$. Given a nonempty index set $\mathcal{J}' \subseteq \mathcal{M}' \triangleq \{1, 2, \ldots, M'\}$, the descent direction $\tilde{\mathbf{d}}(k) = \tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}, \mathcal{J}')$ at the k-th iteration can be represented as

$$\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}') \triangleq \arg\min_{\tilde{\mathbf{d}}} \{ f_a(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}) | \tilde{d}_j = 0, \quad \forall j \notin \mathcal{J}' \} \quad (12)$$

and thus $\tilde{\mathbf{s}}(k+1) = \tilde{\mathbf{s}}(k) + \alpha^{(k)}\tilde{\mathbf{d}}(k)$.

Below we describe two rules of choosing $\tilde{\mathbf{d}}(k)$ (equivalently, $\mathcal{J}'$) and $\alpha^{(k)}$ which are necessary to guarantee convergence[4] of the general block coordinate gradient descent framework.

[4]In the sense that every cluster point is a stationary point, as we will explain in detail.

1) Modified Armijo Rule: Choose $\alpha^{(k)} > 0$ so that the following inequality holds:

$$f(\tilde{\mathbf{s}}(k) + \alpha^{(k)}\tilde{\mathbf{d}}(k)) \leqslant f(\tilde{\mathbf{s}}(k)) + \alpha^{(k)}\eta\Delta^{(k)} \tag{13}$$

where $0 < \eta < 1$, and

$$\Delta^{(k)} \triangleq \nabla g(\tilde{\mathbf{s}}(k))^H\tilde{\mathbf{d}}(k) + \gamma\tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) + h(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) - h(\tilde{\mathbf{s}}(k)) \tag{14}$$

where $0 \leqslant \gamma < 1$; and $\bar{\theta}\mathbf{I}_{M'} \succeq \mathbf{H} \succeq \underline{\theta}\mathbf{I}_{M'}$ where $0 < \underline{\theta} < \bar{\theta}$.

This rule is a kind of generalization from the Armijo rule in smooth optimization about how to choose a proper stepsize. Notice that $\Delta^{(k)}$ is nonpositive and increases with γ (which will be proved in Lemma 2); thus, a larger stepsize will be accepted if we choose either η near 0 or γ near 1.

*2) Modified Gauss-Southwell-*r *Rule:* In each step the index set $\mathcal{J}'$ we choose must satisfy

$$\|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}, \mathcal{J}')\|_2 \geqslant \nu\|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}, \mathcal{M}')\|_2 \tag{15}$$

where $0 < \nu \leqslant 1$, while $\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}, \mathcal{J}')$ and $\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}, \mathcal{M}')$ are defined as in (12).

Indeed, this is also an extension from the Gauss-Southwell-r rule in smooth optimization. It requires that minimizing over partial indices still results in a reasonable stepsize compared to minimizing over all indices.

Next we state the existing convergence proof of the general block coordinate gradient descent framework in [18].

Lemma 1: If the modified Armijo rule and the modified Gauss-Southwell-r rule are satisfied at each iteration, then for the sequence $\{\tilde{\mathbf{s}}(k)\}$ generated by the block coordinate gradient descent framework, every cluster point[5] is a stationary point of $f(\tilde{\mathbf{s}})$.

Proof: See Theorem 1(c) in [18]. ■

Second, we prove that the solution sequence $\{\tilde{\mathbf{s}}(k)\}$ has a limit point under certain conditions. This is important since such a sequence has only one cluster point and the cluster point is the limit point.

Lemma 2: If $f(\tilde{\mathbf{s}})$ is coercive (i.e., $lim_{\tilde{s}_j\to\infty} f(\tilde{\mathbf{s}}) = \infty$ for any $j \in \mathcal{M}'$), and at each iteration the modified Armijo rule is satisfied, then the sequence $\{\tilde{\mathbf{s}}(k)\}$ converges to a certain $\tilde{\mathbf{s}} \in dom f$.

Proof: See Appendix A. ■

Finally, we give the global convergence proof of the proposed GBCD algorithm.

Theorem 1: The sequence $\{\tilde{\mathbf{s}}(k)\}$ generated by the GBCD algorithm using the greedy block selection rule (10) converges to an optimal solution of (6).

Proof: See Appendix B. ■

IV. R-GBCD Algorithm and Performance Guarantee

For a DOA estimation problem, we care more about the right locations of the nonzero supports than their corresponding values. After detecting the nonzero supports (angles), various fine search techniques can be applied in estimating the values (signal strengths). Motivated by this fact, we propose to terminate the GBCD algorithm after finding R nonzero supports without running more redundant iterations to refine the values (recall that R is the number of point sources and can be roughly estimated in advance by some existing methods such as minimum description length (MDL) and Akaike information criterion (AIC) [19]). We call this modified GBCD algorithm as the **R-GBCD algorithm**. It is different from the original GBCD algorithm only in the stopping criterion. An important feature of the R-GBCD algorithm is that, under certain conditions, all nonzero supports found are the actual ones.

For simplicity, we only consider the ideal case that the measurement is noiseless; namely, there is a system of linear equations $\mathbf{X} = \mathbf{BS}$ which has a jointly sparse solution with R rows of nonzeros. We need to figure out whether the R-GBCD algorithm will always succeed in finding the actual nonzero supports. Obviously, such success cannot be achieved for any R and $\mathbf{B}$. Notice that for recovering a single sparse signal from linear measurements, if the signal is "sparse enough" and the measurement matrix is "good enough", algorithms including orthogonal matching pursuit (OMP) and basis pursuit (BP) have a similar performance guarantee [20], [21]. Now we address this issue for the jointly sparse signal recovery problem and the proposed R-GBCD algorithm.

Theorem 2: Consider the system of linear equations $\mathbf{BS} = \mathbf{X}$ where all columns of the basis matrix $\mathbf{B}$ have the same norm. Suppose that the jointly sparsest solution $\mathbf{S}_0$ of the linear equations exists such that its columns $\{\mathbf{s}_{0i}\}_{i=1}^M$ have the same sparsity pattern, and

$$R = \|\mathbf{s}_{0i}\|_0 \leq \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{B})}\right), \quad \forall i \tag{16}$$

where $\mu(\mathbf{B}) = \sup_{i\neq j} \frac{\|\mathbf{b}_i^H\mathbf{b}_j\|_2}{\sqrt{\|\mathbf{b}_i^H\mathbf{b}_i\|_2\|\mathbf{b}_j^H\mathbf{b}_j\|_2}}$ is the mutual coherence of $\mathbf{B}$ [22]. Then, all nonzero supports found by the GBCD algorithm with $\beta = \frac{1}{\|\mathbf{b}_i\|_2^2}$, $\lambda > 0$ are the nonzero supports of $\mathbf{S}_0$.

Proof: See Appendix C.

Next we show that for the noiseless case, $\mathbf{S}_0$ is the unique minimizer of (6) when (16) holds and $\lambda > 0, \lambda \to 0$. Therefore, we can guarantee that GBCD is able to find all R nonzero supports and thus R-GBCD is effective.

Lemma 3: Assume that Λ indexes a linear independent collection of multiple columns in $\mathbf{B}$ and induces a corresponding basis matrix $\mathbf{B}_\Lambda$. Let $\mathbf{S}_*$ minimizes the objective function in (6) over all coefficient vectors supported on Λ. Then the condition

$$ERC(\Lambda) \triangleq \max_{i\notin\Lambda} \|\mathbf{B}_\Lambda^\dagger\mathbf{b}_i\|_1 < 1 \tag{17}$$

guarantees that $\mathbf{S}_*$ is the unique global minimizer of (6). Here ERC stands for *exact recovery coefficient*.

Proof: An analogy to this lemma has been proved in [23] for the case of ℓ_1-norm minimization. Interested readers can easily prove this lemma by following the steps of Lemma 6 in [23]. ■

Lemma 4: For a basis matrix $\mathbf{B}$ with its *mutual coherence* define as in Theorem 2, if (16) is satisfied, then for any index set Λ with cardinality equal to or smaller than R, (17) is satisfied.

[5]A cluster point x of a sequence $\{x(k)\}$ is defined as following: for every neighborhood $\mathcal{X}$ of x, there are infinitely many natural numbers k such that $x(k) \in \mathcal{X}$.

Proof: See Theorem 4.7 in [21]. ■

Remark 1: It is obvious that when $\lambda > 0$ and $\lambda \to 0$, $\mathbf{S}_0$ is a global minimizer of (6) (see also Section VI). Lemma 3 and Lemma 4 indeed show that under the condition (16), (6) has an unique minimizer. Therefore, $\mathbf{S}_0$ is the unique minimizer of (6) when (16) holds and $\lambda > 0, \lambda \to 0$. This way, we can guarantee that the GBCD algorithm is able to find all nonzero supports correctly.

Interestingly, this result is similar to the OMP and BP performance guarantee [20], [21]. Although (16) is a fairly pessimistic bound, it states that the GBCD algorithm can always pick up the actual nonzero supports when an inequality parameterized by the mutual-coherence $\mu(\mathbf{B})$ holds. Notice that in the DOA estimation problem, the basis matrix $\mathbf{B}$ is highly correlated. Therefore, the proposed GBCD algorithm (as well as the R-GBCD algorithm) faces potential performance degradation on estimation accuracy. This motivates a refined algorithm, as demonstrated in the next section.

V. Adaptive Weighting With Music Prior

In order to improve the estimation accuracy, we thereby propose an adaptively weighted form of the GBCD algorithm. Specifically, in computing the descent directions, we multiply the term to maximize in (10) by a weight parameter based on the MUSIC prior. This way, we are likely to choose the block corresponding to the actual angle of arrival. Now the block we choose is

$$\begin{aligned} i_0 &= \arg\max_i \triangle\tilde{d}_i \\ &= \arg\max_i w_i \|\mathbf{s}^i(k+1) - \mathbf{s}^i(k)\|_2, \quad i = 1, \ldots, M \end{aligned} \tag{18}$$

where w_i is the weight parameter corresponding to the i-th block and is determined by some prior. Here we choose the MUSIC prior; the reason will be explained below.

Our main motivation is to exploit the orthogonality between the noise subspace and the array steering vectors. The ideal covariance matrix (when $L \to \infty$) for the received signal can be written as

$$\operatorname{cov}(\mathbf{X}) = \mathbf{B}\mathbf{R}_s\mathbf{B}^H + \sigma^2\mathbf{I}_N \tag{19}$$

where $\mathbf{R}_s$ is the covariance matrix of the true point sources, σ^2 is the noise power and $\mathbf{I}_N$ is the $N \times N$ identity matrix.

By the eigenvalue decomposition, $\operatorname{cov}(\mathbf{X}) = \sum_{i=1}^{N} \lambda_i \mathbf{u}_i \mathbf{u}_i^H$ with λ_i being an eigenvalue and $\mathbf{u}_i$ being an eigenvector, we can divide the whole space $\mathcal{C}^N$ into signal subspace $\mathbf{U}_s = [\mathbf{u}_1, \ldots, \mathbf{u}_R]$ and noise subspace $\mathbf{U}_n = [\mathbf{u}_{R+1}, \ldots, \mathbf{u}_M]$. Suppose that the grid is fine enough so as to cover all the true directions, then we can divide the basis matrix $\mathbf{B}$ into $\mathbf{B}_{\mathcal{J}_0}$ and $\mathbf{B}_{\overline{\mathcal{J}}_0}$, where $\mathbf{B}_{\mathcal{J}_0}$ is the true array manifold, $\mathcal{J}_0$ is the index set corresponding to the true directions, and $\overline{\mathcal{J}}_0 = \frac{\mathcal{M}}{\mathcal{J}_0}$. The classical MUSIC algorithm [2] has shown that

$$\mathbf{U}_n^H \mathbf{B}_{\mathcal{J}_0} = \mathbf{0}, \tag{20}$$

$$\mathbf{U}_n^H \mathbf{B}_{\overline{\mathcal{J}}_0} = \mathbf{Y}, \tag{21}$$

where $\mathbf{Y}$ is some matrix with dimension $(N-R) \times (M-R)$. Notice that in practice, we are estimating the covariance matrix through finite samples; hence, the estimated noise subspace is not exactly the real one. Also, the grid cannot cover all the true directions in general.

Based on the MUSIC prior, we choose the weight parameters $\{w_i\}_{i=1}^M$ as

$$w_i = \frac{1}{\|\mathbf{U}_n^H \mathbf{b}_i\|_2}, \qquad i = 1, \ldots, M. \tag{22}$$

When enough samples are taken, we will have

$$w_i \gg w_j, \quad \forall i \in \mathcal{J}_0, \forall j \in \overline{\mathcal{J}}_0. \tag{23}$$

By this way of choosing weight parameters, $\triangle\tilde{d}_i$ in (18) whose index is within the true nonzero supports of the jointly sparse signals are penalized by a larger weight, while those outside of the true nonzero supports are penalized by smaller weights.

Lastly, notice that the weighted selecting rule of descent directions, when applied to the regular GBCD algorithm in Section III, does not change its global convergence. To see this, we just need to show that this rule actually satisfies the modified Gauss-Southwell-r rule because the modified Armijo rule and the coercive property of the objective function are not interfered here.

At the k-th iteration, we update the block $i_0 = \arg\max_i w_i \|\mathbf{s}^i(k+1) - \mathbf{s}^i(k)\|_2$, which means that

$$\begin{aligned} & w_{i_0} \|\mathbf{s}^{i_0}(k+1) - \mathbf{s}^{i_0}(k)\|_2 \\ & \geqslant \min_i w_i \max_i \|\mathbf{s}^i(k+1) - \mathbf{s}^i(k)\|_2 \qquad (24) \\ & \Rightarrow \|\mathbf{s}^{i_0}(k+1) - \mathbf{s}^{i_0}(k)\|_2 \\ & \geqslant \frac{\min_i w_i}{\max_i w_i} \max_i \|\mathbf{s}^i(k+1) - \mathbf{s}^i(k)\|_2. \qquad (25) \end{aligned}$$

Furthermore, according to (48), we have

$$\begin{aligned} \max_i \|\mathbf{s}^i(k+1) - \mathbf{s}^i(k)\|_2 &= \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_{i_0})\|_2 \\ &\geqslant \frac{1}{\sqrt{M}} \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{M}')\|_2. \end{aligned} \tag{26}$$

This combines with (25) yields

$$\begin{aligned} \|\mathbf{s}^{i_0}(k+1) - \mathbf{s}^{i_0}(k)\|_2 &= \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_{i_0})\|_2 \\ &\geqslant \frac{1}{\sqrt{M}} \frac{\min_i w_i}{\max_i w_i} \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{M}')\|_2 \end{aligned} \tag{27}$$

with $\nu = \frac{1}{\sqrt{M}} \frac{\min_i w_i}{\max_i w_i} > 0$. Here $\mathcal{M}' = \{1, \ldots, RM\}$ and $\tilde{\mathbf{s}} = vec(\mathbf{S})$ are defined in Appendix B.

VI. Regularization Parameter Selection

The regularization parameter λ in (6) plays an important role in the DOA estimation problem, since it represents the tradeoff between data fidelity (the least squares term) and joint sparsity

(the $\ell_{2,1}$-norm term). Here we present a theoretical guidance for choosing an appropriate regularization parameter. An analogy to this work appears in [24], which focuses on the ℓ_1-norm minimization problem. To this end, we rewrite (6) as its equivalent constrained form, which builds the connection between λ and the noise ratio

$$\begin{aligned} \min_{\tilde{\mathbf{s}}} \quad & \|\tilde{\mathbf{B}}\tilde{\mathbf{s}}\|_2 \\ s.t. \quad & \sqrt{\sum_{j=1}^{R}(\tilde{\mathbf{b}}_{i\times j}^H(\tilde{\mathbf{x}}-\tilde{\mathbf{B}}\tilde{\mathbf{s}}))^2} \leqslant \lambda, \quad i=1,\ldots,M \end{aligned} \tag{28}$$

where $\tilde{\mathbf{B}} = \mathbf{I}_R \otimes \mathbf{B}$ is an $NR \times MR$ structured basis matrix, $\tilde{\mathbf{s}} = vec(\mathbf{S}) \in \mathcal{C}^{MR}$, $\tilde{\mathbf{x}} = vec(\mathbf{X}) \in \mathcal{C}^{NR}$, and $\tilde{\mathbf{b}}_{i\times j}$ denotes the $(i \times j)$-th column of $\tilde{\mathbf{B}}$. The equivalence between (6) and (28) is detailed in [6].

When $\mathbf{S}$ (and hence $\tilde{\mathbf{s}} = vec(\mathbf{S})$) is the true value, it is obvious that $\tilde{\mathbf{x}} - \tilde{\mathbf{B}}\tilde{\mathbf{s}}$ is an complex Gaussian random vector with mean $\mathbf{0}$ and covariance matrix $\sigma^2\mathbf{I}_{NR}$ where σ^2 is the noise power. We denote this as $\tilde{\mathbf{x}} - \tilde{\mathbf{B}}\tilde{\mathbf{s}} \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I}_{NR})$. Recalling that $\tilde{\mathbf{B}}$ is a structured matrix and the ℓ_2-norm for each column of $\mathbf{B}$ is N, we further get $\sqrt{\frac{2}{N\sigma^2}}\tilde{\mathbf{b}}_{i\times j}^H(\tilde{\mathbf{x}} - \tilde{\mathbf{B}}\tilde{\mathbf{s}}) \sim \mathcal{N}(0,2)$, and $\frac{2\|\tilde{\mathbf{b}}_{i\times j}^H(\tilde{\mathbf{x}}-\tilde{\mathbf{B}}\tilde{\mathbf{s}})\|_2^2}{N\sigma^2}$ satisfies the $\mathcal{X}^2$ distribution with 2 degrees of freedom. Hence, $\frac{2\sum_{j=1}^{R}(\tilde{\mathbf{b}}_{i\times j}^H(\tilde{\mathbf{x}}-\tilde{\mathbf{B}}\tilde{\mathbf{s}}))^2}{N\sigma^2}$ satisfies the $\mathcal{X}^2$ distribution with $2R$ degrees of freedom.

Now, selection of the regularization parameter is much more transparent. We choose λ high enough so as that the probability of all M inequalities $\sqrt{\sum_{j=1}^{R}(\tilde{\mathbf{b}}_{i\times j}^H(\tilde{\mathbf{x}} - \tilde{\mathbf{B}}\tilde{\mathbf{s}}))^2} \leqslant \lambda$, $i = 1,\ldots,M$ being satisfied, denoted as p_0, is very high. Define the probability that a single equality is satisfied as $\tilde{p_0}$. For simplicity, we consider all M inequalities to be uncorrelated with each other and get $\tilde{p}_0^M = p_0$. Thus, for any $\tilde{p}_0$ such that $Pr(\mathcal{X}_{2R}^2 \leqslant \xi) = \tilde{p}_0$, we have

$$\lambda = \sqrt{\frac{N\sigma^2\xi}{2}}. \tag{29}$$

Note that although the choice of λ depends on our prior knowledge about noise power, we are still able to estimate it well by using the average of $N - R$ smallest eigenvalues in the covariance matrix of the received signals.

VII. Experimental Results

This section presents experimental results to illustrate the performances of the proposed algorithms and compare them with several existing methods. All numerical experiments have been done in Matlab on a laptop with Intel Core i3 CPU and 1.92 G RAM. We consider uncorrelated point sources[6], and the point sources are uncorrelated with noise. Both the signals and the noise are white and follow Gaussian distributions. The input SNR of the r-th point source is defined as $10\log_{10}\left(\frac{\sigma_r^2}{\sigma^2}\right)$, where σ_r^2 and σ^2 are the standard variances of the r-th point source and the noise, respectively. The covariance matrix of the signals is estimated through L snapshots as $\frac{\sum_{t=1}^{L}\mathbf{x}(t)\mathbf{x}^H(t)}{L}$.

We compare the proposed algorithms with BCD-GL (Block Coordinate Descent for Group Lasso) in [15] the L1-SVD algorithm in [5], MUSIC in [2] and Root-MUSIC in [26]. For the GBCD algorithm, the parameter λ in objective function (6) is chosen according to p_0, which is set to be 0.99. In the following experiments, we denote the original GBCD algorithm in Section III and the modified GBCD algorithm in Section IV as GBCD and R-GBCD, respectively; the corresponding versions with weight refinement in Section V are denoted as GBCD+ and R-GBCD+, respectively. Moreover, we use GBCD and GBCD+ when analyzing the convergence behavior and use R-GBCD and R-GBCD+ when analyzing DOA estimation performance. The compared L1-SVD is the constrained version

$$\begin{aligned} \min \quad & \|\mathbf{S}\|_{2,1} \\ s.t. \quad & \|\mathbf{X} - \mathbf{B}\mathbf{S}\|_F^2 \leqslant \beta^2 \end{aligned} \tag{30}$$

where β is chosen such that the probability of $\|\mathbf{N}\|_F \geqslant \beta$ is 0.01 (i.e., the confidence interval is set to be 0.99). The L1-SVD model is solved with the optimization software package SeDuMi.

The experiments are arranged as follows. First, we discuss the convergence results of GBCD and GBCD+ compared to the standard BCD-GL method. Then, we present the spatial spectra of the algorithms and compare their CPU running times. Last, we analyze the performance of R-GBCD and R-GBCD+, and simultaneously illustrate their effectiveness comparing with Root-MUSIC, L1-SVD and the Cramer-Rao lower bound (CRLB). Throughout the simulation, we set the number of array elements to be $N = 11$. The total number of snapshots is $L = 200$ and the initial grid spacing is 1° from 0° to 180°.

A. Convergence Behavior

This subsection analyzes the convergence behavior of GBCD and GBCD+, and compares them with the standard BCD-GL method. The grid is set to be within the range of 0° to 180° with 1° spacing. Five uncorrelated signals impinge from [30° 65° 95° 130° 150°]. The SNR is set to be 10 dB. In Fig. 3, we depict the objective function value versus the number of iterations in 20 iterations. And Fig. 4 is the zoom-in version of Fig. 3 in 200 iterations. Both GBCD and GBCD+ converge very fast, with sharp fall in the first few (say, five) iterations. The standard BCD-GL method converges relatively slower. This performance gap is due to the greedy strategy used by GBCD and GBCD+. Moreover, the zoom-in version also demonstrates that the three methods actually converge to the same value ultmately.

B. Spatial Spectra and Computational Speed

Fig. 5 compares the spatial spectra of MUSIC, L1-SVD, R-GBCD and R-GBCD+. Four narrowband far-field point sources with equal power impinge on the array from different angles [27° 69° 123° 152°]. The SNR is set to be 0 dB. Here the vertical value is defined as $10\log\frac{\|\mathbf{s}^i\|_2}{\max_i\|\|\mathbf{s}^i\|_2}$, $i = 1,2,\ldots,M$ for L1-SVD, R-GBCD and R-GBCD+, with $\mathbf{s}^i$ being the i-th

[6]For simplicity, we only consider that the signals are uncorrelated. When the signals are correlated, the GBCD algorithm still works due to the robustness of the $\ell_{2,1}$-norm minimization formulation. Further, there are existing approaches to improve the performance of MUSIC [25]. Using these techniques, the GBCD algorithm with MUSIC prior also works.

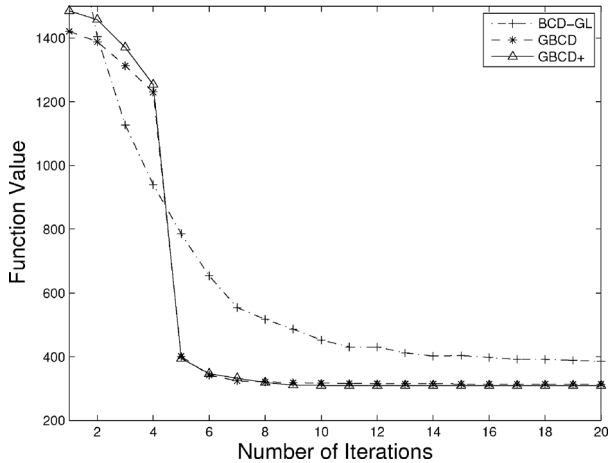

Fig. 3. Objective function value of BCD-GL, GBCD and GBCD+ versus number of iteration with $\mathrm{SNR} = 10$ dB.

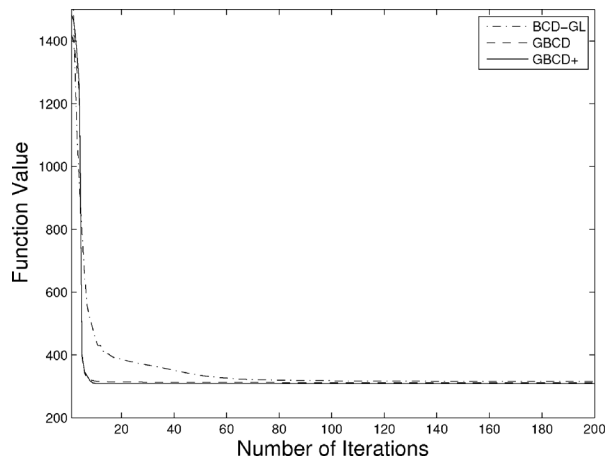

Fig. 4. Zoom in of Fig. 9.

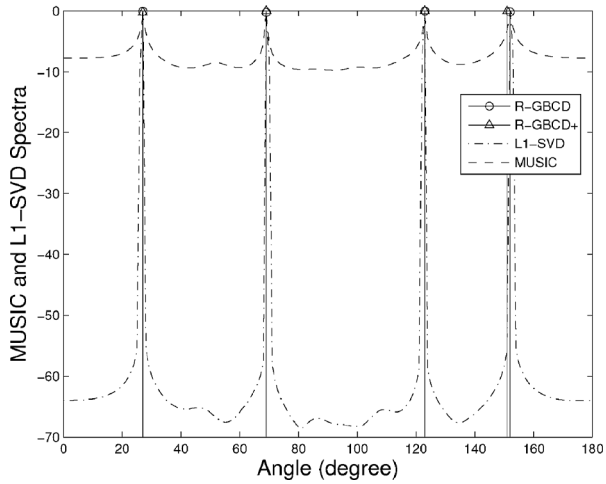

Fig. 5. Spatial spectra of MUSIC and L1-SVD. Those of R-GBCD and R-GBCD+ are also shown by the vertical continuous lines. DOAs: [27° 69° 123° 152°].

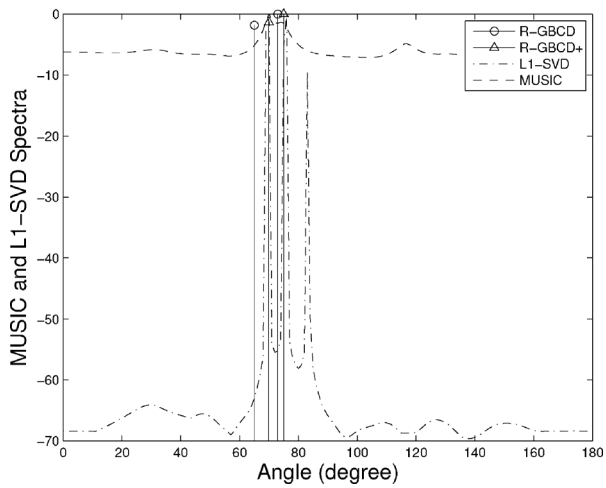

Fig. 6. Spatial spectra of MUSIC and L1-SVD. Those of R-GBCD and R-GBCD+ are also shown by the vertical continuous lines. DOAs: 70° and 75°.

row of the recovered $\mathbf{S}$. For MUSIC, $\|\mathbf{s}^i\|_2$ is $\frac{1}{\|\mathbf{U}_n^H \mathbf{b}_i\|_2}$. We can see that all the four methods are able to detect the correct sources. Both R-GBCD and R-GBCD+ achieve very high spatial resolution, and outperform MUSIC and L1-SVD.

Then we consider two adjacent point sources from 70° and 75°, respectively. As demonstrated in Fig. 6, MUSIC is not able to resolve the two closely spaced signals. L1-SVD identifies two point sources but also bring one false alarm. The two angles found by R-GBCD are biased from the true values. The reason is that the neighboring columns of the basis matrix $\mathbf{B}$ are highly correlated, which hinders the identification of two closely spaced signals. However, after introducing the MUSIC prior, R-GBCD+ achieves accurate estimates without any false alarms. We will revisit the issues of estimation accuracy and resolution in Section VII.C and compare these algorithms numerically.

Next, we compare the running times of the four methods. It is known that when the dimension of the problem is not too large (i.e., N, M, and R do not take too large value), the existing software packages like SeDuMi can solve it quite efficiently. However, for problems with high dimension, the software packages are too time-consuming. In Fig. 7, we present the results of running time versus the number of sources; every point in the figure denotes the average running time of 500 Monte Carlo trials. Here, dimension of the DOA estimation problem grows in proportion to the number of sources. MUSIC yields the fastest computational speed. But for algorithms which solve the $\ell_{2,1}$-norm minimization problem, R-GBCD+ is much faster than the standard software packages used by L1-SVD. The running time of R-GBCD is similar to that of R-GBCD+ since using the MUSIC prior consumes little time.

C. Estimation Accuracy and Resolution

To compare the estimation accuracy of these algorithms, we define the root mean square error (RMSE) of 500 Monte Carlo trials as the performance index

$$RMSE = \sqrt{\sum_{i=1}^{500} \sum_{r=1}^{R} \frac{\left(\hat{\theta}_r(i) - \theta_r\right)^2}{500R}} \tag{31}$$

where $\hat{\theta}_r(i)$ is the estimate of θ_r at the i-th trial.

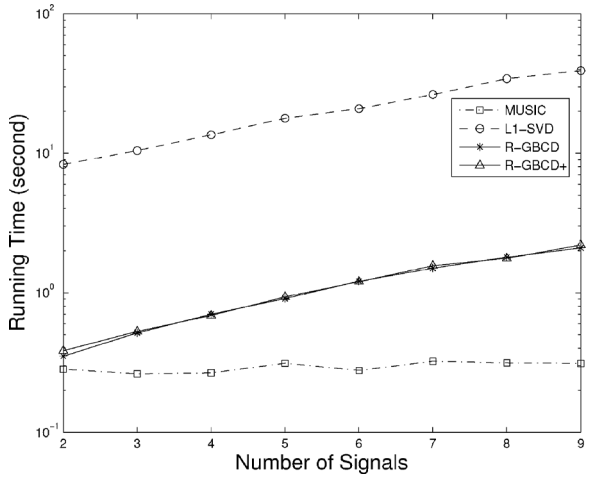

Fig. 7. Running time of MUSIC, L1-SVD, and R-GBCD+ versus number of signals.

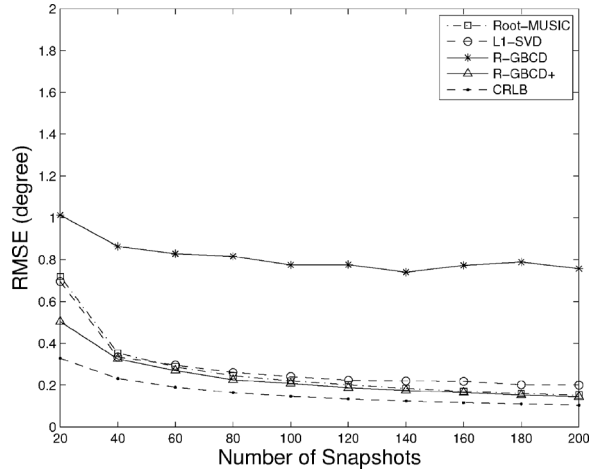

Fig. 9. RMSE of MUSIC, L1-SVD, GBCD and GBCD+ versus number of snapshots with $\mathrm{SNR} = 0$ dB.

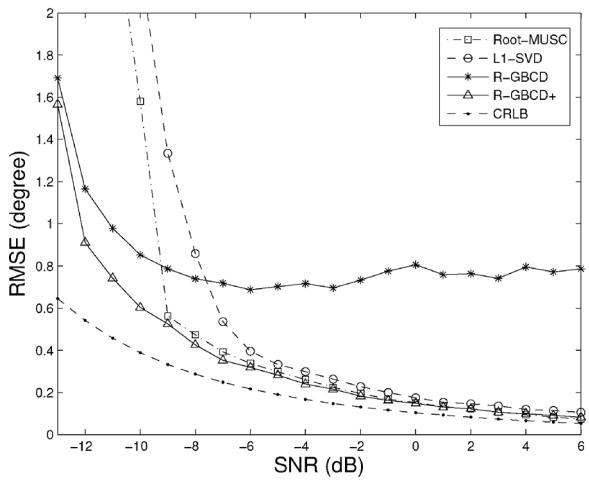

Fig. 8. RMSE of MUSIC, L1-SVD, GBCD and GBCD+ versus input SNR with 200 snapshots.

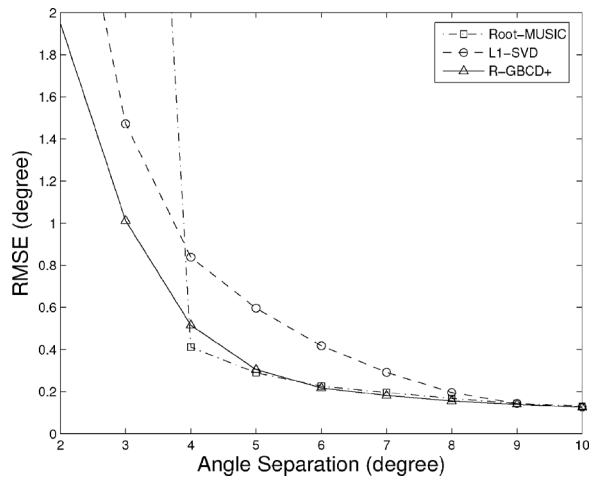

Fig. 10. RMSE of MUSIC, L1-SVD, and R-GBCD+ versus angle separation with $\mathrm{SNR} = 0$ dB.

Note that we cannot make the grid to be very fine uniformly since this will dramatically increase the computational cost. In order to increase the precision as well as avoid too much computation, in the simulation we first use a coarse grid in the range of 0° to 180° with 1° spacing and then set a finer grid around the estimate, similar to that in [5]. In the following experiments, three times of refinement are done for each algorithm with uniform grid spacing 0.2°, 0.1°, and 0.05°, respectively.

The first simulation considers the case of five uncorrelated point sources impinging from five angles [42° 90° 105° 115° 140°]. The RMSE of DOA estimation versus input SNR is shown in Fig. 8 with 200 snapshots, whereas the RMSE of DOA estimation versus the number of snapshots is shown in Fig. 9 with $\mathrm{SNR} = 0$ dB. From Fig. 8, we can see that Root-MUSIC and L1-SVD have similar performance. R-GBCD achieves fairly good estimation accuracy in low SNRs, while R-GBCD+ yields the best performance which is near to the CRLB. This simulation demonstrates that the proposed algorithms are noise resilient. In Fig. 9, R-GBCD+ still has the best performance. Note that Root-MUSIC is relatively inferior when the number of snapshots is small. This is easy to understand since the effectiveness of Root-MUSIC relies on the correct estimate of noise subspace, which is often inaccurate with a small number of snapshots.

The second simulation considers the resolving ability of these methods. As we have analyzed in Section IV, the solution R-GBCD is often biased when two point sources are close. This simulation only compares R-GBCD+, L1-SVD, and Root-MUSIC. Consider two equal-power uncorrelated signals impinging on the array from 90° and $90° + \Delta\theta$. The SNR is 0 dB and the number of snapshots is 200. The RMSE of DOA estimation versus angle separation is presented in Fig. 10. Performance of Root-MUSIC degrades drastically as angle separation is smaller than 4°. The proposed R-GBCD+ algorithm is the most accurate one and outperforms L1-SVD, especially for small angle separation.

VIII. Conclusion

In this paper, we propose a GBCD algorithm for DOA estimation, which is formulated as an $\ell_{2,1}$-norm minimization problem. Different from the existing standard BCD algorithm, the main feature of GBCD is that it adopts a greedy block

selection rule and hence gives preference to joint sparsity directly. Interestingly, this greedy algorithm belongs to a class of block coordinate gradient descent algorithms [18] and satisfies the modified Armijo-rule and the modified Gauss-Southwell-r rule; these properties guarantee its global convergence, as we have proved in the paper. On the other hand, under mild conditions, all nonzero supports found by the GBCD algorithm are the actual ones; this result is the counterpart to that of OMP and BP for ℓ_1-norm minimization. The GBCD algorithm is further refined through using a weighted form of the greedy block selection rule based on the MUSIC prior. Numerical experiments demonstrate that the weighted form of the GBCD method (named as GBCD+) yields faster convergence rate and better recovery performance compared to several traditional DOA estimation methods, especially when the SNR is low and the number of snapshots is small.

Our future research topic is to implement the GBCD algorithm in estimating slowly time-varying point sources. In this case, new observations bring innovation to the current estimate, while the current and future estimates share similar joint sparsity pattern. How to develop an efficient algorithm for this scenario is a challenging but interesting topic.

Appendix A
Proof of Lemma 2

Proof: We first prove that $\Delta^{(k)} \leqslant 0$ in (14). Indeed, for any $0 < \alpha < 1$, based on (12) we have that $f_a(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) \leq f_a(\tilde{\mathbf{s}}(k) + \alpha\tilde{\mathbf{d}}(k))$. Substituting (11) yields

$$\begin{aligned}
&\nabla g(\tilde{\mathbf{s}}(k))^H\tilde{\mathbf{d}}^{(k)} + \frac{1}{2}\tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) + h(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) \\
&\quad\leqslant \nabla g(\tilde{\mathbf{s}}(k))^H(\alpha\tilde{\mathbf{d}}(k)) + \frac{1}{2}(\alpha\tilde{\mathbf{d}}(k))^H\mathbf{H}(\alpha\tilde{\mathbf{d}}(k)) \\
&\qquad + h(\tilde{\mathbf{s}}(k) + \alpha\tilde{\mathbf{d}}(k)) \\
&\quad\leqslant \alpha\nabla g(\tilde{\mathbf{s}}(k))^H\tilde{\mathbf{d}}(k) + \frac{1}{2}\alpha^2\tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) \\
&\qquad + \alpha h(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) + (1-\alpha)h(\tilde{\mathbf{s}}(k)). \qquad (32)
\end{aligned}$$

Rearranging the terms, we have

$$\begin{aligned}
&(1-\alpha)\nabla g(\tilde{\mathbf{s}}(k))^H\tilde{\mathbf{d}}(k) + \frac{1}{2}(1-\alpha^2)\tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) \\
&\qquad +(1-\alpha)(h(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) - h(\tilde{\mathbf{s}}(k))) \leqslant 0. \qquad (33)
\end{aligned}$$

Since $1 - \alpha \geqslant 0$, we divide both sides by $1 - \alpha \geqslant 0$ and let $\alpha \to 1$:

$$\begin{aligned}
&\nabla g(\tilde{\mathbf{s}}(k))^H\tilde{\mathbf{d}}(k) + \tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) \\
&\qquad +h(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) - h(\tilde{\mathbf{s}}(k)) \leqslant 0. \qquad (34)
\end{aligned}$$

Since $0 \leqslant \gamma < 1$, we then have

$$\begin{aligned}
\Delta^{(k)} &= \nabla g(\tilde{\mathbf{s}}(k))^H\tilde{\mathbf{d}}(k) + \gamma\tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) \\
&\quad + h(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) - h(\tilde{\mathbf{s}}(k)) \\
&\leqslant (\gamma - 1)\tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) \leqslant 0. \qquad (35)
\end{aligned}$$

Second, we prove that $\{\tilde{\mathbf{s}}(k)\}$ has a convergent subsequence. Since the modified Armijo rule is satisfied at each step according to the assumptions, it is easy to get for any $0 < \eta < 1$

$$\begin{aligned}
\min_{\tilde{\mathbf{s}}} f(\tilde{\mathbf{s}}) &\leqslant f(\tilde{\mathbf{s}}(k+1)) = f(\tilde{\mathbf{s}}(k) + \alpha^{(k)}\tilde{\mathbf{d}}(k)) \\
&\leqslant f(\tilde{\mathbf{s}}(k)) + \alpha^{(k)}\eta\Delta^{(k)} \leqslant f(\tilde{\mathbf{s}}(k)). \qquad (36)
\end{aligned}$$

Thus, $f(\tilde{\mathbf{s}}(k))$ is nonincreasing and bounded from below, so the sequence $\{f(\tilde{\mathbf{s}}(k))\}$ must converge. Moreover, because $f(\tilde{\mathbf{s}})$ is coercive, according to Weierstrass' Theorem, it must have bounded sublevel sets. These two facts implies that $\{\|\tilde{\mathbf{s}}(k)\|_\infty\}$ is bounded. Thus, there exists a subsequence of $\{\tilde{\mathbf{s}}(k)\}$ that converges to some $\tilde{\mathbf{s}} \in domf$.

Finally we prove that $\{\tilde{\mathbf{s}}(k)\}$ converges to $\tilde{\mathbf{s}}$. According to (35) and (36), we have

$$\begin{aligned}
|f(\tilde{\mathbf{s}}(k+1)) - f(\tilde{\mathbf{s}}(k))| &\geqslant |\alpha^{(k)}\eta\Delta^{(k)}| \\
&\geqslant \alpha^{(k)}\eta(1-\gamma)\tilde{\mathbf{d}}(k)^H\mathbf{H}\tilde{\mathbf{d}}(k) \geqslant \frac{\eta(1-\gamma)\underline{\theta}}{\alpha^{(k)}}\|\alpha^{(k)}\tilde{\mathbf{d}}(k)\|_2^2. \qquad (37)
\end{aligned}$$

The inequality (37), combining with the facts that $\lim_{k\to\infty}|f(\tilde{\mathbf{s}}(k+1)) - f(\tilde{\mathbf{s}}(k))| = 0$ (since $\{f(\tilde{\mathbf{s}}(k))\}$ converges) and that $\alpha^{(k)}\tilde{\mathbf{d}}(k) = \tilde{\mathbf{s}}(k+1) - \tilde{\mathbf{s}}(k)$, leads to

$$\lim_{k\to\infty}\|\tilde{\mathbf{s}}(k+1) - \tilde{\mathbf{s}}(k)\|_2 = 0. \qquad (38)$$

Therefore, $\{\tilde{\mathbf{s}}(k)\}$ is a Cauchy sequence and hence converges to $\tilde{\mathbf{s}}$. ■

Appendix B
Proof of Theorem 1

Proof: The proof is sketched as follows. Since it is obvious that the objective function in (6) is coercive, we need to show that the proposed GBCD algorithm satisfies the modified Armijo rule and the modified Gauss-Southwell-r rule. Under this circumstance, both Lemma 1 and Lemma 2 hold. From Lemma 1, every cluster point of $\tilde{\mathbf{s}}(k)$ is a stationary point of $f(\tilde{\mathbf{s}})$; from Lemma 2, there is only one cluster point, namely, the limit point. Because (6) is a convex program and its stationary points are global optimal points, the above conclusions prove the global convergence of the proposed GBCD algorithm.

Next, we will show that the proposed GBCD algorithm satisfies the modified Armijo rule and the modified Gauss-Southwell-r rule.

To see the first, we recast the objective function $F(\mathbf{S}) = G(\mathbf{S}) + H(\mathbf{S})$ in (7), an equivalent form of (6), as a function with vectorized variables. Define $\tilde{\mathbf{B}} \triangleq \mathbf{I}_R \otimes \mathbf{B}$ as an $NR \times MR$ basis matrix, $\tilde{\mathbf{s}} \triangleq vec(\mathbf{S}) \in \mathcal{C}^{MR}$, and $\tilde{\mathbf{x}} \triangleq vec(\mathbf{X}) \in \mathcal{C}^{NR}$, then the objective function can be written as[7]

$$F(\tilde{\mathbf{s}}) = G(\tilde{\mathbf{s}}) + H(\tilde{\mathbf{s}}) = \frac{1}{2}\|\tilde{\mathbf{x}} - \tilde{\mathbf{B}}\tilde{\mathbf{s}}\|_2^2 + \lambda\sum_{i=1}^{M}\|\tilde{\mathbf{s}}_{\mathcal{J}_i}\|_2 \qquad (39)$$

[7]With a slight abuse of notation, we still use F, G, H, and F_a here.

where $\mathcal{J}_i = \{i, i+M, \ldots, i+(R-1)M\}$. We further define $\mathcal{M}' = \{1, \ldots, RM\}$.

Then at the k-th iteration, the GBCD algorithm is to find a descent direction $\tilde{\mathbf{d}}(k) = \tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_{i_0})$ which is represented as

$$\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_{i_0}) \triangleq \arg\min_{\tilde{\mathbf{d}}} \Big\{ G(\tilde{\mathbf{s}}(k)) + \nabla G(\tilde{\mathbf{s}}(k))^H \tilde{\mathbf{d}} + \frac{1}{2\beta}\tilde{\mathbf{d}}^H\tilde{\mathbf{d}} + H(\tilde{\mathbf{s}}(k)+\tilde{\mathbf{d}}) | \tilde{d}_j = 0, \quad \forall j \notin \mathcal{J}_{i_0} \Big\}. \quad (40)$$

Here

$$\begin{aligned} i_0 &= \arg\max_i \|\tilde{\mathbf{s}}(k+1)_{\mathcal{J}_i} - \tilde{\mathbf{s}}(k)_{\mathcal{J}_i}\|_2 \\ &= \arg\max_i \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_i)\|_2 \end{aligned}$$

and $\beta = \frac{1}{\|\mathbf{b}_i\|_2^2}$ for any i. Because of the special structure of $\tilde{\mathbf{B}} = \mathbf{I}_R \otimes \mathbf{B}$, the principal submatrix of the actual Hessian of $G(\tilde{\mathbf{s}})$ (i.e., $\mathbf{H}_0 = \nabla^2 G(\tilde{\mathbf{s}}) = \tilde{\mathbf{B}}^H\tilde{\mathbf{B}}$) corresponding to the index set $\mathcal{J}_i$, denoted as $\mathbf{H}_{0\mathcal{J}_i\mathcal{J}_i}$, is a multiple of identity matrix

$$\mathbf{H}_{0\mathcal{J}_i\mathcal{J}_i} = \|\mathbf{b}_i\|_2^2 \mathbf{I}_R = \frac{1}{\beta}\mathbf{I}_R. \quad (41)$$

At the k-th iteration, we only update the entries corresponding to the index set $\mathcal{J}_{i_0}$. Therefore, we can directly substitute the Hessian matrix $\mathbf{H}_0$ into (40) and rewrite it as

$$\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_{i_0}) \triangleq \arg\min_{\tilde{\mathbf{d}}} \Big\{ G(\tilde{\mathbf{s}}(k)) + \nabla G(\tilde{\mathbf{s}}(k))^H \tilde{\mathbf{d}} + \frac{1}{2}\tilde{\mathbf{d}}^H\mathbf{H}_0\tilde{\mathbf{d}} + H(\tilde{\mathbf{s}}(k)+\tilde{\mathbf{d}}) | \tilde{d}_j = 0, \quad \forall j \notin \mathcal{J}_{i_0} \Big\}. \quad (42)$$

The descent direction $\tilde{\mathbf{d}}(k)$ at the k-th iteration is then chosen according the above criterion. Since $\mathbf{H}_0$ is the actual Hessian matrix, using the convexity of $G(\tilde{\mathbf{s}})$, we get the following inequality

$$\begin{aligned} &G(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) \\ &\quad \leqslant G(\tilde{\mathbf{s}}(k)) + \nabla G(\tilde{\mathbf{s}}(k))^H \tilde{\mathbf{d}}(k) + \frac{1}{2}\tilde{\mathbf{d}}(k)^H\mathbf{H}_0\tilde{\mathbf{d}}(k). \end{aligned} \quad (43)$$

Recall that $F(\tilde{\mathbf{s}}) = G(\tilde{\mathbf{s}}) + H(\tilde{\mathbf{s}})$. Thus, it is natural to have

$$\begin{aligned} &F(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) \\ &\leqslant F(\tilde{\mathbf{s}}(k)) + \nabla G(\tilde{\mathbf{s}}(k))^H \tilde{\mathbf{d}}^{(k)} + \frac{1}{2}\tilde{\mathbf{d}}(k)^H\mathbf{H}_0\tilde{\mathbf{d}}(k) \\ &\quad + H(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) - H(\tilde{\mathbf{s}}(k)) \\ &\leqslant F(\tilde{\mathbf{s}}(k)) + \eta\nabla G(\tilde{\mathbf{s}}(k))^H \tilde{\mathbf{d}}^{(k)} + \frac{1}{2}\tilde{\mathbf{d}}(k)^H\mathbf{H}_0\tilde{\mathbf{d}}(k) \\ &\quad + H(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) - H(\tilde{\mathbf{s}}(k)) \end{aligned} \quad (44)$$

where $0 < \eta < 1$. The second inequality is due to the fact that

$$\nabla G(\tilde{\mathbf{s}}(k))^H \tilde{\mathbf{d}}^{(k)} + \gamma\tilde{\mathbf{d}}(k)^H\mathbf{H}_0\tilde{\mathbf{d}}(k) + H(\tilde{\mathbf{s}}(k) + \tilde{\mathbf{d}}(k)) - H(\tilde{\mathbf{s}}(k)) \leq 0$$

for any $0 \leqslant \gamma < 1$, which has been proved[8] in Lemma 1. Notice that (44) is just the modified Armijo rule with $\alpha^{(k)} = 1$ and $\gamma = \frac{1}{2}$.

[8]This proof, see (35), only requires that $\tilde{\mathbf{d}}^{(k)}$ is a descent direction and $\mathbf{H}$ approximates $G(\tilde{\mathbf{s}})$.

Second, we show that

$$\begin{aligned} i_0 &= \arg\max_i \|\tilde{\mathbf{s}}(k+1)_{\mathcal{J}_i} - \tilde{\mathbf{s}}(k)_{\mathcal{J}_i}\|_2 \\ &= \arg\max_i \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_i)\|_2 \end{aligned}$$

leads to an update satisfying the modified Gauss-Southwell-r rule.

Since at the k-th iteration, the approximated objective function

$$F_a(\tilde{\mathbf{s}}) \triangleq G(\tilde{\mathbf{s}}(k)) + \nabla G(\tilde{\mathbf{s}}(k))^H(\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(k)) + \frac{1}{2}(\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(k))^H\mathbf{H}_0(\tilde{\mathbf{s}} - \tilde{\mathbf{s}}(k))) + H(\tilde{\mathbf{s}}) \quad (45)$$

is separable to different blocks $\mathcal{J}_i = \{i, i+M, \ldots, i+(R-1)M\}, i = 1, \ldots, M$, is a totally separable problem, we have

$$\|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_{i_0})\|_2 \geqslant \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_i)\|_2, \quad \forall i \neq i_0. \quad (46)$$

Combining with

$$\|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{M}')\|_2 = \sqrt{\sum_{i=1}^{M} \|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_i)\|_2^2} \quad (47)$$

yields

$$\|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{J}_{i_0})\|_2 \geqslant \frac{1}{\sqrt{M}}\|\tilde{\mathbf{d}}_H(\tilde{\mathbf{s}}(k), \mathcal{M}')\|_2 \quad (48)$$

where $\nu = \frac{1}{\sqrt{M}} > 0$.

In sum, we have proved the global convergence of the proposed GBCD algorithm. ■

APPENDIX C
PROOF OF THEOREM 2

Proof: Without loss of generality, suppose that for the jointly sparsest solution $\mathbf{S}_0$ to the system of linear equations, all its R nonzero rows are at the beginning of the vectors. Therefore,

$$\mathbf{X} = \mathbf{B}\mathbf{S}_0 = \sum_{r=1}^{R} \mathbf{b}_r \mathbf{s}_0^r. \quad (49)$$

And at the k-th iteration, there exists an $i_k \in \{1, \ldots, R\}$ such that $\|\mathbf{s}_0^{i_k} - \mathbf{s}^{i_k}(k)\|_2 \geq \|\mathbf{s}_0^i - \mathbf{s}^i(k)\|_2, \forall i \in \{1, \ldots, R\}$.

In the proof, we show that the GBCD algorithm picks up a block corresponding to a correct nonzero support at each iteration. We prove it by induction.

At the 1-st iteration, the initial value is $\mathbf{S}(0) = \mathbf{0}$. Next we prove that the GBCD algorithm will never pick up the j-th block, $\forall j \in \{R+1, \ldots, M\}$. According to (9) and (10), the descent direction for each block is

$$\begin{aligned} \Delta d_i &= \|\mathbf{s}^i(1) - \mathbf{0}\|_2 \\ &= \left\| \frac{\mathbf{p}^i(0)}{\|\mathbf{p}^i(0)\|_2} \max(0, \|\mathbf{p}^i(0)\|_2 - \lambda\beta) \right\|_2, \quad i = 1, \ldots, M. \end{aligned} \quad (50)$$

Thus, in order to avoid picking up the j-th block, $\forall j \in \{R+1,\ldots,M\}$, the sufficient condition is that $\forall j \in \{R+1,\ldots,M\}$

$$\left\| \frac{\mathbf{p}^{i_1}(0)}{\|\mathbf{p}^{i_1}(0)\|_2} \max(0, \|\mathbf{p}^{i_1}(0)\|_2 - \lambda\beta) \right\|_2 \geq \left\| \frac{\mathbf{p}^{j}(0)}{\|\mathbf{p}^{j}(0)\|_2} \max(0, \|\mathbf{p}^{j}(0)\|_2 - \lambda\beta) \right\|_2, \quad (51)$$

meaning that the GBCD algorithm will at least pick up the i_1-th block, if not other $i \in \{1,\ldots,R\}$ and $i \neq i_1$. Note that (51) equivalent to

$$\|\mathbf{p}^{i_1}(0)\|_2 \geq \|\mathbf{p}^{j}(0)\|_2, \quad \forall j \in \{R+1,\ldots,M\}. \quad (52)$$

Now substituting $\mathbf{p}^i(0) = \beta \mathbf{b}_i^H \mathbf{X}$ and (49) into (52), then it is equivalent to

$$\left\| \sum_{r=1}^{R} \mathbf{b}_{i_1}^H \mathbf{b}_r \mathbf{s}_0^r \right\|_2 \geq \left\| \sum_{r=1}^{R} \mathbf{b}_{j}^H \mathbf{b}_r \mathbf{s}_0^r \right\|_2, \ \forall j \in \{R+1,\ldots,M\}. \quad (53)$$

We then find a lower bound for the left-hand-side and an upper bound for the right-hand-side in order to give a sufficient condition. For the left-hand-side, we have

$$\left\| \sum_{r=1}^{R} \mathbf{b}_{i_1}^H \mathbf{b}_r \mathbf{s}_0^r \right\|_2 \geqslant \frac{1}{\beta}\|\mathbf{s}_0^{i_1}\|_2 - \sum_{r=1, r\neq i_1}^{R} \|\mathbf{b}_{i_1}^H \mathbf{b}_r\|_2 \|\mathbf{s}_0^r\|_2 \geqslant \frac{1}{\beta}\|\mathbf{s}_0^{i_1}\|_2 (1 - (R-1)\mu(\mathbf{B})). \quad (54)$$

Here we utilize $\mu(\mathbf{B}) = \sup_{i\neq j} \frac{\|\mathbf{b}_i^H \mathbf{b}_j\|_2}{\sqrt{\|\mathbf{b}_i^H \mathbf{b}_i\|_2 \|\mathbf{b}_j^H \mathbf{b}_j\|_2}} = \sup_{i\neq j} \frac{\|\mathbf{b}_i^H \mathbf{b}_j\|_2}{\beta}$ and $\mathbf{s}_0^{i_1} \geq \mathbf{s}_0^i, \forall i \in \{1,\ldots,R\}$. Similarly, we have for the right-hand-side

$$\left\| \sum_{r=1}^{R} \mathbf{b}_{j}^H \mathbf{b}_r \mathbf{s}_0^r \right\|_2 \leqslant \sum_{r=1}^{R} \|\mathbf{b}_j^H \mathbf{b}_r\|_2 \|\mathbf{s}_0^r\|_2 \leqslant \frac{1}{\beta}\|\mathbf{s}_0^{i_1}\|_2 R\mu(\mathbf{B}), \ \forall j \in \{R+1,\ldots,M\}. \quad (55)$$

Plugging these two bounds into the inequality (53), we obtain the sufficient condition

$$\frac{1}{\beta}\|\mathbf{s}_0^{i_1}\|_2 (1 - (R-1)\mu(\mathbf{B})) \geq \frac{1}{\beta}\|\mathbf{s}_0^{i_1}\|_2 R\mu(\mathbf{B}). \quad (56)$$

This is equivalent to

$$R \leq \frac{1}{2}\left(1 + \frac{1}{\mu(\mathbf{B})}\right). \quad (57)$$

Suppose that up to the $(k-1)$-th iteration, the GBCD algorithm always picks up blocks corresponding to the correct nonzero supports. At the k-th iteration, in order to avoid picking up the j-th block, $\forall j \in \{R+1,\ldots,M\}$, the sufficient condition is that $\forall j \in \{R+1,\ldots,M\}$

$$\left\| \frac{\mathbf{p}^{i_k}(k-1)}{\|\mathbf{p}^{i_k}(k-1)\|_2} \max(0, \|\mathbf{p}^{i_k}(k-1)\|_2 - \lambda\beta) - \mathbf{s}^{i_k}(k-1) \right\|_2 \geq \left\| \frac{\mathbf{p}^{j}(k-1)}{\|\mathbf{p}^{j}(k-1)\|_2} \max(0, \|\mathbf{p}^{j}(k-1)\|_2 - \lambda\beta) - \mathbf{s}^{j}(k-1) \right\|_2. \quad (58)$$

Using the triangle inequality for the left-hand-side yields

$$\left\| \frac{\mathbf{p}^{i_k}(k-1)}{\|\mathbf{p}^{i_k}(k-1)\|_2} \max(0, \|\mathbf{p}^{i_k}(k-1)\|_2 - \lambda\beta) - \mathbf{s}^{i_k}(k-1) \right\|_2 \geq \|\mathbf{p}^{i_k}(k-1) - \mathbf{s}^{i_k}(k-1)\|_2 - \lambda\beta. \quad (59)$$

For the right-hand-side, since $\mathbf{s}^j(k-1) = \mathbf{0}, \forall j \in \{R+1,\ldots,M\}$, any $\|\mathbf{p}^j(k-1) - \mathbf{s}^j(k-1)\|_2 \leq \lambda\beta$ makes (58) automatically hold. Therefore, we only consider the case of $\|\mathbf{p}^j(k-1) - \mathbf{s}^j(k-1)\|_2 \geq \lambda\beta$, which makes the right-hand-side be $\|\mathbf{p}^j(k-1) - \mathbf{s}^j(k-1)\|_2 - \lambda\beta$. Combining the two is the sufficient condition of (58), namely,

$$\|\mathbf{p}^{i_k}(k-1) - \mathbf{s}^{i_k}(k-1)\|_2 \geq \|\mathbf{p}^{j}(k-1) - \mathbf{s}^{j}(k-1)\|_2, \quad \forall j \in \{R+1,\ldots,M\}, \quad (60)$$

or equivalently

$$\left\| \sum_{r=1}^{R} \mathbf{b}_{i_k}^H \mathbf{b}_r (\mathbf{s}_0^r - \mathbf{s}^r(k-1)) \right\|_2 \geq \left\| \sum_{r=1}^{R} \mathbf{b}_{j}^H \mathbf{b}_r (\mathbf{s}_0^r - \mathbf{s}^r(k-1)) \right\|_2, \quad \forall j \in \{R+1,\ldots,M\}. \quad (61)$$

Using the same technique from (54) to (56), similarly we get the sufficient condition as (57). ■

ACKNOWLEDGMENT

The authors are grateful to Prof. W. Yin of Rice University for his constructive comments.

REFERENCES

[1] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
[2] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Anten. Propagat.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[3] R. Roy and T. Kailath, "ESPRIT-Estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 10, pp. 984–995, Oct. 1989.
[4] J. J. Fuchs, "On the application of the global matched filter to DOA estimation with uniform circular arrays," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 702–729, Apr. 2001.
[5] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.
[6] X. Xu, X. Wei, and Z. Ye, "DOA estimation based on sparse signal recovery utilizing weighted ℓ_1-norm penalty," *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 155–158, Mar. 2012.
[7] H. Zhu, G. Leus, and G. B. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.
[8] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, Feb. 2011.
[9] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *J. Roy. Statist. Soc., Ser. B*, vol. 68, no. 1, pp. 49–67, Feb. 2007.
[10] E. van den Berg, M. Schmidt, M. Friedlander, and K. Murphy, "Group sparsity via linear-time projection," Dept. Comput. Sci., Univ. British Columbia, Vancouver, BC, Canada, Tech. Rep. TR-2008-09, Jun. 2008.
[11] J. Liu, S. Ji, and J. Ye, "SLEP: Sparse learning with efficient projections," [Online]. Available: http://www.public.asu.edu/~jye02/Software/SLEP/.
[12] Y. Eldar, P. Kuppinger, and H. Bolcskei, "Block-sparse signals: Uncertainty relations and efficient recovery," *IEEE Trans. Signal Process.*, vol. 58, no. 6, pp. 3042–3054, Jun. 2010.

[13] S. Wright, R. Nowak, and M. Figueiredo, "Sparse reconstruction by separable approximation," *IEEE Trans. Signal Process.*, vol. 57, no. 7, pp. 2479–2493, Jul. 2009.
[14] W. Deng, W. Yin, and Y. Zhang, "Group sparse optimization by alternating direction method," Dept. Comput. Appl. Math., Rice Univ., Houston, TX, Tech. Rep. TR12–13, Jan. 2012.
[15] Z. Qin, K. Scheinberg, and D. Goldfarb, "Efficient block-coordinate descent algorithms for the group lasso," [Online]. Available: http://www.columbia.edu/~zq2107/glasso_bcd_mpc_v2.pdf}
[16] J. S. Sturm, "Using SeDuMi 1.02, a Matlab toolbox for optimization over symmetric cones," [Online]. Available: http://fewcal.kub.nl sturm.
[17] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.
[18] P. Tseng and S. Yun, "A coordinate gradient descent method for nonsmooth separable minimization," *Math. Programming*, vol. 117, no. 1, pp. 387–423, 2009.
[19] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 33, no. 2, pp. 387–392, Feb. 1985.
[20] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Jun. 1999.
[21] M. Elad, *Sparse and Redundant Representations–From Theory to Applications in Signal and Image Processing*. New York: Springer, 2010.
[22] D. Donoho, M. Elad, and V. Temlyakov, "Stable recovery of sparse overcomplete representations in the presense of noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 6–18, Feb. 2006.
[23] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Inf. Theory*, vol. 52, no. 3, pp. 1030–1051, Mar. 2006.
[24] J. J. Fuchs, "The generalized likelihood ratio test and the sparse representations approach," in *Proc. Int. Conf. Image Process.*, Sep. 2010, pp. 245–253.
[25] T. J. Shan, M. Wax, and T. Kailath, "On spatial smoothing for direction-of-arrival estimation of coherent signals," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 3, no. 4, pp. 806–811, Aug. 1985.
[26] B. D. Rao and K. V. S. Hari, "Performance analysis of root-MUSIC," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 12, pp. 1939–1949, Dec. 1989.

Xiaohan Wei received the B.S. degree in electrical engineering and information science from University of Science and Technology of China, Hefei, Anhui, China in 2012.

He is currently working toward the M.S. degree in electrical engineering at the University of Southern California, Los Angeles, CA. His research interests include arrays signal processing, compressed sensing, and convex optimization.

Yabo Yuan is currently working toward the B.S. degree in electrical engineering and information science at the University of Science and Technology of China, Hefei, Anhui, China.

Her research interests include optical signal processing, green optical networks, and convex optimization.

Qing Ling received the B.S. degree in automation and the Ph.D. degree in control theory and control engineering from the University of Science and Technology of China, Hefei, Anhui, in 2001 and 2006, respectively.

From 2006 to 2009, he was a Postdoctoral Research Fellow with the Department of Electrical and Computer Engineering, Michigan Technological University, Houghton. Since 2009, he has been an Assistant Professor with the Department of Automation, University of Science and Technology of China. His current research focuses on decentralized optimization of networked multi-agent systems.