

Article

Off-Grid DOA Estimation Using Alternating Block Coordinate Descent in Compressed Sensing

Weijian Si, Xinggen Qu and Zhiyu Qu *

Department of Information and Communication Engineering, Harbin Engineering University, Harbin 150001, China; E-Mails: swj0418@263.net (W.S.); quxinggen@126.com (X.Q.)

* Author to whom correspondence should be addressed; E-Mail: quzhiyu@hrbeu.edu.cn; Tel.: +86-451-8251-8874.

Academic Editor: Vittorio M. N. Passaro

Received: 1 July 2015 / Accepted: 6 August 2015 / Published: 27 August 2015

Abstract: This paper presents a novel off-grid direction of arrival (DOA) estimation method to achieve the superior performance in compressed sensing (CS), in which DOA estimation problem is cast as a sparse reconstruction. By minimizing the mixed k - l norm, the proposed method can reconstruct the sparse source and estimate grid error caused by mismatch. An iterative process that minimizes the mixed k - l norm alternately over two sparse vectors is employed so that the nonconvex problem is solved by alternating convex optimization. In order to yield the better reconstruction properties, the block sparse source is exploited for off-grid DOA estimation. A block selection criterion is engaged to reduce the computational complexity. In addition, the proposed method is proved to have the global convergence. Simulation results show that the proposed method has the superior performance in comparisons to existing methods.

Keywords: alternating block coordinate descent; block selection criterion; block sparse source; compressed sensing; off-grid direction of arrival (DOA) estimation

1. Introduction

Direction of arrival (DOA) estimation has played an important role in many fields, such as radar, medical imaging and array signal processing [1,2]. In the last decades, many classical methods have been developed, among which multiple signal classification (MUSIC) [3] and estimation of signal

parameter via rotational invariance technique. (ESPRIT) [4] are the most popular and have high resolution for DOA estimation. However, these methods are very sensitive to the number of snapshots, signal to noise ratio (SNR) and the correlation between sources. Small number of snapshots, low SNR and high correlation or coherent sources can all make the performance of these methods degrade severely. More recently, the emerging field of compressed sensing (CS) [5,6] has attracted enormous attention and it can reconstruct the sparse source using nonadaptive linear projection measurement obtained by the measurement matrix that satisfies the restricted isometry property (RIP) [7,8]. Support denotes the set that contains the indices of the nonzero elements in the sparse source. Once the support is determined, the sparse source can be reconstructed.

Due to the fact that sources are intrinsically sparse in the spatial domain, the DOA estimation problem can be regarded as a sparse reconstruction in the framework of CS. The CS-based estimation methods have much better estimation performance than conventional estimation methods. Malioutov *et al.* [9] firstly adopted the sparse signal reconstruction (SSR) perspective for DOA estimation and utilized the singular value decomposition (SVD) of the data matrix to propose l_1 -SVD method. In [10], CS-MUSIC was proposed by revisiting the link between CS and MUSIC. This method identifies the parts of support using CS, after which the remaining parts are estimated by a novel generalized MUSIC criterion. Xu *et al.* [11] utilized the Capon spectrum to design a weighted l_1 -norm penalty in order to further enforce the sparsity and approximate the original l_0 -norm for DOA estimation. Wei *et al.* [12] proposed a modified greedy block coordinate descent (R-GBCD) method and the corresponding version with weight refinement (R-GBCD+) to improve the estimation performance.

The key to guarantee the performance of conventional CS-based estimation methods is that all true DOAs are exactly located on the grid. However, when true DOAs are not on the grid set, the performance may severely degrade due to the grid error caused by mismatch, which is defined as the distance from the true direction to the nearest grid. In order to address this issue, Zhu *et al.* [13] proposed the sparse total least square (STIS) to perform the off-grid DOA estimation, in which perturbations of the model are assumed to be Gaussian. In [14], Yang *et al.* introduced the Bayesian theory in off-grid DOA estimation and proposed an off-grid sparse Bayesian inference based on the singular value decomposition (OGSBI-SVD). Liang *et al.* [15] proposed an off-grid synchronous approach based on distributed compressed sensing to obtain larger array aperture. Zhang *et al.* [16] formulated a novel model based on the sampling covariance matrix and solved the off-grid DOA estimation problem by the block sparse Bayesian method even if the number of sources are unknown.

In this paper, a novel alternating block coordinate descent method called ABCD is proposed for off-grid DOA estimation in CS. The proposed method solves the mixed k - l norm minimization problem to reconstruct the sparse source and estimate the grid error. Since joint estimation will lead to a nonconvex optimization problem, the proposed method adopts an iterative process that minimizes the mixed k - l norm alternately over two sparse vectors. Instead of conventional sparse source, the block sparse source is exploited to achieve better reconstruction properties. The block is updated by the proposed block selection criterion, which can improve efficiency of the proposed method. In addition, we give a detailed derivation process of proving the global convergence of the proposed method. Simulation results illustrate the superior performance of the proposed method as compared with existing methods.

The rest of the paper is organized as follows. An off-grid DOA estimation model is formulated in Sections 2 and 3 introduces the proposed method in detail. The global convergence of the proposed method is proved in Sections 4 and 5 shows the performance of the proposed method. Conclusions are provided in Section 6.

2. Problem Formulation

Consider K far-field narrowband sources $s_1(t), s_2(t), \dots, s_K(t)$ impinging on the uniform linear array (ULA) consisting of M omnidirectional sensors with inter-sensor spacing $\hat{d}$. Assume that each source $s_k(t)$ is located at different direction θ_k with the power σ_k , $k=1, 2, \dots, K$. At time instant t , the received source by ULA can be expressed as

$$\mathbf{x}(t) = \sum_{k=1}^K \mathbf{a}(\theta_k) s_k(t) + \mathbf{n}(t) \quad t=1, 2, \dots, T \quad (1)$$

where $\mathbf{a}(\theta_k) \in \mathbb{C}^{M \times 1}$ and $\mathbf{n}(t) \in \mathbb{C}^{M \times 1}$ denote the steering vector and noise vector, respectively. Since the first point of ULA is set as the origin of sensor array, the m th element of $\mathbf{a}(\theta_k)$ is written as $e^{-j(2\pi/\lambda)\hat{d}(m-1)\cos\theta_k}$ with the wavelength of source λ .

To incorporate the CS theory with the DOA estimation, the entire angular space is divided into a fine grid $\tilde{\theta} = [\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_L]^T$, where $L (L \gg K)$ denotes the number of the grid and $[\cdot]^T$ denotes the transpose. Due to the fact that true directions $\theta = [\theta_1, \theta_2, \dots, \theta_K]^T$ are random in the entire angular space, θ_k for some $k \in \{1, 2, \dots, K\}$ are likely to be not on the grid set. To reduce the grid error caused by mismatch, we formulate the off-grid model, which has a close relationship with the on-grid model. Let $\tilde{\theta}$ satisfy the uniform distribution so that the grid interval $\tau = \tilde{\theta}_{k+1} - \tilde{\theta}_k \propto L^{-1}$. Without loss of generality, by assuming $\theta_k \notin \{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_L\}$ and that $\tilde{\theta}_{l_k} \in \{\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_L\}$ is the nearest grid to θ_k , the steering vector $\mathbf{a}(\theta_k)$ can be approximated as

$$\mathbf{a}(\theta_k) \approx \mathbf{a}(\tilde{\theta}_{l_k}) + \mathbf{b}(\tilde{\theta}_{l_k})(\theta_k - \tilde{\theta}_{l_k}) \quad (2)$$

where $\varepsilon_l = \theta_k - \tilde{\theta}_{l_k} \in [-\frac{\tau}{2}, \frac{\tau}{2}]$ is the grid error and $\mathbf{b}(\tilde{\theta}_{l_k})$ is the partial derivative of $\mathbf{a}(\tilde{\theta}_{l_k})$ with respect to $\tilde{\theta}_{l_k}$. Then, if $\tilde{\theta}_{l_1}, \tilde{\theta}_{l_2}, \dots, \tilde{\theta}_{l_K}$ are respectively nearest grids to the true directions $\theta_1, \theta_2, \dots, \theta_K$, we have

$$\begin{cases} g_l(t) = s_{l_k}(t), & \varepsilon_l = \theta_k - \tilde{\theta}_{l_k}; & l = l_k (k=1, 2, \dots, K) \\ g_l(t) = \varepsilon_l = 0 & ; & elsewhere \end{cases} \quad (3)$$

Then, by imposing the approximation error on the noise, the received source $\mathbf{x}(t)$ can be rewritten as the following sparse form

$$\mathbf{x}(t) = [\mathbf{A} + \mathbf{B} \text{diag}(\varepsilon)] \mathbf{g}(t) + \hat{\mathbf{n}}(t) \quad t=1, 2, \dots, T \quad (4)$$

where $\mathbf{A} = [\mathbf{a}(\tilde{\theta}_1), \mathbf{a}(\tilde{\theta}_2), \dots, \mathbf{a}(\tilde{\theta}_L)]$ is the $M \times L$ array manifold matrix corresponding to all potential directions, which is defined as an overcomplete dictionary in CS, and $\mathbf{B} = [\mathbf{b}(\tilde{\theta}_1), \mathbf{b}(\tilde{\theta}_2), \dots, \mathbf{b}(\tilde{\theta}_L)]$. In addition, the matrix $\hat{\mathbf{N}} = [\hat{\mathbf{n}}(1), \hat{\mathbf{n}}(2), \dots, \hat{\mathbf{n}}(T)]$ is the noise matrix and $\text{diag}(\varepsilon)$ is a diagonal matrix

with ε being the diagonal elements. Since $\mathbf{g}(t) = [g_1(t), g_2(t), \dots, g_L(t)]^T$ has K nonzero elements in L elements, it is a K -sparse vector, where K is referred to as sparsity. More specifically, $\varepsilon = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_L]^T$ is also a K -sparse vector and has the same support as $\mathbf{g}(t)$. It is evident that the on-grid sparse model is a special case by simply setting $\varepsilon = 0$ in Equation (4). Since $\{\mathbf{g}(t)\}_{t=1}^T$ are jointly K -sparse, the matrix $\mathbf{G} = [\mathbf{g}(1), \mathbf{g}(2), \dots, \mathbf{g}(T)] \in \mathbb{C}^{L \times T}$ has K nonzero rows and is called row K -sparse.

To solve the off-grid DOA estimation problem, we need to jointly estimate the support of sparse sources, $\mathbf{g}(t)$, $t = 1, 2, \dots, T$, and grid error ε from the matrix $\mathbf{Y}$ which is given by

$$\mathbf{Y} = [\mathbf{y}(1) \ \mathbf{y}(2) \ \dots \ \mathbf{y}(T)] = \Phi \mathbf{A} \mathbf{G} + \Phi \mathbf{B} \text{diag}(\varepsilon) \mathbf{G} + \tilde{\mathbf{N}} \quad (5)$$

where $\tilde{\mathbf{N}} = \Phi \hat{\mathbf{N}}$ is the $N \times T$ noise matrix, Φ is the $N \times M$ measurement matrix with $N < M$ and N is the number of nonadaptive linear projection measurement.

3. Off-DOA Estimation

In this section, an alternating block coordinate descent (ABCD) method and a block selection criterion are elaborated in the CS scenario. The proposed method not only has the advantages of conventional BCD [17] strategy, but also uses an iterative process that minimizes the mixed k - l norm alternately over two sparse vectors. Note that due to solving the minimization alternately, a tractable convex problem is obtained and the global convergence of ABCD can be easily determined, which is proved in the next section.

By applying the central limit theorem, the components $\hat{\mathbf{n}}(t)$, $t = 1, 2, \dots, T$, of $\hat{\mathbf{N}}$ are independently white Gaussian noise with zero mean and covariance $\sigma^2 \mathbf{I}_M$, where $\mathbf{I}_M$ denotes an $M \times M$ identity matrix. Thus, the covariance matrix of $\mathbf{Y}$ with the size $M \times M$ is expressed as

$$\mathbf{R}_Y = E[\mathbf{y}(t) \mathbf{y}^H(t)] = \Phi \mathbf{A} \mathbf{R}_G \mathbf{A}^H \Phi^H + \Phi \mathbf{B} \text{diag}(\varepsilon) \mathbf{R}_G \text{diag}^H(\varepsilon) \mathbf{B}^H \Phi^H + \sigma^2 \Phi \Phi^H \quad (6)$$

where $\mathbf{R}_G = E[\mathbf{s}(t) \mathbf{s}^H(t)]$ is a $L \times L$ covariance matrix of the sparse source and $(\bullet)^H$ denotes the conjugate transpose. Since all potential directions are one to one corresponding to the powers $\sigma_1^2, \sigma_2^2, \dots, \sigma_L^2$ and we are interested in estimating DOAs, $\mathbf{R}_G$ can be reduced to a diagonal matrix $\mathbf{R}_G = \text{diag}(\mathbf{p})$, where $\mathbf{p} = [\sigma_1^2, \sigma_2^2, \dots, \sigma_L^2]^T$ is a K -sparse vector. Then, by denoting $\hat{\mathbf{A}} = \Phi \mathbf{A} = [\hat{\mathbf{a}}(\tilde{\theta}_1), \hat{\mathbf{a}}(\tilde{\theta}_2), \dots, \hat{\mathbf{a}}(\tilde{\theta}_L)]$ and $\hat{\mathbf{B}} = \Phi \mathbf{B} = [\hat{\mathbf{b}}(\tilde{\theta}_1), \hat{\mathbf{b}}(\tilde{\theta}_2), \dots, \hat{\mathbf{b}}(\tilde{\theta}_L)]$, (6) can be further rewritten as

$$\mathbf{R}_Y = \sum_{l=1}^L \hat{\mathbf{a}}(\tilde{\theta}_l) \hat{\mathbf{a}}^H(\tilde{\theta}_l) \sigma_l^2 + \sum_{l=1}^L \hat{\mathbf{b}}(\tilde{\theta}_l) \hat{\mathbf{b}}^H(\tilde{\theta}_l) \sigma_l^2 \varepsilon_l^2 + \sigma^2 \Phi \Phi^H \quad (7)$$

Due to the vector form of $\mathbf{R}_Y$ in Equation (7), the following measurement vector is given by

$$\mathbf{z} = \text{vec}(\mathbf{R}_Y) = \mathbf{C} \mathbf{p} + \mathbf{D} \mathbf{q} + \sigma^2 \Phi \Phi^H \quad (8)$$

with $\mathbf{C} = [\hat{\mathbf{a}}^H(\tilde{\theta}_1) \otimes \hat{\mathbf{a}}(\tilde{\theta}_1), \hat{\mathbf{a}}^H(\tilde{\theta}_2) \otimes \hat{\mathbf{a}}(\tilde{\theta}_2), \dots, \hat{\mathbf{a}}^H(\tilde{\theta}_L) \otimes \hat{\mathbf{a}}(\tilde{\theta}_L)]$ and $\mathbf{D} = [\hat{\mathbf{b}}^H(\tilde{\theta}_1) \otimes \hat{\mathbf{b}}(\tilde{\theta}_1), \hat{\mathbf{b}}^H(\tilde{\theta}_2) \otimes \hat{\mathbf{b}}(\tilde{\theta}_2), \dots, \hat{\mathbf{b}}^H(\tilde{\theta}_L) \otimes \hat{\mathbf{b}}(\tilde{\theta}_L)]$, where $\otimes$ and $\text{vec}(\bullet)$ denote the Kronecker product and the stack operation by placing the columns of a matrix on the top of one another in order, respectively. Moreover, $\mathbf{q} = [\sigma_1^2 \varepsilon_1^2, \sigma_2^2 \varepsilon_2^2, \dots, \sigma_L^2 \varepsilon_L^2]^T$ is also a K -sparse vector, which has the same support as $\mathbf{p}$. In the conventional sparse source, the nonzero elements of the sparse vector $\mathbf{p}$ or $\mathbf{q}$ can appear anywhere

in the vector. However, in this paper, our goal is to explicitly take block sparse source into account, *i.e.*, the nonzero elements of the sparse vector $\mathbf{p}$ or $\mathbf{q}$ tend to cluster in blocks. The motivations to exploit block sparse source are the following two main reasons. As can be seen in [18], the first reason is that block sparse source has been applied in many applications, such as unions of subspaces and multiband sources [19]. Secondly, block sparse source has better reconstruction properties than sparse source in the conventional sense, which is proved in [20]. To exploit block sparse source, denote $\mathbf{p}[h] = [\sigma_{d(h-1)+1}^2, \sigma_{d(h-1)+2}^2, \dots, \sigma_{dh}^2]^T$ and $\mathbf{q}[h] = [\sigma_{d(h-1)+1}^2 \varepsilon_{d(h-1)+1}^2, \sigma_{d(h-1)+2}^2 \varepsilon_{d(h-1)+2}^2, \dots, \sigma_{dh}^2 \varepsilon_{dh}^2]^T$ as the h th blocks of $\mathbf{p}$ and $\mathbf{q}$ with the length d , respectively, so that we have

$$\mathbf{p} = [\mathbf{p}^T[1], \mathbf{p}^T[2], \dots, \mathbf{p}^T[H]]^T \quad (9)$$

$$\mathbf{q} = [\mathbf{q}^T[1], \mathbf{q}^T[2], \dots, \mathbf{q}^T[H]]^T \quad (10)$$

where $L = dH$. Following the similar manner, the matrices $\mathbf{C}$ and $\mathbf{D}$ can be respectively viewed as a concatenation of block matrices $\mathbf{C}[h]$ and $\mathbf{D}[h]$ of the size $N^2 \times d$, *i.e.*,

$$\mathbf{C} = [\mathbf{C}[1], \mathbf{C}[2], \dots, \mathbf{C}[H]] \quad (11)$$

$$\mathbf{D} = [\mathbf{D}[1], \mathbf{D}[2], \dots, \mathbf{D}[H]] \quad (12)$$

where $\mathbf{C}[h] = [\hat{\mathbf{a}}^H(\tilde{\theta}_{d(h-1)+1}) \otimes \hat{\mathbf{a}}(\tilde{\theta}_{d(h-1)+1}), \hat{\mathbf{a}}^H(\tilde{\theta}_{d(h-1)+2}) \otimes \hat{\mathbf{a}}(\tilde{\theta}_{d(h-1)+2}), \dots, \hat{\mathbf{a}}^H(\tilde{\theta}_{dh}) \otimes \hat{\mathbf{a}}(\tilde{\theta}_{dh})]$ and $\mathbf{D}[h] = [\hat{\mathbf{b}}^H(\tilde{\theta}_{d(h-1)+1}) \otimes \hat{\mathbf{b}}(\tilde{\theta}_{d(h-1)+1}), \hat{\mathbf{b}}^H(\tilde{\theta}_{d(h-1)+2}) \otimes \hat{\mathbf{b}}(\tilde{\theta}_{d(h-1)+2}), \dots, \hat{\mathbf{b}}^H(\tilde{\theta}_{dh}) \otimes \hat{\mathbf{b}}(\tilde{\theta}_{dh})]$. Obviously, the conventional sparse source is a special case of block sparse source by simply setting $d=1$. If $\|\mathbf{p}[h]\|_2$, $h=1, 2, \dots, H$, has at most K nonzero elements, the vector $\mathbf{p}$ is referred to as block K -sparse, where $\|\cdot\|_2$ denotes the Euclidean norm for vectors. In contrast to conventional sparsity, K is called block-sparsity.

Since $\mathbf{p}$ has the same support as $\mathbf{q}$, $\mathbf{p}$ and $\mathbf{q}$ are jointly sparse. Thus, a mixed k - l norm minimization problem [21] is utilized to jointly reconstruct $\mathbf{p}$ and $\mathbf{q}$. Given a $L \times 1$ block sparse vector $\mathbf{p}$, the mixed k - l norm of $\mathbf{p}$ is defined as

$$\|\mathbf{p}\|_{k,l}^l = \sum_{h=1}^H (\|\mathbf{p}[h]\|_k)^l \quad (13)$$

Combining the definition given by Equation (13), this mixed k - l norm minimization problem is formulated as

$$\min_{\mathbf{p}, \mathbf{q}} (\|\mathbf{p}\|_{k,l}^l + \beta \|\mathbf{q}\|_{k,l}^l) \quad s.t. \quad \mathbf{z} = \mathbf{C}\mathbf{p} + \mathbf{D}\mathbf{q} + \sigma^2 \Phi \Phi^H \quad (14)$$

where $0 \leq l \leq 1$. It is worth mentioning that an important class of methods for solving the constrained optimization problem is to form the auxiliary function. By introducing the Lagrange multiplier method, the Lagrange function with respect to Equation (14) is given by

$$\min_{\mathbf{p}, \mathbf{q}} \frac{1}{2} \|\mathbf{z} - \mathbf{C}\mathbf{p} - \mathbf{D}\mathbf{q}\|_2^2 + \beta_p \sum_{h=1}^H (\|\mathbf{p}[h]\|_k)^l + \beta_q \sum_{h=1}^H (\|\mathbf{q}[h]\|_k)^l \quad (15)$$

where β_p and β_q are regularized parameters. As can be seen in Equation (15), the minimization problem with respect to $\mathbf{p}$ and $\mathbf{q}$ is nonconvex so as to make DOA estimation intractable. But note that if we fix

one of two sparse vectors, that is, either $\mathbf{p}$ or $\mathbf{q}$, the minimization problem in Equation (15) turns out to be convex with respect to the other sparse vector alone. Thus, $\mathbf{p}$ and $\mathbf{q}$ can be reconstructed by alternately solving the following two minimization problems under the condition that the other sparse vector is fixed.

$$\min_{\mathbf{p}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{C}\mathbf{p} - \mathbf{D}\mathbf{q}\|_2^2 + \beta_p \sum_{h=1}^H (\|\mathbf{p}[h]\|_k)^l \right\} \quad (16)$$

$$\min_{\mathbf{q}} \left\{ \frac{1}{2} \|\mathbf{z} - \mathbf{C}\mathbf{p} - \mathbf{D}\mathbf{q}\|_2^2 + \beta_q \sum_{h=1}^H (\|\mathbf{q}[h]\|_k)^l \right\} \quad (17)$$

It is clear that Equations (16) and (17) have the same structure and can be solved in a similar manner. In the following, we only need to find the optimal solution for Equation (16). Then, the minimization problem in Equation (17) can be handled in the same way.

The objective function in Equation (16) can be expressed as

$$F(\mathbf{p}) = f(\mathbf{p}) + v(\mathbf{p}) \quad (18)$$

where $f(\mathbf{p}) = \frac{1}{2} \|\mathbf{z} - \mathbf{C}\mathbf{p} - \mathbf{D}\mathbf{q}\|_2^2$ and $v(\mathbf{p}) = \beta_p \sum_{h=1}^H (\|\mathbf{p}[h]\|_k)^l$. Assume that $\mathbf{p}^{(k)}$ is obtained at the k th iteration. Then, by the quadratic approximation of $f(\mathbf{p})$ at the fixed point $\mathbf{p}^{(k)}$, $F(\mathbf{p})$ can be approximated as

$$F(\mathbf{p}) \approx f(\mathbf{p}^{(k)}) + (\mathbf{p} - \mathbf{p}^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{\sum_{h=1}^H \|\mathbf{p}[h] - \mathbf{p}^{(k)}[h]\|_2^2}{2\Sigma} + v(\mathbf{p}) \quad (19)$$

where $\nabla f(\mathbf{p}^{(k)}) = \mathbf{C}^H \cdot [\sum_{h=1}^H (\mathbf{C}[h]\mathbf{p}^{(k)}[h] + \mathbf{D}[h]\mathbf{q}[h]) - \mathbf{z}]$ is the partial derivation of $f(\mathbf{p}^{(k)})$ with respect to $\mathbf{p}^{(k)}$. Σ is set to be $\Sigma = \frac{\|\Delta\|_1}{L}$, where $\|\cdot\|_1$ denotes one norm for vectors, $\Delta = [\Delta_1, \Delta_2, \dots, \Delta_L]^T$ is a $L \times 1$ vector and $\Delta_l = \frac{1}{\|\hat{\mathbf{a}}^H(\tilde{\theta}_l) \otimes \hat{\mathbf{a}}(\tilde{\theta}_l)\|_2^2}$, $l = 1, 2, \dots, L$. Moreover, Δ can be shown to be a block vector, i.e.,

$$\Delta = [\Delta^T[1], \Delta^T[2], \dots, \Delta^T[H]]^T \quad (20)$$

where $\Delta[h]$ is the h th block of Δ with the length d . For the convenience of analysis, denote $\mathbf{J}^{(k)} = \text{diag}(\Delta) \nabla f(\mathbf{p}^{(k)})$, and we can also represent $\mathbf{J}^{(k)}$ as a block vector, i.e.,

$$\mathbf{J}^{(k)} = \text{diag}(\Delta) \mathbf{C}^H \cdot [\sum_{h=1}^H (\mathbf{C}[h]\mathbf{p}^{(k)}[h] + \mathbf{D}[h]\mathbf{q}[h]) - \mathbf{z}] = [(\mathbf{J}^{(k)}[1])^T, (\mathbf{J}^{(k)}[2])^T, \dots, (\mathbf{J}^{(k)}[H])^T]^T \quad (21)$$

where $\mathbf{J}^{(k)}[h]$ is the h th block of $\mathbf{J}^{(k)}$ with the length d . At the k th iteration, $\mathbf{p}^{(k)}$ is updated by minimizing $F(\mathbf{p})$ in Equation (19) so that the next iteration $\mathbf{p}^{(k+1)}$ is given by

$$\mathbf{p}^{(k+1)} = \min_{\mathbf{p}} \left\{ f(\mathbf{p}^{(k)}) + (\mathbf{p} - \mathbf{p}^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{\sum_{h=1}^H \|\mathbf{p}[h] - \mathbf{p}^{(k)}[h]\|_2^2}{2\Sigma} + v(\mathbf{p}) \right\} \quad (22)$$

By further derivation, Equation (22) can be simplified to

$$\mathbf{p}^{(k+1)} = \min_p \left\{ \sum_{h=1}^H \left[\frac{1}{2\Sigma} \left\| \mathbf{p}[h] - \mathbf{u}^{(k)}[h] \right\|_2^2 + \lambda_p \left\| \mathbf{p}[h] \right\|_k^l \right] \right\} \quad (23)$$

where $\mathbf{u}^{(k)}[h] = \mathbf{p}^{(k)}[h] - \mathbf{J}^{(k)}[h]$. As one may note, the objective function in Equation (23) is separable, and thus H blocks of $\mathbf{p}^{(k+1)}$ can be solved in a parallel manner. Although it is hard to solve H blocks, the classical BCD method has provided a critical inspiration, fortunately. The solution to the h th block of $\mathbf{p}^{(k+1)}$ is given by a soft-thresholding operator [22]

$$\mathbf{p}^{(k+1)}[h] = \frac{\mathbf{u}^{(k)}[h]}{\left\| \mathbf{u}^{(k)}[h] \right\|_2} \left(\left\| \mathbf{u}^{(k)}[h] \right\|_2 - \lambda_p \left\| \mathbf{A}[h] \right\|_2 \right) I \left(\left\| \mathbf{u}^{(k)}[h] \right\|_2 \geq \lambda_p \left\| \mathbf{A}[h] \right\|_2 \right), \quad h = 1, 2, \dots, H \quad (24)$$

where $I(\cdot)$ denotes an indicator function. Instead of reconstructing $\mathbf{p}^{(k+1)}$ directly, $\mathbf{p}^{(k+1)}$ can be reconstructed by H blocks, which may be zero vectors during the iteration. It can be seen in Equation (24) that we just need to determine the relation of $\left\| \mathbf{u}^{(k)}[h] \right\|_2$ and $\lambda_p \left\| \mathbf{A}[h] \right\|_2$ to judge whether $\mathbf{p}^{(k+1)}[h]$ is a zero vector. If $\left\| \mathbf{u}^{(k)}[h] \right\|_2$ is less than $\lambda_p \left\| \mathbf{A}[h] \right\|_2$, $\mathbf{p}^{(k+1)}[h]$ must be a zero vector. Therefore, a considerable amount of computations can be avoided in the process of solving H blocks. Similarly, by utilizing the soft-thresholding operator, the solution to the h th block of $\mathbf{q}^{(k+1)}$ is expressed as

$$\mathbf{q}^{(k+1)}[h] = \frac{\bar{\mathbf{u}}^{(k)}[h]}{\left\| \bar{\mathbf{u}}^{(k)}[h] \right\|_2} \left(\left\| \bar{\mathbf{u}}^{(k)}[h] \right\|_2 - \lambda_q \left\| \bar{\mathbf{A}}[h] \right\|_2 \right) I \left(\left\| \bar{\mathbf{u}}^{(k)}[h] \right\|_2 \geq \lambda_q \left\| \bar{\mathbf{A}}[h] \right\|_2 \right), \quad h = 1, 2, \dots, H \quad (25)$$

where $\bar{\mathbf{A}}[h]$ and $\bar{\mathbf{J}}^{(k)}[h]$ are the h th blocks of $\bar{\mathbf{A}}$ and $\bar{\mathbf{J}}$ with the length d , respectively, and $\bar{\mathbf{u}}^{(k)}[h] = \mathbf{q}^{(k)}[h] - \bar{\mathbf{J}}^{(k)}[h]$. Following the similar derivation, we have

$$\bar{\mathbf{J}}^{(k)} = \text{diag}(\bar{\mathbf{A}}) \mathbf{D}^H \cdot \left[\sum_{h=1}^H (\mathbf{C}[h] \mathbf{p}[h] + \mathbf{D}[h] \mathbf{q}^{(k)}[h]) - \mathbf{z} \right] \quad (26)$$

where $\bar{\mathbf{A}} = [\bar{\mathbf{A}}_1, \bar{\mathbf{A}}_2, \dots, \bar{\mathbf{A}}_L]^T$ is a $L \times 1$ vector and $\bar{\mathbf{A}}_l = \frac{1}{\left\| \hat{\mathbf{b}}^H(\tilde{\theta}_l) \otimes \hat{\mathbf{b}}(\tilde{\theta}_l) \right\|_2^2}$, $l = 1, 2, \dots, L$. Based on

Equations (24) and (25), $\mathbf{p}$ and $\mathbf{q}$ can be reconstructed alternately until the following criterion is satisfied

$$\left\| \frac{\boldsymbol{\eta}^{(k+1)} - \boldsymbol{\eta}^{(k)}}{\boldsymbol{\eta}^{(k)}} \right\| \leq \gamma \quad (27)$$

where $\boldsymbol{\eta}^{(k)} = [\mathbf{p}^{(k)}; \mathbf{q}^{(k)}]$ and γ is the small tolerance.

To reduce the computational complexity, a block selection criterion is given. This criterion is of great importance in the whole ABCD method. By utilizing the block selection criterion, we can only update the block that is the closest to $\mathbf{z}$, i.e.,

$$h_0 = \arg \min \left\| \mathbf{w}[h] - \mathbf{z} \right\|_2, \quad h = 1, 2, \dots, H \quad (28)$$

where $\mathbf{w}[h] = \mathbf{C}[h] \mathbf{p}[h] + \mathbf{D}[h] \mathbf{q}[h]$. This means that $\mathbf{p}^{(k+1)}[h_0]$ and $\mathbf{q}^{(k+1)}[h_0]$ at the k th iteration are updated as Equations (24) and (25) while the remaining blocks keep unchanged. The purpose of utilizing this block selection criterion is to avoid the update of repetitive and unnecessary blocks and

reduce the computational complexity. The major steps of reconstructing $\mathbf{p}$ and $\mathbf{q}$ by the ABCD method are given as follows:

Initialization: set $k = 0$, $\mathbf{p}^{(0)} = \mathbf{0}$, $\mathbf{q}^{(0)} = \mathbf{0}$, $\mathbf{C} \in \mathbb{C}^{N^2 \times L}$, $\mathbf{D} \in \mathbb{C}^{N^2 \times L}$, $\mathbf{z} = \text{vec}(\mathbf{R}_Y) \in \mathbb{C}^{N^2 \times 1}$ and $\gamma = 10^{-5}$.

- (1) Calculate $\mathbf{J}^{(k)}$ and $\bar{\mathbf{J}}^{(k)}$ in terms of Equations (21) and (26).
- (2) Due to H blocks of $\mathbf{J}^{(k)}$ and $\bar{\mathbf{J}}^{(k)}$, calculate $\mathbf{u}^{(k)}[h]$ and $\bar{\mathbf{u}}^{(k)}[h]$, $h = 1, 2, \dots, H$.
- (3) Calculate $\mathbf{p}^{(k+1)}[h]$ and $\mathbf{q}^{(k+1)}[h]$, $h = 1, 2, \dots, H$, in terms of (24) and (25).
- (4) Choose the block index h_0 according to (28). Then, $\mathbf{p}^{(k+1)}$ and $\mathbf{q}^{(k+1)}$ are respectively updated as $[\mathbf{p}^{(k)}[1], \dots, \mathbf{p}^{(k)}[h_0 - 1], \mathbf{p}^{(k+1)}[h_0], \mathbf{p}^{(k)}[h_0 + 1], \dots, \mathbf{p}^{(k)}[H]]^T$ and $[\mathbf{q}^{(k)}[1], \dots, \mathbf{q}^{(k)}[h_0 - 1], \mathbf{q}^{(k+1)}[h_0], \mathbf{q}^{(k)}[h_0 + 1], \dots, \mathbf{q}^{(k)}[H]]^T$.
- (5) If $\left\| \frac{\eta^{(k+1)} - \eta^{(k)}}{\eta^{(k)}} \right\| \leq \gamma$, stop the iteration. Otherwise, set $k = k + 1$ and return to step (1).

4. Global Convergence of the ABCD Method

The global convergence of the ABCD method is proved in this section. By combining the existing convergence proof of the general BCD framework [23] with ABCD method, a detailed derivation process for proving the global convergence is shown as follows.

First, we introduce the general BCD framework. Note that $f(\mathbf{p})$ in Equation (18) is a continuous convex function and $v(\mathbf{p})$ in (18) is a non-smooth convex function. Given a fixed point $\mathbf{p}^{(k)}$, $F(\mathbf{p})$ in Equation (18) can be approximated as the following form by exploiting the second order Taylor expansion of $f(\mathbf{p})$ in the general BCD framework.

$$F(\mathbf{p}) \approx f(\mathbf{p}^{(k)}) + (\mathbf{p} - \mathbf{p}^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\mathbf{p} - \mathbf{p}^{(k)})^H \mathbf{H} (\mathbf{p} - \mathbf{p}^{(k)}) + v(\mathbf{p}) \quad (29)$$

where $\mathbf{H} = \nabla^2 f(\mathbf{p}^{(k)}) = \mathbf{C}^H \mathbf{C}$ is the $L \times L$ Hessian matrix of $f(\mathbf{p}^{(k)})$ with respect to $\mathbf{p}^{(k)}$. It is clear that minimizing Equation (29) involves finding the next iteration $\mathbf{p}^{(k+1)}$. Assume that $\chi = [1, 2, \dots, H]$ is an H -dimensional index set and χ_0 is a subset of χ consisting of at most K indexes obtained by Equation (28). Hence, the next iteration $\mathbf{p}^{(k+1)}$ is represented as

$$\mathbf{p}^{(k+1)} = A(\mathbf{p}^{(k)}, \chi_0) = \arg \min_{\mathbf{p}} \{F(\mathbf{p}) / \mathbf{p}^{(k)}[h] = \mathbf{0}, h \in \chi, h \notin \chi_0\} \quad (30)$$

Then, to prove the global convergence, we give the modified Armijo rule and modified Gauss-Southwell- r rule that are prerequisites to guarantee the global convergence. These two rules are described in the following.

- (1) Modified Armijo rule: If $0 < \alpha^{(k)} < 1$, $0 < \xi < 1$ and $0 < \kappa < 1$, the following inequality holds:

$$F(\mathbf{p}^{(k)} + \alpha^{(k)} \mu^{(k)}) \leq F(\mathbf{p}^{(k)}) + \alpha^{(k)} \kappa \Gamma^{(k)} \quad (31)$$

where $\mu^{(k)} = \mathbf{p} - \mathbf{p}^{(k)}$ and $\Gamma^{(k)} = (\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \xi (\mu^{(k)})^H \mathbf{H} \mu^{(k)} + v(\mathbf{p}) - v(\mathbf{p}^{(k)})$.

- (2) Modified Gauss-Southwell- r rule: In the iteration, the index set χ_0 obtained by Equation (28) must satisfy

$$\|A(\mathbf{p}^{(k)}, \chi_0)\|_2 \leq \varsigma \|A(\mathbf{p}^{(k)}, \chi)\|_2 \quad (32)$$

where $0 < \varsigma < 1$.

It is well known that the problem of proving the global convergence is quite complex and intractable. However, fortunately, since Equations (16) and (17) are both convex and have both only one global point, we only need to show that Equations (16) and (17) satisfy the modified Armijo rule and modified Gauss-Southwell- r rule to prove the global convergence of the ABCD method. Furthermore, since Equation (16) has the same structure as Equation (17), it is enough to just prove that Equation (16) satisfies the modified Armijo rule and modified Gauss-Southwell- r rule. Regarding Equation (17), the derivation process is given in the same way.

To see the first, the following inequality holds according to Equation (30).

$$F(\mathbf{p}) = F(\mathbf{p}^{(k)} + \mu^{(k)}) \leq F(\mathbf{p}^{(k)} + \alpha^{(k)} \mu^{(k)}) \quad (33)$$

By substituting Equation (33) into Equation (29), we have

$$\begin{aligned} f(\mathbf{p}^{(k)}) + (\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) + v(\mathbf{p}^{(k)} + \mu^{(k)}) \leq \\ f(\mathbf{p}^{(k)}) + (\alpha^{(k)} \mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\alpha^{(k)} \mu^{(k)})^H \mathbf{H}(\alpha^{(k)} \mu^{(k)}) + v(\mathbf{p}^{(k)} + \alpha^{(k)} \mu^{(k)}) \end{aligned} \quad (34)$$

Based on the fact that $v(\mathbf{p}^{(k)} + \alpha^{(k)} \mu^{(k)}) \leq \alpha^{(k)} v(\mathbf{p}^{(k)} + \mu^{(k)}) + (1 - \alpha^{(k)}) v(\mathbf{p}^{(k)})$, Equation (34) can be further simplified to

$$(1 - \alpha^{(k)}) (\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} [1 - (\alpha^{(k)})^2] (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) + (1 - \alpha^{(k)}) [v(\mathbf{p}^{(k)} + \mu^{(k)}) - v(\mathbf{p}^{(k)})] \leq 0 \quad (35)$$

Since $0 < \alpha^{(k)} < 1$, by dividing by $1 - \alpha^{(k)}$ and setting $\alpha^{(k)} \rightarrow 1$, we have

$$(\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) + v(\mathbf{p}^{(k)} + \mu^{(k)}) - v(\mathbf{p}^{(k)}) \leq 0 \quad (36)$$

For $0 < \xi < 1$, it can be deduced from (36) that

$$\Gamma^{(k)} = (\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \xi (\mu^{(k)})^H \mathbf{H} \mu^{(k)} + v(\mathbf{p}) - v(\mathbf{p}^{(k)}) \leq (\xi - 1) (\mu^{(k)})^H \mathbf{H} \mu^{(k)} \leq 0 \quad (37)$$

Subsequently, by exploiting the convexity of $f(\mathbf{p})$, we obtain

$$f(\mathbf{p}) = f(\mathbf{p}^{(k)} + \mu^{(k)}) \leq f(\mathbf{p}^{(k)}) + (\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) \quad (38)$$

Since $F(\mathbf{p}) = f(\mathbf{p}) + v(\mathbf{p})$, it is nature to have

$$F(\mathbf{p}) = F(\mathbf{p}^{(k)} + \mu^{(k)}) \leq F(\mathbf{p}^{(k)}) + (\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) + v(\mathbf{p}) - v(\mathbf{p}^{(k)}) \quad (39)$$

Following the fact in Equation (37), the following inequality holds

$$\begin{aligned} F(\mathbf{p}^{(k)}) + (\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) + v(\mathbf{p}) - v(\mathbf{p}^{(k)}) \\ \leq F(\mathbf{p}^{(k)}) + \kappa [(\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) + v(\mathbf{p}) - v(\mathbf{p}^{(k)})] \end{aligned} \quad (40)$$

where $0 < \kappa < 1$. Combining Equations (39) and (40), we have

$$F(\mathbf{p}^{(k)} + \mu^{(k)}) \leq F(\mathbf{p}^{(k)}) + \kappa [(\mu^{(k)})^H \nabla f(\mathbf{p}^{(k)}) + \frac{1}{2} (\mu^{(k)})^H \mathbf{H}(\mu^{(k)}) + v(\mathbf{p}) - v(\mathbf{p}^{(k)})] \quad (41)$$

It is worth pointing out that Equation (41) is equal to Equation (31) with $\alpha^{(k)} = 1$ and $\xi = \frac{1}{2}$. Thus, it has been proved that Equation (16) satisfies the modified Armijo rule.

Secondly, in order to prove that Equation (16) satisfies the modified Gauss-Southwell- r rule, the following form is shown

$$h_0 = \arg \min \|\mathbf{w}[h] - \mathbf{z}\|_2 = \arg \min \|A(\mathbf{p}^{(k)}, \chi_i)\|_2 \quad (42)$$

where the cardinality of χ_i is the same as that of χ_0 . Without loss of generality, consider the worst case, *i.e.*, the cardinality of χ_i is the maximization K , and assume $H = KZ$ so that χ is expressed as $\chi = [\chi_1, \chi_2, \dots, \chi_Z]$. Thus, we have

$$\|A(\mathbf{p}^{(k)}, \chi_0)\|_2 \leq \|A(\mathbf{p}^{(k)}, \chi_i)\|_2 \quad (43)$$

Based on the following equation

$$\|A(\mathbf{p}^{(k)}, \chi)\|_2 = \sqrt{\sum_{z=1}^Z \|A(\mathbf{p}^{(k)}, \chi_z)\|_2^2} \quad (44)$$

it is easy to have

$$\|A(\mathbf{p}^{(k)}, \chi_0)\|_2 \leq \frac{1}{\sqrt{Z}} \|A(\mathbf{p}^{(k)}, \chi)\|_2 \quad (45)$$

Since Equation (45) is equal to Equation (32) with $\varsigma = \frac{1}{\sqrt{Z}}$, Equation (16) satisfies the modified Gauss-Southwell- r rule. Therefore, based on the above analysis, the global convergence of the ABCD method has been proved.

5. Simulation Results

This section presents several simulations to validate the superior performance of the proposed method as compared with R-GBCD+ and OGSBI-SVD. The angular space $[-90^\circ, 90^\circ]$ is taken the grid with grid interval $\tau = 3^\circ$ to perform three methods for off-grid targets. We set the length of block, the number of ULA sensors and spacing between adjacent sensors to be $d = 3$, $M = 9$ and $\hat{d} = \frac{\lambda}{2}$, respectively. In the simulation, the root mean squared error (RMSE) and success rate of DOA estimation are two significant performance indexes. RMSE is defined as

$$RMSE = \sqrt{\sum_{i=1}^{Mc} \sum_{k=1}^K \frac{(\bar{\theta}_{k,i} - \theta_k)^2}{K \cdot Mc}} \quad (46)$$

where Mc is the number of Monte Carlo runs and $\bar{\theta}_{k,i}$ is the estimate of θ_k in the i th Monte Carlo run, and success rate is declared if the estimation error is within a certain small Euclidean distance of the true directions.

In the first simulation, we compare the spatial spectra of R-GBCD+, OGSBI-SVD and ABCD. Consider four far-field narrowband sources impinging on the ULA from $[-30.4^\circ -3.8^\circ 10.1^\circ 15.3^\circ]$, where the latter two most closely spaced sources are coherent and the remaining sources are

independent of other sources. Figure 1 presents the spatial spectra of R-GBCD+, OGSBI-SVD and ABCD with SNR 3 dB and number of snapshots 100. For the convenience of analysis, the spatial spectra are normalized. We can see from Figure 1 that the spatial spectra of three methods are able to detect four sources, but the spatial spectrum obtained by R-GBCD+ has obvious bias at the true directions and OGSBI-SVD can yield slight bias in the vicinity of the coherent sources. Note that ABCD has a nearly ideal spatial spectrum, and thus it outperforms R-GBCD+ and OGSBI-SVD in terms of the spatial spectrum.

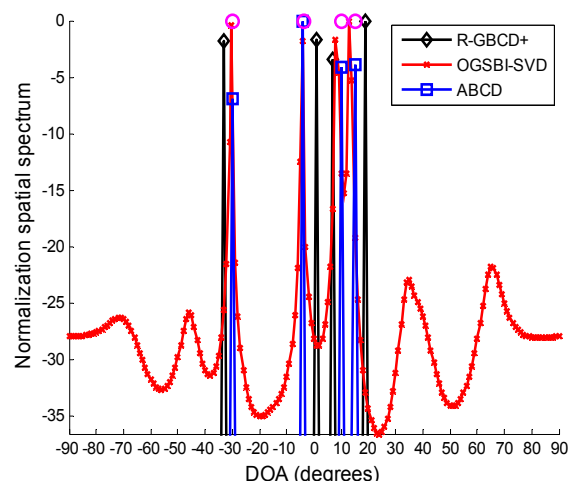


Figure 1. Spatial spectra of R-GBCD+, OGSBI-SVD and ABCD, where the pink circles denote the true DOAs.

The success rates of three methods *vs.* SNR and the number of snapshots are analyzed in the second simulation. The source mode is the same as the first simulation. Figure 2 shows the success rates of three methods *vs.* SNR with the fixed number of snapshots 100, whereas the success rates of three methods *vs.* number of snapshots are depicted with the fixed SNR 0 dB in Figure 3. The following facts can be acquired from Figures 2 and 3 that three methods can estimate correctly for high SNR or large number of snapshots and ABCD has a higher success rate than the other two methods for low SNR or a small number of snapshots.

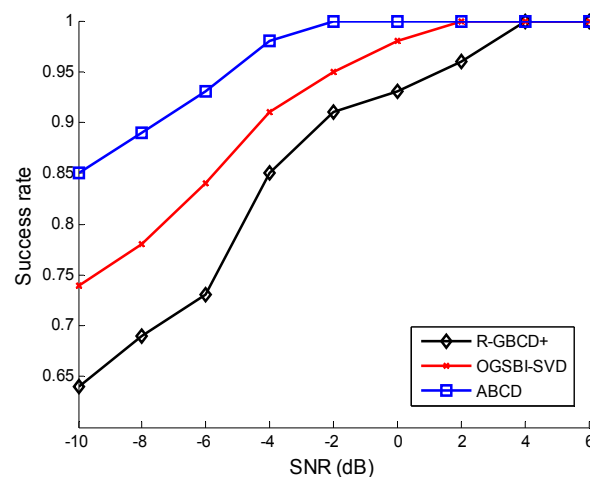


Figure 2. Success rates *vs.* SNR with the fixed number of snapshots 100.

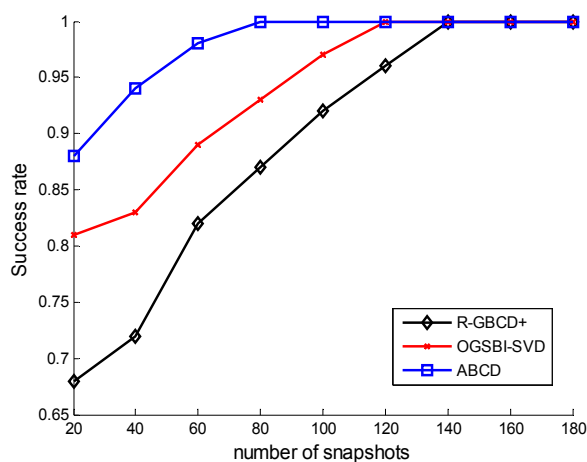


Figure 3. Success rates vs. number of snapshots with the fixed SNR 0 dB.

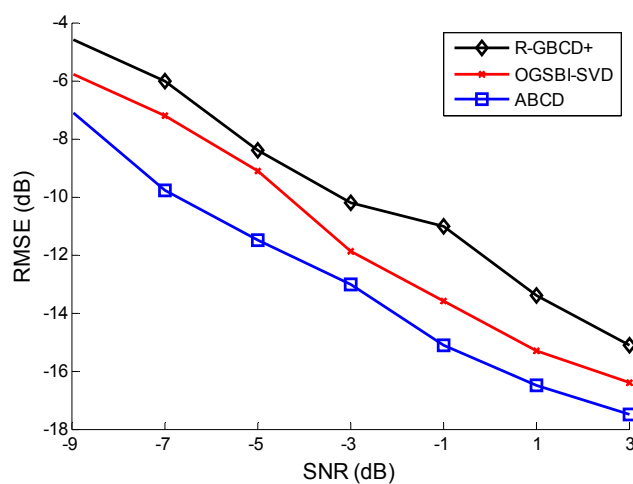


Figure 4. RMSE vs. SNR with the fixed number of snapshots 100.

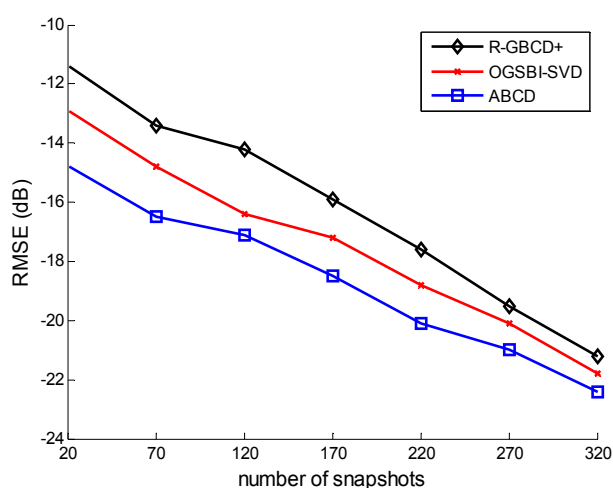


Figure 5. RMSE vs. number of snapshots with the fixed SNR 0 dB.

The third simulation considers the RMSE of three methods vs. SNR and the number of snapshots. All the conditions are the same as the second simulation. Figures 4 and 5 show the RMSE of three methods vs. SNR and the number of snapshots, respectively. It is indicated in Figures 4 and 5 that

ABCD has the best estimation accuracy among all three methods. Moreover, the accuracy of three methods is gradually improving with SNR or the number of snapshots increasing.

Finally, we test the resolving ability by showing the relation between RMSE of three methods and angle separation of sources, which is illustrated in Figure 6. Consider two coherent sources impinging on the ULA from 30.7° and $30.7 + \Delta\theta$, where the step of $\Delta\theta$ is 1° . The SNR is 0 dB and the number of snapshots is 100. As can be seen from Figure 6, the performance of R-GBCD+ and OGSBI-SVD degrades severely as angle separation is 3° , while ABCD can still provide a precise estimation as long as angle separation is no less than 3° . The proposed ABCD is the most accurate method and has higher resolution than the other two methods.

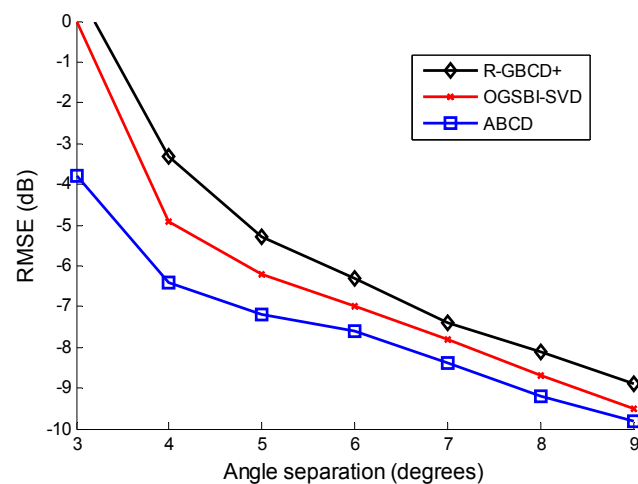


Figure 6. RMSE vs. angle separation with the fixed SNR 0 dB and number of snapshots 100 for coherent sources.

6. Conclusions

In this paper, a novel ABCD method is proposed for off-grid DOA estimation in CS. The proposed method minimizes the mixed k - l norm to reconstruct the sparse source and estimate the grid error. In order to make the minimization problem tractable, an iterative process that minimizes the mixed k - l norm alternately over two sparse vectors is adopted. By reconstructing the block sparse source instead of conventional sparse source, the proposed method can achieve the better reconstruction properties. A block selection criterion is given to update the block so that the proposed method can reduce computational complexity. It is proved that the proposed method has the global convergence. Simulation results show that the proposed method has more notable performance advantages than R-GBCD+ and OGSBI-SVD in terms of spatial spectrum, RMSE and success rate.

Acknowledgments

This work was supported by Aviation Science Foundation of China (201401P6001) and Fundamental Research Funds for the Central Universities (HEUCF150804).

Author Contributions

The main idea was proposed by Weijian Si and Xinggen Qu. Zhiyu Qu conceived the experiments and provided many valuable suggestions. Xinggen Qu wrote the manuscript and all the authors participated in amending the manuscript.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Reddy, V.V.; Mubeen, M.; Ng, B.P.; Reduced-complexity super-resolution DOA estimation with unknown number of sources. *IEEE Signal Process. Lett.* **2015**, *22*, 772–776.
2. Xie, J.; Tao, H.H.; Rao, X.; Su, J.; Passive localization of mixed far-field and near-field sources without estimation the number of sources. *Sensors* **2015**, *15*, 3834–3853.
3. Schmidt, R.O. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Trans. Antennas Propag.* **1986**, *34*, 276–280.
4. Roy, R.; Kailath, T.; ESPRIT-estimation of signal parameters via rotational invariance techniques. *IEEE Trans. Acoust. Speech Signal Process.* **1989**, *37*, 984–995.
5. Donoho, D.L. Compressed sensing. *IEEE Trans. Inf. Theory* **2006**, *52*, 1289–1306.
6. Candès, E.; Tao, T. Near-optimal signal recovery from random projections: Universal encoding strategies? *IEEE Trans. Inf. Theory* **2006**, *52*, 5406–5425.
7. Candès, E. The restricted isometry property and its implications for compressed sensing. *Comptes Rend. Math.* **2008**, *346*, 589–592.
8. Baraniuk, R.; Davenport, M.; DeVore, R.; Wakin, M. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.* **2008**, *28*, 253–263.
9. Malioutov, D.; Cetin, M.; Willsky, A.S. A sparse signal reconstruction perspective for source localization with sensor arrays. *IEEE Trans. Signal Process.* **2005**, *53*, 3010–3022.
10. Kim, J.M.; Lee, O.K.; Ye, J.C. Compressive MUSIC: revisiting the link between compressive sensing and array signal processing. *IEEE Trans. Inf. Theory* **2012**, *58*, 278–301.
11. Xu, X.; Wei, X.H.; Ye, Z.F. DOA estimation based on sparse signal recovery utilizing weighted l_1 -norm penalty. *IEEE Signal Process. Lett.* **2012**, *19*, 155–158.
12. Wei, X.H.; Yuan, Y.B.; Ling, Q. DOA estimation using a greedy block coordinate descent algorithm. *IEEE Trans. Signal Process.* **2012**, *60*, 6382–6394.
13. Zhu, H.; Leus, G.; Giannakis, G. Sparsity-cognizant total least-squares for perturbed compressive sampling. *IEEE Trans. Signal Process.* **2011**, *59*, 2002–2016.
14. Yang, Z.; Xie, L.; Zhang, C. Off-grid direction of arrival estimation using sparse Bayesian Inference. *IEEE Trans. Signal Process.* **2013**, *61*, 38–43.
15. Liang, Y.J.; Ying, R.D.; Lu, Z.Q.; Liu, P.L. Off-grid direction of arrival estimation based on joint spatial sparsity for distributed sparse linear arrays. *Sensors* **2014**, *14*, 21981–22000.
16. Zhang, Y.; Ye, Z.F.; Xu, X.; Hu, N. Off-grid DOA estimation using array covariance matrix and block-sparse Bayesian learning. *Signal Process.* **2014**, *98*, 197–201.

17. Qin, Z.W.; Scheinberg, K.; Goldfarb, D. Efficient block-coordinate descent algorithms for the Group Lasso. *Math. Program. Comput.* **2013**, *5*, 143–169.
18. Eldar, Y.C.; Mishali, M. Robust recovery of signals from a structured union of subspaces. *IEEE Trans. Inf. Theory* **2009**, *55*, 5302–5316.
19. Mishali, M.; Eldar, Y.C. Blind multi-band signal reconstruction: Compressed sensing for analog signals. *IEEE Trans. Signal Process.* **2009**, *57*, 993–1009.
20. Eldar, Y.C.; Kuppinger, P.; Bölcskei, H. Block-sparse signals: uncertainty relations and efficient recovery. *IEEE Trans. Signal Process.* **2010**, *58*, 3042–3054.
21. Zhao, G.H.; Shi, G.M.; Shen, F.F.; Luo, X.; Niu, Y. A sparse representation based DOA estimation algorithm with separable observation model. *IEEE Trans. Antennas Wirel. Propag. Lett.* **2015**, doi:10.1109/LAWP.2015.2413814.
22. Beck, A.; Teboulle, M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.* **2009**, *2*, 183–202.
23. Tseng, P.; Yun, S. A coordinate gradient descent method for nonsmooth separable minimization. *Math. Program.* **2009**, *117*, 387–423.

© 2015 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/4.0/>).