Off-grid DOA Estimation with Nonconvex Regularization via Joint Sparse Representation

Qi Liu[a,*], Hing Cheung So[a,1], Yuantao Gu[b]

[a] *Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China*
[b] *Department of Electronic Engineering and Tsinghua National Laboratory for Information Science and Technology (TNList), Tsinghua University, Beijing, China*

Abstract

In this paper, we address the problem of direction-of-arrival (DOA) estimation using sparse representation. As the performance of on-grid DOA estimation methods will degrade when the unknown DOAs are not on the angular grids, we consider the off-grid model via Taylor series expansion, but dictionary mismatch is introduced. The resulting problem is nonconvex with respect to the sparse signal and perturbation matrix. We develop a novel objective function regularized by the nonconvex sparsity-inducing penalty for off-grid DOA estimation, which is jointly convex with respect to sparse signal and perturbation matrix. Then alternating minimization is applied to tackle this joint sparse representation of the signal recovery and perturbation matrix. Numerical examples are conducted to verify the effectiveness of the proposed method, which achieves more accurate DOA estimation performance and faster implementation than the conventional sparsity-aware and state-of-the-art off-grid schemes.

Keywords: DOA estimation, off-grid model, sparse representation, nonconvex regularization

---

1. Introduction

Direction-of-arrival (DOA) estimation has been extensively studied over the past few decades because of its fundamental role in many signal processing areas ranging from multiple-input multiple-output radar, mobile and wireless communications, channel estimation and sonar to acoustic tracking [1–3].

Recently, sparse representation has attracted increasing interest in statistical signal analysis and parameter estimation. In [4], the concept of sparse representation is extended to address the problem of DOA estimation problem and ℓ_1-SVD algorithm is proposed to reduce the dimension of observations via singular value decomposition (SVD), which can achieve super-resolution performance. A reweighted ℓ_1 norm penalty algorithm [5] exploits the coefficients of the reduced dimension Capon spatial spectrum in constructing the weight matrix to enforce the sparsity of solution, which involves a high computational burden. The methods mentioned above have shown improvements in DOA estimation, but most of them are based on on-grid DOA ℓ_1 norm constrained minimization. Since in practice the unknown DOAs are not always exactly on the sampling grids, their DOA estimation performance will degrade due to errors caused by the mismatches.

[*] Corresponding author
Email addresses: lq_skyven@126.com (Qi Liu), hcso@ee.cityu.edu.hk (Hing Cheung So), gyt@tsinghua.edu.cn (Yuantao Gu)
[1] EURASIP Member

To circumvent this issue, off-grid DOA estimation methods have been developed [6–12]. In [6], a gridless sparse approach via reweighted atomic norm minimization is proposed for off-grid DOA estimation. In [7, 8], alternating minimization is exploited to solve for sparse signal and dictionary mismatch simultaneously, but it suffers from slow convergence. A noise subspace fitting-based off-grid DOA estimation method is derived in [9] using second-order Taylor approximation to achieve higher modeling accuracy. In [10], an analytical performance bound on joint sparse recovery is given and a fast iterative shrinkage-threshold algorithm is implemented to tackle joint sparse recovery with structured dictionary mismatches. In [11], co-prime arrays are considered to increase degrees of the freedom for the grid mismatch and sample covariance matrix is utilized to reduce the effect of noise variance. In [12], a computationally efficient root sparse Bayesian learning (RSBL) method is proposed to eliminate the modeling error when using coarse grid.

Compared with the convex function regularized by least squares problem, it has been demonstrated that utilizing nonconvex functions, such as smoothed ℓ_0 quasi-norm [13], ℓ_p quasi-norm [14] and weak convexity [15], can achieve better sparse signal recovery. In this paper, we develop a novel objective function regularized by the nonconvex sparsity-inducing penalty for off-grid DOA estimation. Our motivation is twofold: (i) to overcome the limitation of the conventional sparsity-based DOA estimation methods that the unknown angles belong to predefined discrete angular grids; and (ii) a proper nonconvex regularization is able to achieve better performance compared with convex relaxation with ℓ_1 norm function. In this study, we first introduce the off-grid model into DOA estimation via first-order Taylor series expansion, which is equivalent to the dictionary mismatch, and then devise an objective function regularized by the nonconvex sparsity-inducing penalty with the least absolute shrinkage and selection operator (LASSO)[16]. The resulting objective function is jointly convex with respect to the sparse signal and perturbation matrix. We follow the rationale of alternating minimization to obtain the sparse signal by alternating direction method of multipliers (ADMM) [17] with incorporating the proximity operator for a fixed perturbation matrix, then update perturbation matrix via fixing the sparse signal and so on. Our results demonstrate that the proposed method outperforms the conventional sparsity-aware and state-of-the-art off-grid schemes.

The rest of this paper is organized as follows. In Section 2, the problem of DOA estimation using sparse representation is formulated. Section 3 introduces the off-grid model and presents our DOA estimation method. In Section 4, numerical examples are conducted to evaluate the performance of the proposed algorithm. Section 5 concludes this paper.

Notation: Lowercase bold-face and uppercase bold-face letters represent vectors and matrices, respectively. $(\cdot)^\dagger$, $(\cdot)^T$ and $(\cdot)^H$ are pseudo-inverse, transpose and conjugate transpose operators, respectively. $\text{vec}(\cdot)$ denotes the vectorization operator which stacks a matrix column by column. $\text{diag}(\cdot)$ is a diagonal matrix composed of the elements of a column vector. $\otimes$ denotes the Kronecker product operator. $\|\cdot\|_1$, $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the ℓ_1 norm, ℓ_2 norm and Frobenious norm, respectively. $\Re$ and $\Im$ take the real and imaginary parts of a complex variable, respectively. $\mathbf{I}_K$ denotes the $K \times K$ identity matrix.

2. Problem Statement

2.1. Signal Model

Consider a uniform linear array (ULA) equipped with M sensors. The inter-element spacing is half-wavelength. The origin is set at the middle point of the ULA. Assume that K narrowband signals from the far-field impinge onto the ULA from unknown and distinct angles of $\theta_1, ..., \theta_K$. The ULA response at the kth target can be expressed as

$$\mathbf{a}(\theta_k) = [e^{-j\pi\frac{(M-1)}{2}\cos(\theta_k)}, ..., e^{j\pi\frac{(M-1)}{2}\cos(\theta_k)}]^T \tag{1}$$

The $M \times 1$ observation vector is:

$$\mathbf{y}_t = \mathbf{A}(\theta)\mathbf{s}_t + \mathbf{n}_t, \qquad t = 1, ..., T \tag{2}$$

where $\mathbf{y}_t = [y_1(t), ..., y_M(t)]^T$, $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), ..., \mathbf{a}(\theta_K)]$ is the array steering matrix, $\mathbf{s}_t = [s_1(t), ..., s_K(t)]^T$ contains the source signal amplitudes, $\mathbf{n}_t = [n_1(t), ..., n_M(t)]^T$ is the complex independent white Gaussian noise vector with zero mean and covariance $\sigma^2\mathbf{I}_M$. Here, T is the number of snapshots, and $y_m(t)$ and $n_m(t)$, $m = 1, ..., M$, are the output and measurement noise of the mth sensor at time t, respectively.

Collecting the T snapshots, the matrix form of (2) can be formulated as a multiple measurement vectors (MMV) model, given by

$$\mathbf{Y} = \mathbf{A}(\theta)\mathbf{S} + \mathbf{N} \tag{3}$$

where $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_T] \in \mathbb{C}^{M\times T}$, $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_T] \in \mathbb{C}^{K\times T}$ and $\mathbf{N} = [\mathbf{n}_1, ..., \mathbf{n}_T] \in \mathbb{C}^{M\times T}$.

In our study, we assume that K is known *a priori* and employ the $M \times K$ measurement matrix $\mathbf{Y_{sv}}$ by thresholding the K largest singular values of the $M \times T$ measurement matrix $\mathbf{Y}$ to reduce computational complexity in directly processing (3), which is analogous to the ℓ_1-SVD algorithm [4]. In summary, the problem of DOA estimation in sparse representation framework is to find the unknown DOAs given K, $\mathbf{Y_{sv}}$ and the mapping $\theta \to \mathbf{A}(\theta)$.

2.2. DOA Estimation in Sparse Representation Framework

Let the set $\mathbf{\Phi} = \{\hat{\theta}_1, ..., \hat{\theta}_N\}$ be the discretized sampling grids of all potential directions in the admissible DOA range $[0, \pi]$, where N is the number of grid points and typically $N \gg M > K$. When the true DOAs are located at (or close to) the sampling grids, the typical DOA estimation model based on the sparse representation framework is linear:

$$\mathbf{Y_{sv}} = \mathbf{A}(\hat{\theta})\hat{\mathbf{S}} + \hat{\mathbf{N}} \tag{4}$$

where $\hat{\mathbf{S}} \in \mathbb{C}^{N\times K}$ is the sparse signal matrix and $\mathbf{A}(\hat{\theta}) = [\mathbf{a}(\hat{\theta}_1), ..., a(\hat{\theta}_N)] \in \mathbb{C}^{M\times N}$. The K rows in $\hat{\mathbf{S}}$ with largest magnitudes are identical to that of $\mathbf{S}$, and the remaining $N - K$ rows in $\hat{\mathbf{S}}$ are regarded as zero. In compressed sensing theory, the main task in (4) is to recover $\hat{\mathbf{S}}$ from the underdetermined system, and DOA estimation is equivalent to finding the positions of K nonzero rows in $\hat{\mathbf{S}}$. The sparse signal recovery can be formulated as the ℓ_0 norm constrained minimization problem:

$$(\ell_0): \quad \min_{\hat{\mathbf{S}}} \ \| \hat{\mathbf{S}} \|_{\text{row},0} \quad \text{s.t.} \quad \mathbf{Y_{sv}} = \mathbf{A}(\hat{\theta})\hat{\mathbf{S}} + \hat{\mathbf{N}} \tag{5}$$

where $\|\cdot\|_{\text{row},0}$ counts the nonzero rows.

Since ℓ_0 norm function is highly discontinuous and nonconvex, solving the ℓ_0 norm constrained minimization problem is known to be NP-hard in general. To address this issue, the ℓ_1 norm, which is the closest convex norm to the ℓ_0 norm, is employed instead. Then the sparse signal recovery problem under the ℓ_1 norm function is:

$$(\ell_1): \quad \min_{\hat{\mathbf{S}}} \ \| \hat{\mathbf{s}}^{\boldsymbol{\ell_2}} \|_1 \quad \text{s.t.} \quad \| \mathbf{Y_{sv}} - \mathbf{A}(\hat{\theta})\hat{\mathbf{S}} \|_F^2 \leq \eta \tag{6}$$

where η is an upper-bound on the noise power, and $\hat{\mathbf{s}}^{\boldsymbol{\ell_2}}$ is a function of $\hat{\mathbf{S}}$ whose the ith element equals the Frobenius norm of the ith row of $\hat{\mathbf{S}}$, i.e., $[\hat{\mathbf{s}}^{\boldsymbol{\ell_2}}]_i = \| \hat{\mathbf{S}}(i,:) \|_2$. Numerical methods [4, 18] have been presented for (6). However, larger coefficients are penalized more heavily in ℓ_1 norm than smaller coefficients, which results to that the sparsest solution of ℓ_1 norm penalty does not approximate the ℓ_0 norm penalty. Nevertheless, reweighted ℓ_1 norm minimization algorithms are designed to tackle this imbalance in (6):

$$(\mathrm{W}\ell_1): \quad \min_{\hat{\mathbf{S}}} \ \| \mathbf{W}(\hat{\mathbf{s}})^{\boldsymbol{\ell_2}} \|_1 \quad \text{s.t.} \quad \| \mathbf{Y_{sv}} - \mathbf{A}(\hat{\theta})\hat{\mathbf{S}} \|_F^2 \leq \eta \tag{7}$$

where W is a weighting matrix and has different forms according to different optimization criteria [19, 20].

To this end, there are two main drawbacks of the DOA estimation methods based on ℓ_1 norm minimization: (i) they recover the DOAs only if the targets exactly correspond to the discretized sampling grids. However, the target positions are not precisely on the grids in practical scenarios and thus DOA estimation bias exists. Moreover, most conventional sparsity-based DOA estimation methods tackle this problem by using dense sampling grids, which lead to highly computational complexity and the estimated DOAs are still constrained on the grids; (ii) they apply toolbox to calculate the ℓ_1 norm constrained minimization problem, such as CVX [21] and Sedumi [22], which cannot tackle the nonconvex optimization problem and is time-consuming, especially for large data size.

3. Algorithm Development

3.1. Off-grid Model

In real scenario, no matter how fine the grid points are, DOAs are almost not located exactly on the discretized sampling grids, which is regarded as the off-grid problem. To address this, off-grid DOA model has been suggested and there are two main ideas. The first applies atomic norm directly on the continuous parameter space for gridless DOA estimation [6, 23], while the second models the off-grid DOA via Taylor series expansion and then handles the resulting dictionary mismatch [7, 10–12]. Note that our development is based on the latter. Here, DOA is decomposed into two parts, namely an integer part of DOA on the grid and a fraction part to complement the on-grid model. Suppose $\theta_k \notin \{\hat{\theta}_1, ..., \hat{\theta}_N\}$ for some $k \in \{1, ..., K\}$. In such a case, DOA θ_k can be rewritten as $\theta_k = \hat{\theta}_k + \delta_k$, where $\hat{\theta}_k$ denotes the nearest grid to θ_k and δ_k is the grid offset.

According to the trigonometric identities, the term $\cos(\hat{\theta}_k + \delta_k)$ can be decomposed as

$$\begin{aligned} \cos(\hat{\theta}_k + \delta_k) &= \cos(\hat{\theta}_k)\cos(\delta_k) - \sin(\hat{\theta}_k)\sin(\delta_k) \\ &\approx \cos(\hat{\theta}_k) - \delta_k \sin(\hat{\theta}_k) \end{aligned} \tag{8}$$

By utilizing first-order Taylor series expansion, the steering vector for the off-grid DOA model is given by

$$\mathbf{a}(\theta_k) \approx \mathbf{a}(\hat{\theta}_k) + \mathbf{b}(\hat{\theta}_k)(\theta_k - \hat{\theta}_k) \tag{9}$$

where $\mathbf{b}(\hat{\theta}_k)$ is the first derivative of $\mathbf{a}(\hat{\theta}_k)$ with respect to $\hat{\theta}_k$. Then the dictionary matrix based on the off-grid DOA model can be corrected as

$$\mathbf{A}(\theta) \approx \mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\mathbf{\Delta} \tag{10}$$

where $\mathbf{\Delta} = \text{diag}(\boldsymbol{\delta})$ with $\boldsymbol{\delta} = [\delta_1, ..., \delta_N]^T$ denotes the perturbation matrix.

As a result, the off-grid DOA estimation based on sparse representation framework is formulated as:

$$\mathbf{Y_{sv}} = [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\mathbf{\Delta}]\hat{\mathbf{S}} + \hat{\mathbf{N}} \tag{11}$$

It is well known that a sparse solution can be obtained by solving a least squares problem with ℓ_1 norm regularization, which is known as the LASSO: $\min_{\hat{\mathbf{S}}} \parallel \mathbf{Y_{sv}} - [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\mathbf{\Delta}]\hat{\mathbf{S}} \parallel_F^2 + \tau \parallel \hat{\mathbf{S}}^{\ell_2} \parallel_1$, where $\tau > 0$ is the trade-off parameter. However, the resulting problem is difficult to solve due to the fact that it is nonconvex with respect to $\hat{\mathbf{S}}$ and $\mathbf{\Delta}$. Therefore, it cannot be directly handled by convex optimization toolboox. Most recently published works [12, 24, 25] tackle it from a sparse Bayesian inference perspective where the Laplace prior is exploited for the signal of interest, which involve high computational complexity.

3.2. Off-grid DOA Estimation with Nonconvex Regularization

To overcome the limitation of the conventional sparsity-based DOA estimation methods, off-grid DOA model is considered, but it is a challenging task because of the presence of the mismatches. We introduce the sparse regularized least squares (SRLS) with ℓ_2 norm to mitigate the mismatches. As a result, sparse signal is obtained by combining the LASSO with the SRLS:

$$\min_{\hat{\mathbf{S}}, \mathbf{\Delta}} \tau \parallel \hat{\mathbf{s}}^{\ell_2} \parallel_1 + \parallel \mathbf{\Delta} \parallel_F^2 + \parallel \mathbf{Y_{sv}} - [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\mathbf{\Delta}]\hat{\mathbf{S}} \parallel_F^2 \tag{12}$$

It has been demonstrated in [15, 26] that the sparsity pattern can be better induced over the ℓ_1 penalty conunterpart, with a proper nonconvex penalty. As far as we know, the SRLS approach has not yet been studied in combination with nonconvex penalty for an underdetermined system. Via adding a nonconvex function $J(\cdot)$, we devise an objective function regularized by the nonconvex sparsity-inducing penalty for off-grid DOA estimation:

$$\min_{\hat{\mathbf{S}}, \mathbf{\Delta}} \lambda J(\hat{\mathbf{s}}_{\text{g}}) + \parallel \mathbf{\Delta} \parallel_F^2 + \parallel \mathbf{Y_{sv}} - [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\mathbf{\Delta}]\hat{\mathbf{S}} \parallel_F^2 \tag{13}$$

in which $\hat{\mathbf{s}}_g = \hat{\mathbf{s}}^{\ell_2}$ and $\lambda > 0$ is the regularization parameter while the sparsity-inducing penalty is defined as

$$J(\hat{\mathbf{s}}_g) = \sum_{i=1}^{N} F(\hat{s}_{g_i}) \tag{14}$$

where $F : \mathbb{C} \to \mathbb{C}^+$ is a weakly convex sparseness function satisfies:

Definition 1.

(a) $F(0) = 0, F(\cdot)$ is even and not identically zero;

(b) $F(\cdot)$ is nondecreasing on $[0, +\infty)$;

(c) The function $\hat{s}_g \to F(\hat{s}_g)/\hat{s}_g$ is nonincreasing on $[0, +\infty)$;

(d) $F(\cdot)$ is weakly convex on $[0, +\infty)$;

The concept of weak convexity is proposed in [27]. Basically, $F(\hat{s}_g)$ is weakly convex if and only if there exists a convex function $H(\hat{s}_g) = F(\hat{s}_g) - \xi \hat{s}_g^2$ when $\xi < 0$. From Lemma 1.1 in [26], $F(\hat{s}_g)/\hat{s}_g \to \alpha$ as $\hat{s}_g \to 0^+$ for $\alpha > 0$. Hence, the nonconvexity of $F(\cdot)$ and $J(\cdot)$ can be defined as $\beta \triangleq -\xi/\alpha$ according to Definition 1 and (14).

Most functions satisfying Definition 1 can be found in Table I of [26]. For example, the weakly convex sparseness function in (14) may be chosen as

$$F(\hat{s}_g) = (\mid \hat{s}_g \mid -\beta \hat{s}_g^2)\mathbf{1}_{|\hat{s}_g| \leq \frac{1}{2\beta}}(\hat{s}_g) + \frac{1}{4\beta}\mathbf{1}_{|\hat{s}_g| > \frac{1}{2\beta}}(\hat{s}_g) \tag{15}$$

where $\mathbf{1}_P(\cdot)$ is the indicator function with value 1 when the argument satisfying P, and 0 otherwise. $F(\cdot)$ in (15) is a continuous piecewise quadratic function.

Now, the task is to estimate $\hat{\theta}$ and δ in each iteration. To solve the nonconvex optimization problem in (13), we apply alternating optimization via minimizing with respect to one variable at each time. To be specific, we first update $\hat{\mathbf{S}}$ by keeping the unknown variable $\boldsymbol{\Delta}$ fixed, and then we do the same for $\boldsymbol{\Delta}$.

By fixing $\boldsymbol{\Delta}$, the joint sparse representation problem in (13) reduces to solving the MMV sparse recovery problem:

$$\min_{\hat{\mathbf{S}}} \lambda J(\hat{\mathbf{s}}_{\mathrm{g}}) + \parallel \mathbf{Y_{sv}} - [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\boldsymbol{\Delta}]\hat{\mathbf{S}} \parallel_F^2 \tag{16}$$

which can be solved using convex optimization toolbox, such as SeDuMi and CVX. Nevertheless, it is time consuming especially when the numbers of sensors and targets are large. For a more efficient implementation, we apply variable splitting and introduce the auxiliary variable $\mathbf{z}$. Then we reformulate (16) as:

$$\min_{\mathbf{z}} \lambda J(\mathbf{z}) + \parallel \tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}} \parallel_2^2 \qquad \text{s.t. } \mathbf{z} = \tilde{\mathbf{s}} \tag{17}$$

where $\tilde{\mathbf{y}} = \mathrm{vec}(\mathbf{Y}_{\mathrm{sv}}) = [\mathbf{y}_{\mathrm{sv}1}^T, \mathbf{y}_{\mathrm{sv}2}^T, ..., \mathbf{y}_{\mathrm{sv}K}^T]^T$, $\tilde{\mathbf{s}} = \mathrm{vec}(\hat{\mathbf{S}}) = [\hat{\mathbf{s}}_1^T, \hat{\mathbf{s}}_2^T, ..., \hat{\mathbf{s}}_K^T]^T$, $\hat{\mathbf{s}}_g = \sqrt{\sum_{k=1}^K (\hat{\mathbf{s}}(k))^2}$ and $\tilde{\mathbf{A}} = \mathbf{I}_K \otimes [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\boldsymbol{\Delta}]$.

ADMM blends the decomposability of dual ascent with the superior convergence property of the multiplier method. We exploit this property with incorporating the proximity operator of weakly function into the framework of augmented Lagrangian to solve (17) such that each iterative step corresponds to a convex optimization.

The augmented Lagrangian is:

$$L(\tilde{\mathbf{s}}, \mathbf{z}, \boldsymbol{\mu}) = \lambda J(\mathbf{z}) + \parallel \tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}} \parallel_2^2 + \boldsymbol{\mu}^T(\mathbf{z} - \tilde{\mathbf{s}}) + \frac{\gamma}{2} \parallel \mathbf{z} - \tilde{\mathbf{s}} \parallel_2^2 \tag{18}$$

where $\gamma > 0$ is a penalty parameter which controls the convergence rate of the algorithm.

Based on the decomposition-coordination procedure of the ADMM, we determine $\{\tilde{\mathbf{s}}, \mathbf{z}, \boldsymbol{\mu}\}$ from (18) via the following steps:

1) With the obtained $\{\tilde{\mathbf{s}}^t, \boldsymbol{\mu}^t\}$ at the tth iteration, the update of $\mathbf{z}^{t+1}$ at the $(t+1)$th iteration is

$$\begin{aligned} \mathbf{z}^{t+1} &= \arg\min_{\mathbf{z}} \ L(\tilde{\mathbf{s}}^t, \mathbf{z}, \boldsymbol{\mu}^t) \\ &= \text{prox}_{\frac{\lambda}{\gamma} J(\cdot)} \left(\tilde{\mathbf{s}}^t + \frac{\boldsymbol{\mu}^t}{\gamma} \right) \end{aligned} \tag{19}$$

where $\text{prox}_{\frac{\lambda}{\gamma} J(\cdot)}(v)$ denotes the proximal operator [28] of $J(\cdot)$. When $\beta < 1/(2\zeta)$,

$$\text{prox}_{\zeta F}(v) = \frac{v - \zeta \text{sign}(v)}{1 - 2\zeta\beta} \mathbf{1}_{\zeta \le |v| \le \frac{1}{2\beta}}(v) + v \mathbf{1}_{|v| > \frac{1}{2\beta}}(v) \tag{20}$$

2) The update of $\tilde{\mathbf{s}}$ is

$$\begin{aligned} \tilde{\mathbf{s}}^{t+1} &= \arg\min_{\tilde{\mathbf{s}}} \ L(\tilde{\mathbf{s}}, \mathbf{z}^t, \boldsymbol{\mu}^t) \\ &= \Pi_{\mathcal{C}} \left(\mathbf{z} - \frac{\boldsymbol{\mu}^t}{\gamma} \right) \end{aligned} \tag{21}$$

where $\Pi_{\mathcal{C}}(\cdot)$ is the Euclidean projection onto $\mathcal{C} = \{\tilde{\mathbf{s}} : \| \tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}} \|_2^2 \le \eta\}$. That is,

$$\Pi_{\mathcal{C}}(\tilde{\mathbf{s}}) = \begin{cases} \tilde{\mathbf{s}}, & \tilde{\mathbf{s}} \in \mathcal{C} \\ \dfrac{\eta}{\| \tilde{\mathbf{s}} \|_2^2} \cdot \tilde{\mathbf{s}}, & \text{otherwise.} \end{cases} \tag{22}$$

In computing (21), Cholesky decomposition can be utilized to improve the computational speed.

3) The update of $\boldsymbol{\mu}$ is

$$\boldsymbol{\mu}^{t+1} = \boldsymbol{\mu}^t - \gamma \left(\mathbf{z}^t - \tilde{\mathbf{y}} + \tilde{\mathbf{A}}\tilde{\mathbf{s}}^t \right) \tag{23}$$

When the variable $\hat{\mathbf{S}}$ has been updated, we minimize over $\boldsymbol{\Delta}$ while keeping the current estimate of $\hat{\mathbf{S}}$ fixed. Then, the problem of off-grid DOA estimation in (13) reduces to:

$$\begin{aligned} \boldsymbol{\Delta}^{t+1} &= \arg\min_{\boldsymbol{\Delta}} \ \| \boldsymbol{\Delta} \|_F^2 + \| \mathbf{Y}_{\mathbf{sv}} - [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\boldsymbol{\Delta}]\hat{\mathbf{S}} \|_F^2 \\ &= \arg\min_{\boldsymbol{\Delta}} \ \| \boldsymbol{\Delta} \|_F^2 + \| \tilde{\mathbf{y}} - \tilde{\mathbf{A}}\tilde{\mathbf{s}} \|_2^2 \end{aligned} \tag{24}$$

Since the quadratic problem in (24) is convex with respect to $\boldsymbol{\Delta}$, it is formulated as a SRLS problem equivalently. The optimal solution to the quadratic problem (24) has a closed-form of:

$$\boldsymbol{\Delta}^{t+1} = \tilde{\mathbf{B}}^{\dagger}[\tilde{\mathbf{y}} - (\mathbf{I}_K \otimes \mathbf{A}(\hat{\theta}))\tilde{\mathbf{s}}^t] \tag{25}$$

where $\tilde{\mathbf{B}} = [\mathbf{B}(\hat{\theta})\text{diag}(\hat{\mathbf{s}}_1^t), \mathbf{B}(\hat{\theta})\text{diag}(\hat{\mathbf{s}}_2^t), ..., \mathbf{B}(\hat{\theta})\text{diag}(\hat{\mathbf{s}}_K^t)]^T$.

This completes one update cycle and the algorithm will terminate once the difference between two consecutive iterations is smaller than a given threshold or if the maximum iteration number is reached.

As for the convergence of the problem in (13), the following result is established.

Theorem 1. For arbitrary starting point, the sequence $\{(\hat{\mathbf{S}}^t, \boldsymbol{\Delta}^t)\}$ generated by our algorithm converges at least to a stationary point of (13).

Proof: In fact, the proposed method utilizing the rationale of alternating optimization suggests that it is the special case of the block coordinate descent algorithm. $\{(\hat{\mathbf{S}}^t, \boldsymbol{\Delta}^t)\} = \arg\ \min\{\| \mathbf{Y}_{\mathbf{sv}} - [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\boldsymbol{\Delta}]\hat{\mathbf{S}} \|_F^2 + \| \boldsymbol{\Delta} \|_F^2 + \lambda J(\hat{\mathbf{s}}_g)\}$. The first two terms of the objective function are differentiable with respect to the corresponding variables, while the remaining term (i.e., the sparsity-inducing penalty) is separable in the entries of $\hat{\mathbf{s}}_g$. $F(\hat{s}_g)$ is continuous and there exists $\alpha > 0$ such that $F(\hat{s}_g) \leq \alpha|\hat{s}_g|$ holds for all $\hat{s}_g \in \mathbb{R}$, which is demonstrated in Section VI-A of [26]. Therefore, the cost function in (13) satisfies those technical assumptions (B1)-(B3) and (C2) in [29]. Moreover, the first term $\| \mathbf{Y_{sv}} - [\mathbf{A}(\hat{\theta}) + \mathbf{B}(\hat{\theta})\mathbf{\Delta}]\hat{\mathbf{S}} \|_F^2$ is Gateaux-differentiable over its open domain. According to Lemma 3.1 in [29], the cost function in (13) is regular at each coordinatewise minimum point. Assuming that the sequence $\{(\hat{\mathbf{S}}^t, \mathbf{\Delta}^t)\}$ utilizing the essential cyclic is defined, each coordinatewise minimum point becomes a stationary point according to Proposition 5.1 and Theorem 5.1 in [29].

4. Numerical Examples

In this section, we present numerical examples for DOA estimation to show the advantages of the proposed method, and to compare it with the conventional on-grid model-based algorithms, including ℓ_1-SVD [4] and Wℓ_1-SVD [20] and the off-grid sparse Bayesian inference (OGSBI) algorithm [25] and RSBL [12]. All algorithms process $\mathbf{Y}_{sv}$ to obtain the DOA estimates. For [4], SeDuMi is used to solve the ℓ_1 norm problem, and the reweighted ℓ_1 norm problem in [20] is tackled by CVX. In all simulation examples, the noise is additive Gaussian white noise and a ULA of $M = 10$ sensors is considered. The direction grid is uniformly divided with resolution of 2° sampling from 0° to 180°, and $T = 200$ snapshots are collected. Assume that two narrowband far-field signals from [66.3°, 80.6°] impinge onto ULA. All results are based on 500 Monte Carlo runs. Our simulations are performed using MATLAB R2015b environment on a system with 3.40GHz intel core i7 CPU and 4 GB RAM, under a 64-bit Windows 7 operating system.

In the first test, we investigate the root mean squared error (RMSE) of the proposed method, ℓ_1-SVD, Wℓ_1-SVD, OGSBI and RSBL versus signal-to-noise ratio (SNR). It can be seen from Fig. 1 that the DOA estimation performance of the proposed method is superior to that of ℓ_1-SVD, Wℓ_1-SVD, OGSBI and RSBL expecially for a higher SNR. We also notice that the on-grid algorithms, i.e., ℓ_1-SVD and Wℓ_1-SVD cannot provide reliable DOA estimation when SNR is above 5 dB.

In the second test, RMSE versus snapshot number with different methods is studied, where SNR is fixed at 0 dB, and the snapshot number is varied from 100 to 600. From Fig. 2, it is observed that the proposed method has better angle estimation performance than other algorithms for all T. The DOA estimation performance of the proposed method gradually improves with the snapshot number.

In the third test, the resolution probability of different algorithms versus SNR is examined and the results are plotted in Fig. 3. The resolution probability is computed as the ratio between the number of successful runs and the total number of the independent runs. A trial is regarded as a successful one when the absolute deviation between the estimated and true DOA is less than 1°. It is concluded that all methods exhibit a 100% correct resolution probability at the high SNR region. We also see that the proposed method has the highest resolution probability at SNR ≥ -6 dB.

In the fourth test, the RMSE versus number of iterations is studied for the proposed method and off-grid algorithms, and the results are shown in Fig. 4. It can be seen that the proposed method is superior to the OGSBI and RSBL in terms of convergence speed and RMSE.

In the fifth test, we compare the CPU run times of different algorithms. The results averaged over 100 trials are tabulated in Table 1. It is observed that the proposed algorithm enjoys more computational attractiveness than ℓ_1-SVD, Wℓ_1-SVD, OGSBI and RSBL.

5. Conclusion

In this paper, we have addressed the problem of DOA estimation in sparse representation framework. A novel objective function regularized by the nonconvex sparsity-inducing penalty has been proposed for off-grid DOA estimation. We follow the rationale of alternating minimization to minimize the resulting objective function, i.e., we first update the sparse signal via ADMM as the solver by fixing perturbation matrix, and then we calculating the perturbation matrix by SRLS when the sparse signal is fixed. Simulation results show that the proposed method provides more accurate DOA estimation and faster implementation compared with several conventional algorithms. Although not shown here, it is worth pointing out that the proposed method can also work for arrays with irregular sensor spacings.

Acknowledgments

This work was supported by a grant from the NSFC/RGC Joint Research Scheme sponsored by the National Natural Science Foundation of China and the Research Grants Council of Hong Kong (61531166005, N_CityU 104/15).

References

[1] E. J. Candès, M. B. Wakin, An introduction to compressive sampling, IEEE Signal Processing Magazine 25 (2) (2008) 21–30.

[2] F. K. W. Chan, H. C. So, L. Huang, L.-T. Huang, Parameter estimation and identifiability in bistatic multiple-input multiple-output radar, IEEE Transactions on Aerospace and Electronic Systems 51 (3) (2015) 2047–2056.

[3] C. Qian, L. Huang, N. D. Sidiropoulos, H. C. So, Enhanced puma for direction-of-arrival estimation and its performance analysis, IEEE Transactions on Signal Processing 64 (16) (2016) 4127–4137.

[4] D. Malioutov, M. Cetin, A. S. Willsky, A sparse signal reconstruction perspective for source localization with sensor arrays, IEEE Transactions on Signal Processing, 53 (8) (2005) 3010–3022.

[5] Q. Liu, X. P. Wang, Direction of arrival estimation via reweighted ℓ_1 norm penalty algorithm for monostatic MIMO radar, Multidimensional Systems and Signal Processing (2016) 1–12.

[6] Z. Yang, L. Xie, Enhancing sparsity and resolution via reweighted atomic norm minimization, IEEE Transactions on Signal Processing 64 (4) (2016) 995–1006.

[7] Z. Yang, C. Zhang, L. Xie, Robustly stable signal recovery in compressed sensing with structured matrix perturbation, IEEE Transactions on Signal Processing 60 (9) (2012) 4658–4671.

[8] H. Zhu, G. Leus, G. B. Giannakis, Sparse regularized total least squares for sensing applications, in: 2010 IEEE 11th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Marrakech, Morocco, June 2010, pp. 1–5.

[9] H. Duan, Z. Qian, Y. Wang, Off-grid DOA estimation based on noise subspace fitting, in: 2015 IEEE International Conference on Digital Signal Processing (DSP), Singapore, July 2015, pp. 675–678.

[10] Z. Tan, P. Yang, A. Nehorai, Joint sparse recovery method for compressed sensing with structured dictionary mismatches, IEEE Transactions on Signal Processing 62 (19) (2014) 4997–5008.

[11] Z. Tan, A. Nehorai, Sparse direction of arrival estimation using co-prime arrays with off-grid targets, IEEE Signal Processing Letters 21 (1) (2014) 26–29.

[12] J. Dai, X. Bao, W. Xu, C. Chang, Root sparse Bayesian learning for off-grid DOA estimation, IEEE Signal Processing Letters 24 (1) (2017) 46–50.

[13] H. Mohimani, M. Babaie-Zadeh, C. Jutten, A fast approach for overcomplete sparse decomposition based on smoothed ℓ_0 norm, IEEE Transactions on Signal Processing 57 (1) (2009) 289–301.

[14] W. J. Zeng, H. C. So, L. Huang, ℓ_p -MUSIC: Robust direction-of-arrival estimator for impulsive noise environments, IEEE Transactions on Signal Processing 61 (17) (2013) 4296–4308.

[15] L. Chen, Y. Gu, Fast sparse recovery via non-convex optimization, in: 2015 IEEE Global Conference on Signal and Information Processing (GlobalSIP), Orlando, FL, USA, December 2015, pp. 1275–1279.

[16] R. Tibshirani, Regression shrinkage and selection via the lasso, Journal of the Royal Statistical Society, Series B 58 (1994) 267–288.

[17] S. Boyd, N. Parikh, E. Chu, B. Peleato, J. Eckstein, Distributed optimization and statistical learning via the alternating direction method of multipliers, Found. Trends Mach. Learn. 3 (1) (2011) 1–122.

[18] J. Dai, X. Xu, D. Zhao, Direction-of-arrival estimation via real-valued sparse representation, IEEE Antennas and Wireless Propagation Letters 12 (2013) 376–379.

[19] E. J. Candès, M. B. Wakin, S. P. Boyd, Enhancing sparsity by reweighted ℓ_1 minimization, Journal of Fourier Analysis and Applications 14 (5) (2008) 877–905.

[20] X. P. Wang, W. Wang, J. Liu, X. Li, J. X. Wang, A sparse representation scheme for angle estimation in monostatic MIMO radar, Signal Processing 104 (2014) 258 – 263.

[21] M. Grant, S. Boyd, CVX: Matlab software for disciplined convex programming, version 2.1, Available: http://cvxr.com/cvx (Mar. 2014).

[22] J. Sturm, Using SeDuMi 1.05, a Matlab Toolbox for Optimization Over Symmetric Cones. Tilburg University, Dept., Econometrics, Tilburg, Netherlands, Available: http://fewcal.kub.nl/sturm/software/sedumi.html (Dec. 2004).

[23] Z. Tan, Y. C. Eldar, A. Nehorai, Direction of arrival estimation using co-prime arrays: A super resolution viewpoint, IEEE Transactions on Signal Processing 62 (21) (2014) 5565–5576.

[24] L. Wang, L. Zhao, G. Bi, C. Wan, L. Zhang, H. Zhang, Novel wideband DOA estimation based on sparse Bayesian learning with Dirichlet process priors, IEEE Transactions on Signal Processing 64 (2) (2016) 275–289.

[25] Z. Yang, L. Xie, C. Zhang, Off-grid direction of arrival estimation using sparse Bayesian inference, IEEE Transactions on Signal Processing 61 (1) (2013) 38–43.

[26] L. Chen, Y. Gu, The convergence guarantees of a non-convex approach for sparse recovery, IEEE Transactions on Signal Processing 62 (15) (2014) 3754–3767.

[27] G. E. Ivanov, Strong and weak convexity for linear differential games, in: Proceedings of 36th IEEE Conference on Decision and Control, Kobe, Japan, Vol. 4, December 1996, pp. 3729–3734 vol.4.

[28] N. Parikh, S. Boyd, Proximal algorithm, Foundations and Trends in Optimization 1 (3) (2014) 127–239.

[29] P. Tseng, Convergence of a block coordinate descent method for nondifferentiable minimization, Journal of Optimization Theory and Applications 109 (3) (2001) 475–494.

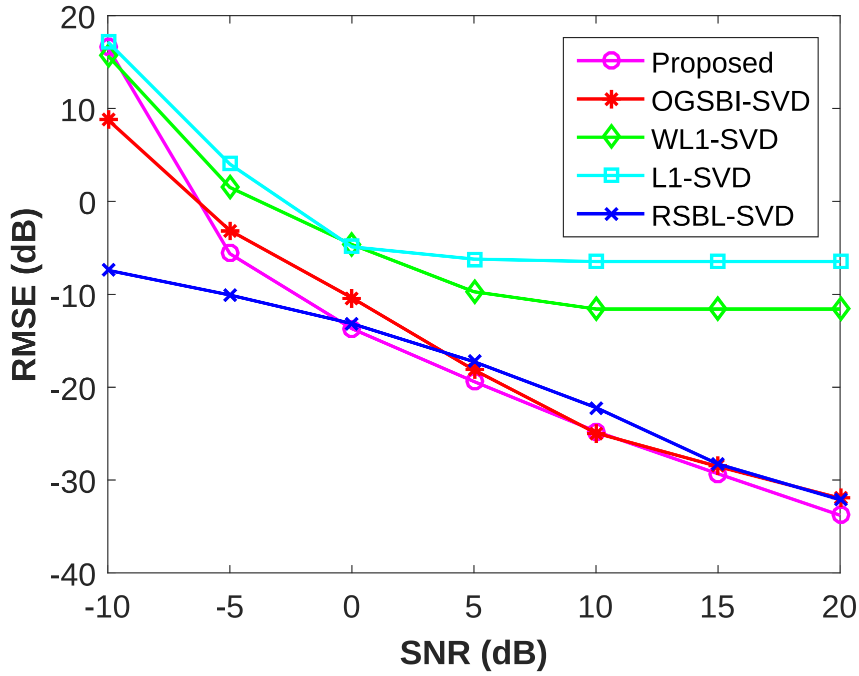

Figure 1: RMSE versus SNR.

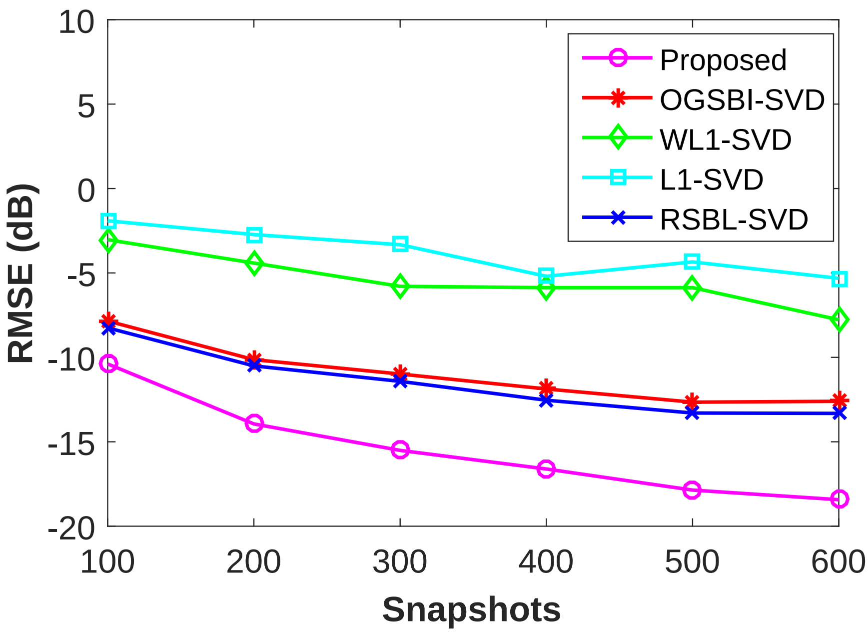

Figure 2: RMSE versus snapshot number.

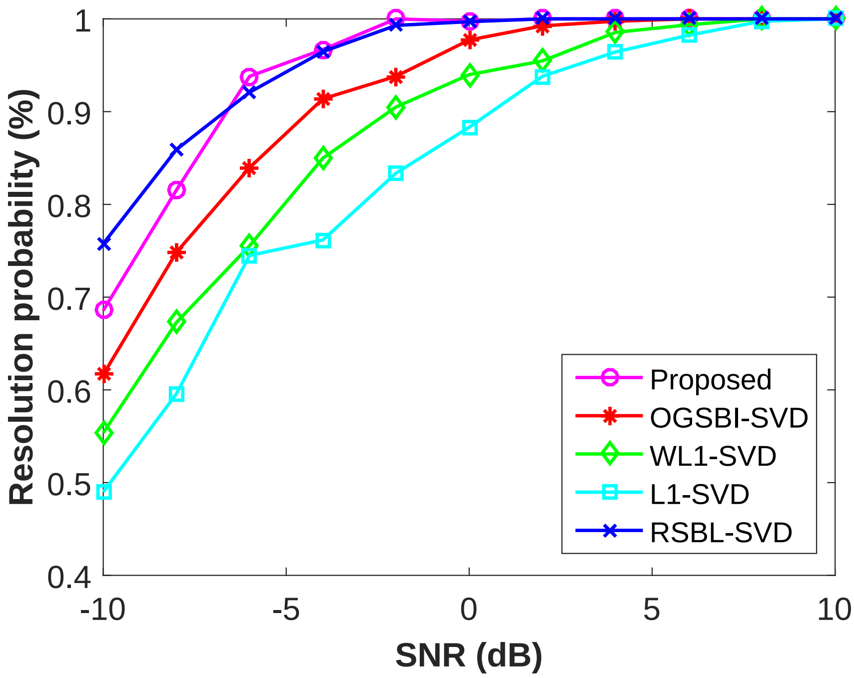

Figure 3: Resolution probability versus SNR.

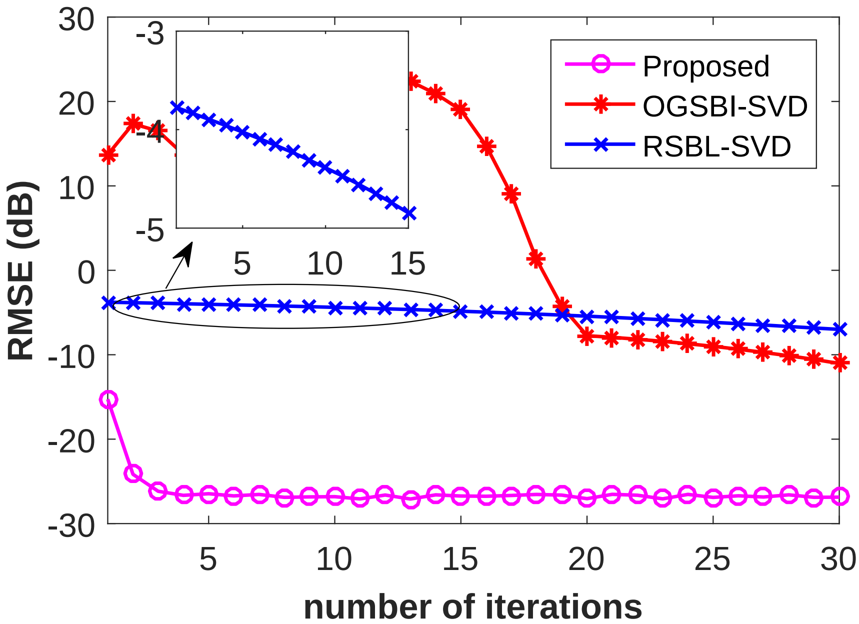

Figure 4: Convergence curves of different off-grid algorithms.

Table 1: Runtime comparison.

Time(s) \ Snapshot / Algorithm	$T = 50$		$T = 200$	
	SNR = 10 dB	SNR = -10 dB	SNR = 10 dB	SNR = -10 dB
Proposed	0.1569	0.1608	0.1629	0.1670
RSBL-SVD	0.2567	0.2640	0.2515	0.2421
OGSBI-SVD	0.2667	0.2688	0.2744	0.2782
ℓ_1-SVD	0.5594	0.5705	0.5507	0.5869
Wℓ_1-SVD	1.8290	1.9113	2.0023	2.1072