Off-grid DOA Estimation Based on Analysis of the Convexity of Maximum Likelihood Function

Liang Liu and Ping Wei

arXiv:1412.6720v1 [cs.IT] 21 Dec 2014

Abstract—**Spatial compressive sensing (SCS) has recently been applied to direction-of-arrival (DOA) estimation owing to advantages over conventional ones. However the performance of compressive sensing (CS)-based estimation methods decreases when true DOAs are not exactly on the discretized sampling grid. We solve the off-grid DOA estimation problem using the deterministic maximum likelihood (DML) estimation method. In this work, we analyze the convexity of the DML function in the vicinity of the global solution. Especially under the condition of large array, we search for an approximately convex range around the ture DOAs to guarantee the DML function convex. Based on the convexity of the DML function, we propose a computationally efficient algorithm framework for off-grid DOA estimation. Numerical experiments show that the rough convex range accords well with the exact convex range of the DML function with large array and demonstrate the superior performance of the proposed methods in terms of accuracy, robustness and speed.**

Index Terms—**Compressive sensing, direction-of-arrival estimation, off-grid model, convexity, deterministic maximum likelihood.**

I. INTRODUCTION

THIS paper addresses the problem of off-grid direction-of-arrival (DOA) estimation through analysis of the convexity of the maximum likelihood (ML) function. The ML criterion has drawn much attention due to its attractive benefits [1]–[5], such as consistency, asymptotic normality, efficiency. However, the ML estimation approach generally requires a multidimensional search at a great computational cost in practical multi-target scenarios. Fortunately, some iterative algorithms, such as the Newton type algorithm [6], [7], are applied to reduce computation cost. However, those iterative algorithms need a good initial value, which must be sufficiently close to the global minimum in order to prevent convergence to a local extremum. This work is dedicated to find out a convergence region by analyzing the convexity of the ML function.

Array processing for DOA estimation has been a topic of intensive research interest during the past two decades [8]. Since the ML approach is often deemed exceedingly complex, several suboptimal methods are proposed, such as subspace based methods, especially MUSIC algorithm [9], which exchange the multidimensional search problem for one-dimensional search problem. Whereas, the performance of MUSIC algorithm is not good as ML estimation in DOA

estimation with snapshot deficient scenario or coherent signals [2]–[4].

Recently, the research on DOA estimation has been advanced owing to the development of methods based on compressed sensing (CS) or sparse signal reconstruction (SSR) [10] by exploiting the spatial sparsity in the array model. The CS-based DOA estimation approaches are extremely attractive for their ability to resolve closely spaced sources, few snapshots and correlated sources. Some CS-based DOA estimation methods have been presented. ℓ_1 optimization (or called basis pursuit, BP) [11]–[13], matching pursuit (MP) [14], [15], and sparse Bayesian inference/learning (SBL) [16]–[18] are proposed in the case of a single measurement vector (SMV). In the case of multiple measurement vectors (MMV), the simultaneous sparse approximation problem arises, and singular value decomposition (SVD) is introduced to reduce complexity and sensitivity against noise [19]. A natural extension of basis pursuit is convex relaxation algorithms, which employ mixed norm $\ell_{p,q}$ optimization in this case [19]–[23]. Simultaneous orthogonal matching pursuit (S-OMP) [20] or multiple response model orthogonal matching pursuit (M-OMP) [24], is a multiple response variant of matching pursuit. Additionally, sparse Bayesian learning for MMV guarantees [25] the simultaneous sparsity by assuming the same sparse prior, such as a Laplace signal prior, for the signals at all snapshots.

All these CS-based DOA estimation methods employ fixed sampling grid and assume that all the true DOAs are exactly located on the selected grid. When the true DOAs are beyond the fixed grid, their performance will degrade due to discretization error. There are still difficulties to set the grid interval in practical situations where the true DOAs are not on the sampling grid. On one hand, a dense sampling grid is necessary for accurate DOA estimation to reduce the gap between the true DOA and its nearest grid point since the estimated DOAs are constrained on the grid. But on the other, a dense sampling grid leads to high computational complexity of recovery algorithm and a highly coherent matrix that violates the condition for the sparse signal recovery.

In [26], an off-grid model for DOA estimation is introduced and a sparse total least squares (STLS) method based on the Gaussian assumption of off-grid distance is proposed. However, the Gaussian condition cannot be satisfied in the off-grid DOA estimation problem. In [27], a new off-grid algorithm termed OGSBI-SVD from a Bayesian perspective is presented by assuming the off-grid distance satisfies uniform distribution. As shown in [27], OGSBI-SVD can exceed a lower bound of root mean squared error (RMSE) that is shared

The authors are with the School of Electronic Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China (e-mail: liu_yinliang@outlook.com; pwei@uestc.edu.cn).

among all on-grid model based methods. But its performance is not ideal especially with high signal to noise ratio (SNR) or large snapshots. Besides, this approach is not robust in the case of large grid interval.

This work proposes a deterministic ML (DML) based approach to mitigate CS-based DOA estimation biases. This approach presents the same performance as DML, whose asymptotic estimation performance often able to achieve the Cramér-Rao bound (CRB). So this approach is no longer limited by the density of sampling grid. This work can be viewed from two aspects, one is that it mitigates CS-based DOA estimation biases through ML estimation, the other is that it realizes the ML estimation through CS-based DOA estimation and Newton iteration. It is the key issue to make the approach convergent. The main contributions of this work are listed as follows. Firstly, we prove that the DML function for DOA estimation is a convex function in the vicinity of the global minimum. Secondly, we come to a conclusion that an approximate convex range of the likelihood function can be obtained under certain conditions, which to some extent suggests how "sufficiently close" to the global minimum the iterations must be initialized. Therefore, we decide the density of sampling grid for CS-based DOA estimation algorithms by our result instead of empiric value. In addition, we propose a class of algorithms for off-grid DOA estimation based on the convexity of the DML function. The proposed algorithms consist of two stages. In the first stage, by setting the grid interval according to our result, CS-based DOA estimation approaches obtain coarse DOA estimates. In the second, we initialize the Newton type iteration with these estimated DOAs to realize the DML estimation. Because of the convexity of the DML function, our proposed methods are likely to converge to the global minimum of the DML function.

The following notations are used in the paper. $(\cdot)^{\mathrm{T}}$ and $(\cdot)^{\mathrm{H}}$ denote the transpose and Hermitian transpose, respectively. $\|\cdot\|_F$, $E(\cdot)$, $\mathrm{Tr}(\cdot)$ and $\Re(\cdot)$ stand for the Frobenius norm, expectation, trace, and real part operators, respectively. $diag(\mathbf{A})$ denotes a column vector composed of the diagonal elements of a matrix $\mathbf{A}$, and $\mathrm{diag}(\mathbf{x})$ is a diagonal matrix with $\mathbf{x}$ being its diagonal elements. x_j is the jth entry of a vector $\mathbf{x}$. $\mathbf{A}_i$, $\mathbf{A}^j$ and A_{ij} are the jth column, ith row and (i,j)th entry of a matrix $\mathbf{A}$, respectively. $x'(\theta)$, $x''(\theta)$ are the 1st, 2nd-order derivative of $x(\theta)$ with respect to θ, respectively. $\odot$ denotes the Hadamard product. $\mathbf{I}_M$ stands for an $M \times M$ identity matrix.

The remainder of the paper is organized as follows. We present the measurement models in Section II, and review CS-based and ML for DOA estimation in Section III. We analyze the convexity of the DML function in Section IV, and derive our CSDML algorithms in Section V. Numerical examples appear in Section VI. We give our conclusions in Section VII.

II. MEASUREMENT MODELS

Consider K narrowband far-field signals impinging on an array composed of $M(M > K)$ sensors. The array output can

be written as [8]

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t), t = 1, \cdots, T, \qquad (1)$$

where $\mathbf{x}(t)$ is the measurement vector at the tth snapshot, $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_K)]$ (denoted as $\mathbf{A}$ for convenience) is the array steering matrix, $\mathbf{a}(\theta_k)$ is the steering vector corresponding to the kth source, $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_K]^{\mathrm{T}}$ is the vector containing the DOAs of all sources, θ_k is the DOA of the kth source, $\mathbf{s}(t)$ is the vector of all signal values at the tth snapshot, T is the total number of temporal measurements, and $\mathbf{n}(t)$ is the $M \times 1$ noise vector following the zero-mean circular complex Gaussian distribution with covariance matrix $E\{\mathbf{n}(t)\mathbf{n}^{\mathrm{H}}(t)\} = \sigma^2\mathbf{I}_M$, σ^2 is noise power. With the further assumption that $\mathbf{s}(t)$ and $\mathbf{n}(t)$ are uncorrelated, the array covariance matrix is given by

$$\mathbf{R} = \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}} + \sigma^2\mathbf{I}_M, \qquad (2)$$

where $\boldsymbol{\Sigma}$ is the source covariance matrix. $\mathbf{R}$ can be estimated by $\widehat{\mathbf{R}} \approx \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^{\mathrm{H}}(t)$.

III. DOA ESTIMATION: CS-BASED DOA ESTIMATION AND MAXIMUM LIKELIHOOD ESTIMATION

In this section we review two kinds of methods for DOA estimation, as we will take advantage of their conclusions in our context.

A. CS-based DOA Estimation

Since DOAs are sparse in the spatial domain, the source localization problem is formulated as a sparse representation problem, or spatial compressive sensing (SCS) problem [19]. To cast the DOA estimation problem in a SCS framework, an overcomplete representation $\boldsymbol{\Theta}$ in terms of all possible source locations was introduced. Let $\boldsymbol{\Theta} = [\Theta_1, \cdots, \Theta_N]$ be a sampling grid of all source locations of interest and $\boldsymbol{\Psi} = [\mathbf{a}(\Theta_1), \cdots, \mathbf{a}(\Theta_N)]$ is the array manifold matrix corresponding to the DOAs $\boldsymbol{\Theta}$. The number of potential source locations N will typically be much greater than the number of sources K or even the number of sensors M. Assuming that the true source directions $\boldsymbol{\theta}$ are contained in $\boldsymbol{\Theta}$, formulation (1) is reformulated using an overcomplete representation as the following form,

$$\mathbf{x}(t) = \boldsymbol{\Psi}\widetilde{\mathbf{s}}(t) + \mathbf{n}(t). \qquad (3)$$

where, $\widetilde{\mathbf{s}}(t)$ is a $N \times 1$ vector containing K non-zero entries, where the nth element $\widetilde{\mathbf{s}}_n(t) = \mathbf{s}_k(t)$ if $\Theta_n = \theta_k$ for $k = 1, \cdots, K$, otherwise $\widetilde{\mathbf{s}}_n(t) = 0$.

In effect, model (3) allows us to exchange the DOA estimation problem for the problem of recovering the sparse signal $\widetilde{\mathbf{s}}$ from the array output $\mathbf{x}$. When the number of snapshots $T > 1$, this problem is a simultaneous sparse approximation problem, which has received a lot of attention recently. And several computationally feasible methods have been presented for estimating the sparse signal, such as matching pursuit (MP), convex relaxation, sparse Bayesian learning (SBL).

B. Maximum Likelihood DOA Estimation

The ML approach is a standard technique in statistical estimation theory. The ML estimation is calculated as the values of the unknown parameters that maximize the likelihood function. This can be interpreted as selecting the set of parameters that make the observed data most probable. ML techniques for the sensor array problem have been studied by a number of researchers, see for example [5], [13], [28]. When applying the ML technique to the sensor array problem, two main methods have been considered, depending on the model assumption on the signal waveforms. When the emitter signals are modeled as Gaussian random processes, a stochastic ML (SML) method is obtained. On the other side,if the emitter signals are modeled as unknown deterministic quantities, the resulting estimator is referred to as the deterministic ML (DML) estimator. According to [29], we know that the DML criterion depends on θ in a simpler way than does the SML criterion. So we prefer the DML estimator in this paper.

The DML DOA estimator is shown as follow [2], [29]

$$\vartheta = \arg \min_{\theta} \operatorname{tr} \left(\mathbf{P}^{\perp} \mathbf{R} \right), \tag{4}$$

where $\mathbf{P}^{\perp} = \mathbf{I} - \mathbf{P}$, $\mathbf{P} = \mathbf{B}\mathbf{B}^{\dagger}$, in which $\mathbf{B}$ is the array steering matrix corresponding to the DOAs estimation $\vartheta = [\vartheta_1, \cdots, \vartheta_K]$, $\mathbf{B}^{\dagger} = \left(\mathbf{B}^{\mathrm{H}} \mathbf{B} \right)^{-1} \mathbf{B}^{\mathrm{H}}$ is the pseudo-inverse of $\mathbf{B}$. However, it is important to note that (4) is a non-linear multidimensional minimization problem, and the criterion function often possesses a large number of local minima.

All methods considered herein require a multidimensional non-linear optimization for computing the signal parameter estimates. Usually, analytical solutions are not available and one has to resort to numerical search techniques. Several optimization methods have appeared in the array processing literature, including different Newton-type techniques. It is well-known that the Newton-type method gives locally a quadratic convergence. So Newton-type algorithms for the DML techniques are described in this subsection. The estimate is iteratively calculated as

$$\vartheta^{k+1} = \vartheta^k - \mathbf{H}^{-1} \nabla, \tag{5}$$

where ϑ^k is the estimate at iteration k, $\mathbf{H}$ represents the Hessian matrix of the criterion function, and ∇ is the gradient. The Hessian and gradient are evaluated at ϑ^k.

The DML gradient is given by [1], [2].

$$\nabla = -2\Re \left(\operatorname{diag} \left(\mathbf{B}^{\dagger} \mathbf{R} \mathbf{P}^{\perp} \mathbf{D} \right) \right). \tag{6}$$

The DML Hessian matrix can be expressed as (See Appendix A)

$$\mathbf{H} = 2\Re \left(\mathbf{C} \right), \tag{7}$$

where

$$\mathbf{C} = \left(\mathbf{D}^{\mathrm{H}} \mathbf{P}^{\perp} \mathbf{D} \right) \odot \left(\mathbf{B}^{\dagger} \mathbf{R} \mathbf{B}^{\dagger \mathrm{H}} \right)^{\mathrm{T}} \tag{8a}$$

$$- \left(\mathbf{D}^{\mathrm{H}} \mathbf{P}^{\perp} \mathbf{R} \mathbf{P}^{\perp} \mathbf{D} \right) \odot \left(\mathbf{B}^{\dagger} \mathbf{B}^{\dagger \mathrm{H}} \right)^{\mathrm{T}} \tag{8b}$$

$$+ \left(\mathbf{B}^{\dagger} \mathbf{D} \right) \odot \left(\mathbf{B}^{\dagger} \mathbf{R} \mathbf{P}^{\perp} \mathbf{D} \right)^{\mathrm{T}} \tag{8c}$$

$$+ \left(\mathbf{B}^{\dagger} \mathbf{D} \right)^{\mathrm{T}} \odot \left(\mathbf{B}^{\dagger} \mathbf{R} \mathbf{P}^{\perp} \mathbf{D} \right) \tag{8d}$$

$$- \mathbf{I} \odot \left(\mathbf{B}^{\dagger} \mathbf{R} \mathbf{P}^{\perp} \mathbf{F} \right)^{\mathrm{T}}, \tag{8e}$$

where $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_K]$, $\mathbf{F} = [\mathbf{F}_1, \cdots, \mathbf{F}_K]$, $\mathbf{D}_i = \mathbf{a}'(\vartheta_i)$, $\mathbf{F}_i = \mathbf{a}''(\vartheta_i)$.

Since this paper needs to analyze the convexity of the DML function, it is necessary to use accurate Hessian matrix expression. For the expression of Hessians matrix of both [1] and [2] is asymptotic, that of our paper differs from them.

IV. THE CONVEXITY OF THE LIKELIHOOD FUNCTION

A. Motivation

Although the Newton method can quickly converge to a local extremum, the quality of the convergence point depends on the shape of the criterion function in question. If ∇ possesses several minima, the iteration must be initialized "sufficiently close" to the global minimum in order to prevent convergence to a local extremum.

How "sufficiently close" to the global minimum will the iteration converge? A sufficient condition is that the likelihood function is convex in this area near the global minimum. So we will analyze the convex of the DML function in this section.

B. The Convexity of the Likelihood Function near the True DOAs

The following results about gradient and Hessian matrix will show that the likelihood function is a convex function near the true DOAs.

Lemma 1. $\mathbf{D}^{\mathrm{H}} \mathbf{P}^{\perp} \mathbf{D} \succ 0$.

Proof: See Appendix B. ∎

Theorem 1. $\lim_{\vartheta \to \theta} \nabla = \mathbf{0}$, $\lim_{\vartheta \to \theta} \mathbf{H} \succ 0$.

Proof: See Appendix C. ∎

Remark 1. Theorem 1 indicates that θ is a minimum of likelihood function. Note that DML function is a continuous function in terms of θ, thus there is a range Ω around the true DOAs θ, where $\forall \vartheta \in \Omega$, $\mathbf{H} \succ 0$.

C. The Convex Range of the Likelihood Function

The likelihood function is a convex function in the vicinity of the global minimum according to subsection IV-B. As long as the initial value is in the convex vicinity of the DML solution, the Newton iteration is feasible. It is necessary to determine the convex range of the DML function. To determine the convex range of the DML function, we need to analyze the positive definiteness of $\mathbf{H}$. Unfortunately, this analysis is rather complicated because $\mathbf{H}$ is a multidimensional nonlinear matrix function in terms of ϑ. We consider a simply scene where array is a uniform linear array (ULA), which is one of the most common array. Without loss of

generality, the norm of the steering vector is normalized. For arbitrary ULAs, whose sensors are located at $d_1, \cdots, d_M$, we have $\mathbf{a}(\theta) = \frac{1}{\sqrt{M}} \exp\left(-j\frac{2\pi}{\lambda} \sin(\theta) [\, d_1 \cdots d_M\,]^{\mathrm{H}}\right)$, $\mathbf{a}'(\theta) = -j\frac{2\pi}{\lambda} \cos(\theta) [\, d_1 \cdots d_M\,]^{\mathrm{H}} \odot \mathbf{a}(\theta)$.

For simplicity, we assume that: (1) the number of sensor elements M is large enough; (2) The interval between adjacent true DOAs $\boldsymbol{\theta}$ is large enough with fixed M; (3) The interval between the estimated DOAs $\boldsymbol{\vartheta}$ and the true DOAs is much smaller comparing to the interval between adjacent DOAs, namely $|\vartheta_i - \theta_i| \ll |\theta_j - \theta_i|$, $|\vartheta_i - \theta_i| \ll |\vartheta_j - \vartheta_i|$, $i \neq j$, where ϑ_i is the estimation of θ_i. According to the assumption (1), (2) and (3), we have $\left|\mathbf{B}_i^{\mathrm{H}}\mathbf{A}_i\right|^2 \gg \left|\mathbf{B}_i^{\mathrm{H}}\mathbf{A}_j\right|^2 \approx 0$, $\left|\mathbf{D}_i^{\mathrm{H}}\mathbf{A}_i\right|^2 \gg \left|\mathbf{D}_i^{\mathrm{H}}\mathbf{A}_j\right|^2 \approx 0$, $\left|\mathbf{D}_i^{\mathrm{H}}\mathbf{D}_i\right|^2 \gg \left|\mathbf{D}_i^{\mathrm{H}}\mathbf{D}_j\right|^2 \approx 0$. Those expressions are shown in APPENDIX G of [2]. We can reformulate them in the form of matrix as follows

$$\mathbf{B}^{\mathrm{H}}\mathbf{A} \approx \left(\mathbf{B}^{\mathrm{H}}\mathbf{A}\right) \odot \mathbf{I}, \tag{9}$$

$$\mathbf{D}^{\mathrm{H}}\mathbf{A} \approx \left(\mathbf{D}^{\mathrm{H}}\mathbf{A}\right) \odot \mathbf{I}, \tag{10}$$

$$\mathbf{D}^{\mathrm{H}}\mathbf{D} \approx \left(\mathbf{D}^{\mathrm{H}}\mathbf{D}\right) \odot \mathbf{I}. \tag{11}$$

Based on (9) (10) (11) and after some matrix manipulations, we further have $\mathbf{B}^{\mathrm{H}}\mathbf{B} = \mathbf{I}$, $\mathbf{B}^{\dagger}\mathbf{B}^{\dagger\mathrm{H}} = \mathbf{I}$, $\mathbf{B}^{\dagger}\mathbf{A} = \left(\mathbf{B}^{\dagger}\mathbf{A}\right) \odot \mathbf{I}$, $\mathbf{D}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{A} = \left(\mathbf{D}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{A}\right) \odot \mathbf{I}$.

Lemma 2. $\mathbf{B}^{\dagger^i}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}_i$ is real, where $\mathbf{B}^{\dagger^i}$ denotes ith row vector of $\mathbf{B}^{\dagger}$.

Proof: See Appendix D. ■

Substituting (2) into (8c), we have

$$(8c) = \left(\left(\mathbf{B}^{\dagger}\mathbf{D}\right) \odot \mathbf{I}\right) \odot \left(\mathbf{B}^{\dagger}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}\right)^{\mathrm{T}}. \tag{12}$$

From (12) we know (8c) is an approximately diagonal matrix, whose ith diagonal element is

$$[(8c)]_{ii} = (\mathbf{B}^{\dagger^i}\mathbf{D}_i)(\mathbf{B}^{\dagger^i}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}_i). \tag{13}$$

Since $\alpha_i = -j2\pi \cos(\vartheta_i)/\lambda$ is a pure imaginary number, $\mathbf{B}_i^{\mathrm{H}}\mathbf{D}_i = \frac{\alpha_i}{M}\sum_{k=1}^{M} d_k$ is a pure imaginary number. Note that $\mathbf{B}^{\dagger^i}\mathbf{D}_i \approx \mathbf{B}_i^{\mathrm{H}}\mathbf{D}_i$, $\mathbf{B}^{\dagger^i}\mathbf{D}_i$ is an approximately imaginary number. Combining Lemma 2, $[(8c)]_{ii}$ is an approximately imaginary number. Namely, (8c) is an approximately diagonal matrix with approximately imaginary numbers being its diagonal elements. So,

$$\Re\left((8c)\right) \approx \mathbf{0}. \tag{14}$$

For similar reasons,

$$\Re\left((8d)\right) \approx \mathbf{0}. \tag{15}$$

Note that the formula (8e) includes $\mathbf{F}_k$, the second derivative of the steering vector $\mathbf{a}(\vartheta_k)$, it is more difficult to take this term into the analysis of formula (8). We will ignore this term without proof. There are two underlying reasons for doing like this. Firstly, the analysis of the positive definiteness of $\mathbf{H}$ will become easier. Secondly, the influence of neglecting (8e) is nearly negligible in terms of the positive definiteness of $\mathbf{H}$ in practice. Therefore, after using the formula (14)-(15) and

neglecting (8e), $\mathbf{H} = 2\Re(\mathbf{C})$ can be approximated as follows

$$\begin{aligned}
\mathbf{C} &\approx (8a) + (8b) \\
&= \left(\mathbf{D}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}\right) \odot \left(\mathbf{B}^{\dagger}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\mathbf{B}^{\dagger\mathrm{H}}\right)^{\mathrm{T}} \\
&\quad - \left(\mathbf{D}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}\right) \odot \left(\mathbf{B}^{\dagger}\mathbf{B}^{\dagger\mathrm{H}}\right)^{\mathrm{T}} \\
&= \left(\left(\mathbf{D}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}\right) \odot \mathbf{I}\right) \odot \left(\left(\left(\mathbf{B}^{\dagger}\mathbf{A}\right) \odot \mathbf{I}\right)\boldsymbol{\Sigma}\left(\left(\mathbf{B}^{\dagger}\mathbf{A}\right) \odot \mathbf{I}\right)^{\mathrm{H}}\right)^{\mathrm{T}} \\
&\quad - \left(\left(\left(\mathbf{D}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{A}\right) \odot \mathbf{I}\right)\boldsymbol{\Sigma}\left(\left(\mathbf{D}^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{A}\right) \odot \mathbf{I}\right)^{\mathrm{H}}\right) \odot \mathbf{I}. \tag{16}
\end{aligned}$$

So $\mathbf{C}$ is an approximately diagonal matrix, whose ith diagonal element is

$$\begin{aligned}
\mathbf{C}_{ii} &= \Sigma_{ii}\left(\mathbf{B}^{\dagger^i}\mathbf{A}_i\right)\left(\mathbf{B}^{\dagger^i}\mathbf{A}_i\right)^{\mathrm{H}}\mathbf{D}_i^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}_i \\
&\quad - \Sigma_{ii}\mathbf{D}_i^{\mathrm{H}}\left(\mathbf{P}^{\perp}\mathbf{A}_i\right)\left(\mathbf{P}^{\perp}\mathbf{A}_i\right)^{\mathrm{H}}\mathbf{D}_i \\
&= \Sigma_{ii}\beta_i\mathbf{D}_i^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}_i - \Sigma_{ii}\left\|\mathbf{P}^{\perp}\mathbf{A}_i\right\|_F^2 \mathbf{D}_i^{\mathrm{H}}\left(\mathbf{e}_i\mathbf{e}_i^{\mathrm{H}}\right)\mathbf{D}_i \\
&= \Sigma_{ii}\left(\beta_i\mathbf{D}_i^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}_i - (1-\beta_i)\mathbf{D}_i^{\mathrm{H}}\left(\mathbf{e}_i\mathbf{e}_i^{\mathrm{H}}\right)\mathbf{D}_i\right), \tag{17}
\end{aligned}$$

where $\beta_i = \left|\mathbf{B}_i^{\mathrm{H}}\mathbf{A}_i\right|^2$, $\mathbf{e}_i \triangleq \frac{\mathbf{P}^{\perp}\mathbf{A}_i}{\|\mathbf{P}^{\perp}\mathbf{A}_i\|_2}$, $\left\|\mathbf{P}^{\perp}\mathbf{A}_i\right\|_2^2 = 1 - \left|\mathbf{B}_i^{\mathrm{H}}\mathbf{A}_i\right|^2 = 1 - \beta_i$.

Lemma 3. $\mathbf{P}^{\perp} - \mathbf{e}_i\mathbf{e}_i^{\mathrm{H}} \succeq 0$.

Proof: See Appendix E. ■

According to Lemma 3, we have

$$\mathbf{D}_i^{\mathrm{H}}\mathbf{P}^{\perp}\mathbf{D}_i - \mathbf{D}_i^{\mathrm{H}}\left(\mathbf{e}_i\mathbf{e}_i^{\mathrm{H}}\right)\mathbf{D}_i \geq 0. \tag{18}$$

If $\beta_i \geq 0.5$, $\beta_i \geq 1 - \beta_i$. Substituting $\beta_i \geq 1 - \beta_i$, $\boldsymbol{\Sigma}_{ii} > 0$ and (18) into (17), we have $\mathbf{C}_{ii} \geq 0$. Note that both $\mathbf{C}$ and $\mathbf{H}$ are approximately diagonal matrices, we have the following result

$$\mathbf{H} \succeq 0, when \left|\mathbf{B}_i^{\mathrm{H}}\mathbf{A}_i\right|^2 \geq 0.5, i = 1, 2, \cdots, K. \tag{19}$$

Remark 2. Since we use the approximation (9)(10)(11) and neglect (8e) to obtain (19), the convex range given by (19) is a rough range. Even so, the result is often useful to roughly determine convex range of the DML function. In fact, the rough range accords well with the real range, as long as the hypothesis conditions are fully satisfied. This will be shown by simulations. The significance of the result lies in its giving some guidance for the iterative algorithm initialization in practice. As long as the initial value for the iterative algorithm is in the above convex range, the algorithm will converge to the global minimum of the DML function. The approximate convex range is acquired through solving the inequation $\left|\mathbf{B}_i^{\mathrm{H}}\mathbf{A}_i\right|^2 \geq 0.5$ using the numerical method. In this way, the range is slightly different when θ_i varies. The range is almost same as long as θ_i is not a big angle. So, we take the width where $\theta_i = 0$ as the range. Namely, the approximately convex range is about half of 3dB (half-power) beamwidth $\mathrm{BW}_{0.5}$ of the array around the true DOAs.

V. CSDML: Off-Grid DOA Estimation Based on the Convexity of DML Function

CS-based DOA estimation methods based on on-grid model can provide estimated DOAs which are on its nearest grid point since the estimated DOAs are constrained on the grid. And the estimated DOAs can be uesd to initialize Newton iteration.

TABLE I
ALGORITHM CSDML

1) Set grid interval r according to (19) and (20), obtain Θ and Ψ;
2) Calculate θ_c through the on-grid methods such as MP, convex relaxation, SBL and so on;
3) Set $\vartheta^0 = \theta_c$ for Newton iteration;
4) At ϑ^k, calculate ∇ and $\mathbf{H}$ according to (6) and (7), respectively;
5) Update ϑ^{k+1} according to (5);
6) Check the convergence criterion for the kth iteration. Terminate the process when $\left\| \vartheta^{k+1} - \vartheta^k \right\|_2 \leq \tau$ or the maximum number of iterations is reached, where τ is a predefined small value. Otherwise, return to step (4) and continue the whole process again.

Since the result (19) indicates that by and large the convex range of the likelihood function is half of the 3dB beamwidth of the array, we can set grid interval r for the CS-based DOA estimation as

$$r = \gamma \frac{\mathrm{BW}_{0.5}}{2}, \tag{20}$$

where $\gamma < 1$ is a small regulation parameter. When the DOAs are more adjacent, γ is set to a smaller number. We recommend to set $\gamma = 0.5$ usually since it can give an accurate yet fast DOA estimation. Then the DOAs θ_c estimated by the CS-based DOA estimation methods are most likely in the convex range of the DML function so that Newton iteration always converge to the global minimum rather than a local extremum.

Based on the above ideas, we propose a class of algorithms for off-grid DOA estimation, outlined in Table I. The proposed algorithms consist of two stages. By setting the grid interval according to our result (19), CS-based DOA estimation approaches obtain coarse DOA estimates in the first stage. Next, we initialize the Newton type iteration with these estimated DOAs to realize the DML estimation.

Remark 3. It is worthy to note that all kinds of DOA estimation methods can be applied in the first stage of the proposed approach, including the CS-based methods, such as MP, convex relaxation, SBL. As long as their DOA estimates are already in the convex vicinity of the DML solution, the Newton type iteration will converge to the global minimum of the DML function and achieve the same performance in the final. We can choose the computationally effective methods in practice. Usually, OMP is decidedly less costly than all of these methods.

We assume $N \gg M \gg K$, where N denotes the number of sampling grid. Note that each iteration of both M-BP and M-SBL can also be computed in $\mathcal{O}(NM^2)$ using the implementation as [25]. After SVD is introduced, each M-OMP iteration can be computed in $\mathcal{O}(NMK)$ using the implementation as [24] and standard techniques for least-squares problems (see [30], Chapter 5, for extensive details). Each Newton iteration can be computed in $\mathcal{O}(M^2K)$, which is not significant compared to M-BP, M-SBL or M-OMP iteration.

VI. SIMULATION

This section includes two parts. Firstly, we compare the convex rough range of ours with convex exact range of the DML function. Secondly, we present the numerical simulation results to illustrate the performance of the proposed algorithm. In our experiments, we consider $K = 2$ narrowband far-field non-coherent signals with the equal power impinging on a ULA composed of M sensors which are separated by a half wavelength of the signal. The SNR is defined as $\mathrm{SNR}_i = 10\log_{10}(E(|\mathbf{s}_i(t)|^2)/\sigma^2)$, where SNR_i denotes ith signal's SNR. The RMSE is defined as $\mathrm{RMSE} = \sqrt{\frac{1}{N_m K} \sum_{i=1}^{N_m} \sum_{k=1}^{K} \left(\theta_k^i - \widehat{\theta}_k^i\right)^2}$, where the superscript i refers to the ith trial, N_m denotes the number of Monte Carlo tests. The simulations are performed using MATLAB2012B running on an Intel Core 2 Duo, 3.20 GHz processor with 3 GB of memory, under Windows7 32bit.

A. Convexity of the Likelihood Function

In order to measure the similarity of the range based on the (19) and the exact one, We define two measures as

$$\mathrm{IRR} = \sqrt[K]{M(\Omega_I)/M(\Omega_R)}, \tag{21}$$

$$\mathrm{IAR} = \sqrt[K]{M(\Omega_I)/M(\Omega_A)}, \tag{22}$$

where, Ω_R represents a set, which subjects to $\forall \vartheta \in \Omega_R$, $\mathbf{H} \succ 0$ and is obtained from experiment. Ω_A represents another set, which approximately subjects to $\forall \vartheta \in \Omega_A$, $\mathbf{H} \succ 0$, and is obtained from (19). And $\Omega_I = \Omega_R \cap \Omega_A$. $M(\Omega)$ represents the measure of Ω. Taking K-th root of the ratio is to eliminate the influence of the dimension K. Obviously, $0 \leq \mathrm{IRR} \leq 1$, $0 \leq \mathrm{IAR} \leq 1$. When Ω_A is closer to Ω_R, both IRR and IAR are closer to 1, and it means that the estimated range and the actual range is consistent. Otherwise, they are closer to 0.

In the simulations, to acquire the Ω_R, firstly, we divide the area near the real DOAs into uniformly-spaced grid, then we calculate the eigenvalues of $\mathbf{H}$ corresponding to the value of the DOAs grid. If the eigenvalues of $\mathbf{H}$ at a grid are all non-negative, the ML function is convex at this grid, which belongs to Ω_R. We regard the number of elements of Ω_R as $M(\Omega_R)$. We can easily obtain Ω_A and $M(\Omega_A)$ through the 3dB beamwidth $\mathrm{BW}_{0.5}$.

We carry out 3 experiments in this subsection. The number of snapshots is fixed at $T = 200$ in each experiment. 500 Monte Carlo trials are simulated except the first one. Fig.1 is the result of an experiment with SNR $= 10$ dB, $M = 8$ and $\theta = [0° \ 30°]^{\mathrm{T}}$. $\mathrm{BW}_{0.5} \approx 12.8°$ when $M = 8$. It intuitively shows the convex rough range and the exact range of the DML function. Despite that the rough range based on (19) is slightly different, two ranges still agree well.

Fig.2 depicts the IRR and IAR versus SNR, with $\theta = [-7.5° \ 7.5°]^{\mathrm{T}}$ and $M = 8$. It is shown that the mean of IRR and IAR slightly increase with the increasing of SNR, but nearly remains 0.73 and 0.84, respectively. As long as (2) is satisfied strictly, the convex rough is always invariable. But IRR and IAR fluctuate more remarkably around the mean value with low SNR.
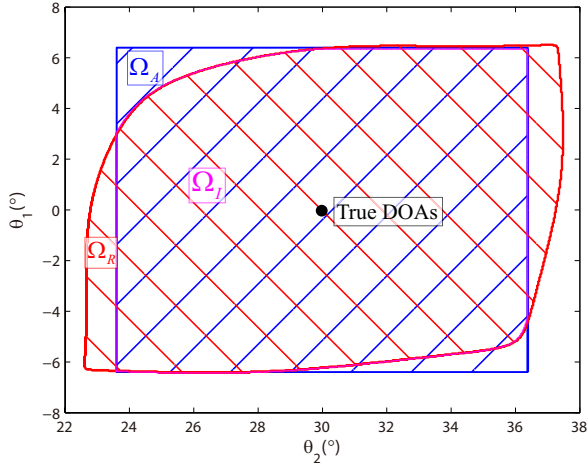
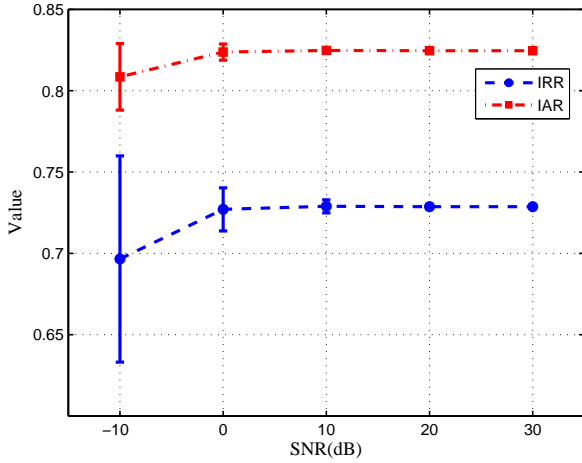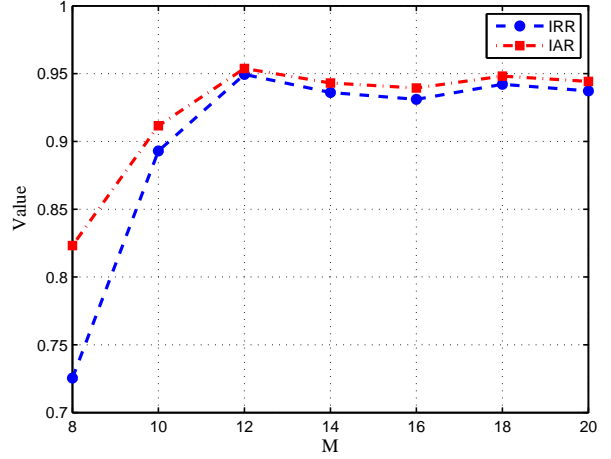Fig. 1.   The rough and exact convex range of the DML function.

Fig. 3.   ICR and IAR versus the number of sensors.

kinds of the CS-based methods are optional in the first stage of our proposed approach. In our simulations, we take OMP [24] and SBL [25] as representatives of on-grid model based methods. There are two underlying reasons for doing like this. Firstly, OMP is computationally effective method. Secondly, both SBL and OGSBI-SVD are from a Bayesian perspective. We do not choose ℓ_1-SVD because it costs too much usually. Since OGSBI-SVD uses the SVD to reduce the computational workload of the signal recovery process and the sensitivity to noise. For the sake of fair, the SVD is introduced for OMP and SBL, too. We compare OGSBI-SVD, SBL, OMP with our proposed algorithms ,CSDML(SBL) and CSDML(OMP), in terms of RMSE and computational time with respect to SNR, snapshots and grid interval, where CSDML(SBL) and CSDML(OMP) denote our proposed algorithms using SBL and OMP in the first stage, respectively. 1000 Monte Carlo trials for each experiment are simulated in this subsection.

It should be noted that there exists a Lower Bound for the RMSE of the on-Grid methods (GLB) regardless of the SNR since the best DOA estimate that those methods can obtain is the grid point nearest to the true DOA. In fact, the lower bound is shared among all on-grid model based methods including ℓ_1-SVD, SBL, OMP and so on. In the case of the uniformly distributed DOA, the lower bound is $\text{GLB} = \frac{r}{2\sqrt{3}}$. In the case of the fixed DOA, the lower bound is $\text{GLB} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\theta_k - \widehat{\theta}_k\right)^2}$, where $\widehat{\theta}_k$ is the grid point nearest to θ_k.

In the first two experiments, we consider that $\boldsymbol{\theta} = [2.37°, 30.82°]$ and $r = 2°$, which is equal to that λ is set 0.313 since $\text{BW}_{0.5} \approx 12.8°$ for $M = 8$. Fig.4 depicts the RMSE versus SNR, with the number of snapshots fixed at $T = 200$, and Fig.5 shows the RMSE versus the number of snapshots, with SNR fixed at 10 dB. It is shown that the performance of the off-grid based methods, OGSBI-SVD and CSDML, are superior to that of the on-grid method, SBL and OMP, in terms of RMSE in Fig.4 and 5. The performance of the on-grid model will degrade due to discretization error when the true DOAs are beyond the fixed sampling grid. Fig.4 and 5

Fig. 2.   IRR and IAR versus SNR.

Fig.3 shows the IRR and IAR versus the number of sensors M , with SNR fixed at 10 dB, DOA fixed at $[-10°, 10°]$. Fig.3 shows that by and large the mean of IRR and IAR increases with the number of sensors. Namely, the convex rough range and the exact range of the DML function agree better with more sensors. In Fig.3, the means of both IRR and IAR are about 0.93 when $M \geq 12$. Namely, the convex rough range and the exact range of the DML function agree very well when DOA interval is larger than $2.5\text{BW}_{0.5}$ ($\text{BW}_{0.5} \approx 8°$ when $M = 12$).

B. DOA Estimation Performance

In this subsection, we consider the uniform sampling grid as $\{-90°, -90° + r, -90° + 2r, \cdots, 90°\}$, with r being the grid interval. We fix the number of sensors of ULA at $M = 8$. We set the parameters of OGSBI-SVD as [27]. It's worth pointing out that it is necessary to set the origin at the middle point of the ULA to reduce the approximation error for OGSBI-SVD, however is unnecessary for our proposed algorithms. We assume that K is known in our simulation. Note that all
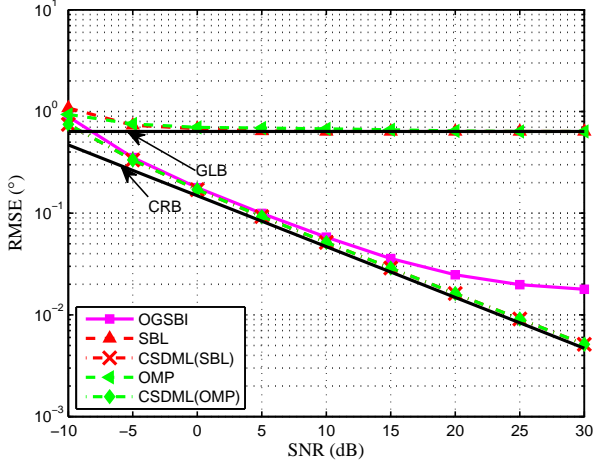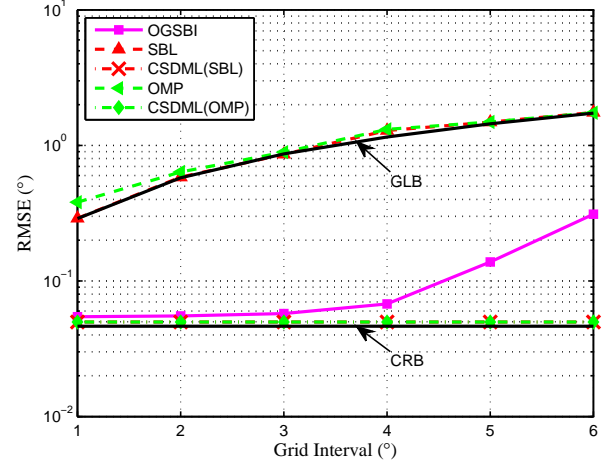
Fig. 4. RMSE of DOA estimates versus SNR.

Fig. 6. RMSE of DOA estimates versus the grid interval.
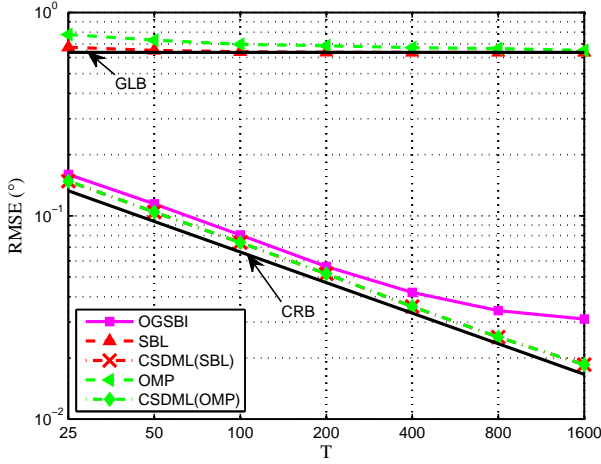
Fig. 5. RMSE of DOA estimates versus the number of snapshots.

TABLE II
AVERAGED CPU TIMES OF OGSBI-SVD AND CSDML WITH RESPECT TO
r. TIME UNIT: sec.

r	$1°$	$2°$	$3°$	$4°$	$5°$	$6°$
OGSBI-SVD	1.1e+0	1.3e-1	5.1e-2	3.3e-2	2.8e-2	2.1e-2
SBL	1.6e-1	1.6e-2	7.7e-3	6.1e-3	5.0e-3	3.8e-3
DML	5.0e-4	6.0e-4	5.9e-4	6.2e-4	6.2e-4	6.3e-4
CSDML(SBL)	1.6e-1	1.7e-2	8.3e-3	6.7e-3	5.6e-3	4.4e-3
OMP	4.3e-4	3.4e-4	3.4e-4	3.2e-4	3.2e-4	3.2e-4
DML	5.5e-4	5.6e-4	5.9e-4	6.2e-4	6.6e-4	6.3e-4
CSDML(OMP)	9.8e-4	9.0e-4	9.3e-4	9.4e-4	9.8e-4	9.5e-4

show that although SBL and OMP have different performance, CSDML(SBL) and CSDML(OMP) have same performance, which outperforms OGSBI-SVD under the same simulation conditions, especially with high SNR or large snapshots, where OGSBI-SVD has the constant performances, however both CSDML(SBL) and CSDML(OMP) are almost able to achieve the CRB.

The sensitivity to grid interval of different methods is studied in the third experiment. The SNR and the number of snapshots are fixed at 10 dB and 200, respectively. When $M = 8$, we have $\text{BW}_{0.5} \approx 12.8°$. Therefore the grid interval r is selected from $1°$ to $6°$ with an interval of $1°$. It is equal to that λ is set from 0.156 to 0.938 with an interval of 0.156. In each trial, $K = 2$ directions of θ_1, θ_2 are uniformly generated within direction intervals $[-3°, 3°]$ and $[27°, 33°]$ respectively. The RMSE versus different grid intervals is depicted in Fig.6, which shows that CSDML has superior performance than OGSBI-SVD. This simulation also demonstrates that CSDML is more robust than OGSBI-SVD, even in the case of very coarse grid. Table II presents the

averaged CPU times of OGSBI-SVD and CSDML (excluding the SVD process) with respect to r. Their CPU times decrease as the grid gets coarser for both OGSBI-SVD and CSDML. However, CSDML, especially CSDML(OMP), is many times faster than OGSBI-SVD especially in the case of very fine grid interval. Additionally, the CPU times of Newton iteration are much less comparing with SBL or OGSBI-SVD, for Newton iteration gives locally a quadratic convergence. The table also shows that CSDML(OMP) is less costly than CSDML(SBL) since OMP is less costly than SBL.

VII. CONCLUSIONS

In this work, we analyzed the convexity of the DML function in the vicinity of the global solution and found out a rough range with large array, which is the half of 3dB beamwidth around the true DOAs to generally guarantee the DML function convex. It to some extent answered the question that how close it should be to the global minimum to make sure the iteration feasible. After that, we proposed algorithms for off-grid DOA estimation based on the convexity of the DML function. We illustrated by simulations that the approximately convex range agrees well with the exact convex range of the DML function with large array. The simulations demonstrated that the proposed approaches outperform standard CS methods based on on-grid model and the OGSBI-SVD based on off-

grid model. The performance of the proposed approaches is close to the CRB, and exceeds that of the OGSBI-SVD, especially with high SNR and large snapshots. Moreover, the proposed approaches are robust in the case of large grid interval, whereas the OGSBI-SVD is not. Besides, the proposed approaches are many times faster than OGSBI-SVD in all case of grid interval.

From a new perspective, this work provided a computationally efficient algorithm framework including initial value calculation and Newton iteration to realize DML estimation. Unlike other approaches, our algorithms do not need a global/multidimensional search and the convex analysis of the DML function would guarantee the Newton iteration convergence. This algorithm frame can be applied to DOA estimation, frequency estimation and so on.

APPENDIX A
DERIVATION OF THE HESSIAN MATRIX

For convenience, the following matrices are defined as

$$\mathbf{E}_i \triangleq \frac{\partial \mathbf{B}}{\partial \vartheta_i} = [0 \; \mathbf{D}_i \; 0],$$

$$\mathbf{G}_{i,j} \triangleq \frac{\partial^2 \mathbf{B}}{\partial \vartheta_i \partial \vartheta_j} = \begin{cases} [\mathbf{0} \; \mathbf{F}_i \; \mathbf{0}], & j = i \\ \mathbf{0}, & j \neq i \end{cases}.$$

According to the results in [31], [32], we have

$$\frac{\partial^2 \mathbf{P}}{\partial \vartheta_i \partial \vartheta_j} = 2\Re \left(-\mathbf{P}^\perp \mathbf{E}_j \mathbf{B}^\dagger \mathbf{E}_i \mathbf{B}^\dagger \right.$$
$$+ \mathbf{P}^\perp \mathbf{E}_i \mathbf{B}^\dagger \mathbf{B}^{\dagger H} \mathbf{E}_j^H \mathbf{P}^\perp$$
$$- \mathbf{B}^{\dagger H} \mathbf{E}_j^H \mathbf{P}^\perp \mathbf{E}_i \mathbf{B}^H$$
$$- \mathbf{P}^\perp \mathbf{E}_i \mathbf{B}^\dagger \mathbf{E}_j \mathbf{B}^\dagger$$
$$\left. + \mathbf{P}^\perp \mathbf{G}_{i,j} \mathbf{B}^\dagger \right).$$

Therefore, we have

$$H_{ji} = \frac{\partial^2 \mathrm{tr}\left(\mathbf{P}^\perp \mathbf{R}\right)}{\partial \vartheta_i \partial \vartheta_j} = -\frac{\partial^2 \mathrm{tr}\left(\mathbf{P}\mathbf{R}\right)}{\partial \vartheta_i \partial \vartheta_j}$$
$$= 2\Re \left[\mathrm{tr}\left(+ \mathbf{B}^\dagger \mathbf{R} \mathbf{B}^{\dagger H} \mathbf{E}_j^H \mathbf{P}^\perp \mathbf{E}_i \right.\right.$$
$$- \mathbf{B}^\dagger \mathbf{B}^{\dagger H} \mathbf{E}_j^H \mathbf{P}^\perp \mathbf{R} \mathbf{P}^\perp \mathbf{E}_i$$
$$+ \mathbf{B}^\dagger \mathbf{R} \mathbf{P}^\perp \mathbf{E}_j \mathbf{B}^\dagger \mathbf{E}_i$$
$$+ \mathbf{B}^\dagger \mathbf{R} \mathbf{P}^\perp \mathbf{E}_i \mathbf{B}^\dagger \mathbf{E}_j$$
$$\left.\left. - \mathbf{B}^\dagger \mathbf{R} \mathbf{P}^\perp \mathbf{G}_{i,j} \right) \right]. \tag{23}$$

By comparison, (7) is the matrix form of (23).

APPENDIX B
PROOF OF LEMMA 1

Suppose arbitrary vector $\mathbf{x} \neq \mathbf{0}$. $\mathbf{P}^\perp = \mathbf{I} - \mathbf{B}\mathbf{B}^\dagger$ is the orthogonal projector matrix and the rank of $\mathbf{B}$ is K, therefore its eigenvalues are composed of ones and zeros, and the number of them is $M - K$ and K, respectively. Thus $\lambda_{\min}\left(\mathbf{P}^\perp\right) = 0$.

Obviously, $\mathbf{D}^H \mathbf{P}^\perp \mathbf{D}$ is a Hermitian matrix. And according to the definition of the Rayleigh quotient and Min-max theorem in [33], we have

$$\mathbf{x}^H \mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\mathbf{x} = (\mathbf{D}\mathbf{x})^H \mathbf{P}^\perp (\mathbf{D}\mathbf{x}) \geq \lambda_{\min}\left(\mathbf{P}^\perp\right) \|\mathbf{D}\mathbf{x}\|^2.$$

If and only if $\mathbf{D}\mathbf{x}$ is the eigenvector of $\mathbf{P}^\perp$ corresponding to the smallest eigenvalue 0, $\mathbf{x}^H \mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\mathbf{x} = 0$. Otherwise, $\mathbf{x}^H \mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\mathbf{x} > 0$.

Since $\mathbf{P}^\perp \mathbf{B} = \mathbf{0}$, then $\mathbf{a}(\vartheta_i)$, column vector of $\mathbf{B}$, is eigenvector of $\mathbf{P}^\perp$ corresponding to eigenvalue 0. If and only if $\mathbf{D}\mathbf{x} \in R(\mathbf{B})$, where $R(\mathbf{B})$ denotes the column space of $\mathbf{B}$, $(\mathbf{D}\mathbf{x})^H \mathbf{P}^\perp (\mathbf{D}\mathbf{x}) = 0$. However, $\mathbf{D}_i = \mathbf{a}'(\vartheta_i)$ and $\mathbf{B}_i = \mathbf{a}(\vartheta_i)$, thus $\mathbf{D}\mathbf{x} \in R(\mathbf{B})$ is usually not satisfied, and $(\mathbf{D}\mathbf{x})^H \mathbf{P}^\perp (\mathbf{D}\mathbf{x}) > 0$. Namely, $\forall \mathbf{x} \neq 0, (\mathbf{D}\mathbf{x})^H \mathbf{P}^\perp (\mathbf{D}\mathbf{x}) > 0$. Therefore $\mathbf{D}^H \mathbf{P}^\perp \mathbf{D} \succ 0$.

APPENDIX C
PROOF OF THEOREM 1

Considering (2) and after some matrix manipulations, we have

$$\mathbf{B}^\dagger \mathbf{R} \mathbf{P}^\perp = \mathbf{B}^\dagger \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \mathbf{P}^\perp, \tag{24}$$

$$(8a) + (8b) = \left(\mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\right) \odot \left(\mathbf{B}^\dagger \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \mathbf{B}^{\dagger H}\right)^\mathrm{T}$$
$$- \left(\mathbf{D}^H \mathbf{P}^\perp \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \mathbf{P}^\perp \mathbf{D}\right) \odot \left(\mathbf{B}^\dagger \mathbf{B}^{\dagger H}\right)^\mathrm{T} \tag{25}$$

Taking the limit of (24) and (25), respectively, and note that $\lim_{\vartheta \to \theta} \mathbf{A}^H \mathbf{P}^\perp = \mathbf{0}$, we have

$$\lim_{\vartheta \to \theta} \mathbf{B}^\dagger \mathbf{R} \mathbf{P}^\perp = \mathbf{0} \Rightarrow \lim_{\vartheta \to \theta} \nabla = \mathbf{0}, \tag{26}$$

$$\lim_{\vartheta \to \theta} ((8a) + (8b)) = \left(\mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\right) \odot \boldsymbol{\Sigma}^\mathrm{T}. \tag{27}$$

According to the theorem on Hadamard product in [33], combining Lemma 1 and $\boldsymbol{\Sigma}^\mathrm{T} \succ 0$, we obtain

$$\left(\mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\right) \odot \boldsymbol{\Sigma}^\mathrm{T} \succ 0. \tag{28}$$

Note that $\mathbf{D}^H \mathbf{P}^\perp \mathbf{D} \odot \boldsymbol{\Sigma}^\mathrm{T}$ is a Hermitian matrix, thus we have

$$\Re\left(\left(\mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\right) \odot \boldsymbol{\Sigma}^\mathrm{T}\right) \succ 0. \tag{29}$$

Combining the results of (26) and (27), we obtain

$$\lim_{\vartheta \to \theta} \mathbf{H} = 2\Re\left(\left(\mathbf{D}^H \mathbf{P}^\perp \mathbf{D}\right) \odot \boldsymbol{\Sigma}^\mathrm{T}\right) \succ 0. \tag{30}$$

We thus prove Theorem 1.

APPENDIX D
PROOF OF LEMMA 2

In this subsection, $\mathbf{A}_{-j}$, $\mathbf{A}^{-i}$ and $A_{i(-j)}$ are the jth to last column, ith to last row and ith row jth to last column of a matrix $\mathbf{A}$, respectively. Let $\mathbf{X} = \mathbf{B}^\dagger \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \mathbf{P}^\perp \odot \mathbf{B}^\mathrm{T}$, we have $X_{ik} = \left(\mathbf{B}^\dagger \mathbf{A} \boldsymbol{\Sigma} \mathbf{A}^H \left(\mathbf{P}^\perp\right)_k \mathbf{B}^k\right)_{ii}$.

For arbitrary ULA, whose sensors are located at d_1, $\cdots$, d_M, we make a transformation of $\mathbf{A} = \mathbf{U}\mathbf{G}$, where $\mathbf{G}$ is a diagonal matrix, whose ith diagonal element is $\exp\left(\left(-j\pi\left(d_1 + d_M\right)\sin\left(\theta_i\right)\right)/(2\lambda)\right)$, and $\mathbf{U}$ is an array steering matrix, whose origin is at the middle point of the ULA. Therefore, $\mathbf{U} = \mathbf{J}\mathbf{U}^*$, $\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}^H = \left(\mathbf{G}\boldsymbol{\Sigma}\mathbf{G}^H\right)^*$, where $\mathbf{J}$ is the $M \times M$ exchange matrix whose entries all are zero except the one in the $(i, M - i + 1)$th position for $i = 1, \cdots, M$. Substituting them into $\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^H$, we have

$$\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^H = \mathbf{J}\left(\mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^H\right)^*\mathbf{J}. \tag{31}$$

Similarly, we let $\mathbf{B} = \mathbf{VH}$ and we have $\mathbf{V} = \mathbf{JV}^*$. According to the above results and after some matrix manipulations, we obtain

$$\mathbf{B}^\dagger = \mathbf{H}^{-1}\mathbf{V}^\dagger, \tag{32}$$

$$\mathbf{V}^k = \left(\mathbf{V}^{-k}\right)^*, \tag{33}$$

$$\mathbf{V}^\dagger \mathbf{J} = \left(\mathbf{V}^\dagger\right)^*, \tag{34}$$

$$\mathbf{J}\mathbf{P}^\perp{}_k = \left(\mathbf{P}^\perp{}_{-k}\right)^*. \tag{35}$$

Substituting (32) into X_{ik} and $X_{i(-k)}$, and considering the form of $\mathbf{H}$, we have

$$X_{ik} = \left(\mathbf{V}^\dagger \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\left(\mathbf{P}^\perp\right)_k \mathbf{V}^k\right)_{ii}, \tag{36}$$

$$X_{i(-k)} = \left(\mathbf{V}^\dagger \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\left(\mathbf{P}^\perp\right)_{-k} \mathbf{V}^{-k}\right)_{ii}. \tag{37}$$

Substituting (31)(33)(34)(35) into (36), we have

$$X_{ik} = \left(\mathbf{V}^\dagger \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\left(\mathbf{P}^\perp\right)_{-k} \mathbf{V}^{-k}\right)_{ii}^*. \tag{38}$$

Comparing (38) with (37), we have $X_{i(-k)} = X_{ik}^*$. So we obtain

$$\Re\left(X_{i(-k)}\right) = \Re\left(X_{ik}\right). \tag{39}$$

Since $\mathbf{P}^\perp \mathbf{B} = \mathbf{0}$, $\sum_{k=1}^{M}\left(\mathbf{P}^\perp\right)_k B_{ki} = \mathbf{0}$, $\sum_{k=1}^{M} X_{ik} = 0$,

$$\sum_{k=1}^{M} \Re(X_{ik}) = 0. \tag{40}$$

Note that $d_1, \cdots, d_M$ is a arithmetic progression, and according to (39) and (40), we acquire

$$\Re\left(\sum_{k=1}^{M} d_k X_{ik}\right) = 0. \tag{41}$$

Namely, $\sum_{k=1}^{M} d_k X_{ik}$ is a pure imaginary number.

Since $D_{ki} = \alpha_i d_k B_{ki}$, where $\alpha_i = -j2\pi \cos\left(\vartheta_i\right)/\lambda$ is a pure imaginary number. And combining (41), we have proved that $\mathbf{B}^{\dagger^i} \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^{\mathrm{H}}\mathbf{P}^\perp \mathbf{D}_i = \alpha_i \sum_{k=1}^{M} d_k X_{ik}$ is real number.

APPENDIX E
PROOF OF LEMMA 3

The eigenvalues of $\mathbf{P}^\perp$ are composed of $M - K$ ones and K zeros. Because $\mathbf{P}^\perp \mathbf{e}_i = \mathbf{e}_i$, $\mathbf{e}_i$ is normalized eigenvector of $\mathbf{P}^\perp$ corresponding to eigenvalue 1. Therefore, $\mathbf{P}^\perp$ can be eigendecomposed into $\mathbf{P}^\perp = \sum_{i=1}^{M-K} \mathbf{q}_i \mathbf{q}_i^{\mathrm{H}}$, where $\mathbf{q}_1 = \mathbf{e}_i$. Thus, $\mathbf{P}^\perp - \mathbf{e}_i \mathbf{e}_i^{\mathrm{H}} = \sum_{i=2}^{M-K} \mathbf{q}_i \mathbf{q}_i^{\mathrm{H}}$. So the eigenvalues of $\mathbf{P}^\perp - \mathbf{e}_i \mathbf{e}_i^{\mathrm{H}}$ are composed of $M - K - 1$ ones and $K + 1$ zeros. Therefore, $\mathbf{P}^\perp - \mathbf{e}_i \mathbf{e}_i^{\mathrm{H}} \succeq 0$.

ACKNOWLEDGMENT

REFERENCES

[1] U. Sandkuhler and J. Bohme, "Accuracy of maximum-likelihood estimates for array processing," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 12, Apr 1987, pp. 2015–2018.

[2] P. Stoica and N. Arye, "Music, maximum likelihood, and cramer-rao bound," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[3] P. Stoica and K. Sharman, "Maximum likelihood methods for direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 7, pp. 1132–1143, Jul 1990.

[4] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound: further results and comparisons," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec 1990.

[5] ——, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1783–1795, Oct 1990.

[6] P. E. Gill, W. Murray, and M. H. Wright, *Practical optimization*. Academic press, 1981.

[7] J. E. Dennis Jr and R. B. Schnabel, *Numerical methods for unconstrained optimization and nonlinear equations.* Siam, 1996.

[8] H. Krim and M. Viberg, "Two decades of array signal processing research: the parametric approach," *IEEE Signal Process. Mag.*, vol. 13, no. 4, pp. 67–94, Jul 1996.

[9] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar 1986.

[10] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, April 2006.

[11] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.

[12] E. J. Candès, J. K. Romberg, and T. Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Commun. Pure Appl. Math.*, vol. 59, no. 8, pp. 1207–1223, 2006.

[13] M. Li and Y. Lu, "Maximum likelihood doa estimation in unknown colored noise fields," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 44, no. 3, pp. 1079–1090, July 2008.

[14] S. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, vol. 41, no. 12, pp. 3397–3415, Dec 1993.

[15] R. DeVore and V. Temlyakov, "Some remarks on greedy algorithms," *Adv. Comput. Math.*, vol. 5, no. 1, pp. 173–187, 1996.

[16] M. E. Tipping, "Sparse bayesian learning and the relevance vector machine," *J. Mach. Learn. Res.*, vol. 1, pp. 211–244, Sep. 2001.

[17] S. Ji, Y. Xue, and L. Carin, "Bayesian compressive sensing," *IEEE Trans. Signal Process.*, vol. 56, no. 6, pp. 2346–2356, June 2008.

[18] S. Babacan, R. Molina, and A. Katsaggelos, "Bayesian compressive sensing using laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan 2010.

[19] D. Malioutov, M. Cetin, and A. 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.

[20] S. Cotter, B. Rao, K. Engan, and K. Kreutz-Delgado, "Sparse solutions to linear inverse problems with multiple measurement vectors," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2477–2488, July 2005.

[21] Y. Eldar and M. Mishali, "Robust recovery of signals from a structured union of subspaces," *IEEE Trans. Inf. Theory*, vol. 55, no. 11, pp. 5302–5316, Nov 2009.

[22] J. A. Tropp, "Algorithms for simultaneous sparse approximation. Part II: Convex relaxation," *Signal Process.*, vol. 86, no. 3, pp. 589 – 602, 2006.

[23] M. Hyder and K. Mahata, "A robust algorithm for joint-sparse recovery," *IEEE Signal Process. Lett.*, vol. 16, no. 12, pp. 1091–1094, Dec 2009.

[24] J. A. Tropp, A. C. Gilbert, and M. J. Strauss, "Algorithms for simultaneous sparse approximation. Part I: Greedy pursuit," *Signal Process.*, vol. 86, no. 3, pp. 572 – 588, 2006.

[25] D. Wipf and B. Rao, "An empirical bayesian strategy for solving the simultaneous sparse approximation problem," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3704–3716, July 2007.

[26] H. Zhu, G. Leus, and G. Giannakis, "Sparsity-cognizant total least-squares for perturbed compressive sampling," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2002–2016, May 2011.

[27] Z. Yang, L. Xie, and C. Zhang, "Off-grid direction of arrival estimation using sparse bayesian inference," *IEEE Trans. Signal Process.*, vol. 61, no. 1, pp. 38–43, Jan 2013.

[28] M. Pesavento and A. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1310–1324, Jul 2001.

[29] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, in *Radar Array Processing*. Springer Berlin Heidelberg, 1993, pp. 99–151.

[30] G. H. Golub and C. F. van Van Loan, *Matrix computations*. The Johns Hopkins University Press, 1996.

[31] G. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," *SIAM J. Numer. Anal.*, vol. 10, no. 2, pp. 413–432, 1973.

[32] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.

[33] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge University Press, 1985.