Underdetermined DOA Estimation for Wideband Signals Using Robust Sparse Covariance Fitting

Zhen-Qing He, *Student Member, IEEE*, Zhi-Ping Shi, *Member, IEEE*, Lei Huang, *Member, IEEE*, and Hing Cheung So, *Senior Member, IEEE*

***Abstract*—From the co-array perspective, sparse spatial sampling can significantly increase the degrees-of-freedom (DOFs), enabling us to perform underdetermined direction-of-arrival (DOA) estimation. By leveraging the increased DOFs from the sparse spatial sampling, we develop a new underdetermined DOA estimation method for wideband signals, named wideband *sparse* spectrum fitting (W-SpSF) estimator. In W-SpSF, we formulate a sparse reconstruction problem that includes a quadratic (ℓ_2) weighted covariance fitting term added to a sparsity-promoting $(\ell_{2,1})$ regularizer. Meanwhile, the optimal regularization parameter of W-SpSF is studied to ensure robust sparse recovery. Numerical results enabled nested arrays demonstrate that the W-SpSF estimator outperforms the spatial smoothing based MUSIC algorithm and works well in nonuniform noise environment.**

***Index Terms*—Co-array, direction-of-arrival (DOA) estimation, sparse spectrum fitting (SpSF), wideband signal.**

I. Introduction

THE topic of target localization with less sensors than sources, i.e., underdetermined direction-of-arrival (DOA) estimation, has been receiving considerable interest in recent years [1]–[9]. An effective approach for this problem is to construct a new *virtual* array with a higher degrees-of-freedom (DOFs) than that obtained from the physical array. From the co-array perspective, sparse spatial sampling can provide a significant improvement in DOFs, and typical schemes of which include nested arrays [2] and co-prime arrays [3], [4]. Indeed, since the nested/co-prime arrays share a basic co-array configuration of *virtual* uniform linear array (ULA), their combinations with the standard MUSIC [10] have been extended to various scenarios, e.g., $2q$th-order cumulant statistics [5], two-dimensional arrays [6], [7], vector-sensor arrays [8] and wideband (WB) signals [9]. Most of them, however, resort to the spatial smoothing technique for establishing a rank-aware covariance matrix, thereby resulting in some aperture loss.

More recently, by vectorizing the covariance matrix, a series of sparse signal recovery (SSR) based approaches [11]–[18] have also been devised for DOA estimation, which have their roots in the sparse spectrum/covariance fitting criterion. These methods, explicitly or implicitly, become aware of the increased DOFs from co-array. For instance, to tackle the underdetermined case, the SSR based technique is tailored in [14] for nested arrays and in [15]–[18] for co-prime arrays, respectively. However, these techniques (albeit without aperture loss) concern only the narrowband signals and the literature for WB signals is less abundant. Although WB DOA estimation using SSR is not new (see, e.g., [19]–[21]), previous works seldom address the underdetermined case from the difference co-array perspective.

In this letter, inspired by the increased DOFs from the difference co-array in sparse spatial sampling, we reformulate the underdetermined DOA estimation problem for WB signals in the SSR framework. We first devise a unified *Gaussian-noise-aware* sparse model by prewhitening the perturbed errors resulting from finite samples (snapshots). Then an approach called **W**ideband **Sp**arse **S**pectrum **F**itting (W-SpSF) is developed using $\ell_{2,1}$-norm sparse regularization plus a weighted covariance fitting (least squares) criterion. In the W-SpSF estimator, the choice of the regularization parameter is also discussed to ensure robust sparse recovery. The W-SpSF approach has many merits such as without *a priori* knowledge on the source number, suppressing the spatial aliasing and being applicable to nonuniform noise. Simulation results corroborate the effectiveness of the W-SpSF.

II. Signal Model and Problem Statement

Consider a linear array of M sensors whose locations are given by the set $\mathcal{D} = \{d_m : m \in [M]\}$, $[M] \triangleq \{1, \ldots, M\}$. And suppose Q (possibly $Q \geq M$) WB signals impinging on this array from the far-field directions $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_Q]^T$ where $(\cdot)^T$ is the transpose. Each sensor signal after time-sampling is partitioned into L segments, and a K-point discrete Fourier transform (DFT) is applied to each segment. The WB array output can be modeled as [22], [23]:

$$\boldsymbol{x}_l(f_k) = \boldsymbol{A}_k(\boldsymbol{\theta})\boldsymbol{s}_l(f_k) + \boldsymbol{n}_l(f_k), k \in [K], l \in [L] \quad (1)$$

where $\boldsymbol{x}_l(f_k) \in \mathbb{C}^M$, $\boldsymbol{s}_l(f_k) \in \mathbb{C}^Q$ and $\boldsymbol{n}_l(f_k) \in \mathbb{C}^M$ are the DFT coefficients of the received data, source signals and additive noise, respectively, $\boldsymbol{A}_k(\boldsymbol{\theta}) \triangleq \boldsymbol{A}(f_k, \boldsymbol{\theta}) = [\boldsymbol{a}_k(\theta_1), \ldots, \boldsymbol{a}_k(\theta_Q)] \in$

Manuscript received June 24, 2014; revised September 01, 2014; accepted September 10, 2014. Date of publication September 16, 2014; date of current version October 13, 2014. This work was supported in part by the National Natural Science Foundation of China under Grants 61222106 and 61171187, and by the Shenzhen Kongqie talent program under Grant KQC201109020061A. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Phillip Ainsleigh.

Z.-Q. He and Z.-P. Shi are with the National Key Laboratory of Science and Technology on Communications, University of Electronic Science and Technology of China, Chengdu 611731, China.

L. Huang is with the Department of Electronic and Information Engineering, Harbin Institute of Technology Shenzhen Graduate School, Shenzhen 518055, China (e-mail: dr.lei.huang@ieee.org).

H. C. So is with the Department of Electronic Engineering, City University of Hong Kong, Hong Kong, China.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/LSP.2014.2358084

$\mathbb{C}^{M\times Q}$ is the steering matrix at the frequency bin f_k with the $M\times 1$ steering vector

$$\boldsymbol{a}_k(\theta_q)=\left[e^{\frac{-\jmath 2\pi f_k d_1\sin(\theta_q)}{c}},\ldots,e^{\frac{-\jmath 2\pi f_k d_M\sin(\theta_q)}{c}}\right]^T \tag{2}$$

where $\jmath=\sqrt{-1}$, and c is the propagation speed.

The signals and noises, within different frequency bins, are assumed to be complex circular *Gaussian* distributed and independent of each other [9]. Then the kth narrowband output covariance matrix is

$$\boldsymbol{R}_k=\mathbb{E}\{\boldsymbol{x}_l(f_k)\boldsymbol{x}_l^H(f_k)\}=\boldsymbol{A}_k(\boldsymbol{\theta})\boldsymbol{\mathcal{R}}_k\boldsymbol{A}_k^H(\boldsymbol{\theta})+\boldsymbol{\mathcal{Q}}_k \tag{3}$$

where $\mathbb{E}\{\cdot\}$ and $(\cdot)^H$ represent the expectation operator and conjugate transpose, respectively, $\boldsymbol{\mathcal{R}}_k=\mathbb{E}\{\boldsymbol{s}_l(f_k)\boldsymbol{s}_l^H(f_k)\}=\mathrm{diag}(\boldsymbol{\rho}_k)$ is the kth signal covariance matrix with $\boldsymbol{\rho}_k=[\rho_{k,1}^2,\ldots,\rho_{k,Q}^2]^T$ being the source power vector, and $\boldsymbol{\mathcal{Q}}_k=\mathbb{E}\{\boldsymbol{n}_l(f_k)\boldsymbol{n}_l^H(f_k)\}=\mathrm{diag}(\boldsymbol{\sigma}_k)$ is the kth noise covariance matrix with $\boldsymbol{\sigma}_k=[\sigma_{k,1}^2,\ldots,\sigma_{k,M}^2]^T$ being the noise power vector. Vectorizing (3), we get

$$\boldsymbol{\upsilon}_k\triangleq\mathrm{vec}(\boldsymbol{R}_k)=\boldsymbol{\mathcal{A}}_k(\boldsymbol{\theta})\boldsymbol{\rho}_k+\mathbb{1}_k,k\in[K] \tag{4}$$

where $\mathbb{1}_k=[\sigma_{k,1}^2\boldsymbol{e}_1^T,\ldots,\sigma_{k,M}^2\boldsymbol{e}_M^T]^T$, $\boldsymbol{e}_m$ is the mth vector in the canonical basis of $\mathbb{R}^M$, $\boldsymbol{\mathcal{A}}_k(\boldsymbol{\theta})\triangleq\boldsymbol{A}_k^*(\boldsymbol{\theta})\odot\boldsymbol{A}_k(\boldsymbol{\theta})=[\boldsymbol{a}_k^*(\theta_1)\otimes\boldsymbol{a}_k(\theta_1),\ldots,\boldsymbol{a}_k^*(\theta_Q)\otimes\boldsymbol{a}_k(\theta_Q)]\in\mathbb{C}^{M^2\times Q}$ where $(\cdot)^*$, $\odot$ and $\otimes$ represent the complex conjugate, Khatri-Rao product and Kronecker product, respectively. The increased DOFs directly rely on the distinct rows of $\boldsymbol{\mathcal{A}}_k(\boldsymbol{\theta})$, which is equivalent to the cardinality of the location set, $\mathcal{D}_u=\{d_i-d_j:i,j\in[M]\}$, of the difference co-array [2]. The problem at hand is to determine $\boldsymbol{\theta}$ without the need of the knowledge on the source number Q and noise variances $\{\boldsymbol{\sigma}_k\}_{k=1}^K$.

III. Wideband DOA Estimation Using Robust Sparse Covariance Fitting

A. Noise Reduction and Prewhitening

Note that the noise-like component $\mathbb{1}_k$ in (4) consists of the unknown terms $\{\sigma_{k,m}^2\}_{m=1}^M$ and 0 otherwise. In fact, the indices of $\{\sigma_{k,m}^2\}_{m=1}^M$ in $\mathbb{1}_k$ correspond exactly with the zero locations of the difference co-array. Furthermore, these zero locations have M occurrences (i.e., redundancy) and inherently lead to only one DOF instead of M DOFs. To mitigate the effect of $\mathbb{1}_k$, we therefore eliminate the M entries of $\boldsymbol{\upsilon}_k$ indexed by the positions of $\{\sigma_{k,m}^2\}_{m=1}^M$ within $\mathbb{1}_k$. This can be achieved by the following linear operation:

$$\boldsymbol{y}_k\triangleq\boldsymbol{J}\boldsymbol{\upsilon}_k=\boldsymbol{B}_k(\boldsymbol{\theta})\boldsymbol{\rho}_k,k\in[K] \tag{5}$$

where $\boldsymbol{J}=[\boldsymbol{J}_1,\ldots,\boldsymbol{J}_{M-1}]^T\in\mathbb{R}^{M(M-1)\times M^2}$, $\boldsymbol{J}_m=[\bar{\boldsymbol{e}}_{m(M+1)-M+1},\ldots,\bar{\boldsymbol{e}}_{m(M+1)}]\in\mathbb{R}^{M^2\times M}$ with $\bar{\boldsymbol{e}}_i,i=m(M+1)-M+1,\ldots,m(M+1)$, being the ith vector in the canonical basis of $\mathbb{R}^{M^2}$, and $\boldsymbol{B}_k(\boldsymbol{\theta})=\boldsymbol{J}\boldsymbol{\mathcal{A}}_k(\boldsymbol{\theta})\in\mathbb{C}^{M(M-1)\times Q}$ is the *virtual* steering matrix.

Having the L available segments (frequency snapshots), $\boldsymbol{\upsilon}_k$ is estimated by $\check{\boldsymbol{\upsilon}}_k=\boldsymbol{\upsilon}_k+\Delta\boldsymbol{\upsilon}_k=\mathrm{vec}(\check{\boldsymbol{R}}_k)$ where $\check{\boldsymbol{R}}_k=\frac{1}{L}\sum_{l=1}^L\boldsymbol{x}_l(f_k)\boldsymbol{x}_l^H(f_k)$ and $\Delta\boldsymbol{\upsilon}_k=\check{\boldsymbol{\upsilon}}_k-\boldsymbol{\upsilon}_k$ is the estimation error. Accordingly, we have

$$\check{\boldsymbol{y}}_k=\boldsymbol{J}\check{\boldsymbol{\upsilon}}_k=\boldsymbol{y}_k+\Delta\boldsymbol{y}_k \tag{6}$$

where $\Delta\boldsymbol{y}_k=\boldsymbol{J}\Delta\boldsymbol{\upsilon}_k$ is the corresponding estimation error of $\boldsymbol{y}_k$. Actually, $\Delta\boldsymbol{\upsilon}_k$ is asymptotically (in L) complex normal ($\mathcal{CN}$) distributed (see, e.g., [21], [24], [25]), viz., $\Delta\boldsymbol{\upsilon}_k\sim\mathcal{CN}(\boldsymbol{0},\frac{1}{L}\boldsymbol{R}_k^T\otimes\boldsymbol{R}_k)$ which further leads to

$$\Delta\boldsymbol{y}_k\sim\mathcal{CN}\left(\boldsymbol{0},\boldsymbol{W}_k\right),\boldsymbol{W}_k\triangleq\frac{1}{L}\boldsymbol{J}\left(\boldsymbol{R}_k^T\otimes\boldsymbol{R}_k\right)\boldsymbol{J}^T \tag{7}$$

where $\boldsymbol{0}$ is a zero vector/matrix with proper dimension which, for simplicity, is determined from the context. Therefore, the error $\Delta\boldsymbol{y}_k$ in (6) can be, at least approximately, viewed as an additive spatially correlated (colored) *Gaussian* noise. To further alleviate this correlation, a prewhitening procedure is considered here, that is,

$$\hat{\boldsymbol{y}}_k=\boldsymbol{W}_k{}^{-\frac{1}{2}}\check{\boldsymbol{y}}_k=\boldsymbol{W}_k{}^{-\frac{1}{2}}\boldsymbol{B}_k(\boldsymbol{\theta})\boldsymbol{\rho}_k+\boldsymbol{\epsilon}_k,k\in[K] \tag{8}$$

where $\boldsymbol{\epsilon}_k\triangleq\boldsymbol{W}_k{}^{-\frac{1}{2}}\Delta\boldsymbol{y}_k\sim\mathcal{CN}(\boldsymbol{0},\boldsymbol{I}_{M(M-1)})$, and $\boldsymbol{I}_{M(M-1)}$ is the $M(M-1)\times M(M-1)$ identity matrix.

The new observation (8), reminiscent of the initial model (1), can be viewed as a long received model with standard *Gaussian* noise. In the following, we will adopt (8) to address the problem of WB DOA estimation because compared to the ideal model (4), it is not corrupted by the noise variances anymore, thereby allowing easy access to sparse modeling without introducing additional parameters.

B. Wideband Sparse Spectrum Fitting

Following sparse localization framework (see, e.g., [12]–[21]), all K vectors $\{\hat{\boldsymbol{y}}_k\}_{k=1}^K$ in (8) can be readily rewritten as a unified *noise-aware* sparse model

$$\hat{\boldsymbol{y}}=\boldsymbol{\Phi}(\boldsymbol{\Omega})\boldsymbol{p}^\circ+\boldsymbol{\epsilon}. \tag{9}$$

Here, $\hat{\boldsymbol{y}}\triangleq[\hat{\boldsymbol{y}}_1^T,\ldots,\hat{\boldsymbol{y}}_K^T]^T$, $\boldsymbol{\epsilon}\triangleq[\boldsymbol{\epsilon}_1^T,\ldots,\boldsymbol{\epsilon}_K^T]^T$, $\boldsymbol{\Phi}(\boldsymbol{\Omega})\triangleq\mathrm{blkdiag}\{\boldsymbol{W}_1^{-\frac{1}{2}}\boldsymbol{B}_1(\boldsymbol{\Omega}),\ldots,\boldsymbol{W}_K^{-\frac{1}{2}}\boldsymbol{B}_K(\boldsymbol{\Omega})\}$ where $\mathrm{blkdiag}\{\cdot\}$ denotes the block diagonal operator is the overcomplete basis in which each column vector is normalized by its ℓ_2-norm, $\boldsymbol{\Omega}\triangleq\{\Omega_i:i\in[N]\}$ is the sampling grid set covering all the potential direction domain, and $\boldsymbol{p}^\circ=[(\boldsymbol{p}_1^\circ)^T,\ldots,(\boldsymbol{p}_K^\circ)^T]^T$ where $\boldsymbol{p}_k^\circ\in\mathbb{R}^N$ is the kth narrowband sparse spatial spectrum. When $N\gg Q$ such that $\{\theta_q:q\in[Q]\}\subset\boldsymbol{\Omega}$, $\{\boldsymbol{p}_k^\circ\}_{k=1}^K$ share the same sparse support (i.e., nonzero index) and hence $\boldsymbol{P}^\circ\triangleq[\boldsymbol{p}_1^\circ,\ldots,\boldsymbol{p}_K^\circ]$ is a *joint* row Q-sparse matrix. That is, the sparsity of $\boldsymbol{P}^\circ$ (or $\boldsymbol{p}^\circ$) behaves only in the spatial domain instead of the frequency domain. This motivates us to employ the prevalent $\ell_{2,1}$-norm minimization formulation for *joint* sparse recovery [19], [26]–[28]:

$$\min_{\boldsymbol{p}}\frac{1}{2}\|\hat{\boldsymbol{y}}-\boldsymbol{\Phi}(\boldsymbol{\Omega})\boldsymbol{p}\|_2^2+\tau\|\boldsymbol{p}\|_{2,1},\boldsymbol{p}\succeq\boldsymbol{0} \tag{10}$$

where $\|\boldsymbol{p}\|_{2,1}\triangleq\sum_{i=1}^N\|\boldsymbol{p}^i\|_2$ is the $\ell_{2,1}$-norm of $\boldsymbol{p}$, $\boldsymbol{p}^i=[p_i,p_{N+i},\ldots,p_{(K-1)N+i}]^T\in\mathbb{R}^K$ (p_i is the ith entry of $\boldsymbol{p}$), $\succeq$ denotes $\geq$ with an elementwise operation, and τ is a nonnegative regularization parameter. The convex formulation (10), also referred to as the **W**ideband **Sp**arse **S**pectrum **F**itting (W-SpSF) estimator[1], can be solved by interior-point algorithm based software packages such as SeDuMi [29] and CVX [30], with a complexity of $\mathcal{O}((NK)^3)$. Actually, the quadratic term of (10), i.e., $\frac{1}{2}\|\hat{\boldsymbol{y}}-\boldsymbol{\Phi}(\boldsymbol{\Omega})\boldsymbol{p}\|_2^2=\frac{1}{2}\sum_{k=1}^K\|\boldsymbol{W}_k^{-\frac{1}{2}}[\hat{\boldsymbol{y}}_k-\boldsymbol{B}_k(\boldsymbol{\Omega})\boldsymbol{p}_k]\|_2^2$, is essentially a weighted least squares fitting criterion which can

[1]Here we use its rudiment from the narrowband SpSF estimator [13].

induce an asymptotically unbiased estimator [31] and hence is desirable to debias the solution. The sparse-inducing solution of (10), however, is highly sensitive to τ as this parameter balances the sparsity and the data fidelity (least squares term) and its poor choice may lead to a non-sparse solution. Therefore, an appropriate selection of τ is required to guarantee robust sparse recovery, which will be investigated in the next subsection.

Remark on Parameter Identifiability: Based on Corollary 1 of [13], the maximum number of sources, which can be uniquely identified by W-SpSF, is $\frac{\tilde{N}}{2K}$ if any $\tilde{N}$ columns of $\boldsymbol{\Phi}(\boldsymbol{\Omega})$ are linearly independent (assume here that τ has been well chosen to promote sparse solution). With no ambiguity in $\boldsymbol{\Phi}(\boldsymbol{\Omega})$, we have $\tilde{N} = K \cdot \mathrm{DOF}_{\mathcal{A}}$ where $\mathrm{DOF}_{\mathcal{A}}$ denotes the DOFs of $\boldsymbol{\mathcal{A}}_k(\boldsymbol{\theta})$. This indicates that more degrees mean better identifiability, which agrees with our intuition. As a result, the proposed W-SpSF estimator is able to accommodate the enhanced DOFs from arbitrary array geometries (e.g., nested arrays [2], co-prime arrays [3], [4] and spatial compressive sensing based random arrays [32]) to improve its parameter identifiability considerably.

C. Regularization Parameter Selection

Here we analyze how to choose a proper τ via duality theory. Unlike [26], [27] which determine an appropriate choice of τ by revisiting the minimization (10) from its dual maximization, we address the selection of τ directly via the Karush-Kuhn-Tucker (KKT) conditions [33], [34], under an additional non-negative constraint. We start with the real-valued reformulation of problem (10):

$$\min_{\boldsymbol{p},\boldsymbol{z}} \frac{1}{2}\|\boldsymbol{z}\|_2^2 + \tau \sum_{i=1}^{N} \|\boldsymbol{p}^i\|_2 \quad \text{s.t.} \quad \boldsymbol{z} = \hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}, \boldsymbol{p} \succeq \boldsymbol{0} \tag{11}$$

where $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega}) = [\mathrm{Re}(\boldsymbol{\Phi}^T(\boldsymbol{\Omega})), \mathrm{Im}(\boldsymbol{\Phi}^T(\boldsymbol{\Omega}))]^T$, $\hat{\boldsymbol{z}} = [\mathrm{Re}(\hat{\boldsymbol{y}}^T), \mathrm{Im}(\hat{\boldsymbol{y}}^T)]^T$, $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ represent the real and imaginary parts, respectively. The associated *Lagrangian* is

$$\begin{aligned}&\frac{1}{2}\|\boldsymbol{z}\|_2^2 + \tau \sum_{i=1}^{N} \|\boldsymbol{p}^i\|_2 + \boldsymbol{\mu}^T[\boldsymbol{z} - \hat{\boldsymbol{z}} + \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}] - \boldsymbol{\lambda}^T\boldsymbol{p} \\ &\quad = \underbrace{\frac{1}{2}\|\boldsymbol{z}\|_2^2 + \boldsymbol{\mu}^T(\boldsymbol{z} - \hat{\boldsymbol{z}})}_{\mathcal{L}(\boldsymbol{z})} + \sum_{i=1}^{N} h_i(\boldsymbol{p}^i)\end{aligned}$$

where $\boldsymbol{\mu}$ and $\boldsymbol{\lambda}$ are vectors of *Lagrangian multipliers* associated with $\boldsymbol{z} = \hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}$ and $\boldsymbol{p} \succeq \boldsymbol{0}$, respectively, $h_i(\boldsymbol{p}^i) \triangleq \tau\|\boldsymbol{p}^i\|_2 + (\boldsymbol{\eta}^i - \boldsymbol{\lambda}^i)^T\boldsymbol{p}^i$, $\boldsymbol{\eta}^i \triangleq \tilde{\boldsymbol{\Phi}}_i^T(\boldsymbol{\Omega})\boldsymbol{\mu}$, $\boldsymbol{\lambda}^i$ and $\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})$ are sub-vector and sub-matrix extracted, respectively, from $\boldsymbol{\lambda}$ and columns of $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})$, corresponding to $\boldsymbol{p}^i$ within $\boldsymbol{p}$.

By the KKT conditions [33], [34], we get

$$\boldsymbol{z} = -\boldsymbol{\mu}, \tag{12a}$$

$$\boldsymbol{0} \in \partial h_i(\boldsymbol{p}^i), i \in [N], \tag{12b}$$

$$\boldsymbol{z} = \hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}, \tag{12c}$$

$$\boldsymbol{p} \succeq \boldsymbol{0}, \boldsymbol{\lambda} \succeq \boldsymbol{0},$$

$$p_{i\times j}\lambda_{i\times j} = 0, i \in [N], j \in [K], \tag{12d}$$

where $\partial h_i(\boldsymbol{p}^i)$ denotes the subgradient set of h_i at $\boldsymbol{p}^i$ (see [34] for a definition), $p_{i\times j}$ and $\lambda_{i\times j}$ represent the jth element of $\boldsymbol{p}^i$ and $\boldsymbol{\lambda}^i$, respectively. The condition (12a) is due to the gradient operation: $\nabla\mathcal{L}(\boldsymbol{z}) = \boldsymbol{0}$. In addition, $\boldsymbol{0} \in \partial h_i(\boldsymbol{p}^i)$ means that $h_i(\boldsymbol{p}^i)$ reaches its minimum at $\boldsymbol{p}^i$, for which the following lemma provides a necessary and sufficient condition.

Lemma 1: The convex function, $h_i(\boldsymbol{p}^i)$, has a minimum value $(> -\infty)$ *if and only if* $(\boldsymbol{\lambda}^i - \boldsymbol{\eta}^i)^T\boldsymbol{p}^i \le \tau\|\boldsymbol{p}^i\|_2$.

Proof: Note that $h_i(\boldsymbol{p}^i)$ consists of a *cone (conical surface)* $\tau\|\boldsymbol{p}^i\|_2$ and a hyperplane $(\boldsymbol{\eta}^i - \boldsymbol{\lambda}^i)^T\boldsymbol{p}^i$, both passing through the origin. Thus, the necessary and sufficient condition for $h_i(\boldsymbol{p}_i)$ reaching its minimum $(\neq -\infty)$ is that the hyperplane, $(\boldsymbol{\eta}^i - \boldsymbol{\lambda}^i)^T\boldsymbol{p}^i$, must be located below the *cone*, $\tau\|\boldsymbol{p}^i\|_2$, i.e., $h_i(\boldsymbol{p}^i) \ge 0$ and hence $(\boldsymbol{\lambda}^i - \boldsymbol{\eta}^i)^T\boldsymbol{p}^i \le \tau\|\boldsymbol{p}^i\|_2$. ∎

Combing Lemma 1, (12a), (12c) and (12d), the condition (12b) can be readily simplified as

$$\tau\|\boldsymbol{p}^i\|_2 \ge [\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}]^T\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})\boldsymbol{p}^i, i \in [N]. \tag{13}$$

Since $\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}$ is equivalent to the random noise $\tilde{\boldsymbol{\epsilon}} \triangleq [\mathrm{Re}(\boldsymbol{\epsilon}^T), \mathrm{Im}(\boldsymbol{\epsilon}^T)]^T$, τ should be chosen such that the inequality (13) holds with high probability as nearly a sure event. The condition (13), however, involves the unknown parameter $\boldsymbol{p}^i$ (or $\|\boldsymbol{p}^i\|_2$), which raises difficulty for choosing a proper τ. To circumvent this issue, we introduce the following theorem which characterizes the condition (13) in a probabilistic sense and hence provides a turning point for the regularization parameter selection.

Theorem 1: Denote the jth column of $\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})$ by $\tilde{\boldsymbol{\phi}}_{i\times j}$, and define $2N$ events:

$$\begin{aligned}&\Delta_i : [\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}]^T\tilde{\boldsymbol{\Phi}}_i^T(\boldsymbol{\Omega})\boldsymbol{p}^i \le \tau\|\boldsymbol{p}^i\|_2,\ i \in [N], \\ &\Delta'_i : \sum_{j=1}^{K}\left[\tilde{\boldsymbol{\phi}}_{i\times j}^T(\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p})\right]^2 \le \tau^2, i \in [N],\end{aligned} \tag{14}$$

then $\Pr(\Delta'_i) \le \Pr(\Delta_i), i \in [N]$, where $\Pr(\cdot)$ represents the probability of event.

Proof: Using the *Cauchy–Schwarz* inequality $[\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}]^T\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})\boldsymbol{p}^i \le \|[\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}]^T\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})\|_2\|\boldsymbol{p}^i\|_2$, we obtain

$$\begin{aligned}\tau^2 &\ge \left\|[\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}]^T\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})\right\|_2^2 \\ &= \sum_{j=1}^{K}\left[\tilde{\boldsymbol{\phi}}_{i\times j}^T(\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p})\right]^2 \\ &\Leftrightarrow \tau\|\boldsymbol{p}^i\|_2 \ge \|[\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}]^T\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})\|_2\|\boldsymbol{p}^i\|_2 \\ &\Rightarrow \tau\|\boldsymbol{p}^i\|_2 \ge [\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}]^T\tilde{\boldsymbol{\Phi}}_i(\boldsymbol{\Omega})\boldsymbol{p}^i,\end{aligned}$$

i.e., $\Delta'_i \Rightarrow \Delta_i$ and hence $\Delta'_i \subset \Delta_i$, which ultimately leads to $\Pr(\Delta'_i) \le \Pr(\Delta_i),\ i \in [N]$. ∎

From Theorem 1, if $\Pr(\Delta'_i) \to 1$, then $\Pr(\Delta_i) \to 1$. Therefore, we select the parameter τ such that the N events, $\{\Delta'_i : i \in [N]\}$, i.e., all inequalities in (14), hold with high probability. Note that the signals and noises are statistically independent of each other. We thus have $\mathbb{E}\{\boldsymbol{\epsilon}_{k_1}\boldsymbol{\epsilon}_{k_2}^H\} = \boldsymbol{0}, k_1 \ne k_2, k_1, k_2 \in [K]$, and hence $\boldsymbol{\epsilon} \sim \mathcal{CN}(\boldsymbol{0}, \boldsymbol{I}_{KM(M-1)})$ which further yields $\tilde{\boldsymbol{\epsilon}} = [\mathrm{Re}(\boldsymbol{\epsilon}^T), \mathrm{Im}(\boldsymbol{\epsilon}^T)]^T \sim \mathcal{N}(\boldsymbol{0}, \frac{1}{2}\boldsymbol{I}_{2KM(M-1)})$ where $\mathcal{N}(\cdot)$ denotes the normal distribution. Recall that all the column vectors of $\tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})$ have been normalized to one. We naturally have $\tilde{\boldsymbol{\phi}}_{i\times j}^T(\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p}) \sim \mathcal{N}(0, \frac{1}{2})$ and $2\sum_{j=1}^{K}[\tilde{\boldsymbol{\phi}}_{i\times j}^T(\hat{\boldsymbol{z}} - \tilde{\boldsymbol{\Phi}}(\boldsymbol{\Omega})\boldsymbol{p})]^2 \sim \chi^2(K)$, where $\chi^2(K)$ denotes the chi-square distribution with K DOFs. For simplicity, we simultaneously set all N independent events $\{\Delta'_i : i \in [N]\}$ with high probability, defined by P_τ,

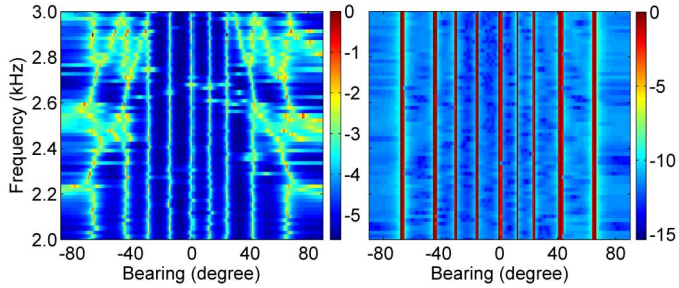

Fig. 1. Frequency-bearing images in uniform noise case. $M = 6, Q = 9, L = 400, K = 60$, SNR $= 0$ dB. Left: SS-MUSIC. Right: W-SpSF.

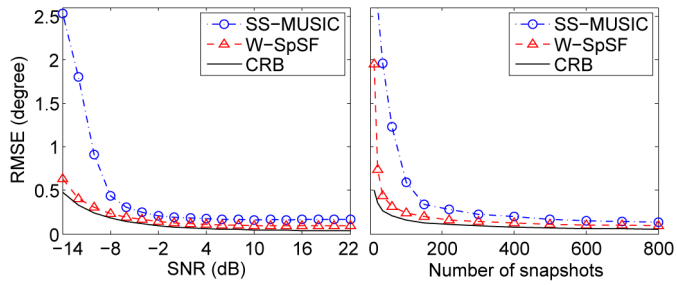

Fig. 2. RMSEs in nonuniform noise case. $M = 6, Q = 7, K = 10$. Left: $L = 400$. Right: SNR $= 0$ dB.

and each event Δ'_i with equal probability $P_\tau^{\frac{1}{N}}$. Eventually, the choice of τ is

$$\tau = \sqrt{\frac{\xi}{2}}, P_\tau^{\frac{1}{N}} \triangleq \Pr\left\{\mathcal{F}_{\chi^2}(K) \le \xi\right\} \tag{15}$$

where $\mathcal{F}_{\chi^2}(K)$ denotes the cumulative distribution function of $\chi^2(K)$.

IV. Simulation Results

In our simulations, the signals and noises within different frequency bins are assumed to be complex *Gaussian* distributed and with power vectors $\boldsymbol{\rho} = [\rho_1^2, \ldots, \rho_Q^2]^T$ and $\boldsymbol{\sigma} = [\sigma_1^2, \ldots, \sigma_M^2]^T$, respectively. Accordingly, the WB stochastic Cramér-Rao bound (CRB) is derived in the Appendix, which can be applied to the underdetermined case. The ideal weighted matrix $\boldsymbol{W}_k$, defined in (7), is replaced by $\check{\boldsymbol{W}}_k \triangleq \frac{1}{L}\boldsymbol{J}(\check{\boldsymbol{R}}_k^T \otimes \check{\boldsymbol{R}}_k)\boldsymbol{J}^T$. We further assume $\rho_1^2 = \cdots = \rho_Q^2 = \rho^2$ and define the signal-to-noise ratio (SNR) as SNR $\triangleq \frac{\rho^2}{M}\sum_{m=1}^{M}\frac{1}{\sigma_m^2}$ [35]. Throughout the simulations, we consider a 2-level nested array of 6 sensors with locations $\mathcal{D} = \{0, d, 2d, 3d, 7d, 11d\}$ where d is the basic element-spacing. The probability P_τ is set to be 0.999. In addition, $c = 340$ m/s and each signal has the common center frequency of $f_C = 2.5$ kHz.

Case 1: Aliasing-free Test. In this case, there are $Q = 9$ signals with a common bandwidth of 1 kHz from directions $[-68°, -45°, -30°, -15°, 0°, 12°, 24°, 42°, 65°]$. The uniform noise case with variance 10 is considered here and the grid spacing is set as 1° within $[-90°, 90°]$. The basic element spacing, d, is set as 8 cm, i.e., the half-wavelength of 2.125 kHz. As such, the spatial aliasing may occur when $f > 2.125$ kHz for the SS-MUSIC [9], which, however, can be well suppressed by the W-SpSF approach as long as d is less than the half-wavelength of the lowest frequency [20, Theorem 2]. The frequency-bearing images are plotted in Fig. 1 at $L = 400$, $K = 60$ and SNR $= 0$ dB. It is seen that spatial aliasing appears in the SS-MUSIC (left subplot) when $f > 2.125$ kHz, but no aliasing is observed in the W-SpSF estimator (right subplot).

Case 2: Estimation Accuracy. In this case, we examine the accuracy of W-SpSF in terms of root mean square error (RMSE) by averaging the results of 500 independent trials. We consider the nonuniform noise case with $\boldsymbol{\sigma} = [9.8, 6.5, 10.2, 8.4, 12.6, 7.5]^T$, and set $d = 5$ cm to avoid the spatial aliasing. Assume that $Q = 7$ signals with the same bandwidth as in Case 1 impinging from directions $[-45°, -30°, -15°, 0°, 10°, 24°, 40°]$. To improve the estimation precision, we set a finer grid spacing of 0.1° around the estimated peaks, but the initial grid remains unchanged as in Case 1. In Fig. 2, we plot the RMSEs versus the SNR with $L = 400$ (left subplot) and versus the number of snapshots with SNR $= 0$ dB (right subplot), under a common condition of $K = 10$. It is observed that the proposed W-SpSF estimator has a lower RMSE which is much closer to the CRB than that of the SS-MUSIC.

V. Conclusion

This letter presents the **W**ideband **Sp**arse **S**pectrum **F**itting (W-SpSF) estimator which can efficiently accommodate the increased DOFs from sparse spatial sampling to perform underdetermined DOA estimation. To guarantee robust sparse recovery against the parameter sensitivity of W-SpSF, the best regularization parameter is obtained by the KKT conditions. Theoretical analysis and simulation results illustrate that W-SpSF exhibits excellent estimation performance and is superior to the SS-MUSIC. Nevertheless, due to modeling error (outlier) W-SpSF cannot provide reliable DOA estimation for correlated signals, which will be addressed as our future work.

Appendix
Stochastic CRB for Wideband Signals

Stack all the array outputs $\{\boldsymbol{x}_l(f_k)\}_{k=1}^K$ as a vector $\boldsymbol{x}_l \triangleq [\boldsymbol{x}_l^T(f_1), \ldots, \boldsymbol{x}_l^T(f_K)]^T, l \in [L]$. With the statistic assumptions in Section II, the resultant covariance matrix of $\boldsymbol{x}_l$ is

$$\boldsymbol{R} = \mathbb{E}\{\boldsymbol{x}_l\boldsymbol{x}_l^H\} = \text{blkdiag}\{\boldsymbol{R}_1, \ldots, \boldsymbol{R}_K\}. \tag{16}$$

Define the parameter vector as $\boldsymbol{\psi} \triangleq [\boldsymbol{\theta}^T, \boldsymbol{\delta}^T]^T$, $\boldsymbol{\delta} \triangleq [\boldsymbol{\rho}^T, \boldsymbol{\sigma}^T]^T$. Recall that the CRB is the inverse of the Fisher information matrix (FIM). That is, we only need to derive its FIM. Using the block diagonal structure of $\boldsymbol{R}$, the (i,j)th entry of the associated FIM, denoted by $\boldsymbol{F}$, is given by [36]

$$\begin{aligned}[\boldsymbol{F}]_{i,j} &= L \cdot \text{tr}\left(\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \psi_i}\boldsymbol{R}^{-1}\frac{\partial \boldsymbol{R}}{\partial \psi_j}\right) \\ &= \sum_{k=1}^{K}\left[L \cdot \text{tr}\left(\boldsymbol{R}_k^{-1}\frac{\partial \boldsymbol{R}_k}{\partial \psi_i}\boldsymbol{R}_k^{-1}\frac{\partial \boldsymbol{R}_k}{\partial \psi_j}\right)\right] \\ &\triangleq \sum_{k=1}^{K}[\boldsymbol{F}_k]_{i,j},\ i,\ j \in [2Q+M]\end{aligned} \tag{17}$$

where $\text{tr}(\cdot)$ is the trace operator and ψ_i is the ith element of $\boldsymbol{\psi}$. Thus, the WB FIM, $\boldsymbol{F}$, is the sum of all the associated narrowband FIMs, $\{\boldsymbol{F}_k\}_{k=1}^K$. Then using the result in [14] on the narrowband FIM, we can easily obtain the relevant WB CRB.

REFERENCES

[1] W.-K. Ma, T.-H. Hsieh, and C.-Y. Chi, "Direction-of-arrival estimation of quasi-stationary signals with less sensors than sources and unknown spatial noise covariance: A Khatri-Rao subspace approach," *IEEE Trans. Signal Process.*, vol. 58, no. 4, pp. 2168–2180, Apr. 2010.

[2] P. Pal and P. P. Vaidyanathan, "Nested arrays: A novel approach to array processing with enhanced degrees of freedom," *IEEE Trans. Signal Process.*, vol. 58, no. 8, pp. 4167–4181, Aug. 2010.

[3] P. Pal and P. P. Vaidyanathan, "Coprime sampling and the MUSIC algorithm," in *Proc. 2011 IEEE Digital Signal Processing Workshop and IEEE Signal Processing Education Workshop (DSP/SPE)*, Sedona, AZ, USA, Jan. 2011, pp. 289–294.

[4] P. P. Vaidyanathan and P. Pal, "Sparse sensing with co-prime samplers and arrays," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 573–586, Feb. 2011.

[5] P. Pal and P. P. Vaidyanathan, "Multiple level lested array: An efficient geometry for $2q$th order cumulant based array processing," *IEEE Trans. Signal Process.*, vol. 60, no. 3, pp. 1253–1269, Mar. 2012.

[6] P. Pal and P. P. Vaidyanathan, "Nested arrays in two dimensions, Part I: Geometrical considerations," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4694–4705, Sep. 2012.

[7] P. Pal and P. P. Vaidyanathan, "Nested arrays in two dimensions, Part II: Application to two dimensional array processing," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4706–4718, Sep. 2012.

[8] K. Han and A. Nehorai, "Nested vector-sensor array processing via tensor modeling," *IEEE Trans. Signal Process.*, vol. 62, no. 10, pp. 2542–2553, May 2014.

[9] K. Han and A. Nehorai, "Wideband Gaussian source processing using a linear nested array," *IEEE Signal Process. Lett.*, vol. 20, no. 11, pp. 1110–1113, Nov. 2013.

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

[11] L. Blanco and M. Nájar, "Sparse covariance fitting for direction of arrival estimation," *EURASIP J. Adv. Signal Process.*, vol. 2012, no. 1, pp. 1–11, May 2012.

[12] N. Hu, Z. Ye, X. Xu, and M. Bao, "DOA estimation for sparse array via sparse signal reconstruction," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 2, pp. 760–773, Apr. 2013.

[13] J. Zheng and M. Kaveh, "Sparse spatial spectral estimation: A covariance fitting algorithm, performance and regularization," *IEEE Trans. Signal Process.*, vol. 61, no. 11, pp. 2767–2777, Jun. 2013.

[14] Z.-Q. He, Z.-P. Shi, and L. Huang, "Covariance sparsity-aware DOA estimation for nonuniform noise," *Digit. Signal Process.*, vol. 28, pp. 75–81, May 2014.

[15] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, Vancouver, BC, Canada, May 2013, pp. 3967–3971.

[16] Z. Tan, Y. C. Eldar, and A. Nehorai, "Direction of arrival estimation using co-prime arrays: A super resolution viewpoint," *IEEE Trans. Signal Process.*, vol. 62, pp. 5565–5576, 2014.

[17] Y. D. Zhang, S. Qin, and M. G. Amin, "DOA estimation exploting coprime arrays with sparse sensor spacing," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Process. (ICASSP)*, Florence, Italy, May 2014, pp. 2267–2271.

[18] Z. Tan and A. Nehorai, "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," *IEEE Signal Process. Lett.*, vol. 2, no. 1, pp. 26–29, Jan. 2014.

[19] D. Malioutov, M. Çetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 3010–3022, Aug. 2005.

[20] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "Direction-of-arrival estimation of wideband signals via covariance matrix sparse representation," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4256–4270, Sep. 2011.

[21] Z.-M. Liu, Z.-T. Huang, and Y.-Y. Zhou, "Sparsity-inducing direction finding for narrowband and wideband signals based on array covariance vectors," *IEEE Trans. Wireless Commun.*, vol. 12, no. 8, pp. 3896–3907, Aug. 2013.

[22] M. A. Doron and A. J. Weiss, "On focusing matrices for wideband array processing," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1295–1302, Jun. 1992.

[23] T. S. Lee, "Efficient wideband source localization using beamforming invariance technique," *IEEE Trans. Signal Process.*, vol. 42, no. 6, pp. 1376–1387, Jun. 1994.

[24] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digit. Signal Process.*, vol. 8, pp. 185–210, Jul. 1998.

[25] H. Li, P. Stoica, and J. Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Trans. Signal Process.*, vol. 47, no. 5, pp. 1314–1323, May 1999.

[26] X. Xu, X. Wei, and Z. Ye, "DOA estimation based on sparse signal recovery utilizing weighted ℓ_1-norm penalty," *IEEE Signal Process. Lett.*, vol. 19, no. 3, pp. 155–158, Mar. 2012.

[27] X. Wei, Y. Yuan, and Q. Ling, "DOA estimation using a greedy block coordinate descent algorithm," *IEEE Trans. Signal Process.*, vol. 60, no. 12, pp. 6382–6394, Dec. 2012.

[28] J. Yin and T. Chen, "Direction-of-arrival estimation using a sparse representation of array covariance vectors," *IEEE Trans. Signal Process.*, vol. 59, no. 9, pp. 4489–4493, Sep. 2011.

[29] J. F. Sturm, "Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones," *Optim. Method Softw.*, vol. 11–12, pp. 625–653, Aug. 1999.

[30] M. Grant and S. Boyd, "CVX: Matlab software for disciplined convex programming, version 1.22," [Online]. Available: http://cvxr.com/cvx 2012

[31] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ, USA: Prentice-Hall, 1993.

[32] M. Rossi, A. Haimovich, and Y. C. Eldar, "Spatial compressive sensing for MIMO radar," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 419–430, Jan. 2014.

[33] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[34] S. Boyd, "Lecture Notes for EE364B: Convex optimization II (lectures)," 2007 [Online]. Available: http://www.stanford.edu/class/ee364b/

[35] M. Pesavento and A. B. 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.

[36] P. Stoica, E. G. Larsson, and A. B. Gershman, "The stochastic CRB for array processing: A textbook derivation," *IEEE Signal Process. Lett.*, vol. 8, no. 5, pp. 148–150, May 2001.