An Asymptotically Efficient Weighted Least Squares Estimator for Co-Array-Based DoA Estimation

Saeid Sedighi, *Student Member, IEEE,* M. R. Bhavani Shankar, *Senior Member, IEEE,*
and Björn Ottersten, *Fellow, IEEE*

Abstract—Co-array-based Direction of Arrival (DoA) estimation using Sparse Linear Arrays (SLAs) has recently gained considerable interest in array processing thanks to its capability of providing enhanced degrees of freedom. Although the literature presents a variety of estimators in this context, none of them are proven to be statistically efficient. This work introduces a novel estimator for the co-array-based DoA estimation employing the Weighted Least Squares (WLS) method. An analytical expression for the large sample performance of the proposed estimator is derived. Then, an optimal weighting is obtained so that the asymptotic performance of the proposed WLS estimator coincides with the Cramér-Rao Bound (CRB), thereby ensuring asymptotic statistical efficiency of resulting WLS estimator. This implies that the proposed WLS estimator has a significantly better performance compared to existing methods. Numerical simulations are provided to validate the analytical derivations and corroborate the improved performance.

Index Terms—Sparse linear arrays, direction of arrival (DoA) estimation, weighted least squares estimator, Cramér-Rao bound (CRB), performance analysis, asymptotically statistically efficient estimator.

I. INTRODUCTION

The problem of Direction of Arrival (DoA) estimation is of central importance in the field of array processing with many applications in radar, sonar, and wireless communications [2]–[4]. Estimating DoAs using Uniform Linear Arrays (ULAs) is well investigated in the literature; a number of algorithms such as Maximum Likelihood (ML) estimation, MUSIC, ESPRIT and subspace fitting have been presented and their performance thoroughly analyzed [5]–[8]. However, it is widely known that ULAs are not capable of identifying more sources than the number of physical elements in the array [3], [7].

To transcend this limitation, exploitation of Sparse Linear Arrays (SLAs) with particular geometries, such as Minimum Redundancy Arrays (MRAs) [9], co-prime arrays [10] and nested arrays [11] has been proposed. These architectures can dramatically boost the degrees of freedom of the array for uncorrelated source signals such that a significantly larger number of sources than the number of physical elements in the array can be identified. In addition, the enhanced degrees of freedom provided by these SLAs can improve the resolution performance appreciably compared to ULAs [11]. These features have spurred further research on DoA estimation using SLAs in recent years. A detailed study on DoA estimation via SLAs through an analysis of the Cramér-Rao Bound (CRB) has been conducted in [12]. The findings in [12] provide valuable insights into the performance limits of DoA estimation via SLAs and are useful for benchmarking. Further, a number of DoA estimators based on the difference co-array of SLAs have been proposed in the literature. In general, existing co-array-based estimators can be classified under two main groups: 1) Sparsity-Based Methods (SBMs); 2) Augmented Covariance-Based Methods (ACBMs).

SBMs first discretize the angular domain into a grid and then estimate DoAs by imposing sparsity constraints on source profiles and exploiting the compressive sensing recovery techniques [13]–[16]. Such estimators are susceptible to grid mismatch leading to significant performance degradation when DoAs do not lie on the predefined grid [17]. To alleviate this problem, the authors in [18] include a first order approximation of grid mismatch in the model through Taylor series expansion and then estimate DoAs and mismatch variables jointly. Although the algorithm proposed in [18] shows an improvement over that of conventional sparsity-based methods, it is still restricted by higher-order mismatch terms. The shortcoming of discretization approach is overcome by a grid-less sparsity-based algorithm in [19]. This algorithm, named as Sparse and Parametric Approach (SPA), employs the covariance fitting criteria and semidefinite programming. Although SPA does not suffer from grid mismatch, its asymptotic performance (for the large number of snapshots) is not guaranteed to achieve the CRB.

In the second approach, DoAs are estimated by applying conventional subspace methods such as MUSIC, ESPRIT on an Augmented Sample Covariance Matrix (ASCM) obtained from the original sample covariance matrix by exploiting the difference co-array structure [11], [20], [21]. Two different ways of constructing the ASCM are given in the literature, namely, 1) the direct augmentation approach [22]; 2) spatial smoothing approach [11]. Both approaches span identical subspaces resulting in the same estimation performance, but the former is computationally efficient. A main drawback of this family of algorithms is the need for prior knowledge on the exact number of sources, which may not be available in practice. On the contrary, SBMs need information on an

upper bound on the number of sources. The performance of such estimators, called Co-Array-Based MUSIC (CAB-MUSIC) and Co-Array-Based ESPRIT (CAB-ESPRIT), has been rigorously analyzed in [20] and [21], respectively. An existence of a considerable gap between their performance and the CRB is revealed when the number of sources is greater than one. Further, CAB-MUSIC and CAB-ESPRIT suffer from an inherent performance loss when applied to SLAs with holes in their difference co-arrays, such as co-prime arrays, since they discard the information contained in the non-contiguous segment of the difference co-array. To avoid this performance loss, array interpolation-based algorithms has been used for co-prime arrays in [23]–[25] where a convex optimization problem is formulated to recover the covariance matrix of a virtual ULA interpolated from the co-prime array. The performance of interpolation-based algorithms has been analyzed in [26]. In addition, the authors of this paper have recently proposed a least squares estimator capable of exploiting the information contained in the non-contiguous ULA segment of the difference co-array [27]. Although the proposed estimators in [23], [24], [27] succeed in dealing with the performance loss due to omission of the non-contiguous segment of difference co-array, they are still unable to achieve the CRB performance.

It is known that the covariance matrix of uncorrelated signals received by a linear array is structured, e.g, Hermitian Toeplitz for ULA [28]. The structure in the covariance matrix is shown to be highly beneficial in obtaining an enhanced covariance matrix estimate compared to the conventional sample covariance matrix [28]. This, in turn, could yield better DoA estimates through an application of MUSIC. While the discussions in [28] are restricted to ULAs, the approach can be straightforwardly used for SLAs to obtain an enhanced covariance matrix estimate and subsequently construct the ASCM. The performance of such an estimator, which we call it Structured CAB-MUSIC (SCAB-MUSIC), has been never investigated in the literature for SLAs. However, our results for this method, reported in the ensuing Section V, reveal that SCAB-MUSIC does not attain the CRB.

The performance gap between the estimators available in the literature and the CRB motivates the current work on designing an asymptotically statistically efficient estimator for co-array-based DoA estimation via SLAs. To close this gap, in this paper, we propose a Weighted Least Squares (WLS) approach to DoA estimation using SLAs. We analytically prove that the proposed approach can yield an estimator that asymptotically achieves the corresponding CRB for any SLA configuration. As a consequence, the resulting WLS estimator exhibits enhanced performance compared to the existing algorithms in the literature. Accordingly, the contributions of the paper are described as follows:

- For any given feasible weighting matrix, we formulate the WLS approach towards estimating the DoAs and the ancillary variables − source powers and noise variance.
- We first provide a consistent estimate of the noise variance which is applicable to the difference co-array model. Making use of this consistent noise variance estimate, we derive WLS estimates of the signal powers and concentrate the WLS objective on the DoAs.

- The proposed estimator is the minimizer of the aforementioned concentrated WLS objective. Key attributes of this WLS estimator are studied for any feasible weighting matrix by proving consistency, asymptotic unbiasedness and then deriving a closed-form expression for the asymptotic covariance matrix of DoA estimation errors.
- Considering the earlier asymptotic unbiasedness of the proposed WLS estimator and noting that the covariance matrix of any unbiased estimator is lower bounded by the CRB, the optimal weighting matrix should be the one that renders the resulting covariance matrix of DoA estimation errors and CRB identical. To derive this optimal weighting matrix, we reformulate the CRB expression given in [12] in a form suitable for establishing equality of the CRB and the covariance matrix of DoA estimation errors.
- The new expression for the CRB is exploited to analytically obtain the optimal weighting that results in the equivalence of the asymptotic performance of the proposed WLS estimator and the CRB.
- With the framework to obtain asymptotically efficient WLS estimate provided, we now consider the key aspect of implementing the minimization of the WLS objective that leads to the proposed WLS estimate. This, typically needs computationally complex minimization of a multimodal objective function. The quality of the solutions of the iterative algorithms used for minimizing such multimodal functions highly depends on the initialization such that the global minima potentially achieved in case a very good initial point, which is close enough to the global minima, is available. This motivates us to introduce two efficient algorithms for solving the optimization problem. The first algorithm is applicable to SLAs with hole-free co-arrays, such as MRA and nested arrays. This method recasts the optimization problem as a quadratic optimization problem followed by rooting a polynomial function. This leads to a significant reduction in computations, rendering the complexity of the proposed estimator comparable to that of the other techniques such as CAB-MUSIC, CAB-ESPIRIT and SCAB-MUSIC while the WLS estimator enjoys a better performance compared to them. The second algorithm can be used for SLAs with holes in their co-arrays such as co-prime arrays. This algorithm recasts the optimization problem as a polynomial optimization problem followed by rooting a polynomial function where the global minima of the introduced polynomial optimization problem is guaranteed to be attained by using Lasserre's Semidefinite Programming (SDP) relaxation given in [29].
- Further, we validate the analytical derivations through numerical simulations and compare the performance of the proposed WLS estimator with those proposed in the literature. Numerical results confirm asymptotic efficiency of the proposed WLS estimator and illustrate its superior performance in terms of estimation accuracy and resolution compared to the existing estimators in the literature.

Organization: Section II describes the co-array system

model. In Section III, the proposed WLS framework is presented and the form of the WLS estimates of DoAs is derived. The performance of the proposed WLS estimator is analytically evaluated and its asymptotic statistical efficiency is proved in Section IV. The simulation results and related discussions are included in Section V. Finally, Section VI concludes the paper.

Notation: Vectors and matrices are referred to by lower- and upper-case bold-face, respectively. The superscripts $*$, T, H denote the conjugate, transpose and Hermitian (conjugate transpose) operations, respectively. $\|\mathbf{a}\|_2$ stands for the ℓ_2-norm of $\mathbf{a}$. $[\mathbf{A}]_{i,j}$ and $[\mathbf{a}]_i$ indicate the $(i,j)^{\text{th}}$ and i^{th} entry of $\mathbf{A}$ and $\mathbf{a}$, respectively. $\hat{\mathbf{A}}$ and $\hat{\mathbf{a}}$ denote the estimate of $\mathbf{A}$ and $\mathbf{a}$, respectively. $(a_1, a_2, \cdots, a_n)$ is an n-tuple with elements of $a_1, a_2, \cdots, a_n$. $|\mathbb{A}|$ represents the cardinality of the set $\mathbb{A}$. $\text{diag}(\mathbf{a})$ is a diagonal matrix whose diagonal entries are equal to the elements of $\mathbf{a}$. The $M \times M$ identity matrix is denoted by $\mathbf{I}_M$. The real and image part of a are denoted by $\text{Re}\{a\}$ and $\text{Im}\{a\}$, respectively. $\mathbb{E}\{.\}$ stands for the statistical expectation. $\delta[.]$ is Kroneckers delta. $\otimes$ and $\odot$ represent Kronecker and Khatri-Rao products, respectively. $\text{tr}(\mathbf{A})$, $\text{rank}(\mathbf{A})$ and $\text{vec}(\mathbf{A}) = \begin{bmatrix} \mathbf{a}_1^T & \mathbf{a}_2^T & \cdots & \mathbf{a}_n^T \end{bmatrix}^T$ denote the trace, rank and vectorization operations, respectively. $\mathbf{A}^\dagger$ and $\Pi_{\mathbf{A}}^\perp$ indicate the pseudoinverse and the projection matrix onto the null space of the full column rank matrix $\mathbf{A}^H$, respectively. $\mathcal{R}(\mathbf{A}), \mathcal{N}(\mathbf{A})$, respectively, represent the range and null spaces of $\mathbf{A}$.

II. CO-ARRAY SYSTEM MODEL

We consider an SLA with M elements located at positions $\left(m_1 \frac{\lambda}{2}, \ m_2 \frac{\lambda}{2}, \cdots, m_M \frac{\lambda}{2}\right)$ with $m_i \in \mathbb{M}$. Here $\mathbb{M}$ is a set of integers with cardinality $|\mathbb{M}| = M$, and λ represents the wavelength of the incoming signals. It is assumed K narrowband signals with distinct DoAs $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 & \theta_2 & \cdots & \theta_K \end{bmatrix}^T$ impinge on the SLA from far field. Accordingly, the vector of signals received by the SLA at time instance t can be modeled as

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{x}(t) + \mathbf{n}(t) \in \mathbb{C}^{M \times 1}, \qquad t = 1, \cdots, N, \quad (1)$$

where $\mathbf{x}(t) \in \mathbb{C}^{K \times 1}$ denotes the vector of source signals, $\mathbf{n}(t) \in \mathbb{C}^{M \times 1}$ is additive noise, and $\mathbf{A}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{a}(\theta_1), & \mathbf{a}(\theta_2), & \cdots & \mathbf{a}(\theta_K) \end{bmatrix} \in \mathbb{C}^{M \times K}$ represents the SLA steering matrix where

$$\mathbf{a}(\theta_i) = \begin{bmatrix} e^{\text{j}\pi \sin \theta_i m_1} & e^{\text{j}\pi \sin \theta_i m_2} & \cdots & e^{\text{j}\pi \sin \theta_i m_M} \end{bmatrix}^T, \quad (2)$$

is the SLA manifold vector for the ith signal. Further, the following assumptions are made on source signals and noise:

A1 The noise vector follows a zero-mean circular complex Gaussian distribution with the covariance matrix, $\mathbb{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \sigma^2 \mathbf{I}_M$.

A2 The source signal vector is modeled as a zero-mean circular complex Gaussian random vector with covariance matrix $\mathbb{E}\{\mathbf{x}(t)\mathbf{x}^H(t)\} = \text{diag}(\mathbf{p})$ where $\mathbf{p} = \begin{bmatrix} p_1, & p_2, & \cdots & p_K \end{bmatrix}^T \in \mathbb{R}_{>0}^{K \times 1}$ (i. e., $p_l > 0 \ \forall l$).

A3 Source and noise vectors are mutually independent.

A4 There is no temporal correlation between the snapshots, i.e., $\mathbb{E}\{\mathbf{n}(t_1)\mathbf{n}^H(t_2)\} = \mathbb{E}\{\mathbf{x}(t_1)\mathbf{x}^H(t_2)\} = \mathbf{0}$ when $t_1 \neq t_2$ and $\mathbf{0}$ is an all zero matrix of appropriate dimensions.
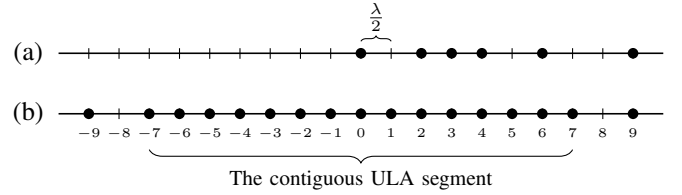

Fig. 1. Array geometry of a co-prime array with $M = 6$ elements: (a) physical array with $\mathbb{M} = \{0, 2, 3, 4, 6, 9\}$; (b) difference co-array with $\mathbb{D} = \{0, 1, 2, 3, 4, 5, 6, 7, 9\}$ and $v = 8$.

Based on the above assumptions, the covariance matrix of the received signals, i.e., $\mathbf{R} = \mathbb{E}\{\mathbf{y}(t)\mathbf{y}^H(t)\}$, is given by

$$\mathbf{R} = \mathbf{A}(\boldsymbol{\theta})\text{diag}(\mathbf{p})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2 \mathbf{I}_M \in \mathbb{C}^{M \times M}. \quad (3)$$

Following [11], [12], [20], the difference co-array model of the SLA is obtained by vectorizing the covariance matrix in (3), which results in

$$\mathbf{r} \doteq \text{vec}(\mathbf{R}) = \left(\mathbf{A}^*(\boldsymbol{\theta}) \odot \mathbf{A}(\boldsymbol{\theta})\right)\mathbf{p} + \sigma^2 \text{vec}(\mathbf{I}_M),$$
$$= \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} + \sigma^2 \mathbf{J}\mathbf{g} \in \mathbb{C}^{M^2 \times 1}, \quad (4)$$

where $\mathbf{A}_d(\boldsymbol{\theta}) \in \mathbb{C}^{(2D-1) \times K}$ corresponds to the steering matrix of the difference co-array whose elements are located at $(-\ell_{D-1}\frac{\lambda}{2}, \cdots, 0, \cdots, \ell_{D-1}\frac{\lambda}{2})$ with $\ell_i \in \mathbb{D} = \{|m_p - m_q| | m_p, m_q \in \mathbb{M}\}$ and $D = |\mathbb{D}|$. Further, $\mathbf{g} \in \{0, 1\}^{(2D-1) \times 1}$ is a column vector with $[\mathbf{g}]_i = \delta[i - D]$, and the selection matrix $\mathbf{J}$ is represented as follows

Definition 1. The binary matrix $\mathbf{J} \in \{0, 1\}^{M^2 \times (2D-1)}$ is defined as [12]

$$\mathbf{J} = \begin{bmatrix} \text{vec}(\mathbf{L}_{D-1}^T) & \cdots & \text{vec}(\mathbf{L}_0) & \cdots & \text{vec}(\mathbf{L}_{D-1}), \end{bmatrix}, \quad (5)$$

where $[\mathbf{L}_n]_{p,q} = \begin{cases} 1, & \text{if } m_p - m_q = \ell_n, \\ 0, & \text{otherwise}, \end{cases}$ with $1 \leq p, q \leq M$ and $0 \leq n \leq D - 1$.

The difference co-array model in (4) can be perceived to be the response of a virtual array whose steering matrix is given by $\mathbf{A}_d(\boldsymbol{\theta})$ to the parameter vector with signal powers $\mathbf{p}$ in presence of the noise vector $\sigma^2 \text{vec}(\mathbf{I}_M)$. This virtual array includes a contiguous ULA segment around the origin with the size of $2v - 1$ where v is the largest integer such that $\{0, 1, \cdots, v - 1\} \subseteq \mathbb{D}$. An illustrative example of an SLA, the corresponding difference co-array and its contiguous ULA segment is provided in Fig. 1. It has been shown in [10]–[12] that the size of the contiguous ULA segment of the difference co-array plays a crucial role in the number of identifiable sources such that K distinct sources are identifiable if $K \leq v - 1$. Hence, in case the SLA is designed properly such that $v > M$, we are able to identify more sources than the number of physical elements in the SLA, exploiting the source signal covariance matrix structure efficiently.

III. CO-ARRAY-BASED WLS ESTIMATOR

The problem under consideration is the estimation of the unknown parameters in (4) − DoAs, signal powers and the noise variance − using array observations, i.e., $\{\mathbf{y}(t)\}_{t=1}^N$. Of these, the DoAs are of primary interest and the other parameters are of subordinate interest. However, the estimation of the secondary parameters is essential for accurate DoA

estimation. In what follows, we first propose a heuristic, but, consistent estimate of the noise variance. We then derive the WLS estimates of DoAs and source signal powers exploiting the proposed consistent estimate of the noise variance.

A. Estimation of the Noise Variance

Let $\hat{\mathbf{R}}$ denote the sample covariance matrix, defined as

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=1}^{N} \mathbf{y}(t)\mathbf{y}^{H}(t) \in \mathbb{C}^{M \times M}, \tag{6}$$

and $\hat{\mathbf{r}} = \text{vec}(\hat{\mathbf{R}})$ denote its vectorized form. In addition, let $\hat{\mathbf{R}}_v$ be the augmented sample covariance matrix, which is constructed as follows [22]

$$\hat{\mathbf{R}}_v = \begin{bmatrix} \mathbf{T}_v\mathbf{J}^{\dagger}\hat{\mathbf{r}} & \mathbf{T}_{v-1}\mathbf{J}^{\dagger}\hat{\mathbf{r}} & \cdots & \mathbf{T}_1\mathbf{J}^{\dagger}\hat{\mathbf{r}} \end{bmatrix} \in \mathbb{C}^{v \times v}, \tag{7}$$

where $\mathbf{T}_i$ is a selection matrix defined as

$$\mathbf{T}_i = \begin{bmatrix} \mathbf{0}_{v \times (i+D-v-1)} & \mathbf{I}_v & \mathbf{0}_{v \times (D-i)} \end{bmatrix} \in \{0,1\}^{v \times (2D-1)}, \tag{8}$$

Then, we are able to obtain a consistent estimate of the noise variance, as stated in the following lemma.

Lemma 1. *If $K \leq v - 1$, a consistent estimate of the noise variance is given by*

$$\widehat{\sigma^2} = \frac{\text{vec}^{H}(\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^{H})\mathbf{T}\mathbf{J}^{\dagger}\hat{\mathbf{r}}}{v - K}, \tag{9}$$

where $\hat{\mathbf{U}}_n$ represents the eigenvectors of the augmented sample covariance matrix $\hat{\mathbf{R}}_v$ corresponding to its $v - K$ smallest eigenvalues and $\mathbf{T} = \begin{bmatrix} \mathbf{T}_v^{T} & \mathbf{T}_{v-1}^{T} & \cdots & \mathbf{T}_1^{T} \end{bmatrix}^{T} \in \mathbb{C}^{v^2 \times (2D-1)}$.

Proof. See Appendix A □

Remark 1 (Efficiency of the Noise Estimate). It can be demonstrated that $\widehat{\sigma^2}$, while being consistent, is not statistically efficient, meaning that $\frac{\mathbb{E}\{(\widehat{\sigma^2} - \sigma^2)^2\}}{CRB(\sigma^2)} > 1$ where $CRB(\sigma^2)$ represents the CRB of σ^2. We will not dwell on this further since estimation of σ^2 is not the main aim here. A consistent estimate of σ^2 suffices for our purpose.

Remark 2. Employing Lemma 7 in Appendix G, it can easily be shown that $\widehat{\sigma^2}^{*} = \widehat{\sigma^2}$, implying that $\widehat{\sigma^2}$ is real-valued. It is not, however, guaranteed to be positive for a small number of snapshots. Nonetheless, considering the fact that $\widehat{\sigma^2}$ is a consistent estimate of $\sigma^2 > 0$ ensures that $\widehat{\sigma^2}$ is positive when the number of snapshots is adequately large. As a consequence, the asymptotic performance of the DoA estimator, which will be introduced in the next subsection, will not be affected.

B. WLS Estimates of DoAs

To estimate source DoAs from (4), it is possible to formulate the co-array-based LS estimates of $\boldsymbol{\theta}$, $\mathbf{p}$ as

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_{ls} \\ \hat{\mathbf{p}}_{ls} \end{bmatrix} = \underset{\boldsymbol{\theta},\mathbf{p}}{\arg\min} \|\hat{\mathbf{r}} - \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} - \widehat{\sigma^2}\text{vec}(\mathbf{I}_M)\|_2^2. \tag{10}$$

However, our investigations, presented in [27], indicate that the LS estimates of DoAs do not show a significant performance improvement in terms of MSE compared to the existing

algorithms. Thus it would be useful to introduce a weighting in the above criterion to achieve better performance. Hence, we propose the following WLS estimator instead

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_{wls} \\ \hat{\mathbf{p}}_{wls} \end{bmatrix} = \underset{\boldsymbol{\theta},\mathbf{p}}{\arg\min} \left\| \mathbf{W}^{\frac{1}{2}} \left(\hat{\mathbf{r}} - \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p} - \widehat{\sigma^2}\text{vec}(\mathbf{I}_M) \right) \right\|_2^2. \tag{11}$$

where $\mathbf{W}$ is a positive definite weighting matrix. The weighting matrix $\mathbf{W}$ should be determined to minimize the MSE of DoA estimates. For the time being, we defer problem of finding the optimal weighting matrix until Section IV-B and proceed with the derivation of the WLS estimator for DoAs.

Inserting (9) into (11) and performing certain standard algebraic manipulations leads to

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_{wls} \\ \hat{\mathbf{p}}_{wls} \end{bmatrix} = \underset{\boldsymbol{\theta},\mathbf{p}}{\arg\min} \|\mathbf{W}^{\frac{1}{2}}(\hat{\mathbf{Q}}\hat{\mathbf{r}} - \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p})\|_2^2, \text{ where} \tag{12}$$

$$\hat{\mathbf{Q}} \doteq \mathbf{I}_{M^2} - \frac{\text{vec}(\mathbf{I}_M)\text{vec}^{H}(\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^{H})\mathbf{T}\mathbf{J}^{\dagger}}{v - K}. \tag{13}$$

Solving (12) with respect to $\mathbf{p}$ yields

$$\hat{\mathbf{p}}_{wls} = \left(\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) \right)^{\dagger} \mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}. \tag{14}$$

Remark 3 (Consistency of Signal Power Estimates). Based on (48) and (47) in Appendix B, it is readily deduced that $\hat{\mathbf{p}}_{wls}$ is a consistent estimator of $\mathbf{p}$ iff $\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})$ has full column rank. Clearly, $\mathbf{W}$ is positive definite by definition and it was shown in [12] that $\mathbf{J}$ has full column rank. These imply that $\hat{\mathbf{p}}_{wls}$ is consistent iff $\mathbf{A}_d(\boldsymbol{\theta})$ has full column rank. Following the same approach used in the proof of Lemma 3 in Appendix B, it can be shown that $\mathbf{A}_d(\boldsymbol{\theta})$ has full column rank if $K \leq 2v - 1$. This condition is weaker than the identifiability condition, i.e., $K \leq v - 1$, given in the literature [10]–[12]. Hence, the consistency of $\hat{\mathbf{p}}_{wls}$ is guaranteed in practice.

Remark 4. Making use of Lemma 8 in Appendix G, it can be shown that $\hat{\mathbf{p}}_{wls}$ is a real number, i.e., $\hat{\mathbf{p}}_{wls}^{*} = \hat{\mathbf{p}}_{wls}$. Hence, the same considerations mentioned in Remarks 2 for $\widehat{\sigma^2}$ are applicable to $\hat{\mathbf{p}}_{wls}$ as well.

Finally, inserting (14) into (12), concentrates the WLS objective on DoAs and the WLS estimator of $\boldsymbol{\theta}$ follows as

$$\hat{\boldsymbol{\theta}}_{wls} = \underset{\boldsymbol{\theta}}{\arg\min} \|\Pi_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}^{\perp}\mathbf{W}^{\frac{1}{2}}\hat{\mathbf{Q}}\hat{\mathbf{r}}\|_2^2. \tag{15}$$

In general, the above problem can be solved iteratively by using either gradient descent or Newton's methods [30]. The gradient and Hessian of the objective function, needed for implementing the aforementioned methods, are given in Appendix C. However, finding the global minimum in (15) through these algorithms is not guaranteed due to multimodality of the objective function. The quality of the solution is susceptible to the initial point with the global minima potentially achieved in case a very good initial point, which is close enough to the global minima, is available. This motivates us to introduce two efficient algorithms for solving the optimization problem (15), which are presented in next. The first algorithm is applicable to SLAs with hole-free co-arrays, such as MRA and nested arrays, while the second one can be used for SLAs with holes in their co-arrays such as co-prime arrays.

Remark 5. We should remark that the proposed estimator, like ACBMs, requires uncorrelated sources and the exact knowledge of their number.

C. WLS Implementation for SLAs With Hole-Free Co-arrays

When the SLA has no holes in its difference co-array, it is possible to recast (15) as a quadratic optimization problem followed by rooting a polynomial through reparameterization of the objective function. The main idea is similar to the technique used in [31], [32]. Indeed, for such kind of SLAs, the objective function in (15) can be reparameterized in terms of the coefficient of the following polynomial

$$f(z) = \sum_{n=0}^{K} \gamma_{K-n} z^n = \gamma_0 \prod_{k=1}^{K} (z_k - e^{\mathrm{j}\pi \sin \theta_k}). \quad (16)$$

To show that, let define

$$\mathbf{\Gamma}_{A_d}^H = \begin{bmatrix} \gamma_K & \cdots & \gamma_1 & \gamma_0 & 0 & \cdots & 0 \\ 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 \end{bmatrix} \in \mathbb{C}^{(2D-1-K)\times(2D-1)}, \quad (17)$$

and $\mathbf{J}_\perp \in \mathbb{R}^{M^2 \times (M^2-2D+1)}$ be obtained from QR-factorization of $\mathbf{J}$ as $\mathbf{J} = \begin{bmatrix} \mathbf{J}_\parallel & \mathbf{J}_\perp \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}^T & \mathbf{0}^T \end{bmatrix}^T$. It is readily confirmed that $\begin{bmatrix} \mathbf{J}_\perp & \mathbf{J}^{\dagger H} \mathbf{\Gamma}_{A_d} \end{bmatrix}^H \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}) = 0$ due to the Vadermonde structure of $\mathbf{A}_d(\boldsymbol{\theta})$ for the aforementioned SLAs. This implies that the columns of $\mathbf{\Gamma} \doteq \begin{bmatrix} \mathbf{J}_\perp & \mathbf{J}^{\dagger H} \mathbf{\Gamma}_{A_d} \end{bmatrix}$ span the null space of $\mathbf{A}_d^H(\boldsymbol{\theta}) \mathbf{J}^H$. Hence, considering the fact that $\Pi^\perp_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{W}^{-\frac{1}{2}} \mathbf{\Gamma}}$ [33], the objective in (15) can be rewritten as

$$\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{\Gamma} (\mathbf{\Gamma}^H \mathbf{W}^{-1} \mathbf{\Gamma})^{-1} \mathbf{\Gamma}^H \hat{\mathbf{Q}} \hat{\mathbf{r}}. \quad (18)$$

Minimization of (18) with respect to the free parameters in $\mathbf{\Gamma}$ leads to the estimates of $\gamma_0, \gamma_1, \cdots, \gamma_K$ from which the WLS estimates of DoAs can be obtained through finding the roots of the polynomial (16). However, the reparameterized optimization problem is still complicated due to multimodality of the objective function and, moreover, the constraint on $\gamma_0, \gamma_1, \cdots, \gamma_K$ arising from the fact that the roots of the polynomial (16) should lie on the unit circle.

The multimodal objective function (18) can be relaxed through replacing $(\mathbf{\Gamma}^H \mathbf{W}^{-1} \mathbf{\Gamma})^{-1}$ with its consistent estimate. This relaxation does not affect the asymptotic behavior of the objective function [5], [31], [32], [34], but converts the objective function to a quadratic function with respect to $\gamma_0, \gamma_1, \cdots, \gamma_K$. A consistent estimate of $(\mathbf{\Gamma}^H \mathbf{W}^{-1} \mathbf{\Gamma})^{-1}$ can be obtained in two ways: 1) making use of CAB-MUSIC or CAB-ESPIRIT to derive an initial consistent estimate of $\boldsymbol{\theta}$; 2) through minimizing $\|\mathbf{\Gamma}^H \hat{\mathbf{Q}} \hat{\mathbf{r}}\|^2 = \|\mathbf{\Gamma}_{A_d}^H \mathbf{J}^\dagger \hat{\mathbf{Q}} \hat{\mathbf{r}}\|^2$ with respect to $\gamma_0, \gamma_1, \cdots, \gamma_K$. Following the same arguments provided in Appendix B and the fact that there is a one-to-one mapping from $\boldsymbol{\theta}$ to $\gamma_0, \gamma_1, \cdots, \gamma_K$ [31], it can easily be shown that minimizing $\|\mathbf{\Gamma}_{A_d}^H \mathbf{J}^\dagger \hat{\mathbf{Q}} \hat{\mathbf{r}}\|^2$ with respect to $\gamma_0, \gamma_1, \cdots, \gamma_K$ gives consistent estimates of $\gamma_0, \gamma_1, \cdots, \gamma_K$. Once consistent estimates of $\gamma_0, \gamma_1, \cdots, \gamma_K$ are given, an initial consistent estimates of $\boldsymbol{\theta}$ can be obtained by solving the polynomial equation (16).

In addition, we need to ensure that the roots of the polynomial lie on the unit circle. Following the methodology in [31], [32], this is addressed by imposing the conjugate symmetric

constraint, i.e., $\gamma_n = \gamma_{K-n}^*$ for $n = 0, 1, \cdots, K$. While this constraint is only necessary, following [31], this relaxation tends to be tight in the asymptotic regimes. Further, to avoid $\gamma_i = 0$, $\forall i$, an additional constraint is required. Herein, we employ the linear constraints, i.e., $\mathrm{Re}\{\gamma_0\} = 1$ or $\mathrm{Im}\{\gamma_0\} = 1$ [31], [32] for simplicity.

Accordingly, the proposed procedure for estimating $\boldsymbol{\theta}$ can be summarized as follows:

1) Compute a consistent estimate of $\boldsymbol{\theta}$ by using CAB-MUSIC, CAB-ESPRIT or minimizing the quadratic function $\|\mathbf{\Gamma}_{A_d}^H \mathbf{J}^\dagger \hat{\mathbf{Q}} \hat{\mathbf{r}}\|$.
2) Based on the initial consistent estimate of $\boldsymbol{\theta}$ and the sample covarince matrix, i.e., $\hat{\mathbf{R}}$, calculate a consistent estimate of $(\mathbf{\Gamma}^H \mathbf{W}^{-1} \mathbf{\Gamma})^{-1}$.
3) Minimize the quadratic criterion

$$\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{\Gamma} (\hat{\mathbf{\Gamma}}^H \hat{\mathbf{W}}^{-1} \hat{\mathbf{\Gamma}})^{-1} \mathbf{\Gamma}^H \hat{\mathbf{Q}} \hat{\mathbf{r}}. \quad (19)$$

with respect to $\gamma_0, \cdots, \gamma_K$ considering the conjugate symmetric constraint, i.e., $\gamma_n = \gamma_{K-n}^*$ for $n = 0, \cdots, K$, besides the linear constraints $\mathrm{Re}\{\gamma_0\} = 1$ or $\mathrm{Im}\{\gamma_0\} = 1$.
4) Obtain the WLS estimate of $\boldsymbol{\theta}$ by rooting $f(z)$ in (16).

D. WLS Implementation for SLAs With Holes in Co-arrays

For these SLAs, it is possible to recast (15) as a polynomial optimization problem followed by rooting $f(z)$ given in (16). This can be done, similar to Section III-C, through reparameterization of (15) in terms of the coefficients of $f(z)$, i.e., $\gamma_0, \gamma_1, \cdots, \gamma_K$, by finding a set of bases spanning the null space of $\mathbf{A}_d^H(\boldsymbol{\theta}) \mathbf{J}^H$. Then, the introduced polynomial optimization problem can be globally solved by using the algorithm proposed in [29].

Let define $\tilde{\mathbf{\Gamma}}_{A_d} = \begin{bmatrix} \tilde{\mathbf{\Gamma}} & \mathbf{\Delta}_1 & \mathbf{\Delta}_2 \end{bmatrix} \in \mathbb{C}^{(2D-1)\times(2D-1-K)}$ such that the matrices $\tilde{\mathbf{\Gamma}} \in \mathbb{C}^{(2D-1)\times(2v-1-K)}$, $\mathbf{\Delta}_1 \in \mathbb{C}^{(2D-1)\times(D-v)}$ and $\mathbf{\Delta}_2 \in \mathbb{C}^{(2D-1)\times(D-v)}$ are given by

$$\tilde{\mathbf{\Gamma}}^H = \begin{bmatrix} 0 & \cdots & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & & & \ddots & \ddots & \ddots & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \gamma_K & \cdots & \gamma_1 & \gamma_0 & 0 & \cdots & 0 \end{bmatrix}, \quad (20)$$

$$\mathbf{\Delta}_1^H = \begin{bmatrix} \delta_K^1 & \cdots & \delta_1^1 & \delta_0^1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \delta_K^2 & \cdots & \delta_1^2 & \delta_0^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & & \ddots & \ddots & & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & \delta_K^{D-v} & \cdots & \delta_1^{D-v} & \delta_0^{D-v} & 0 & \cdots & 0 \end{bmatrix}, \quad (21)$$

$$\mathbf{\Delta}_2^H = \begin{bmatrix} 0 & \cdots & 0 & \delta_K^1 & \cdots & \delta_1^1 & \delta_0^1 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & \delta_K^2 & \cdots & \delta_1^2 & \delta_0^2 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \ddots & & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & \delta_K^{D-v} & \cdots & \delta_1^{D-v} & \delta_0^{D-v} \end{bmatrix}, \quad (22)$$

where $\gamma_0, \cdots, \gamma_K$ are the coefficients of $f(z)$ given in (16) and $\delta_0^i, \cdots, \delta_K^i$ are the coefficients of the following polynomial

$$q_i(z) = \sum_{n=0}^{K} \delta_{K-n}^i z^{\ell_{a_n^i}} = f(z) g_i(z) \quad (23)$$

with $a_n^i = D - 1 - K - i + n$ for $i \in \{1, \cdots, D-v\}$ and $g_i(z) = \frac{\sum_{n=0}^{K} \delta_{K-n}^i z^{\ell_{a_n^i}}}{f(z)}$. From (16), it is observed that $f(e^{\mathrm{j}\pi \sin \theta_k}) = 0$ for $k \in \{1, \cdots, K\}$, which in turn implies that $q_i(e^{\mathrm{j}\pi \sin \theta_k}) = 0$ for $k \in \{1, \cdots, K\}$. Hence, it is easily checked that $\begin{bmatrix} \mathbf{J}_\perp & \mathbf{J}^{\dagger H} \mathbf{\Gamma}_{A_d} \end{bmatrix}^H \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}) = 0$, indicating that

the columns of $\mathbf{\Gamma} \doteq \begin{bmatrix} \mathbf{J}_\perp & \mathbf{J}^{\dagger H}\mathbf{\Gamma}_{A_d} \end{bmatrix}$ span the null space of $\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H$. Hence, considering $\Pi^\perp_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{W}^{-\frac{1}{2}}\tilde{\mathbf{\Gamma}}}$ [33], the objective in (15) can be rewritten as

$$\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\tilde{\mathbf{\Gamma}}(\tilde{\mathbf{\Gamma}}^H\mathbf{W}^{-1}\tilde{\mathbf{\Gamma}})^{-1}\tilde{\mathbf{\Gamma}}^H\hat{\mathbf{Q}}\hat{\mathbf{r}}. \quad (24)$$

Here the minimization should be done with respect to the free parameters in $\tilde{\mathbf{\Gamma}}$, i.e., $\gamma_0,\cdots,\gamma_K$ and $\delta_0^i,\cdots,\delta_K^i$ $\forall i$ from which the WLS etimates of DoAs can be obtained through finding the roots of the polynomial (16). Akin to Section III-C, (24) can be relaxed through replacing $(\tilde{\mathbf{\Gamma}}^H\mathbf{W}^{-1}\tilde{\mathbf{\Gamma}})^{-1}$ with its consistent estimate without affecting the asymptotic behavior of the objective function [5], [31], [32], [34]. However, the reparameterized optimization problem is still complicated despite this relaxation due to the existing constraints on $\gamma_0,\cdots,\gamma_K$ and $\delta_0^i,\cdots,\delta_K^i$ $\forall i$. It can be further simplified if the parameters $\delta_0^i,\cdots,\delta_K^i$ $\forall i$ are somehow expressed in terms of $\gamma_0,\cdots,\gamma_K$. This parametrizes the optimization problem only in terms of the desired parameters, i.e., $\gamma_0,\cdots,\gamma_K$ and thereby eliminates the constraints corresponding to $\delta_0^i,\cdots,\delta_K^i$ $\forall i$. Towards this, in what follows, we use the fact that the remainder $h_i(z)$ generated by division of $q_i(z)$ by $f(z)$ is zero according to (23).

Let $\tilde{n}$ be an integer such that $\ell_{a_{\tilde{n}}^i} < K$ and $\ell_{a_{\tilde{n}+1}^i} \geq K$. It is then possible to rewrite $q_i(z)$ as $q_i(z) = \sum_{n=0}^{\tilde{n}} \delta_{K-n}^i z^{a_n^i} + \sum_{n=\tilde{n}+1}^{K} \delta_{K-n}^i z^{a_n^i}$. Making use of polynomial long division and after some tedious calculations, the remainder $h_i(z)$ resulting from division of $q_i(z)$ by $f(z)$ takes the form

$$h_i(z) = \sum_{n=0}^{K-1} c_{K-1-n}^i z^n, \quad \text{where} \quad (25)$$

$$c_{K-1-n}^i = \begin{cases} \delta_{K-n}^i - \tilde{a}_n(\gamma_0,\cdots,\gamma_K,\delta_{K-\tilde{n}-1}^i,\cdots,\delta_0^i) & \text{if} \quad n \in \mathbb{S} \\ \bar{a}_n(\gamma_0,\cdots,\gamma_K,\delta_{K-\tilde{n}-1}^i,\cdots,\delta_0^i) & \text{if} \quad n \notin \mathbb{S} \end{cases} \quad (26)$$

with $\mathbb{S} = \{\ell_{a_0^i},\cdots,\ell_{a_{\tilde{n}}^i}\}$, $\tilde{a}_n(.)$ and $\bar{a}_n(.)$ being linear functions of $\delta_{K-\tilde{n}-1}^i,\cdots,\delta_0^i$ for $n = 0,\cdots,K-1$ whose coefficients are obtained during long division. $h_i(z)$ is identically zero $\forall i$ if and only if $c_n^i = 0$ $\forall n,i$. Letting c_n^i's equal to zero for each i results in K linear equations with respect to $\delta_K^i,\cdots,\delta_0^i$. Considering the fact that δ_0^i can be chosen arbitrarily, the solution of these K linear equations provides us with the values of $\delta_K^i,\cdots,\delta_1^i$ based on $\gamma_0,\cdots,\gamma_K$ and δ_0^i. Through an appropriate choice of δ_0^i, the parameters $\delta_K^i,\cdots,\delta_1^i$ can be expressed as polynomial functions of $\gamma_0,\cdots,\gamma_K$. Hence, $\tilde{\mathbf{\Gamma}}$ is parameterized only in terms of the coefficients $\gamma_0,\cdots,\gamma_K$.

Consequently, the optimization problem (23) is converted to a polynomial optimization with respect to $\gamma_0,\gamma_1,\cdots,\gamma_K$. Similar to Section III-C, conjugate symmetry constraint is imposed on $\{\gamma_i\}$ towards ensuring roots of $f(z)$ on the unit circle. Then, the resulting polynomial optimization can be solved by using the Lasserre's SDP relaxation given in [29]. It is proved in [29] that the Lasserre's SDP relaxation attains the global minima of the polynomial optimization if the order of relaxation is big enough. However, evidently, this solution exhibits higher complexity than the descent algorithms mentioned earlier due to the nature of operations and increased number of variables.

Accordingly, the proposed procedure for estimating $\boldsymbol{\theta}$ can be summarized as follows:

1) Compute a consistent estimate of $\boldsymbol{\theta}$ by using CAB-MUSIC, CAB-ESPRIT.
2) Based on the initial consistent estimate of $\boldsymbol{\theta}$ and the sample covariance matrix, i.e., $\hat{\mathbf{R}}$, calculate a consistent estimate of $(\mathbf{\Gamma}^H\mathbf{W}^{-1}\mathbf{\Gamma})^{-1}$.
3) Compute the coefficients c_n^i's $\forall i,n$, using polynomial long division.
4) Let c_n^i's equal to zero for each i and solve the K resulting linear equations with respect to $\delta_K^i,\cdots,\delta_0^i$. This step gives us the values of $\delta_K^i,\cdots,\delta_1^i$ based on $\gamma_0,\cdots,\gamma_K$ and δ_0^i.
5) Choose the free variable δ_0^i such that $\delta_0^i, \delta_K^i,\cdots,\delta_0^i$ are obtained as polynomial functions of $\gamma_0,\cdots,\gamma_K$.
6) Find the minima of the following polynomial criterion

$$\hat{\mathbf{r}}^H\hat{\mathbf{Q}}^H\tilde{\mathbf{\Gamma}}(\hat{\tilde{\mathbf{\Gamma}}}^H\hat{\mathbf{W}}^{-1}\hat{\tilde{\mathbf{\Gamma}}})^{-1}\tilde{\mathbf{\Gamma}}^H\hat{\mathbf{Q}}\hat{\mathbf{r}}. \quad (27)$$

with respect to $\gamma_0,\cdots,\gamma_K$ considering the conjugate symmetric constraint and the linear constraints $\text{Re}\{\gamma_0\} = 1$ or $\text{Im}\{\gamma_0\} = 1$ by using the Lasserre's SDP relaxation.
7) Obtain the WLS estimate of $\boldsymbol{\theta}$ by rooting $f(z)$ in (16).

IV. ASYMPTOTIC PERFORMANCE ANALYSIS

The asymptotic behavior of the proposed WLS estimator for a large number of samples is analyzed in this section and its asymptotic statistical efficiency when an optimal weighting matrix being selected is shown.

A. Asymptotic Performance

In this subsection, we analyze the key attributes of the proposed WLS estimator including consistency, bias and estimation errors for an arbitrary weighting matrix. Asymptotics are resorted to yield tractable analytical results. We start with proving the consistency of the DoA estimates provided by the proposed WLS estimator.

Theorem 1. $\hat{\boldsymbol{\theta}}_{wls}$ is a consistent estimate of $\boldsymbol{\theta}$ if $K \leq v - 1$.

Proof. See Appendix B ∎

Remark 6 (Unbiasedness). It readily follows from Theorem 1 that $\hat{\boldsymbol{\theta}}_{wls}$ is asymptotically unbiased as well.

Remark 7. Note that the sufficient condition for the consistency of $\hat{\boldsymbol{\theta}}_{wls}$, i.e., $K \leq v - 1$, given in Theorem 1, is equivalent to the sufficient condition for source identifiability given in [10]–[12].

The consistency of $\hat{\boldsymbol{\theta}}_{wls}$ can be utilized to obtain the covariance matrix of DoA estimation errors through a Taylor series expansion approach when $N \to \infty$. This is detailed in the following theorem.

Theorem 2. Let $\mathbf{K}_M$ denote the commutation matrix defined according to Definition 2 in Appendix G. If $K \leq v - 1$ and $\mathbf{K}_M\mathbf{W} = \mathbf{W}^*\mathbf{K}_M$, the asymptotic ($N \to \infty$) covariance matrix of the WLS estimator $\hat{\boldsymbol{\theta}}_{wls}$ is given by (28), shown

$$\mathbf{C}_{wls} = \mathbb{E}\left\{(\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})^H\right\} = \frac{1}{\pi^2 N}\text{diag}^{-1}(\mathbf{p})\left(\boldsymbol{\Omega}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\boldsymbol{\Omega}\right)^{-1}$$
$$\times \left(\boldsymbol{\Omega}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{Q}\mathbf{M}^2\mathbf{Q}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\boldsymbol{\Omega}\right)\left(\boldsymbol{\Omega}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\boldsymbol{\Omega}\right)^{-1}\text{diag}^{-1}(\mathbf{p}), \quad (28)$$

at the top of the next page, where

$$\mathbf{Q} \doteq \mathbf{I}_{M^2} - \frac{\text{vec}(\mathbf{I}_M)\mathbf{b}^H}{v - K}, \quad (29)$$

$$\mathbf{b} \doteq \mathbf{J}^{\dagger H}\mathbf{T}^H\text{vec}\left(\mathbf{U}_n\mathbf{U}_n^H\right), \quad (30)$$

$$\mathbf{M}^2 = \mathbf{R}^T \otimes \mathbf{R}, \quad (31)$$

$$\boldsymbol{\Omega} \doteq \mathbf{J}\text{diag}\left(\mathbf{d}\right)\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta}), \quad (32)$$

with $\mathbf{d} = \begin{bmatrix} -\ell_{D-1} & \cdots & \ell_0 & \cdots & \ell_{D-1} \end{bmatrix}^T$, $\mathbf{U}_n$ *being given in Appendix A, and* $\boldsymbol{\Phi}(\boldsymbol{\theta}) = \text{diag}(\begin{bmatrix} \cos\theta_1 & \cdots & \cos\theta_K \end{bmatrix}^T)$.

Proof. See Appendix C   □

Remark 8. In Theorem 2, it is assumed that $\mathbf{K}_M\mathbf{W} = \mathbf{W}^*\mathbf{K}_M$ because it simplifies the expression for the covariance matrix of DoA estimation errors. Further, it is shown in Appendix E (See Lemma 9) that the optimal weighting matrix, to be introduced later in Theorem 4, fulfills the constraint. Thus the constraint is not restrictive. However, it is fairly straightforward to also obtain the covariance matrix of $\hat{\boldsymbol{\theta}}_{wls}$ for the weighting matrices not satisfying the aforementioned constraint by exploiting the derivations given in Appendix C.

B. Optimal Weighting Matrix and Achieving CRB

The results presented in Section IV-A are valid for an arbitrary weighting matrix. However, it is of interest to find an optimal weighting matrix providing the best DoA estimation performance in terms of MSE. In this regard, we resort to the fact that $\mathbf{C}_{wls}$, given in Theorem 2, is lower bounded by the CRB. Thus, a good way of finding the optimal weighting matrix is to seek a weighting matrix rendering $\mathbf{C}_{wls}$ to be identical to the CRB. Accordingly, in what follows, we first reformulate the CRB expression given in [12] in a form suitable for establishing equality of $\mathbf{C}_{wls}$ and the CRB in Theorem 3. Then, we show in Theorem 4 that there is a weighting matrix enabling $\mathbf{C}_{wls}$ to coincide with the CRB.

Theorem 3. *The CRB expression given in [12, Eq. (49)] can be reformulated as the following form*

$$CRB(\boldsymbol{\theta}) = \left(\text{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{F}\left(\mathbf{F}^H\mathbf{H}\mathbf{F}\right)^{-1}\mathbf{F}^H\boldsymbol{\Omega}\text{diag}(\mathbf{p})\right)^{-1}, \quad (33)$$

where $\mathbf{F} \in \mathbb{C}^{M^2 \times (M^2-K)}$ *is any matrix whose columns span the null space of* $\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H$ *and*

$$\mathbf{H} = \mathbf{Q}\mathbf{M}^2\mathbf{Q}^H + \frac{\mathbf{M}^2\mathbf{b}\mathbf{b}^H\mathbf{M}^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}. \quad (34)$$

Proof. See Appendix D   □

Theorem 4. *If the weighting matrix is selected as follows*

$$\mathbf{W}_{\text{opt}} = \left(\Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}\mathbf{S}\Pi^{\perp}_{\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})} + \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^T\right)^{-1}, \quad (35)$$

where

$$\mathbf{S} = \mathbf{Q}\mathbf{M}^2\mathbf{Q}^H + \mathbf{b}\mathbf{b}^H, \quad (36)$$

then we have $\mathbf{C}_{wls} = CRB(\boldsymbol{\theta})$.

Proof. See Appendix E   □

Remark 9. We note that the optimal weighting matrix given in Theorem 4 depends on the true value of the parameters. However, in practice, it can be replaced with a consistent estimate without affecting the asymptotic performance of the WLS estimator [5], [31], [32], [34]. To this end, we can first use any other consistent estimator like CAB-MUSIC or CAB-ESPRIT to obtain an initial estimate of $\boldsymbol{\theta}$. Then, we compute a consistent estimate of the optimal weighting matrix based on the initial estimate of $\boldsymbol{\theta}$ and the sample covariance matrix $\hat{\mathbf{R}}$. Finally, we use the WLS estimator given in (15) to derive asymptotically statistically efficient estimates of DoAs. This procedure may be iterated with $\mathbf{W}_{\text{opt}}$ and $\boldsymbol{\theta}$ alternatively estimated. This may enhance the estimation accuracy especially at low snapshots.

We are also required to make sure that the optimal weighting matrix is positive definite, and moreover, its estimate, obtained using the approach explained in Remark 9, is always nonsingular regardless of the available number of snapshots. The following Lemma addresses this concern.

Lemma 2. $\mathbf{W}_{\text{opt}}$ *is positive definite and its estimate, obtained through either CAB-MUSIC or CAB-ESPRIT, is nonsingular regardless of the available number of snapshots.*

Proof. See Appendix F   □

V. SIMULATION RESULTS

In this section, we provide some numerical results to validate the analytical results obtained in Section III as well as to assess the performance of the proposed estimator. Further, we compare the performance of the WLS estimator proposed in this paper with that of CAB-MUSIC [11], [20], CAB-ESPRIT [21], SPA [19] and SCAB-MUSIC [28]; and we will show that the WLS estimator yields better performance in terms of resolution, estimation accuracy and statistical efficiency.

A. General Set-up

In all experiments, each simulated point has been computed by 5000 Monte Carlo repetitions. In addition, it is assumed that the K independent sources are located at $\{-60° + 120°(k - 1)/(K - 1)|k = 0, 1, \cdots, K - 1\}$. All sources have equal powers, i.e., $p_k = p \ \forall k$, and the SNR is defined as $10\log\frac{p}{\sigma^2}$. Throughout this section, we use three different types of SLAs with $M = 6$ physical elements and the following geometries:

$$\mathbb{M}_{\text{nested}} : \{1, 2, 3, 4, 8, 12\}, \quad (37)$$

$$\mathbb{M}_{\text{co-prime}} : \{0, 2, 3, 4, 6, 9\}, \quad (38)$$

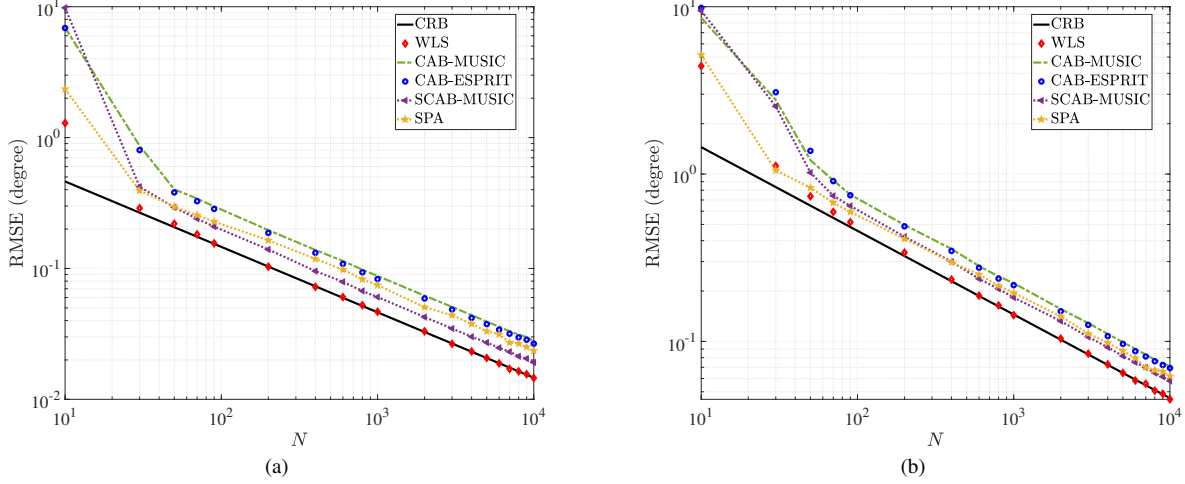$$\mathbb{M}_{\text{MRA}} : \{0, 1, 6, 9, 11, 13\}. \quad (39)$$

Fig. 2. RMSE in degree for θ_2 versus the number of snapshots for a nested array with $M = 6$ elements and configuration given in (37), SNR = 3 dB, and: (a) $K = 4 < M$; (b) $K = 7 > M$.
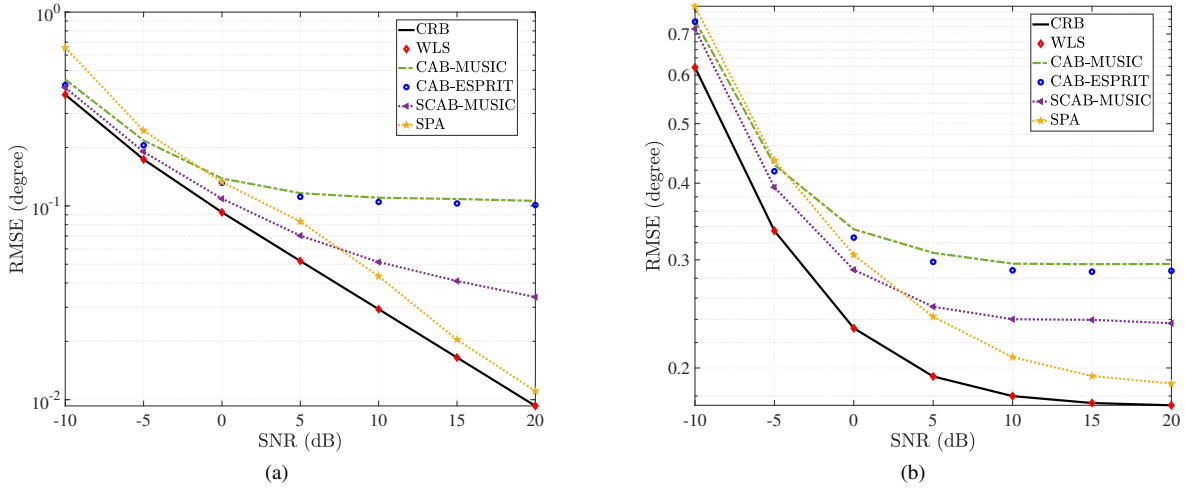

Fig. 3. RMSE in degree for θ_2 versus SNR for a nested array with $M = 6$ elements and configuration given in (37), $N = 500$, and: (a) $K = 4 < M$; (b) $K = 7 > M$.

These SLAs generate the difference co-arrays as:

$$\mathbb{D}_{\text{nested}} : \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11\}, \tag{40}$$

$$\mathbb{D}_{\text{co-prime}} : \{0, 1, 2, 3, 4, 5, 6, 7, 9\}, \tag{41}$$

$$\mathbb{D}_{\text{MRA}} : \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13\}. \tag{42}$$

The optimization problem (15) for MRAs and nested arrays is solved through the algorithm described in Section III-C; and for co-prime arrays it is solved by using the algorithm given in Section III-D. In both cases, CAB-ESPRIT is used to derive a consistent estimate of $(\boldsymbol{\Gamma}^H \mathbf{W}^{-1} \boldsymbol{\Gamma})^{-1}$. Further, we take the grid from $-90°$ to $90°$ with step size $0.001°$ to implement CAB-MUSIC and SCAB-MUSIC. Moreover, all estimators but SPA need an exact knowledge of the exact number of sources. Hence, for a fair comparison, SPA is also assumed to know the exact number of sources in all simulations. For this end, it is implemented by applying MUSIC on the augmented covaraince matrix estimate obtained from the SPA algorithm.

B. MSE vs. the Number of Snapshots

Fig. 2 depicts the Root-Mean-Squares-Error (RMSE) for θ_2 in degree versus the number of snapshots for the nested array in (37). The SNR is assumed to be 3 dB. In addition, noting $M = 6$, two different scenarios are considered: (a) $K = 4 < M$, and (b) $K = 7 > M$. Fig. 2 illustrates a close agreement between the RMSE of the proposed WLS estimator and the CRB when about 70 or more snapshots are available, indicating asymptotic statistical efficiency of the WLS estimator. Further, a considerable gap is observed between the performance of CAB-MUSIC (CAB-ESPRIT) and that of the WLS estimator (the CRB). For instance, at $N = 400$, Figs 3a and 3b show a performance gain of roughly 2.6 dB and 2 dB, respectively, in terms of the RMSE when the WLS estimator is used. It is also observed that SCAB-MUSIC and SPA outperform CAB-MUSIC CAB-ESPRIT, but their performance is inferior to that of the WLS estimator and they are unable to attain the CRB.

Fig. 2 also shows that when a small number of snapshots is available, for example less than 70, all estimators are confronted with substantial performance degradation. Performance loss of the subspace methods, i.e., CAB-MUSIC, CAB-ESPRIT and SCAB-MUSIC, is justified by the subspace swap arising from the inaccurate estimate of the resulting augmented

covariance matrix is this case. Further, the underlying reasons for performance degradation of the WLS estimator in such a regime are twofold. Firstly, as mentioned in Remarks 1 and 4, the estimates of σ^2 and $\mathbf{p}$ are not precise and might even yield negative values in this case. Consequently, the value of $\hat{\mathbf{Q}}\hat{\mathbf{r}}$ significantly deviates from its asymptotic value, i.e., $\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})\mathbf{p}$, which, in turn, causes the minimizer of (15) to diverge from the true value of $\boldsymbol{\theta}$. Secondly, there exists a poor estimate of the optimal weighting matrix, i.e., $\mathbf{W}_{\text{opt}}$. This has a detrimental effect on the performance of the WLS estimator. However, it is seen that the proposed WLS estimator still has superior performance compared to the other estimators even in low snapshot paradigm.

C. MSE vs. SNR

Fig. 3 shows the RMSE for θ_2 in degree versus SNR for the same setup used for Fig. 2. The number of snapshots is considered to be $N = 500$. It is seen in Figs. 3a and Fig. 3b that the RMSE of the WLS estimator perfectly matches the CRB for the considered range of SNR as a consequence of its asymptotic statistical efficiency. However, the other estimators are not capable of attaining the CRB.

Fig. 3a demonstrates that for $K = 4 < M$ the RMSEs of the WLS estimator and SPA, like the CRB, tend to decay to zero as SNR increases while the RMSEs of CAB-MUSIC, CAB-ESPRIT and SCAB-MUSIC tend to get saturated at the high SNR regime. For instance, the RMSEs of CAB-MUSIC and CAB-ESPRIT get saturated at SNR around 5 dB. The underlying cause for this saturation behavior of CAB-MUSIC and CAB-ESPRIT was already explained in [20, Corollary 2] and [21], respectively, through an analysis of their MSE expressions. This saturation behaviour of CAB-MUSIC, CAB-ESPRIT and SCAB-MUSIC renders them highly inefficient as SNR increases while the WLS estimator remains statistically efficient for the considered range of SNR.

Fig. 3b shows that when $K = 7 < M$, the RMSEs of all the estimators as well as the CRB get saturated at the high SNR regime. The saturation point for the WLS estimator, SPA and the CRB is at the SNR around 15 dB while for CAB-MUSIC, CAB-ESPRIT and SCAB-MUSIC it happens at the SNR around 10 dB. Nonetheless, the WLS estimator still pefroms better than all the other estimators under this condition. For example, at $\text{SNR} = 15$, the performance gains of about 2.2 dB and 1.3 dB are attained in terms of RMSE compared to CAB-MUSIC (CAB-ESPRIT) and SCAB-MUSIC, respectively.

D. Impact of Different SLA Configurations

In Fig. 4, we plot the RMSE for θ_2 in degree versus SNR for different types of SLAs given in (37)-(39). The rest of parameters are equal to those in Fig. 3. It is readily observed that there is a good agreement between the RMSE of the WLS estimator and the CRB regardless of the array geometry. These simulations corroborate the analytical results where the asymptotic equality of the CRB and the MSE of the WLS estimator is shown considering a generic SLA. Another observation is that amongst these three SLAs, MRA is endowed with the least RMSE followed by the nested and co-prime arrays, respectively. This follows from the distinction

between the size of their corresponding difference co-array in comparison to each other. The difference co-arrays for these SLAs are given in (40)-(42). Indeed, the array with a bigger difference co-array size brings about the lower RMSE.

E. Resolution Probability

Figs. 5 and 6 depict the probability of resolution versus SNR for the proposed WLS estimator, CAB-MUSIC, CAB-ESPRIT, SCAB-MUSIC and SPA. The co-prime and nested arrays with the configurations given in (38) and (37) are considered in Figs. 5 and 6, respectively. The number of snapshots is considered to be $N = 500$. In addition, we consider two sources with equal powers, located at $\theta_1 = 20° - \frac{\Delta\theta}{2}$ and $\theta_2 = 20° + \frac{\Delta\theta}{2}$ with: (a) $\Delta\theta = 1°$, and (b) $\Delta\theta = 2°$. We define the two sources as being resolvable if $\max_{i \in \{1,2\}} |\hat{\theta}_i - \theta_i| < \frac{\Delta\theta}{2}$ [35]. Figs. 5 and 6 demonstrate that the WLS estimator has the best resolution performance amongst all the estimators while the SPA resolution performance is inferior to that of all the other ones. Furthermore, CAB-MUSIC and CAB-ESPRIT perform almost equivalently and SCAB-MUIC performs slightly better than them. When $\Delta\theta = 1°$, all the estimators but the WLS are unable to resolve the sources with a probability of 1 even at $\text{SNR} = 17$ dB while the WLS estimator could achieve a resolution probability of 1 at $\text{SNR} = 13$ dB. In case $\Delta\theta$ is increased to $2°$, the WLS achieves SNR gains of 3 dB and 6 dB compared to SCAB-MUSIC, to attain a resolution probability of 1, when co-prime and nested arrays are used respectively. A comparison of Figs. 5 and 6 indicates that the probability of resolution for the WLS, CAB-MUSIC and CAB-ESPRIT increases when the co-prime array is replaced with the nested array. This can be justified comparing the aperture size of the co-prime and nested arrays. Indeed, the nested array enjoys a bigger aperture compared to the co-prime array; thus it is expected to demonstrate a better probability of resolution. However, it is observed that the resolution probabilities of SCAB-MUSIC and SAP decline as the nested array is used instead of the co-prime array. This behavior arises from the structure exhibited by the covarinace matrix in co-prime arrays compared to that of nested arrays. Hence, SCAB-MUSIC and SAP, utilizing a structured estimate of the covariance matrix, are capable of providing a better estimate of the covariance matrix when the co-prime array is used. This naturally leads to a better covaraince matrix estimate and hence better DoA estimates.

VI. CONCLUSION

In this paper, a novel WLS estimator for the co-array-based DoA estimation via SLAs was proposed and its performance is thoroughly analyzed. It was shown that the proposed WLS estimator provides consistent estimates of DoAs of identifiable sources for any SLAs. Further, an asymptotic closed-form expression for the resulting covariance matrix of DoA estimation errors was derived and it was analytically proved that it asymptotically coincides with the CRB in case the optimal weighting matrix is selected. This implies that the proposed WLS estimator is asymptotically statistically efficient. It thus closes an important gap in the co-array-based DoA estimation. Simulation results demonstrated superior performance of the
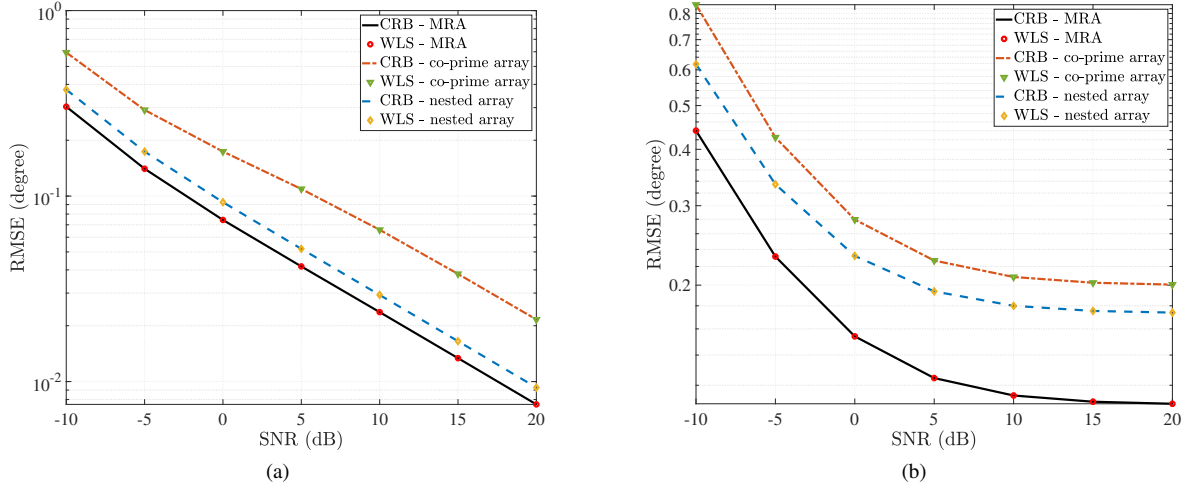
Fig. 4. RMSE in degree for θ_2 versus SNR for SLAs with $M = 6$ elements and different configurations, $N = 500$, and: (a) $K = 4 < M$; (b) $K = 7 > M$.
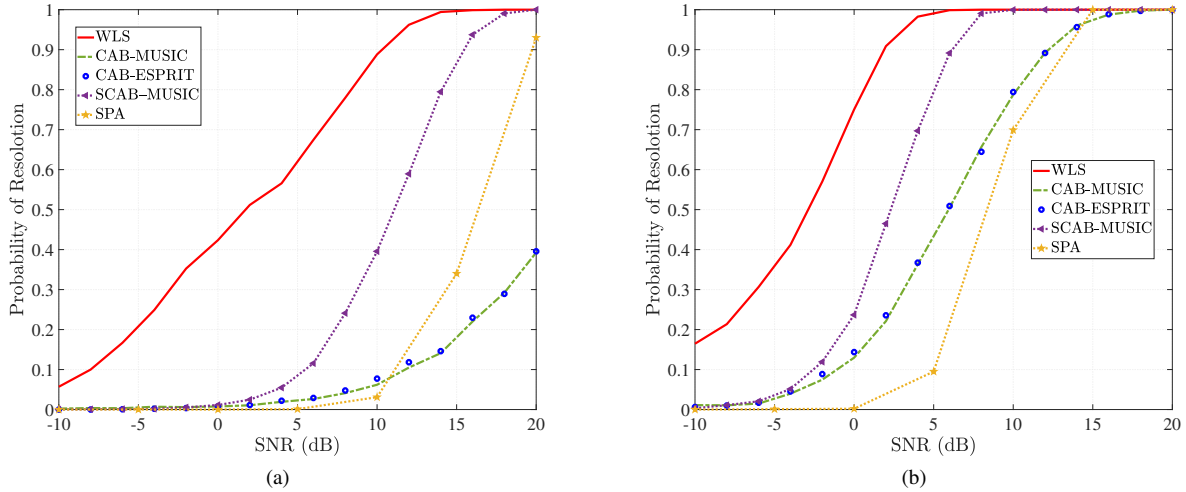
Fig. 5. Probability of resolution versus SNR for a co-prime array with $M = 6$ elements and configuration given in (38), $N = 500$, and: (a) $\Delta\theta = 1°$; (b) $\Delta\theta = 2°$.
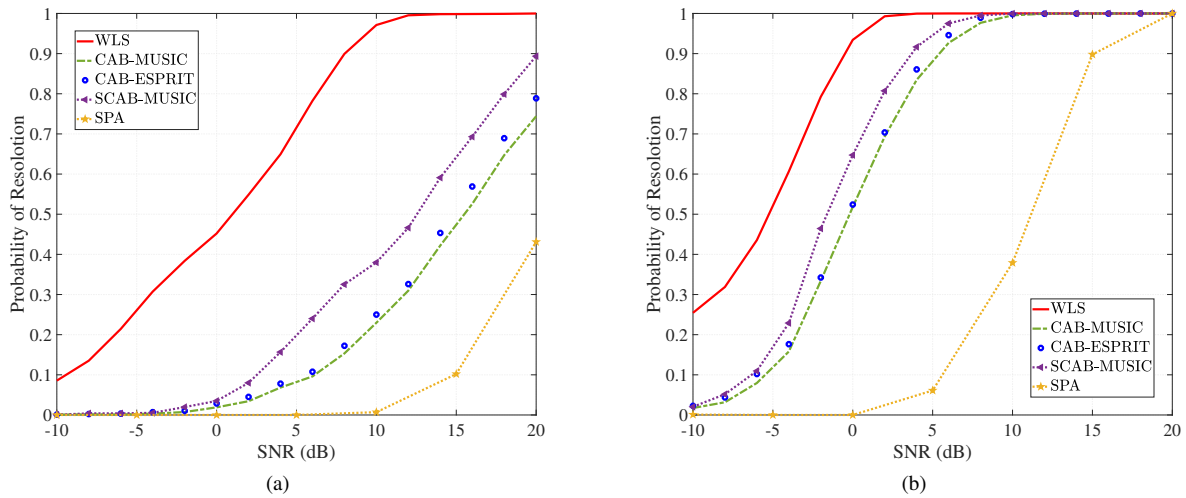
Fig. 6. Probability of resolution versus SNR for a nested array with $M = 6$ elements and configuration given in (37), $N = 500$, and: (a) $\Delta\theta = 1°$; (b) $\Delta\theta = 2°$.

proposed WLS estimator compared to the existing algorithms in the literature in terms of estimation accuracy and resolution.

APPENDIX A
PROOF OF LEMMA 1

It is well-known that the sample covariance $\hat{\mathbf{R}}$ is a consistent estimate of $\mathbf{R}$ under the current assumption [36], implying that

$\lim_{N \to \infty} \hat{\mathbf{r}} = \mathbf{r}$. As a consequence, considering (7), we obtain

$$\lim_{N \to \infty} \hat{\mathbf{R}}_v = \begin{bmatrix} \mathbf{T}_v \mathbf{J}^\dagger \mathbf{r} & \mathbf{T}_{v-1} \mathbf{J}^\dagger \mathbf{r} & \cdots & \mathbf{T}_1 \mathbf{J}^\dagger \mathbf{r} \end{bmatrix}. \quad (43)$$

On the other hand, it has been proved in [22] that

$$\lim_{N \to \infty} \hat{\mathbf{R}}_v = \mathbf{A}_v(\boldsymbol{\theta}) \mathrm{diag}(\mathbf{p}) \mathbf{A}_v^H(\boldsymbol{\theta}) + \sigma^2 \mathbf{I}_v \doteq \mathbf{R}_v, \quad (44)$$

where the matrix $\mathbf{R}_v \in \mathbb{C}^{v \times v}$ has the same structure as the covariance matrix of signals received by a contiguous ULA whose elements are located at $\left(0, \frac{\lambda}{2}, \lambda, \cdots, (v-1)\frac{\lambda}{2}\right)$ and $\mathbf{A}_v(\boldsymbol{\theta}) \in \mathbb{C}^{v \times K}$ is its corresponding steering matrix. Therefore, in case $K \le v - 1$, exploiting the eigendecomposition, it is possible for (44) to be expressed as $\mathbf{R}_v = \mathbf{U}_s \boldsymbol{\Lambda}_s \mathbf{U}_s^H + \sigma^2 \mathbf{U}_n \mathbf{U}_n^H$, where $\mathbf{U}_s$ and $\mathbf{U}_n$ represent the eigenvectors of $\mathbf{R}_v$ corresponding to its K largest and $v - K$ smallest eigenvalues, respectively. From (44), $\hat{\mathbf{R}}_v$ can be deemed to be a perturbed version of $\mathbf{R}_v$. Therefore, we have

$$\lim_{N \to \infty} \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H = \mathbf{U}_n \mathbf{U}_n^H. \quad (45)$$

Eventually, making use of (43), (44), (45) and the fact that $\mathbf{U}_n^H \mathbf{A}_v(\boldsymbol{\theta}) = 0$, we obtain

$$\lim_{N \to \infty} \hat{\sigma}^2 = \frac{\mathrm{vec}^H(\mathbf{U}_n \mathbf{U}_n^H) \mathbf{T} \mathbf{J}^\dagger \mathbf{r}}{v - K}$$
$$= \frac{\mathrm{vec}^H(\mathbf{U}_n \mathbf{U}_n^H) \mathrm{vec}(\begin{bmatrix} \mathbf{T}_v \mathbf{J}^\dagger \mathbf{r} & \mathbf{T}_{v-1} \mathbf{J}^\dagger \mathbf{r} & \cdots & \mathbf{T}_1 \mathbf{J}^\dagger \mathbf{r} \end{bmatrix})}{v - K}$$
$$= \frac{\sigma^2 \mathrm{tr}(\mathbf{U}_n^H \mathbf{U}_n)}{v - K} = \sigma^2. \quad (46)$$

APPENDIX B
PROOF OF THEOREM 1

Let define $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \|\Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^\perp \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}} \hat{\mathbf{r}}\|_2^2$, $\tilde{L}(\boldsymbol{\theta}) = \lim_{N \to \infty} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ and use $\boldsymbol{\theta}_0$ to distinguish the actual DoA vector from a generic vector $\boldsymbol{\theta}$. Since the derivatives of $\mathbf{A}_d(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ are bounded, $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ converges uniformly to $\tilde{L}(\boldsymbol{\theta})$ as $N \to \infty$ [34]. Thus, $\hat{\boldsymbol{\theta}}_{wls}$ also converges to the minimizing argument of $\tilde{L}(\boldsymbol{\theta})$ as $N \to \infty$.

It readily follows from (45) that $\lim_{N \to \infty} \hat{\mathbf{Q}} = \mathbf{Q}$. In addition, from (46) and (4)

$$\mathbf{Q}\mathbf{r} = \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_0) \mathbf{p} \quad (47)$$

is readily checked. Hence, considering the fact that $\lim_{N \to \infty} \hat{\mathbf{r}} = \mathbf{r}$ and making use of continuous differentiability of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$, we obtain

$$\tilde{L}(\boldsymbol{\theta}) = L(\boldsymbol{\theta}, \lim_{N \to \infty} \hat{\mathbf{r}}, \lim_{N \to \infty} \hat{\mathbf{Q}}) = \|\Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d}^\perp \mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_0) \mathbf{p}\|_2^2. \quad (48)$$

It is evident from (48) that $\tilde{L}(\boldsymbol{\theta}_0) = 0$ and $\tilde{L}(\boldsymbol{\theta}) \ge 0$, implying $\tilde{L}(\boldsymbol{\theta})$ has a global minima at $\boldsymbol{\theta}_0$. Consequently, consistency of $\hat{\boldsymbol{\theta}}_{wls}$ follows if $\boldsymbol{\theta}_0$ is the unique solution to $\tilde{L}(\boldsymbol{\theta}) = 0$ with respect to $\boldsymbol{\theta}$. According to (48), $\tilde{L}(\boldsymbol{\theta})$ is equal to zero if and only if

$$\Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^\perp \mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_0) \mathbf{p} = 0. \quad (49)$$

In what follows, we employ the method of proof by contradiction to complete the proof. Let assume that $\boldsymbol{\theta}_1$ is a solution to (49) but $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_0$. This means that $\boldsymbol{\theta}_1$ could differ from $\boldsymbol{\theta}_0$ at q DoAs where $1 \le q \le K$. Substituting $\boldsymbol{\theta}_1$ into (49) yields

$$\Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_1)}^\perp \mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_0') \mathbf{p}' = 0. \quad (50)$$

where $\boldsymbol{\theta}_0' \in \mathbb{C}^{q \times 1}$ consists of those elements of $\boldsymbol{\theta}_0$ which are not shared with $\boldsymbol{\theta}_1$, and $\mathbf{A}_d(\boldsymbol{\theta}_0') \in \mathbb{C}^{q \times 1}$ and $\mathbf{p}' \in \mathbb{C}^{q \times 1}$ are the corresponding blocks of $\mathbf{A}_d(\boldsymbol{\theta}_0)$ and $\mathbf{p}$. The expression on the left side of the above equation can be rewritten as follows

$$\underbrace{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}_0') & \mathbf{A}_d(\boldsymbol{\theta}_1) \end{bmatrix} \begin{bmatrix} \mathbf{I}_K \\ -\left(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_1)\right)^\dagger \mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_0') \end{bmatrix} \mathbf{p}'}_{\aleph}. \quad (51)$$

Since the weighting matrix $\mathbf{W}$ is positive definite by definition and $\mathbf{J}$ a is full column rank matrix [12], it is concluded that (49) is zero if and only if the term $\aleph$ in (51) is zero.

Lemma 3. $\begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}_0') & \mathbf{A}_d(\boldsymbol{\theta}_1) \end{bmatrix}$ *has full column rank if* $K \le v - 1$.

Proof. Let $\mathbf{A}_\vartheta(\boldsymbol{\theta}) \in \mathbb{C}^{(2v-1) \times K}$ denote the steering matrix corresponding to the contiguous ULA segment of the difference co-array. Since $\begin{bmatrix} \mathbf{A}_\vartheta(\boldsymbol{\theta}_0') & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1) \end{bmatrix}$ is a sub-matrix of $\begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}_0') & \mathbf{A}_d(\boldsymbol{\theta}_1) \end{bmatrix}$, it is sufficient to show that $\begin{bmatrix} \mathbf{A}_\vartheta(\boldsymbol{\theta}_0') & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1) \end{bmatrix}$ is full column rank instead of $\begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}_0') & \mathbf{A}_d(\boldsymbol{\theta}_1) \end{bmatrix}$.

It is possible to decompose $\begin{bmatrix} \mathbf{A}_\vartheta(\boldsymbol{\theta}_0') & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1) \end{bmatrix}$ as follows

$$\begin{bmatrix} 1 & \cdots & 1 & 1 & \cdots & 1 \\ \alpha_1 & \cdots & \alpha_q & \beta_1 & \cdots & \beta_K \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \alpha_1^{2(v-1)} & \cdots & \alpha_q^{2(v-1)} & \beta_1^{2(v-1)} & \cdots & \beta_K^{2(v-1)} \end{bmatrix}$$
$$\times \mathrm{diag}(\begin{bmatrix} \alpha_1^{1-v} & \cdots & \alpha_q^{1-v} & \beta_1^{1-v} & \cdots & \beta_K^{1-v} \end{bmatrix}). \quad (52)$$

where $\alpha_i = e^{\mathrm{j}\pi \sin[\boldsymbol{\theta}_0']_i}$ and $\beta_i = e^{\mathrm{j}\pi \sin[\boldsymbol{\theta}_1]_i}$. The second matrix in (52) is a $(K+q) \times (K+q)$ diagonal matrix and thus full rank. However, the first one is a $(2v-1) \times (K+q)$ Vandermonde matrix which has full column rank for distinct DoAs iff $K + q \le 2v - 1$. This condition is fulfilled for all admissible q iff $K \le v - 1$. Hence, it follows that $\begin{bmatrix} \mathbf{A}_\vartheta(\boldsymbol{\theta}_0') & \mathbf{A}_\vartheta(\boldsymbol{\theta}_1) \end{bmatrix}$ and, in turn, $\begin{bmatrix} \mathbf{A}_d(\boldsymbol{\theta}_0') & \mathbf{A}_d(\boldsymbol{\theta}_1) \end{bmatrix}$ have full column rank if $K \le v - 1$. $\square$

According to Lemma 3, on condition that $K \le v - 1$, the term $\aleph$ in (51) is equal to zero iff

$$\begin{bmatrix} \mathbf{I}_K \\ -\left(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\tilde{\boldsymbol{\theta}})\right)^\dagger \mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}_0) \end{bmatrix} \mathbf{p}' = 0, \quad (53)$$

implying that $\mathbf{p}' = 0$, which is in contradiction with the definition of $\mathbf{p}$ given in **A2**. Therefore, it is concluded that $\boldsymbol{\theta}_0$ has to be the unique solution to $\tilde{L}(\boldsymbol{\theta}) = 0$ with respect to $\boldsymbol{\theta}$ if $K \le v - 1$, which indicates the consistency of $\hat{\boldsymbol{\theta}}_{wls}$. $\blacksquare$

APPENDIX C
PROOF OF THEOREM 2

The proof is composed of four steps. First, a closed-form error expression for the DoA estimates is given through a Taylor series expansion method. The given closed-form expression involves the gradient and Hessian of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$. Hence, the corresponding gradient and Hessian is computed at the second and third steps, respectively. Finally, the covariance matrix of DoA estimation errors is obtained by combining the net results of preceding steps.

A. Closed-form expression for DoA estimation errors

From (15), we know that $\hat{\boldsymbol{\theta}}_{wls}$ is a critical point of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$, thus we have $\nabla_{\boldsymbol{\theta}} L(\hat{\boldsymbol{\theta}}_{wls}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = 0$ where $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ denotes the gradient of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$. $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ is a real-valued[1] function on $\mathbb{R}^K$, thereby applying Taylor theorem [37, Ch. 6, Theorem 12] around the true value of $\boldsymbol{\theta}$ yields

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) + \nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})(\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})$$
$$+ \left(\mathbf{I}_K \otimes (\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})^T \right) \mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}})(\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta}) = 0, \quad (54)$$

where $\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ denotes the Hessian matrix of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$, the matrix $\mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ is given in [37, Ch. 6, Definition 2], and $\boldsymbol{\theta}^* = \boldsymbol{\theta}(1 - t) + t\hat{\boldsymbol{\theta}}_{wls}$ for some $t \in (0, 1)$. From (54), we have

$$\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta} = -\left(\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) + \overbrace{\left(\mathbf{I}_K \otimes (\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})^T \right) \mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}})}^{\hbar} \right)^{-1}$$
$$\times \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}). \quad (55)$$

Noting that $\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H = \mathbf{U}_n \mathbf{U}_n^H + \mathcal{O}(\frac{1}{\sqrt{N}})$ [38], it is readily concluded that $\hat{\mathbf{Q}} = \mathbf{Q} + \mathcal{O}(\frac{1}{\sqrt{N}})$ for large N. Consequently, considering the fact that $\hat{\mathbf{r}} = \mathbf{r} + \mathcal{O}(\frac{1}{\sqrt{N}})$ for large N [38], [39] and making use of continuous differentiability of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$, it can readily be shown that $\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \mathbf{r}, \mathbf{Q}) + \mathcal{O}(\frac{1}{\sqrt{N}})$ and $\mathbf{H}(\boldsymbol{\theta}^*, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \mathbf{H}(\boldsymbol{\theta}^*, \mathbf{r}, \mathbf{Q}) + \mathcal{O}(\frac{1}{\sqrt{N}})$ for large N. On the other hand, since $\hat{\boldsymbol{\theta}}_{wls}$ is a consistent estimate of $\boldsymbol{\theta}$ according to Theorem 1, there exists $a > 0$ such that $\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta} = \mathcal{O}(\frac{1}{N^a})$ for large N. Hence, it follows that the Hessian term in (55) converges to a constant value as $N \to \infty$ while the term $\hbar$ converges to zero as $N \to \infty$. Therefor, it is possible to neglect the term $\hbar$ compared to the Hessian term in (55) in the asymptotic regime ($N \to \infty$), leading to

$$\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta} \simeq - \left(\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \mathbf{r}, \mathbf{Q}) \right)^{-1} \nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}). \quad (56)$$

B. Derivation of $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$

Taking derivative of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$ and exploiting the following expression for the derivative of projection matrix $\Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp}$ [40]

$$\frac{\partial \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp}}{\partial \theta_k} = -(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp}$$
$$- \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}} \mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k} (\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger}, \quad (57)$$

we get

$$\frac{\partial L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})}{\partial \theta_k} = -2\text{Re}\{ \hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}$$
$$\times (\mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}} \hat{\mathbf{r}} \}. \quad (58)$$

Making use of Lemma 8 in Appendix G, it can be readily shown that the term inside the $\text{Re}\{.\}$ operator is real-valued in case $\mathbf{K}_M \mathbf{W} = \mathbf{W}^* \mathbf{K}_M$. Hence, considering (14), (58) can be written as follows

$$\frac{\partial L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})}{\partial \theta_k} = -2\hat{\mathbf{p}}_{ls}^H (\mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}} \hat{\mathbf{r}} \quad (59)$$
$$= -\text{j} 2\pi [\hat{\mathbf{p}}_{ls}^H]_k \cos \theta_k \mathbf{a}_d^H(\theta_k) \text{diag}(\mathbf{d}) \mathbf{J}^H \mathbf{W}^{\frac{1}{2}} \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}} \hat{\mathbf{r}},$$

[1] The fact that $\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ is a real-valued function will be shown later in (58)

where

$$\mathbf{a}_d(\theta_k) = \left[e^{-\text{j}\pi \sin \theta_k \ell_{D-1}} \quad \cdots \quad 1 \quad \cdots \quad e^{\text{j}\pi \sin \theta_k \ell_{D-1}} \right]^T. \quad (60)$$

From (59) and using $\lim_{N \to \infty} \hat{\mathbf{p}}_{ls} = \mathbf{p}$ and $\lim_{N \to \infty} \hat{\mathbf{Q}} = \mathbf{Q}$, the gradient of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$ is given by

$$\nabla_{\boldsymbol{\theta}} L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) \quad (61)$$
$$\simeq -\text{j} 2\pi \text{diag}(\mathbf{p}) \boldsymbol{\Phi}(\boldsymbol{\theta}) \mathbf{A}_d^H(\boldsymbol{\theta}) \text{diag}(\mathbf{d}) \mathbf{J}^H \mathbf{W}^{\frac{1}{2}} \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}} \mathbf{Q} \hat{\mathbf{r}}.$$

where $\boldsymbol{\Phi}(\boldsymbol{\theta}) = \text{diag}(\left[\cos \theta_1 \quad \cdots \quad \cos \theta_K \right]^T).$

C. Derivation of $\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \mathbf{r}, \mathbf{Q})$

Taking the derivative of (58) with respect to θ_l and making use of (57) and the following expression for the derivative of $(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}$

$$\frac{\partial (\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}}{\partial \theta_l} = \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}} \mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_l} (\mathbf{A}_d^H(\boldsymbol{\theta}) \mathbf{J}^H \mathbf{W} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{-1}$$
$$- (\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_l})^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}, \quad (62)$$

leads to (63), which is shown at the top of this page. Given the fact that $\tilde{L}(\boldsymbol{\theta}_0) = 0$, it is possible to neglect the first four terms on the right had side of (63) compared to the last term as $N \to \infty$. Thus, by replacing $(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}} \hat{\mathbf{r}}$ from (14), we have

$$\frac{\partial^2 L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})}{\partial \theta_l \partial \theta_k} \simeq -2\pi^2 [\hat{\mathbf{p}}_{ls}^H]_k \cos \theta_k \mathbf{a}_d^H(\theta_k) \text{diag}(\mathbf{d}) \mathbf{J}^H \mathbf{W}^{\frac{1}{2}} \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}}$$
$$\times \mathbf{J} \text{diag}(\mathbf{d}) \mathbf{a}_d(\theta_l) \cos \theta_l [\hat{\mathbf{p}}_{ls}]_l. \quad (64)$$

Based on (64) and considering $\lim_{N \to \infty} \hat{\mathbf{p}}_{ls} = \mathbf{p}$ and $\lim_{N \to \infty} \hat{\mathbf{Q}} = \mathbf{Q}$, the Hessian matrix of $L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}})$ with respect to $\boldsymbol{\theta}$ is obtained as follows

$$\lim_{N \to \infty} \nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \hat{\mathbf{r}}, \hat{\mathbf{Q}}) = \nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}, \mathbf{r}, \mathbf{Q}) \quad (65)$$
$$\simeq -2\pi^2 \text{diag}(\mathbf{p}) \boldsymbol{\Phi}(\boldsymbol{\theta}) \mathbf{A}_d^H(\boldsymbol{\theta}) \text{diag}(\mathbf{d}) \mathbf{J}^H \mathbf{W}^{\frac{1}{2}} \Pi_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}^{\perp} \mathbf{W}^{\frac{1}{2}}$$
$$\times \mathbf{J} \text{diag}(\mathbf{d}) \mathbf{A}_d(\boldsymbol{\theta}) \boldsymbol{\Phi}(\boldsymbol{\theta}) \text{diag}(\mathbf{p}).$$

D. Calculation of the error covariance matrix

Combining (56), (61) and (65), it is possible to derive an asymptotic ($N \to \infty$) expression for the covariance matrix of $\hat{\boldsymbol{\theta}}_{wls}$ given at the bottom of this page. It is shown in [41] that

$$\mathbb{E}\{\hat{\mathbf{r}} \hat{\mathbf{r}}^H\} = \mathbf{r} \mathbf{r}^H + \frac{1}{N} \left(\mathbf{R}^T \otimes \mathbf{R} \right). \quad (67)$$

Inserting (67) into (66) and exploiting (47) gives (28).

APPENDIX D
PROOF OF THEOREM 3

Before proceeding to the main proof, let us first introduce the following preliminary lemmas.

Lemma 4. *The vector $\mathbf{b}$ belongs to the range space of the matrix $\mathbf{F}$, i.e., $\mathbf{b} \in \mathcal{R}(\mathbf{F})$.*

Proof. Recall $\mathbf{A}_v(\boldsymbol{\theta})$ from Appendix A, which is defined as the steering matrix of a ULA whose elements are located at $\left(0, \frac{\lambda}{2}, \lambda, \cdots, (v - 1)\frac{\lambda}{2} \right)$. It is observe that there is a selection matrix $\mathbf{Z} \in \{0, 1\}^{K^2 \times K}$ such that

$$\mathbf{A}_v^*(\boldsymbol{\theta}) \odot \mathbf{A}_v(\boldsymbol{\theta}) = (\mathbf{A}_v^*(\boldsymbol{\theta}) \otimes \mathbf{A}_v(\boldsymbol{\theta})) \mathbf{Z}. \quad (68)$$

$$
\begin{aligned}
\frac{\partial^2 L(\boldsymbol{\theta},\hat{\mathbf{r}},\hat{\mathbf{Q}})}{\partial \theta_l \partial \theta_k} =\ &-2\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_l} (\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H \mathbf{W}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{-1} (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}}\hat{\mathbf{r}} \\
&+2\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_l})^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}}\hat{\mathbf{r}} \\
&-2\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J}\frac{\partial^2 \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_l \partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}}\hat{\mathbf{r}} \\
&+2\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_l})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \hat{\mathbf{Q}}\hat{\mathbf{r}} \\
&+2\hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H} (\mathbf{J}\frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} \mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_l} (\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger} \hat{\mathbf{Q}}\hat{\mathbf{r}}.
\end{aligned}
\tag{63}
$$

Using (115), (68) and the fact that $\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n = 0$ leads to

$$
\mathbf{b}^H \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}) = \mathrm{vec}^H(\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H \mathbf{A}_v(\boldsymbol{\theta}))\mathbf{Z} = 0. \tag{69}
$$

(69) shows that $\mathbf{b}$ belongs to the null space of $\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H$. We know $\mathcal{N}(\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H) = \mathcal{R}(\mathbf{F})$. This concludes the proof. $\square$

Lemma 5. *The vector* $\mathbf{b}$ *belongs to the null space of the matrix* $\boldsymbol{\Omega}^H \doteq \boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H$, *i.e.,* $\mathbf{b} \in \mathcal{N}(\boldsymbol{\Omega}^H)$.

Proof. Let $\mathbf{a}_v(\theta_k) \doteq \begin{bmatrix} 1 & e^{\mathrm{j}\pi\sin\theta_k} & \cdots & e^{\mathrm{j}\pi\sin\theta_k(v-1)} \end{bmatrix}^T$ denote the kth column of $\mathbf{A}_v(\boldsymbol{\theta})$ and $\mathbf{v} \doteq \begin{bmatrix} 0 & 1 & \cdots & v-1 \end{bmatrix}^T$. Exploiting (60) and (115), we can show that

$$
\mathbf{T}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta}) \tag{70}
$$
$$
= \frac{-\mathrm{j}}{\pi} \begin{bmatrix} \frac{\partial \mathbf{T}\mathbf{a}_d(\theta_1)}{\partial \theta_1} & \frac{\partial \mathbf{T}\mathbf{a}_d(\theta_2)}{\partial \theta_2} & \cdots & \frac{\partial \mathbf{T}\mathbf{a}_d(\theta_K)}{\partial \theta_K} \end{bmatrix}
$$
$$
= \frac{-\mathrm{j}}{\pi} \begin{bmatrix} \frac{\partial \mathbf{a}_v^*(\theta_1)\otimes\mathbf{a}_v(\theta_1)}{\partial \theta_1} & \frac{\partial \mathbf{a}_v^*(\theta_2)\otimes\mathbf{a}_v(\theta_2)}{\partial \theta_2} & \cdots & \frac{\partial \mathbf{a}_v^*(\theta_K)\otimes\mathbf{a}_v(\theta_K)}{\partial \theta_K} \end{bmatrix}
$$
$$
= \frac{-\mathrm{j}}{\pi} (\mathrm{diag}(\mathbf{v})\mathbf{A}_v^*(\boldsymbol{\theta})\odot\mathbf{A}_v(\boldsymbol{\theta}) + \mathbf{A}_v^*(\boldsymbol{\theta})\odot\mathrm{diag}(\mathbf{v})\mathbf{A}_v(\boldsymbol{\theta}))\boldsymbol{\Phi}(\boldsymbol{\theta}).
$$

Exploiting (68), (70) and the fact that $\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n = 0$ gives

$$
\mathbf{b}^H \boldsymbol{\Omega} = \frac{1}{\mathrm{j}\pi} \big(\mathrm{vec}^H(\mathbf{A}_v^H(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H\mathrm{diag}(\mathbf{v})\mathbf{A}_v(\boldsymbol{\theta}))\mathbf{Z}\boldsymbol{\Phi}(\boldsymbol{\theta}) \tag{71}
$$
$$
+ \mathrm{vec}^H(\mathbf{A}_v^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{v})\mathbf{U}_n\mathbf{U}_n^H\mathbf{A}_v(\boldsymbol{\theta}))\mathbf{Z}\boldsymbol{\Phi}(\boldsymbol{\theta}) \big) = 0.
$$

This completes the proof. $\square$

Now, let define $\boldsymbol{\Psi} \doteq \begin{bmatrix} \mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) & \mathrm{vec}(\mathbf{I}_M) \end{bmatrix}$, the CRB expression given in [12, Theorem 2] can then be rewritten as follows

$$
CRB^{-1}(\boldsymbol{\theta}) = \frac{1}{\pi^2 N}\mathrm{diag}(\mathbf{p})\boldsymbol{\Omega}^H \mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{M}^{-1}\boldsymbol{\Psi}}\mathbf{M}^{-1}\boldsymbol{\Omega}\mathrm{diag}(\mathbf{p}). \tag{72}
$$

Based on the projection decomposition theorem [33], we have

$$
\Pi^{\perp}_{\mathbf{M}^{-1}\boldsymbol{\Psi}} = \Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta})} - \Pi_{\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta})}\mathbf{M}^{-1}\mathrm{vec}(\mathbf{I}_M)}
$$
$$
= \Pi_{\mathbf{MF}} - \Pi_{\Pi_{\mathbf{MF}}\mathbf{M}^{-1}\mathrm{vec}(\mathbf{I}_M)}, \tag{73}
$$

where the last equality is obtained by using the fact that $\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{MF}}$ [33]. Substituting (73) into (72) yields

$$
CRB^{-1}(\boldsymbol{\theta}) = \frac{1}{\pi^2 N}\mathrm{diag}(\mathbf{p})\boldsymbol{\Omega}^H \mathbf{F}\Big[(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1} \tag{74}
$$
$$
- \frac{(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M)\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}}{\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M)} \Big] \mathbf{F}^H\boldsymbol{\Omega}\mathrm{diag}(\mathbf{p}).
$$

Let decompose the vector $\mathrm{vec}(\mathbf{I}_M)$ as a sum of two vectors $\mathbf{h}_{\parallel}$ and $\mathbf{h}_{\perp}$ where

$$
\mathbf{h}_{\parallel} \doteq \mathrm{vec}(\mathbf{I}_M) - \frac{(v-K)\mathbf{M}^2\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}, \tag{75}
$$
$$
\mathbf{h}_{\perp} \doteq \frac{(v-K)\mathbf{M}^2\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}. \tag{76}
$$

It follows from (30), (89) and the definitions of $\mathbf{h}_{\parallel}$ and $\mathbf{h}_{\perp}$, given above, that

$$
\begin{aligned}
\mathbf{h}_{\perp}^H\mathbf{M}^{-2}\mathbf{h}_{\parallel} &= \frac{(v-K)\mathbf{b}^H\mathrm{vec}(\mathbf{I}_M)}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} - \frac{(v-K)^2\mathbf{b}^H\mathbf{M}^2\mathbf{b}}{(\mathbf{b}^H\mathbf{M}^2\mathbf{b})^2} \\
&= \frac{(v-K)\mathrm{vec}^H(\mathbf{U}_n\mathbf{U}_n^H)\mathbf{Tg}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} - \frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} \\
&= \frac{(v-K)\mathrm{vec}^H(\mathbf{U}_n\mathbf{U}_n^H)\mathrm{vec}(\mathbf{I}_v)}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} - \frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} = 0. \tag{77}
\end{aligned}
$$

Accordingly and making use of the fact that $\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{MF}}$, we obtain

$$
\begin{aligned}
&\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M) \\
&= \mathrm{vec}^H(\mathbf{I}_M)\mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta})}\mathbf{M}^{-1}\mathrm{vec}(\mathbf{I}_M) \\
&= \mathrm{vec}^H(\mathbf{I}_M)\mathbf{M}^{-2}\mathrm{vec}(\mathbf{I}_M) - \mathbf{h}_{\parallel}^H\mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta})}\mathbf{M}^{-1}\mathbf{h}_{\parallel} \\
&\quad - \mathbf{h}_{\perp}^H\mathbf{M}^{-1}(\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H}\mathbf{A}_d(\boldsymbol{\theta})^H\mathbf{J}^H\mathbf{M}^{-2}\mathbf{h}_{\perp} \\
&\quad - \mathbf{h}_{\parallel}^H\mathbf{M}^{-1}(\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger H}\mathbf{A}_d(\boldsymbol{\theta})^H\mathbf{J}^H\mathbf{M}^{-2}\mathbf{h}_{\perp} \\
&\quad - \mathbf{h}_{\perp}^H\mathbf{M}^{-2}\mathbf{JA}_d(\boldsymbol{\theta})(\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta}))^{\dagger}\mathbf{M}^{-1}\mathbf{h}_{\perp}. \tag{78}
\end{aligned}
$$

From the definition of $\mathbf{h}_{\perp}$ in (76) and Lemma 4, we have

$$
\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H\mathbf{M}^{-2}\mathbf{h}_{\perp} = \frac{(v-K)\mathbf{A}_d^H(\boldsymbol{\theta})\mathbf{J}^H\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} = \mathbf{0}. \tag{79}
$$

Inserting (79) into (78) leads the last three terms in (78) to vanish. Then, exploiting $\Pi^{\perp}_{\mathbf{M}^{-1}\mathbf{JA}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{MF}}$ once again yields

$$
\begin{aligned}
&\mathrm{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathrm{vec}(\mathbf{I}_M) \tag{80} \\
&= \mathrm{vec}^H(\mathbf{I}_M)\mathbf{M}^{-2}\mathrm{vec}(\mathbf{I}_M) - \mathbf{h}_{\parallel}^H\mathbf{M}^{-1}\Pi^{\perp}_{\mathbf{MF}}\mathbf{M}^{-1}\mathbf{h}_{\parallel}.
\end{aligned}
$$

Given $\mathrm{vec}(\mathbf{I}_M) = \mathbf{h}_{\parallel} + \mathbf{h}_{\perp}$ and from (77), it is observed that

$$
\mathrm{vec}^H(\mathbf{I}_M)\mathbf{M}^{-2}\mathrm{vec}(\mathbf{I}_M) = \mathbf{h}_{\parallel}^H\mathbf{M}^{-2}\mathbf{h}_{\parallel} + \mathbf{h}_{\perp}^H\mathbf{M}^{-2}\mathbf{h}_{\perp}. \tag{81}
$$

$$
\begin{aligned}
&\mathbb{E}\left\{ (\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_{wls} - \boldsymbol{\theta})^H \right\} \simeq \mathbb{E}\left\{ (\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}}))^{-1}\nabla_{\boldsymbol{\theta}}L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}})(\nabla_{\boldsymbol{\theta}}L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}}))^H (\nabla_{\boldsymbol{\theta}}^2 L(\boldsymbol{\theta}_0,\hat{\mathbf{r}},\hat{\mathbf{Q}}))^{-1} \right\} \\
&= \frac{1}{\pi^2}(\mathrm{diag}(\mathbf{p})\boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p}))^{-1} \\
&\quad \times \mathrm{diag}(\mathbf{p})\boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{Q}\mathbb{E}\{\hat{\mathbf{r}}\hat{\mathbf{r}}^H\}\mathbf{Q}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p}) \\
&\quad \times (\mathrm{diag}(\mathbf{p})\boldsymbol{\Phi}^H(\boldsymbol{\theta})\mathbf{A}_d^H(\boldsymbol{\theta})\mathrm{diag}(\mathbf{d})\mathbf{J}^H\mathbf{W}^{\frac{1}{2}}\Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}}\mathbf{JA}_d(\boldsymbol{\theta})}\mathbf{W}^{\frac{1}{2}}\mathbf{J}\mathrm{diag}(\mathbf{d})\mathbf{A}_d(\boldsymbol{\theta})\boldsymbol{\Phi}(\boldsymbol{\theta})\mathrm{diag}(\mathbf{p}))^{-1}.
\end{aligned}
\tag{66}
$$

Inserting (81) into (80) leads to

$$\text{vec}^H(\mathbf{I}_M)\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\text{vec}(\mathbf{I}_M)$$
$$= \frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} + \mathbf{h}_\parallel^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_\parallel. \quad (82)$$

Further, making use of (79) and Lemma 5, we can show that

$$\boldsymbol{\Omega}^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_\perp = \boldsymbol{\Omega}^H\mathbf{M}^{-1}\Pi_{\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}^\perp\mathbf{M}^{-1}\mathbf{h}_\perp$$
$$= \boldsymbol{\Omega}^H\mathbf{M}^{-2}\mathbf{h}_\perp$$
$$\quad - \boldsymbol{\Omega}^H\mathbf{M}^{-1}(\mathbf{M}^{-1}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}\mathbf{A}_d(\boldsymbol{\theta})^H\mathbf{J}\mathbf{M}^{-2}\mathbf{h}_\perp$$
$$= \frac{(v-K)\boldsymbol{\Omega}^H\mathbf{b}}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} = \mathbf{0}. \quad (83)$$

Substituting (80) and (83) into (74) gives

$$CRB^{-1}(\boldsymbol{\theta}) = \frac{1}{\pi^2 N}\text{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{F}\left[(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\right.$$
$$\left. - \frac{(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_\parallel\mathbf{h}_\parallel^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}}{\frac{(v-K)^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}} + \mathbf{h}_\parallel^H\mathbf{F}(\mathbf{F}^H\mathbf{M}^2\mathbf{F})^{-1}\mathbf{F}^H\mathbf{h}_\parallel}\right]\mathbf{F}^H\boldsymbol{\Omega}\text{diag}(\mathbf{p})$$
$$= \frac{1}{\pi^2 N}\text{diag}(\mathbf{p})\boldsymbol{\Omega}^H\mathbf{F} \quad (84)$$
$$\times \left(\mathbf{F}^H\left(\mathbf{M}^2 + \frac{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}{(v-K)^2}\mathbf{h}_\parallel\mathbf{h}_\parallel^H\right)\mathbf{F}\right)^{-1}\mathbf{F}^H\boldsymbol{\Omega}\text{diag}(\mathbf{p}),$$

where the last equality is obtained by using the matrix inversion lemma [40]. Now, given the definition of $\mathbf{h}_\parallel$ in (75), it is possible to show that

$$\mathbf{M}^2 + \frac{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}{(v-K)^2}\mathbf{h}_\parallel\mathbf{h}_\parallel^H = \mathbf{Q}\mathbf{M}^2\mathbf{Q} + \frac{\mathbf{M}^2\mathbf{b}\mathbf{b}^H\mathbf{M}^2}{\mathbf{b}^H\mathbf{M}^2\mathbf{b}}. \quad (85)$$

Inserting (85) into (84) completes the proof.

APPENDIX E
PROOF OF THEOREM 4

The proof is compromised of two steps. The first step involves simplification of the covariance matrix of DoA estimation errors through inserting the optimal weighting matrix given in Theorem 4 into (28). At the second step, we simplify the CRB expression given in Theorem 3 through doing some algebraic manipulations and show that the CRB coincides with the simplified covariance matrix of DoA estimation errors given in the first step.

A. Simplification of the errors covariance matrix

Considering the expression for $\mathbf{W}_{\text{opt}}$ from Theorem 4 and exploiting $\mathbf{F}^H\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta}) = 0$ and $\Pi_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}^\perp = \Pi_{\mathbf{W}_{\text{opt}}^{-\frac{1}{2}}\mathbf{F}}$ results in

$$\mathbf{W}_{opt}^{\frac{1}{2}}\Pi_{\mathbf{W}_{opt}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}^\perp\mathbf{W}_{opt}^{\frac{1}{2}} = \mathbf{F}(\mathbf{F}^H\mathbf{S}\mathbf{F})^{-1}\mathbf{F}^H. \quad (86)$$

Let now introduce the following Lemma, whereby we can proceed further with simplification of (86).

Lemma 6. The matrix $\mathbf{Q}\mathbf{M}^2\mathbf{Q}^H \in \mathbb{C}^{M^2\times M^2}$ can be decomposed as $\mathbf{V}\mathbf{V}^H$, where $\mathbf{V}^H \in \mathbb{C}^{M^2\times M^2}$ is a singular matrix with rank $M^2 - 1$ whose null space is spanned by $\mathbf{b}$.

Proof. For any given matrices $\mathbf{B}$ and $\mathbf{C}$, it is known that $\text{rank}(\mathbf{B} - \mathbf{C}) \geq |\text{rank}(\mathbf{B}) - \text{rank}(\mathbf{C})|$ [40]. Hence, recalling the definition of $\mathbf{Q}$ in (29), we have

$$\text{rank}(\mathbf{Q}) \geq M^2 - 1. \quad (87)$$

On the other hand, given (30), we obtain

$$\mathbf{b}^H\mathbf{Q} = \mathbf{b}^H - \frac{\text{vec}^H\left(\mathbf{U}_n\mathbf{U}_n^H\right)\mathbf{T}\mathbf{g}}{v-K}\mathbf{b}^H. \quad (88)$$

Let $\mathbf{g}' \in \{0,1\}^{(2v-1)\times 1}$ be a column vector with $[\mathbf{g}']_i = \delta[i-v]$. Utilizing (116), (114) and [12, Corollary 3], we observe

$$\mathbf{T}\mathbf{g} = \mathbf{T}'\mathbf{g}' = \text{vec}(\mathbf{I}_v). \quad (89)$$

where $\mathbf{T}' \in \{0,1\}^{v^2\times 2v-1}$ is defined in (114) in Appendix G. Inserting (89) into (88) gives

$$\mathbf{b}^H\mathbf{Q} = \mathbf{b}^H - \frac{\text{tr}(\mathbf{U}_n\mathbf{U}_n^H)}{v-K}\mathbf{b}^H = \mathbf{b}^H - \mathbf{b}^H = 0. \quad (90)$$

which implies that

$$\text{rank}(\mathbf{Q}) \leq M^2 - 1. \quad (91)$$

Comparing (87) and (91) concludes that $\text{rank}(\mathbf{Q}) = M^2 - 1$. The fact $\text{rank}(\mathbf{Q}) = M^2 - 1$ implies that the dimension of the null space of $\mathbf{Q}^H$ is equal to 1. In addition, it follows from (88) that $\mathbf{b} \in \mathcal{N}(\mathbf{Q}^H)$. This means that the vector $\mathbf{b}$ spans the null space of $\mathbf{Q}^H$.

Let now define the matrix $\mathbf{V} = \mathbf{Q}\mathbf{M}$. Since $\mathbf{R}$ is positive definite, the matrix $\mathbf{M} = (\mathbf{R}^T \otimes \mathbf{R})^{\frac{1}{2}}$ is also positive definite, implying that $\mathcal{R}(\mathbf{V}) = \mathcal{R}(\mathbf{Q})$. Hence, $\mathbf{V}^H$ is also a matrix of rank $M^2 - 1$ whose null space is spanned by the vector $\mathbf{b}$. $\square$

It follows from Lemma 4 and Lemma 6 that $\mathcal{N}(\mathbf{V}^H) \subset \mathcal{R}(\mathbf{F})$ and, in turn, $\mathcal{N}(\mathbf{F}^H) \subset \mathcal{R}(\mathbf{V})$. This implies that a K-dimensional subspace of $\mathcal{R}(\mathbf{V})$ is a spanning set for the null space of $\mathbf{F}^H$. Hence, making use of Lemma 6 and recalling the definition of $\mathbf{S}$ in (36) result in

$$\mathbf{F}^H\mathbf{S}\mathbf{F} = \mathbf{F}^H\Pi_{\mathbf{F}}\mathbf{V}\mathbf{V}^H\Pi_{\mathbf{F}}\mathbf{F} + \mathbf{F}^H\mathbf{b}\mathbf{b}^H\mathbf{F}$$
$$= \mathbf{F}^H\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{V}}^H \\ \mathbf{b}^H\end{bmatrix}\mathbf{F}, \quad (92)$$

where $\bar{\mathbf{V}} = \mathbf{U}_v\boldsymbol{\Lambda}_v \in \mathbb{C}^{M^2\times(M^2-K-1)}$ with $\mathbf{U}_v \in \mathbb{C}^{M^2\times(M^2-K-1)}$ being comprised of left-singular vectors of $\Pi_{\mathbf{F}}\mathbf{V}$ and the diagonal matrix $\boldsymbol{\Lambda}_v \in \mathbb{C}^{(M^2-K-1)\times(M^2-K-1)}$ its corresponding singular values, meaning that

$$\mathcal{R}(\bar{\mathbf{V}}) \subset \mathcal{R}(\mathbf{F}), \quad (93)$$
$$\mathcal{R}(\bar{\mathbf{V}}) \subset \mathcal{R}(\mathbf{V}), \quad (94)$$
$$\text{rank}(\bar{\mathbf{V}}) = M^2 - K - 1. \quad (95)$$

It follows from (94) and Lemma 6 that $\mathbf{b} \notin \mathcal{R}(\bar{\mathbf{V}})$. From (95) and the fact that $\mathbf{b} \notin \mathcal{R}(\bar{\mathbf{V}})$, it can be deduced that

$$\text{rank}(\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix}) = M^2 - K = \text{rank}(\mathbf{F}). \quad (96)$$

Comparing (93), (95) and (96) proves that $\mathcal{R}(\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix}) = \mathcal{R}(\mathbf{F})$. Hence, there is a full rank matrix $\mathbf{X} \in \mathbb{C}^{(M^2-K)\times(M^2-K)}$ such that

$$\begin{bmatrix}\bar{\mathbf{V}} & \mathbf{b}\end{bmatrix} = \mathbf{F}\mathbf{X}. \quad (97)$$

Inserting (92) and (97) into (86) yields

$$\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\Pi_{\mathbf{W}_{\text{opt}}^{\frac{1}{2}}\mathbf{J}\mathbf{A}_d(\boldsymbol{\theta})}^\perp\mathbf{W}_{\text{opt}}^{\frac{1}{2}} = \mathbf{F}^{\dagger H}\mathbf{X}^{-H}\mathbf{X}^{-1}\mathbf{F}^\dagger = (\mathbf{F}\mathbf{X})^{\dagger H}(\mathbf{F}\mathbf{X})^\dagger$$
$$= \begin{bmatrix}\bar{\mathbf{V}}^{\dagger H} & \mathbf{b}^{\dagger H}\end{bmatrix}\begin{bmatrix}\bar{\mathbf{V}}^\dagger \\ \mathbf{b}^\dagger\end{bmatrix}. \quad (98)$$

Considering (98), Lemma 5 and $\mathcal{R}(\mathbf{b}^{\dagger H}) = \mathcal{R}(\mathbf{b})$ gives

$$\mathbf{\Omega}^H \mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} = \mathbf{\Omega}^H \begin{bmatrix} \bar{\mathbf{V}}^{\dagger H} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}^{\dagger} \\ \mathbf{b}^{\dagger} \end{bmatrix}$$
$$= \mathbf{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger}. \qquad (99)$$

Moreover, given (92) and using $\Pi^{\perp}_{\mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \mathbf{JA}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{W}_{\mathrm{opt}}^{-\frac{1}{2}} \mathbf{F}}$, we have

$$\Pi^{\perp}_{\mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \mathbf{Q}\mathbf{M}^2\mathbf{Q}^H \mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \mathbf{JA}_d(\boldsymbol{\theta})}$$
$$= \mathbf{W}_{\mathrm{opt}}^{-\frac{1}{2}} \mathbf{F}(\mathbf{F}^H \mathbf{W}_{\mathrm{opt}}^{-1} \mathbf{F})^{-1} \mathbf{F}^H \bar{\mathbf{V}}\bar{\mathbf{V}}^H \mathbf{F}(\mathbf{F}^H \mathbf{W}_{\mathrm{opt}}^{-1} \mathbf{F})^{-1} \mathbf{F}^H \mathbf{W}_{\mathrm{opt}}^{-\frac{1}{2}}$$
$$= \Pi^{\perp}_{\mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \bar{\mathbf{V}}\bar{\mathbf{V}}^H \mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}_{\mathrm{opt}}^{\frac{1}{2}} \mathbf{JA}_d(\boldsymbol{\theta})}. \qquad (100)$$

Eventually, substituting (99) and (100) into (28), we find

$$\mathbf{C}_{wls} = \frac{1}{\pi^2 N} \mathrm{diag}^{-1}(\mathbf{p}) \left(\mathbf{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \mathbf{\Omega}\right)^{-1} \times$$
$$\left(\mathbf{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \bar{\mathbf{V}}\bar{\mathbf{V}}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \mathbf{\Omega}\right) \left(\mathbf{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \mathbf{\Omega}\right)^{-1} \mathrm{diag}^{-1}(\mathbf{p})$$
$$= \frac{1}{\pi^2 N} \left(\mathrm{diag}(\mathbf{p}) \mathbf{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \mathbf{\Omega} \mathrm{diag}(\mathbf{p})\right)^{-1} \qquad (101)$$

B. Simplification of the CRB expression

Now, we will show that the CRB expression given in Lemma 3 is also reduced to (101). Let decompose the vector $\frac{\mathbf{M}^2\mathbf{b}}{\sqrt{\mathbf{b}^H \mathbf{M}^2 \mathbf{b}}}$ as a sum of two vectors $\mathbf{q}_{\parallel}$ and $\mathbf{q}_{\perp}$ where

$$\mathbf{q}_{\parallel} = \frac{\sqrt{\mathbf{b}^H \mathbf{M}^2 \mathbf{b}}}{\|\mathbf{b}\|^2} \mathbf{b}, \quad \mathbf{q}_{\perp} = \frac{\mathbf{M}^2\mathbf{b}}{\sqrt{\mathbf{b}^H \mathbf{M}^2 \mathbf{b}}} - \frac{\sqrt{\mathbf{b}^H \mathbf{M}^2 \mathbf{b}}}{\|\mathbf{b}\|^2} \mathbf{b}. \qquad (102)$$

It is observed that $\mathbf{q}_{\parallel}$ is a scaled version of $\mathbf{b}$ and that $\mathbf{b}^H \mathbf{q}_{\perp} = 0$. Hence, recalling (34) and taking account of (92), we have

$$\mathbf{F}^H \mathbf{H} \mathbf{F} = \mathbf{F}^H \mathbf{V}\mathbf{V}\mathbf{F} + \mathbf{F}^H (\mathbf{q}_{\parallel} + \mathbf{q}_{\perp})(\mathbf{q}_{\parallel} + \mathbf{q}_{\perp})^H \mathbf{F}$$
$$= \mathbf{F}^H \begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\parallel} + \bar{\mathbf{q}}_{\perp} \end{bmatrix} \begin{bmatrix} \bar{\mathbf{V}}^H \\ \mathbf{q}_{\parallel}^H + \bar{\mathbf{q}}_{\perp}^H \end{bmatrix} \mathbf{F}, \qquad (103)$$

where $\bar{\mathbf{q}}_{\perp} = \Pi_{\mathbf{F}} \mathbf{q}_{\perp}$. By definition, it is evident that $\bar{\mathbf{q}}_{\perp} \in \mathcal{R}(\mathbf{F})$ and $\mathbf{b}^H \bar{\mathbf{q}}_{\perp} = 0$, meaning that $\bar{\mathbf{q}}_{\perp} \in \mathcal{R}(\bar{\mathbf{V}})$. In addition, since $\mathbf{b} \notin \mathcal{R}(\bar{\mathbf{V}})$, $\mathbf{q}_{\parallel} \notin \mathcal{R}(\bar{\mathbf{V}})$ in turn. In consequence, considering (93) and (95), it can be inferred that $\mathcal{R}(\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\parallel} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}) = \mathcal{R}(\mathbf{F})$ implying that there is a full rank matrix $\mathbf{D} \in \mathbb{C}^{(M^2-K) \times (M^2-K)}$ such that

$$\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\parallel} + \bar{\mathbf{q}}_{\perp} \end{bmatrix} = \mathbf{F}\mathbf{D}. \qquad (104)$$

By inserting (104) into (103) and doing some calculations similar to (98), we obtain

$$\mathbf{F}(\mathbf{F}^H \mathbf{H} \mathbf{F})^{-1} \mathbf{F}^H = \begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\parallel} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}^{\dagger H} \begin{bmatrix} \bar{\mathbf{V}}^H \\ \mathbf{q}_{\parallel}^H + \bar{\mathbf{q}}_{\perp}^H \end{bmatrix}^{\dagger}. \qquad (105)$$

Since $\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\parallel} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}$ is a full column rank matrix, its pseudoinverse by definition can be computed as

$$\begin{bmatrix} \bar{\mathbf{V}} & \mathbf{q}_{\parallel} + \bar{\mathbf{q}}_{\perp} \end{bmatrix}^{\dagger} = \begin{bmatrix} \bar{\mathbf{V}}\bar{\mathbf{V}}^H & \bar{\mathbf{V}}^H \bar{\mathbf{q}}_{\perp} \\ \bar{\mathbf{q}}_{\perp}^H \bar{\mathbf{V}} & \|\mathbf{q}_{\parallel}\|^2 + \|\bar{\mathbf{q}}_{\perp}\|^2 \end{bmatrix}^{-1} \begin{bmatrix} \bar{\mathbf{V}}^H \\ \mathbf{q}_{\parallel}^H + \bar{\mathbf{q}}_{\perp}^H \end{bmatrix} =$$

$$\begin{bmatrix} \bar{\mathbf{V}}^{\dagger} + \frac{\bar{\mathbf{V}}^{\dagger} \bar{\mathbf{q}}_{\perp} \bar{\mathbf{q}}_{\perp}^H \Pi_{\bar{\mathbf{V}}} - \bar{\mathbf{V}}^{\dagger} \bar{\mathbf{q}}_{\perp} (\mathbf{q}_{\parallel}^H + \bar{\mathbf{q}}_{\perp}^H)}{\|\mathbf{q}_{\parallel}\|^2 + \|\bar{\mathbf{q}}_{\perp}\|^2 - \bar{\mathbf{q}}_{\perp}^H \Pi_{\bar{\mathbf{V}}} \bar{\mathbf{q}}_{\perp}} \\ \frac{-\bar{\mathbf{q}}_{\perp}^H \Pi_{\bar{\mathbf{V}}} + \mathbf{q}_{\parallel}^H + \bar{\mathbf{q}}_{\perp}^H}{\|\mathbf{q}_{\parallel}\|^2 + \|\bar{\mathbf{q}}_{\perp}\|^2 - \bar{\mathbf{q}}_{\perp}^H \Pi_{\bar{\mathbf{V}}} \bar{\mathbf{q}}_{\perp}} \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{V}}^{\dagger} - \frac{\bar{\mathbf{V}}^{\dagger} \bar{\mathbf{q}}_{\perp} \mathbf{q}_{\parallel}^H}{\|\mathbf{q}_{\parallel}\|^2} \\ \mathbf{q}_{\parallel}^{\dagger} \end{bmatrix},$$
$$(106)$$

where that last equality is obtained by exploiting the fact that $\bar{\mathbf{q}}_{\perp}$ belongs to the range space of $\bar{\mathbf{V}}$. Now, inserting (105) and (106) into the CRB expression given in (33) gives

$$CRB(\boldsymbol{\theta}) = (\mathrm{diag}(\mathbf{p}) \mathbf{\Omega}^H \begin{bmatrix} \bar{\mathbf{V}}^{\dagger H} - \frac{\mathbf{q}_{\parallel} \bar{\mathbf{q}}_{\perp}^H \bar{\mathbf{V}}^{\dagger H}}{\|\mathbf{q}_{\parallel}\|^2} & \mathbf{q}_{\parallel}^{\dagger H} \end{bmatrix}$$
$$\times \begin{bmatrix} \bar{\mathbf{V}}^{\dagger} - \frac{\bar{\mathbf{V}}^{\dagger} \bar{\mathbf{q}}_{\perp} \mathbf{q}_{\parallel}^H}{\|\mathbf{q}_{\parallel}\|^2} \\ \mathbf{q}_{\parallel}^{\dagger} \end{bmatrix} \mathbf{\Omega} \mathrm{diag}(\mathbf{p}))^{-1}$$
$$= \frac{1}{\pi^2 N} \left(\mathrm{diag}(\mathbf{p}) \mathbf{\Omega}^H (\bar{\mathbf{V}}\bar{\mathbf{V}}^H)^{\dagger} \mathbf{\Omega} \mathrm{diag}(\mathbf{p})\right)^{-1}, \qquad (107)$$

where the last equality is obtained by using Lemma 5 and the fact that $\mathbf{q}_{\parallel}$ is a scaled version of $\mathbf{b}$.

Eventually, comparing (101) into (107) concludes the proof.

APPENDIX F
PROOF OF LEMMA 2

We first prove positive definiteness of $\mathbf{W}_{\mathrm{opt}}$. It follows from (92) to (97) in Appendix E-A that

$$\Pi^{\perp}_{\mathbf{JA}_d(\boldsymbol{\theta})} \mathbf{S} \Pi^{\perp}_{\mathbf{JA}_d(\boldsymbol{\theta})} = \Pi_{\mathbf{F}} \mathbf{S} \Pi_{\mathbf{F}} = \mathbf{F}\mathbf{X}\mathbf{X}^H \mathbf{F}^H. \qquad (108)$$

Consequently, we have

$$\mathbf{W}_{\mathrm{opt}}^{-1} = \begin{bmatrix} \mathbf{FX} & \mathbf{JA}_d(\boldsymbol{\theta}) \end{bmatrix} \begin{bmatrix} \mathbf{FX} & \mathbf{JA}_d(\boldsymbol{\theta}) \end{bmatrix}^H. \qquad (109)$$

The matrices $\mathbf{FX} \in \mathbb{C}^{M^2 \times (M^2-K)}$ and $\mathbf{JA}_d(\boldsymbol{\theta}) \in \mathbb{C}^{M^2 \times K}$ are full column rank by definition. Further, it is easily checked that $\mathbf{FX}$ and $\mathbf{JA}_d(\boldsymbol{\theta})$ are orthogonal subspaces, i.e., $\mathbf{X}^H \mathbf{F}^H \mathbf{JA}_d(\boldsymbol{\theta}) = \mathbf{0}$. Hence, $\begin{bmatrix} \mathbf{FX} & \mathbf{JA}_d(\boldsymbol{\theta}) \end{bmatrix} \in \mathbb{C}^{M^2 \times M^2}$ is full rank. This implies that $\mathbf{W}_{\mathrm{opt}}$ is also full rank and thereby positive definite.

Now we show that the estimate of $\mathbf{W}_{\mathrm{opt}}$, obtained from either CAB-MUSIC or CAB-ESPRIT, is always positive definite regardless of the available number of snapshots. Let assume that an arbitrary number of snapshots is available and $\hat{\mathbf{S}} = \hat{\mathbf{Q}} \hat{\mathbf{M}}^2 \hat{\mathbf{Q}}^H + \hat{\mathbf{b}}\hat{\mathbf{b}}^H$ denotes the estimate of $\mathbf{S}$, obtained based on the sample covarinace matrix. Further, let $\hat{\boldsymbol{\theta}}$ be an estimate of $\boldsymbol{\theta}$ given by CAB-MUSIC or CAB-ESPRIT and $\hat{\mathbf{F}}$ is the estimate of $\mathbf{F}$ obtained from $\hat{\boldsymbol{\theta}}$. Following similar kind of arguments and derivations provided from Lemma 6 to (97) in Appendix E-A, it can readily be shown that

$$\Pi^{\perp}_{\mathbf{JA}_d(\hat{\boldsymbol{\theta}})} \hat{\mathbf{S}} \Pi^{\perp}_{\mathbf{JA}_d(\hat{\boldsymbol{\theta}})} = \Pi_{\hat{\mathbf{F}}} \hat{\mathbf{S}} \Pi_{\hat{\mathbf{F}}} = \hat{\mathbf{F}}\hat{\mathbf{X}}\hat{\mathbf{X}}^H \hat{\mathbf{F}}^H, \qquad (110)$$

where $\hat{\mathbf{X}} \in \mathbb{C}^{(M^2-K) \times (M^2-K)}$ is a full rank matrix. Hence, using (110), we observe that

$$\hat{\mathbf{W}}_{opt}^{-1} = \begin{bmatrix} \hat{\mathbf{F}}\hat{\mathbf{X}} & \mathbf{JA}_d(\hat{\boldsymbol{\theta}}) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{F}}\hat{\mathbf{X}} & \mathbf{JA}_d(\hat{\boldsymbol{\theta}}) \end{bmatrix}^H. \qquad (111)$$

Once again, we note that $\hat{\mathbf{F}}\hat{\mathbf{X}} \in \mathbb{C}^{M^2 \times (M^2-K)}$ and $\mathbf{JA}_d(\hat{\boldsymbol{\theta}}) \in \mathbb{C}^{M^2 \times K}$ are full column rank by definition, and moreover, they span orthogonal subspaces. Hence, $\begin{bmatrix} \hat{\mathbf{F}}\hat{\mathbf{X}} & \mathbf{JA}_d(\hat{\boldsymbol{\theta}}) \end{bmatrix} \in \mathbb{C}^{M^2 \times M^2}$ and in turn $\hat{\mathbf{W}}_{\mathrm{opt}}$ is full rank.

APPENDIX G
COMMUTATION MATRIX AND SOME RELEVANT LEMMAS

Definition 2. Let $\mathbf{B}$ be any matrix in $\mathbb{R}^{p \times p}$. Then, there exists a permutation matrix $\mathbf{K}_p \in \{0,1\}^{p^2 \times p^2}$ such that $\text{vec}(\mathbf{B}^T) = \mathbf{K}_p \text{vec}(\mathbf{B})$. This matrix, called the commutation matrix, is an involutory and symmetric matrix, i.e., $\mathbf{K}_p = \mathbf{K}_p^T = \mathbf{K}_p^{-1}$ [37].

Lemma 7. $\mathbf{K}_M \mathbf{J}^{\dagger H} \mathbf{T}^H \mathbf{K}_v = \mathbf{J}^{\dagger H} \mathbf{T}^H$ where $\mathbf{K}_M \in \{0,1\}^{M^2 \times M^2}$ and $\mathbf{K}_v \in \{0,1\}^{v^2 \times v^2}$ are commutation matrices defined according to Definition 2.

Proof. It has been proved in [12] that $\mathbf{J}$ has orthogonal columns. Further, recalling the definition of $\mathbf{J}$ given in Definition 1, since $\|\text{vec}(\mathbf{L}_i^T)\|^2 = \|\text{vec}(\mathbf{L}_i)\|^2$ for $0 < n < D-1$, it is readily confirmed that

$$(\mathbf{J}^H \mathbf{J})^{-1} = \text{diag}\left(\left[\tfrac{1}{\|\text{vec}(\mathbf{L}_{D-1})\|^2} \cdots \tfrac{1}{\|\text{vec}(\mathbf{L}_0)\|^2} \cdots \tfrac{1}{\|\text{vec}(\mathbf{L}_{D-1})\|^2}\right]^T\right). \quad (112)$$

Recalling $\mathbf{A}_\vartheta$ and $\mathbf{A}_v$ in the proof of Lemmas 1 and 3, and based on [12, Appendix B], it is straightforward to show that

$$\mathbf{A}_v^*(\boldsymbol{\theta}) \odot \mathbf{A}_v(\boldsymbol{\theta}) = \mathbf{T}' \mathbf{A}_\vartheta(\boldsymbol{\theta}), \quad (113)$$

where $\mathbf{T}' \in \{0,1\}^{v^2 \times 2v-1}$ can be defined like $\mathbf{J}$ as

$$\mathbf{T}' = \begin{bmatrix} \text{vec}(\mathbf{G}_{v-1}^T) & \cdots & \text{vec}(\mathbf{G}_0) & \cdots & \text{vec}(\mathbf{G}_{v-1}) \end{bmatrix}, \quad (114)$$

where $[\mathbf{G}_n]_{p,q} = \begin{cases} 1, & \text{if } p-q=n, \\ 0, & \text{otherwise.} \end{cases}$. On the other hand, comparing the vectorized form of (43) and (44) gives

$$\mathbf{T}\mathbf{A}_d(\boldsymbol{\theta}) = \mathbf{A}_v^*(\boldsymbol{\theta}) \odot \mathbf{A}_v(\boldsymbol{\theta}). \quad (115)$$

Making use of (113), (115) and the fact that $\mathbf{A}_\vartheta(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{0}_{(2v-1) \times (D-v)} & \mathbf{I}_{2v-1} & \mathbf{0}_{(2v-1) \times (D-v)} \end{bmatrix} \mathbf{A}_d(\boldsymbol{\theta})$ results in

$$\mathbf{T} = \begin{bmatrix} \mathbf{0}_{v^2 \times (D-v)} & \mathbf{T}' & \mathbf{0}_{v^2 \times (D-v)} \end{bmatrix}. \quad (116)$$

Combining (5), (112), (114) and (116) gives

$$\mathbf{K}_M \mathbf{J}^{\dagger H} \mathbf{T}^H \mathbf{K}_v = \mathbf{J}^{\dagger H} \mathbf{T}^H = \frac{\text{vec}(\mathbf{L}_0)\text{vec}^T(\mathbf{G}_0)}{\|\text{vec}(\mathbf{L}_0)\|^2} \quad (117)$$
$$+ \sum_{i=1}^{v-1} \frac{\text{vec}(\mathbf{L}_i)\text{vec}^T(\mathbf{G}_i) + \text{vec}(\mathbf{L}_i^T)\text{vec}^T(\mathbf{G}_i^T)}{\|\text{vec}(\mathbf{L}_i)\|^2}. \quad \square$$

Lemma 8. *Let* $\mathbf{K}_M \in \{0,1\}^{M^2 \times M^2}$ *be the commutation matrix as defined according to Definition 2 and* $\mathbf{K}_M \mathbf{W} = \mathbf{W}^* \mathbf{K}_M$. *Then, it follows that*

a) $\hat{\mathbf{Q}}^* \hat{\mathbf{r}}^* = \mathbf{K}_M \hat{\mathbf{Q}} \hat{\mathbf{r}}$,

b) $\hat{\mathbf{r}}^T \hat{\mathbf{Q}}^T \mathbf{K}_M = \hat{\mathbf{r}}^H \hat{\mathbf{Q}}^H$,

c) $(\mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^T = (\mathbf{J} \frac{\partial \mathbf{A}_d(\boldsymbol{\theta})}{\partial \theta_k})^H \mathbf{K}_M$,

d) $\mathbf{K}_M \mathbf{W}^{\frac{1}{2}*}(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger T} = \mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H}$,

e) $\mathbf{K}_M \mathbf{W}^{\frac{1}{2}*}\Pi^{\perp *}_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}*} \mathbf{K}_M = \mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}}$.

Proof. The proof of each item is given in the following:

a) and b): Since $\hat{\mathbf{R}}$ is Hermitian [36], it readily follows that $\hat{\mathbf{r}}^* = \mathbf{K}_M \hat{\mathbf{r}}$. Hence, recalling (13) and making use of Lemma 7 and the fact that $\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H$ is Hermitian, we obtain

$$\hat{\mathbf{Q}}^* \hat{\mathbf{r}}^* = \mathbf{K}_M\left(\hat{\mathbf{r}} - \frac{\text{vec}(\mathbf{I}_M)\text{vec}^H(\hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H)\mathbf{K}_v \mathbf{T} \mathbf{J}^\dagger \mathbf{K}_M \hat{\mathbf{r}}}{v-K}\right) = \mathbf{K}_M \hat{\mathbf{Q}} \hat{\mathbf{r}}. \quad (118)$$

Further, transposing (118) results in item b).

c): Let $\mathbf{a} \in \mathbf{C}^{K \times 1}$ be an arbitrary vector. From (4), we find

$$\mathbf{K}_M \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})\mathbf{a} = \text{vec}\left(\mathbf{A}^*(\boldsymbol{\theta})\text{diag}(\mathbf{a})\mathbf{A}^T(\boldsymbol{\theta})\right) = (\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^* \mathbf{a}. \quad (119)$$

Since $\mathbf{a}$ is an arbitrary vector, it can be concluded that

$$\mathbf{K}_M \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}) = (\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^*. \quad (120)$$

Eventually, replacing both side of (120) with their conjugate transposes and taking derivative with respect to θ_k gives c).

d): Using (120) and $\mathbf{K}_M \mathbf{W} = \mathbf{W}^* \mathbf{K}_M$, it is observed that

$$\mathbf{K}_M \mathbf{W}^{\frac{1}{2}}(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger H} = \quad (121)$$
$$\mathbf{W}^*(\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^*\left((\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^H \mathbf{W} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})\right)^{-1} = \mathbf{W}^{\frac{1}{2}*}(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger T}.$$

Multiplying both sides of (121) by $\mathbf{K}_M$ leads to d).

e) and f): Making use of (120), transpose of (121) and the fact that $\mathbf{W} = \mathbf{K}_M \mathbf{W}^* \mathbf{K}_M$, we get

$$\mathbf{W}^{\frac{1}{2}} \Pi^{\perp}_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}} = \quad (122)$$
$$\mathbf{K}_M \mathbf{W}^* \mathbf{K}_M - \mathbf{K}_M \mathbf{W}^*(\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^*(\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}))^{\dagger *} \mathbf{W}^{\frac{1}{2}*} \mathbf{K}_M$$
$$= \mathbf{K}_M \mathbf{W}^{\frac{1}{2}*} \Pi^{\perp *}_{\mathbf{W}^{\frac{1}{2}} \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} \mathbf{W}^{\frac{1}{2}*} \mathbf{K}_M. \quad \square$$

Lemma 9. $\mathbf{K}_M \mathbf{W}_{\text{opt}} = \mathbf{W}_{\text{opt}}^* \mathbf{K}_M$

Proof. Since $\mathbf{K}_M$ is involutory and symmetric, to prove the lemma, we can equivalently show that $\mathbf{K}_M \mathbf{W}_{\text{opt}}^* \mathbf{K}_M = \mathbf{W}_{\text{opt}}$. From (120), it is readily observed that

$$\mathbf{K}_M \Pi^{\perp *}_{\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} = \Pi^{\perp}_{\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})}. \quad (123)$$

Now, it follows from (120) and (123) that

$$\mathbf{K}_M \mathbf{W}_{\text{opt}}^* \mathbf{K}_M = \quad (124)$$
$$\left(\Pi^{\perp}_{\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} \mathbf{K}_M \mathbf{S}^* \mathbf{K}_M \Pi^{\perp}_{\mathbf{J} \mathbf{A}_d(\boldsymbol{\theta})} + \mathbf{J} \mathbf{A}_d(\boldsymbol{\theta}) \mathbf{A}_d^H(\boldsymbol{\theta}) \mathbf{J}^T\right)^{-1}.$$

In consequence, $\mathbf{K}_M \mathbf{W}_{\text{opt}}^* \mathbf{K}_M = \mathbf{W}_{\text{opt}}$ iff $\mathbf{K}_M \mathbf{S}^* \mathbf{K}_M = \mathbf{S}$. Exploiting Lemma 7 and recalling (30), we find

$$\mathbf{K}_M \mathbf{b}^* = \mathbf{K}_M \mathbf{J}^{\dagger H} \mathbf{T}^H \mathbf{K}_v \text{vec}\left(\mathbf{U}_n \mathbf{U}_n^H\right) = \mathbf{b}. \quad (125)$$

Further, recalling (29) and using transpose of (125), we have

$$\mathbf{K}_M \mathbf{Q}^* = \mathbf{K}_M - \frac{\mathbf{K}_M \text{vec}(\mathbf{I}_M)\mathbf{b}^T}{v-K} = \mathbf{Q} \mathbf{K}_M \quad (126)$$

Now recalling (36) and making use of (125) and (126) gives

$$\mathbf{K}_M \mathbf{S}^* \mathbf{K}_M = \mathbf{K}_M \mathbf{Q}^* \mathbf{M}^{2*} \mathbf{Q}^T \mathbf{K}_M + \mathbf{K}_M \mathbf{b}^* \mathbf{b}^T \mathbf{K}_M \quad (127)$$
$$= \mathbf{Q} \mathbf{K}_M \mathbf{M}^{2*} \mathbf{K}_M \mathbf{Q}^H + \mathbf{b} \mathbf{b}^H = \mathbf{S},$$

where the last equality is obtained by using the fact that $\mathbf{K}_M \mathbf{M}^{2*} \mathbf{K}_M = \mathbf{K}_M (\mathbf{R} \otimes \mathbf{R}^T) \mathbf{K}_M = \mathbf{R}^T \otimes \mathbf{R} = \mathbf{M}^2$. $\quad \square$

REFERENCES

[1] S. Sedighi, R. B. S. Mysore, and B. Ottersten, "A statistically efficient estimator for co-array based DoA estimation," in 52nd *Annual Asilomar Conference on Signals, Systems, and Computers*, October 2018. [Online]. Available: http://hdl.handle.net/10993/37660

[2] H. Van Trees, *Optimum Array Processing (Detection, Estimation, and Modulation Theory, Part IV)*. New York: John Wiley and Sons Inc., 2002.

[3] S. S. Haykin, J. Litva, , and T. J. Shepherd, Eds., *Radar Array Processing*. Berlin, Germany: Springer-Verlag, 1993.

[4] B. Ottersten, "Array processing for wireless communications," in *Proceedings of 8th Workshop on Statistical Signal and Array Processing*, Jun 1996, pp. 466–473.

[5] A. Paulraj, B. Ottersten, R. Roy, A. Swindlehurst, G. Xu, and T. Kailath, "16 subspace methods for directions-of-arrival estimation," *Handbook of Statistics*, vol. 10, pp. 693–739, 1993.

[6] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for daa estimation algorithms: unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst*, vol. 29, no. 4, pp. 1170–1184, Oct 1993.

[7] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 38, no. 10, pp. 1783–1795, Oct 1990.

[8] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Transactions on Signal Processing*, vol. 39, no. 11, pp. 2436–2449, Nov 1991.

[9] A. Moffet, "Minimum-redundancy linear arrays," *IEEE Transactions on Antennas and Propagation*, vol. 16, no. 2, pp. 172–175, Mar 1968.

[10] 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.

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

[12] C. L. Liu and P. Vaidyanathan, "Cramér-Rao bounds for coprime and other sparse arrays, which find more sources than sensors," *Digital Signal Processing*, vol. 61, pp. 43 – 61, 2017.

[13] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DoA estimation using co-prime arrays," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013, pp. 3967–3971.

[14] Q. Shen, W. Liu, W. Cui, and S. Wu, "Underdetermined DoA estimation under the compressive sensing framework: A review," *IEEE Access*, vol. 4, pp. 8865–8878, 2016.

[15] P. Pal and P. P. Vaidyanathan, "Pushing the limits of sparse support recovery using correlation information," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 711–726, Feb 2015.

[16] ——, "Correlation-aware techniques for sparse support recovery," in *2012 IEEE Statistical Signal Processing Workshop (SSP)*, Aug 2012, pp. 53–56.

[17] Y. Chi, A. Pezeshki, L. Scharf, and R. Calderbank, "Sensitivity to basis mismatch in compressed sensing," in *2010 IEEE International Conference on Acoustics, Speech and Signal Processing*, March 2010, pp. 3930–3933.

[18] Z. Tan and A. Nehorai, "Sparse direction of arrival estimation using co-prime arrays with off-grid targets," *IEEE Signal Processing Letters*, vol. 21, no. 1, pp. 26–29, Jan 2014.

[19] Z. Yang, L. Xie, and C. Zhang, "A discretization-free sparse and parametric approach for linear array signal processing," *IEEE Transactions on Signal Processing*, vol. 62, no. 19, pp. 4959–4973, Oct 2014.

[20] M. Wang and A. Nehorai, "Coarrays, MUSIC, and the Cramér-Rao bound," *IEEE Trans. Signal Process.*, vol. 65, no. 4, pp. 933–946, Feb 2017.

[21] J. Steinwandt, F. Roemer, and M. Haardt, "Performance analysis of ESPRIT-type algorithms for co-array structures," in *2017 IEEE 7th International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP)*, Dec 2017, pp. 1–5.

[22] C. L. Liu and P. P. Vaidyanathan, "Remarks on the spatial smoothing step in coarray MUSIC," *IEEE Signal Processing Letters*, vol. 22, no. 9, pp. 1438–1442, Sept 2015.

[23] C. Zhou, Y. Gu, X. Fan, Z. Shi, G. Mao, and Y. D. Zhang, "Direction-of-arrival estimation for coprime array via virtual array interpolation," *IEEE Transactions on Signal Processing*, vol. 66, no. 22, pp. 5956–5971, Nov 2018.

[24] C. Zhou, Y. Gu, Z. Shi, and Y. D. Zhang, "Off-grid direction-of-arrival estimation using coprime array interpolation," *IEEE Signal Processing Letters*, vol. 25, no. 11, pp. 1710–1714, Nov 2018.

[25] H. Qiao and P. Pal, "Gridless line spectrum estimation and low-rank toeplitz matrix compression using structured samplers: A regularization-free approach," *IEEE Transactions on Signal Processing*, vol. 65, no. 9, pp. 2221–2236, May 2017.

[26] ——, "Unified analysis of co-array interpolation for direction-of-arrival estimation," in *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, March 2017, pp. 3056–3060.

[27] S. Sedighi, R. B. S. Mysore, S. Maleki, and B. Ottersten, "Consistent least squares estimator for co-array-based DoA estimation," in *2018 IEEE 10th Sensor Array and Multichannel Signal Processing Workshop (SAM)*, July 2018, pp. 524–528.

[28] Hongbin Li, P. Stoica, and Jian Li, "Computationally efficient maximum likelihood estimation of structured covariance matrices," *IEEE Transactions on Signal Processing*, vol. 47, no. 5, pp. 1314–1323, May 1999.

[29] J. Nie, "An exact jacobian SDP relaxation for polynomial optimization," *Mathematical Programming*, vol. 137, no. 1-2, pp. 225–255, 2013.

[30] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.

[31] P. Stoica and K. C. Sharman, "Novel eigenanalysis method for direction estimation," *IEE Proceedings F - Radar and Signal Processing*, vol. 137, no. 1, pp. 19–26, Feb 1990.

[32] M. Jansson, B. Goransson, and B. Ottersten, "A subspace method for direction of arrival estimation of uncorrelated emitter signals," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 945–956, April 1999.

[33] H. Yanai, K. Takeuchi, and Y. Takane, *Projection Matrices, Generalized Inverse Matrices, and Singular Value decomposition*. USA: New York: Springer-Verlag, 2014.

[34] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1122–1135, May 1991.

[35] M. Kaveh and A. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 34, no. 2, pp. 331–341, April 1986.

[36] M. Bilodeau and D. Brenner, *Theory of multivariate statistics*. New York: Springer-Verlag, 1999.

[37] J. R. Magnus and H. Neudecker, *Matrix differential calculus with applications in statistics and econometrics*. USA: New York: John Wiley and Sons, Inc., 1988.

[38] P. Stoica and A. Nehorai, "Music, maximum likelihood, and cramer-rao bound," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 5, pp. 720–741, May 1989.

[39] S. M. Kay, *Fundamentals of statistical signal processing, Volume I: Estimation theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.

[40] C. D. Meyer, *Matrix analysis and applied linear algebra*. Philadelphia: Society for Industrial and Applied Mathematics, 2000.

[41] A. J. Weiss and B. Friedlander, "Performance analysis of spatial smoothing with interpolated arrays," *IEEE Transactions on Signal Processing*, vol. 41, no. 5, pp. 1881–1892, May 1993.

Saeid Sedighi (S'17) was born in Tehran, Iran, in 1987. He received the M.Sc. degree in Electrical Engineering from Imam Khomeini International University (IKIU), Qazvin, Iran, in 2012. From October 2012 to January 2015, he was a research assistant in the Signal Processing and Information Theory (SPIT) lab at the Electrical Engineering Department, Imam Khomeini International University (IKIU). He is currently working toward the Ph.D. degree in the Interdisciplinary Centre for Security, Reliability and Trust (SnT), University of Luxembourg, Luxembourg under the supervision of Prof. Björn Ottersten. His research interests are statistical signal processing, array processing, radar, sparse sensing, cognitive radio, detection and estimation theory.

M. R. Bhavani Shankar received Masters and Ph. D in Electrical Communication Engineering from Indian Institute of Science, Bangalore in 2000 and 2007 respectively. He was a Post Doc at the ACCESS Linnaeus Centre, Signal Processing Lab, Royal Institute of Technology (KTH), Sweden from 2007 to September 2009. He joined SnT in October 2009 as a Research Associate and is currently a Research Scientist at SnT. He was with Beceem Communications, Bangalore from 2006 to 2007 as a Staff Design Engineer working on Physical Layer algorithms for WiMAX compliant chipsets. He was a visiting student at the Communication Theory Group, ETH Zurich, headed by Prof. Helmut Blcskei during 2004. Prior to joining Ph. D, he worked on Audio Coding algorithms in Sasken Communications, Bangalore as a Design Engineer from 2000 to 2001. His research interests include Design and Optimization of MIMO Communication Systems, Automotive Radar and Array Processing, polynomial signal processing, Satellite communication systems, Resource Allocation, Game Theory and Fast Algorithms for Structured Matrices. He is currently on the Executive Committee of the IEEE Benelux joint chapter on communications and vehicular technology, member of the EURASIP Technical Area Committee on Theoretical and Methodological Trends in Signal Processing and serves as handling editor for Elsevier Signal Processing. He was a co-recipient of the 2014 Distinguished Contributions to Satellite Communications.

Björn Ottersten (S'87–M'89–SM'99–F'04) was born in Stockholm, Sweden, in 1961. He received the M.S. degree in electrical engineering and applied physics from Linkping University, Linkping, Sweden, in 1986, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, USA, in 1990. He has held research positions with the Department of Electrical Engineering, Linkping University, the Information Systems Laboratory, Stanford University, the Katholieke Universiteit Leuven, Leuven, Belgium, and the University of Luxembourg, Luxembourg. From 1996 to 1997, he was the Director of Research with ArrayComm, Inc., a start-up in San Jose, CA, USA, based on his patented technology. In 1991, he was appointed a Professor of signal processing with the Royal Institute of Technology (KTH), Stockholm, Sweden. From 1992 to 2004, he was the Head of the Department for Signals, Sensors, and Systems, KTH, and from 2004 to 2008, he was the Dean of the School of Electrical Engineering, KTH. He is currently the Director for the Interdisciplinary Centre for Security, Reliability and Trust, University of Luxembourg. He was a recipient of the IEEE Signal Processing Society Technical Achievement Award in 2011 and the European Research Council advanced research grant twice, in 20092013 and in 20172022. He has co-authored journal papers that received the IEEE Signal Processing Society Best Paper Award in 1993, 2001, 2006, and 2013, and seven IEEE conference papers best paper awards. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the Editorial Board of the IEEE Signal Processing Magazine. He is currently a member of the editorial boards of EURASIP Signal Processing Journal, EURASIP Journal of Advances Signal Processing and Foundations and Trends of Signal Processing. He is a fellow of EURASIP.