IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received November 12, 2018, accepted December 20, 2018, date of publication January 7, 2019, date of current version January 23, 2019.

Digital Object Identifier 10.1109/ACCESS.2018.2889958

Unified Subspace Fitting Framework and Its Performance Analysis for Direct Position Determination in the Presence of Multipath Propagation

JIANPING DU[1,2], HONGYI YU[1,2], DING WANG[1,2], DALONG ZHANG[3], AND GUANGYI LIU[1,2]

[1]National Digital Switching System Engineering and Technological Research Center, Zhengzhou 450002, China
[2]Zhengzhou Institute of Information Science and Technology, Zhengzhou 450002, China
[3]School of Information Engineering, Zhengzhou University, Zhengzhou 450000, China

Corresponding author: Ding Wang (wang_ding814@aliyun.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61201381, Grant 61401513, and Grant 61501517, and in part by the China Postdoctoral Science Foundation under Grant 2016M592989.

ABSTRACT The cost functions and their performances of direct position determination (DPD) methods in the presence of multipath propagation are investigated. We first establish a general DPD (GDPD) model in the presence of multi-path propagation and point out that the existing cost functions cannot get the emitter positions correctly because of the singularity of the manifold matrix in a multipath propagation scenario. Eight cost functions are developed for the GDPD model and formulated in a unified subspace fitting (USF)-based framework, which provides insight into their algebraic and asymptotic relations. Moreover, we derive the closed-form expressions of the asymptotic distributions of the estimation errors, which are optimized by those cost functions. Besides, the optimal cost function for achieving an optimal asymptotic performance is derived based on the optimization theory. Finally, the numerical simulations and Cramér–Rao lower bound are provided to verify the analytical results and show that: 1) the cost functions which work well in the single-path DPD model cannot find the emitters correctly in a multipath scenario; 2) the signal subspace fitting cost functions and noise subspace fitting cost functions, which are proposed in this paper, find the emitters accurately in the multipath propagation scenarios; 3) the optimal-weighted-subspace-fitting cost function holds the best asymptotic performance under the USF framework; and 4) the asymptotic performance of a multiple dimension cost function is better than a 1D cost function.

INDEX TERMS Asymptotic distribution, direct position determination, multipath propagation, subspace fitting, maximum likelihood, multiple signal classification, weighted subspace fitting.

I. INTRODUCTION

Most of the position systems are designed for the LoS (Line-of-Sight) emitters, such as GPS (Global Position System), LoRa (Long Range) positioning system. The research about NLoS (None-Line-of-Sight) emitters positioning has attracted much attention in recent years. Aero platforms (satellite, EWA: Early Warning Air-plane, etc.) are adopted to extend the positioning area for the positioning of NLoS emitters. We are interested in the positioning of NLoS emitters assisted by aero platforms (UAV: Unmanned Aerial Vehicle, satellite, etc.) in this paper. There are two ways to get NLoS emitter positions, the first is to sample directly and synchronously on aero platforms (e.g. WSN: Wireless Sensor Network), and the second is to forward the signal to the receivers without any processing except frequency-shifting and power amplification (e.g. satellite). Compared with the first scheme, the transponders in the second scheme only need forwarding the received signal to the receivers, and does not need transmitting the sampled data to the positioning center (See FIGURE 1).

Additional communication overhead and synchronization cost are not required for transponders in the second scheme, and it is important for a transponder which is installed on a UAV or satellite.

Multipath propagation is the most difficult problem in this positioning system. The signals received by each antenna are

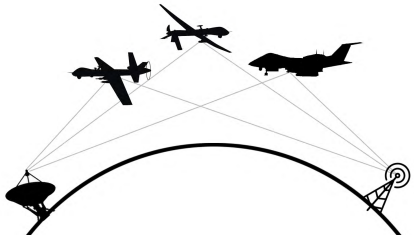


FIGURE 1. Position System assisted with transponders .

the superposition of multiple delay and frequency-shifting signals (See FIGURE 1). Most of existing positioning methods are designed by developing statistical measurements to against multi-path effects. For example, using CAF (Cross Ambiguity Function) method to estimate time-frequency difference, using Received Signal Strength Indication (RSSI) fingerprinting to match energy intensity, and using UWB (Ultra Wide Bandwidth) signal to separate difference paths from received signal. The common feature of these methods is to reduce the influence of secondary path on main path, and then locate it as a single path location problem.

When CAF is adopted for estimating the time-frequency difference in the presence of multipath propagation, multiple peaks will be generated because of the multi-path propagation. The matching problem between peaks of the CAF and their relevant paths difference will be very difficult. In addition, in a low SNR scenario, the two-step positioning method may not be able to accurately estimate the time-frequency difference.

Unlike existing methods to reduce multipath interference, the Direct Position Determination (DPD) method makes use of the multipath propagation to improve the reliability and availability of the position system. DPD can theoretically achieve a better positioning performance. Du *et al.* [1] established a GDPD (General Direct Positioning Determination) mathematical model and algorithm of forwarding positioning system, and compared the positioning performance of different cost functions by Monte Carlo simulation, but theoretical analyses of cost functions are absent.

DPD methods obtain the emitter positions directly for achieving better performances [2]–[7] than two-step positioning methods. A DPD method uses the observations from all sensors together, and establishes a cost function that depend only on the emitter for achieving a global optimal estimations. DPD methods overcome the problem of associating estimated parameters with their relevant sources, and are shown to outperform two-step methods (TDOA: Time-Difference-Of-Arrival, FDOA: Frequency-Difference-Of-Arrival, DOA: Direction-Of-Arrival etc.), especially in low Signal-to-Noise Ratio (SNR) scenarios [8], [9]. The key idea of a DPD method is that the two-step optimizations (measure estimations and position optimizations) are combined into a one-step optimization (position optimizations) to obtain a theoretical optimal solution.

In DPD methods, The Maximum Likelihood (ML) and MUltiple SIgnal Classification (MUSIC) are widely used as cost functions. A DPD method with an ML cost function mainly focuses on the positioning of known waveform signals. ML establishes a likelihood function of the received signals, which determined only by emitter positions, as the cost function, and obtains the maximum likelihood estimate of the emitter positions. A DPD method with a MUSIC cost function is mainly for the unknown waveform signal positioning, and the cost function is formed by maximizing the projection of the array manifold vectors onto the signal subspace. Weiss et al. compared the positioning performances of different cost functions by numerical simulation in [10]. The results shown that when the number of snapshots was sufficient, ML and MUSIC can effectively close with Cramér-Rao Lower Bound (CRLB). A Signal Subspace Projection-MUltiple SIgnal Classification (SSP-MUSIC) cost function is developed in their works. The SSP-MUSIC maximized the array manifold projection onto the signal subspace for achieving the optimal estimations of emitter position. In a direction finding application, since the path attenuation of each antenna has been calibrated in advance, the maximization of the signal subspace projection is equivalent with the minimization of the Noise Subspace Projection-MUSIC (NSP-MUSIC) of the array manifold space. A GDPD model which took the multipath propagation into account was established in [1]. Amar and Weiss pointed out that the cost function, proposed in [10], which works well for single-path propagation DPD method, would be failed in finding the emitter positions. In a multipath propagation scenario, the array manifold matrix would be singular or nearly singular when there are two or more paths with the same propagation delays from the candidates to receiving stations. In this case, the difference between an SSP-MUSIC cost function and a NSP-MUSIC cost function turns to be significant, and fake solutions will be find for those candidates. Although existing literatures have found that performances of different cost functions are different, there are few theoretical analyses on the performance of different cost functions. The performances of those cost functions were verified by Monte Carlo simulations rather than a theoretical analysis in the existing literatures. A cost function is regarded as a reasonable cost function, if the error covariance matrix closes to the CRLB with the increases of snapshots number.

There have been many results about the asymptotic performance for the single parameter estimation problem in the direction finding application. The main research results of the existing literature are concentrated in three areas: (1) the asymptotic distribution analysis of DOA estimation errors for specific signals, (2) the asymptotic distribution analysis of DOA estimation errors with model errors, and (3) the asymptotic distribution analysis of DOA estimation errors under Subspace Fitting framework.

Most of existing studies focused on the asymptotic distribution analysis of DOA estimation error for specific signals. Hamza and Buckley [11] analyzed the effects of a limited number of snapshots on a general class of multiple dimension

signal subspace estimation methods. Kaveh and Barabell [12] analyzed the statistical performance of the music and the minimum-norm algorithms in resolving plane waves in noise. Li *et al.* [13] and Dauxois *et al.* [14] studied the asymptotic performance by Principal Component Analysis (PCA) method for non-circular signals in the presence of circular white Gaussian noise. Wang and Kaveh [15], [16] presented an analytical evaluation of detection and estimation performances of narrow-band signal and coherent wide-band system subspace processing for multiple source direction finding. An asymptotic analysis was presented of a class of high-resolution estimators for resolving correlated and coherent plane waves in noise for direction finding application in [17]. Stoica and Nehorai [18] proposed a numerical and analytical study of conditional and unconditional DOA estimation. An analytical performance evaluation of the errors of the direction-of-arrival estimates obtained by the MUSIC algorithm for uncorrelated sources was studied by Porat and Friedlander [19], [20]. Delmas and Meurisse [21] studied the asymptotic performance of direction finding algorithms with temporally correlated narrowband signal. Zhou *et al.* [22] analyzed the asymptotic performance of DOA estimation based on the G-MUSIC algorithm for a single source using RLA (Random-Linear-Array).

The asymptotic distribution of DOA estimation error in the presence of model errors has also attracted the interest of some scholars. Swindlehurst and Kailath [23], [24] studied the performance of subspace algorithms for the situations on which the noise covariance and array response are perturbed from their assumed values. Ferreol *et al.* [25] proposed an asymptotic performance analysis of subspace DOA estimation in the presence of modeling errors.

Some literatures focused on the asymptotic distribution of DOA estimation error under the Subspace Fitting framework. Ottersten *et al.* [26] proposed an asymptotic analysis of the ML and Weighted Subspace Fitting (WSF) methods for the deterministic emitter signals. Five methods of DOA estimation which were derived from ML principle and analytic results on their theoretical performance levels were proposed by Stotica and Sharman [27]. Moulines and Cardoso [28] and Cardoso and Moulines [29] derived and worked out closed-form expressions of the asymptotic covariance of MUSIC-like DOA estimates based on two fourth-order cumulate matrices. Bengtsson and Ottersten introduced a general class of Subspace Fitting (SF) algorithms for consistent estimation of parameters from a possibly full-rank data model. The asymptotic performances of the algorithms are analyzed, and an optimally weighted algorithm is derived in [30]. Ottersten *et al.* [31], [32] proposed several estimation methods as solutions to different versions of a basic subspace fitting problem, and the asymptotic performance of the multidimensional subspace fitting methods was investigated. They studied the asymptotic robustness of sensor array processing methods further, and pointed out that the asymptotic properties of essentially all DOA estimation methods based on a multidimensional search, depend only on the second order properties of the emitter signals in [33]. Stoical and Nehorai [34], [35] analyzed the asymptotic distribution of the MUSIC algorithm for the single parameter estimation problem in the direction finding application. The minimization of NSP-MUSIC was adopted as the cost function in their works. The asymptotic distribution neglected the terms with $O(1/N)$, and it was only a upper bound of the asymptotic distribution.

Viberg *et al.* [36], [37] introduced a Weighted Subspace Fitting (WSF) method. Viberg analyzed the asymptotic distribution of the ML, Multiple Dimension-MUSIC (MD-MUSIC), and WSF for the direction finding problem, and obtained the optimal weights of a WSF for achieving the best asymptotic distribution performance. Authors derived a close form of the asymptotic performance of a signal subspace fitting MUSIC cost function. The asymptotic distribution performances of an MD-MUSIC and an ML were compared by numerical simulations. Numerical simulation results showed that the asymptotic distribution performance of an MD-MUSIC method was worse than an ML method. It was unexpected since the One Dimension-MUSIC (1D-MUSIC) was known to have the same asymptotic performance as the ML for uncorrelated sources [34]. An MD-MUSIC method not only takes the correlation between sources signals into account, but also extends the searching dimensions, and it theoretically obtains a better performance than a 1D-MUSIC method. However, reasons of the unexpected result was not analyzed further in their works.

Viberg and Ottersten [36] maximized the projection onto the signal subspace to obtain the emitter position. It is well known that the SSP-MUSIC is a simplified version of the NSP-MUSIC. Obviously, an SSP-MUSIC holds a worse performance than an NSP-MUSIC. It is regrettable that Stoica did not give the asymptotic distribution result of an NSP-MUSIC with $O(1/N)$, and did not compare the asymptotic distribution performance between an NSP-MUSIC and an ML theoretically. It is necessary to analyses the second moment statistical properties of eigen-vectors for getting an $O(1/N)$ asymptotic distribution performance. If an eigenvalue is different with other eigenvalues (eigenvalues related to the signal subspace), the second moment statistical properties of the eigen-vector correspond to the eigenvalue have been well studied in [38] and [39], and the results are adopted for getting the asymptotic distribution properties of an SSP-MUSIC in [36]. However, the eigenvalues correspond to the noise subspace are multiple roots, and there are few results on the second moment statistical properties of eigen-vectors corresponding to the multiple roots. Arie et al. pointed out that the classical derivation of the asymptotic efficiency uses a first-order perturbation analysis, relying on a ''small-errors'' assumption, which under sub-asymptotic conditions turns inaccurate, rendering the ML generally biased and inefficient. A tensor formulation of higher-order derivatives to derive a tractable formulation of a higher (up to the third) order perturbation analysis was proposed for predicting the bias and Mean Squared Error (MSE) matrix of the ML of parameter

vectors in general non-linear models under sub-asymptotic conditions in [40].

The above literatures focused on the asymptotic performance for the direction finding application. It was assumed that the path attenuations were well calibrated or known precisely in advance. However, the path attenuations are unknown in a GDPD application, and some simplification methods in the existing literatures can not work well in a GDPD model.

The motivation of this paper is to develop the cost functions for the DPD method in multipath propagation scenarios, and to analysis the asymptotic distribution of those cost functions. A GDPD positioning model in the presence of multi-path propagation is established as the base positioning model. We develop eight cost functions under the Unified Subspace Fitting (USF) framework, and derive the closed-form expressions of asymptotic distributions of those cost functions. MUSIC and ML cost functions are unified into the Weighted Signal Subspace Fitting (WSSF) framework, and the optimal weights of a WSSF cost function for achieving the optimal asymptotic performance is derived base on algebra and optimization theories. Finally, the asymptotic distribution performances of different cost functions and the CRLB of the position estimations are compared by numerical simulations.

This paper is organized as: First, a GDPD model in presence of multi-path propagation is established in the section 2; Then, A geometric understanding of USF and eight cost functions are proposed in section 3 and section 4; Next, the asymptotic distribution properties of Multiple Dimension functions and CRLB are studied in section 5; Afterwards, The optimal weights are optimized for achieving the best asymptotic distribution performance in the section 6; Finally, a numerical simulation is taken to verify above results in the section 7 and a conclusion is proposed in the section 8.

II. GDPD MODEL IN THE PRESENCE OF MULTI-PATH PROPAGATION

The GDPD model in presence of multi-path propagation follows the model in [1], [41], and [42]. For the convenience of reading, we give the relevant variables descriptions and formulas again.

Consider D emitters are located at $\mathbf{p}_{\mathrm{e}}$, and L passive transponders are placed at $\mathbf{p}_{\mathrm{t}}$, where

$$\mathbf{p}_{\mathrm{e}} = [\mathbf{p}_{\mathrm{e}}^T(1), \mathbf{p}_{\mathrm{e}}^T(2), \ldots, \mathbf{p}_{\mathrm{e}}^T(D)]^T, \tag{1}$$

$$\mathbf{p}_{\mathrm{t}} = [\mathbf{p}_{\mathrm{t}}^T(1), \mathbf{p}_{\mathrm{t}}^T(2), \ldots, \mathbf{p}_{\mathrm{t}}^T(L)]^T, \tag{2}$$

and $\mathbf{p}_{\mathrm{e}}(d) \triangleq [p_{\mathrm{e}}(d,x), p_{\mathrm{e}}(d,y), p_{\mathrm{e}}(d,z)]^T$ and $\mathbf{p}_{\mathrm{t}}(\ell) \triangleq [p_{\mathrm{t}}(\ell,x), p_{\mathrm{t}}(\ell,y), p_{\mathrm{t}}(\ell,z)]^T$ are the 3D coordinates of emitter positions and transponder positions. The signals transmitted by the emitters are reflected by the transponders and intercepted by N receiving arrays. The raw signal samples are transferred "in-band" (i.e., as multipath), and required tightly synchronized in time and frequency [41]. Each array includes M antennas, and the centers of arrays are located at $\mathbf{p}_{\mathrm{r}} = [\mathbf{p}_{\mathrm{r}}^T(1), \mathbf{p}_{\mathrm{r}}^T(2), \ldots, \mathbf{p}_{\mathrm{r}}^T(N)]^T$, and $\mathbf{p}_{\mathrm{r}}(N) \triangleq [p_{\mathrm{r}}(n,x), p_{\mathrm{r}}(n,y), p_{\mathrm{r}}(n,z)]^T$. It is assumed that the locations of transponders and receivers are known a priori(e.g. transponders are installed on UAVs whose positions are known a priori), but the signal waveforms are unknown. The scenario is depicted in figure 2 [1].

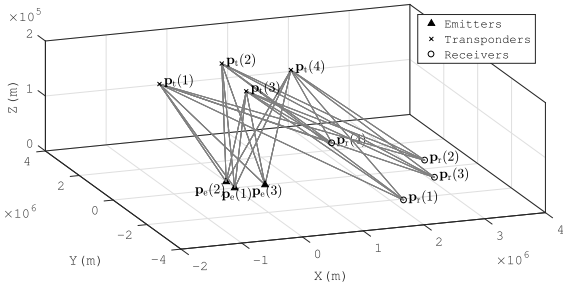


FIGURE 2. Multiple-path positioning problem with static transponders/receivers.

Denote the signal propagation delay between the dth emitter and the ℓth transponder by $\bar{\tau}_{d,\ell} = \frac{1}{\mathrm{c}} \|\mathbf{p}_{\mathrm{e}}(d) - \mathbf{p}_{\mathrm{t}}(\ell)\|_F$. Let $\tilde{\tau}_{\ell,n,m} = \frac{1}{\mathrm{c}} \|\mathbf{p}_{\mathrm{r}}(n,m) - \mathbf{p}_{\mathrm{t}}(\ell)\|_F$ denotes the propagation delay between the ℓth transponder and the mth antenna in the nth receiver, where $\mathrm{c} = 3 \times 10^8$ m/s, and $\mathbf{p}_{\mathrm{r}}(n,m)$ is the position of the mth antenna in the nth array. Denote

$$\tilde{\boldsymbol{\tau}}_{\ell,n} = [\tilde{\tau}_{\ell,n,1}, \tilde{\tau}_{\ell,n,1}, \ldots, \tilde{\tau}_{\ell,n,M}]^T, \tag{3}$$

where $\tilde{\boldsymbol{\tau}}_{\ell,n}$ is an $M \times 1$ vectors which represents the Delay Of Arrical (DOA) from the transponder ℓ to the receiving array n.

The path attenuation from the dth emitter to the nth receiver array which reflected by the ℓth transponder is denoted by $\alpha_{d,\ell,n}$. We assume that the receiving arrays have been corrected for all antennas, and each antenna in an array shares the same path attenuation.

We assume that the path attenuations, $\alpha_{d,\ell,n}$, remains constant during the observation time interval and the noise level σ is known. The frequency-domain model for the kth Discrete Fourier Transform (DFT) coefficient of the nth receiving array is given by:

$$\mathbf{r}_n(k) = \sum_{\ell=1}^{L}\sum_{d=1}^{D} \alpha_{d,\ell,n}\tilde{\mathbf{a}}_{\ell,n}(k)e^{-i\omega_k\bar{\tau}_{\ell,d}}s_d(k) + \mathbf{n}(k), \tag{4}$$

$$\tilde{\mathbf{a}}_{\ell,n}(k) = e^{-i\omega_k\tilde{\boldsymbol{\tau}}_{\ell,n}}, \tag{5}$$

$$s_d(k) = \tilde{s}_d(k)e^{-i\omega_k t_d} \tag{6}$$

$$\omega_k = \frac{2\pi k}{T}, 0 \le k \le K-1, \tag{7}$$

where t_d is the unknown transmit time of the emitter d, $\tilde{s}_d(k)$ is the kth Fourier coefficient of the dth source signal $\check{s}_d(t)$, $t \in [0, T]$. $\mathbf{r}_n(k)$ and $\mathbf{n}(k)$ are $M \times 1$ vectors of the kth Fourier coefficients of $\bar{\mathbf{r}}_n(t)$ and $\bar{\mathbf{n}}(t)$, $t \in [0, T]$. $\tilde{\mathbf{a}}_{\ell,n}(k)$ is an $M \times 1$ vector, and it denotes the generalized array response of the nth receiver at frequency ω_k.

Denote the unknown parameter vector $\boldsymbol{\theta} = [\mathbf{p}_{\mathrm{e}}^T, \bar{\boldsymbol{\alpha}}^T]^T$, where $\mathbf{p}_{\mathrm{e}}$ are emitter positions vector, and $\bar{\boldsymbol{\alpha}}$ are path

attenuations vector. Make (4) into matrix form

$$\mathbf{r}(k) = \mathbf{A}(k, \boldsymbol{\theta})\mathbf{s}(k) + \mathbf{n}(k) \tag{8}$$

where

$$\mathbf{r}(k) \triangleq [\mathbf{r}_1^T(k), \mathbf{r}_2^T(k), \ldots, \mathbf{r}_N^T(k)]^T, \tag{9}$$

$$\mathbf{r}_n^T(k) = [\mathbf{r}_{n1}^T(k), \mathbf{r}_{n2}^T(k), \ldots, \mathbf{r}_{nM}^T(k)]^T, \tag{10}$$

$$\mathbf{A}(k, \boldsymbol{\theta}) \triangleq \boldsymbol{\Gamma}(\mathbf{p}_e, k)\boldsymbol{\alpha}, \tag{11}$$

$$\boldsymbol{\Gamma}(\mathbf{p}_e, k) \triangleq \tilde{\mathbf{A}}(\tilde{\boldsymbol{\tau}}, k)\mathbf{V}(\bar{\boldsymbol{\tau}}, k), \tag{12}$$

$$\tilde{\mathbf{A}}(\tilde{\boldsymbol{\tau}}, k) = \begin{bmatrix} \tilde{\mathbf{A}}_1(\tilde{\boldsymbol{\tau}}, k) & 0 & \ldots & 0 \\ 0 & \tilde{\mathbf{A}}_2(\tilde{\boldsymbol{\tau}}, k) & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \tilde{\mathbf{A}}_N(\tilde{\boldsymbol{\tau}}, k) \end{bmatrix}, \tag{13}$$

$$\tilde{\mathbf{A}}_n(\tilde{\boldsymbol{\tau}}, k) = [\tilde{\mathbf{a}}_{1,n}(k), \tilde{\mathbf{a}}_{2,n}(k), \ldots, \tilde{\mathbf{a}}_{L,n}(k)], \tag{14}$$

where $\tilde{\mathbf{a}}_{\ell,n}(k) \triangleq \tilde{\mathbf{a}}_{\ell,n}(\tilde{\boldsymbol{\tau}}_{\ell,n}, k)$.

$$\mathbf{V}(\bar{\boldsymbol{\tau}}, k) = \mathbf{I}_N \otimes \bar{\mathbf{V}}(\bar{\boldsymbol{\tau}}, k), \tag{15}$$

$$\bar{\mathbf{V}}(\bar{\boldsymbol{\tau}}, k) = [\bar{\mathbf{V}}_1(\bar{\boldsymbol{\tau}}, k), \bar{\mathbf{V}}_2(\bar{\boldsymbol{\tau}}, k), \ldots, \bar{\mathbf{V}}_D(\bar{\boldsymbol{\tau}}, k)], \tag{16}$$

$$\bar{\mathbf{V}}_d(\bar{\boldsymbol{\tau}}, k) = \mathrm{diag}([e^{-i\omega_k \bar{\tau}_{d,1}}, e^{-i\omega_k \bar{\tau}_{d,2}}, \ldots, e^{-i\omega_k \bar{\tau}_{d,L}}]), \tag{17}$$

$$\boldsymbol{\alpha} = \begin{bmatrix} \boldsymbol{\alpha}_1 \\ \boldsymbol{\alpha}_2 \\ \vdots \\ \boldsymbol{\alpha}_N \end{bmatrix}, \tag{18}$$

$$\boldsymbol{\alpha}_n = \begin{bmatrix} \boldsymbol{\alpha}_{1n} & 0 & \ldots & 0 \\ 0 & \boldsymbol{\alpha}_{2n} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \boldsymbol{\alpha}_{Dn} \end{bmatrix}, \tag{19}$$

$$\boldsymbol{\alpha}_{dn} = [\alpha_{d1n}, \alpha_{d2n}, \ldots, \alpha_{dLn}]^T, \tag{20}$$

$$\mathbf{s}(k) \triangleq [s_1(k), s_2(k), \ldots, s_D(k)]^T, \tag{21}$$

where $\otimes$ is the kronecker product, and $\mathbf{I}_N$ is an identify matrix with size of $N \times N$.

The covariance matrix of received data is

$$\begin{aligned} \mathbf{R}(k) &= E\{\mathbf{r}(k)\mathbf{r}^H(k)\} \\ &= \mathbf{A}(k, \boldsymbol{\theta})\mathbf{P}(k)\mathbf{A}^H(k) + \sigma^2(k)\mathbf{I}, \end{aligned} \tag{22}$$

where $\mathbf{P}(k) \triangleq \mathbf{s}(k)\mathbf{s}^H(k)$, and $\mathbf{R}(k)$ is estimated by

$$\hat{\mathbf{R}}(k) = \frac{1}{J}\sum_{j=1}^{J} \mathbf{r}_j(k)\mathbf{r}_j^H(k). \tag{23}$$

Denote the eigenvectors of $\hat{\mathbf{R}}(k)$ by $\hat{\mathbf{U}}(k)$.

$$\hat{\mathbf{U}}(k) \triangleq [\hat{\mathbf{U}}_s, \hat{\mathbf{U}}_n], \tag{24}$$

where $\hat{\mathbf{U}}_s(k) = [\hat{\mathbf{u}}_s(1), \hat{\mathbf{u}}_s(2), \ldots, \hat{\mathbf{u}}_s(D)]$ is consisted by eigenvectors correspond to the maximal Dth eigenvalues $\hat{\Lambda}_s = [\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_D]^T$, and $\hat{\mathbf{U}}_n(k) = [\hat{\mathbf{u}}_n(D+1), \hat{\mathbf{u}}_n(D+2), \ldots, \hat{\mathbf{u}}_n(NM)]$ is consisted by the other $NM - D$ eigenvectors $\hat{\Lambda}_n = [\hat{\lambda}_{D+1}, \hat{\lambda}_{D+2}, \ldots, \hat{\lambda}_{NM}]^T$.

III. A GEOMETRIC UNDERSTANDING OF USF

The Subspace Fitting (SF) approach was first described for single parameter estimation by Schmidt [43], [44] and formalized by Cadzow [45]. Based on the works of Schmidt and Cadzow, we extend the SF approach for single parameter model to the GDPD model and named Unified Subspace Fitting Framework (USF). The differences between the USF and the SF are: (1) the dimension of unknown parameters is increased; (2) the array manifold matrix may be singular due to multipath propagation. Conclusions that are applicable in a single parameter SF approach turns to be no longer applicable in a USF framework because of the differences. Therefore, it is necessary to study the theories and methods of USF framework.

A USF cost function measures the geometric relations between the subspace spanned manifold vectors, the signals subspace and the noise subspace. In order to give readers a more intuitive understanding of USF framework, we propose a geometric interpretation of USF framework.

A. POSITIONING PROBLEM

Denote $\boldsymbol{\theta}^*$ as the true value of unknown parameters. It is seen from (8) that each snapshot $\mathbf{r}(k)$ from the array is just a linear combination of the columns of $\mathbf{A}(k, \boldsymbol{\theta}^*)$, In other words, each observation is constrained to lie in the d-dimensional subspace of $\mathbb{C}^{MN}$ defined by the d columns of $\mathbf{A}(k, \boldsymbol{\theta}^*)$. A geometric schematic of Subspace Fitting Framework is given in FIGURE 3.

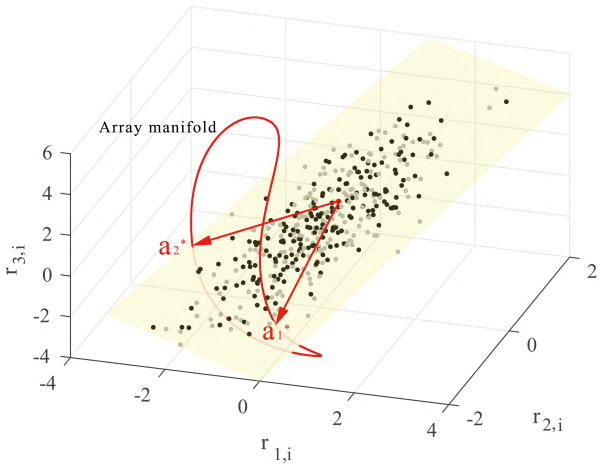


FIGURE 3. Geometric interpretation of positioning problem.

Assume that there are three receivers and two emitters. 200 observations are collected by each receiver which are denoted by $\mathbf{r}_i = [r_{1,i}, r_{2,i}, r_{3,i}]^T$, $i = 1, 2, \ldots, 200$. An observation vector $\mathbf{r}_i$ corresponds to a dot in the FIGURE 3. Substitute the true parameters $\boldsymbol{\theta}^*$ into the manifold matrix $\mathbf{A}(k, \boldsymbol{\theta}^*)$, and columns of $\mathbf{A}(k, \boldsymbol{\theta}^*)$ are denoted as $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$.

From (8), the received signals are represented by

$$\mathbf{r}_i = s_{1,i}\mathbf{a}_1^* + s_{2,i}\mathbf{a}_2^* + \mathbf{n}_i, \tag{25}$$

where $s_{1,i}$ and $s_{2,i}$ are the ith samples of source 1 and source 2, $i = 1, 2, \ldots, 200$. the received signal $\mathbf{r}_i$ is a linear combination of $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$, so the dots in FIGURE 3 are distributed around the subspace spanned by $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$.

Parameters estimation problem for signal model (25) is to find the optimal vectors $\mathbf{a}_1^*$ and $\mathbf{a}_2^*$ which are determined by the unknown parameters $\boldsymbol{\theta}$. From a geometric point of view, it is to search for the optimal plane which is determined by $\boldsymbol{\theta}$, so that the dots are near the plane. We should design a reasonable cost function to measure the performance of a parameter vector $\boldsymbol{\theta}$ to obtain $\boldsymbol{\theta}^*$.

B. MAXIMUM LIKELIHOOD

The most intuitive approach is to fit a plane, which is spanned by $\mathbf{a}_1(\boldsymbol{\theta})$ and $\mathbf{a}_2(\boldsymbol{\theta})$ over the dots and minimize the fitting error:

$$\hat{\boldsymbol{\theta}} = \arg\min g_m(\boldsymbol{\theta}) = \|\mathbf{r}_i - s_{1,i}\mathbf{a}_1(\boldsymbol{\theta}) - s_{2,i}\mathbf{a}_2(\boldsymbol{\theta})\|_F^2, \quad (26)$$

where $\|\cdot\|_F$ means the Frobenius norm. The cost function (26) is known as the ML method if the noise is the AWGN. If the source signal waveforms are unknown, $s_{1,i}$ and $s_{2,i}$ in (26) are required to be optimized. There will be a large scale searching to obtain the waveforms parameters.

C. SIGNAL SUBSPACE PROJECTION

If $s_{1,i}$ and $s_{2,i}$ are incoherent over i, the PCA (Principal Component Analysis) method can find the optimal fitting plane of programming (26) without the optimization of signal waveforms. Eigenvectors, correspond to the largest two eigen values of the covariance matrix of received signals, span the signal subspace (the yellow plane in FIGURE 4), and the noise subspace is orthogonal with the signal subspace.

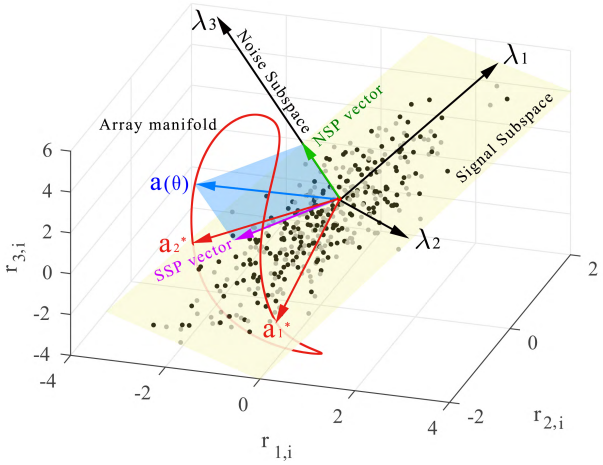


FIGURE 4. Geometric interpretation of 1D-SP-MUSIC.

If we search $\mathbf{a}(\boldsymbol{\theta})$ along the array manifold space (the red curve), and maximize the projection of $\mathbf{a}(\boldsymbol{\theta})$ onto the signal subspace (the green vector):

$$[\hat{\boldsymbol{\theta}}] = \arg\max g_{ps}(\boldsymbol{\theta}) = \|\mathbf{a}(\boldsymbol{\theta})\mathbf{u}_s\|_F^2, \quad (27)$$

we can find local optimal solutions at $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{a}}_2$ which follow $\hat{\mathbf{a}}_1 \to \mathbf{a}_1^*$ and $\hat{\mathbf{a}}_2 \to \mathbf{a}_2^*$. The cost function which maximizes the projections onto the signal subspace one by one is known as the One Dimension-SSP-MUSIC (1D-SSP-MUSIC).

If we serach $\mathbf{a}(\boldsymbol{\theta})$ along the array manifold space, and minimize the projection onto the noise subspace (the purple vector):

$$[\hat{\boldsymbol{\theta}}] = \arg\min g_{pn}(\boldsymbol{\theta}) = \|\mathbf{a}(\boldsymbol{\theta})\mathbf{u}_n\|_F^2. \quad (28)$$

We can find local optimal solutions at $\hat{\mathbf{a}}_1$ and $\hat{\mathbf{a}}_2$ which follow $\hat{\mathbf{a}}_1 \to \mathbf{a}_1^*$ and $\hat{\mathbf{a}}_2 \to \mathbf{a}_2^*$. The cost function which minimizes the projections onto the noise subspace one by one is known as the One Dimension-NSP-MUSIC (1D-NSP-MUSIC).

Remark: If the length of $\mathbf{a}(\boldsymbol{\theta})$ has been normalized to 1, the maximization of the projection onto the signal subspace is equivalent with the minimization of the projection onto the noise subspace, that is 1D-SSP-MUSIC is equivalent with 1D-NSP-MUSIC is the length of manifold vector has been normalize to 1.

D. SIGNAL SUBSPACE FITTING

If we search $\mathbf{a}_1(\boldsymbol{\theta})$ and $\mathbf{a}_2(\boldsymbol{\theta})$ simultaneously, and fit the subspace which is spanned by $\mathbf{a}_1(\boldsymbol{\theta})$,$\mathbf{a}_2(\boldsymbol{\theta})$ (the blue plane in FIGURE 5) to the signal subspace (the yellow plane in FIGURE 5), it is named as Signal Subspace Fitting (SSF) method.

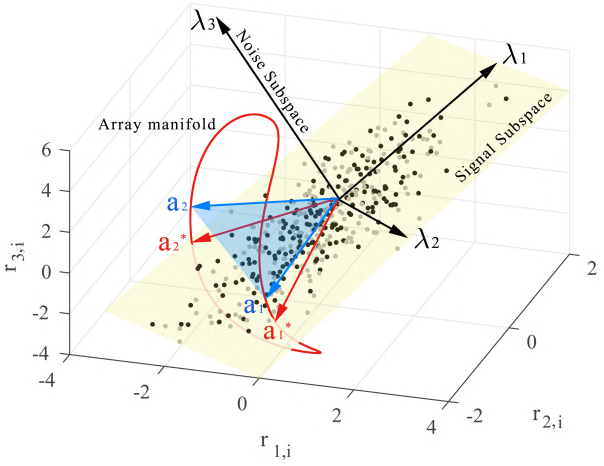


FIGURE 5. Geometric interpretation of SSF.

In the SSF method, fitting the plane spanned by $\mathbf{a}_1(\boldsymbol{\theta})$ and $\mathbf{a}_2(\boldsymbol{\theta})$ to the plane spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$ is equivalent to measuring the coplanarity of those vectors, where $\mathbf{u}_1$ and $\mathbf{u}_2$ is the eigen vectors corresponds to λ_1 and λ_2. If $\mathbf{u}_d, d = 1, 2$, is a linear combination of $\mathbf{a}_1(\boldsymbol{\theta})$ and $\mathbf{a}_2(\boldsymbol{\theta})$, vector $\mathbf{u}_d$ is in the plane spanned by $\mathbf{a}_1(\boldsymbol{\theta})$ and $\mathbf{a}_2(\boldsymbol{\theta})$. In the Multiple Dimension-Signal Subspace Fitting (MD-SSF), we measure the coplanarity of the fitting by

$$[\hat{\boldsymbol{\theta}}] = \arg\min g_s(\boldsymbol{\theta}) = \|\mathbf{U}_s - \mathbf{A}(\boldsymbol{\theta})\mathbf{T}\|_F^2, \quad (29)$$

where $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}_1(\boldsymbol{\theta}), \mathbf{a}_2(\boldsymbol{\theta})]$, $\mathbf{U}_s = [\mathbf{u}_1, \mathbf{u}_2]$, and $\mathbf{T}$ is 2×2 linear transform matrix.

If the manifold vector $\mathbf{a}(\boldsymbol{\theta})$ is in the plane spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$, $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{U}_s\mathbf{T}$. However, if $\mathbf{a}(\boldsymbol{\theta})$ satisfies $\mathbf{a}(\boldsymbol{\theta}) = \mathbf{U}_s\mathbf{T}$, $\mathbf{a}(\boldsymbol{\theta})$ not always in the plane spanned by $\mathbf{u}_1$ and $\mathbf{u}_2$. If $\mathbf{a}(\boldsymbol{\theta}) = 0$ and $\mathbf{T} = 0$, the equation is satisfied, but we do not find the true manifold vector. We can not minimize $\|\mathbf{a}(\boldsymbol{\theta}) - \mathbf{U}_s\mathbf{T}\|$ to obtain the true manifold vector unless constrains $\|\mathbf{a}(\boldsymbol{\theta})\|_F = 1$.

Since the lengths of the column vectors in $\mathbf{U}_s$ have been normalized to 1, we can get the true manifold subspace $\mathbf{A}(\boldsymbol{\theta})$ without the normalization of column vectors of matrix $\mathbf{A}(\boldsymbol{\theta})$.

If we search a manifold vector $\mathbf{a}(\boldsymbol{\theta})$ in the array manifold space, the cost function turns to be a One Dimension-SSF (1D-SSF) cost function:

$$[\hat{\boldsymbol{\theta}}] = \arg\min g_s(\boldsymbol{\theta}) = \|\mathbf{U}_s - \mathbf{a}(\boldsymbol{\theta})\mathbf{T}\|_F^2, \tag{30}$$

and there are two local optimal solutions of programming (30) which corresponds to $\hat{\mathbf{a}}_1(\boldsymbol{\theta})$ and $\hat{\mathbf{a}}_2(\boldsymbol{\theta})$.

The geometric understanding of $\min g_{s,1}(\boldsymbol{\theta}) = \|\mathbf{u}_1 - \mathbf{a}(\boldsymbol{\theta})t_1\|_F$ in (30) is to minimize distance between $\mathbf{u}_1$ and scaled $\mathbf{a}(\boldsymbol{\theta})$. if the manifold subspace is orthogonal with the signal subspace, the projection of $\mathbf{u}_1$ onto the manifold subspace is the optimal estimation of $\mathbf{a}(\boldsymbol{\theta})t_1$, and it is the true estimation parameters vector $\boldsymbol{\theta}^*$.

Since the manifold subspace can not always be orthogonal with the signal subspace, 1D-SSF is a simplified version of MD-SSF, and holds a worse performance than an MD-SSF.

E. NOISE SUBSPACE FITTING

If we search a subspace which is spanned by $\mathbf{a}_1(\boldsymbol{\theta})$, $\mathbf{a}_2(\boldsymbol{\theta})$ and measure the orthogonality with the noise subspace, it is named as Noise Subspace Fitting (NSF) method.

In a direction finding application, the cost function of a MUSIC method is

$$\hat{\theta} = \arg\max g_1(\theta) = \frac{1}{\mathbf{a}^H(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H\mathbf{a}(\boldsymbol{\theta})}. \tag{31}$$

From a geometric point of view, the MUSIC method searches the manifold vector $\mathbf{a}(\boldsymbol{\theta})$ who has the minimal projection length onto the noise subspace. Since $\mathbf{a}^H(\boldsymbol{\theta})\mathbf{a}(\boldsymbol{\theta}) = 1$ always holds, the orthogonality of the two subspaces is correctly defined by $\mathbf{a}^H(\boldsymbol{\theta})\mathbf{U}_s\mathbf{U}_s^H\mathbf{a}(\boldsymbol{\theta})$.

In a single path propagation DPD model, the cost function of a MUSIC method is

$$\hat{\theta} = \arg\max g_2(\theta) = \frac{1}{\boldsymbol{\alpha}^H\mathbf{A}(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H\mathbf{A}^H(\boldsymbol{\theta})\boldsymbol{\alpha}}, \tag{32}$$

where $\boldsymbol{\alpha}$ is the unknown path attenuations column vector, $\mathbf{A}(\boldsymbol{\theta})$ is a diagonal matrix with full rank. From a geometric point of view, $\boldsymbol{\alpha}^H\mathbf{A}(\boldsymbol{\theta})\mathbf{U}_n\mathbf{U}_n^H\mathbf{A}^H(\boldsymbol{\theta})\boldsymbol{\alpha}$ is the projection length of vector $\boldsymbol{\alpha}^H\mathbf{A}(\boldsymbol{\theta})$ onto vector $\mathbf{U}_n$. Since the length of vector $\boldsymbol{\alpha}^H\mathbf{A}(\boldsymbol{\theta})$ is $\boldsymbol{\alpha}^H\mathbf{A}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta})\boldsymbol{\alpha} = \boldsymbol{\alpha}^H\boldsymbol{\alpha} = 1$, the orthogonality of the two subspaces is correctly defined by the MUSIC cost function.

MUSIC cost functions (32) for DPD method are widely used in the existing literatures [2], [3], [5], [8], but it will be failed in a multipath propagation scenario.

In a multipath propagation positioning application, the matrix $\mathbf{A}(\boldsymbol{\theta})$ turns to be a diagonal block matrix. The matrix $\mathbf{A}(\boldsymbol{\theta})$ will be singular if there are two different paths from an emitter to a receiving antenna with the same propagation delay. In this case, we can not guarantee that the length of manifold vector is 1, or even $\exists\boldsymbol{\alpha}$ makes $\boldsymbol{\alpha}^H\mathbf{A}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta})\boldsymbol{\alpha} = 0$. The manifold vector space will be a surface rather than a curve (the red surface in FIGURE 6).

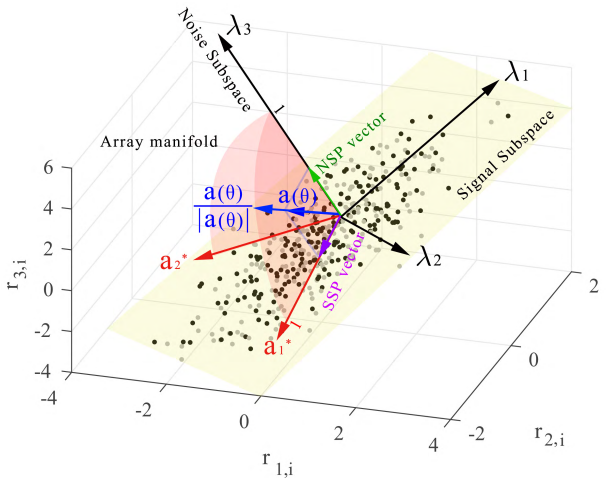


FIGURE 6. Geometric interpretation of NSF.

In a GDPD application, NSP-MUSIC only finds the array manifold vector whose projection length is zero because that its own length is zero, instead of finding the array manifold vector who orthogonal to the noise subspace. In this paper, we normalize the length of $\mathbf{a}(\boldsymbol{\theta})$ to 1 in the cost function to overcome the problem of NSP-MUSIC:

$$\bar{\mathbf{a}}(\boldsymbol{\theta}) \triangleq \frac{\mathbf{a}(\boldsymbol{\theta})}{\|\mathbf{a}(\boldsymbol{\theta})\|_F}, \tag{33}$$

Following the principle that if a space $\mathbf{A}$ is orthogonal with a anther space $\mathbf{B}$, $\mathbf{a}$ is orthogonal with $\mathbf{b}$, where $\mathbf{a}$ is any vector in space $\mathbf{A}$, and $\mathbf{b}$ is any vector in $\mathbf{B}$, We measure the orthogonality of the subspace spanned by $\mathbf{A}(\boldsymbol{\theta})$ and the noise subspace by:

$$\hat{\boldsymbol{\theta}} = \arg\min g_n = \|\bar{\mathbf{A}}^H(\boldsymbol{\theta})\mathbf{U}_n\|_F^2, \tag{34}$$

where $\bar{\mathbf{A}}(\boldsymbol{\theta}) \triangleq [\bar{\mathbf{a}}_1, \bar{\mathbf{a}}_2]$. Since $\bar{\mathbf{a}}_d$ has been normalized to 1, and $\|\bar{\mathbf{u}}_n\|_F = 1$, $\|\bar{\mathbf{a}}_d\bar{\mathbf{u}}_n\|_F^2$ measures the orthogonality of those vectors correctly in a GDPD model in the presence of multipath propagation.

IV. COST FUNCTIONS OF USF

We have studied the intuitive geometric interpretations of USF in section III. We will propose general models of cost functions under the USF framework in this section.

A. SIGNAL SUBSPACE FITTING METHOD

The basic Signal Subspace Fitting (SSF) method is posed as

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] = \arg\min \bar{s}(\boldsymbol{\theta}, \mathbf{T}) = \|\mathbf{U}_s - \mathbf{A}(\boldsymbol{\theta})\mathbf{T}\|_F^2. \tag{35}$$

The columns of $\mathbf{U}_s$ span the signal subspace, $\boldsymbol{\theta}$ is the unknown parameters vector and $\mathbf{A}(\boldsymbol{\theta})$ is the array manifold matrix

which determined by $\boldsymbol{\theta}$. $\|\cdot\|_F$ is the Frobenius norm of a matrix. The SSF method fits the manifold matrix $\mathbf{A}(\boldsymbol{\theta})$ to the signal subspace $\mathbf{U}_s$ by optimizing $\boldsymbol{\theta}$ and $\mathbf{T}$, and the fitting Root-Mean-Square Error (RMSE) is defined as the cost function of an SSF method.

The optimization of programming (35) is separable in $\boldsymbol{\theta}$ and $\mathbf{T}$. For a fixed $\boldsymbol{\theta}$, By substituting the pseudo-inverse solution, $\hat{\mathbf{T}} = \mathbf{A}^{+}(\boldsymbol{\theta})\mathbf{U}_s$, back into (35), we obtain the following equivalent programming

$$\hat{\boldsymbol{\theta}} = \arg\max C_1(\boldsymbol{\theta}) = \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(\boldsymbol{\theta})\mathbf{U}_s\mathbf{U}_s^H\}, \tag{36}$$

where $\mathbf{P}_{\mathbf{A}}(\boldsymbol{\theta}) \triangleq \mathbf{A}(\boldsymbol{\theta})\mathbf{A}^{+}(\boldsymbol{\theta})$ is the projection matrix that projects onto the subspace spanned by the columns of $\mathbf{A}(\boldsymbol{\theta})$, and $\mathbf{A}^{+}(\boldsymbol{\theta}) \triangleq [\mathbf{A}^H(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})]^{-1}\mathbf{A}^H(\boldsymbol{\theta})$is the pseudo-inverse of $\mathbf{A}(\boldsymbol{\theta})$. (36) is a unified cost function of the Subspace Fitting Model, and the different definitions of $\mathbf{U}_s$ and $\mathbf{A}(\boldsymbol{\theta})$ lead to the different cost functions.

1) ONE DIMENSION-SIGNAL SUBSPACE FITTING

For the uncorrelated D source signals scenario, there are D intersections between the manifold subspace and the signal subspace. If we search the vector in the manifold space over the available region, and we can find D vectors witch lie on the signal subspace. One Dimension-Signal Subspace Fitting (1D-SSF) method searches the optimal manifold vector in the manifold space one by one to obtain the D emitter positions. The combinations of multiple emitters are not taken into account in this method. Set $\mathbf{U}_s \triangleq \hat{\mathbf{U}}_s(k)$, $\mathbf{A}(\boldsymbol{\theta}) \triangleq \mathbf{a}(k,\boldsymbol{\theta})$ in the (35), where $\mathbf{a}(k,\boldsymbol{\theta})$ is a column of $\mathbf{A}(k,\boldsymbol{\theta})$, 1D-SSF is defined as:

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] = \arg\min \bar{s}_1(\boldsymbol{\theta}, \mathbf{T}) = \sum_{k=1}^{K} \|\hat{\mathbf{U}}_s(k) - \mathbf{a}(k,\boldsymbol{\theta})\mathbf{T}(k)\|_F^2, \tag{37}$$

where $\boldsymbol{\theta} \triangleq [\mathbf{p}_e, \boldsymbol{\alpha}]$, $\mathbf{a}(k,\boldsymbol{\theta})$ corresponds to the manifold vector of parameters $\boldsymbol{\theta}$. Substitute the optimal estimation of $\mathbf{T}(k)$ into (37), and follows (36)

$$\hat{\boldsymbol{\theta}} = \arg\max s_1(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\}, \tag{38}$$

where $\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta}) \triangleq \mathbf{a}(k,\boldsymbol{\theta})\mathbf{a}^{+}(k,\boldsymbol{\theta})$ is the projection matrix of $\mathbf{a}(k,\boldsymbol{\theta})$, and $\mathbf{a}^{+}(k,\boldsymbol{\theta}) \triangleq \left[\mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{a}(k,\boldsymbol{\theta})\right]^{-1}\mathbf{a}^H(k,\boldsymbol{\theta})$ is the pseudo-inverse of $\mathbf{a}(k,\boldsymbol{\theta})$.

2) MULTIPLE DIMENSIONS-SIGNAL SUBSPACE FITTING

All emitters are taken into account in the Multiple Dimensions Signal Subspace Fitting (MD-SSF) cost function. The manifold matrix $\mathbf{A}(\boldsymbol{\theta}) \triangleq \mathbf{A}(k,\boldsymbol{\theta})$ is a $MN \times D$ matrix. Set $\mathbf{U}_s \triangleq \hat{\mathbf{U}}_s(k)$, and adopt the subspace fitting formulation, the method is posed as:

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] = \arg\min \bar{s}_2(\boldsymbol{\theta}, \mathbf{T}) = \sum_{k=1}^{K} \|\hat{\mathbf{U}}_s(k) - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{T}(k)\|_F^2. \tag{39}$$

Substitute the optimal estimation of $\mathbf{T}(k)$ into (39)

$$\hat{\boldsymbol{\theta}} = \arg\max s_2(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\}, \tag{40}$$

where $\hat{\mathbf{U}}_s(k)$ is defined in (24), and $\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta}) \triangleq \mathbf{A}(k,\boldsymbol{\theta})\mathbf{A}^{+}(k,\boldsymbol{\theta})$ is the projection matrix of $\mathbf{A}(k,\boldsymbol{\theta})$, and $\mathbf{A}^{+}(k,\boldsymbol{\theta}) \triangleq \left[\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})\right]^{-1}\mathbf{A}^H(k,\boldsymbol{\theta})$ is the pseudo-inverse of $\mathbf{A}(k,\boldsymbol{\theta})$.

3) ONE DIMENSION-WEIGHTED SIGNAL SUBSPACE FITTING

Set $\mathbf{U}_s \triangleq \hat{\mathbf{U}}_s(k)\mathbf{W}^{1/2}(k)$, and $\mathbf{A}(\boldsymbol{\theta}) \triangleq \mathbf{a}(k,\boldsymbol{\theta})$, where $\mathbf{a}(k,\boldsymbol{\theta})$ is a column of $\mathbf{A}(k,\boldsymbol{\theta})$, the cost function of a Weighted Signal Subspace Fitting (1D-WSSF) is defined by:

$$\begin{aligned}[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] &= \arg\min \bar{s}_3(\boldsymbol{\theta}, \mathbf{T}) \\ &= \sum_{k=1}^{K} \|\hat{\mathbf{U}}_s(k)\mathbf{W}^{1/2}(k) - \mathbf{a}(k,\boldsymbol{\theta})\mathbf{T}(k)\|_F^2.\end{aligned} \tag{41}$$

Substitute the optimal estimation of $\mathbf{T}(k)$ into (41):

$$\hat{\boldsymbol{\theta}} = \arg\max s_3(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\mathbf{W}(k)\hat{\mathbf{U}}_s^H(k)\}, \tag{42}$$

where $\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})$ has been defined in 1D-SSF, $\mathbf{W}^{1/2}(k)$ is a diagonal matrix with size of $D \times D$, and the dth element in the diagonal is the weight of the dth eigenvector in the signal subspace ($d = 1, 2, \ldots, D$).

4) MULTIPLE DIMENSION WEIGHTED SIGNAL SUBSPACE FITTING

Set $\mathbf{U}_s \triangleq \hat{\mathbf{U}}_s(k)\mathbf{W}^{1/2}(k)$, and $\mathbf{A}(\boldsymbol{\theta}) \triangleq \mathbf{A}(k,\boldsymbol{\theta})$, the cost function of Multiple Dimension Weighted Signal Subspace Fitting (MD-WSSF) is defined by:

$$\begin{aligned}[\hat{\boldsymbol{\theta}}, \hat{\mathbf{T}}] &= \arg\min \bar{s}_4(\boldsymbol{\theta}, \mathbf{T}) \\ &= \sum_{k=1}^{K} \|\hat{\mathbf{U}}_s(k)\mathbf{W}^{1/2}(k) - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{T}(k)\|_F^2.\end{aligned} \tag{43}$$

Substitute the optimal estimation of $\mathbf{T}(k)$ into (43):

$$\hat{\boldsymbol{\theta}} = \arg\max s_4(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\mathbf{W}(k)\hat{\mathbf{U}}_s^H(k)\}, \tag{44}$$

where $\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta}) = \mathbf{A}(k,\boldsymbol{\theta})\mathbf{A}^{+}(k,\boldsymbol{\theta})$, and $\mathbf{W}^{1/2}(k)$ has been defined in 1D-WSSF.

5) DETERMINISTIC MAXIMUM LIKELIHOOD

We choose $\mathbf{U}_s \triangleq \mathbf{r}(k)$, $\mathbf{A}(\boldsymbol{\theta}) \triangleq \mathbf{A}(k,\boldsymbol{\theta})$ and $\mathbf{T} = \mathbf{S}(k)$ in (35), and it gives the Deterministic Maximum Likelihood (DML) cost function. For an AWGN channel, cost function of the DML method is

$$[\hat{\boldsymbol{\theta}}, \hat{\mathbf{S}}] = \arg\min \bar{s}_5(\boldsymbol{\theta}, \mathbf{S}) = \sum_{k=1}^{K} \|\mathbf{r}(k) - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)\|_F^2. \tag{45}$$

Follows with (36),

$$\hat{\boldsymbol{\theta}} = \arg\max s_5(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\mathbf{r}(k)\mathbf{r}^H(k)\} = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{R}}(k)\}, \tag{46}$$

where $\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta}) = \mathbf{A}(k,\boldsymbol{\theta})\mathbf{A}^{+}(k,\boldsymbol{\theta})$, The ML method fits a $3D + DNL$-dimensional subspace spanned by the columns of $\mathbf{A}(k)$ to the received data $\mathbf{r}(k)$.

B. NOISE SUBSPACE FITTING METHOD

Since the noise subspace of the received signals is orthogonal with the signal subspace, fitting the manifold space onto the signal subspace is equivalent to minimize the manifold projection length onto the noise subspace.

The angle between $\mathbf{a}(k,\boldsymbol{\theta})$ and $\mathbf{u}_n(k,i)$ is:

$$\cos(\beta_i) = \frac{\|\mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{u}_n(k,i)\|_F}{\|\mathbf{a}(k,\boldsymbol{\theta})\|_F \|\mathbf{u}_n(k,i)\|_F}, \tag{47}$$

where $\mathbf{a}(k,\boldsymbol{\theta})$ is a vector in the manifold matrix, and $\mathbf{u}_n(k,i)$ is the ith column of $\mathbf{U}_n(k)$. If $\cos(\beta_i) = 0$, the vector $\mathbf{a}(k,\boldsymbol{\theta})$ is orthogonal with $\mathbf{u}_n(k,i)$. If the space $\mathbf{A}(k,\boldsymbol{\theta})$ is orthogonal with the space $\mathbf{U}_n(k)$, each vector in the space $\mathbf{A}(k,\boldsymbol{\theta})$ is orthogonal with all vectors in the space $\mathbf{U}_n(k)$. In a direction finding application or a single path DPD application, $\forall\boldsymbol{\theta}$, $\|\mathbf{a}(k,\boldsymbol{\theta})\|_F = 1$, and $\|\mathbf{u}_n(k,i)\|_F = 1$. In this case, the measurement of the orthogonality is simplified as

$$\cos(\beta_i) = \mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{u}_n(k,i). \tag{48}$$

The NSF methods for direction finding or single-path propagation applications minimize $\cos(\beta_i)$ over $\boldsymbol{\theta}$, and get the optimal estimation of $\boldsymbol{\theta}$:

$$\hat{\boldsymbol{\theta}} = \arg\min n(\boldsymbol{\theta}) = \|\mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{u}_n(k,i)\|_F^2. \tag{49}$$

However, the length of a manifold vector in a GDPD model can not always keep 1 [1]. In a scenario with multipath propagation, $\exists\boldsymbol{\theta}$ and d, makes the length of the dth column vector in the manifold matrix $\mathbf{A}(k,\boldsymbol{\theta})$ satisfies $\|\mathbf{a}_d(k,\boldsymbol{\theta})\| = 0$, and leads to $\|\hat{\mathbf{u}}_n^H(k,i)\mathbf{a}_d(k,\boldsymbol{\theta})\| = 0, \forall i$. In this case, programming (49) reaches the minimum value, since $\|\cdot\|_F \geq 0$. However, $\boldsymbol{\theta}$ is a fake solution (In FIGURE 3, the manifold curve intersects the signal space near to zero).

The length of a manifold vector should be normalized in the Noise Subspace Fitting method to avoid finding the fake solutions.

1) ONE DIMENSION-NOISE SUBSPACE FITTING

One Dimension-Noise Subspace Fitting (1D-NSF) method normalize the length of all vectors in the manifold matrix, and minimize the manifold vector projection length onto the noise subspace.

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min \bar{u}_1(\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{MN-D} \frac{\|\hat{\mathbf{u}}_n^H(k,i)\mathbf{a}(k,\boldsymbol{\theta})\|_F^2}{\|\mathbf{a}(k,\boldsymbol{\theta})\|_F^2} \\ &= \sum_{k=1}^{K} \frac{\|\hat{\mathbf{U}}_n^H(k)\mathbf{a}(k,\boldsymbol{\theta})\|_F^2}{\|\mathbf{a}(k,\boldsymbol{\theta})\|_F^2} \\ &= \sum_{k=1}^{K} \frac{\mathbf{a}^H(k,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\mathbf{a}(k,\boldsymbol{\theta})}{\mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{a}(k,\boldsymbol{\theta})}, \\ &= \sum_{k=1}^{K} \mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\right\}.\end{aligned} \tag{50}$$

2) MULTIPLE DIMENSION-NOISE SUBSPACE FITTING

Multiple Dimension-Noise Subspace Fitting (MD-NSF) method normalize the length of all vectors in the manifold matrix, and minimize the manifold matrix projection length onto the noise subspace.

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min \bar{u}_2(\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{MN-D}\sum_{d=1}^{D} \frac{\|\hat{\mathbf{u}}_n^H(k,i)\mathbf{a}_d(k,\boldsymbol{\theta})\|_F^2}{\|\mathbf{a}_d(k,\boldsymbol{\theta})\|_F^2} \\ &= \sum_{k=1}^{K}\sum_{d=1}^{D} \frac{\|\hat{\mathbf{U}}_n^H(k)\mathbf{a}_d(k,\boldsymbol{\theta})\|_F^2}{\|\mathbf{a}_d(k,\boldsymbol{\theta})\|_F^2} \\ &= \sum_{k=1}^{K}\sum_{d=1}^{D} \frac{\mathbf{a}_d^H(k,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\mathbf{a}_d(k,\boldsymbol{\theta})}{\mathbf{a}_d^H(k,\boldsymbol{\theta})\mathbf{a}_d(k,\boldsymbol{\theta})} \\ &= \sum_{k=1}^{K}\sum_{d=1}^{D} \mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,d,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\right\}.\end{aligned} \tag{51}$$

3) MULTIPLE DIMENSION-WEIGHTED NOISE SUBSPACE FITTING

Weight the D manifold vectors as:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min \bar{u}_3(\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K}\sum_{i=1}^{MN-D}\sum_{d=1}^{D} \frac{\|\hat{\mathbf{u}}_n^H(k,i)\mathbf{a}_d(k,\boldsymbol{\theta})w_d^{\frac{1}{2}}\|_F^2}{\|\mathbf{a}_d(k,\boldsymbol{\theta})\|_F^2} \\ &= \sum_{k=1}^{K}\sum_{d=1}^{D} \frac{\|\hat{\mathbf{U}}_n^H(k)\mathbf{a}_d(k,\boldsymbol{\theta})w_d^{\frac{1}{2}}\|_F^2}{\|\mathbf{a}_d(k,\boldsymbol{\theta})\|_F^2} \\ &= \sum_{k=1}^{K}\sum_{d=1}^{D} w_d \frac{\mathbf{a}^H(k,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\mathbf{a}_d(k,\boldsymbol{\theta})}{\mathbf{a}_d^H(k,\boldsymbol{\theta})\mathbf{a}_d(k,\boldsymbol{\theta})} \\ &= \sum_{k=1}^{K}\sum_{d=1}^{D} w_d \mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,d,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\right\}.\end{aligned} \tag{52}$$

C. UNIFIED SUBSPACE FITTING COST FUNCTION FAMILY

The MUSIC method is the most commonly used cost function in the DPD framework for unknown waveform signals. In a

single path propagation positioning application, the length of a manifold vector is constrained to 1, and the existing cost functions ignore the lengths of manifold vectors. Weiss et al. simplified the Noise Subspace Projection to the Signal Subspace Projection further for reducing the computational complexity.

We pointed out that the length of a manifold vector can not be ignored in a GDPD model in the presence of multipath propagation, and proposed 8 cost functions which can be used in a GDPD model. The eight cost functions are listed in Table 1.

TABLE 1. Unified subspace fitting (USF) cost functions.

USF				
SSF			NSF	
	1D-SSF	1D-WSSF	1D-NSF	
DML	MD-SSF	MD-WSSF	MD-NSF	MD-WNSF

The USF framework is divided into the Signal Subspace Fitting (SSF) methods and the Noise Subspace Fitting (NSF) methods. If a column vector of the manifold matrix is adopted to measure the relation of the array manifold space and the signal/noise subspace, it is named 1D-Subspace Fitting method, otherwise, it is named MD-Subspace Fitting method.

V. ASYMPTOTIC DISTRIBUTION ANALYSIS

We will analysis the asymptotic distributions of the estimation errors which are optimized by those cost functions in this section.

For a 3D geolocation application, there are $3 + LN$ unknown parameters of each emitter, and they are organized in a vector which is denoted as

$$\hat{\boldsymbol{\theta}} = \left[\hat{\mathbf{p}}_{\mathrm{e}}^T(x) \quad \hat{\mathbf{p}}_{\mathrm{e}}^T(y) \quad \hat{\mathbf{p}}_{\mathrm{e}}^T(z) \quad \hat{\boldsymbol{\alpha}}_1^T \quad \hat{\boldsymbol{\alpha}}_2^T \quad \dots \quad \hat{\boldsymbol{\alpha}}_L^T\right] \triangleq [\hat{\boldsymbol{\theta}}_1^T, \hat{\boldsymbol{\theta}}_2^T, \dots, \hat{\boldsymbol{\theta}}_{3+LN}^T]^T, \tag{53}$$

where

$$\hat{\mathbf{p}}_{\mathrm{e}}(\nu) \triangleq [\hat{p}_{\mathrm{e}}(\nu,1), \hat{p}_{\mathrm{e}}(\nu,2), \cdots, \hat{p}_{\mathrm{e}}(\nu,D)]^T, \tag{54}$$

$$\hat{\boldsymbol{\alpha}}_\ell \triangleq [\hat{\boldsymbol{\alpha}}_{\ell,1}^T, \hat{\boldsymbol{\alpha}}_{\ell,1}^T, \cdots, \hat{\boldsymbol{\alpha}}_{\ell,N}^T]^T,$$

$$\hat{\boldsymbol{\alpha}}_{\ell,n} \triangleq [\hat{\alpha}_{\ell,1,1}, \hat{\alpha}_{\ell,1,2}, \cdots, \hat{\alpha}_{\ell,1,N}]^T, \tag{55}$$

$\nu \in \{x, y, z\}$, and $\hat{\boldsymbol{\theta}}_i$ is a $D\times 1$ column vector, which represents the ith unknown parameter vector.

A. ASYMPTOTIC DISTRIBUTION OF THE MD-WSSF

The cost function of MD-WSSF is defined in (44), and the asymptotic distribution of an MD-WSSF cost function is given by Theorem 1.

Theorem 1: Assume that $\mathbf{W}(k)$, $k = 1, 2, \dots, K$, *are Hermitian positive definite matrices.* $\hat{\boldsymbol{\theta}}$ *and* $\boldsymbol{\theta}_0$ *are the optimal solution and true values of programming* (44). *The normalized estimation error,* $\sqrt{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, *is asymptotically normal distributed with zero mean and covariance matrix* $\boldsymbol{\Sigma}$, *which is given by:*

$$\boldsymbol{\Sigma} = (\bar{\mathbf{C}}'')^{-1}\mathbf{Q}(\bar{\mathbf{C}}'')^{-1}, \tag{56}$$

$$\bar{\mathbf{C}}'' = -2\sum_{k=1}^{K} \mathrm{Re}\left\{\mathbf{F}(k,\boldsymbol{\theta}) \odot [\mathbf{1}_{3+LN} \otimes \mathbf{E}^T(k)]\right\}, \tag{57}$$

$$\mathbf{Q} = 2\sum_{k=1}^{K} \sigma_k^2 \mathrm{Re}\{\mathbf{F}(k,\boldsymbol{\theta}) \odot [\mathbf{1}_{3+LN} \otimes \mathbf{G}^T(k)]\}, \tag{58}$$

$$\mathbf{F}(k,\boldsymbol{\theta}) = \mathbf{A}_{\boldsymbol{\theta}}^H(k)\mathbf{P}_{\mathbf{A}}^{\perp}(k)\mathbf{A}_{\boldsymbol{\theta}}(k), \tag{59}$$

$$\mathbf{A}_{\boldsymbol{\theta}}(k) = \left[\mathbf{A}_{\boldsymbol{\theta}_1}(k), \mathbf{A}_{\boldsymbol{\theta}_2}(k), \dots, \mathbf{A}_{\boldsymbol{\theta}_{3+LN}}(k)\right], \tag{60}$$

$$\mathbf{A}_{\boldsymbol{\theta}_i}(k) = \left[\frac{\partial \mathbf{a}_1(k)}{\partial\theta_i(1)}, \frac{\partial \mathbf{a}_2(k)}{\partial\theta_i(2)}, \dots, \frac{\partial \mathbf{a}_D(k)}{\partial\theta_i(D)}\right], \tag{61}$$

$$\mathbf{G}(k,\boldsymbol{\theta}) \triangleq \mathbf{A}^{+}(k)\mathbf{U}_s(k)\mathbf{W}(k)\boldsymbol{\Lambda}_s(k)\tilde{\boldsymbol{\Lambda}}^{-2}(k) \times \mathbf{W}(k)\mathbf{U}_s(k)^H\mathbf{A}^{+H}(k), \tag{62}$$

$$\mathbf{E}(k,\boldsymbol{\theta}) \triangleq \mathbf{A}^{+}(k)\mathbf{U}_s(k)\mathbf{W}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k), \tag{63}$$

$$\boldsymbol{\Lambda}_s(k) \triangleq \mathrm{diag}[\lambda_1(k), \lambda_2(k), \dots, \lambda_D(k)], \tag{64}$$

$$\tilde{\boldsymbol{\Lambda}}(k) \triangleq [\boldsymbol{\Lambda}_s(k) - \sigma_k^2\mathbf{I}], \tag{65}$$

where $\mathbf{a}_d(k)$ *is the* d*th column of* $\mathbf{A}(k,\boldsymbol{\theta})$, $\lambda_d(k)$ *is the* d*th eigenvalue of the covariance matrix* $\mathbf{R}$, σ_k *is the noise standard deviation of frequency* ω_k. $\mathbf{1}_{3+LN}$ *is a* $(3+LN)\times(3+LN)$ *matrix with all ones,* $\mathbf{I}$ *is a* $D \times D$ *identity matrix.*

Proof: Since $\hat{\boldsymbol{\theta}}$ the optimal solution of programming (44), we have $\mathbf{C}'(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. The first order of Taylor series expansion of $\mathbf{C}'(\hat{\boldsymbol{\theta}})$ around the true $\boldsymbol{\theta}_0$ leads to

$$\mathbf{0} = \mathbf{C}'(\boldsymbol{\theta}_0) + \mathbf{C}''(\boldsymbol{\theta}_\xi)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0), \tag{66}$$

where $\boldsymbol{\theta}_\xi$ is a point on the line segment joining $\boldsymbol{\theta}_0$ and $\hat{\boldsymbol{\theta}}$.

Denote the limiting second order derivative of $\mathbf{C}(\boldsymbol{\theta})$ by

$$\bar{\mathbf{C}}''(\boldsymbol{\theta}) \triangleq \lim_{J\to\infty} \mathbf{C}''(\boldsymbol{\theta}). \tag{67}$$

Since

$$\|\mathbf{C}''(\boldsymbol{\theta}_\xi) - \bar{\mathbf{C}}''(\boldsymbol{\theta})\|_F \le \|\mathbf{C}''(\boldsymbol{\theta}_\xi) - \mathbf{C}''(\boldsymbol{\theta}_0)\|_F + \|\mathbf{C}''(\boldsymbol{\theta}_0) - \bar{\mathbf{C}}(\boldsymbol{\theta}_0)\|_F. \tag{68}$$

Reference [36, Th. 1] proved that $\hat{\boldsymbol{\theta}}$ obtained from (44) converges w.p.1 to $\boldsymbol{\theta}_0$ as $J \to \infty$. Since $C''(\boldsymbol{\theta})$ is continuous by assumption, the first term on the right side $\|\mathbf{C}''(\boldsymbol{\theta}_\xi) - \mathbf{C}''(\boldsymbol{\theta}_0)\|_F \to 0$ because of the limiting definition of (67).

Consequently, $\mathbf{C}''(\boldsymbol{\theta}_\xi) \to \bar{\mathbf{C}}''(\boldsymbol{\theta}_0)$ w.p.1. and we assume that $\bar{\mathbf{C}}''(\boldsymbol{\theta}_0)$ is invertible. For large J we then have

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\{\bar{\mathbf{C}}''(\boldsymbol{\theta}_0)\}^{-1}\mathbf{C}'(\boldsymbol{\theta}_0) + o(\mathbf{C}'(\boldsymbol{\theta}_0)). \tag{69}$$

The first order deviation of the cost function $\mathbf{C}'(\boldsymbol{\theta}_0)$ has been defined in (197), and the second order deviation $\mathbf{C}''(\boldsymbol{\theta}_0)$ defined in (215).

It follows from Lemma 4 that for large number of snapshots J, the estimated $\hat{\mathbf{U}}_s(k) = \mathbf{U}_s(k) + o_p(J^{-1/2})$, and we get the ηth element of $\mathbf{C}'(\boldsymbol{\theta}_0)$:

$$C_\eta(\boldsymbol{\theta}_0) = 2\sum_{k=1}^{K}\sum_{d=1}^{D} \mathrm{Re}[\mathbf{w}^{dH}(k)\mathbf{U}_s^H(k) \times \mathbf{A}^{+H}(\boldsymbol{\theta}_0,k)\mathbf{A}_\eta^H(\boldsymbol{\theta}_0,k)\mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}_0,k)\hat{\mathbf{u}}_{s,d}(k)] + o_p(J^{-1/2}), \tag{70}$$

where $\hat{\mathbf{u}}_{s,d}(k)$ is the dth column of $\hat{\mathbf{U}}_s(k)$. Define the signal space estimation error vector $\tilde{\mathbf{u}}_{s,d}(k) \triangleq \hat{\mathbf{u}}_{s,d}(k) - \mathbf{u}_{s,d}(k)$, and notice that $\mathbf{P}_{\mathbf{A}}^{\perp}(k,\boldsymbol{\theta}_0)\mathbf{u}_{s,d}(k) = 0, d = 1, 2, \ldots, D$, the (70) turns to be:

$$C_\eta(\boldsymbol{\theta}_0) = 2\sum_{k=1}^{K}\sum_{d=1}^{D}\mathrm{Re}\Big[\mathbf{w}^{dH}(k)\mathbf{U}_s^H(k) \times \mathbf{A}^{+H}(\boldsymbol{\theta}_0,k)\mathbf{A}_\eta^H(\boldsymbol{\theta}_0,k) \times \mathbf{P}_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}_0,k)\tilde{\mathbf{u}}_{s,d}(k)\Big] + o_p(J^{-1/2}). \quad (71)$$

Since $\tilde{\mathbf{u}}_{s,d}(k)$ is asymptotically normal, the gradient, $C'(\boldsymbol{\theta}_0)$, is also asymptotically normal:

$$\sqrt{J}C'(\boldsymbol{\theta}_0) \in \mathrm{AsN}(\mathbf{0}, \mathbf{Q}), \quad (72)$$

where AsN$(\cdot)$ means the Asymptotically Normal Distribution,and the ηth row and ξth column element of matrix $\mathbf{Q}$ is defined by:

$$Q_{\eta\xi} = \lim_{J\to\infty} J \cdot E\left\{C_\eta C_\xi\right\}. \quad (73)$$

The normalized estimation error, $\sqrt{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$, is asymptotically normal distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}$, which is given by:

$$\boldsymbol{\Sigma} = (\bar{\mathbf{C}}'')^{-1}\mathbf{Q}(\bar{\mathbf{C}}'')^{-1}. \quad (74)$$

Substitute (71) into (73) as:

$$Q_{\eta\xi} = 4\lim_{J\to\infty} J \cdot E\left\{\sum_{k_i=1}^{K}\sum_{d_i=1}^{D}\mathrm{Re}\left[\Psi_{\xi,d_i}(k_1)\tilde{\mathbf{u}}_{s,d_i}(k_i)\right] \cdot \sum_{k_j=1}^{K}\sum_{d_j=1}^{D}\mathrm{Re}\left[\Psi_{\eta,d_j}(k_j)\tilde{\mathbf{u}}_{s,d_j}(k_j)\right]\right\}, \quad (75)$$

where $\Psi_{\xi,d_i}(k) \triangleq \mathbf{w}_{d_i}^H(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k)\mathbf{A}_\xi^H(k)\mathbf{P}_{\mathbf{A}}^{\perp}(k)$, $\Psi_{\eta,d_j}(k) \triangleq \mathbf{w}_{d_j}^H(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k)\mathbf{A}_\eta^H(k)\mathbf{P}_{\mathbf{A}}^{\perp}(k)$. Since $\mathrm{Re}\{x\}\mathrm{Re}\{y\} = \frac{1}{2}\mathrm{Re}(xy^T + xy^H)$,

(75) turns to be

$$Q_{\eta\xi} = 2\lim_{J\to\infty}\sum_{k=1}^{K}\sum_{d_i=1}^{D}\sum_{d_j=1}^{D} \times \mathrm{Re}\left\{J\Psi_{\xi,d_i}(k)E[\tilde{\mathbf{u}}_{s,d_i}(k)\tilde{\mathbf{u}}_{s,d_j}^T(k)]\Psi_{\eta,d_j}^T(k) + J\Psi_{\xi,d_i}(k)E[\tilde{\mathbf{u}}_{s,d_i}(k)\tilde{\mathbf{u}}_{s,d_j}^T(k)]\Psi_{\eta,d_j}^H(k)\right\}. \quad (76)$$

Apply Lemma 4 to (76)

$$Q_{\eta\xi} = 2\sum_{k=1}^{K}\sum_{d_i=1}^{D}\mathrm{Re}\Big\{ - \sum_{\substack{d_j=1\\ d_j\neq d_i}}^{D}\left[\Lambda(\lambda_{d_i},\lambda_{d_j},k)\Psi_{\xi,d_i}(k)\mathbf{u}_{s,d_j}(k)\mathbf{u}_{s,d_i}^T(k)\Psi_{\eta,d_j}^T(k)\right] + \sum_{\substack{d=1\\ d\neq d_i}}^{D}\left[\Lambda(\lambda_{d_i},\lambda_d,k)\Psi_{\xi,d_i}(k)\mathbf{u}_{s,d}(k)\mathbf{u}_{s,d}^H(k)\Psi_{\eta,d_j}^H(k)\right] + \Lambda(\lambda_{d_i},\sigma_k^2,k)\Psi_{\xi,d_i}(k)\mathbf{U}_n(k)\mathbf{U}_n^H(k)\Psi_{\eta,d_j}^H(k)\Big\}, \quad (77)$$

where $\Lambda(\lambda_i,\lambda_j,k) \triangleq \frac{\lambda_i(k)\lambda_j(k)}{[\lambda_i(k)-\lambda_j(k)]^2}$.

Since $\mathbf{P}_{\mathbf{A}}^{\perp}(k)\mathbf{u}_{s,d}(k) = 0$, we get $\mathbf{u}_{s,d}^T(k)\mathbf{P}_{\mathbf{A}}^{\perp T}(k) = 0$, and $\mathbf{u}_{s,d}^H(k)\mathbf{P}_{\mathbf{A}}^{\perp H}(k) = 0$. Equation (77) turns to be

$$Q_{\eta\xi} = 2\sum_{k=1}^{K}\sum_{d=1}^{D}\mathrm{Re}\left\{\Lambda(\lambda_d,\sigma_k^2,k)\Psi_{\xi,d}(k)\mathbf{U}_n(k)\mathbf{U}_n^H(k)\Psi_{\eta,d}^H(k)\right\}. \quad (78)$$

Noting that

$$\mathbf{P}_{\mathbf{A}}^{\perp}(k)\mathbf{U}_n(k)\mathbf{U}_n^H(k) = \mathbf{P}_{\mathbf{A}}^{\perp}(k)[\mathbf{I} - \mathbf{U}_s(k)\mathbf{U}_s^H(k)] = \mathbf{P}_{\mathbf{A}}^{\perp}(k), \quad (79)$$

$$Q_{\eta\xi} = 2\sum_{k=1}^{K}\sigma_k^2\mathrm{Re}\left\{\mathrm{tr}[\mathbf{D}_{\xi\eta}(k)\mathbf{G}(k,\boldsymbol{\theta})]\right\}, \quad (80)$$

where

$$\mathbf{D}_{\xi\eta}(k) \triangleq \mathbf{A}_\xi^H(k)\mathbf{P}_{\mathbf{A}}^{\perp}(k)\mathbf{A}_\eta(k); \quad (81)$$

$$\mathbf{G}(k,\boldsymbol{\theta}) \triangleq \mathbf{A}^{+}(k)\mathbf{U}_s(k)\mathbf{W}(k)\boldsymbol{\Lambda}_s(k)\tilde{\boldsymbol{\Lambda}}^{-2}(k) \times \mathbf{W}(k)\mathbf{U}_s(k)^H\mathbf{A}^{+H}(k); \quad (82)$$

$$\boldsymbol{\Lambda}_s(k) \triangleq \mathrm{diag}[\lambda_1(k),\lambda_2(k),\ldots,\lambda_D(k)]; \quad (83)$$

$$\tilde{\boldsymbol{\Lambda}}(k) \triangleq \boldsymbol{\Lambda}_s(k) - \sigma_k^2\mathbf{I}. \quad (84)$$

The matrix form of (80) is

$$\mathbf{Q} = 2\sum_{k=1}^{K}\sigma_k^2\mathrm{Re}\{\mathbf{F}(k,\boldsymbol{\theta}) \odot [\mathbf{1}_{3+LN} \otimes \mathbf{G}^T(k)]\}, \quad (85)$$

where $\mathbf{1}_{3+LN}$ is a $(3+LN)\times(3+LN)$ matrix with all ones, and $\mathbf{F}(k,\boldsymbol{\theta})$ is defined in (211)

Substitute $\mathbf{Q}(\boldsymbol{\theta}_0)$ in (85) and $\bar{\mathbf{C}}''(\boldsymbol{\theta}_0)$ in (215) into (69):

$$\sqrt{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \in \mathrm{AsN}(0, \boldsymbol{\Sigma}), \quad (86)$$

where $\boldsymbol{\Sigma} = (\bar{\mathbf{C}}'')^{-1}\mathbf{Q}(\bar{\mathbf{C}}'')^{-H}$. □

Theorem 1 is the base theorem for the asymptotic distribution analysis of the SSF framework, and it gives the asymptotic distribution performance of an MD-WSSF cost function with $O(J^{-1})$.

B. ASYMPTOTIC DISTRIBUTION OF MD-SSF

If the weights of the signal subspace are same, that is $\mathbf{W}_k = \mathbf{I}$, MD-WSSF degenerates to MD-SSF. Substitute $\mathbf{W}_k = \mathbf{I}$ into theorem 1, and get the asymptotic distribution of an MD-SSF estimator from theorem 2.

Theorem 2: The MUSIC ($\mathbf{W}_k = \mathbf{I}$) estimation errors $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ are asymptotically (for large J) joinly Gaussian distributed

with zero means and the asymptotically covariances are given in (56),(57),(58) *with*

$$\mathbf{G}(k,\boldsymbol{\theta}) = [\mathbf{A}^2(k)\mathbf{S}(k)\mathbf{A}^2(k)]^{-1} + \sigma^2[\mathbf{A}^2(k)\mathbf{S}^H(k)\mathbf{A}^2(k)\mathbf{S}(k)\mathbf{A}^2(k)]^{-1}, \quad (87)$$

$$\mathbf{E}(k,\boldsymbol{\theta}) = [\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}. \quad (88)$$

where $\mathbf{A}^2(k) \triangleq \mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})$.

Proof: The covariance of received signals is

$$\begin{aligned}&\mathbf{A}(k,\boldsymbol{\theta})\mathbf{S}(k)\mathbf{A}^H(k) + \sigma_k^2\mathbf{I}\\ &= \mathbf{R}(k)\\ &= \mathbf{U}_s(k)\boldsymbol{\Lambda}_s(k)\mathbf{U}_s^H(k) + \mathbf{U}_n(k)\boldsymbol{\Lambda}_n(k)\mathbf{U}_n^H(k)\\ &= \mathbf{U}_s(k)\boldsymbol{\Lambda}_s(k)\mathbf{U}_s^H(k) + \sigma_k^2\mathbf{U}_n(k)\mathbf{U}_n^H(k). \end{aligned} \quad (89)$$

Denote $\tilde{\boldsymbol{\Lambda}}(k) \triangleq \boldsymbol{\Lambda}_s(k) - \sigma_k^2\mathbf{I}$,

$$\mathbf{A}(k,\boldsymbol{\theta})\mathbf{S}(k)\mathbf{A}^H(k) = \mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}(k)\mathbf{U}_s^H(k); \quad (90)$$

$$\tilde{\boldsymbol{\Lambda}}(k) = \mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})\mathbf{S}(k)\mathbf{A}^H(k)\mathbf{U}_s(k); \quad (91)$$

$$[\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\tilde{\boldsymbol{\Lambda}}(k)[\mathbf{A}^H(k)\mathbf{U}_s(k)]^{-1} = \mathbf{S}(k); \quad (92)$$

$$\mathbf{A}^H(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^{-1}(k)\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta}) = \mathbf{S}^{-1}(k). \quad (93)$$

$$\begin{aligned}&\tilde{\boldsymbol{\Lambda}}^{-1}(k) = [\mathbf{A}^H(k)\mathbf{U}_s(k)]^{-1}\mathbf{S}^{-1}(k)[\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1};\\ &\tilde{\boldsymbol{\Lambda}}^{-2}(k) = \left\{[\mathbf{A}^H(k)\mathbf{U}_s(k)]^{-1}\mathbf{S}^{-1}(k)[\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\right\}^2; \end{aligned} \quad (94)$$

$$\begin{aligned}&\mathbf{A}^H(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^{-2}(k)\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})\\ &\quad= \mathbf{S}^{-1}(k)[\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}[\mathbf{A}^H(k)\mathbf{U}_s(k)]^{-1}\mathbf{S}^{-H}(k);\\ &\quad= \mathbf{S}^{-1}(k)[\mathbf{A}^H(k)\mathbf{U}_s(k)\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\mathbf{S}^{-H}(k). \end{aligned} \quad (95)$$

Since

$$\begin{aligned}\mathbf{P}_{\mathbf{A}}(k)\mathbf{U}_s(k) &= \mathbf{U}_s(k);\\ \mathbf{A}(k,\boldsymbol{\theta})[\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\mathbf{A}^H(k)\mathbf{U}_s(k) &= \mathbf{U}_s(k);\\ \mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})[\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\mathbf{A}^H(k)\mathbf{U}_s(k) &= \mathbf{I}_D;\\ \mathbf{A}^H(k)\mathbf{U}_s(k)\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta}) &= \mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta}). \end{aligned} \quad (96)$$

substitute (96) into (95)

$$\begin{aligned}&\mathbf{A}^H(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^{-2}(k)\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})\\ &\quad= \mathbf{S}^{-1}(k)[\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\mathbf{S}^{-H}(k). \end{aligned} \quad (97)$$

Notice that

$$\begin{aligned}&\boldsymbol{\Lambda}_s(k)\tilde{\boldsymbol{\Lambda}}^{-2}(k)\\ &= \mathrm{diag}\left\{\frac{\lambda_1(k)}{[\lambda_1(k)-\sigma_k^2]^2}, \frac{\lambda_2(k)}{[\lambda_2(k)-\sigma_k^2]^2}, \ldots, \frac{\lambda_D}{[\lambda_D(k)-\sigma_k^2]^2}\right\}\\ &= \mathrm{diag}\left[\frac{1}{\lambda_1(k)-\sigma_k^2}, \frac{1}{\lambda_2(k)-\sigma_k^2}, \ldots, \frac{1}{\lambda_D(k)-\sigma_k^2}\right]\\ &\quad+ \mathrm{diag}\left\{\frac{\sigma_k^2}{[\lambda_1(k)-\sigma_k^2]^2}, \frac{\sigma_k^2}{[\lambda_2(k)-\sigma_k^2]^2}, \ldots, \frac{\sigma_k^2}{[\lambda_D(k)-\sigma_k^2]^2}\right\}\\ &= \tilde{\boldsymbol{\Lambda}}^{-1}(k) + \sigma_k^2\tilde{\boldsymbol{\Lambda}}^{-2}(k). \end{aligned} \quad (98)$$

Substitute (98) and $\mathbf{W}(k) = \mathbf{I}$ into (62), and apply (93)(97)

$$\begin{aligned}\mathbf{G}(k,\boldsymbol{\theta}) &= [\mathbf{A}^2(k)]^{-1}\mathbf{S}^{-1}(k)[\mathbf{A}^2(k)]^{-1}\\ &\quad+ \sigma_k^2[\mathbf{A}^2(k)]^{-1}\mathbf{S}^{-1}(k)[\mathbf{A}^2(k)]^{-1}\mathbf{S}^{-H}(k)[\mathbf{A}^2(k)]^{-1}\\ &= [\mathbf{A}^2(k)\mathbf{S}(k)\mathbf{A}^2(k)]^{-1}\\ &\quad+ \sigma_k^2[\mathbf{A}^2(k)\mathbf{S}^H(k)\mathbf{A}^2(k)\mathbf{S}(k)\mathbf{A}^2(k)]^{-1}. \end{aligned} \quad (99)$$

where $\mathbf{A}^2(k) \triangleq \mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})$. Substitute (96) into (210)

$$\begin{aligned}\mathbf{E}(k,\boldsymbol{\theta}) &= \mathbf{A}^+(k)\mathbf{U}_s(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k)\\ &= \mathbf{A}^+(k)\mathbf{A}^{+H}(k)\\ &= [\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}. \end{aligned} \quad (100)$$

From Theorem 1, the normalized estimation error, $\sqrt{J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$, is asymptotically normal distributed with zero mean and covariance matrix $\boldsymbol{\Sigma}$, which is given by:

$$\boldsymbol{\Sigma} = (\bar{\mathbf{C}}'')^{-1}\mathbf{Q}(\bar{\mathbf{C}}'')^{-1}; \quad (101)$$

$$\bar{\mathbf{C}}'' = -2\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{F}(k,\boldsymbol{\theta}) \odot [\mathbf{1}_{3+LN} \otimes \mathbf{E}^T(k)]\right\}; \quad (102)$$

$$\mathbf{Q} = 2\sum_{k=1}^{K}\sigma_k^2\mathrm{Re}\{\mathbf{F}(k,\boldsymbol{\theta}) \odot [\mathbf{1}_{3+LN} \otimes \mathbf{G}^T(k)]\}. \quad (103)$$

where $\mathbf{G}(k,\boldsymbol{\theta})$ is defined in (99), $\mathbf{E}(k,\boldsymbol{\theta})$ is defined in (100), and $\mathbf{F}$ is defined in (59) □

C. ASYMPTOTIC DISTRIBUTION OF DML

From another point of view, DML can be regarded as a WSSF method from Lemma 1 in Appendix A .

Follow from Theorem 1, substitute the weight matrix of a DML estimator $\mathbf{W}(k) = \tilde{\boldsymbol{\Lambda}}(k)$ into Theorem 1, and get the asymptotic distribution of a DML estimator from Theorem 3.

Theorem 3: The ML ($\mathbf{W}(k) = \tilde{\boldsymbol{\Lambda}}(k)$) *estimation errors* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}$ *are asymptotically (for large J) joinly Gaussian distributed with zero means and the asymptotically covariances are given in* (56),(57),(58) *with*

$$\mathbf{G}(k,\boldsymbol{\theta}) = \mathbf{S}(k) + \sigma_k^2[\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}, \quad (104)$$

$$\mathbf{E}(k,\boldsymbol{\theta}) = \mathbf{S}(k). \quad (105)$$

Proof: Substitute $\mathbf{W}(k) = \tilde{\boldsymbol{\Lambda}}(k)$ into (62)

$$\begin{aligned}\mathbf{G}(k,\boldsymbol{\theta}) &= \mathbf{A}^+(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}(k)\boldsymbol{\Lambda}_s(k)\tilde{\boldsymbol{\Lambda}}^{-2}(k)\tilde{\boldsymbol{\Lambda}}(k)\mathbf{U}_s(k)^H\mathbf{A}^{+H}(k)\\ &= \mathbf{A}^+(k)\mathbf{U}_s(k)\boldsymbol{\Lambda}_s(k)\mathbf{U}_s(k)^H\mathbf{A}^{+H}(k)\\ &= \mathbf{A}^+(k)\mathbf{U}_s(k)[\tilde{\boldsymbol{\Lambda}}(k) + \sigma_k^2\mathbf{I}]\mathbf{U}_s(k)\mathbf{A}^{+H}(k)\\ &= \mathbf{A}^+(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}(k)\mathbf{U}_s(k)\mathbf{A}^{+H}(k)\\ &\quad+ \sigma_k^2\mathbf{A}^+(k)\mathbf{U}_s(k)\mathbf{U}_s(k)\mathbf{A}^{+H}(k)\\ &= \mathbf{A}^+(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}(k)\mathbf{U}_s(k)\mathbf{A}^{+H}(k)\\ &\quad+ \sigma_k^2[\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}. \end{aligned} \quad (106)$$

From (93), we get

$$\begin{aligned}&\mathbf{A}^H(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^{-1}(k)\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta}) = \mathbf{S}^{-1}(k)\\ &\tilde{\boldsymbol{\Lambda}}^{-1}(k) = [\mathbf{A}^H(k)\mathbf{U}_s(k)]^{-1}\mathbf{S}^{-1}(k)[\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\\ &\tilde{\boldsymbol{\Lambda}}(k) = \mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})\mathbf{S}(k)\mathbf{A}^H(k)\mathbf{U}_s(k). \end{aligned} \quad (107)$$

Substitute (107) into the first item of the right part in (106)

$$\begin{aligned}&\mathbf{A}^{+}(k)\mathbf{U}_s(k)\tilde{\mathbf{\Lambda}}(k)\mathbf{U}_s(k)\mathbf{A}^{+H}(k)\\&\quad=\mathbf{A}^{+}(k)\mathbf{U}_s(k)\mathbf{U}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})\mathbf{S}(k)\\&\mathbf{A}^H(k)\mathbf{U}_s(k)\mathbf{U}_s(k)\mathbf{A}^{+H}(k)\\&\quad=\mathbf{S}(k).\end{aligned}\tag{108}$$

Substitute (108) into (106)

$$\mathbf{G}(k,\boldsymbol{\theta})=\mathbf{S}(k)+\sigma_k^2[\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}.\tag{109}$$

Substitute $\mathbf{W}(k)=\tilde{\mathbf{\Lambda}}(k)$ into (210)

$$\mathbf{E}(k,\boldsymbol{\theta})\triangleq\mathbf{A}^{+}(k)\mathbf{U}_s(k)\tilde{\mathbf{\Lambda}}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k)=\mathbf{S}(k).\tag{110}$$

□

D. ASYMPTOTIC DISTRIBUTION OF 1D-NSF

Theorem 4: The asymptotically covariance of the estimation errors of a 1D-NSF cost function is the same as that of a 1D-SSF cost function.

Proof: The cost function of 1D-NSF is

$$\begin{aligned}\hat{\boldsymbol{\theta}}&=\arg\min\bar{u}_1(\boldsymbol{\theta})\\&=\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\right\}\\&=\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\left[\mathbf{I}-\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\right]\right\}\\&=\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\right\}-\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\right\}.\end{aligned}\tag{111}$$

Since

$$\begin{aligned}&\mathrm{tr}\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\}\\&\quad=\mathrm{tr}\left\{\mathbf{a}(k,\boldsymbol{\theta})[\mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{a}(k,\boldsymbol{\theta})]^{-1}\mathbf{a}^H(k,\boldsymbol{\theta})\right\}\\&\quad=\mathrm{tr}\left\{\mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{a}(k,\boldsymbol{\theta})[\mathbf{a}^H(k,\boldsymbol{\theta})\mathbf{a}(k,\boldsymbol{\theta})]^{-1}\right\}\\&\quad=1.\end{aligned}\tag{112}$$

Move the constant term in the (111)

$$\hat{\boldsymbol{\theta}}=\arg\max\tilde{u}_1(\boldsymbol{\theta})=\sum_{k=1}^{K}\mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\right\}.\tag{113}$$

It is the same as the programming (38) of 1D-SSF. □

E. ASYMPTOTIC DISTRIBUTION OF MD-WNSF

The cost function of an MD-WNSF is given by

$$\begin{aligned}\hat{\boldsymbol{\theta}}&=\arg\min\bar{u}_3(\boldsymbol{\theta})\\&=\sum_{k=1}^{K}\sum_{d=1}^{D}w_d\frac{\mathbf{a}_d^H(k,\boldsymbol{\theta})\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\mathbf{a}_d(k,\boldsymbol{\theta})}{\mathbf{a}_d^H(k,\boldsymbol{\theta})\mathbf{a}_d(k,\boldsymbol{\theta})},\\&=\sum_{k=1}^{K}\sum_{d=1}^{D}w_d\mathrm{tr}\left\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta},d)\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k)\right\},\\&=\sum_{k=1}^{K}\left\{\sum_{d=1}^{D}w_d\mathrm{tr}\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta},d)\}\right.\\&\quad\left.-\sum_{d=1}^{D}w_d\mathrm{tr}\left[\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta},d)\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\right]\right\},\end{aligned}\tag{114}$$

where $\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta},d)\triangleq\mathbf{a}_d^H(k,\boldsymbol{\theta})[\mathbf{a}_d^H(k,\boldsymbol{\theta})\mathbf{a}_d(k,\boldsymbol{\theta})]^{-1}\mathbf{a}_d^H(k,\boldsymbol{\theta})$ denotes the projection matrix of $\mathbf{a}_d(k,\boldsymbol{\theta})$, and

$$\begin{aligned}&\mathrm{tr}\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta},d)\}\\&\quad=\mathrm{tr}\left\{\mathbf{a}_d(k,\boldsymbol{\theta})[\mathbf{a}_d^H(k,\boldsymbol{\theta})\mathbf{a}_d(k,\boldsymbol{\theta})]^{-1}\mathbf{a}_d^H(k,\boldsymbol{\theta})\right\}\\&\quad=\mathrm{tr}\left\{\mathbf{a}_d^H(k,\boldsymbol{\theta})\mathbf{a}_d(k,\boldsymbol{\theta})[\mathbf{a}_d^H(k,\boldsymbol{\theta})\mathbf{a}_d(k,\boldsymbol{\theta})]^{-1}\right\}\\&\quad=1.\end{aligned}\tag{115}$$

$\sum_{d=1}^{D}w_d=1$, substitute (115) into (114), and move the constant items in the (114):

$$\begin{aligned}\hat{\boldsymbol{\theta}}&=\arg\max u_3(\boldsymbol{\theta})\\&=\sum_{k=1}^{K}\left\{\sum_{d=1}^{D}w_d\mathrm{tr}\left[\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta},d)\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\right]\right\}.\end{aligned}\tag{116}$$

The cost function of 1D-SSF is given by

$$\hat{\boldsymbol{\theta}}=\arg\max s_4(\boldsymbol{\theta})=\sum_{k=1}^{K}\mathrm{tr}\{\mathbf{P}_{\mathbf{a}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)\}.\tag{117}$$

Replace $\mathbf{A}(k,\boldsymbol{\theta})$ by $\mathbf{a}(k,\boldsymbol{\theta})$ and set $\mathbf{W}(k)=\mathbf{I}_D$ in Theorem 1, we will get the asymptotic distribution of a 1D-SSF cost function. From (116), the asymptotic distribution of an MD-NWNSF can be derived from the conclusion of 1D-SSF.

Theorem 5: The asymptotically covariance of the estimation errors of cost function

$$\hat{\boldsymbol{\theta}}_d=\arg\max z_d=C_d(\boldsymbol{\theta}_d)\tag{118}$$

is defined by

$$\mathbf{\Sigma}_d(\boldsymbol{\theta}_d^*)=\mathbf{C}_d''(\boldsymbol{\theta}_d^*)^{-1}\mathbf{Q}_d(\boldsymbol{\theta}_d^*)\mathbf{C}_d''(\boldsymbol{\theta}_d^*)^{-1}.\tag{119}$$

where $\mathbf{C}_d'(\boldsymbol{\theta}_d^*)$ and $\mathbf{C}_d''(\boldsymbol{\theta}_d^*)$ is the first and second derivative the cost function $C_d(\boldsymbol{\theta}_d^*)$ around the true $\boldsymbol{\theta}_d^*$. $\mathbf{Q}_d(\boldsymbol{\theta}_d^*)$ is the covariance matrix of the random vector $\sqrt{J}\mathbf{C}_d'(\boldsymbol{\theta}_d^*)$, and $\mathbf{C}_{d_i}'(\boldsymbol{\theta}_d^*)$ is independent with $\mathbf{C}_{d_j}'(\boldsymbol{\theta}_d^*)$ if $i\neq j$. The asymptotically covariance of the estimation errors of cost function

$$\hat{\boldsymbol{\theta}}=\arg\max V=\sum_{i=1}^{D}w_dC_d(\boldsymbol{\theta}_d)\tag{120}$$

is given by

$$\mathbf{\Sigma}_v(\boldsymbol{\theta}^*)=\mathbf{V}(\boldsymbol{\theta}^*)^{-1}\bar{\mathbf{Q}}(\boldsymbol{\theta}^*)\mathbf{V}(\boldsymbol{\theta}^*)^{-1},\tag{121}$$

where

$$\begin{aligned}\boldsymbol{\theta}&=[\boldsymbol{\theta}_1^T,\boldsymbol{\theta}_2^T,\ldots,\boldsymbol{\theta}_D^T]^T,\\\mathbf{V}(\boldsymbol{\theta}^*)&\triangleq\sum_{d=1}^{D}w_d\mathbf{V}_d''(\boldsymbol{\theta}_d^*),\end{aligned}\tag{122}$$

$$\bar{\mathbf{Q}}(\boldsymbol{\theta}^*)\triangleq\sum_{d=1}^{D}w_d\bar{\mathbf{Q}}_d(\boldsymbol{\theta}_d^*),\tag{123}$$

where the element at row $d+Di$ and column $d+Dj$ of matrix $\mathbf{V}''_d(\boldsymbol{\theta}^*_d)$ is the element at row i and column j of matrix $\mathbf{C}''_d(\boldsymbol{\theta}^*_d)$, and the other elements are 0. The element at row $d+Di$ and column $d+Dj$ of matrix $\bar{\mathbf{Q}}''_d(\boldsymbol{\theta}^*_d)$ is the element at row i and column j of matrix $\mathbf{Q}''_d(\boldsymbol{\theta}^*_d)$, and the other elements are 0.

Proof: The first order derivative of V is

$$\mathbf{V}'(\boldsymbol{\theta}^*) = \sum_{i=1}^{D} w_d \mathbf{V}'_d(\boldsymbol{\theta}^*), \tag{124}$$

where the element at row $d+Di$ of $\mathbf{V}'_d(\boldsymbol{\theta}^*)$ is the element at row i of $\mathbf{C}'_d(\boldsymbol{\theta}^*_d)$, and the other elements of $\mathbf{V}'_d(\boldsymbol{\theta}^*)$ are 0;

The second derivative of V is

$$\mathbf{V}''(\boldsymbol{\theta}) = \sum_{i=1}^{D} w_d \mathbf{V}''_d(\boldsymbol{\theta}^*), \tag{125}$$

where the element at row $d+Di$ and column $d+Dj$ of matrix $\mathbf{V}''_d(\boldsymbol{\theta}^*)$ is the element at row i and column j of matrix $\mathbf{C}''_d(\boldsymbol{\theta}^*_d)$, and the other elements are 0.

Since $\mathbf{C}'_{d_i}(\boldsymbol{\theta}_{d_i})$ is independent with $\mathbf{C}'_{d_j}(\boldsymbol{\theta}^*_{d_j})$ if $d_i \neq d_j$, the covariance matrix of the random vector $\sqrt{J}\mathbf{V}'(\boldsymbol{\theta}^*)$ is

$$\bar{\mathbf{Q}}(\boldsymbol{\theta}^*) = \sum_{i}^{D} w_d \mathbf{Q}_d(\boldsymbol{\theta}^*_d). \tag{126}$$

The asymptotically covariance of the estimation errors of cost function $\hat{\boldsymbol{\theta}} = \arg\max V = \sum_{i=1}^{D} w_d C_d(\boldsymbol{\theta})$ is

$$\boldsymbol{\Sigma}_v(\boldsymbol{\theta}^*) = \mathbf{V}(\boldsymbol{\theta}^*)^{-1}\bar{\mathbf{Q}}(\boldsymbol{\theta}^*)\mathbf{V}(\boldsymbol{\theta}^*)^{-1}, \tag{127}$$

where $\mathbf{V}(\boldsymbol{\theta}^*) \triangleq \sum_{d=1}^{D} w_d \mathbf{C}''_d(\boldsymbol{\theta}^*_d)$, $\bar{\mathbf{Q}}(\boldsymbol{\theta}^*) \triangleq \sum_{d=1}^{D} w_d \mathbf{Q}_d(\boldsymbol{\theta}^*_d)$. □

The asymptotic distribution of MD-WSSF has been studies in Theorem 1, and that of a 1D-SSF can be derived easily from the Theorem. Substitute items in the asymptotic distribution analysis of a 1D-SSF cost function $\mathbf{C}''_d(\boldsymbol{\theta}^*_d)^{-1}$ and $\mathbf{Q}_d(\boldsymbol{\theta}^*_d)$ into Theorem 5, we will get the asymptotic distribution of an MD-WNSF cost function.

F. CRLB ANALYSIS

We studied the asymptotic distribution of MD-WSSF, MD-SSF, DML, and MD-WNSF in Theorem 1, Theorem 2, Theorem 3 and Theorem 5. The CRLB of the parameters estimation is given in this section.

Theorem 6: The signal model is defined as (8)*. The CRLB of the emitter position parameters and path attenuations is given by*

$$\mathrm{CRLB}(\boldsymbol{\theta}) = \frac{\sigma^2}{2}\left\{\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{F}(k,\boldsymbol{\theta}) \odot \left\{\mathbf{1}_{3+LN} \otimes [\mathbf{s}(k)\mathbf{s}^H(k)]\right\}\right\}\right\}^{-1}, \tag{128}$$

where $\mathbf{1}_{3+LN}$ is a $(3+LN)\times(3+LN)$ matrix with all ones, and

$$\mathbf{F}(k,\boldsymbol{\theta}) = \mathbf{A}^H_{\boldsymbol{\theta}}(k)\mathbf{P}^{\perp}_{\mathbf{A}}(k)\mathbf{A}_{\boldsymbol{\theta}}(k), \tag{129}$$

where $\mathbf{A}_{\boldsymbol{\theta}}(k)$ was defined in (60)*.*

Proof: Stoica and Nehorai [34] proposed the CRLB result in the application of direction finding, and Du *et al.* [1] analyzed the CRLB of multiple emitters positioning in the presence of multi-path propagation. We study the CRLB of path attenuations and the position estimations in this section. The unknown parameter vector for a multiple unknown waveform signals positioning application is denoted as $\bar{\boldsymbol{\theta}} = [\bar{\mathbf{s}}(1), \tilde{\mathbf{s}}(1), \bar{\mathbf{s}}(2), \tilde{\mathbf{s}}(2), \ldots, \bar{\mathbf{s}}(K), \tilde{\mathbf{s}}(K), \boldsymbol{\theta}]^T$, where $\bar{\mathbf{s}}(k) = [\bar{s}_1(k), \bar{s}_2(k), \ldots, \bar{s}_D(k)]$ are real parts of D signals at frequency k and $\tilde{\mathbf{s}}(k)$ are imaginary parts, $\boldsymbol{\theta}$ is defined in (53).

Follow from the Appendix C, the Fisher Information Matrix of the observed data is

$$\mathbf{I}_{\bar{\theta}} = \left[\begin{array}{cccc:c} \mathbf{G}(1) & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}(1) \\ \mathbf{0} & \mathbf{G}(2) & \cdots & \mathbf{0} & \mathbf{V}(2) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{G}(K) & \mathbf{V}(K) \\ \hdashline \mathbf{V}^H(1) & \mathbf{V}^H(2) & \cdots & \mathbf{V}^H(K) & \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{array}\right] \tag{130}$$

$$\triangleq \begin{bmatrix} \mathbf{G} & \mathbf{V} \\ \mathbf{V}^H & \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{bmatrix}, \tag{131}$$

where $\mathbf{G}(k,\boldsymbol{\theta}) \triangleq \begin{bmatrix} \mathbf{I}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}(k) & \mathbf{I}_{\bar{\mathbf{s}}\tilde{\mathbf{s}}}(k) \\ \mathbf{I}^H_{\bar{\mathbf{s}}\tilde{\mathbf{s}}}(k) & \mathbf{I}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(k) \end{bmatrix}$, $\mathbf{V}(k) = \begin{bmatrix} \mathbf{I}_{\bar{\mathbf{s}}\boldsymbol{\theta}}(k) \\ \mathbf{I}_{\tilde{\mathbf{s}}\boldsymbol{\theta}}(k) \end{bmatrix}$. The CRLB of $\boldsymbol{\theta}$ is

$$\mathrm{CRLB}^{-1}(\boldsymbol{\theta}) = \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathbf{V}^H\mathbf{G}^{-1}\mathbf{V}, \tag{132}$$

Since $\mathbf{G}$ is a block diagonal matrix,

$$\mathbf{V}^H\mathbf{G}^{-1}\mathbf{V} = \sum_{k=1}^{K}\mathbf{V}^H(k)\mathbf{G}^{-1}(k)\mathbf{V}(k). \tag{133}$$

The second derivative of Log-likelihood function were derived in Appendix C

$$\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} = \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{S}^H(k)\mathbf{A}^H_{\boldsymbol{\theta}}(k)\mathbf{A}_{\boldsymbol{\theta}}(k)\mathbf{S}(k)\right\}; \tag{134}$$

$$\mathbf{S}(k) = \mathbf{I}_{3+LN} \otimes \mathbf{S}(k); \tag{135}$$

$$\mathbf{S}(k) = \mathrm{diag}\{\mathbf{s}(k)\}; \tag{136}$$

$$\mathbf{V}(k) = \begin{bmatrix} \frac{2}{\sigma^2}\mathrm{Re}\left\{\mathbf{A}^H(k)\mathbf{A}_{\boldsymbol{\theta}}(k)\mathbf{S}(k)\right\} \\ -\frac{2}{\sigma^2}\mathrm{Im}\left\{\mathbf{A}^H(k)\mathbf{A}_{\boldsymbol{\theta}}(k)\mathbf{S}(k)\right\} \end{bmatrix}; \tag{137}$$

$$\mathbf{G}(k,\boldsymbol{\theta}) = \begin{bmatrix} \mathrm{Re}\{\mathbf{H}(k)\} & -\mathrm{Im}\{\mathbf{H}(k)\} \\ \mathrm{Im}\{\mathbf{H}(k)\} & \mathrm{Re}\{\mathbf{H}(k)\} \end{bmatrix}; \tag{138}$$

$$\mathbf{H}(k) = \frac{2}{\sigma^2}\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta}). \tag{139}$$

where $\mathbf{I}_{3+LN}$ is a $(3+LN)\times(3+LN)$ identify matrix. Since

$$\mathbf{G}^{-1}(k) = \begin{bmatrix} \mathrm{Re}\{\mathbf{H}^{-1}(k)\} & \mathrm{Im}\{\mathbf{H}^{-1}(k)\} \\ -\mathrm{Im}\{\mathbf{H}^{-1}(k)\} & \mathrm{Re}\{\mathbf{H}^{-1}(k)\} \end{bmatrix}, \tag{140}$$

$$\mathbf{V}^H(k)\mathbf{G}^{-1}(k)\mathbf{V}(k) = \frac{2}{\sigma^2}\mathrm{Re}\left\{\mathbf{S}^H(k)\mathbf{A}^H_{\boldsymbol{\theta}}(k)\mathbf{A}(k,\boldsymbol{\theta})[\mathbf{A}^2(k)]^{-1}\mathbf{A}^H(k)\mathbf{A}_{\boldsymbol{\theta}}(k)\mathbf{S}(k)\right\}. \tag{141}$$

CRLB of $\boldsymbol{\theta}$ is

$$\begin{aligned}\mathrm{CRLB}^{-1}(\boldsymbol{\theta}) &= \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} - \mathbf{V}^H\mathbf{G}^{-1}\mathbf{V},\\ &= \frac{2}{\sigma^2}\mathrm{Re}\left\{\mathbf{S}^H(k)\mathbf{A}_{\boldsymbol{\theta}}^H(k)\mathbf{P}_{\mathbf{A}}^{\perp}(k)\mathbf{A}_{\boldsymbol{\theta}}(k)\mathbf{S}(k)\right\}\\ &= \frac{2}{\sigma^2}\mathrm{Re}\left\{[\mathbf{A}_{\boldsymbol{\theta}}^H(k)\mathbf{P}_{\mathbf{A}}^{\perp}(k)\mathbf{A}_{\boldsymbol{\theta}}(k)]\right.\\ &\quad\left.\odot\left\{\mathbf{1}_{3+LN}\otimes[\mathbf{s}(k)\mathbf{s}^H(k)]\right\}\right\}. \qquad (142)\end{aligned}$$

where

$$\mathbf{P}_{\mathbf{A}}^{\perp}(k) = \mathbf{I} - \mathbf{P}_{\mathbf{A}}(k), \qquad (143)$$

$$\mathbf{P}_{\mathbf{A}}(k) = \mathbf{A}(k,\boldsymbol{\theta})[\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\mathbf{A}^H(k), \qquad (144)$$

$$\mathbf{F}(k,\boldsymbol{\theta}) \triangleq \mathbf{A}_{\boldsymbol{\theta}}^H(k)\mathbf{P}_{\mathbf{A}}^{\perp}(k)\mathbf{A}_{\boldsymbol{\theta}}(k) \qquad (145)$$

and $\mathbf{1}_{3+LN}$ is a $(3+LN)\times(3+LN)$ matrix with all ones. Denote,

$$\mathrm{CRLB}^{-1}(\boldsymbol{\theta}) = \frac{2}{\sigma^2}\sum_{k=1}^{K}\left\{\mathrm{Re}\left\{\mathbf{F}(k)\odot\left\{\mathbf{1}_{3+LN}\otimes[\mathbf{s}(k)\mathbf{s}^H(k)]\right\}\right\}\right\}. \qquad (146)$$

□

We discussed the asymptotic distribution of MD-WSSF, MD-NSF, MD-SSF and DML in this section, and the CRLB of the unknown parameters was studied in addition.

VI. THE OPTIMAL WEIGHTS OF THE MD-WSSF

The asymptotic distribution properties were given in the above Theorems. The DML and MD-SSF are regarded as special cases in the WSF framework. The optimal weights of a WSSF framework are optimized in this section.

Theorem 7: For all Hermitian matrices $\mathbf{W}(k)$

$$\boldsymbol{\Sigma}[\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)] \le \boldsymbol{\Sigma}(\mathbf{W}(k)), \qquad (147)$$

where $\boldsymbol{\Sigma}$ *is defined in* (56)

Proof: Set $\mathbf{W}(k) = \tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)$, and substitute it into Theorem 1:

$$\mathbf{E}(k,\boldsymbol{\theta}) = \mathbf{A}^+(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k), \qquad (148)$$

$$\begin{aligned}\mathbf{G}(k,\boldsymbol{\theta}) &= \mathbf{A}^+(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)\mathbf{U}_s(k)^H\mathbf{A}^{+H}(k) \qquad (149)\\ &= \mathbf{E}(k,\boldsymbol{\theta}),\end{aligned}$$

$$\boldsymbol{\Sigma}[\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)] = (\bar{\mathbf{C}}'')^{-1}\mathbf{Q}(\bar{\mathbf{C}}'')^{-1}, \qquad (150)$$

(151)

where

$$\bar{\mathbf{C}}'' = -2\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{F}(k,\boldsymbol{\theta})\odot[\mathbf{1}_{3+LN}\otimes\mathbf{E}^T(k)]\right\}, \qquad (152)$$

$$\mathbf{Q} = 2\sum_{k=1}^{K}\sigma_k^2\mathrm{Re}\{\mathbf{F}(k,\boldsymbol{\theta})\odot[\mathbf{1}_{3+LN}\otimes\mathbf{E}^T(k)]\}. \qquad (153)$$

Since the noise is AWGN, denote $\sigma_k = \sigma, k = 1, 2, \ldots, K$,

$$\boldsymbol{\Sigma}[\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)] = \frac{\sigma^2}{2}\left\{\sum_{k=1}^{K}\mathrm{Re}\{\mathbf{F}(k,\boldsymbol{\theta})\odot\{\mathbf{1}_{3+LN}\otimes\mathbf{H}(k)\}\}\right\}^{-1}, \qquad (154)$$

where $\mathbf{H}(k) \triangleq [\mathbf{A}^+(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k)]$. Follow from Lemma 5

$$\boldsymbol{\Sigma}(\mathbf{W}) \ge \frac{\sigma^2}{2}\left\{\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{F}(k,\boldsymbol{\theta})\odot\{\mathbf{1}_{3+LN}\otimes[\mathbf{E}(k)\mathbf{G}^{-1}(k)\mathbf{E}(k)]\}\right\}\right\}^{-1}, \qquad (155)$$

where

$$\begin{aligned}&\mathbf{E}(k,\boldsymbol{\theta})\mathbf{G}^{-1}(k)\mathbf{E}(k,\boldsymbol{\theta})\\ &= \mathbf{A}^+(k)\mathbf{U}_s(k)\mathbf{W}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k)\\ &\quad\times[\mathbf{A}^+(k)\mathbf{U}_s(k)\mathbf{W}(k)\boldsymbol{\Lambda}_s(k)\tilde{\boldsymbol{\Lambda}}^{-2}(k)\mathbf{W}(k)\mathbf{U}_s(k)^H\mathbf{A}^{+H}(k)]^{-1}\\ &\quad\times\mathbf{A}^+(k)\mathbf{U}_s(k)\mathbf{W}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k)\\ &= \mathbf{A}^+(k)\mathbf{U}_s(k)\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k). \qquad (156)\end{aligned}$$

Substitute (156) into (155)

$$\begin{aligned}\boldsymbol{\Sigma}(\mathbf{W}) &\ge \frac{\sigma^2}{2}\left\{\sum_{k=1}^{K}\mathrm{Re}\{\mathbf{F}(k,\boldsymbol{\theta})\odot\{\mathbf{1}_{3+LN}\otimes[\mathbf{H}(k)]\}\}\right\}^{-1}\\ &= \boldsymbol{\Sigma}[\tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)]. \qquad (157)\end{aligned}$$

□

We can get the optimal weights of an MD-WSSF cost function from Theorem 7. Substitute the optimal weights $\mathbf{W}_k = \tilde{\boldsymbol{\Lambda}}^2(k)\boldsymbol{\Lambda}_s^{-1}(k)$ into the cost function of a MD-WSSF estimator, and get the best asymptotic distribution performance in the WSSF framework.

VII. NUMERICAL SIMULATIONS

Some numerical examples are presented in this section to compare the performances of the discussed cost functions. The simulation scenarios in this paper follow from [1]. Three emitters are located at [0, 0, 0], [100, 0, 0] and [0, 100, 0], and four receiving arrays are located at [2200, −2100, 0], [3300, 600, 0], [3100, −700, 0], and [2300, 2500, 0]. Four transponders are located at [−1210, 100, 200],[100, 1120, 200], (−200, −1040, 200) and [970, 160, 200] to transmit the signals. The layout of elements in the scenario are given in FIGURE 7 (All are measured in km).

Each receiving array is a Uniform Circular Array (UCA) with eight antennas and radius of 30 m. The band width is 8 kHz. The carrier frequency is 10 MHz. The complex-valued signal frequency coefficients subject to $\|\mathbf{s}_d\|_F^2 = 1$. The path attenuation coefficients are drawn from a uniform

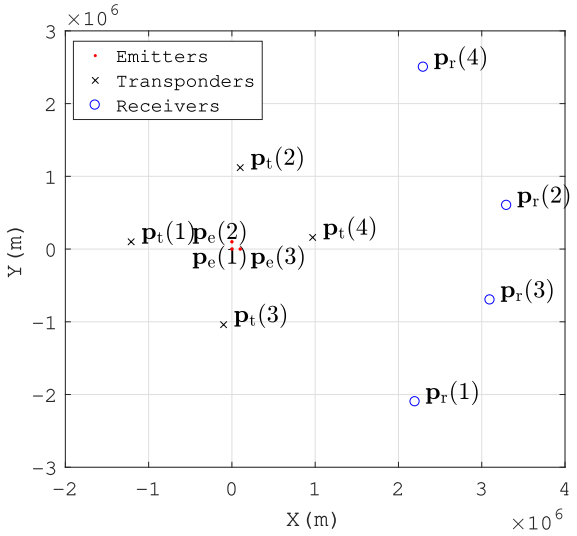


FIGURE 7. Scenario Layout.

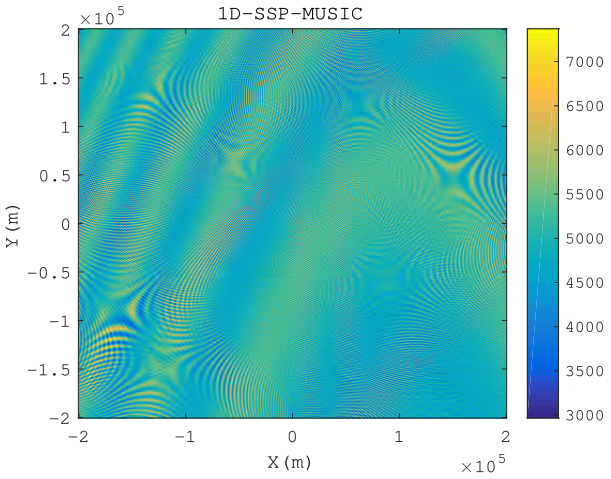


FIGURE 8. Spatial spectrum of 1D-SSP-MUSIC (SNR = 10dB).

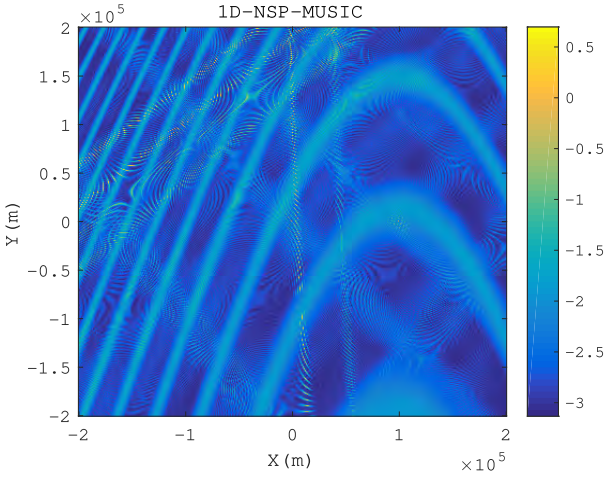


FIGURE 9. Spatial spectrum of 1D-NSP-MUSIC (SNR = 10dB).

distribution between 0 and 1. The SNR is defined in terms of "Post Processing SNR", which is given by

$$\mathrm{SNR} \triangleq \frac{E\left\{\sum_{k=1}^{K} \|\mathbf{A}(k)\mathbf{s}(k)\|_F^2\right\}}{K\sigma^2}. \tag{158}$$

Root-Mean-Squared Error (RMSE) of the estimated position is given by

$$\mathrm{RMSE}(\mathbf{p}_e) \triangleq \sqrt{\frac{\sum_{i=1}^{N_s} \|\hat{\mathbf{p}}_e - \mathbf{p}_e\|_F^2}{N_s D}}, \tag{159}$$

where N_s is the number of Monte-Carlo runs (N_s is set as 200 in this section), D is the number of emitters, $\mathbf{p}_e$ are the real emitter positions and $\hat{\mathbf{p}}_e$ are the estimated emitter positions.

An scalar quantity of the error covariance matrix corresponding to RMSE is defined as

$$\Sigma(\mathbf{p}_e) \triangleq \sqrt{\frac{\mathrm{tr}\{\mathbf{\Sigma}(\mathbf{p}_e)\}}{D}}. \tag{160}$$

A. SPATIAL SPECTRUM OF 1D-NSF AND 1D-MUSIC

We compare the spatial spectrum of a 1D-NSF cost function and that of a 1D-MUSIC. The 1D-MUSIC cost function was proposed by Weiss [2], and it was widely used in the existing DPD methods. The 1D-MUSIC method maximizes the manifold projection length onto the signal subspace (1D-SSP-MUSIC) or minimizes the manifold projection length onto the noise subspace (1D-NSP-MUSIC) without normalizations.

Because the cost function of a 1D-NSF and the cost function of a 1D-SSF are equivalent, only the spatial spectrum of the 1D-NSF cost function is given. Since it is a three-dimensional emitters position searching problem, the spatial spectrum should be a three-dimensional function. Only a slice of $z = 0$ in the three-dimensional spatial spectrum is given for a more intuitive comparison (See FIGURE 8, FIGURE 9, FIGURE 10, and FIGURE 11).

Since 1D-SSP-MUSIC and 1D-NSP-MUSIC do not normalize the length of the manifold vector, and they can not find emitters correctly. Three peaks can be clearly found in the spatial spectrum of the 1D-NSF method, and each of them corresponds to an emitter location.

B. STANDARD DEVIATIONS VERSUS SNR

We have got the theoretical asymptotic distribution performances of DPD methods in section 4. The asymptotic performances of them are compared in FIGURE 12.

The number of snapshots J is 1000 and the number of frequency K is 32 in the numerical simulation. From FIGURE 12, the Multiple Dimension-Optimal Weighted Signal Subspace Fitting (MD-OWSSF) cost function gets the best performance than other cost functions. The Maximum Likelihood (ML) cost function had a better performance than the Multiple Dimension-Signal Subspace Fitting (MD-SSF)

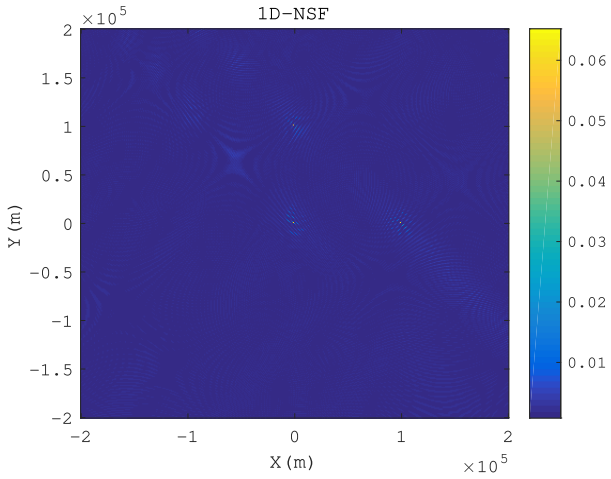


FIGURE 10. Spatial spectrum of 1D-NSP (SNR = 10dB).

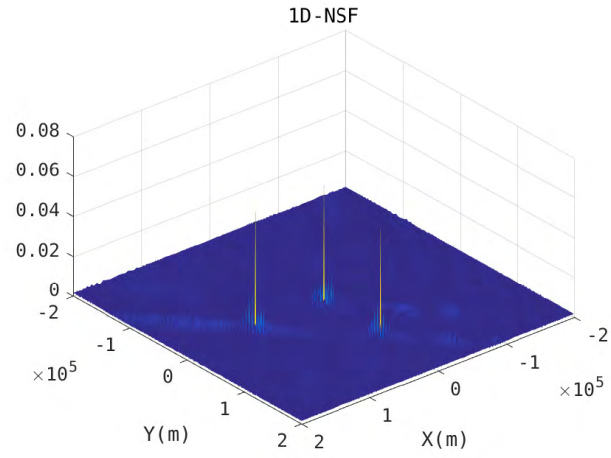


FIGURE 11. Spatial spectrum of 1D-NSP (SNR = 10dB, 3D view).

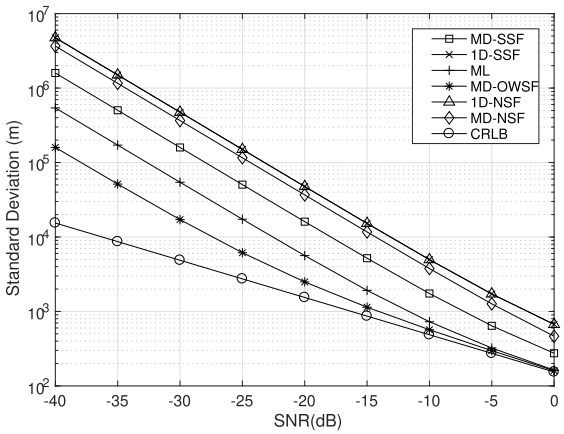


FIGURE 12. Standard deviations of DPD cost functions versus SNR (theoretical result).

cost function. The multiple dimension cost functions get better performances than one dimension cost functions in the numerical simulation.

C. STANDARD DEVIATIONS VERSUS NUMBER OF SNAPSHOTS

The asymptotic distribution of DPD methods are analyzed in this section. FIGURE 13 gives standard deviations of DPD cost functions versus the number of snapshots.

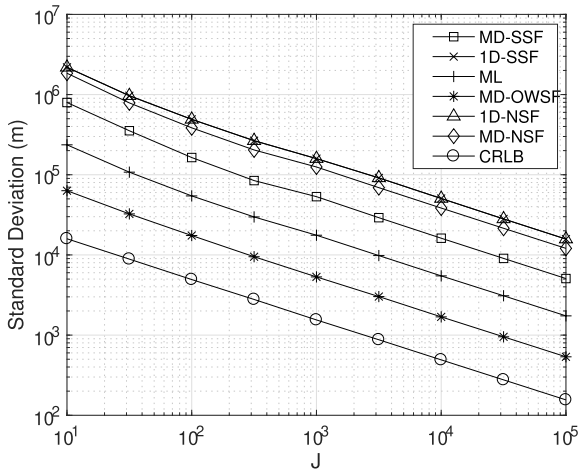


FIGURE 13. Standard deviations of DPD cost functions versus the number of snapshots (SNR = −30dB).

From FIGURE 13, the error standard deviations decrease with the increase of the number of snapshots. Because of the limitation of paper length, the theoretical comparison analysis of different cost functions is not discussed here. The numerical simulation results can only show the performance comparison of those methods in this scenario, and the simulation results can not apply to other scenarios. However, the theoretical asymptotic performances of those cost functions proposed in this paper can get the asymptotic performances for a given scenario.

D. PERFORMANCES AND THE NUMBER OF RECEIVING STATIONS

In the positioning system described in this paper, the more receiving stations, the higher positioning accuracy. FIGURE 14 shows the performance comparison of different cost functions when there is only one receiving stations (The receiving station is placed at [2200, −2100, 0]km).

When there is only one receiving station, the positioning performance decreases compared to that of the four receiving stations in FIGURE 12. The OWSSF cost function is less affected by the decrease of the number of receiving stations than other cost functions.

E. PERFORMANCES AND THE NUMBER OF EMITTERS

When transponders and receiving stations are fixed, the fewer emitter, the higher positioning accuracy. FIGURE 15 shows the performance comparison of different cost functions when there is only one emitter (The emitter is placed at [0, 0, 0]km).

When there is only one emitter, performances of all cost functions are improved, and the performances difference of

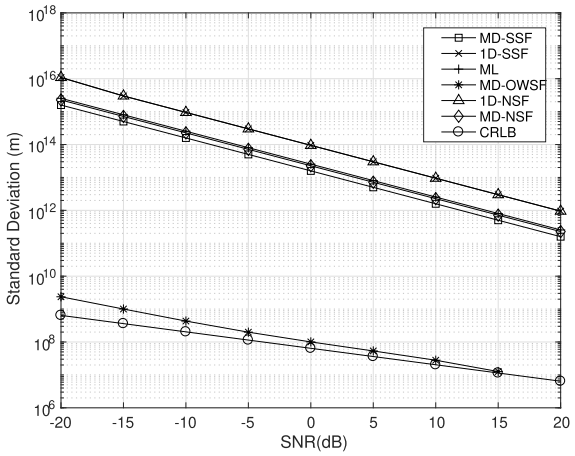


FIGURE 14. Performances of one receiving station.

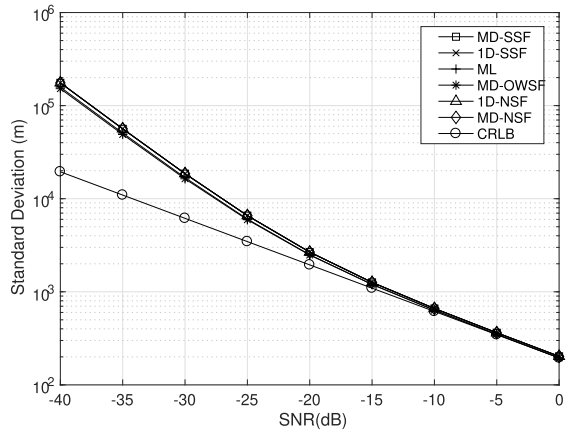


FIGURE 15. Performances of one emitter.

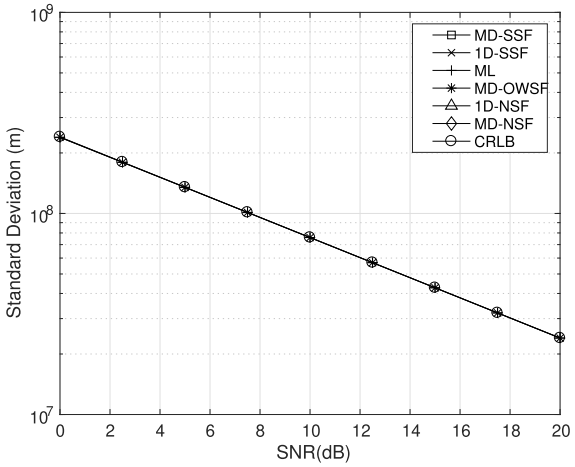


FIGURE 16. Performances of the minimum system.

those cost functions are not significant. It is because that an MD-SSF method degenerates to a 1D-SSF method when there is only one emitter.

F. PERFORMANCES OF THE MINIMUM SYSTEM

The minimum system consists of one emitter, three transponders and one receiving station. FIGURE 16 shows the positioning performance of different cost functions for the minimum system (Three transponders are placed at $[-1210, 100, 200]$, $[200, 1120, 200]$, $[-200, -1040, 200]$, one emitter is placed at $[0, 0, 0]$, and one receiving station is placed at $[2200, -2100, 0]$(km)).

As can be seen from FIGURE 16, all cost functions have the same positioning performance at the minimum system, and all are close to CRLB.

VIII. CONCLUSIONS

This paper attempts to collect the General Direct Position Determination (GDPD) cost functions in a unified framework. Some algebraic asymptotic distribution conclusions are presented in the unified framework. The main contributions of this paper are:

We analyzed the reason why the existing Direct Position Determination (DPD) cost function can not be used in a multipath propagation scenario, and proposed eight cost functions for DPD model with multipath propagation. The eight cost functions were collected in a Unified Subspace Fitting (USF) framework for the General Direct Position Determine (GDPD) model. Furthermore, the USF framework was divided into Signal Subspace Fitting (SSF) framework, and Noise Subspace Fitting (NSF) framework. MD-MUSIC and DML were studied as special cases of a Multiple Dimension-Weighted Signal Subspace Fitting (MD-WSSF) method.

A unified derivation of the asymptotic distribution properties of MD-WSSF cost functions for GDPD model was proposed. The second moment of the signal subspace eigenvectors were adopted to analysis the covariance matrix of the estimations which were optimized by an MD-WSSF cost function. Besides, the estimation covariance matrices of the Deterministic Maximum Likelihood (DML) and the Multiple Dimension-SSF (MD-SSF) were derived from the unified derivation process. The Cramér-Rao Lower Bound (CRLB) in presence of multi-path propagation was derived. It was assumed that the noise levels were known in advance, but the complex source signal waveforms and the path attenuations were unknown in the model. Some algebraic Lemmas were proposed and proved to assistant the derivation of the CRLB. We compared the asymptotic distribution performances of cost functions in the USF framework by numerical simulations. The numerical simulation results showed that the Multiple Dimension-Optimal Weighted Signal Subspace Fitting (MD-OWSSF) cost function holds the best asymptotic distribution performance in the USF framework.

We discussed the asymptotic distribution properties of different cost functions in this paper, but a theoretical analysis of performance comparisons has not been done yet. We would keep on investigating in the asymptotic distribution performance comparisons in the future work.

APPENDIX A
BASIC LEMMAS

We introduce some Lemmas firstly:

Lemma 1: The Deterministic ML method has the same asymptotic distribution as the following estimator:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\min \bar{m}(\boldsymbol{\theta}) = \sum_{k=1}^{K} \|\hat{\mathbf{U}}_s(k)\tilde{\Lambda}_k^{\frac{1}{2}} - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{T}(k)\|_F^2 \\ &= \arg\max m(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\tilde{\Lambda}(k)\hat{\mathbf{U}}_s^H(k)\},\end{aligned} \tag{161}$$

where $\tilde{\boldsymbol{\Lambda}}(k) = \boldsymbol{\Lambda}_s(k) - \sigma_k^2\mathbf{I}$.

Proof: The Deterministic ML criterion function is

$$\hat{\boldsymbol{\theta}} = \arg\max s_5(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{R}}(k)\}, \tag{162}$$

where

$$\hat{\mathbf{R}}(k) = \hat{\mathbf{U}}_s(k)\hat{\boldsymbol{\Lambda}}_s(k)\hat{\mathbf{U}}_s^H(k) + \hat{\mathbf{U}}_n(k)\hat{\boldsymbol{\Lambda}}_n(k)\hat{\mathbf{U}}_n^H(k). \tag{163}$$

Since $\hat{\boldsymbol{\Lambda}}_n(k)$ can be replaced by $\sigma_k^2\mathbf{I}$ without affecting the asymptotic properties [34]. (163) turns to be

$$\begin{aligned}\hat{\mathbf{R}}(k) &= \hat{\mathbf{U}}_s(k)\hat{\boldsymbol{\Lambda}}_s(k)\hat{\mathbf{U}}_s^H(k) + \sigma_k^2\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n^H(k) \\ &= \hat{\mathbf{U}}_s(k)\hat{\boldsymbol{\Lambda}}_s(k)\hat{\mathbf{U}}_s^H(k) + \sigma_k^2[\mathbf{I} - \hat{\mathbf{U}}_s(k)\hat{\mathbf{U}}_s^H(k)] \\ &= \sigma_k^2\mathbf{I} + \hat{\mathbf{U}}_s(k)[\hat{\boldsymbol{\Lambda}}_s(k) - \sigma_k^2\mathbf{I}]\hat{\mathbf{U}}_s^H(k).\end{aligned} \tag{164}$$

Substitute (164) into (162)

$$\begin{aligned}\hat{\boldsymbol{\theta}} = \arg\max s_5(\boldsymbol{\theta}) = \sum_{k=1}^{K} &\mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\sigma_k^2 \\ &+ \mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)[\hat{\boldsymbol{\Lambda}}_s(k) - \sigma_k^2\mathbf{I}]\hat{\mathbf{U}}_s^H(k)\},\end{aligned} \tag{165}$$

and notice that the first item in the cost function:

$$\begin{aligned}&\mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\sigma_k^2\} \\ &\quad= \sigma_k^2\mathrm{tr}\{\mathbf{A}(k,\boldsymbol{\theta})[\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\mathbf{A}^H(k,\boldsymbol{\theta})\} \\ &\quad= \sigma_k^2\mathrm{tr}\{\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})[\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\} \\ &\quad= \sigma_k^2 D,\end{aligned} \tag{166}$$

is a constant item, and remove it from the objective function

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= \arg\max m(\boldsymbol{\theta}) \\ &= \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)[\hat{\boldsymbol{\Lambda}}_s(k) - \sigma_k^2\mathbf{I}]\hat{\mathbf{U}}_s^H(k)\}.\end{aligned} \tag{167}$$

By [36, Lemma 5], replacement of the weighting matrix $\hat{\mathbf{W}}(k) = \hat{\boldsymbol{\Lambda}}_s(k) - \sigma_k^2\mathbf{I}$ by $\mathbf{W}(k) = \boldsymbol{\Lambda}_s(k) - \sigma_k^2\mathbf{I}$ does not change the asymptotic distribution of the estimate. □

Lemma 2: $\mathbf{A}$ *is a complex matrix with size of* $N \times N$, *and* $\mathbf{C}$ *is a Hermitian matrix with size of* $N \times N$, *then*

$$\mathrm{tr}[(\mathbf{A} + \mathbf{A}^H)\mathbf{C}] = 2\mathrm{Re}\{\mathrm{tr}[\mathbf{AC}]\}. \tag{168}$$

Proof: Denote $\mathbf{a}_n$ is the nth row of $\mathbf{A}$, $\mathbf{c}_n$ is the nth column of $\mathbf{C}$, and $\bar{\mathbf{c}}_n$ is the nth row of $\mathbf{C}$.

$$\mathbf{A} = \begin{bmatrix}\mathbf{a}_1 \\ \mathbf{a}_2 \\ \vdots \\ \mathbf{a}_N\end{bmatrix}, \tag{169}$$

$$\mathbf{C} = \begin{bmatrix}\mathbf{c}_1 & \mathbf{c}_2 & \cdots & \mathbf{c}_N\end{bmatrix} = \begin{bmatrix}\bar{\mathbf{c}}_1 \\ \bar{\mathbf{c}}_2 \\ \vdots \\ \bar{\mathbf{c}}_N\end{bmatrix}. \tag{170}$$

The left term of (168) is expanded as

$$\mathrm{tr}[(\mathbf{A} + \mathbf{A}^H)\mathbf{C}] = \mathrm{tr}[\mathbf{AC}] + \mathrm{tr}[\mathbf{A}^H\mathbf{C}], \tag{171}$$

where

$$\mathrm{tr}[\mathbf{AC}] = \sum_{n=1}^{N} \mathbf{a}_n\mathbf{c}_n, \tag{172}$$

$$\begin{aligned}\mathrm{tr}[\mathbf{A}^H\mathbf{C}] &= \mathrm{tr}[\mathbf{C}\mathbf{A}^H] \\ &= \sum_{n=1}^{N} (\bar{\mathbf{c}}_n\mathbf{a}_n^H).\end{aligned} \tag{173}$$

Since $\mathbf{C}$ is a Hermitian matrix, $\bar{\mathbf{c}}_n = \mathbf{c}_n^H$.

$$\mathrm{tr}[\mathbf{A}^H\mathbf{C}] = \sum_{n=1}^{N} (\bar{\mathbf{c}}_n\mathbf{a}_n^H) = \sum_{n=1}^{N} (\mathbf{c}_n^H\mathbf{a}_n^H) = \{\mathrm{tr}[\mathbf{AC}]\}^H. \tag{174}$$

Substitute (174) into (171)

$$\begin{aligned}\mathrm{tr}[\mathbf{AC}] + \mathrm{tr}[\mathbf{A}^H\mathbf{C}] &= \mathrm{tr}[\mathbf{AC}] + \{\mathrm{tr}[\mathbf{AC}]\}^H \\ &= 2\mathrm{Re}\{\mathrm{tr}[\mathbf{AC}]\}.\end{aligned} \tag{175}$$

□

Lemma 3: $\mathbf{A}$,$\mathbf{B}$,$\mathbf{C}$ *are real symmetric matrices. Denote*

$$\mathbf{U} = \begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{C}\end{bmatrix}, \tag{176}$$

Matrix $\mathbf{C} - (\mathbf{B}\mathbf{A}^{-1}\mathbf{B})$ *is a semi-definite matrix, if* $\mathbf{U}$ *is a semi-definite matrix,*

Proof: Denote

$$\mathbf{V} = \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}\end{bmatrix}. \tag{177}$$

Since

$$\begin{aligned}&\begin{bmatrix}\mathbf{A}^{-\frac{1}{2}} & \mathbf{0} \\ -\mathbf{B}\mathbf{A}^{-1} & \mathbf{I}\end{bmatrix}\begin{bmatrix}\mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{C}\end{bmatrix}\begin{bmatrix}\mathbf{A}^{-\frac{1}{2}} & \mathbf{0} \\ -\mathbf{B}\mathbf{A}^{-1} & \mathbf{I}\end{bmatrix}^T \\ &\qquad= \begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}\end{bmatrix},\end{aligned} \tag{178}$$

$\mathbf{U}$ and $\mathbf{V}$ are congruence. If $\mathbf{U}$ is a semi-definite matrix, $\mathbf{V}$ is a semi-definite matrix too. $\forall \mathbf{x} \in \mathbb{R}$

$$\begin{aligned}&\mathbf{x}(\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B})\mathbf{x}^T \\ &\qquad= \begin{bmatrix}\mathbf{0} & \mathbf{x}\end{bmatrix}\begin{bmatrix}\mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}\end{bmatrix}\begin{bmatrix}\mathbf{0} \\ \mathbf{x}^T\end{bmatrix}.\end{aligned} \tag{179}$$

Since $\mathbf{V}$ is a semi-definite matrix, $\mathbf{x}(\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B})\mathbf{x}^T \geq 0$, and $\mathbf{C} - \mathbf{B}\mathbf{A}^{-1}\mathbf{B}$ is a semi-definite matrix □

APPENDIX B
DERIVATIVES OF AN MD-WSSF COST FUNCTION

The cost function of an MD-WSSF method is defined as

$$\hat{\boldsymbol{\theta}} = \arg\max C_7(\boldsymbol{\theta}) = \sum_{k=1}^{K} \mathrm{tr}\{\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\mathbf{W}(k)\hat{\mathbf{U}}_s^H(k)\}. \tag{180}$$

A. THE FIRST DERIVATIVES OF A MANIFOLD MATRIX

The unknown parameter vector is defined as (53). Denote θ_η as the ηth element of $\boldsymbol{\theta}$, The first order deviation of $\mathbf{A}(k,\boldsymbol{\theta})$ is denoted by

$$\mathbf{A}_\eta(k) \triangleq \frac{\partial \mathbf{A}(k,\boldsymbol{\theta})}{\partial \theta_\eta}. \tag{181}$$

If θ_η is an position real parameter

$$\frac{\partial \mathbf{A}(k,\boldsymbol{\theta})}{\partial \theta_\eta} = \tilde{\mathbf{A}}(k,\boldsymbol{\theta})\frac{\partial \mathbf{V}(k,\boldsymbol{\theta})}{\partial \theta_\eta}\boldsymbol{\alpha}, \tag{182}$$

where

$$\frac{\partial \mathbf{V}(k,\boldsymbol{\theta})}{\partial \theta_\eta} = \mathbf{I}_N \otimes \left[0,\ldots,\frac{\partial \bar{\mathbf{V}}_d(k,\boldsymbol{\theta})}{\partial \theta_\eta},\ldots,0\right],$$

$$\frac{\partial \bar{\mathbf{V}}_d(k,\boldsymbol{\theta})}{\partial \theta_\eta} = \mathrm{diag}\left[\frac{\partial e^{-i\omega_k \bar{\tau}_{1d}}}{\partial \theta_\eta}, \frac{\partial e^{-i\omega_k \bar{\tau}_{2d}}}{\partial \theta_\eta}, \ldots, \frac{\partial e^{-i\omega_k \bar{\tau}_{Ld}}}{\partial \theta_\eta}\right].$$

If $\theta_\eta = \alpha_{d,\ell,n}$ is a path attenuation parameter with complex value which represents to the path attenuation from emitter d, relayed by transponder ℓ, to the receiver array n.

$$\frac{\partial \mathbf{A}(k,\boldsymbol{\theta})}{\partial \theta_\eta} = \mathbf{A}(k,\boldsymbol{\theta})\mathbf{V}(k,\boldsymbol{\theta})\frac{\partial \boldsymbol{\alpha}}{\partial \theta_\eta}, \tag{183}$$

where

$$\frac{\partial \boldsymbol{\alpha}}{\partial \theta_\eta} = \mathbf{I}_{d\ell n}, \tag{184}$$

where the ith row, jth column of $\mathbf{I}_{d\ell n}$ is defined as

$$\mathbf{I}_{d\ell n}(i,j) = \begin{cases} 1 & i = (n-1)LD + (d-1)L + \ell, j = d; \\ 0 & \text{else.} \end{cases} \tag{185}$$

B. THE FIRST DERIVATIVES OF A MANIFOLD VECTOR

The dth column of $\mathbf{A}(k,\boldsymbol{\theta})$ is

$$\mathbf{a}_d(k,\boldsymbol{\theta}) = \tilde{\mathbf{A}}(k,\boldsymbol{\theta})\mathbf{V}_d(k,\boldsymbol{\theta})\boldsymbol{\alpha}_d, \tag{186}$$

where $\mathbf{V}_d(k,\boldsymbol{\theta}) = \mathbf{I}_N \otimes \bar{\mathbf{V}}_d(k,\boldsymbol{\theta}), \boldsymbol{\alpha}_d = [\boldsymbol{\alpha}_{d1}^T, \boldsymbol{\alpha}_{d2}^T, \ldots, \boldsymbol{\alpha}_{dN}^T]^T$.

If $\theta_\eta = p_{d,\zeta}$ are position parameters ($\zeta \in \{1,2,3\}$ which represents coordinates x, y, z) of emitter d,

$$\frac{\partial \mathbf{a}_d(k,\boldsymbol{\theta})}{\partial p_{d,\zeta}} = \tilde{\mathbf{A}}(k,\boldsymbol{\theta})\frac{\partial \mathbf{V}_d(k,\boldsymbol{\theta})}{\partial p_{d,\zeta}}\boldsymbol{\alpha}_d, \tag{187}$$

where

$$\frac{\partial \mathbf{V}_d(k,\boldsymbol{\theta})}{\partial p_{d,\zeta}} = \mathbf{I}_N \otimes \frac{\partial \bar{\mathbf{V}}_d(k,\boldsymbol{\theta})}{\partial p_{d,\zeta}}, \tag{188}$$

$$\frac{\partial \bar{\mathbf{V}}_d(k,\boldsymbol{\theta})}{\partial p_{d,\zeta}} = \mathrm{diag}\left\{\left[\frac{\partial e^{-i\omega_k \bar{\tau}_{1,d}}}{\partial p_{d,\zeta}}, \frac{\partial e^{-i\omega_k \bar{\tau}_{2,d}}}{\partial p_{d,\zeta}}, \ldots, \frac{\partial e^{-i\omega_k \bar{\tau}_{L,d}}}{\partial p_{d,\zeta}}\right]\right\}. \tag{189}$$

where $\bar{\tau}_{\ell,d}$ is the propagation delay from the emitter d to the transponder ℓ, and $\eta = 3(\zeta-1)+d$.

If $\theta_\eta = \alpha_{d,\ell,n}$ are path attenuation parameters

$$\frac{\partial \mathbf{a}_d}{\partial \alpha_{d,\ell,n}} = \tilde{\mathbf{A}}\mathbf{V}_d\frac{\partial \boldsymbol{\alpha}_d}{\partial \alpha_{d,\ell,n}}, \tag{190}$$

where

$$\frac{\partial \boldsymbol{\alpha}_d}{\partial \alpha_{d,\ell,n}} = [\underbrace{0,\ldots,1}_{(n-1)L+\ell},0,\ldots,0]^T, \tag{191}$$

and $\eta = 3 + (n-1)L + \ell$.

C. THE FIRST DERIVATIVES OF THE PROJECTION MATRIX

Denote the first partial derivative of the projection matrix $\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})$ with respect to independent variable θ_η by

$$\mathbf{P}_{\mathbf{A}\eta}(k,\boldsymbol{\theta}) \triangleq \frac{\partial \mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})}{\partial \theta_\eta} = \frac{\partial \mathbf{A}(k,\boldsymbol{\theta})}{\partial \theta_\eta}\mathbf{A}(k,\boldsymbol{\theta})^+ + \mathbf{A}(k,\boldsymbol{\theta})\frac{\partial \mathbf{A}(k,\boldsymbol{\theta})^+}{\partial \boldsymbol{\theta}_\eta}, \tag{192}$$

The first order derivative of the pseudo-inverse of $\mathbf{A}(k,\boldsymbol{\theta})$:

$$\frac{\partial \mathbf{A}(k,\boldsymbol{\theta})^+}{\partial \theta_\eta} = [\mathbf{A}(k,\boldsymbol{\theta})^H\mathbf{A}(k,\boldsymbol{\theta})]^{-1}\mathbf{A}_\eta^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}^\perp(k,\boldsymbol{\theta}) - \mathbf{A}(k,\boldsymbol{\theta})^+\mathbf{A}_\eta(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})^+. \tag{193}$$

Substitute (193) into (192):

$$\mathbf{P}_{\mathbf{A}\eta}(k,\boldsymbol{\theta}) = \mathbf{A}(k,\boldsymbol{\theta})^{+H}\mathbf{A}_\eta^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}^\perp(k,\boldsymbol{\theta}) + \mathbf{P}_{\mathbf{A}}^\perp(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})^+. \tag{194}$$

D. THE FIRST DERIVATIVE OF THE COST FUNCTION

The cost function is defined in (180). Denote C_η as the ηth component of the gradient $C'(\boldsymbol{\theta}_0)$ and $\boldsymbol{\theta}_0$ are the true values of the unknown parameters:

$$C_\eta = \sum_{k=1}^{K} C_\eta(k,\boldsymbol{\theta}), \tag{195}$$

where

$$C_\eta(k,\boldsymbol{\theta}) = \mathrm{tr}\{\mathbf{P}_{\mathbf{A}\eta}(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\mathbf{W}(k)\hat{\mathbf{U}}_s^H(k)\}. \tag{196}$$

Substitute (194) into (196)

$$C_\eta(k,\boldsymbol{\theta}) = 2\mathrm{Re}\left\{\mathrm{tr}\left[\mathbf{W}(k)\hat{\mathbf{U}}_s^H(k)\mathbf{A}(k,\boldsymbol{\theta})^{+H}\mathbf{A}_\eta^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}^\perp(k,\boldsymbol{\theta})\hat{\mathbf{U}}_s(k)\right]\right\}. \tag{197}$$

E. THE SECOND DERIVATIVE OF A PROJECTION MATRIX

$$\frac{\partial^2 \mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})}{\partial \theta_\eta \partial \theta_\xi} = \left\{\mathbf{P}_{\mathbf{A}\xi}^\perp(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta}) + \mathbf{P}_{\mathbf{A}}{}^\perp(k,\boldsymbol{\theta})\mathbf{A}_{\eta\xi}(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta}) + \mathbf{P}_{\mathbf{A}}{}^\perp(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\mathbf{A}_\xi^+(k,\boldsymbol{\theta})\right\} + \{\cdots\}^H, \tag{198}$$

where $\{\cdots\}^H$ means that the same expression appears again with complex conjugate and transpose. and

$$\begin{aligned}
\mathbf{A}_\eta(k,\boldsymbol{\theta}) &\triangleq \frac{\partial \mathbf{A}(k,\boldsymbol{\theta})}{\partial \theta_\eta};\\
\mathbf{A}_\xi(k,\boldsymbol{\theta}) &\triangleq \frac{\partial \mathbf{A}(k,\boldsymbol{\theta})}{\partial \theta_\xi};\\
\mathbf{A}_{\eta\xi}(k,\boldsymbol{\theta}) &\triangleq \frac{\partial^2 \mathbf{A}(k,\boldsymbol{\theta})}{\partial \theta_\eta \partial \theta_\xi};\\
\mathbf{A}^+_\xi(k,\boldsymbol{\theta}) &\triangleq \frac{\partial \mathbf{A}^+(k,\boldsymbol{\theta})}{\partial \theta_\xi};\\
\mathbf{P}_{\mathbf{A}\xi}^{\perp}(k,\boldsymbol{\theta}) &\triangleq \frac{\partial \mathbf{P}_{\mathbf{A}}^{\perp}(k,\boldsymbol{\theta})}{\partial \theta_\xi}.
\end{aligned} \tag{199}$$

Substitute (193) and (194) into (198):

$$\begin{aligned}
&\frac{\partial^2 \mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta})}{\partial\theta_\eta\partial\theta_\xi}\\
&= \Big\{-\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\xi(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta})\\
&\quad - \mathbf{A}^{+H}(k,\boldsymbol{\theta})\mathbf{A}_\xi^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta})\\
&\quad + \mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_{\eta\xi}(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta})\\
&\quad + \mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})[\mathbf{A}^2(k,\boldsymbol{\theta})]^{-1}\mathbf{A}_\xi^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\\
&\quad -\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta})\mathbf{A}_\xi(k,\boldsymbol{\theta})\mathbf{A}^+(k,\boldsymbol{\theta})\Big\}\\
&\quad + \{\cdots\}^H.
\end{aligned} \tag{200}$$

F. THE SECOND DERIVATIVE OF THE MD-WSSF COST FUNCTION

The element in the ηth row ξth column of $\mathbf{C}''$, evaluated in $\boldsymbol{\theta}_0$ is given by

$$C_{\eta\xi}(\boldsymbol{\theta}_0) = \lim_{N\to\infty}\sum_{k=1}^{K}\mathrm{tr}\left\{\frac{\partial^2\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta}_0)}{\partial\theta_\eta\partial\theta_\xi}\hat{\mathbf{U}}_s(k)\mathbf{W}(k)\hat{\mathbf{U}}_s(k)^H\right\}. \tag{201}$$

Apply (200) to (201) and $\mathbf{P}_{\mathbf{A}}{}^{\perp}\mathbf{U}_s(k)=0$:

$$\begin{aligned}
C_{\eta\xi} &= \sum_{k=1}^{K}\mathrm{tr}\left\{\frac{\partial^2\mathbf{P}_{\mathbf{A}}(k,\theta_0)}{\partial\theta_\eta\partial\theta_\xi}\mathbf{U}_s(k)\mathbf{W}(k)\mathbf{U}_s(k)^H\right\}\\
&= -\sum_{k=1}^{K}\mathrm{tr}\Big\{[\mathbf{A}_\xi^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\\
&\qquad + \mathbf{A}_\eta^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\xi(k,\boldsymbol{\theta})]\\
&\qquad \mathbf{A}^+(k,\boldsymbol{\theta})\mathbf{U}_s(k)\mathbf{W}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k,\boldsymbol{\theta})\Big\}.
\end{aligned} \tag{202}$$

Apply Lemma 2 in Appendix B

$$\begin{aligned}
C_{\eta\xi} = -2\sum_{k=1}^{K}\mathrm{Re}\Big\{\mathrm{tr}\Big[&\mathbf{A}_\xi^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\\
&\mathbf{A}^+(k,\boldsymbol{\theta})\mathbf{U}_s(k)\mathbf{W}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k,\boldsymbol{\theta})\Big]\Big\}.
\end{aligned} \tag{203}$$

The second derivative matrix of $C(k,\boldsymbol{\theta})$ is denoted as

$$\mathbf{C}''(k,\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{C}''_{1,1}(k,\boldsymbol{\theta}) & \cdots & \mathbf{C}''_{1,3+LN}(k,\boldsymbol{\theta})\\ \vdots & \ddots & \vdots\\ \mathbf{C}''_{3+LN,1}(k,\boldsymbol{\theta}) & \cdots & \mathbf{C}''_{3+LN,3+LN}(k,\boldsymbol{\theta})\end{bmatrix}, \tag{204}$$

where

$$\mathbf{C}''_{i,j}(k,\boldsymbol{\theta}) \triangleq \left[\frac{\partial C(k,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_i^T}\right]^H\frac{\partial C(k,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}_j^T}. \tag{205}$$

Denote ξ is an unknown parameter in $\boldsymbol{\theta}_i$ corresponding to the βth emitter, and η is an unknown parameter in $\boldsymbol{\theta}_j$ corresponding to the γth emitter. We have

$$\mathbf{A}_\xi(k,\boldsymbol{\theta}) = [\underbrace{\mathbf{0},\cdots,\mathbf{0},\frac{\partial\mathbf{a}_\beta(k,\boldsymbol{\theta})}{\partial\xi}}_{i\text{ columns}},\mathbf{0},\cdots,\mathbf{0}]; \tag{206}$$

$$\mathbf{A}_\eta(k,\boldsymbol{\theta}) = [\underbrace{\mathbf{0},\cdots,\mathbf{0},\frac{\partial\mathbf{a}_\gamma(k,\boldsymbol{\theta})}{\partial\eta}}_{j\text{ columns}},\mathbf{0},\cdots,\mathbf{0}]. \tag{207}$$

where $\mathbf{a}_\beta(k,\boldsymbol{\theta})$ and $\mathbf{a}_\gamma(k,\boldsymbol{\theta})$ are the βth and the γth columns of $\mathbf{A}(k,\boldsymbol{\theta})$. Denote

$$\begin{aligned}
\mathbf{D}_{\xi,\eta}(k,\boldsymbol{\theta}) &\triangleq \mathbf{A}_\xi^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_\eta(k,\boldsymbol{\theta})\\
&= \begin{Bmatrix}\mathbf{0} & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & d_{i,j}^{\beta,\gamma}(k,\boldsymbol{\theta}) & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & \mathbf{0}\end{Bmatrix},
\end{aligned} \tag{208}$$

where $d_{i,j}^{\beta\gamma}(k,\boldsymbol{\theta}) = \frac{\partial\mathbf{a}_\beta^H(k,\boldsymbol{\theta})}{\partial\xi}\mathbf{P}_{\mathbf{A}}{}^{\perp}\frac{\partial\mathbf{a}_\gamma(k,\boldsymbol{\theta})}{\partial\eta}$. $\mathbf{D}_{\xi\eta}(k,\boldsymbol{\theta})$ is a matrix with all zeros except the element in the row i column j. Substitute (208) into (203), and denote the βth row γth column element of $\mathbf{C}_{i,j}(k,\boldsymbol{\theta})$ as $C_{i,j}^{\beta,\gamma}(k,\boldsymbol{\theta})$.

$$\begin{aligned}
C_{\xi,\eta} = C_{i,j}^{\beta\gamma} &= -2\sum_{k=1}^{K}\mathrm{Re}\left[\mathrm{tr}\left\{\mathbf{D}_{i,j}^{\beta\gamma}(k,\boldsymbol{\theta})\mathbf{E}(k,\boldsymbol{\theta})\right\}\right]\\
&= \sum_{k=1}^{K}\mathrm{Re}\left[d_{i,j}^{\beta,\gamma}(k,\boldsymbol{\theta})e_{\gamma,\beta}(k,\boldsymbol{\theta})\right],
\end{aligned} \tag{209}$$

where $e_{\gamma,\beta}(k,\boldsymbol{\theta})$ the element at row γ column β of $\mathbf{E}(k,\boldsymbol{\theta})$, and

$$\mathbf{E}(k,\boldsymbol{\theta}) \triangleq \mathbf{A}^+(k,\boldsymbol{\theta})\mathbf{U}_s(k)\mathbf{W}(k)\mathbf{U}_s^H(k)\mathbf{A}^{+H}(k). \tag{210}$$

where

$$\mathbf{F}(k,\boldsymbol{\theta}) = \mathbf{A}_{\boldsymbol{\theta}}^H(k,\boldsymbol{\theta})\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_{\boldsymbol{\theta}}(k,\boldsymbol{\theta}). \tag{211}$$

Matrix $\mathbf{F}$ is denoted as

$$\mathbf{F}(k,\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{F}_{1,1}(k,\boldsymbol{\theta}) & \cdots & \mathbf{F}_{1,3+LN}(k,\boldsymbol{\theta})\\ \vdots & \ddots & \vdots\\ \mathbf{F}_{3+LN,1}(k,\boldsymbol{\theta}) & \cdots & \mathbf{F}_{3+LN,3+LN}(k,\boldsymbol{\theta})\end{bmatrix}, \tag{212}$$

where the ith row j column block matrix $\mathbf{F}_{i,j}(k,\boldsymbol{\theta})$ is defined as

$$\mathbf{F}_{i,j}(k,\boldsymbol{\theta}) \triangleq \left[\mathbf{A}_{\boldsymbol{\theta}_i}(k,\boldsymbol{\theta})\right]^H \mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta})\mathbf{A}_{\boldsymbol{\theta}_j}(k,\boldsymbol{\theta}), \quad (213)$$

where $\mathbf{A}_{\boldsymbol{\theta}_i}(k,\boldsymbol{\theta})$ and $\mathbf{A}_{\boldsymbol{\theta}_j}(k,\boldsymbol{\theta})$ are defined in (206) and (207); the βth row γth column element of $\mathbf{F}_{i,j}(k,\boldsymbol{\theta})$ is denoted as $f_{i,j}^{\beta,\gamma}(k,\boldsymbol{\theta})$. Since (213) and (208), we get $f_{i,j}^{\beta,\gamma}(k,\boldsymbol{\theta}) = d_{i,j}^{\beta,\gamma}(k,\boldsymbol{\theta})$. Substitute (213) into (209)

$$C_{\xi,\eta}(k,\boldsymbol{\theta}) = -2\mathrm{Re}\left[f_{i,j}^{\beta,\gamma}(k,\boldsymbol{\theta})e_{\gamma,\beta}(k,\boldsymbol{\theta})\right], \quad (214)$$

Make (214) into matrix form

$$\mathbf{C}''(k,\boldsymbol{\theta}) = -2\mathrm{Re}\left\{\mathbf{F}(k,\boldsymbol{\theta}) \odot [\mathbf{1}_{3+LN} \otimes \mathbf{E}^T(k,\boldsymbol{\theta})]\right\}, \quad (215)$$

where $\odot$ represents the Hadamard-Schur product, $\otimes$ is the Kronecker product, $\mathbf{1}_{3+LN}$ is a $(3+LN)\times(3+LN)$ matrix with all ones.

G. THE SECOND DERIVATIVE OF MD-NSF

The element in the ηth row ξth column of $\mathbf{C}''$, evaluated in $\boldsymbol{\theta}_0$ is given by

$$C_{\eta\xi}(\boldsymbol{\theta}_0) = \lim_{N\to\infty}\sum_{k=1}^{K}\mathrm{tr}\left\{\frac{\partial^2\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta}_0)}{\partial\theta_\eta\partial\theta_\xi}\hat{\mathbf{U}}_n(k)\hat{\mathbf{U}}_n(k)^H\right\}. \quad (216)$$

Apply (200) to (216) and $\mathbf{A}^H(k,\theta_0)\mathbf{U}_n(k) = 0$:

$$\begin{aligned} C_{\eta\xi} &= \sum_{k=1}^{K}\mathrm{tr}\left\{\frac{\partial^2\mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta}_0)}{\partial\theta_\eta\partial\theta_\xi}\mathbf{U}_n(k)\mathbf{U}_n(k)^H\right\} \\ &= \sum_{k=1}^{K}\mathrm{tr}\Big\{[\mathbf{A}_\eta(k,\boldsymbol{\theta}_0)\mathbf{E}(k,\boldsymbol{\theta}_0)\mathbf{A}_\xi^H(k,\boldsymbol{\theta}_0) \\ &\quad + \mathbf{A}_\xi(k,\boldsymbol{\theta}_0)\mathbf{E}(k,\boldsymbol{\theta}_0)\mathbf{A}_\eta^H(k,\boldsymbol{\theta}_0)] \\ &\quad \mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta}_0)\mathbf{U}_n(k)\mathbf{U}_n^H(k)\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta}_0)\Big\}. \end{aligned} \quad (217)$$

where $\mathbf{E}(k,\boldsymbol{\theta}_0) \triangleq [\mathbf{A}(k,\boldsymbol{\theta}_0)^H\mathbf{A}(k,\boldsymbol{\theta}_0)]^{-1}$. Since

$$\begin{aligned} \mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta}_0)\mathbf{U}_n(k) &= [\mathbf{I} - \mathbf{P}_{\mathbf{A}}(k,\boldsymbol{\theta}_0)]\mathbf{U}_n(k) \\ &= \mathbf{U}_n(k), \end{aligned} \quad (218)$$

$$\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta}_0)\mathbf{U}_n(k)\mathbf{U}_n^H(k)\mathbf{P}_{\mathbf{A}}{}^{\perp}(k,\boldsymbol{\theta}_0) = \mathbf{U}_n(k)\mathbf{U}_n^H(k), \quad (219)$$

apply Lemma 2 in Appendix B, (217) turns

$$\begin{aligned} C_{\eta\xi} &= 2\sum_{k=1}^{K}\mathrm{Re}\Big\{\mathrm{tr}\Big\{\mathbf{A}_\eta(k,\boldsymbol{\theta}_0)\mathbf{E}(k,\boldsymbol{\theta}_0)\mathbf{A}_\xi^H(k,\boldsymbol{\theta}_0) \\ &\quad \mathbf{U}_n(k)\mathbf{U}_n^H(k)\Big\}\Big\} \\ &= 2\sum_{k=1}^{K}\mathrm{Re}\Big\{\mathrm{tr}\Big\{\mathbf{A}_\xi^H(k,\boldsymbol{\theta}_0)\mathbf{U}_n(k,\boldsymbol{\theta}_0)\mathbf{U}_n^H(k)\mathbf{A}_\eta(k,\boldsymbol{\theta}_0) \\ &\quad \mathbf{E}(k,\boldsymbol{\theta}_0)\}\Big\}, \end{aligned} \quad (220)$$

Make (220) into matrix from:

$$\mathbf{C}''(k,\boldsymbol{\theta}_0) = -2\mathrm{Re}\left\{\mathbf{F}(k,\boldsymbol{\theta}_0) \odot [\mathbf{1}_{3+LN} \otimes \mathbf{E}^T(k,\boldsymbol{\theta}_0)]\right\}, \quad (221)$$

where

$$\mathbf{F}(k,\boldsymbol{\theta}_0) = \mathbf{A}_{\boldsymbol{\theta}}^H(k,\boldsymbol{\theta}_0)\mathbf{U}_n(k)\mathbf{U}_n^H(k)\mathbf{A}_{\boldsymbol{\theta}}(k,\boldsymbol{\theta}_0), \quad (222)$$

$$\mathbf{E}(k,\boldsymbol{\theta}_0) = [\mathbf{A}^H(k,\boldsymbol{\theta}_0)\mathbf{A}(k,\boldsymbol{\theta}_0)]^{-1}. \quad (223)$$

APPENDIX C
FISHER INFORMATION MATRIX

The log-likelihood function of the observations is

$$\begin{aligned} \log[\mathrm{p}(\mathbf{r}|\bar{\boldsymbol{\theta}})] &= KMN\log(\pi\sigma^2) \\ &\quad + \frac{1}{\sigma^2}\sum_{k=1}^{K}||\mathbf{r}(k) - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)||^2. \end{aligned} \quad (224)$$

Move the constant item

$$L = -\frac{1}{\sigma^2}\sum_{k=1}^{K}[\mathbf{r}(k) - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H[\mathbf{r}(k) - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]. \quad (225)$$

A. THE FIRST DERIVATIVES OF THE LOG-LIKELIHOOD FUNCTION

If $\bar{\theta}_i$ is a real value parameter, the first derivative of the log likelihood function is:

$$\begin{aligned} \frac{\partial L}{\partial\bar{\theta}_i} &= \frac{1}{\sigma^2}\sum_{k=1}^{K}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\theta}_i}\mathbf{e}(k,\boldsymbol{\theta})\right. \\ &\quad \left. + \mathbf{e}^H(k,\boldsymbol{\theta})\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\bar{\theta}_i}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\theta}_i}\mathbf{e}(k,\boldsymbol{\theta})\right\}, \end{aligned} \quad (226)$$

where $\mathbf{e}(k,\boldsymbol{\theta}) = \mathbf{r}(k) - \mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)$.

If $\bar{\theta}_j$ is an imaginary value parameter, the first derivative of the log likelihood function is

$$\begin{aligned} \frac{\partial L}{\partial\bar{\theta}_j} &= \frac{1}{\sigma^2}\sum_{k=1}^{K}\left\{-i\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\theta}_j}\mathbf{e}(k,\boldsymbol{\theta})\right. \\ &\quad \left. + i\mathbf{e}^H(k,\boldsymbol{\theta})\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\bar{\theta}_j}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Im}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\theta}_j}\mathbf{e}(k,\boldsymbol{\theta})\right\}, \end{aligned} \quad (227)$$

B. THE SECOND DERIVATIVES OF THE LOG-LIKELIHOOD FUNCTION

Notice the following results:

$$\begin{cases} \mathrm{Re}(x)\mathrm{Re}(y^T) = \frac{1}{2}[\mathrm{Re}(xy^T) + \mathrm{Re}(xy^H)]; \\ \mathrm{Im}(x)\mathrm{Im}(y^T) = -\frac{1}{2}[\mathrm{Re}(xy^T) - \mathrm{Re}(xy^H)]; \\ \mathrm{Re}(x)\mathrm{Im}(y^T) = \frac{1}{2}[\mathrm{Im}(xy^T) - \mathrm{Im}(xy^H)]. \end{cases} \quad (228)$$

If $\bar{\theta}_i$ and $\bar{\theta}_j$ are real parameters, the second derivatives are:

$$\begin{aligned}\frac{\partial^2 L}{\partial\theta_i\partial\theta_j} &= \frac{4}{\sigma^4}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}\mathbf{e}(k_1)\right\} \\ &\quad\times\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]^H}{\partial\bar{\theta}_j}\mathbf{e}(k_2)\right\}^T \\ &= \frac{2}{\sigma^4}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \\ &\left\{\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}[\mathbf{e}(k_1)\mathbf{e}^T(k_2)]\ \frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]^{HT}}{\partial\bar{\theta}_j}\right\}\right. \\ &\left.+\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}[\mathbf{e}(k_1)\mathbf{e}^H(k_2)]\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]}{\partial\bar{\theta}_j}\right\}\right\}.\end{aligned} \tag{229}$$

The fisher information matrix corresponding to θ_i and θ_j which are real value parameters

$$\begin{aligned}\mathbf{I}_{ij} &= E\left\{\frac{\partial^2 L}{\partial\bar{\theta}_i\partial\bar{\theta}_j}\right\} \\ &= \frac{2}{\sigma^4}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \\ &\quad\times\left\{\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}E\{\mathbf{e}(k_1)\mathbf{e}^T(k_2)\}\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]^{HT}}{\partial\bar{\theta}_j}\right\}\right. \\ &\quad\left.+\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}E\{\mathbf{e}(k_1)\mathbf{e}^H(k_2)\}\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]}{\partial\bar{\theta}_j}\right\}\right\}.\end{aligned} \tag{230}$$

Since $E\{\mathbf{e}(k_1)\mathbf{e}^T(k_2)\} = \mathbf{0}, \forall k_1, k_2$; $E\{\mathbf{e}(k_1)\mathbf{e}(k_2)^H\} = \mathbf{0}, \forall k_1 \neq k_2$, and $E\{\mathbf{e}(k,\boldsymbol{\theta})\mathbf{e}^H(k,\boldsymbol{\theta})\} = \sigma^2\mathbf{I}$

$$\mathbf{I}_{ij} = \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\theta}_i}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\bar{\theta}_j}\right\}. \tag{231}$$

If $\bar{\theta}_i$ and $\bar{\theta}_j$ are imaginary parameters, the second derivatives are:

$$\begin{aligned}\frac{\partial^2 L}{\partial\bar{\theta}_i\partial\bar{\theta}_j} &= \frac{2}{\sigma^4}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \\ &\quad\times\left\{-\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}\mathbf{e}(k_1)\mathbf{e}^T(k_2)\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]^{HT}}{\partial\bar{\theta}_j}\right\}\right. \\ &\quad\left.+\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}\mathbf{e}(k_1)\mathbf{e}^H(k_2)\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]}{\partial\bar{\theta}_j}\right\}\right\}.\end{aligned} \tag{232}$$

The fisher information matrix corresponding to $\bar{\theta}_i$ and $\bar{\theta}_j$ which are imaginary value parameters:

$$\mathbf{I}_{ij} = \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\theta}_i}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\bar{\theta}_j}\right\}. \tag{233}$$

If $\bar{\theta}_i$ is a real value parameter and $\bar{\theta}_j$ is an imaginary parameter, the second derivative are:

$$\begin{aligned}&\frac{\partial^2 L}{\partial\bar{\theta}_i\partial\bar{\theta}_j} \\ &= \frac{2}{\sigma^4}\sum_{k_1=1}^{K}\sum_{k_2=1}^{K} \\ &\quad\times\left\{\mathrm{Im}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}\mathbf{e}(k_1)\mathbf{e}^T(k_2)\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]^{HT}}{\partial\bar{\theta}_j}\right\}\right. \\ &\quad\left.-\mathrm{Im}\left\{\frac{\partial[\mathbf{A}(k_1)\mathbf{s}(k_1)]^H}{\partial\bar{\theta}_i}\mathbf{e}(k_1)\mathbf{e}^H(k_2)\frac{\partial[\mathbf{A}(k_2)\mathbf{s}(k_2)]}{\partial\bar{\theta}_j}\right\}\right\}.\end{aligned} \tag{234}$$

The fisher information matrix corresponding to $\bar{\theta}_i$ and $\bar{\theta}_j$ which $\bar{\theta}_i$ is a real parameter and $\bar{\theta}_j$ is an imaginary value parameter:

$$\mathbf{I}_{ij} = -\frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Im}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\theta}_i}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\bar{\theta}_j}\right\}. \tag{235}$$

C. ELEMENTS OF THE FISHER INFORMATION MATRIX

$$\begin{aligned}\mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} &\triangleq E\left\{\frac{\partial L}{\partial\boldsymbol{\theta}}^H\frac{\partial L}{\partial\boldsymbol{\theta}}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\boldsymbol{\theta}}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\boldsymbol{\theta}}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{S}^H(k)\mathbf{A}_{\boldsymbol{\theta}}^H(k,\boldsymbol{\theta})\mathbf{A}_{\boldsymbol{\theta}}(k,\boldsymbol{\theta})\mathbf{S}(k)\right\},\end{aligned} \tag{236}$$

where $\mathbf{S}(k) = \mathbf{I}_{3+LN}\otimes\mathbf{S}(k)$, $\mathbf{S}(k) = \mathrm{diag}\{\mathbf{s}(k)\}$.

$$\begin{aligned}\mathbf{I}_{\bar{\mathbf{s}}\boldsymbol{\theta}}(k) &\triangleq E\left\{\frac{\partial L}{\partial\bar{\mathbf{s}}(k)}^H\frac{\partial L}{\partial\boldsymbol{\theta}}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\mathbf{s}}(k)}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\boldsymbol{\theta}}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}_{\boldsymbol{\theta}}(k,\boldsymbol{\theta})\mathbf{S}(k)\right\},\end{aligned} \tag{237}$$

$$\begin{aligned}\mathbf{I}_{\tilde{\mathbf{s}}\boldsymbol{\theta}}(k) &\triangleq E\left\{\frac{\partial L}{\partial\tilde{\mathbf{s}}(k)}^H\frac{\partial L}{\partial\boldsymbol{\theta}}\right\} \\ &= -\frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Im}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial[\tilde{\mathbf{s}}(k)i]}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\boldsymbol{\theta}}\right\} \\ &= -\frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Im}\left\{\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}_{\boldsymbol{\theta}}(k,\boldsymbol{\theta})\mathbf{S}(k)\right\},\end{aligned} \tag{238}$$

$$\begin{aligned}\mathbf{I}_{\bar{\mathbf{s}}\bar{\mathbf{s}}}(k) &\triangleq E\left\{\frac{\partial L}{\partial\bar{\mathbf{s}}(k)}^H\frac{\partial L}{\partial\bar{\mathbf{s}}(k)}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial\bar{\mathbf{s}}(k)}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial\bar{\mathbf{s}}(k)}\right\} \\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})\right\},\end{aligned} \tag{239}$$

$$\begin{aligned}\mathbf{I}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(k) &\triangleq E\left\{\frac{\partial L}{\partial \tilde{\mathbf{s}}(k)}^H \frac{\partial L}{\partial \tilde{\mathbf{s}}(k)}\right\}\\ &= -\frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Im}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial \tilde{\mathbf{s}}(k)}\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]}{\partial[\tilde{\mathbf{s}}(k)i]}\right\}\\ &= -\frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Im}\left\{\mathbf{A}^H(k,\boldsymbol{\theta})\mathbf{A}(k,\boldsymbol{\theta})\right\}, \qquad (240)\end{aligned}$$

$$\begin{aligned}\mathbf{I}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(k) &\triangleq E\left\{\frac{\partial L}{\partial \tilde{\mathbf{s}}(k)}^H \frac{\partial L}{\partial \tilde{\mathbf{s}}(k)}\right\}\\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\frac{\partial[\mathbf{A}(k,\boldsymbol{\theta})\mathbf{s}(k)]^H}{\partial[\tilde{\mathbf{s}}(k)i]}\frac{\partial[\mathbf{A}(k)\mathbf{s}(k)]}{\partial[\tilde{\mathbf{s}}(k)i]}\right\}\\ &= \frac{2}{\sigma^2}\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{A}^H(k)\mathbf{A}(k,\boldsymbol{\theta})\right\}, \qquad (241)\end{aligned}$$

The Fisher Information Matrix is

$$\mathbf{I}_{\bar{\boldsymbol{\theta}}} = \left[\begin{array}{cccc:c} \mathbf{G}(1) & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{V}(1)\\ \mathbf{0} & \mathbf{G}(2) & \cdots & \mathbf{0} & \mathbf{V}(2)\\ \vdots & \vdots & \ddots & \vdots & \vdots\\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{G}(K) & \mathbf{V}(K)\\ \hdashline \mathbf{V}^H(1) & \mathbf{V}^H(2) & \cdots & \mathbf{V}^H(K) & \mathbf{I}_{\boldsymbol{\theta}\boldsymbol{\theta}} \end{array}\right], \qquad (242)$$

where $\mathbf{G}(k,\boldsymbol{\theta}) \triangleq \begin{bmatrix}\mathbf{I}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(k) & \mathbf{I}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(k)\\ \mathbf{I}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(k) & \mathbf{I}_{\tilde{\mathbf{s}}\tilde{\mathbf{s}}}(k)\end{bmatrix}$, $\mathbf{V}(k) \triangleq \begin{bmatrix}\mathbf{I}_{\tilde{\mathbf{s}}\boldsymbol{\theta}}(k)\\ \mathbf{I}_{\tilde{\mathbf{s}}\boldsymbol{\theta}}(k)\end{bmatrix}$.

APPENDIX D
LEMMAS FOR ASYMPTOTIC ANALYSIS

A. COVARIANCE OF THE EIGENVECTOR ESTIMATIONS

Lemma 4: Denote that $\mathbf{R}$ *is the covariance matrix of received data. The columns of* $\mathbf{U}$ *is the eigenvectors of* $\mathbf{R}$. $\mathbf{U}_s = [\mathbf{u}_s(1), \mathbf{u}_s(2), \ldots, \mathbf{u}_s(D)]$ *is consisted by the eigenvectors corresponding to the D signals.* $\hat{\mathbf{u}}_s(d)$ *is a estimation of* $\mathbf{u}_s(d)$ *from* $\hat{\mathbf{R}}$. *The eigenvector estimation errors* $\tilde{\mathbf{u}}_s(d) = \hat{\mathbf{u}}(d) - \mathbf{u}_s(d)$ *are asymptotically (for large N, where N is the number of snapshots) jointly Gaussian distributed with zero means and covariance matrices given by*

$$\begin{aligned}&E\{[\hat{\mathbf{u}}_s(d_i) - \mathbf{u}_s(d_i)][\hat{\mathbf{u}}_s(d_j) - \mathbf{u}_s(d_j)]^H\}\\ &= \frac{\lambda_s(d_i)}{N}\left\{\sum_{\substack{d=1\\ d\neq d_i}}^{D}\frac{\lambda_s(d)}{[\lambda_s(d)-\lambda_s(d_i)]^2}\mathbf{u}_s(d)\mathbf{u}_s^H(d)\right.\\ &\quad\left.+\frac{\sigma^2}{[\sigma^2-\lambda_s(d_i)]^2}\mathbf{U}_n\mathbf{U}_n^H\right\}\delta_{d_i,d_j} + o(N^{-1}), \qquad (243)\end{aligned}$$

$$\begin{aligned}&E\left\{[\hat{\mathbf{u}}_s(d_i) - \mathbf{u}_s(d_i)][\hat{\mathbf{u}}_s(d_j) - \mathbf{u}_s(d_j)]^T\right\}\\ &= -\frac{\lambda_s(d_i)\lambda_s(d_j)}{N[\lambda_s(d_i)-\lambda_s(d_j)]^2}\mathbf{u}_s(d_j)\mathbf{u}_s^T(d_i)(1-\delta_{d_i,d_j}) + o(N^{-1}),\end{aligned} \qquad (244)$$

where $\lambda_s(d)$ *is the* d*th eigenvalue.* σ *is the noise standard deviation, and* δ_{d_i,d_j} *is the Kronecker delta (=1 if* $d_i = d_j$*, and =0 if* $d_i \neq d_j$*).* $\mathbf{U}_n$ *is consisted by the eigenvectors corresponding to the noise.*

Proof: This Lemma was proposed and had been proved in [27], [38], [39], and [46]–[50] . □

APPENDIX E
LEMMAS FOR THE OPTIMAL WEIGHTS OF WEIGHTED SIGNAL SUBSPACE FITTING

Lemma 5: Let $\mathbf{A}_k$, $\mathbf{B}_k$ *and* $\mathbf{C}_k$ *be* $m \times m$ *(Hermitian) Positive semi-definite matrices (*$\mathbf{B}_k$ *can be Hermitian only). Denote* $\mathbf{1}$ *is a* $n \times n$ *square matrix with all ones. Then, assuming that the inverses appearing below exist, it holds that*

$$\begin{aligned}&\left\{\sum_{k=1}^{K}\mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}\otimes\mathbf{B}_k)]\right\}^{-1}\left\{\sum_{k=1}^{K}\mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}\otimes\mathbf{C}_k)]\right\}\\ &\quad\times\left\{\sum_{k=1}^{K}\mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}\otimes\mathbf{B}_k)]\right\}^{-1}\\ &\geq\left\{\sum_{k=1}^{K}\mathrm{Re}\left\{\mathbf{A}_k \odot [\mathbf{1}\otimes(\mathbf{B}_k\mathbf{C}_k^{-1}\mathbf{B}_k)]\right\}\right\}^{-1}. \qquad (245)\end{aligned}$$

Proof: Notice that

$$\mathbf{D}_k = \begin{bmatrix}\mathbf{A}_k & \mathbf{A}_k\\ \mathbf{A}_k & \mathbf{A}_k\end{bmatrix} = \begin{bmatrix}\mathbf{I}_m\\ \mathbf{I}_m\end{bmatrix}\mathbf{A}\begin{bmatrix}\mathbf{I}_m & \mathbf{I}_m\end{bmatrix}, \qquad (246)$$

where $\mathbf{I}_m$ is an $m \times m$ identity matrix. Denote

$$\begin{aligned}\mathbf{F}_k &= \begin{bmatrix}\mathbf{1}\otimes\mathbf{C}_k & \mathbf{1}^2\otimes\mathbf{B}_k\\ \mathbf{1}^2\otimes\mathbf{B}_k & \mathbf{1}^3\otimes(\mathbf{B}_k\mathbf{C}_k^{-1}\mathbf{B}_k)\end{bmatrix}\\ &= \begin{bmatrix}\mathbf{I}_{mn}\\ \mathbf{1}\otimes(\mathbf{B}_k\mathbf{C}_k^{-1})\end{bmatrix}(\mathbf{1}\otimes\mathbf{C}_k)\begin{bmatrix}\mathbf{I}_{mn} & \mathbf{1}\otimes(\mathbf{C}_k^{-1}\mathbf{B}_k)\end{bmatrix},\end{aligned} \qquad (247)$$

where $\mathbf{I}_{mn}$ is an $mn \times mn$ identity matrix. Since $\mathbf{C}_k$ and $\mathbf{B}_k$ are Hermitian matrices, $\mathbf{D}_k$ and $\mathbf{F}_k$ are semi-definite matrices. Follow from [34, Lemma A.1] (Let $\mathbf{D}_k, \mathbf{F}_k \in \mathbb{C}^{m\times m}$ be two (Hermitian) positive semi-definite matrices. Then the matrix $\mathbf{D}_k \odot \mathbf{F}_k$ is positive semi-definite too.), and the fact that $\mathrm{Re}\{\mathbf{D}_k \odot \mathbf{F}_k\} \geq 0$ if $\mathbf{D}_k \odot \mathbf{F}_k \geq 0$. Sum $\mathrm{Re}\{\mathbf{D}_k \odot \mathbf{F}_k\}$ over k, we get

$$\sum_{k=1}^{K}\mathrm{Re}(\mathbf{D}_k \odot \mathbf{F}_k) = \begin{bmatrix}\mathbf{M}_{\mathbf{AC}} & \mathbf{M}_{\mathbf{AB}}\\ \mathbf{M}_{\mathbf{AB}} & \mathbf{M}_{\mathbf{BCB}}\end{bmatrix} \geq 0. \qquad (248)$$

where

$$\mathbf{M}_{\mathbf{AC}} \triangleq \sum_{k=1}^{K}\mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}\otimes\mathbf{C}_k)], \qquad (249)$$

$$\mathbf{M}_{\mathbf{AB}} \triangleq \sum_{k=1}^{K}\mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}^2\otimes\mathbf{B}_k)], \qquad (250)$$

$$\mathbf{M}_{\mathbf{BCB}} \triangleq \sum_{k=1}^{K}\mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}^3\otimes\mathbf{B}_k\mathbf{C}_k^{-1}\mathbf{B}_k)]. \qquad (251)$$

Since $\mathbf{1}^m = N^{m-1}\mathbf{1}$, (248) turns to be

$$\sum_{k=1}^{K} \mathrm{Re}(\mathbf{D}_k \odot \mathbf{F}_k) = \begin{bmatrix} \mathbf{M}_{\mathbf{AC}} & n\bar{\mathbf{M}}_{AB} \\ n\bar{\mathbf{M}}_{AB} & n^2\bar{\mathbf{M}}_{BCB} \end{bmatrix} \geq 0. \quad (252)$$

where

$$\bar{\mathbf{M}}_{AB} \triangleq \sum_{k=1}^{K} \mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}^2 \otimes \mathbf{B}_k)], \quad (253)$$

$$\bar{\mathbf{M}}_{BCB} \triangleq \sum_{k=1}^{K} \mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1}^3 \otimes \mathbf{B}_k\mathbf{C}_k^{-1}\mathbf{B}_k)]. \quad (254)$$

From Lemma 3,

$$\begin{aligned} &\left\{\sum_{k=1}^{K} \mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1} \otimes \mathbf{B}_k)]\right\}^{-1} \left\{\sum_{k=1}^{K} \mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1} \otimes \mathbf{C}_k)]\right\} \\ &\quad \times \left\{\sum_{k=1}^{K} \mathrm{Re}[\mathbf{A}_k \odot (\mathbf{1} \otimes \mathbf{B}_k)]\right\}^{-1} \\ &\geq \left\{\sum_{k=1}^{K} \mathrm{Re}\left\{\mathbf{A}_k \odot [\mathbf{1} \otimes (\mathbf{B}_k\mathbf{C}_k^{-1}\mathbf{B}_k)]\right\}\right\}^{-1}. \end{aligned} \quad (255)$$

□

REFERENCES

[1] J. Du, D. Wang, W. Yu, and H. Yu, "Direct position determination of unknown signals in the presence of multipath propagation," *Sensors*, vol. 18, no. 3, p. E892, Mar. 2018.

[2] A. J. Weiss, "Direct position determination of narrowband radio frequency transmitters," *IEEE Signal Process. Lett.*, vol. 11, no. 5, pp. 513–516, May 2004.

[3] A. Amar and A. J. Weiss, "Direct position determination of multiple radio signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Montreal, PQ, Canada, May 2004, p. ii-81.

[4] A. Amar and A. J. Weiss, "Localization of narrowband radio emitters based on Doppler frequency shifts," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5500–5508, Nov. 2008.

[5] A. Weiss, "Direct geolocation of wideband emitters based on delay and Doppler," *IEEE Trans. Signal Process.*, vol. 59, no. 6, pp. 2513–2521, Jun. 2011.

[6] T. Tirer and A. J. Weiss, "High resolution direct position determination of radio frequency sources," *IEEE Signal Process. Lett.*, vol. 23, no. 2, pp. 192–196, Feb. 2016.

[7] T. Tirer and A. J. Weiss, "High resolution localization of narrowband radio emitters based on Doppler frequency shifts," *Signal Process.*, vol. 141, pp. 288–298, Dec. 2017.

[8] A. Amar and A. J. Weiss, "Advances in direct position determination," in *Proc. Sensor Array Multichannel Signal Process. Workshop*, Barcelona, Spain, Jul. 2004, pp. 584–588.

[9] A. Amar and A. J. Weiss, "New asymptotic results on two fundamental approaches to mobile terminal location," in *Proc. 3rd Int. Symp. Commun., Control Signal Process.*, St Julians, Malta, Mar. 2008, pp. 1320–1323.

[10] A. Amar and A. J. Weiss, "Direct position determination (DPD) of multiple known and unknown radio-frequency signals," in *Proc. 12th Eur. Signal Process. Conf.*, Sep. 2004, pp. 1115–1118.

[11] R. Hamza and K. Buckley, "Second-order statistical analysis of totally weighted subspace fitting methods," *IEEE Trans. Signal Process.*, vol. 42, no. 9, pp. 2520–2524, Sep. 1994.

[12] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 2, pp. 331–341, Apr. 1986.

[13] X.-L. Li, M. Anderson, and T. Adali, "Principal component analysis for noncircular signals in the presence of circular white Gaussian noise," in *Proc. 44th Conf. Rec. Asilomar Conf. Signals, Syst. Comput.*, Nov. 2010, pp. 1796–1801.

[14] J. Dauxois, A. Pousse, and Y. Romain, "Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference," *J. Multivariate Anal.*, vol. 12, no. 1, pp. 136–154, Mar. 1982.

[15] H. Wang and M. Kaveh, "On the performance of signal-subspace processing—Part I: Narrow-band systems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1201–1209, Oct. 1986.

[16] H. Wang and M. Kaveh, "On the performance of signal-subspace processing—Part II: Coherent wide-band systems," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 11, pp. 1583–1591, Nov. 1987.

[17] S. U. Pillai and B. H. Kwon, "Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 8, pp. 1176–1189, Aug. 1989.

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

[19] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-36, no. 4, pp. 532–544, Apr. 1988.

[20] B. Porat and B. Friedlander, "On the asymptotic relative efficiency of the MUSIC algorithm," in *Proc. Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, New York, NY, USA, vol. 4, Apr. 1988, pp. 2376–2379.

[21] J. P. Delmas and Y. Meurisse, "Asymptotic performance analysis of DOA finding algorithms with temporally correlated narrowband signals," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2669–2674, Sep. 2000.

[22] H.-F. Zhou, L. Huang, H. C. So, and J. Li, "Performance analysis of G-MUSIC based DOA estimator with random linear array: A single source case," *Signal Process.*, vol. 142, pp. 513–521, Jan. 2018.

[23] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors. I. The MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1758–1774, Jul. 1992.

[24] A. L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model error. II. Multidimensional algorithms," *IEEE Trans. Signal Process.*, vol. 41, no. 9, pp. 2882–2890, Sep. 1993.

[25] A. Ferreol, P. Larzabal, and M. Viberg, "On the asymptotic performance analysis of subspace DOA estimation in the presence of modeling errors: Case of MUSIC," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 907–920, Mar. 2006.

[26] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Process.*, vol. 40, no. 3, pp. 590–600, Mar. 1992.

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

[28] E. Moulines and J.-F. Cardoso, "Direction finding algorithms using fourth order statistics: Asymptotic performance analysis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, San Francisco, CA, USA, vol. 2, Mar. 1992, pp. 437–440.

[29] J.-F. Cardoso and E. Moulines, "Asymptotic performance analysis of direction-finding algorithms based on fourth-order cumulants," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 214–224, Jan. 1995.

[30] M. Bengtsson and B. Ottersten, "A generalization of weighted subspace fitting to full-rank models," *IEEE Trans. Signal Process.*, vol. 49, no. 5, pp. 1002–1012, May 2001.

[31] B. Ottersten, B. Wahlberg, M. Viberg, and T. Kailath, "Stochastic maximum likelihood estimation in sensor arrays by weighted subspace fitting," in *Proc. 33rd Asilomar Conf. Signals, Syst. Comput.*, Pacific Grove, CA, USA, Nov. 1989, pp. 599–603.

[32] B. Ottersten and M. Viberg, "Analysis of subspace fitting based methods for sensor array processing," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Glasgow, U.K., vol. 4, 1989, pp. 2807–2810.

[33] B. Ottersten and M. Viberg, "Asymptotic robustness of sensor array processing methods," in *Proc. Int. Conf. Acoust., Speech, Signal Process.*, Albuquerque, NM, USA, vol. 5, Apr. 1990, pp. 2635–2638.

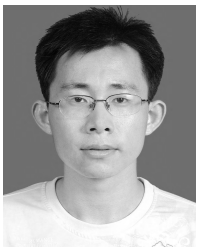
[34] P. Stoica and N. Arye, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 37, no. 5, pp. 720–741, May 1989.

[35] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound: Further results and comparisons," *IEEE Trans. Acoust., Speech Signal Process.*, vol. 38, no. 12, pp. 2140–2150, Dec. 1990.

[36] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 5, pp. 1110–1121, May 1991.
[37] M. Viberg, B. Ottersten, and T. Kailath, "Detection and estimation in sensor arrays using weighted subspace fitting," *IEEE Trans. Signal Process.*, vol. 39, no. 11, pp. 2436–2449, Nov. 1991.
[38] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 122–148, 1963.
[39] G. P. Gupta, "Asymptotic theory for principal component analysis in the complex case," *Int. Indian Stat. Assoc.*, vol. 3, no. 1, pp. 97–106, 1965.
[40] A. Yeredor, A. Weiss, and A. J. Weiss, "High-order analysis of the efficiency gap for maximum likelihood estimation in nonlinear Gaussian models," *IEEE Trans. Signal Process.*, vol. 66, no. 18, pp. 4782–4795, Sep. 2018.
[41] O. Bar-Shalom and A. J. Weiss, "Transponder-aided single platform geolocation," *IEEE Trans. Signal Process.*, vol. 61, no. 5, pp. 1239–1248, Mar. 2013.
[42] O. Bar-Shalom and A. J. Weiss, "Emitter geolocation using single moving receiver," *Signal Process.*, vol. 105, pp. 70–83, Dec. 2014.
[43] R. O. Schmidt, "A signal subspace approach to multiple emitter location spectral estimation," Ph.D. dissertation, Dept. Elect. Eng., Stanford Univ., Stanford, CA, USA, 1981.
[44] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. AP-34, no. 3, pp. 276–280, Mar. 1986.
[45] J. A. Cadzow, "A high resolution direction-of-arrival algorithm for narrowband coherent and incoherent sources," *IEEE Trans. Acoust., Speech Signal Process.*, vol. ASSP-36, no. 7, pp. 965–979, Jul. 1988.
[46] A. W. Davis, "Asymptotic theory for principal component analysis: Non-normal case," *Austral. J. Statist.*, vol. 19, no. 3, pp. 206–212, 1977.
[47] D. J. Jeffries and D. R. Farrier, "Asymptotic results for eigenvector methods," *IEE Proc. F Commun., Radar Signal Process.*, vol. 132, no. 7, pp. 589–594, Dec. 1985.
[48] B. Friedlander and A. J. Weiss, "On the second-order statistics of the eigenvectors of sample covariance matrices," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3136–3139, Nov. 1998.
[49] D. Paul, "Asymptotic of sample eigen structure for a large dimensional spiked covariance model," *Statistica Sinica*, vol. 17, no. 4, pp. 1617–1642, 2007.
[50] K. Sharman, T. Durrani, M. Wax, and T. Kailath, "Asymptotic performance of eigenstructure spectral analysis methods," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, San Diego, CA, USA, Mar. 1984, pp. 440–443.

JIANPING DU received the B.S. degree in applied mathematics and the M.S. degree in operation research from the Zhengzhou Institute of Information Science and Technology, Zhengzhou, in 2003 and 2006, respectively. Since 2006, he has been a Lecturer with the Communication Engineering Department, Zhengzhou Institute of Information Science and Technology. He has authored three books and more than 20 articles. He holds six patents. His research interests include mathematical modeling, high-resolution emitters positioning methods, array signal processing, and short wave emitter positioning.

HONGYI YU received the M.Sc. degree from the Department of Communication Engineering, Zhengzhou Information Science and Technology Institute, China, in 1987, and the Ph.D. degree from Xidian University, in 1999. He is currently a Professor with the National Digital Switching System Engineering and Technological Research Center, Zhengzhou, China. His main research interests are in the areas of wireless communication theory, visible light communications, and signal processing.



DING WANG graduated from the Department of Communication Engineering, Zhengzhou Institute of Information Science and Technology, China, in 2004. He received the M.S. degree in military communication and the Ph.D. degree in communication and system engineering from the Zhengzhou Institute of Information Science and Technology, in 2007 and 2011, respectively. From 2012 to 2015, he was an Assistant Professor with the National Digital Switching System Engineering and Technological Research Center and the Zhengzhou Institute of Information Science and Technology, where he has been an Associate Professor since 2016. His research activities are focused on array signal processing and passive location.

DALONG ZHANG received the Ph.D. degree in communication and information system from Information Engineering University, Henan, China, in 2008. He is currently an Associate Professor with the School of Information Engineering, Zhengzhou University, Henan. His research interests include wireless communications networks, the Internet of Vehicles, sensor networks, the Internet of Things, and satellite positioning systems.

GUANGYI LIU received the B.S. degree in communication engineering from Information Engineering University, Zhengzhou, China, in 2005, and the M.S. and Ph.D. degrees in communication engineering from the National Digital Switching System Engineering and Technological Research and Development Center, Zhengzhou, China, in 2009 and 2015, respectively. His research interests include wireless communications and signal analysis.