An ESPRIT-Based Approach for 2-D Localization of Incoherently Distributed Sources in Massive MIMO Systems

Anzhong Hu, Tiejun Lv, *Senior Member, IEEE*, Hui Gao, *Member, IEEE*,
Zhang Zhang, and Shaoshi Yang, *Member, IEEE*

Abstract—In this paper, an approach of estimating signal parameters via rotational invariance technique (ESPRIT) is proposed for two-dimensional (2-D) localization of incoherently distributed (ID) sources in large-scale/massive multiple-input multiple-output (MIMO) systems. The traditional ESPRIT-based methods are valid only for one-dimensional (1-D) localization of the ID sources. By contrast, in the proposed approach the signal subspace is constructed for estimating the nominal azimuth and elevation direction-of-arrivals and the angular spreads. The proposed estimator enjoys closed-form expressions and hence it bypasses the searching over the entire feasible field. Therefore, it imposes significantly lower computational complexity than the conventional 2-D estimation approaches. Our analysis shows that the estimation performance of the proposed approach improves when the large-scale/massive MIMO systems are employed. The approximate Cramér-Rao bound of the proposed estimator for the 2-D localization is also derived. Numerical results demonstrate that albeit the proposed estimation method is comparable with the traditional 2-D estimators in terms of performance, it benefits from a remarkably lower computational complexity.

Index Terms—Large-scale/massive multiple-input multiple-output (LS-MIMO/massive MIMO), very large arrays, two-dimensional (2-D) localization, direction-of-arrival (DOA), angular spread.

I. INTRODUCTION

Multiple-input multiple-output (MIMO) techniques represent a family of ground-breaking advances in wireless communications during the past two decades. This is because they are capable of providing more degrees of freedom to significantly improve the system's data rate and link reliability [1]. Recently, the massive MIMO, which is also known as the large-scale MIMO, has been attracting increasing attentions owing to its unprecedented potential of high spectral efficiency [2]-[8]. In massive MIMO systems, the base station (BS) is equipped with a hundred or a few hundred antennas, and

A. Hu, T. Lv, and H. Gao are with the Key Laboratory of Trustworthy Distributed Computing and Service, Ministry of Education, and also with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing, China 100876 (email: huanzhong@bupt.edu.cn, lvtiejun@bupt.edu.cn, huigao@bupt.edu.cn).

Z. Zhang is with the Research and Innovation Centre, Alcatel-Lucent Shanghai Bell Co. Ltd., Shanghai, China 201206 (email: chang.zhang@alcatel-sbell.com.cn).

S. Yang is with the School of Electronics and Computer Science, University of Southampton, SO17 1BJ Southampton, U.K. (email: sy7g09@ecs.soton.ac.uk).

serves tens of user terminals (UTs) simultaneously. When the angular spreads are not wide enough, the performance of these systems will degrade significantly, and hence a beamforming approach is proposed for achieving directional antenna gain [9]. Additionally, instead of the one-dimensional linear array, the antenna arrays of the massive MIMO systems are expected to be implemented in more than one dimension because of the constraint concerning the array aperture. Consequently, the beamforming may be required to operate in two dimensions which correspond to the azimuth and elevation directions [2], some examples include the three-dimensional beamforming approaches of [10], [11]. The performance of the beamforming-based systems closely relies on the accuracy of the estimated angular parameters, i.e., the location parameters. For example, $0.1°$ and $0.04°$ estimation errors cause 20 dB and 3 dB reductions of the output signal-to-noise ratio (SNR), respectively [12], [13], and the influence of estimation error becomes significant when the number of the antennas increases [14]. Therefore, as opposed to the one-dimensional (1-D) localization problem where only the azimuth angular parameters need to be estimated, in this paper we focus on the problem of two-dimensional (2-D) localization of distributed sources in the context of the massive MIMO systems, where both the azimuth and elevation angular parameters have to be estimated.

The localization of point sources, i.e., the direction-of-arrival (DOA) estimation, has been of interest to the signal processing community for decades [15]. When the signal of each source emits from a single DOA and the DOAs of all the sources can be distinguished, the sources are assumed to be point sources, and this case corresponds to the line-of-sight transmission scenario [16]. When the signal of each source emits from an angular region, the sources are assumed to be distributed sources, and this case corresponds to the multipath transmission scenario [17]. Obviously, the distributed sources model is more appropriate for cellular wireless systems, where signals are usually transmitted via multipath.

The distributed sources can be categorized into coherently distributed (CD) sources and incoherently distributed (ID) sources [18], which are valid for slowly time-varying channels and rapidly time-varying channels, respectively. In cellular mobile communication systems, rapidly time-varying channels are typically more appropriate to characterize the realistic cir-

cumstances. Additionally, the classical localization approaches for point sources have been successfully generalized to the scenario of the CD sources [16], [18]-[20]. However, the researches on the localization of the ID sources are less adequate [21]-[41]. For example, [21] is entirely limited to the 1-D localization scenario, [22] is only suitable for the single-source localization, and the performance of [23] depends on the accuracy of the initial estimates of location parameters. Therefore, the localization of the ID sources needs to be investigated more extensively. Furthermore, the localization approaches for ID sources can be categorized into parametric approaches and non-parametric approaches. The non-parametric approaches, such as the beamforming approach and the Capon spectrum approach in [24], are shown to perform worse than the parametric ones. Hence, we concentrate on the parametric approaches for ID sources in this paper.

Although most of the traditional parametric approaches are proposed for 1-D localization of the ID sources, some of them can be extended to the 2-D scenario. Among the existing approaches for 2-D localization of the ID sources, the maximum likelihood (ML) estimator of [25] is optimal, while the approximate ML estimator of [26] exhibits suboptimal performance with lower complexity. However, in these ML-based estimators, the 2-D nominal DOAs and angular spreads of all the UTs are estimated by searching exhaustively over the feasible field. The prohibitive complexity makes these estimators infeasible in large-scale systems. Another approximate ML estimator reduces the searching dimension by using the simplified signal model proposed in [27], but this estimator is limited to the single-source assumption [28]. For the sake of reducing the computational complexity, the least-squares (LS) criterion based estimators are proposed by using the covariance matrix matching technique in [25], [29]-[35]. Nevertheless, these estimators are either restricted to the single-source case [29]-[35] or too complicated due to the same search dimension as faced by the approximate ML estimator [25].

On the other hand, the subspace based approaches and the beamforming approaches for localization are of reduced complexity compared with the ML-based approaches and the LS-based approaches, though they are less attractive in performance. Similar to the philosophy of the multiple signal classification method [42], in the subspace based approaches, the signal parameters are estimated by exploiting the fact that the columns of the noise-free covariance matrix of the received signals are orthogonal to those of the pseudonoise subspace [18], [37]-[40]. Additionally, by employing the minimum variance distortionless response beamforming for localization of the ID sources, the generalized Capon estimator is derived [39]. In these approaches, although the 2-D nominal DOAs and angular spreads of only a single UT need to be estimated by searching, their complexity is still very high.

The estimation of signal parameters via rotational invariance technique (ESPRIT) [43]-[46] is also a subspace based approach, and has been employed for the 1-D localization of the ID sources in [19]. However, the method proposed in [19] cannot be extended to 2-D localization owing to the mutual coupling of the 2-D angular parameters. Another ESPRIT based approach [41] decouples the estimation of the 2-D

nominal DOAs by changing the projection of the incident signals. However, in [41] the nominal azimuth DOA still has to be estimated with searching, and the estimation of the angular spreads is not considered. In addition, the approaches proposed in [19], [41] depend on the assumption that the distance between adjacent antennas is much shorter than the wavelength.

In this paper, an ESPRIT-based approach is proposed for 2-D localization of multiple ID sources in the massive MIMO systems employing very large uniform rectangular arrays (URAs). We reveal that the array response matrix is linearly related to the signal subspace. After dividing the URA into three subarrays, the array response matrices of the three subarrays are also shown to be linearly related with each other. Relying on these linear relations, the 2-D nominal DOAs and angular spreads are estimated by the signal subspace. To be more specific, the main contributions of this paper are listed as follows.

1) As opposed to that of the existing works [19], [41], the distance between adjacent antennas is not constrained. In addition, the 2-D angular parameters are decoupled by the proposed algorithm and estimated without searching. These two issues have not been investigated in the existing ESPRIT-based approaches, and the latter is particularly crucial to 2-D localization.
2) The impact of the number of the BS antennas on the performance of the proposed approach is analyzed in the context of the massive MIMO systems. It is proved that the estimated signal subspace tends to be in the same subspace as the array response matrix when the number of the BS antennas increases. Therefore, the estimation performance improves when the number of the BS antennas increases, which is particularly beneficial for the massive MIMO systems.
3) The approximate Cramér-Rao bound (CRB) for the estimation of the 2-D angular parameters is derived, whereas the known CRB is only valid for the estimation of the 1-D angular parameters.
4) It is shown that the proposed approach is of significantly lower complexity than both the LS based covariance matching approaches and the subspace based approaches. This is because the proposed estimator has closed-form expressions. This advantage is particularly attractive in the massive MIMO systems, because the potentially prohibitive computational complexity is one of the major challenges faced by the massive MIMO systems.

The rest of this paper is organized as follows. In Section II, the system model and the major assumptions are given. In section III, we present the proposed ESPRIT-based approach. In Section IV, the analysis of the proposed approach is provided. More specifically, the impact of the number of the BS antennas on the performance is analyzed, and the approximate CRB for the 2-D estimation is derived. In addition, the computational complexity of the proposed approach is compared with that of other well-known approaches. Numerical results are given in Section V, and the conclusions are drawn in Section VI.
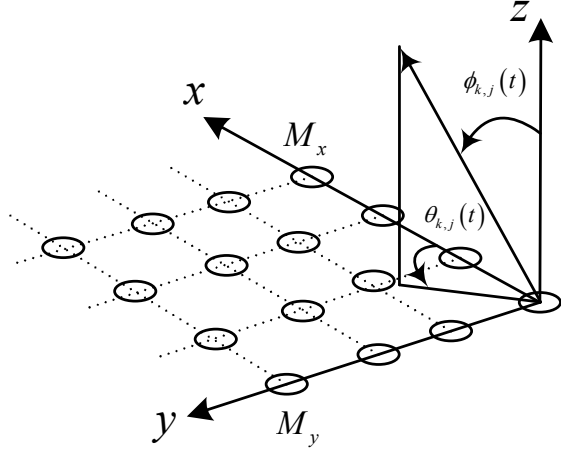
Notations: Lower-case (upper-case) boldface symbols de-

Fig. 1. Array geometry of the URA considered. The direction of the incident path is projected onto the array plane. The angle from the x-axis to the projected line is the azimuth DOA, $\theta_{k,j}(t)$, and the angle from the z-axis to the incident path is the elevation DOA, $\phi_{k,j}(t)$. The range of the two parameters are $0 \leq \theta_{k,j}(t) < \pi$ and $0 \leq \phi_{k,j}(t) < \pi/2$.

note vectors (matrices); $\mathbf{I}_K$ represents the $K \times K$ identity matrix, and $\mathbf{0}_{M \times K}$ represents an $M \times K$ zero matrix; $\mathrm{diag}(\cdot)$ is a diagonal matrix and the values in the brackets are the diagonal elements; $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$, $(\cdot)^\dagger$, and $\mathbb{E}\{\cdot\}$ denote the conjugate, the transpose, the conjugate transpose, the pseudoinverse, and the expectation, respectively; $[\cdot]_{j,k}$, $\mathrm{tr}(\cdot)$, and $||\cdot||_{\mathrm{F}}$ represent the (j,k)th entry, the trace, and the Frobenius norm of a matrix, respectively; $\odot$ is the Hadamard product operator; $[\cdot]_j$ is the jth element of a vector; i is the imaginary unit; and $\delta(\cdot)$ is the Kronecker delta function.

II. SYSTEM MODEL

Consider a URA with $M = M_{\mathrm{x}}M_{\mathrm{y}}$ antennas, where M_{x} and M_{y} are the numbers of antennas in the x-direction and the y-direction, respectively, as shown in Fig. 1.

The transmitted signals of all the UTs are in the same frequency band. In the presence of scattering, the received signal at the antenna array is given by [35]

$$\mathbf{x}(t) = \sum_{k=1}^{K} s_k(t) \sum_{j=1}^{N_k} \gamma_{k,j}(t)\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t)) + \mathbf{n}(t) \in \mathbb{C}^{M \times 1},$$
(1)

where K is the number of the UTs, $s_k(t)$ is the complex-valued signal transmitted by the kth UT, and N_k is the number of multipaths of the kth UT; $t = 1, 2, \cdots, T$ is the sampling time, where T is the number of received signal snapshots; $\gamma_{k,j}(t)$, $\theta_{k,j}(t)$, and $\phi_{k,j}(t)$ are the complex-valued path gain, the real-valued azimuth DOA, and the real-valued elevation DOA of the jth path from the kth UT, respectively, which satisfy $0 \leq \theta_{k,j}(t) < \pi$ and $0 \leq \phi_{k,j}(t) < \pi/2$ as shown in Fig. 1; and $\mathbf{n}(t) \in \mathbb{C}^{M \times 1}$ is the complex-valued additive noise. It should be noted that the ranges of the DOAs are the localization ranges of the array, which means that sources out of these ranges cannot be localized by the array. The array manifold, $\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t)) \in \mathbb{C}^{M \times 1}$, is the response of the array corresponding to the azimuth and elevation DOAs of

$\theta_{k,j}(t)$ and $\phi_{k,j}(t)$. With respect to the antenna at the origin of the axes, the mth element of $\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t))$ is defined as [47]

$$[\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t))]_m = \exp\Big(iu\sin(\phi_{k,j}(t))\big[(m_{\mathrm{x}} - 1) \\ \times \cos(\theta_{k,j}(t)) + (m_{\mathrm{y}} - 1)\sin(\theta_{k,j}(t))\big]\Big), \\ m = (m_{\mathrm{y}} - 1)M_{\mathrm{x}} + m_{\mathrm{x}}, \ m_{\mathrm{x}} = 1, 2, \cdots, M_{\mathrm{x}}, \\ m_{\mathrm{y}} = 1, 2, \cdots, M_{\mathrm{y}},$$
(2)

where $u = 2\pi d/\lambda$, d is the distance between two adjacent antennas, λ is the wavelength. We can see that $[\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t))]_m$ corresponds to the response of the $(m_{\mathrm{x}}, m_{\mathrm{y}})$th antenna element in the coordinate system shown in Fig. 1. The azimuth and elevation DOAs can be expressed as [35]

$$\theta_{k,j}(t) = \bar{\theta}_k + \tilde{\theta}_{k,j}(t), \tag{3}$$
$$\phi_{k,j}(t) = \bar{\phi}_k + \tilde{\phi}_{k,j}(t), \tag{4}$$

where $\bar{\theta}_k$ and $\bar{\phi}_k$ are the real-valued nominal azimuth DOA and the real-valued nominal elevation DOA for the kth UT, and they are the means of $\theta_{k,j}(t)$ and $\phi_{k,j}(t)$, respectively; $\tilde{\theta}_{k,j}(t)$ and $\tilde{\phi}_{k,j}(t)$ are the corresponding real-valued random angular deviations with zero mean and standard deviations σ_{θ_k} and σ_{ϕ_k}, which are referred to as the angular spreads. We emphasize that the task of localization is to estimate the 2-D nominal DOAs, $\bar{\theta}_k$, $\bar{\phi}_k$, and the 2-D angular spreads, σ_{θ_k}, σ_{ϕ_k}, $k = 1, 2, \cdots, K$, with the aid of the received signal snapshots, $\mathbf{x}(t)$, $t = 1, 2, \cdots, T$. Because the signals of the K UTs are transmitted at the same frequency band and the same time, the received snapshot signals from one UT cannot be extracted from $\mathbf{x}(t)$, $t = 1, 2, \cdots, T$, regardless of whether the transmitted signals are pilots or data symbols[1]. As a result, the 2-D angular parameters of the K UTs can only be estimated jointly.

In this paper, the following initial assumptions are considered.
1) The angular deviations, $\tilde{\theta}_{k,j}(t)$ and $\tilde{\phi}_{k,j}(t)$, $k = 1, 2, \cdots, K$, $j = 1, 2, \cdots, N_k$, $t = 1, 2, \cdots, T$, are temporally independent and identically distributed (i.i.d.) Gaussian random variables with covariances

$$\mathbb{E}\left\{\tilde{\theta}_{k,j}(t)\tilde{\theta}_{\tilde{k},\tilde{j}}(\tilde{t})\right\} = \sigma_{\theta_k}^2 \delta(k - \tilde{k})\delta(j - \tilde{j})\delta(t - \tilde{t}), \tag{5}$$

and

$$\mathbb{E}\left\{\tilde{\phi}_{k,j}(t)\tilde{\phi}_{\tilde{k},\tilde{j}}(\tilde{t})\right\} = \sigma_{\phi_k}^2 \delta(k - \tilde{k})\delta(j - \tilde{j})\delta(t - \tilde{t}), \tag{6}$$

respectively, where the angular spreads, σ_{θ_k} and σ_{ϕ_k}, are far less than one.
2) The path gains, $\gamma_{k,j}(t)$, $k = 1, 2, \cdots, K$, $j = 1, 2, \cdots, N_k$, $t = 1, 2, \cdots, T$, are temporally i.i.d. complex-valued zero-mean random variables, whose covariance is

$$\mathbb{E}\left\{\gamma_{k,j}(t)\gamma_{\tilde{k},\tilde{j}}^*(\tilde{t})\right\} = \frac{\sigma_{\gamma_k}^2}{N_k}\delta(k - \tilde{k})\delta(j - \tilde{j})\delta(t - \tilde{t}). \tag{7}$$

---

[1]When the UTs transmit orthogonal pilots, the BS can correlate the received signals with the known pilots of one UT to extract the signal of that UT. Then, the BS only obtains a rank-1 covariance matrix which is not capable of performing the 2-D localization of the UT.

Note that if the path gain factors of different paths are uncorrelated, the sources are said to be ID [18].

3) The noise, $\mathbf{n}(t)$, $t = 1, 2, \cdots, T$ are composed of temporally and spatially i.i.d. complex-valued circularly symmetric zero-mean Gaussian variables, whose covariance matrix is given by

$$\mathbb{E}\left\{\mathbf{n}(t)\mathbf{n}^H(\tilde{t})\right\} = \sigma_{\mathrm{n}}^2 \mathbf{I}_M \delta(t - \tilde{t}). \tag{8}$$

4) The transmitted signals, $s_k(t)$, $k = 1, 2, \cdots, K$, $t = 1, 2, \cdots, T$, are modeled as deterministic ones with constant absolute values, and we denote

$$S_k = |s_k(t)|^2 \tag{9}$$

as the transmitted signal power of the kth UT.

5) The angular deviations, the path gains, the noise, and the transmitted signals are uncorrelated from each other.

6) The array is calibrated, which means the response of the array is known. Hence, the array manifold for any 2-D DOAs, cf. (2), is known *a priori*. The number of the BS antennas M is much larger than the number of the UTs K.

7) The number of multipaths $N_k, \forall k$, is large.

With these assumptions and using the central limit theorem, it can be verified that the received signal vector $\mathbf{x}(t)$ in (1) is a zero-mean circularly symmetric complex-valued Gaussian vector [22], [25], [30], [48].

III. THE ESPRIT-BASED APPROACH

The existing subspace based and covariance matching approaches are complicated for the 2-D localization in the massive MIMO systems due to the exhausted multidimensional search for estimating the angular parameters. Although the traditional 1-D ESPRIT-based approach avoids searching over the parameter space [19], the angular parameters are mutually coupled when this approach is employed in the 2-D localization straightforwardly. The existing 2-D ESPRIT-based approach decouples the angular parameters, but the azimuth nominal DOA is still estimated with searching, and the angular spreads are not estimated [41]. Hence, in this section, the expression of the signal subspace is first derived, which is the foundation of the ESPRIT-based approaches. Then, the signal subspace based ESPRIT approach is proposed for estimating the 2-D angular parameters without searching.

A. The Signal Subspace

It can be seen that the array manifold $\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t))$ in (2) is a function of the azimuth and elevation DOAs. With the first order Taylor series expansion of $\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t))$ around the nominal DOAs, $\bar{\theta}_k, \bar{\phi}_k$, it can be approximated as

$$\begin{aligned} \mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t)) &= \mathbf{a}(\bar{\theta}_k + \tilde{\theta}_{k,j}(t), \bar{\phi}_k + \tilde{\phi}_{k,j}(t)) \\ &\approx \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k) + \frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k} \tilde{\theta}_{k,j}(t) \\ &\quad + \frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\phi}_k} \tilde{\phi}_{k,j}(t), \end{aligned} \tag{10}$$

where the remainder of the series is omitted. It is assumed that the standard deviations of $\tilde{\theta}_{k,j}(t)$ and $\tilde{\phi}_{k,j}(t)$, i.e., σ_{θ_k} and σ_{ϕ_k}, are sufficiently small. Thus, the approximation is almost

true. Then, the received signal given by (1) can be rewritten as

$$\begin{aligned} \mathbf{x}(t) &\approx \sum_{k=1}^{K} \left(\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k) c_{k,1}(t) + \frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k} c_{k,2}(t) \right. \\ &\quad \left. + \frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\phi}_k} c_{k,3}(t) \right) + \mathbf{n}(t) \in \mathbb{C}^{M \times 1}, \quad \tag{11} \end{aligned}$$

where

$$\begin{aligned} c_{k,1}(t) &= s_k(t) \sum_{j=1}^{N_k} \gamma_{k,j}(t), \\ c_{k,2}(t) &= s_k(t) \sum_{j=1}^{N_k} \gamma_{k,j}(t) \tilde{\theta}_{k,j}(t), \end{aligned}$$

and

$$c_{k,3}(t) = s_k(t) \sum_{j=1}^{N_k} \gamma_{k,j}(t) \tilde{\phi}_{k,j}(t).$$

As a result, if $\mathbf{n}(t)$ is not taken into account, the received signal is linearly related to the array manifold $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$ and its partial derivatives. Therefore, it can be concisely expressed as

$$\mathbf{x}(t) \approx \mathbf{A}\mathbf{c}(t) + \mathbf{n}(t), \tag{12}$$

where

$$\begin{aligned} \mathbf{A} = &\left[\mathbf{a}(\bar{\theta}_1, \bar{\phi}_1), \mathbf{a}(\bar{\theta}_2, \bar{\phi}_2), \cdots, \mathbf{a}(\bar{\theta}_K, \bar{\phi}_K), \right. \\ &\frac{\partial \mathbf{a}(\bar{\theta}_1, \bar{\phi}_1)}{\partial \bar{\theta}_1}, \frac{\partial \mathbf{a}(\bar{\theta}_2, \bar{\phi}_2)}{\partial \bar{\theta}_2}, \cdots, \frac{\partial \mathbf{a}(\bar{\theta}_K, \bar{\phi}_K)}{\partial \bar{\theta}_K}, \\ &\left. \frac{\partial \mathbf{a}(\bar{\theta}_1, \bar{\phi}_1)}{\partial \bar{\phi}_1}, \frac{\partial \mathbf{a}(\bar{\theta}_2, \bar{\phi}_2)}{\partial \bar{\phi}_2}, \cdots, \frac{\partial \mathbf{a}(\bar{\theta}_K, \bar{\phi}_K)}{\partial \bar{\phi}_K} \right] \in \mathbb{C}^{M \times 3K} \end{aligned} \tag{13}$$

denotes the array response matrix of the URA, and

$$\begin{aligned} \mathbf{c}(t) = &[c_{1,1}(t), c_{2,1}(t), \cdots, c_{K,1}(t), c_{1,2}(t), c_{2,2}(t), \cdots, \\ &c_{K,2}(t), c_{1,3}(t), c_{2,3}(t), \cdots, c_{K,3}(t)]^T \in \mathbb{C}^{3K \times 1}. \end{aligned}$$

It should be noted that $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$ is obtained by changing the DOAs, $\theta_{k,j}(t), \phi_{k,j}(t)$, in (2) to the nominal DOAs, $\bar{\theta}_k, \bar{\phi}_k$. We can see that $\mathbf{A}$ is only determined by the nominal DOAs, $\bar{\theta}_k, \bar{\phi}_k, k = 1, 2, \cdots, K$. Thus, these nominal DOAs might be obtained from $\mathbf{A}$.

Based on the properties of $\tilde{\theta}_{k,j}(t), \tilde{\phi}_{k,j}(t), \gamma_{k,j}(t)$, and $s_k(t)$ that are given in (5), (6), (7), and (9), respectively, and the assumption that the transmitted signals, the path gains, and the angular deviations are uncorrelated from each other, the variances of $c_{k,1}(t)$, $c_{k,2}(t)$, and $c_{k,3}(t)$ are obtained as

$$\begin{aligned} \mathbb{E}\left\{c_{k,1}(t)c_{k,1}^*(t)\right\} &= S_k \sigma_{\gamma_k}^2, \tag{14} \\ \mathbb{E}\left\{c_{k,2}(t)c_{k,2}^*(t)\right\} &= S_k \sigma_{\gamma_k}^2 \sigma_{\theta_k}^2, \tag{15} \end{aligned}$$

and

$$\mathbb{E}\left\{c_{k,3}(t)c_{k,3}^*(t)\right\} = S_k \sigma_{\gamma_k}^2 \sigma_{\phi_k}^2, \tag{16}$$

respectively. Additionally, the covariance is

$$\mathbb{E}\left\{c_{k,l}(t)c_{\tilde{k},\tilde{l}}^*(t)\right\} = 0, \forall k \neq \tilde{k}, \text{or } l \neq \tilde{l}. \tag{17}$$

Therefore, the covariance matrix of $\mathbf{c}(t)$ is

$$\boldsymbol{\Lambda}_{\mathrm{c}} = \mathbb{E}\left\{\mathbf{c}(t)\mathbf{c}^H(t)\right\} \in \mathbb{R}^{3K \times 3K}, \qquad (18)$$

which is a diagonal matrix with $[\boldsymbol{\Lambda}_{\mathrm{c}}]_{k,k} = S_k\sigma_{\gamma_k}^2$, $[\boldsymbol{\Lambda}_{\mathrm{c}}]_{K+k,K+k} = [\boldsymbol{\Lambda}_{\mathrm{c}}]_{k,k}\sigma_{\theta_k}^2$, $[\boldsymbol{\Lambda}_{\mathrm{c}}]_{2K+k,2K+k} = [\boldsymbol{\Lambda}_{\mathrm{c}}]_{k,k}\sigma_{\phi_k}^2$, $k = 1, 2, \cdots, K$. Therefore, the angular spreads, $\sigma_{\theta_k}, \sigma_{\phi_k}, k = 1, 2, \cdots, K$, can be obtained from $\boldsymbol{\Lambda}_{\mathrm{c}}$.

From (12), we can see that $\mathbf{A}$ and $\boldsymbol{\Lambda}_{\mathrm{c}}$ might be obtained from the covariance matrix of $\mathbf{x}(t)$. Since the signal and the noise are uncorrelated from each other, and satisfy (8) and (18), the covariance matrix of the received signal $\mathbf{x}(t)$ given by (12) is thus expressed as

$$\mathbf{R}_{\mathbf{x}} = \mathbb{E}\left\{\mathbf{x}(t)\mathbf{x}^H(t)\right\} \approx \mathbf{A}\boldsymbol{\Lambda}_{\mathrm{c}}\mathbf{A}^H + \sigma_{\mathrm{n}}^2\mathbf{I}_M \in \mathbb{C}^{M \times M}. \quad (19)$$

It can be seen that $\mathbf{R}_{\mathbf{x}}$ is a normal matrix, i.e., $\mathbf{R}_{\mathbf{x}}\mathbf{R}_{\mathbf{x}}^H = \mathbf{R}_{\mathbf{x}}^H\mathbf{R}_{\mathbf{x}}$. Because $\mathbf{A}\boldsymbol{\Lambda}_{\mathrm{c}}\mathbf{A}^H$ is positive semi-definite and $\sigma_{\mathrm{n}}^2 > 0$, $\mathbf{R}_{\mathbf{x}}$ is positive definite. Thus, the eigenvalue-decomposition (EVD) of $\mathbf{R}_{\mathbf{x}}$ is also the singular value decomposition of $\mathbf{R}_{\mathbf{x}}$. Let $\mathbf{A}$ be a full rank matrix. Then, the largest $3K$ eigenvalues of $\mathbf{R}_{\mathbf{x}}$ are larger than σ_{n}^2, and the other $M - 3K$ eigenvalues of $\mathbf{R}_{\mathbf{x}}$ approximately equal σ_{n}^2. In the next section, we will prove that for the massive MIMO systems $\mathbf{A}$ is indeed a full rank matrix. Hence, the EVD of $\mathbf{R}_{\mathbf{x}}$ can be written as

$$\mathbf{R}_{\mathbf{x}} \approx [\mathbf{E}_{\mathrm{s}}, \mathbf{E}_{\mathrm{n}}] \begin{bmatrix} \boldsymbol{\Sigma}_{\mathrm{s}} & \mathbf{0}_{3K \times (M-3K)} \\ \mathbf{0}_{(M-3K) \times 3K} & \sigma_{\mathrm{n}}^2\mathbf{I}_{M-3K} \end{bmatrix} [\mathbf{E}_{\mathrm{s}}, \mathbf{E}_{\mathrm{n}}]^H$$
$$= \mathbf{E}_{\mathrm{s}}\boldsymbol{\Sigma}_{\mathrm{s}}\mathbf{E}_{\mathrm{s}}^H + \sigma_{\mathrm{n}}^2\mathbf{E}_{\mathrm{n}}\mathbf{E}_{\mathrm{n}}^H, \qquad (20)$$

where $\mathbf{E}_{\mathrm{s}} \in \mathbb{C}^{M \times 3K}$ and $\mathbf{E}_{\mathrm{n}} \in \mathbb{C}^{M \times (M-3K)}$ are composed of the eigenvectors of $\mathbf{R}_{\mathbf{x}}$, and $\boldsymbol{\Sigma}_{\mathrm{s}} \in \mathbb{R}^{3K \times 3K}$ is a diagonal matrix comprising the largest $3K$ eigenvalues of $\mathbf{R}_{\mathbf{x}}$. It can be seen that $[\mathbf{E}_{\mathrm{s}}, \mathbf{E}_{\mathrm{n}}] \in \mathbb{C}^{M \times M}$ is a unitary matrix, which satisfies

$$\mathbf{I}_M = [\mathbf{E}_{\mathrm{s}}, \mathbf{E}_{\mathrm{n}}][\mathbf{E}_{\mathrm{s}}, \mathbf{E}_{\mathrm{n}}]^H = \mathbf{E}_{\mathrm{s}}\mathbf{E}_{\mathrm{s}}^H + \mathbf{E}_{\mathrm{n}}\mathbf{E}_{\mathrm{n}}^H,$$

which means that

$$\mathbf{E}_{\mathrm{n}}\mathbf{E}_{\mathrm{n}}^H = \mathbf{I}_M - \mathbf{E}_{\mathrm{s}}\mathbf{E}_{\mathrm{s}}^H. \qquad (21)$$

Hence, substituting (21) into (20) yields

$$\mathbf{R}_{\mathbf{x}} \approx \mathbf{E}_{\mathrm{s}}\tilde{\boldsymbol{\Sigma}}_{\mathrm{s}}\mathbf{E}_{\mathrm{s}}^H + \sigma_{\mathrm{n}}^2\mathbf{I}_M, \qquad (22)$$

where $\tilde{\boldsymbol{\Sigma}}_{\mathrm{s}} = \boldsymbol{\Sigma}_{\mathrm{s}} - \sigma_{\mathrm{n}}^2\mathbf{I}_{3K} \in \mathbb{R}^{3K \times 3K}$. Then, from (19) and (22), we obtain

$$\mathbf{A}\boldsymbol{\Lambda}_{\mathrm{c}}\mathbf{A}^H \approx \mathbf{E}_{\mathrm{s}}\tilde{\boldsymbol{\Sigma}}_{\mathrm{s}}\mathbf{E}_{\mathrm{s}}^H. \qquad (23)$$

It is known that the diagonal elements of $\boldsymbol{\Sigma}_{\mathrm{s}}$ are larger than σ_{n}^2, which means $\tilde{\boldsymbol{\Sigma}}_{\mathrm{s}}$ has full rank. Hence, according to the definition of subspace, $\mathbf{E}_{\mathrm{s}}$ and $\mathbf{A}$ are approximately in the same subspace, i.e.,

$$\mathbf{A} \approx \mathbf{E}_{\mathrm{s}}\mathbf{T}, \qquad (24)$$

where $\mathbf{T} \in \mathbb{C}^{3K \times 3K}$ is a full rank matrix. Additionally, $\mathbf{E}_{\mathrm{s}}$ and $\mathbf{E}_{\mathrm{n}}$ are termed as the signal subspace and the noise subspace, respectively. It is obvious that the signal subspace can be obtained from the received signal snapshots $\mathbf{x}(t)$, $t = 1, 2, \cdots, T$, and $\mathbf{A}$ is linearly related to the signal subspace. Hence, the linear relation in (24) will be used for estimating the nominal DOAs, and the estimation approach will be given in the next subsection.
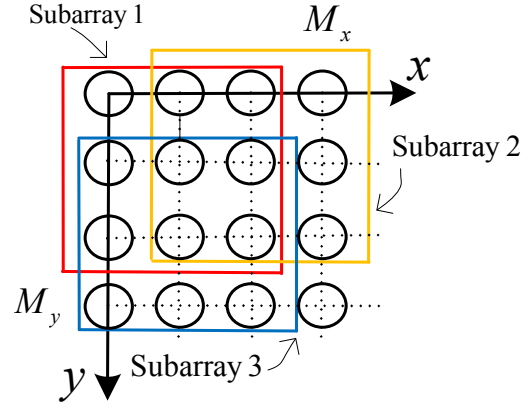

Fig. 2. Subarrays of the URA considered. Subarray 2 is the shift of Subarray 1 in the x-direction with distance d, and Subarray 3 is the shift of Subarray 1 in the y-direction with distance d.

B. The Proposed Estimator

Similar to the practice in the general ESPRIT methods, the antenna array is divided into several subarrays in the proposed estimator as well. Then, the linear relations between the array response matrices of the subarrays can be tactfully constructed for estimating angular parameters. For the estimation of both the elevation and azimuth nominal DOAs, which are coupled in the array manifold, the array has to be divided into at least three subarrays to decouple the 2-D nominal DOAs. This is because obtaining the 2-D nominal DOAs needs at least two different functions of them, which can only be derived from at least two different linear relations between the subarrays, and at least three subarrays are needed to obtain the two linear relations. Although the URA can be divided into more than three subarrays, the computational complexity of estimation increases when the number of the subarrays increases, which constitutes one of the main challenges in the context of the massive MIMO systems. In addition, only one antenna is not used with the three-subarray division, which is rather small in comparison with the total number of antennas M. Therefore, the URA is divided into three subarrays, as shown in Fig. 2. Thus, the proposed approach uses almost all of the BS antennas with low computational complexity.

In order to obtain the linear relations between the array response matrices of the subarrays, these array response matrices need to be derived. From (13), it can be seen that the array response matrix $\mathbf{A}$ is constructed by the array manifold $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$ and its partial derivatives. Similarly, the array response matrix of each subarray is also constructed by the array manifold of the subarray and its partial derivatives. Hence, the array manifold of each subarray and its partial derivatives will be derived first. The array manifold of the lth subarray corresponding to the nominal DOAs $\bar{\theta}_k, \bar{\phi}_k$, cf. (13), is denoted as $\mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k) \in \mathbb{C}^{\tilde{M} \times 1}$, $l = 1, 2, 3$, where $\tilde{M} = (M_{\mathrm{x}} - 1)(M_{\mathrm{y}} - 1)$. Note that $\mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k)$ is obtained by selecting the elements of $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$ that correspond to the lth subarray and keeping these selected elements in the same order as in $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$. In other words, $\mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k)$ can be written

as

$$\mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k) = \mathbf{J}_l \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k), \qquad (25)$$

where $\mathbf{J}_l \in \mathbb{R}^{\tilde{M} \times M}$ is the selection matrix that assigns the elements of $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$ to the lth subarray, and is defined as

$$[\mathbf{J}_l]_{\tilde{m}, \tilde{n}} = \begin{cases} 1, & \tilde{n} = \tilde{m} + \lfloor \frac{\tilde{m}-1}{M_{\mathrm{x}}-1} \rfloor + d_l, \tilde{m} = 1, 2, \cdots, \tilde{M}, \\ 0, & \text{otherwise}, \end{cases}$$

in which $d_1 = 0$, $d_2 = 1$, and $d_3 = M_{\mathrm{x}}$. In the above equation, the floor operator makes $\lfloor \frac{\tilde{m}-1}{M_{\mathrm{x}}-1} \rfloor = n, \forall n = 0, 1, \cdots, M_{\mathrm{y}} - 2$ when $\tilde{m} = n(M_{\mathrm{x}}-1)+1, n(M_{\mathrm{x}}-1)+2, \cdots, (n+1)(M_{\mathrm{x}}-1)$. It can be seen that for the $\tilde{m}$th row of $\mathbf{J}_l$, only the $(\tilde{m} + \lfloor \frac{\tilde{m}-1}{M_{\mathrm{x}}-1} \rfloor + d_l)$th entry is one, and the other entries are zeros. Thus, $\mathbf{J}_l$ assigns the $(\tilde{m} + \lfloor \frac{\tilde{m}-1}{M_{\mathrm{x}}-1} \rfloor + d_l)$th entry of $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$ to the $\tilde{m}$th entry of $\mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k)$, and this coincides with the relation between the subarrays and the URA. From the definition of $\mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k)$, it can be found that the array manifolds of different subarrays are linearly related as

$$\mathbf{a}_q(\bar{\theta}_k, \bar{\phi}_k) = F_q(\bar{\theta}_k, \bar{\phi}_k) \mathbf{a}_1(\bar{\theta}_k, \bar{\phi}_k), \qquad (26)$$

where $q = 2, 3$, and

$$F_2(\bar{\theta}_k, \bar{\phi}_k) = \exp(iu \sin(\bar{\phi}_k)\cos(\bar{\theta}_k)), \qquad (27)$$
$$F_3(\bar{\theta}_k, \bar{\phi}_k) = \exp(iu \sin(\bar{\phi}_k)\sin(\bar{\theta}_k)). \qquad (28)$$

Note that $F_2(\bar{\theta}_k, \bar{\phi}_k)$ and $F_3(\bar{\theta}_k, \bar{\phi}_k)$ are two different functions of the 2-D nominal DOAs, $\bar{\theta}_k$, $\bar{\phi}_k$, and can be exploited to estimate these nominal DOAs. After computing the partial derivatives of $\mathbf{a}_q(\bar{\theta}_k, \bar{\phi}_k)$, we can see that they are related to $\mathbf{a}_1(\bar{\theta}_k, \bar{\phi}_k)$ and its partial derivatives as

$$\frac{\partial \mathbf{a}_q(\bar{\theta}_k, \bar{\phi}_k)}{\partial \theta} = F_q(\bar{\theta}_k, \bar{\phi}_k) \frac{\partial \mathbf{a}_1(\bar{\theta}_k, \bar{\phi}_k)}{\partial \theta} + \frac{\partial F_q(\bar{\theta}_k, \bar{\phi}_k)}{\partial \theta} \mathbf{a}_1(\bar{\theta}_k, \bar{\phi}_k), \qquad (29)$$

and

$$\frac{\partial \mathbf{a}_q(\bar{\theta}_k, \bar{\phi}_k)}{\partial \phi} = F_q(\bar{\theta}_k, \bar{\phi}_k) \frac{\partial \mathbf{a}_1(\bar{\theta}_k, \bar{\phi}_k)}{\partial \phi} + \frac{\partial F_q(\bar{\theta}_k, \bar{\phi}_k)}{\partial \phi} \mathbf{a}_1(\bar{\theta}_k, \bar{\phi}_k). \qquad (30)$$

In the existing ESPRIT-based approaches [19], [41], the partial derivatives, $\partial F_q(\bar{\theta}_k, \bar{\phi}_k)/\partial \theta$ and $\partial F_q(\bar{\theta}_k, \bar{\phi}_k)/\partial \phi$, are approximated as zero, which is based on the assumption that the distance between adjacent antennas d is much shorter than the wavelength λ. In fact, d might not satisfy this assumption. These partial derivatives in (29) and (30) do not vanish. Therefore, this restriction on d is not needed in our derivation.

By replacing $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$, $\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)/\partial \bar{\theta}_k$, and $\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)/\partial \bar{\phi}_k$ in (13) with $\mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k)$, $\partial \mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k)/\partial \bar{\theta}_k$, and $\partial \mathbf{a}_l(\bar{\theta}_k, \bar{\phi}_k)/\partial \bar{\phi}_k$, respectively, the array response matrix of the lth subarray is expressed as

$$\mathbf{A}_l = \left[\mathbf{a}_l(\bar{\theta}_1, \bar{\phi}_1), \mathbf{a}_l(\bar{\theta}_2, \bar{\phi}_2), \cdots, \mathbf{a}_l(\bar{\theta}_K, \bar{\phi}_K), \right.$$

$$\frac{\partial \mathbf{a}_l(\bar{\theta}_1, \bar{\phi}_1)}{\partial \bar{\theta}_1}, \frac{\partial \mathbf{a}_l(\bar{\theta}_2, \bar{\phi}_2)}{\partial \bar{\theta}_2}, \cdots, \frac{\partial \mathbf{a}_l(\bar{\theta}_K, \bar{\phi}_K)}{\partial \bar{\theta}_K},$$

$$\left. \frac{\partial \mathbf{a}_l(\bar{\theta}_1, \bar{\phi}_1)}{\partial \bar{\phi}_1}, \frac{\partial \mathbf{a}_l(\bar{\theta}_2, \bar{\phi}_2)}{\partial \bar{\phi}_2}, \cdots, \frac{\partial \mathbf{a}_l(\bar{\theta}_K, \bar{\phi}_K)}{\partial \bar{\phi}_K} \right] \in \mathbb{C}^{\tilde{M} \times 3K}. (31)$$

Then, from (26) and (29)-(31), we can see that the array response matrices of the subarrays are linearly related as

$$\mathbf{A}_q = \mathbf{A}_1 \mathbf{\Phi}_{q,1}, \qquad (32)$$

where

$$\mathbf{\Phi}_{q,1} = \begin{bmatrix} \mathbf{\Lambda}_{q,1} & \mathbf{\Lambda}_{q,2} & \mathbf{\Lambda}_{q,3} \\ \mathbf{0}_{K \times K} & \mathbf{\Lambda}_{q,1} & \mathbf{0}_{K \times K} \\ \mathbf{0}_{K \times K} & \mathbf{0}_{K \times K} & \mathbf{\Lambda}_{q,1} \end{bmatrix} \in \mathbb{C}^{3K \times 3K}, \qquad (33)$$

$$\mathbf{\Lambda}_{q,1} = \mathrm{diag}\left(F_q(\bar{\theta}_1, \bar{\phi}_1), F_q(\bar{\theta}_2, \bar{\phi}_2), \cdots, F_q(\bar{\theta}_K, \bar{\phi}_K) \right) \in \mathbb{C}^{K \times K},$$

$$\mathbf{\Lambda}_{q,2} = \mathrm{diag}\left(\frac{\partial F_q(\bar{\theta}_1, \bar{\phi}_1)}{\partial \bar{\theta}_1}, \frac{\partial F_q(\bar{\theta}_2, \bar{\phi}_2)}{\partial \bar{\theta}_2}, \cdots, \frac{\partial F_q(\bar{\theta}_K, \bar{\phi}_K)}{\partial \bar{\theta}_K} \right) \in \mathbb{C}^{K \times K},$$

and

$$\mathbf{\Lambda}_{q,3} = \mathrm{diag}\left(\frac{\partial F_q(\bar{\theta}_1, \bar{\phi}_1)}{\partial \bar{\phi}_1}, \frac{\partial F_q(\bar{\theta}_2, \bar{\phi}_2)}{\partial \bar{\phi}_2}, \cdots, \frac{\partial F_q(\bar{\theta}_K, \bar{\phi}_K)}{\partial \bar{\phi}_K} \right) \in \mathbb{C}^{K \times K}.$$

From (27), (28), and (33), we know that the diagonal elements of $\mathbf{\Phi}_{q,1}$ are functions of the 2-D nominal DOAs and can be expressed as

$$[\mathbf{\Phi}_{2,1}]_{k+(l-1)K, k+(l-1)K} = \exp(iu \sin(\bar{\phi}_k)\cos(\bar{\theta}_k)), (34)$$
$$[\mathbf{\Phi}_{3,1}]_{k+(l-1)K, k+(l-1)K} = \exp(iu \sin(\bar{\phi}_k)\sin(\bar{\theta}_k)), (35)$$

where $l = 1, 2, 3$. Hence, the diagonal elements of $\mathbf{\Phi}_{q,1}$ will be used for estimating the nominal DOAs.

On the other hand, the array response matrix $\mathbf{A}_l$ of the lth subarray is also linearly related to the signal subspace $\mathbf{E}_\mathrm{s}$. By substituting (25) into (31), the array response matrix of the lth subarray is expressed as

$$\mathbf{A}_l = \mathbf{J}_l \mathbf{A} \qquad (36)$$
$$\approx \mathbf{J}_l \mathbf{E}_\mathrm{s} \mathbf{T} \qquad (37)$$
$$= \mathbf{E}_l \mathbf{T}, \quad l = 1, 2, 3, \qquad (38)$$

where (37) is derived by substituting (24) into (36), and

$$\mathbf{E}_l \triangleq \mathbf{J}_l \mathbf{E}_\mathrm{s} \in \mathbb{C}^{\tilde{M} \times 3K}, \quad l = 1, 2, 3, \qquad (39)$$

are termed as the selected signal subspaces. It can be seen that the array response matrix $\mathbf{A}_l$ of the lth subarray, cf. (36), and the selected signal subspace $\mathbf{E}_l$ of the lth subarray, cf. (39), are selected in the same way. Because the signal subspace $\mathbf{E}_\mathrm{s}$ and the array response matrix $\mathbf{A}$ are linearly related, cf. (24), we discover that the selected signal subspace $\mathbf{E}_l$ and the array response matrix $\mathbf{A}_l$ of the lth subarray are linearly related. Therefore, it is proved that $\mathbf{E}_l, l = 1, 2, 3$, are linearly related with each other in a similar way to $\mathbf{A}_l$ in (32), which is exploited to obtain the diagonal elements of $\mathbf{\Phi}_{2,1}$ and $\mathbf{\Phi}_{3,1}$. These diagonal elements are different functions of the 2-D nominal DOAs, cf. (33).

Because only the selected signal subspace $\mathbf{E}_l$ can be obtained from the received signal snapshots, $\mathbf{A}_l$ in (32) needs to be written as the linear transformation of $\mathbf{E}_l$ for obtaining

the diagonal elements of $\mathbf{\Phi}_{2,1}$ and $\mathbf{\Phi}_{3,1}$. Combining (32) and (38), we get

$$\mathbf{A}_1 \approx \mathbf{E}_1\mathbf{T}, \tag{40}$$

$$\mathbf{A}_1\mathbf{\Phi}_{2,1} \approx \mathbf{E}_2\mathbf{T}, \tag{41}$$

and

$$\mathbf{A}_1\mathbf{\Phi}_{3,1} \approx \mathbf{E}_3\mathbf{T}. \tag{42}$$

Note that $\mathbf{A}_1$ is linearly related to all the selected signal subspaces. Substituting (40) into (41) and (42) yields

$$\mathbf{E}_1\mathbf{\Psi}_1 \approx \mathbf{E}_2, \tag{43}$$

and

$$\mathbf{E}_1\mathbf{\Psi}_2 \approx \mathbf{E}_3, \tag{44}$$

where

$$\mathbf{\Psi}_1 = \mathbf{T}\mathbf{\Phi}_{2,1}\mathbf{T}^{-1} \in \mathbb{C}^{3K \times 3K}, \tag{45}$$

and

$$\mathbf{\Psi}_2 = \mathbf{T}\mathbf{\Phi}_{3,1}\mathbf{T}^{-1} \in \mathbb{C}^{3K \times 3K}. \tag{46}$$

Obviously, the diagonal elements of $\mathbf{\Phi}_{2,1}$ and $\mathbf{\Phi}_{3,1}$ are the eigenvalues of $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$, respectively, which is because $\mathbf{\Phi}_{2,1}$ and $\mathbf{\Phi}_{3,1}$ are upper triangular matrices. Therefore, in order to estimate the diagonal elements of $\mathbf{\Phi}_{2,1}$ and $\mathbf{\Phi}_{3,1}$, $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ need to be estimated from the selected signal subspaces $\mathbf{E}_l, l = 1, 2, 3$. According to (43) and (44), they can be obtained by employing the well-known total least-squares (TLS) criterion [45]. First, compute the EVD as

$$[\mathbf{E}_1, \mathbf{E}_2]^H[\mathbf{E}_1, \mathbf{E}_2] = \mathbf{E}_{\mathrm{x}}\mathbf{\Lambda}_{\mathrm{x}}\mathbf{E}_{\mathrm{x}}^H \in \mathbb{C}^{6K \times 6K}, \tag{47}$$

$$[\mathbf{E}_1, \mathbf{E}_3]^H[\mathbf{E}_1, \mathbf{E}_3] = \mathbf{E}_{\mathrm{y}}\mathbf{\Lambda}_{\mathrm{y}}\mathbf{E}_{\mathrm{y}}^H \in \mathbb{C}^{6K \times 6K}, \tag{48}$$

where the columns of $\mathbf{E}_{\mathrm{x}} \in \mathbb{C}^{6K \times 6K}$ and $\mathbf{E}_{\mathrm{y}} \in \mathbb{C}^{6K \times 6K}$ are the eigenvectors of the left-hand side matrices of (47) and (48), respectively, while the diagonal elements of $\mathbf{\Lambda}_{\mathrm{x}} \in \mathbb{C}^{6K \times 6K}$ and $\mathbf{\Lambda}_{\mathrm{y}} \in \mathbb{C}^{6K \times 6K}$ are their respective eigenvalues, which are placed in descending order from the upper left corner. Then, $\mathbf{E}_{\mathrm{x}}$ and $\mathbf{E}_{\mathrm{y}}$ are partitioned as

$$\mathbf{E}_{\mathrm{x}} = \begin{bmatrix} \mathbf{E}_{\mathrm{x}11} & \mathbf{E}_{\mathrm{x}12} \\ \mathbf{E}_{\mathrm{x}21} & \mathbf{E}_{\mathrm{x}22} \end{bmatrix}, \quad \mathbf{E}_{\mathrm{y}} = \begin{bmatrix} \mathbf{E}_{\mathrm{y}11} & \mathbf{E}_{\mathrm{y}12} \\ \mathbf{E}_{\mathrm{y}21} & \mathbf{E}_{\mathrm{y}22} \end{bmatrix}, \tag{49}$$

where $\mathbf{E}_{\mathrm{x}ab} \in \mathbb{C}^{3K \times 3K}, \mathbf{E}_{\mathrm{y}ab} \in \mathbb{C}^{3K \times 3K}, a, b = 1, 2$. Finally, $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ can be estimated as

$$\hat{\mathbf{\Psi}}_1 = -\mathbf{E}_{\mathrm{x}12}\mathbf{E}_{\mathrm{x}22}^{-1} \in \mathbb{C}^{3K \times 3K}, \tag{50}$$

$$\hat{\mathbf{\Psi}}_2 = -\mathbf{E}_{\mathrm{y}12}\mathbf{E}_{\mathrm{y}22}^{-1} \in \mathbb{C}^{3K \times 3K}, \tag{51}$$

and we have $\hat{\mathbf{\Psi}}_1 \approx \mathbf{\Psi}_1$, $\hat{\mathbf{\Psi}}_2 \approx \mathbf{\Psi}_2$.

For estimating the nominal DOAs, we calculate the EVD of $\hat{\mathbf{\Psi}}_1$ and $\hat{\mathbf{\Psi}}_2$ as

$$\hat{\mathbf{\Psi}}_1 = \mathbf{T}_1\mathbf{\Lambda}_1\mathbf{T}_1^{-1}, \tag{52}$$

$$\hat{\mathbf{\Psi}}_2 = \mathbf{T}_2\mathbf{\Lambda}_2\mathbf{T}_2^{-1}, \tag{53}$$

where $\mathbf{T}_1 \in \mathbb{C}^{3K \times 3K}$ and $\mathbf{T}_2 \in \mathbb{C}^{3K \times 3K}$ are composed of the eigenvectors of $\hat{\mathbf{\Psi}}_1$ and $\hat{\mathbf{\Psi}}_2$, respectively, while $\mathbf{\Lambda}_1 \in \mathbb{C}^{3K \times 3K}$ and $\mathbf{\Lambda}_2 \in \mathbb{C}^{3K \times 3K}$ are diagonal matrices whose diagonal elements are the corresponding eigenvalues, which are placed in descending order from the upper left corner. From the previous analysis, the diagonal elements of $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ can

be taken as the estimates of the diagonal elements of $\mathbf{\Phi}_{2,1}$ and $\mathbf{\Phi}_{3,1}$. However, the diagonal elements of $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are in different order compared with the diagonal elements of $\mathbf{\Phi}_{2,1}$ and $\mathbf{\Phi}_{3,1}$, which means the diagonal elements of $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ are mismatched. Therefore, these elements should be matched before the nominal DOAs are estimated.

From the definition of $\mathbf{\Phi}_{q,1}$ given in (33), we can see that $\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1} \in \mathbb{C}^{3K \times 3K}$ and $\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}^{-1} \in \mathbb{C}^{3K \times 3K}$ are also upper triangular matrices, and their diagonal elements satisfy $[\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}]_{p,p} = [\mathbf{\Phi}_{2,1}]_{p,p}[\mathbf{\Phi}_{3,1}]_{p,p}$ and $[\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}^{-1}]_{p,p} = [\mathbf{\Phi}_{2,1}]_{p,p}/[\mathbf{\Phi}_{3,1}]_{p,p}$, $p = 1, 2, \cdots, 3K$, respectively. In addition, according to (45) and (46), we have

$$\hat{\mathbf{\Psi}}_3 = \hat{\mathbf{\Psi}}_1\hat{\mathbf{\Psi}}_2 \approx \mathbf{T}\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}\mathbf{T}^{-1} \in \mathbb{C}^{3K \times 3K}, \tag{54}$$

$$\hat{\mathbf{\Psi}}_4 = \hat{\mathbf{\Psi}}_1\hat{\mathbf{\Psi}}_2^{-1} \approx \mathbf{T}\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}^{-1}\mathbf{T}^{-1} \in \mathbb{C}^{3K \times 3K}. \tag{55}$$

Therefore, the eigenvalues of $\hat{\mathbf{\Psi}}_3$ are approximately the diagonal elements of $\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}$. In addition, denote the EVD of $\mathbf{\Psi}_3$ as

$$\hat{\mathbf{\Psi}}_3 = \mathbf{T}_3\mathbf{\Lambda}_3\mathbf{T}_3^{-1}, \tag{56}$$

where $\mathbf{T}_3 \in \mathbb{C}^{3K \times 3K}$ is composed of the eigenvectors of $\mathbf{\Psi}_3$, and $\mathbf{\Lambda}_3 \in \mathbb{C}^{3K \times 3K}$ is a diagonal matrix composed of the eigenvalues of $\mathbf{\Psi}_3$. From (54) and (56), we have

$$\mathbf{\Lambda}_3 = \mathbf{T}_3^{-1}\hat{\mathbf{\Psi}}_3\mathbf{T}_3 \approx \mathbf{T}_3^{-1}\mathbf{T}\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}\mathbf{T}^{-1}\mathbf{T}_3, \tag{57}$$

in which the diagonal elements of $\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}$ approximately formulate the diagonal elements of $\mathbf{\Lambda}_3$. Similarly, denote

$$\tilde{\mathbf{\Psi}}_4 = \mathbf{T}_3^{-1}\hat{\mathbf{\Psi}}_4\mathbf{T}_3 \in \mathbb{C}^{3K \times 3K}. \tag{58}$$

Substituting (55) into (58) yields

$$\tilde{\mathbf{\Psi}}_4 \approx \mathbf{T}_3^{-1}\mathbf{T}\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}^{-1}\mathbf{T}^{-1}\mathbf{T}_3. \tag{59}$$

Comparing (57) and (59), we know that the diagonal elements of $\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}^{-1}$ approximately formulate the diagonal elements of $\tilde{\mathbf{\Psi}}_4$ in the same manner as formulating the diagonal elements of $\mathbf{\Lambda}_3$ with the diagonal elements of $\mathbf{\Phi}_{2,1}\mathbf{\Phi}_{3,1}$. More specifically, if $[\mathbf{\Lambda}_3]_{p,p} \approx [\mathbf{\Phi}_{2,1}]_{c_p,c_p}[\mathbf{\Phi}_{3,1}]_{c_p,c_p}, \forall p$, where $c_p \in \{1, 2, \cdots, 3K\}$ and it varies with p. Then, we have $[\tilde{\mathbf{\Psi}}_4]_{p,p} \approx [\mathbf{\Phi}_{2,1}]_{c_p,c_p}/[\mathbf{\Phi}_{3,1}]_{c_p,c_p}$. These facts can be exploited to match the diagonal elements of $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$. A matching algorithm is proposed as follows.

Algorithm 1: Matching of the Eigenvalues

Step 1) Calculate the EVD of $\hat{\mathbf{\Psi}}_1$ and $\hat{\mathbf{\Psi}}_2$ using (52) and (53).

Step 2) Calculate the EVD of $\hat{\mathbf{\Psi}}_3$ as (56). Calculate $\tilde{\mathbf{\Psi}}_4$ as (58). Set $p = 0$.

Step 3) $p \leftarrow p + 1$.

- *Calculate the product $\beta_{p,\tilde{p},1} = [\mathbf{\Lambda}_1]_{p,p}[\mathbf{\Lambda}_2]_{\tilde{p},\tilde{p}}$ and the quotient $\beta_{p,\tilde{p},2} = [\mathbf{\Lambda}_1]_{p,p}/[\mathbf{\Lambda}_2]_{\tilde{p},\tilde{p}}$, where $\tilde{p} = 1, 2, \cdots, 3K$.*

- *Match $\beta_{p,\tilde{p},1}$ and $[\mathbf{\Lambda}_3]_{p',p'}$, $\beta_{p,\tilde{p},2}$ and $[\tilde{\mathbf{\Psi}}_4]_{p',p'}$, $p' = 1, 2, \cdots, 3K$, according to the LS criterion, i.e., find the corresponding relation between $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ based on $(x_p, y_p) = \arg\min_{\tilde{p},p'} \mu_{p,\tilde{p},p'}$, where $\mu_{p,\tilde{p},p'} = |\beta_{p,\tilde{p},1} - [\mathbf{\Lambda}_3]_{p',p'}|^2 + \left|\beta_{p,\tilde{p},2} - [\tilde{\mathbf{\Psi}}_4]_{p',p'}\right|^2.$*

- *Set $[\tilde{\boldsymbol{\Lambda}}_2]_{p,p} = [\boldsymbol{\Lambda}_2]_{x_p,x_p}$, where $\tilde{\boldsymbol{\Lambda}}_2 \in \mathbb{C}^{3K \times 3K}$ is a diagonal matrix.*

Step 4) Repeat Step 3 until $p = 3K$.

Step 5) Sort the diagonal elements of $\boldsymbol{\Lambda}_1$ in descending order, and the pth largest element of $\boldsymbol{\Lambda}_1$ is given by $\lambda_{1,p} = [\boldsymbol{\Lambda}_1]_{z_p,z_p}, p = 1, 2, \cdots, 3K$, where $z_p \in \{1, 2, \cdots, 3K\}$ represents the position of the pth largest element of $\boldsymbol{\Lambda}_1$ and can be obtained by sorting the diagonal elements of $\boldsymbol{\Lambda}_1$. Additionally, the diagonal elements of $\tilde{\boldsymbol{\Lambda}}_2$ are sorted according to $z_p, p = 1, 2, \cdots, 3K$, and the pth sorted element is given by $\lambda_{2,p} = [\tilde{\boldsymbol{\Lambda}}_2]_{z_p,z_p}, p = 1, 2, \cdots, 3K$.

Remark 1: In the second item of Step 3, for a given p, all the combinations of $\tilde{p}$ and p' are used for calculating $\mu_{p,\tilde{p},p'}$, and the combination of $\tilde{p}$ and p' that corresponds to the minimum of $\mu_{p,\tilde{p},p'}$ is assigned to (x_p, y_p). Since any two UTs are distinguished by at least one of the 2-D nominal DOAs, if two diagonal elements of $\boldsymbol{\Lambda}_3$ are of the same value, the corresponding two diagonal elements of $\tilde{\boldsymbol{\Psi}}_4$ will not be of the same value. Similarly, if two elements in $\beta_{p,\tilde{p},1}, p, \tilde{p} = 1, 2, \cdots, 3K$, are of the same value, the corresponding two elements in $\beta_{p,\tilde{p},2}, p, \tilde{p} = 1, 2, \cdots, 3K$, will not be of the same value. Hence, the elements can be matched without ambiguity. Meanwhile, the above matching algorithm is presented in this way for clarity. Actually, it can be simplified. In Step 3, the calculations of the already selected diagonal elements of $\boldsymbol{\Lambda}_2$, $\boldsymbol{\Lambda}_3$, and $\tilde{\boldsymbol{\Psi}}_4$ can be omitted in the subsequent iterations.

From the matching algorithm, we can see that $\lambda_{1,p}, \lambda_{2,p}, p = 1, 2, \cdots, 3K$, are the estimates of the diagonal elements of $\boldsymbol{\Phi}_{q,1}, q = 2, 3$, given by (33), though these two groups of elements may be different in order. Without loss of generality, we can denote $\lambda_{1,3(k-1)+l}$ and $\lambda_{2,3(k-1)+l}$, $l = 1, 2, 3$ as the estimates of $[\boldsymbol{\Phi}_{2,1}]_{k+(l-1)K,k+(l-1)K}$ and $[\boldsymbol{\Phi}_{3,1}]_{k+(l-1)K,k+(l-1)K}$, $l = 1, 2, 3$, respectively. According to the expressions of the diagonal elements of $\boldsymbol{\Phi}_{2,1}$ and $\boldsymbol{\Phi}_{3,1}$ given in (34) and (35), we have

$$\lambda_{1,3(k-1)+l} \approx \exp(iu\sin(\bar{\phi}_k)\cos(\bar{\theta}_k)), \quad (60)$$

$$\lambda_{2,3(k-1)+l} \approx \exp(iu\sin(\bar{\phi}_k)\sin(\bar{\theta}_k)), \quad (61)$$

where $l = 1, 2, 3$. Then, the estimates of the nominal DOAs, $\bar{\theta}_k$ and $\bar{\phi}_k$, can be expressed as

$$\hat{\bar{\theta}}_k = \frac{1}{3}\sum_{l=1}^{3}\arctan\left(\frac{\ln\left(\lambda_{2,3(k-1)+l}\right)}{\ln\left(\lambda_{1,3(k-1)+l}\right)}\right), \quad (62)$$

$$\hat{\bar{\phi}}_k = \frac{1}{3}\sum_{l=1}^{3}\arcsin\left(\frac{1}{u}\sqrt{-\sum_{a=1}^{2}\left(\ln\left(\lambda_{a,3(k-1)+l}\right)\right)^2}\right), \quad (63)$$

where $k = 1, 2, \cdots, K$.

From (19), we see that $\boldsymbol{\Lambda}_c$ can be estimated as

$$\hat{\boldsymbol{\Lambda}}_c = \hat{\mathbf{A}}^{\dagger}(\mathbf{R}_x - \hat{\sigma}_n^2\mathbf{I}_M)\left(\hat{\mathbf{A}}^H\right)^{\dagger} \in \mathbb{C}^{3K \times 3K}, \quad (64)$$

where $\hat{\sigma}_n^2$ is the estimate of the variance of the noise, and it is the average of the smallest $M - 3K$ eigenvalues of $\mathbf{R}_x$. In addition, $\hat{\mathbf{A}} \in \mathbb{C}^{M \times 3K}$ is the estimate of $\mathbf{A}$, and it may

be obtained by replacing the nominal DOAs in $\mathbf{A}$ with the estimated nominal DOAs. From the definition of $\boldsymbol{\Lambda}_c$ in (18), the angular spreads, σ_{θ_k} and σ_{ϕ_k}, can be estimated as

$$\hat{\sigma}_{\theta_k} = \sqrt{\frac{[\hat{\boldsymbol{\Lambda}}_c]_{K+k,K+k}}{[\hat{\boldsymbol{\Lambda}}_c]_{k,k}}}, \quad (65)$$

$$\hat{\sigma}_{\phi_k} = \sqrt{\frac{[\hat{\boldsymbol{\Lambda}}_c]_{2K+k,2K+k}}{[\hat{\boldsymbol{\Lambda}}_c]_{k,k}}}, \quad (66)$$

where $k = 1, 2, \cdots, K$. It is obvious that the accuracy of the estimated angular spreads depends on the estimated nominal DOAs.

In practice, the covariance matrix $\mathbf{R}_x$ may be estimated as

$$\hat{\mathbf{R}}_x = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^H(t) \in \mathbb{C}^{M \times M}. \quad (67)$$

For clarity, the proposed estimation approach is summarized as follows.

Algorithm 2: Estimation of the Nominal DOAs and the Angular Spreads

Step 1) Calculate the sample covariance matrix, $\hat{\mathbf{R}}_x$, according to (67).

Step 2) Calculate the EVD of $\hat{\mathbf{R}}_x$ according to (20), and find $\mathbf{E}_s$ that corresponds to the largest $3K$ eigenvalues.

Step 3) Calculate the selected matrices, $\mathbf{E}_l, l = 1, 2, 3$, according to (39); and estimate the transform matrices, $\boldsymbol{\Psi}_1$ and $\boldsymbol{\Psi}_2$, based on the TLS criterion, which entails performing the EVD according to (47) and (48), partitioning the matrices according to (49), and calculating the transform matrices according to (50) and (51).

Step 4) Match the eigenvalues using Algorithm 1.

Step 5) Estimate the nominal DOAs with (62) and (63), the diagonal matrix $\hat{\boldsymbol{\Lambda}}_c$ with (64), and the angular spreads with (65) and (66).

Remark 2: As opposed to traditional approaches, such as the existing subspace-based, the LS-based covariance matching, the ML-based approaches, and the existing 2-D ESPRIT-based approach, where the searching of angular parameters is typically inevitable, the proposed estimator dispenses with the complicated searching due to its closed-form expression. Therefore, the proposed approach imposes significantly lower computational complexity.

In the next section, the computational complexity and performance analyses of the proposed approach will be provided.

IV. ANALYSIS OF THE PROPOSED APPROACH

In this section, the impact of the number of the BS antennas M on the rank of $\mathbf{A}$ and on the performance of the proposed estimator is investigated. Then, the approximate CRB concerning the covariances of the estimation errors is derived to measure the performance of the proposed estimator from another perspective. Finally, the computational complexity of the proposed approach is analyzed, and is compared with that of the existing approaches. It is shown that the estimation performance improves as M increases, and the proposed approach is of much lower complexity.

A. The Impact of the Number of the BS Antennas

Note that the covariance matrix $\mathbf{R_x}$ can only be estimated with the aid of the sample covariance matrix $\hat{\mathbf{R}}_\mathbf{x}$. According to (12) and (67), we have

$$\hat{\mathbf{R}}_\mathbf{x} \approx \mathbf{A}\hat{\mathbf{R}}_\mathrm{c}\mathbf{A}^H + \hat{\mathbf{R}}_\mathrm{n}, \qquad (68)$$

where

$$\hat{\mathbf{R}}_\mathrm{c} = \frac{1}{T}\sum_{t=1}^{T}\mathbf{c}(t)\mathbf{c}^H(t) \in \mathbb{C}^{3K \times 3K}, \qquad (69)$$

and

$$\hat{\mathbf{R}}_\mathrm{n} = \frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{A}\mathbf{c}(t)\mathbf{n}^H(t) + \mathbf{n}(t)\mathbf{c}^H(t)\mathbf{A}^H + \mathbf{n}(t)\mathbf{n}^H(t)\right)$$
$$\in \mathbb{C}^{M \times M} \qquad (70)$$

are the estimates of $\mathbf{\Lambda}_\mathrm{c}$ and $\sigma_\mathrm{n}^2\mathbf{I}_M$ invoked in (19), respectively.

Because $\hat{\mathbf{R}}_\mathbf{x}$ is a normal positive semi-definite matrix, its EVD is similar to the EVD of $\mathbf{R_x}$ characterized in (20), and is given by

$$\hat{\mathbf{R}}_\mathbf{x} = \hat{\mathbf{E}}_\mathrm{s}\hat{\mathbf{\Sigma}}_\mathrm{s}\hat{\mathbf{E}}_\mathrm{s}^H + \hat{\mathbf{E}}_\mathrm{n}\hat{\mathbf{\Sigma}}_\mathrm{n}\hat{\mathbf{E}}_\mathrm{n}^H, \qquad (71)$$

where $\hat{\mathbf{E}}_\mathrm{s} \in \mathbb{C}^{M \times 3K}$ and $\hat{\mathbf{E}}_\mathrm{n} \in \mathbb{C}^{M \times (M-3K)}$ are composed of the eigenvectors of $\hat{\mathbf{R}}_\mathbf{x}$, while $\hat{\mathbf{\Sigma}}_\mathrm{s} \in \mathbb{R}^{3K \times 3K}$ and $\hat{\mathbf{\Sigma}}_\mathrm{n} \in \mathbb{R}^{(M-3K) \times (M-3K)}$ are diagonal matrices with their diagonal elements being the eigenvalues of $\hat{\mathbf{R}}_\mathbf{x}$. The diagonal elements of $\hat{\mathbf{\Sigma}}_\mathrm{s}$ are the largest $3K$ eigenvalues of $\hat{\mathbf{R}}_\mathbf{x}$, and $\hat{\mathbf{E}}_\mathrm{s}$ is the estimate of the signal subspace $\mathbf{E}_\mathrm{s}$. Because the number of received signal snapshots T is finite, $\hat{\mathbf{R}}_\mathrm{c}$ and $\hat{\mathbf{R}}_\mathrm{n}$ are random matrices, and their eigenvalues are also random variables. Consequently, $\hat{\mathbf{E}}_\mathrm{s}$ and $\mathbf{A}$ might not be in the same subspace as in (24), albeit the linear relationship is crucial to the estimation performance. A proposition is given below to show the impact of M on the relation between $\mathbf{A}$ and $\hat{\mathbf{E}}_\mathrm{s}$ subject to finite T.

Proposition 1: As the number of the BS antennas $M \to \infty$, $\mathbf{A}$ approximates to a full rank matrix, and $\hat{\mathbf{E}}_\mathrm{s}$ tends almost surely to be in the same subspace as $\mathbf{A}$.

Proof: Please see Appendix A. ∎

Remark 3: From the above analysis, we know that when the number of the BS antennas M grows without bound, $\mathbf{A}$ approximates to a full rank matrix, and the estimation accuracy of $\hat{\mathbf{E}}_\mathrm{s}$ improves almost surely. Therefore, the performance of the proposed estimator subject to finite T becomes better and better when the number of the BS antennas M increases.

B. Approximate CRB of the Proposed Estimator

For the proposed estimator, the approximate CRB concerning the covariance matrix of the error of the estimated signal parameter vector $\mathbf{u}$, whose specific form is defined in Appendix B, is given as follows.

$$\mathbf{C} = \left(\mathbf{J}_{\mathbf{u},\mathbf{u}} - \mathbf{J}_{\mathbf{u},\mathbf{v}}\mathbf{J}_{\mathbf{v},\mathbf{v}}^{-1}\mathbf{J}_{\mathbf{u},\mathbf{v}}^T\right)^{-1} \in \mathbb{R}^{4K \times 4K}, \qquad (72)$$

which means

$$\mathbb{E}\left\{(\hat{\mathbf{u}} - \mathbf{u})(\hat{\mathbf{u}} - \mathbf{u})^T\right\} \geq \mathbf{C}. \qquad (73)$$

A detailed derivation of (72) and the definitions of the variables used in (72) and (73) can be found in Appendix B.

Remark 4: The derived approximate CRB plays a very important role for measuring the quality of the proposed estimator. It provides us with a measure of the spread of the error. In the simulation results of Section V, we will plot the approximate CRB as a reference to see how well the proposed estimator works.

C. Complexity Analysis

In this paper, the notation $O(n)$ means that complexity of the arithmetic is linear in $n \in \mathbb{R}^+$ [49, p. 5]. The number of snapshots T is fixed, and the complexities of various algorithms considered are compared in the asymptotic sense as $M \to \infty$.

The complexities of Step 1, Step 2, and Step 3 in Algorithm 2 are $O(M^2T)$, $O(M^3)$, and $O(MK^2)$, respectively, and the total complexities of other steps in Algorithm 2 are $O(K^3)$. Since M is far larger than K, the complexity of the proposed approach is roughly characterized as $O(M^3 + M^2T + MK^2) \to O(M^3)$ as $M \to \infty$.

Amongst the existing estimators, there is only a covariance matching estimator (COMET) [35] proposed for the 2-D localization of the ID sources. Although the known subspace based approaches [37], [40], the generalized Capon beamforming approach [39] and the ML approach [25] are proposed for the 1-D localization of the ID sources, they can be modified for the corresponding 2-D localization. In contrast to the proposed estimator, these approaches search over all the possible combinations of the nominal DOAs and the angular spreads to obtain the estimates. Therefore, their computational complexity is unbearable for the massive MIMO systems.

For the COMET approach [35], the parameter estimation criterion is

$$\arg\min_{\mathbf{u}'} \mathrm{tr}\left(\left(\sum_{k=1}^{K}\hat{s}_k\mathbf{B}_k + \hat{\sigma}_\mathrm{n}^2\mathbf{I}_M - \hat{\mathbf{R}}_\mathbf{x}\right)^2\right), \qquad (74)$$

where $\mathbf{B}_k \in \mathbb{C}^{M \times M}$ (defined in (98)) is a function of $\tilde{\mathbf{u}}_k = [\bar{\theta}_k, \bar{\phi}_k, \sigma_{\theta_k}, \sigma_{\phi_k}]^T \in \mathbb{R}^{4 \times 1}, k = 1, 2, \cdots, K$, and $\mathbf{u}' = [\tilde{\mathbf{u}}_1^T, \tilde{\mathbf{u}}_2^T, \cdots, \tilde{\mathbf{u}}_K^T]^T \in \mathbb{R}^{4K \times 1}$. When the calculations of $\hat{s}_k, \mathbf{B}_k, k = 1, 2, \cdots, K$, and of $\hat{\sigma}_\mathrm{n}$ are ignored, the computational complexity of this approach is $O(D_1M^3 + M^2T) \to O(D_1M^3)$ as $M \to \infty$, where D_1 is the search dimension for estimating the nominal DOAs and the angular spreads of all the UTs.

When the subspace based approach of [40] is modified for the 2-D localization, the estimation criterion is given as

$$\arg\min_{\tilde{\mathbf{u}}} \left\|\hat{\mathbf{R}}_\mathbf{x}^{-1}\mathbf{B}\right\|_\mathrm{F}^2, \qquad (75)$$

where $\mathbf{B} \in \mathbb{C}^{M \times M}$ (defined by omitting the subscript k of $\mathbf{B}_k$) is a function of $\tilde{\mathbf{u}} = [\bar{\theta}, \bar{\phi}, \sigma_\theta, \sigma_\phi]^T \in \mathbb{R}^{4 \times 1}$. By searching the K local minima, the angular parameters of all the K UTs can be estimated. Hence, the computational complexity of this approach is $O(D_2M^3 + M^2T) \to O(D_2M^3)$ as $M \to \infty$, where D_2 is the search dimension for estimating the nominal DOAs and the angular spreads of a single UT. It is obvious that $D_1 = D_2^K$, which implies that $D_1 \gg D_2$.

TABLE I
COMPUTATIONAL COMPLEXITY COMPARISON OF LOCALIZATION
APPROACHES

COMET [35]	$O(D_1 M^3)$
Subspace estimator [40]	$O(D_2 M^3)$
DISPARE [37]	$O(D_2 M^3)$
The proposed estimator	$O(M^3)$

* $D_1 \gg D_2 \gg 1$.

The dispersed signal parametric estimation (DISPARE) approach advocated in [37] is based on subspace fitting. When this approach is modified for the 2-D localization, the estimation criterion becomes

$$\arg\min_{\tilde{\mathbf{u}}} \left\| \hat{\mathbf{E}}_n^H \mathbf{B} \right\|_F^2, \qquad (76)$$

where $\tilde{\mathbf{u}}$ is defined below (75), $\hat{\mathbf{E}}_n \in \mathbb{C}^{M \times N_n}$ corresponds to the pseudonoise subspace, and $N_n \approx M - 3K \to O(M)$ is the dimension of this subspace as $M \to \infty$. Hence, the computational complexity of this approach is also $O(D_2 M^3 + M^2 T) \to O(D_2 M^3)$ as $M \to \infty$.

For the sake of clarity, the computational complexities of all these approaches are summarized in Table I. It can be easily seen that the complexity of the proposed approach is significantly lower than that of other approaches.

V. NUMERICAL RESULTS

In this section, we provide numerical results to illustrate the performance of the proposed approach, of the subspace based approach [40], and of the DISPARE [37]. Additionally, the approximate CRB of the proposed estimator is also calculated for comparison. In particular, the dimension of the pseudosignal space is chosen as the number of eigenvalues that collectively contain 95% of the sum of the eigenvalues in the DISPARE approach. The COMET approach of [35] is not considered in our simulations because of its prohibitive computational complexity.

The simulation parameters of the first three simulations as shown in Fig. 3, Fig. 4, and Fig. 5 are given as follows. The number of the UTs is $K = 2$, the number of multipaths is $N_k = 50, k = 1, 2$, and the transformed distance between any two adjacent antennas, cf. the sentence below (2), is $u = \pi$ radians. The nominal azimuth DOAs of the two UTs are $\bar{\theta}_1 = 10°$, $\bar{\theta}_2 = 50°$, and the corresponding nominal elevation DOAs are $\bar{\phi}_1 = 30°$, $\bar{\phi}_2 = 40°$. The azimuth angular spreads are $\sigma_{\theta_k} = 1°, k = 1, 2$, and the elevation angular spreads are $\sigma_{\phi_k} = 1°, k = 1, 2$. The path gain variances are $\sigma_{\gamma_k}^2 = 1, k = 1, 2$, and the noise variance is $\sigma_n^2 = 1$. The transmitted signals, $s_k(t), k = 1, 2$, are BPSK modulated. It can be seen that the average received SNR from each UT is S_k, where S_k is the transmitted signal power. The number of snapshots is $T = 500$. For [40] and [37], the search range of the nominal azimuth DOAs are set as $[\bar{\theta}_k - 1°, \bar{\theta}_k + 1°], k = 1, 2$, the search range of the nominal elevation DOAs are set as $[\bar{\phi}_k - 1°, \bar{\phi}_k + 1°], k = 1, 2$, and the search range of the angular spreads is set as $[0.2°, 2°]$. The values out of these ranges need not to be searched because the minima can only be achieved in these search ranges. In addition, the search step size of the

nominal DOAs and the angular spreads is $0.2°$. The number of simulation trials is 200. The metric of root mean square error (RMSE) is evaluated for the estimates of various source parameters.

Given the number of snapshots $T = 500$, a rough estimate of the delay required for obtaining these snapshots in a typical scenario is also presented here. We consider a Long Term Evolution (LTE) uplink system, which operates at 2 GHz, the channel bandwidth is 2.5 MHz, and the sampling rate is 3.84 MHz [50]. In order to obtain uncorrelated snapshots in (1), the delay is approximately $500/3.84 \times 10^{-6} \approx 1.3 \times 10^{-4}$ s. When the distance between one UT and the BS is 1 km, and the speed of the UT is 134 m/s (this may be the scenario of high speed railway user, and is the worst scenario for obtaining temporarily uncorrelated snapshots), the maximum change of the nominal azimuth DOA after sampling the T snapshots is $134 \times 1.3 \times 10^{-4}/10^3/\pi \times 180 \approx 0.001°$, where 10^3 m is the distance between one UT and the BS. When the speed of the UT is slower than 134 m/s, the delay is acceptable and the proposed approach is applicable in practice.

Subject to these simulation parameters, the complexities of the estimation approaches considered can be compared explicitly. The search dimensions are $D_1 = (11 \times 10)^4 = 1.4641 \times 10^8$, $D_2 = (11 \times 10)^2 = 1.21 \times 10^4$, where 11 is calculated from the search of the nominal DOA, i.e., $(1° - (-1°))/0.2° + 1$, and 10 is calculated from the search of the angular spread, i.e., $(2° - 0.2°)/0.2° + 1$. When $M = 100$, the complexities of the COMET and the DISPARE are roughly computed as $O(1.4641 \times 10^{14} + 5.0 \times 10^6) = O(1.4641 \times 10^{14})$ and $O(1.21 \times 10^{10} + 5.0 \times 10^6) = O(1.21 \times 10^{10})$, respectively, while the computational complexity of the proposed approach is roughly $O(1.0 \times 10^6 + 5.0 \times 10^6 + 4.0 \times 10^2) = O(6.0 \times 10^6)$, where $O(5.0 \times 10^6)$ is the complexity of calculating the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{x}}$ in (67). Hence, the complexity of the proposed approach is lower than 0.1% of the complexities of the existing approaches. Obviously, in terms of implementation, these existing approaches are significantly more complicated than the proposed method in the context of the massive MIMO systems. In Fig. 3, the base 10 logarithms of the computational complexities in big O notation versus the number of the BS antennas M for these three approaches are shown. We can see that the complexity of the proposed approach is significantly lower than that of other approaches. In addition, the complexity of the proposed approach for $M = 100$ is close to that of the DISPARE with $M = 9$, and is even significantly lower than that of the COMET with $M = 9$. Therefore, in certain configurations, employing the proposed approach in the massive MIMO systems does not even impose a higher complexity than employing these benchmark search-based approaches in traditional small-scale MIMO systems.

In the second test as shown in Fig. 4, the average received SNR from each UT is 10 dB, which means $S_k = 0.2, k = 1, 2$. The numbers of the BS antennas in the x-direction and the y-direction satisfy $M_x = M_y = \sqrt{M}$. The RMSEs of the estimated nominal DOAs and angular spreads versus the number of the BS antennas M are plotted in Fig. 4. It can be observed that the RMSEs of these estimated parameters of the proposed approach decrease rapidly as M increases, while the RMSEs
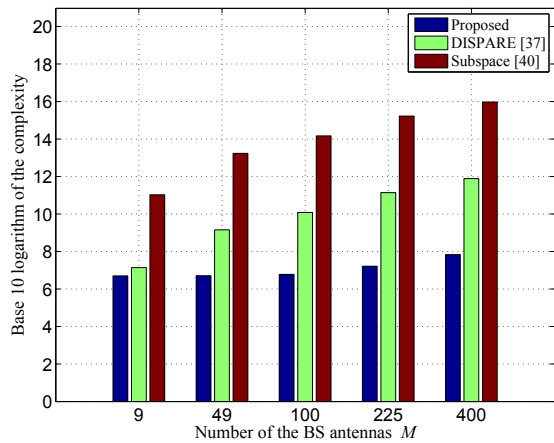
Fig. 3. Comparison of "computational complexities versus the number of the BS antennas M", for the estimates of the angular parameters of two sources when using different estimation methods. The y-axis represents the base 10 logarithm of the computational complexity in big O notation.
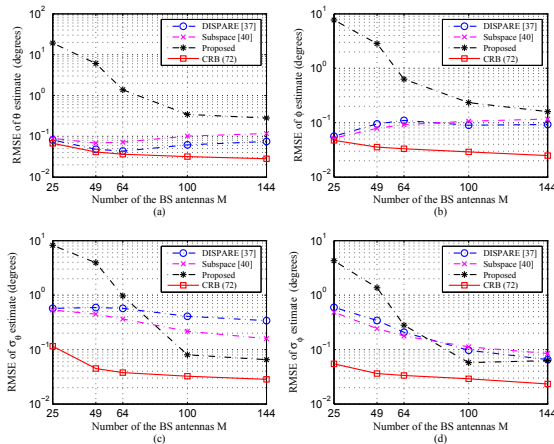

Fig. 5. Comparison of "RMSEs versus average received SNR from each UT" for the estimates of the angular parameters of two sources when using different estimation methods, and the number of the BS antennas is $M = 100$. (a), (b), (c), and (d) correspond to the estimation of the nominal azimuth DOA, the nominal elevation DOA, the azimuth angular spread, and the elevation angular spread, respectively.

It can be seen that the RMSEs of the proposed approach also decrease rapidly when the SNR increases, while the RMSEs of other approaches decrease slowly. These results demonstrate that the performance of the proposed estimator is deteriorated when the power of the received noise is high, and the effect of increasing the SNR is similar to the effect of increasing the number of the BS antennas, as compared with Fig. 4. Therefore, the proposed approach can potentially trade for good performance in low SNR scenarios by employing a large number of the BS antennas. In other words, for the massive MIMO systems the transmitted power can be significantly reduced due to an unprecedented high number of the BS antennas.

In the fourth example as shown in Fig. 6, some of the parameters are changed for evaluating the performance of these approaches with the increased number of the UTs. The number of the UTs is modified as $K = 5$, and the number of multipaths is $N_k = 50, k = 1, 2, \cdots, 5$. The nominal azimuth DOAs of the five UTs are $\bar{\theta}_1 = 10°$, $\bar{\theta}_2 = 50°$, $\bar{\theta}_3 = 130°$, $\bar{\theta}_4 = 110°$, $\bar{\theta}_5 = 30°$, and the corresponding nominal elevation DOAs are $\bar{\phi}_1 = 30°$, $\bar{\phi}_2 = 40°$, $\bar{\phi}_3 = 70°$, $\bar{\phi}_4 = 80°$, $\bar{\phi}_5 = 50°$. The numbers of the BS antennas in the x-direction and the y-direction are $M_x = 10$ and $M_y = 10$, respectively. The average received SNR from each UT is 30 dB. In this simulation, the search step size of the nominal DOAs and the angular spreads is $0.01°$, and the search range of the angular spreads is set as $[0.01°, 1°]$. The azimuth angular spreads of these UTs are the same as the elevation angular spreads, and vary with each data point in Fig. 6. The RMSEs of the estimated nominal DOAs and angular spreads versus the angular spread, are plotted in Fig. 6. It is observed that the RMSEs of the proposed approach achieve their minima when the angular spread is in the middle of the range. When the angular spreads are small, the expectations of the diagonal elements of $\hat{\mathbf{R}}_c$ that is formulated in (68) are small. Thus, the impact of the noise becomes the dominant

Fig. 4. Comparison of "RMSEs versus the number of the BS antennas M", for the estimates of the angular parameters of two sources when using different estimation methods, and the average received SNR from each UT is 10 dB. (a), (b), (c), and (d) correspond to the estimation of the nominal azimuth DOA, the nominal elevation DOA, the azimuth angular spread, and the elevation angular spread, respectively.

of these estimated parameters of the subspace based approach and the DISPARE are almost invariant because they have achieved their best performance when M is not so large. These results coincide with our analysis of the impact of the number of the BS antennas M on the estimation performance. More specifically, for the proposed approach, as M increases, the estimated signal subspace tends to be in the same subspace as the array response matrix $\mathbf{A}$. Thus, the estimation performance improves. In addition, it is easy to observe that when $M = 144$ the RMSEs of the estimated azimuth and elevation DOAs of the proposed approach are close to those of [40] and [37], while the RMSEs of the estimated angular spreads of the proposed approach are superior to those of the latter.

In the third test as shown in Fig. 5, the numbers of the BS antennas in the x-direction and the y-direction are $M_x = 10$ and $M_y = 10$, respectively, and hence $M = 100$. The RMSEs of the estimated nominal DOAs and angular spreads versus the average received SNR from each UT are depicted in Fig. 5.
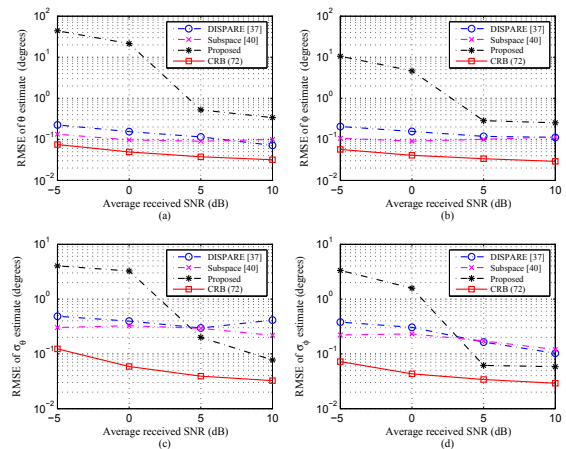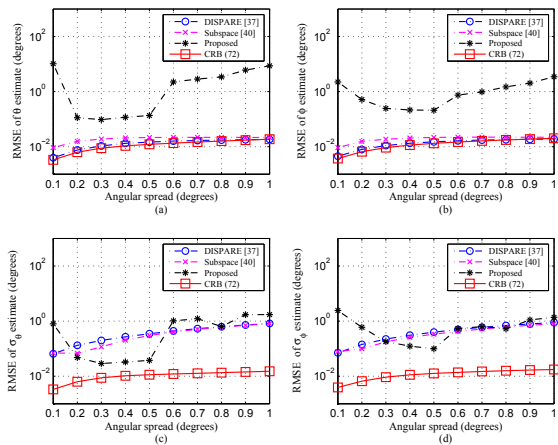
Fig. 6. RMSEs versus angular spread concerning the estimates of the angular parameters of five sources, while the number of the BS antennas is $M = 100$, and the average received SNR from each UT is 30 dB. (a), (b), (c), and (d) correspond to the estimation of the nominal azimuth DOA, the nominal elevation DOA, the azimuth angular spread, and the elevation angular spread, respectively.
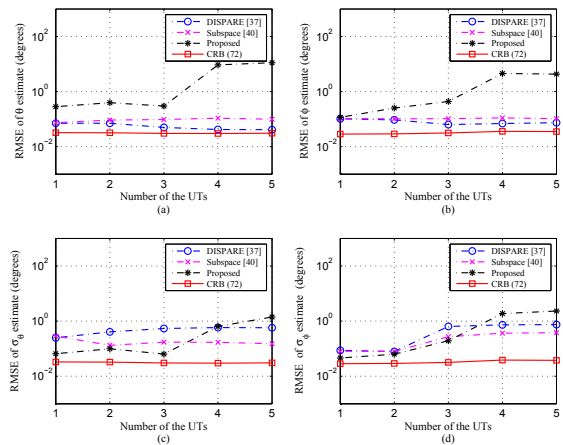

Fig. 7. RMSEs versus the number of the UTs concerning the estimates of the angular parameters, while the number of the BS antennas is $M = 100$, and the average received SNR from each UT is 10 dB. (a), (b), (c), and (d) correspond to the estimation of the nominal azimuth DOA, the nominal elevation DOA, the azimuth angular spread, and the elevation angular spread, respectively.

factor. When the angular spreads are large, the remainder of the Taylor series in (10) cannot be omitted. Thus, the estimation performance degrades. However, the performance of [40] and [37] are mainly dominated by the remainder of the Taylor series rather than by the noise. This is because the Taylor series approximation of [40] and [37] is different from that of the proposed approach, and the former imposes less impacts on the estimation performance. Hence, the proposed approach is best suitable for localization of multiple UTs when the angular spreads of these UTs remain in the modest region.

In the fifth test as shown in Fig. 7, some of the parameters are changed for evaluating the performance of these approaches when the number of the UTs increases. The nominal azimuth DOAs, the nominal elevation DOAs, and the number of the BS antennas are the same as in the third example. The average received SNR from each UT is 10 dB. The azimuth angular spreads are $\sigma_{\theta_k} = 1°, k = 1, 2, \cdots, 5$, and the elevation angular spreads are $\sigma_{\phi_k} = 1°, k = 1, 2, \cdots, 5$. The RMSEs of the estimated nominal DOAs and the angular spreads offered by the proposed approach and the approaches in [40] and [37] versus the number of the UTs, are plotted in Fig. 7. It is observed that the RMSEs of the proposed approach increase as the number of the UTs increases. This is because the increase of the number of the UTs causes the sum of the remainder of the Taylor series in (10) increases. As a result, the performance of the proposed approach degrades as the number of the UTs increases. In contrast, the RMSEs of [40] and [37] increase slowly as the number of the UTs increases. This is because the nominal DOAs of the UTs are only estimated by searching around the true values in these approaches, which is based on the assumption that the coarse estimates of the nominal DOAs have been obtained.

In the sixth test as shown in Fig. 8, the simulation parameters are the same as those in the third test, except that the average received SNR is 10 dB. The RMSEs of the estimated nominal DOAs and of the estimated angular spreads attained

by the proposed approach and by the approaches of [40] and [37] versus the number of scatterers, are plotted in Fig. 8. Note that the number of scatterers is the same as the number of multipaths. It can be seen that the RMSEs of these approaches are almost invariant with the number of scatterers. The path gains are temporally independent. Thus, the T snapshots of the received signal in (1) are independent of each other as long as the number of the multipaths is no less than one. Since the multipaths cannot be distinguished in the received signal, when the number of the multipaths increases, the number of independent snapshots remains invariant. Note that the average received SNR remains constant in the simulation in order to evaluate the impact of the number of scatterers. From (67), it is known that the sample covariance matrix $\hat{\mathbf{R}}_{\mathbf{x}}$ is directly related to the number of independent snapshots T and the estimation accuracy of $\hat{\mathbf{R}}_{\mathbf{x}}$ is crucial to the estimation performance. As a result, the number of scatterers imposes little impact on the estimation performance of these approaches.

VI. CONCLUSIONS

In this paper, we have proposed an ESPRIT-based approach for the 2-D localization of multiple ID sources in the massive MIMO systems. The proposed approach does not constrain the distance between adjacent antennas and decouples the 2-D angular parameters. Therefore, it is feasible for the 2-D localization. Our analysis has shown that the performance of the proposed approach improves as the number of the BS antennas increases, and the computational complexity of the proposed approach is significantly lower than that of other approaches. For example, in some representative scenarios as considered, the complexity of the proposed estimator is less than 0.1% of that of the existing methods. In addition, the simulation results have demonstrated that the performance of the proposed approach is comparable to that of other approaches in the massive MIMO systems. The extension of the proposed approach to the scenario where sources having large angular spreads may be addressed in our future work.
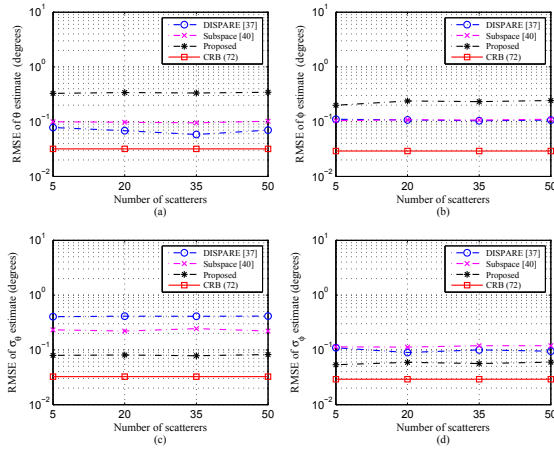
Fig. 8. RMSEs versus the number of scatterers concerning the estimates of the angular parameters, while the number of the BS antennas is $M = 100$, and the average received SNR from each UT is 10 dB. (a), (b), (c), and (d) correspond to the estimation of the nominal azimuth DOA, the nominal elevation DOA, the azimuth angular spread, and the elevation angular spread, respectively.

APPENDIX A
PROOF OF PROPOSITION 1

First, the norms of the column vectors of $\mathbf{A}$ are given for normalizing the column vectors of $\mathbf{A}$. By changing the DOAs, $\theta_{k,j}(t), \phi_{k,j}(t)$, in (2) to the nominal DOAs $\bar{\theta}_k, \bar{\phi}_k, \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$ is obtained, and we have

$$\left[\frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k}\right]_m = iu\sin(\bar{\phi}_k)\Big[-(m_x-1)\sin(\bar{\theta}_k) \\ +(m_y-1)\cos(\bar{\theta}_k)\Big]\times[\mathbf{a}(\bar{\theta}_k,\bar{\phi}_k)]_m, \quad (77)$$

$$\left[\frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\phi}_k}\right]_m = iu\cos(\bar{\phi}_k)\big[(m_x-1)\cos(\bar{\theta}_k) \\ +(m_y-1)\sin(\bar{\theta}_k)\big]\times[\mathbf{a}(\bar{\theta}_k,\bar{\phi}_k)]_m, \quad (78)$$

where m, m_x, and m_y are defined in (2). From (13), we can see that there are only three types of column vectors in $\mathbf{A}$, and the norms of these three kinds of column vectors are given by

$$r_{k,1} = \sqrt{\mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)} = \sqrt{M}, \quad (79)$$

$$r_{k,2} = \sqrt{\left(\frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k}\right)^H \frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k}} = \sqrt{M\tilde{M}}\tilde{r}_{k,2}, (80)$$

$$r_{k,3} = \sqrt{\left(\frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\phi}_k}\right)^H \frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\phi}_k}} = \sqrt{M\tilde{M}}\tilde{r}_{k,3}, (81)$$

where $\tilde{r}_{k,2}$ and $\tilde{r}_{k,3}, k = 1, 2, \cdots, K$, do not tend to infinity or zero as $M \to \infty$. Hence, the norms of the first K columns of $\mathbf{A}$ are proportional to $\sqrt{M}$, and the norms of the last $2K$ columns of $\mathbf{A}$ tend to be proportional to M as $M \to \infty$.

For measuring the angles between the column vectors of $\mathbf{A}$, the normalized inner products of these column vectors are derived. From (13), there are only five kinds of inner products for the column vectors of $\mathbf{A}$, and they are given by

$$t_{k,k',1} = \mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)\mathbf{a}(\bar{\theta}_{k'}, \bar{\phi}_{k'}), \quad (82)$$

$$t_{k,k',2} = \mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)\frac{\partial \mathbf{a}(\bar{\theta}_{k'}, \bar{\phi}_{k'})}{\partial \bar{\theta}_{k'}}, \quad (83)$$

$$t_{k,k',3} = \left(\frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k}\right)^H \frac{\partial \mathbf{a}(\bar{\theta}_{k'}, \bar{\phi}_{k'})}{\partial \bar{\theta}_{k'}}, \quad (84)$$

$$t_{k,k,1} = \mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)\frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k} = M\tilde{t}_{k,k,1}v_k, \quad (85)$$

and

$$t_{k,k,2} = \left(\frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\theta}_k}\right)^H \frac{\partial \mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)}{\partial \bar{\phi}_k} = M\tilde{M}\tilde{t}_{k,k,2}, (86)$$

where $k = 1, 2, \cdots, K$, $k' = 1, 2, \cdots, K$, $k' \neq k$, $v_k = g_{1,k}M_x + g_{2,k}M_y + g_{3,k}$; $t_{k,k',1}, t_{k,k',2}, t_{k,k',3}$, $\tilde{t}_{k,k,1}, \tilde{t}_{k,k,2}, g_{1,k}, g_{2,k}$, and $g_{3,k}$ do not tend to infinity as $M \to \infty$. It should be noted that the inner product $\mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)(\partial\mathbf{a}(\bar{\theta}_{k'}, \bar{\phi}_{k'})/\partial\bar{\phi}_{k'})$ is similar to (83); the inner products $(\partial\mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)/\partial\bar{\phi}_k)(\partial\mathbf{a}(\bar{\theta}_{k'}, \bar{\phi}_{k'})/\partial\bar{\phi}_{k'})$ and $(\partial\mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)/\partial\bar{\theta}_k)(\partial\mathbf{a}(\bar{\theta}_{k'}, \bar{\phi}_{k'})/\partial\bar{\phi}_{k'})$ are similar to (84); and the inner product $\mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)(\partial\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)/\partial\bar{\phi}_k)$ is similar to (85).

It can be found that when the inner products of the column vectors of $\mathbf{A}$ in (82)-(84) are normalized by the norms in (79)-(81), these normalized inner products tend to zero when $M \to \infty$. Thus, the column vectors, $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$, $\partial\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)/\partial\bar{\theta}_k$, and $\partial\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)/\partial\bar{\phi}_k$, tend to be orthogonal to any such column vector with different k. On the other hand, it can be easily found that the column vectors, $\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)$, $\partial\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)/\partial\bar{\theta}_k$, and $\partial\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)/\partial\bar{\phi}_k$, are linearly independent. Therefore, the rank of $\mathbf{A}$ tends to $3K$ as $M \to \infty$. Then, the column vectors of $\mathbf{A}$ can be orthonormalized by employing the QR decomposition as

$$\mathbf{A} = \tilde{\mathbf{A}}\mathbf{T}_A, \quad (87)$$

where $\tilde{\mathbf{A}} \in \mathbb{C}^{M\times 3K}$ satisfies $\tilde{\mathbf{A}}^H\tilde{\mathbf{A}} = \mathbf{I}_{3K}$, and $\mathbf{T}_A \in \mathbb{C}^{3K\times 3K}$ is an upper triangular matrix. It is easy to find that $\mathbf{T}_A$ tends to be a full rank matrix when $M \to \infty$, which means the condition number of $\mathbf{T}_A$ does not tend to infinity when $M \to \infty$. Then, $\mathbf{A}\hat{\mathbf{R}}_c\mathbf{A}^H$ in (68) can be written as

$$\mathbf{A}\hat{\mathbf{R}}_c\mathbf{A}^H = \tilde{\mathbf{A}}\left(\mathbf{T}_A\hat{\mathbf{R}}_c\mathbf{T}_A^H\right)\tilde{\mathbf{A}}^H. \quad (88)$$

It can be seen that $\mathbf{A}\hat{\mathbf{R}}_c\mathbf{A}^H$ has at most $3K$ nonzero eigenvalues, which are also the eigenvalues of $\mathbf{T}_A\hat{\mathbf{R}}_c\mathbf{T}_A^H$. Since $\hat{\mathbf{R}}_c$ is normal, i.e., $\hat{\mathbf{R}}_c\hat{\mathbf{R}}_c^H = \hat{\mathbf{R}}_c^H\hat{\mathbf{R}}_c$, the EVD of this matrix can be written as $\hat{\mathbf{R}}_c = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H$, where $\mathbf{U}_s \in \mathbb{C}^{3K\times 3K}$ is composed of the eigenvectors of $\hat{\mathbf{R}}_c$, and is a unitary matrix; $\mathbf{\Lambda}_s \in \mathbb{R}^{3K\times 3K}$ is a diagonal matrix composed of the eigenvalues of $\hat{\mathbf{R}}_c$. Then, we have $\hat{\mathbf{R}}_c = \mathbf{SS}^H$, where $\mathbf{S} = \mathbf{U}_s\mathbf{\Lambda}_s^{1/2} \in \mathbb{C}^{3K\times 3K}$. Because $\hat{\mathbf{R}}_c$ is a random matrix, $\mathbf{S}$ is also a random matrix. Then, the trace of $\mathbf{A}\hat{\mathbf{R}}_c\mathbf{A}^H$ can be

expressed as

$$\text{tr}\left(\mathbf{A}\hat{\mathbf{R}}_{\text{c}}\mathbf{A}^H\right) = ||\mathbf{AS}||_{\text{F}}^2 = \sum_{p=1}^{3K} w_p \left(\sum_{m=1}^{M} |[\mathbf{A}]_{m,p}|^2\right)$$
$$+ \sum_{\substack{p_2 \neq p_1 \\ p_2=1}}^{3K} \sum_{p_1=1}^{3K} y_{p_1,p_2} \left(\sum_{m=1}^{M} [\mathbf{A}]_{m,p_1}[\mathbf{A}]_{m,p_2}^*\right), \quad (89)$$

where $w_p = \sum_{p'=1}^{3K}|[\mathbf{S}]_{p,p'}|^2$, $y_{p_1,p_2} = \sum_{p'=1}^{3K}[\mathbf{S}]_{p_1,p'}[\mathbf{S}]_{p_2,p'}^*$. According to the norms of the column vectors of $\mathbf{A}$ in (79)-(81), we know that

$$\sum_{m=1}^{M} |[\mathbf{A}]_{m,p}|^2 = \begin{cases} M, & p \in \mathcal{P}_1 \\ M\tilde{M}\tilde{r}_{p-K,2}^2, & p \in \mathcal{P}_2 \\ M\tilde{M}\tilde{r}_{p-2K,3}^2, & p \in \mathcal{P}_3, \end{cases}$$

where $\mathcal{P}_1 = \{1, 2, \cdots, K\}, \mathcal{P}_2 = \{K + 1, K + 2, \cdots, 2K\}, \mathcal{P}_3 = \{2K + 1, 2K + 2, \cdots, 3K\}$. Meanwhile, due to the inner products of the column vectors of $\mathbf{A}$ in (82)-(86), we know that

$$\sum_{m=1}^{M} [\mathbf{A}]_{m,p_1}[\mathbf{A}]_{m,p_2}^*$$
$$= \begin{cases} t_{p_1,p_2,1}, & p_1, p_2 \in \mathcal{P}_1, p_1 \neq p_2 \\ t_{p_1,p_2-K,2}, & p_1 \in \mathcal{P}_1, p_2 \in \mathcal{P}_2, p_1 \neq p_2 - K \\ t_{p_1-K,p_2-K,3}, & p_1, p_2 \in \mathcal{P}_2, p_1 \neq p_2 \\ M\tilde{t}_{p_1,p_1,1}v_{p_1}, & p_1 \in \mathcal{P}_1, p_2 = p_1 + K \\ M\tilde{M}\tilde{t}_{p_1-K,p_1-K,2}, & p_1 \in \mathcal{P}_2, p_2 = p_1 + K. \end{cases}$$

According to the statement below (86), $\sum_{m=1}^{M}[\mathbf{A}]_{m,p_1}[\mathbf{A}]_{m,p_2}^*$ for any other combination of p_1 and p_2 is similar to one of the results above. As a result, the trace in (89) can be re-expressed as

$$\text{tr}\left(\mathbf{A}\hat{\mathbf{R}}_{\text{c}}\mathbf{A}^H\right) = M\left(\tilde{M}a_{\text{s}} + M_{\text{x}}b_{\text{s}} + M_{\text{y}}c_{\text{s}} + d_{\text{s}}\right), (90)$$

where a_{s} is a linear function of $\tilde{r}_{p-K,2}^2$, $\tilde{r}_{p-K,3}^2$, and the variables that are similar to $\tilde{t}_{p_1-K,p_1-K,2}$ (these variables have similar expressions); b_{s} and c_{s} are linear functions of variables like $\tilde{t}_{p_1,p_1,1}$; d_{s} is a linear function of variables like $\tilde{t}_{p_1,p_1,1}$, $t_{p_1,p_2,1}$, $t_{p_1,p_2-K,2}$, and $t_{p_1-K,p_2-K,3}$. It can be seen that a_{s}, b_{s}, c_{s}, and d_{s} are random variables. In addition, it is not difficult to verify that the means and variances of these random variables do not tend to infinity and a_{s} tends to be a positive number as $M \to \infty$.

On the other hand, according to the norms of the column vectors of $\mathbf{A}$, the trace of $\hat{\mathbf{R}}_{\text{n}}$ can be expressed as

$$\text{tr}\left(\hat{\mathbf{R}}_{\text{n}}\right) = M\left(M_{\text{x}}a_{\text{n}} + M_{\text{y}}b_{\text{n}} + \tilde{c}_{\text{n}}\right). \quad (91)$$

Similarly, it can be verified that the means and variances of a_{n}, b_{n}, and $\tilde{c}_{\text{n}}$ do not tend to infinity when $M \to \infty$.

From (68), we have

$$\text{tr}\left(\hat{\mathbf{R}}_{\text{x}}\right) = \text{tr}\left(\mathbf{A}\hat{\mathbf{R}}_{\text{c}}\mathbf{A}^H\right) + \text{tr}\left(\hat{\mathbf{R}}_{\text{n}}\right). \quad (92)$$

Because $\hat{\mathbf{R}}_{\text{x}}$ is a positive semi-definite matrix, the eigenvalue of $\hat{\mathbf{R}}_{\text{x}}$ can be expressed as

$$\lambda_{\text{x}m} = \xi_{\text{s}m}\text{tr}\left(\mathbf{A}\hat{\mathbf{R}}_{\text{c}}\mathbf{A}^H\right) + \xi_{\text{n}m}\text{tr}\left(\hat{\mathbf{R}}_{\text{n}}\right),$$

where $m = 1, 2, \cdots, M$, $0 \leq \xi_{\text{s}m} \leq 1$, $0 \leq \xi_{\text{n}m} \leq 1$, and $\sum_{m=1}^{M}\xi_{\text{s}m} = \sum_{m=1}^{M}\xi_{\text{n}m} = 1$. Because $\mathbf{A}\hat{\mathbf{R}}_{\text{c}}\mathbf{A}^H$ is a positive semi-definite matrix, and has at most $3K$ nonzero eigenvalues, we have $\xi_{\text{s}n_p} \geq 0$, where $n_p \in \{1, 2, \cdots, M\}, p = 1, 2, \cdots, 3K$, are different values, and $\xi_{\text{s}m'} = 0$ for $m' \in \{1, 2, \cdots, M\}, m' \neq n_p$. It can be seen that the eigenvectors that correspond to the eigenvalues $\lambda_{\text{x}n_p}, p = 1, 2, \cdots, 3K$, are in the column space of $\mathbf{A}$, and the other eigenvectors are in the null space of $\mathbf{A}$.

Since the condition number of $\mathbf{T}_{\text{A}}$ does not tend to infinity when $M \to \infty$, $\xi_{\text{s}n_p}$ does not tend to zero as $M \to \infty$. Then, from (90) and (91), we have

$$\frac{\lambda_{\text{x}n_p}}{\lambda_{\text{x}m'}} = \frac{\xi_{\text{s}n_p}\text{tr}\left(\mathbf{A}\hat{\mathbf{R}}_{\text{c}}\mathbf{A}^H\right) + \xi_{\text{n}n_p}\text{tr}\left(\hat{\mathbf{R}}_{\text{n}}\right)}{\xi_{\text{n}m'}\text{tr}\left(\hat{\mathbf{R}}_{\text{n}}\right)} \quad (93)$$

$$\overset{\text{a.s.}}{\to} \infty, \quad \text{as} \quad M \to \infty, \quad m' \neq n_p, \quad (94)$$

where $\overset{\text{a.s.}}{\to}$ denotes the almost sure convergence. Note that $\xi_{\text{s}m'}$ vanishes in (93) because $\xi_{\text{s}m'} = 0$ for $m' \neq n_p$. It can be seen that $\lambda_{\text{x}n_p}, p = 1, 2, \cdots, 3K$, tend almost surely to be the largest $3K$ eigenvalues of $\hat{\mathbf{R}}_{\text{x}}$, which means the columns of $\hat{\mathbf{E}}_{\text{s}}$ tend almost surely to be in the column space of $\mathbf{A}$ as $M \to \infty$. Therefore, we have proved that $\hat{\mathbf{E}}_{\text{s}}$ tends almost surely to be in the same subspace as $\mathbf{A}$ when $M \to \infty$.

Appendix B
Derivation of the Approximate CRB

First, the array manifold, cf. (2), for $\theta_{k,j}(t)$ and $\phi_{k,j}(t)$ is approximated by

$$[\mathbf{a}(\theta_{k,j}(t), \phi_{k,j}(t))]_m \approx \exp\left(iu\sin(\bar{\phi}_k)\left[(m_{\text{x}} - 1)\cos(\bar{\theta}_k)\right.\right.$$
$$\left.+(m_{\text{y}} - 1)\sin(\bar{\theta}_k)\right]\right) \times \exp\left(iu\tilde{\phi}_{k,j}(t)\cos(\bar{\phi}_k)\right.$$
$$\times \left[(m_{\text{x}} - 1)\cos(\bar{\theta}_k) + (m_{\text{y}} - 1)\sin(\bar{\theta}_k)\right]\right) \times \exp\left(iu\tilde{\theta}_{k,j}(t)\right.$$
$$\left.\times \sin(\bar{\phi}_k)\left[-(m_{\text{x}} - 1)\sin(\bar{\theta}_k) + (m_{\text{y}} - 1)\cos(\bar{\theta}_k)\right]\right), \quad (95)$$

where $m, m_{\text{x}}, m_{\text{y}}$ are defined in (2). This approximation is similar to that in [35]. The Taylor series expansion of (95) is different from that given in (10). According to (1) and (95), the covariance matrix $\mathbf{R}_{\text{x}}$ given by (19) can be reformulated as

$$\mathbf{R}_{\text{x}} \approx \sum_{k=1}^{K} \sigma_k^2 \mathbf{\Xi}_k + \sigma_{\text{n}}^2 \mathbf{I}_M, \quad (96)$$

where $\sigma_k^2 = S_k\sigma_{\gamma_k}^2$. It can be easily found that $\mathbf{\Xi}_k$ can be written as

$$\mathbf{\Xi}_k = \left(\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)\mathbf{a}^H(\bar{\theta}_k, \bar{\phi}_k)\right) \odot \mathbf{B}_k = \mathbf{D}_k\mathbf{B}_k\mathbf{D}_k^H, \quad (97)$$

where $\mathbf{D}_k = \text{diag}(\mathbf{a}(\bar{\theta}_k, \bar{\phi}_k)) \in \mathbb{C}^{M \times M}$, and each entry of $\mathbf{B}_k \in \mathbb{R}^{M \times M}$ equals

$$[\mathbf{B}_k]_{m,n} = \exp\left(-\left(\sigma_{\phi_k}^2\cos^2(\bar{\phi}_k)\left[\delta_{\text{x}}\cos(\bar{\theta}_k) + \delta_{\text{y}}\sin(\bar{\theta}_k)\right]^2\right.\right.$$
$$\left.\left.+\sigma_{\theta_k}^2\sin^2(\bar{\phi}_k)\left[-\delta_{\text{x}}\sin(\bar{\theta}_k) + \delta_{\text{y}}\cos(\bar{\theta}_k)\right]^2\right) \times \frac{1}{2}u^2\right). (98)$$

At the end of Section II, it is already verified that the received signal $\mathbf{x}(t)$ in (1) is a zero-mean circularly symmetric

complex-valued Gaussian vector. Then, the Fisher information matrix (FIM) can be used to derive the CRB. Because the received signal is approximated with the aid of (95), we can only derive the approximate FIM and the approximate CRB. Let us define $\mathbf{u} = [\mathbf{u}_{\bar{\theta}}^T, \mathbf{u}_{\bar{\phi}}^T, \mathbf{u}_{\sigma_\theta}^T, \mathbf{u}_{\sigma_\phi}^T]^T \in \mathbb{R}^{4K \times 1}$, $\mathbf{v} = [\sigma_1^2, \sigma_2^2, \cdots, \sigma_K^2, \sigma_n^2]^T \in \mathbb{R}^{(K+1) \times 1}$, and $\boldsymbol{\xi} = [\mathbf{u}^T, \mathbf{v}^T]^T \in \mathbb{R}^{(5K+1) \times 1}$, where $\mathbf{u}_{\bar{\theta}} = [\bar{\theta}_1, \bar{\theta}_2, \cdots, \bar{\theta}_K]^T \in \mathbb{R}^{K \times 1}$, $\mathbf{u}_{\bar{\phi}} = [\bar{\phi}_1, \bar{\phi}_2, \cdots, \bar{\phi}_K]^T \in \mathbb{R}^{K \times 1}$, $\mathbf{u}_{\sigma_\theta} = [\sigma_{\theta_1}, \sigma_{\theta_2}, \cdots, \sigma_{\theta_K}]^T \in \mathbb{R}^{K \times 1}$, and $\mathbf{u}_{\sigma_\phi} = [\sigma_{\phi_1}, \sigma_{\phi_2}, \cdots, \sigma_{\phi_K}]^T \in \mathbb{R}^{K \times 1}$, the approximate (finite-sample) FIM $\mathbf{J}_{\boldsymbol{\xi},\boldsymbol{\xi}} \in \mathbb{R}^{(5K+1) \times (5K+1)}$ is then expressed as [51, p. 525]

$$[\mathbf{J}_{\boldsymbol{\xi},\boldsymbol{\xi}}]_{q,q'} = T \operatorname{tr}\left(\mathbf{R_x}^{-1} \frac{\partial \mathbf{R_x}}{\partial [\boldsymbol{\xi}]_q} \mathbf{R_x}^{-1} \frac{\partial \mathbf{R_x}}{\partial [\boldsymbol{\xi}]_{q'}} \right), \quad (99)$$

where $q = 1, 2, \cdots, 5K+1$, $q' = 1, 2, \cdots, 5K+1$, and T is the number of received signal snapshots. From (96) and (97), the following partial derivatives may be obtained, which are

$$\frac{\partial \mathbf{R_x}}{\partial \bar{\theta}_k} \approx \sigma_k^2 \Big(\mathbf{D}_{\bar{\theta}_k} \mathbf{D}_k \mathbf{B}_k \mathbf{D}_k^H - \mathbf{D}_k \mathbf{B}_k \mathbf{D}_k^H \mathbf{D}_{\bar{\theta}_k} + \mathbf{D}_k (\mathbf{B}_k \odot \mathbf{B}_{\bar{\theta}_k}) \mathbf{D}_k^H \Big),$$

$$\frac{\partial \mathbf{R_x}}{\partial \bar{\phi}_k} \approx \sigma_k^2 \Big(\mathbf{D}_{\bar{\phi}_k} \mathbf{D}_k \mathbf{B}_k \mathbf{D}_k^H - \mathbf{D}_k \mathbf{B}_k \mathbf{D}_k^H \mathbf{D}_{\bar{\phi}_k} + \mathbf{D}_k (\mathbf{B}_k \odot \mathbf{B}_{\bar{\phi}_k}) \mathbf{D}_k^H \Big),$$

$$\frac{\partial \mathbf{R_x}}{\partial \sigma_{\theta_k}} \approx \sigma_k^2 \left(\mathbf{D}_k (\mathbf{B}_k \odot \mathbf{B}_{\sigma_\theta,k}) \mathbf{D}_k^H \right),$$

$$\frac{\partial \mathbf{R_x}}{\partial \sigma_{\phi_k}} \approx \sigma_k^2 \left(\mathbf{D}_k (\mathbf{B}_k \odot \mathbf{B}_{\sigma_\phi,k}) \mathbf{D}_k^H \right),$$

$$\frac{\partial \mathbf{R_x}}{\partial \sigma_k^2} \approx \mathbf{D}_k \mathbf{B}_k \mathbf{D}_k^H,$$

and

$$\frac{\partial \mathbf{R_x}}{\partial \sigma_n^2} \approx \mathbf{I}_M,$$

where $\mathbf{D}_{\bar{\theta}_k} \in \mathbb{C}^{M \times M}$, $\mathbf{D}_{\bar{\phi}_k} \in \mathbb{C}^{M \times M}$, $\mathbf{B}_{\bar{\theta}_k} \in \mathbb{C}^{M \times M}$, $\mathbf{B}_{\bar{\phi}_k} \in \mathbb{C}^{M \times M}$, $\mathbf{B}_{\sigma_\theta,k} \in \mathbb{C}^{M \times M}$, and $\mathbf{B}_{\sigma_\phi,k} \in \mathbb{C}^{M \times M}$ are defined as

$$[\mathbf{D}_{\bar{\theta}_k}]_{m,m} = iu \sin(\bar{\phi}_k)[-(m_{\mathrm{x}} - 1) \sin(\bar{\theta}_k) + (m_{\mathrm{y}} - 1) \cos(\bar{\theta}_k)],$$

$$[\mathbf{D}_{\bar{\phi}_k}]_{m,m} = iu \cos(\bar{\phi}_k)[(m_{\mathrm{x}} - 1) \cos(\bar{\theta}_k) + (m_{\mathrm{y}} - 1) \sin(\bar{\theta}_k)],$$

$$[\mathbf{B}_{\bar{\theta}_k}]_{m,n} = -\frac{1}{2} u^2 \big[-\sigma_{\phi_k}^2 \cos^2(\bar{\phi}_k) + \sigma_{\theta_k}^2 \sin^2(\bar{\phi}_k) \big] \times \big[(\delta_{\mathrm{x}}^2 - \delta_{\mathrm{y}}^2) \sin(2\bar{\theta}_k) - 2\delta_{\mathrm{x}} \delta_{\mathrm{y}} \cos(2\bar{\theta}_k) \big],$$

$$[\mathbf{B}_{\bar{\phi}_k}]_{m,n} = -\frac{1}{2} u^2 \sin(2\bar{\phi}_k) \big(-\sigma_{\phi_k}^2 \left[\delta_{\mathrm{x}} \cos(\bar{\theta}_k) + \delta_{\mathrm{y}} \sin(\bar{\theta}_k) \right]^2 + \sigma_{\theta_k}^2 \left[-\delta_{\mathrm{x}} \sin(\bar{\theta}_k) + \delta_{\mathrm{y}} \cos(\bar{\theta}_k) \right]^2 \big),$$

$$[\mathbf{B}_{\sigma_\theta,k}]_{m,n} = -u^2 \sigma_{\theta_k} \sin^2(\bar{\phi}_k) \left[-\delta_{\mathrm{x}} \sin(\bar{\theta}_k) + \delta_{\mathrm{y}} \cos(\bar{\theta}_k) \right]^2,$$

and

$$[\mathbf{B}_{\sigma_\phi,k}]_{m,n} = -u^2 \sigma_{\phi_k} \cos^2(\bar{\phi}_k) \left[\delta_{\mathrm{x}} \cos(\bar{\theta}_k) + \delta_{\mathrm{y}} \sin(\bar{\theta}_k) \right]^2,$$

respectively. Note that $\mathbf{D}_{\bar{\theta}_k}, \mathbf{D}_{\bar{\phi}_k}$ are diagonal matrices. Similar to (99), $\mathbf{J}_{\mathbf{u},\mathbf{u}} \in \mathbb{R}^{4K \times 4K}$, $\mathbf{J}_{\mathbf{u},\mathbf{v}} \in \mathbb{R}^{4K \times (K+1)}$, and

$\mathbf{J}_{\mathbf{v},\mathbf{v}} \in \mathbb{R}^{(K+1) \times (K+1)}$ can be defined, and they are related to $\mathbf{J}_{\boldsymbol{\xi},\boldsymbol{\xi}}$ as

$$\mathbf{J}_{\boldsymbol{\xi},\boldsymbol{\xi}} = \left[\begin{array}{cc} \mathbf{J}_{\mathbf{u},\mathbf{u}} & \mathbf{J}_{\mathbf{u},\mathbf{v}} \\ \mathbf{J}_{\mathbf{u},\mathbf{v}}^T & \mathbf{J}_{\mathbf{v},\mathbf{v}} \end{array} \right]. \quad (100)$$

Then, by the simple block matrix inversion lemma [52], the approximate CRB concerning the covariance matrix of the estimation error of the angular parameter vector $\mathbf{u}$ is obtained as (72) and (73).

REFERENCES

[1] D. Gesbert, M. Kountouris, R. W. Heath, C.-B. Chae, and T. Sälzer, "Shifting the MIMO paradigm," *IEEE Signal Processing Mag.*, vol. 24, no. 5, pp. 36–46, Sep. 2007.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Mag.*, vol. 30, no. 1, pp. 40–60, Jan. 2013.

[3] T. L. Marzetta, "Noncooperative cellular wireless with unlimited numbers of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 9, pp. 3590–3600, Nov. 2010.

[4] J. Zhang, C.-K. Wen, S. Jin, X. Gao, and K.-K. Wong, "On capacity of large-scale MIMO multiple access channels with distributed sets of correlated antennas," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 133–148, Feb. 2013.

[5] L. Dai, Z. Wang, and Z. Yang, "Spectrally efficient time-frequency training OFDM for mobile large-scale MIMO systems," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 251–263, Feb. 2013.

[6] X. Su, J. Zeng, L.-P. Rong, and Y.-J. Kuang, "Investigation on key technologies in large-scale MIMO," *J. Comput. Sci. Technology*, vol. 28, no. 3, pp. 412–419, May 2013.

[7] W. Ding and T. Lv, "Multiuser MIMO using block diagonalization: How many users should be served?" in *Proc. IEEE Global Telecommun. Conf. (GLOBECOM 2013)*, Atlanta, GA, Dec. 2013, pp. 4315–4320.

[8] H. Liu, H. Gao, A. Hu, and T. Lv (2014, Mar. 19). "Low-complexity transmission mode selection in MU-MIMO systems," [Online]. Available: http://arxiv.org/abs/1403.4374.

[9] O. N. Alrabadi, E. Tsakalaki, H. Huang, and G. F. Pedersen, "Beamforming via large and dense antenna arrays above a clutter," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 2, pp. 314–325, Feb. 2013.

[10] J. Koppenborg, H. Halbauer, S. Saur, and C. Hoek, "3D beamforming trials with an active antenna array," in *Proc. Int. ITG Workshop Smart Antennas*, Dresden, Germany, Mar. 2012, pp. 110–114.

[11] H. Halbauer, S. Saur, J. Koppenborg, and C. Hoek, "3D beamforming: Performance improvement for cellular networks," *Bell Labs Tech. J.*, vol. 18, no. 2, pp. 37–56, Sep. 2013.

[12] L. C. Godara, "A robust adaptive array processor," *IEEE Trans. Circuits Syst.*, vol. 34, pp. 721–730, Jul. 1987.

[13] C. L. Zahm, "Effects of errors in the direction of incidence on the performance of an adaptive array," *Proc. IEEE*, vol. 60, no. 8, pp. 1008–1009, Aug. 1972.

[14] J. W. Kim and C. K. Un, "A robust adaptive array based on signal subspace approach," *IEEE Trans. Signal Process.*, vol. 41, pp. 3166–3171, Nov. 1993.

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

[16] Y. U. Lee, J. Choi, I. Song, and S. R. Lee, "Distributed source modeling and direction-of-arrival estimation techniques," *IEEE Trans. Signal Process.*, vol. 45, pp. 960–969, Apr. 1997.

[17] D. Astély and B. Ottersten, "The effects of local scattering on direction of arrival estimation with MUSIC," *IEEE Trans. Signal Process.*, vol. 47, pp. 3220–3234, Apr. 1999.

[18] S. Valaee, B. Champagne, and P. Kabal, "Parametric localization of distributed sources," *IEEE Trans. Signal Process.*, vol. 43, pp. 2144–2153, Sep. 1995.

[19] S. Shahbazpanahi, S. Valaee, and M. H. Bastani, "Distributed source localization using ESPRIT algorithm," *IEEE Trans. Signal Process.*, vol. 49, pp. 2169–2178, Oct. 2001.

[20] J. Lee, I. Song, H. Kwon, and S. R. Lee, "Low-complexity estimation of 2-D DOA for coherently distributed sources," *Signal Process.*, vol. 83, no. 8, pp. 1789–1802, Aug. 2003.

[21] Z. Zheng and G. Li, "Low-complexity estimation of DOA and angular spread for an incoherently distributed source," *Wireless Personal Commun.*, vol. 70, no. 4, pp. 1653–1663, Jun. 2013.

[22] M. Bengtsson and B. Ottersten, "Low-complexity estimators for distributed sources," *IEEE Trans. Signal Process.*, vol. 48, pp. 2185–2194, Aug. 2000.

[23] S. Shahbazpanahi, S. Valaee, and A. B. Gershman, "A covariance fitting approach to parametric localization of multiple incoherently distributed sources," *IEEE Trans. Signal Process.*, vol. 52, pp. 592–600, Mar. 2004.

[24] M. Tapio, "Direction and spread estimation of spatially distributed signals via the power azimuth spectrum," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP 2002)*, Orlando, FL, May 2002, pp. 3005–3008.

[25] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Process.*, vol. 50, no. 1/2, pp. 57–69, Apr. 1996.

[26] B. T. Sieskul, "An asymptotic maximum likelihood for joint estimation of nominal angles and angular spreads of multiple spatially distributed sources," *IEEE Trans. Signal Process.*, vol. 59, pp. 1534–1538, Mar. 2010.

[27] M. Ghogho, O. Besson, and A. Swami, "Estimation of direction of arrival of multiple scattered sources," *IEEE Trans. Signal Process.*, vol. 49, pp. 2467–2480, Nov. 2001.

[28] O. Besson, F. Vincent, and P. Stoica, "Approximate maximum likelihood estimators for array processing in multiplicative noise environments," *IEEE Trans. Signal Process.*, vol .48, pp. 2506–2518, Sep. 2000.

[29] A. B. Gershman, C. F. Mecklenbräuker, and J. F. Böhme, "Matrix fitting approach to direction of arrival estimation with imperfect coherence of wavefronts," *IEEE Trans. Signal Process.*, vol. 45, pp. 1894–1899, Jul. 1997.

[30] O. Besson and P. Stoica, "Decoupled estimation of DOA and angular spread for a spatially distributed source," *IEEE Trans. Signal Process.*, vol. 48, pp. 1872–1882, Jul. 2000.

[31] O. Besson, P. Stoica, and A. B. Gershman, "Simple and accurate direction of arrival estimator in the case of imperfect spatial coherence," *IEEE Trans. Signal Process.*, vol. 49, pp. 730–737, Jul. 2001.

[32] A. Monakov and O. Besson, "Direction finding for an extended target with possibly non-symmetric spatial spectrum," *IEEE Trans. Signal Process.*, vol. 52, pp. 283–287, Jan. 2004.

[33] A. Zoubir, Y. Wang, and P. Chargé, "A modified COMET-EXIP method for estimating a scattered source," *Signal Process.*, vol. 86, no. 4, pp. 733–743, Apr. 2006.

[34] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing applications," *Digital Signal Process.*, vol. 8, no. 3, pp. 185–210, Jul. 1998.

[35] H. Boujemâa, "Extension of COMET algorithm to multiple diffuse source localization in azimuth and elevation," *European Trans. Telecommun.*, vol. 16, no. 6, pp. 557–566, Nov./Dec. 2005.

[36] M. Bengtsson and B. Ottersten, "A generalization of weighted subspace fitting to full-rank models," *IEEE Trans. Signal Process.*, vol. 49, pp. 1002–1012, May 2001.

[37] Y. Meng, P. Stoica, and K. M. Wong, "Estimation of the directions of arrival of spatially dispersed signals in array processing," *IEE Proc. Radar, Sonar and Navigation*, vol. 143, no. 1, pp. 1–9, Feb. 1996.

[38] M. Bengtsson, "Antenna array signal processing for high rank models," Ph.D. dissertation, Signals, Sensors, Syst. Dept., Royal Institute of Technology, Stockholm, Sweden, 1999.

[39] A. Hassanien, S. Shahbazpanahi, and A. B. Gershman, "A generalized Capon estimator for localization of multiple spread sources," *IEEE Trans. Signal Process.*, vol. 52, pp. 280–283, Jan. 2004.

[40] A. Zoubir, Y. Wang, and P. Chargé, "Efficient subspace-based estimator for localization of multiple incoherently distributed sources," *IEEE Trans. Signal Process.*, vol. 56, pp. 532–542, Feb. 2008.

[41] J. Zhou, Z. Zheng, and G. Li, "Low-complexity estimation of the nominal azimuth and elevation for incoherently distributed sources," *Wireless Personal Commun.*, vol. 71, no. 3, pp. 1777–1793, Aug. 2013.

[42] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, pp. 276–280, Mar. 1986.

[43] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proc. IEEE*, vol. 74, no. 7, 1044–1046, Jul. 1986.

[44] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT–A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, pp. 1340–1342, Oct. 1986.

[45] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 37, pp. 984–995, Jul. 1989.

[46] M. Haardt, M. D. Zoltowski, C. P. Mathews, and J. A. Nossek, "2D unitary ESPRIT for efficient 2D parameter estimation," in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process. (ICASSP 1995)*, Detroit, MI, May 1995, pp. 2096–2099.

[47] P. Heidenreich, A. M. Zoubir, and M. Rübsamen, "Joint 2-D DOA estimation and phase calibration for uniform rectangular arrays," *IEEE Trans. Signal Process.*, vol. 60, pp. 4683–4693, Sep. 2012.

[48] P. Zetterberg, "Mobile cellular communications with base station antenna arrays: Spectrum efficiency, algorithms and propagation models," Ph.D. dissertation, Royal Institute of Technology, Stockholm, Sweden, 1997.

[49] G. H. Golub and C. F. V. Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: The Johns Hopkins University Press, 1996.

[50] B. Karakaya, H. Arslan, and H. A. Cirpan, "Channel estimation for LTE uplink in high Doppler spread," in *Proc. IEEE Wireless Commun. Networking Conf. (WCNC 2008)*, Las Vegas, NV, Mar. 2008, pp. 1126–1130.

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

[52] S. E. Jo, S. W. Kim, and T. J. Park, "Equally constrained affine projection algorithm," in *Proc. 38th Asilomar Conf. Signals, Syst. and Comput.*, Pacific Grove, CA, Nov. 2004, pp. 955–959.