Inconsistency of ESPRIT DoA Estimation for Large Arrays and a Correction via RMT

Wei Yang*, Zhengyu Wang*, Xiaoyi Mai†, Zenan Ling*, Robert C. Qiu*, Zhenyu Liao*
*School of Electronic Information and Communications (EIC), Huazhong University of Science and Technology, Wuhan, China
Email: {yangwei_eic, wangzhengyu, lingzenan, caiming, zhenyu_liao}@edu.hust.edu
†Institut de Mathématiques de Toulouse (IMT), University of Toulouse-Jean Jaurès, Toulouse, France
Email: xiaoyi.mai@math.univ-toulouse.fr

***Abstract*—Recent advances in random matrix theory (RMT) have shed light on the *inadequacy* of the sample covariance matrix (SCM) as an estimator for population covariance in large-dimensional scenarios. However, the applicability of this insight to the widely used SCM-based ESPRIT Direction of Arrival (DoA) estimation technique remains *unresolved*. This paper addresses this gap by investigating the asymptotic behavior of ESPRIT in the context of *large arrays with limited samples*, where the number of samples/snapshots T and the number of sensors N are both large and comparable. We demonstrate that classical ESPRIT yields *inconsistent* DoA estimates as $N, T \to \infty$ at the same pace. We propose an improved G-ESPRIT method and prove its asymptotic consistency in the same setting. Numerical simulations are presented to validate our theoretical assertions.**

***Index Terms*—DoA estimation, ESPRIT, random matrix theory, sample covariance matrix, subspace method**

I. Introduction

Commonly used subspace Direction-of-Arrival (DoA) methods such as MUSIC [1] and ESPRIT [2] propose to retrieve DoA structural information from the sample covariance matrix (SCM) of received signals. For a SCM computed from T snapshots on an array of N sensors, it is now well known that in the *large array and limited sample regime* when T is *not much larger* than N, the SCM is a poor estimator of the population covariance in an eigenspectral sense (see [3] and Section II-B below for a brief review). As a consequence, one should *not* expect that subspace methods could provide consistent estimates of the true DoAs in the setting where N, T are both large and comparable.

With the progress of random matrix theory (RMT) over the past decade, many methods in statistics, signal processing, and machine learning have been revisited in the large-dimensional setting, resulting in novel insights and improved algorithms better suited for large-dimensional data [3], [4]. In the case of subspace DoA methods, it has been shown in [5] that the popular MUSIC method, despite the eigenspectral inconsistency of SCM in the large N, T regime, still provides consistent DoA estimates in widely spaced DoA scenario (see Assumption 3 below for a precise definition). In the case of closely spaced sources, however, the classical MUSIC approach *is bound to fail*, and improved estimators such as G-MUSIC should be used instead [5], [6].

In this paper, we propose to analyze the equally popular DoA subspace method ESPRIT [2] (to be reviewed in Section II-A below) in the *large array and limited sample* regime. While it has been empirically observed that ESPRIT outperforms MUSIC in some cases [7] but not in others [8], its theoretical assessment in the large-dimensional setting **remains an open problem**, see [9], due to its mathematically involved form compared to, e.g., the MUSIC method.

Our main results can be summarized as follows:

1) we prove in the *large array and limited sample* regime that the standard ESPRIT method (Algorithm 1) provides **inconsistent estimates** of the true DoAs (Theorem 2), unless the sources are uncorrelated (Remark 3); and
2) we propose an improved method called **G-ESPRIT** (Algorithm 2) and show it provides asymptotically consistent DoA estimates in the same setting as $N, T \to \infty$.

II. Models and Preliminaries

A. ESPRIT DoA Estimation

We consider a unitary linear array (ULA) of N sensors that receives K narrow-band and far-field source signals with DoAs $\theta_1, \ldots, \theta_K$. The signal $\mathbf{x}(t) \in \mathbb{C}^N$ received by this array of sensors at time $t = 1, \ldots, T$ is given by

$$\mathbf{x}(t) = \textstyle\sum_{k=1}^{K} \mathbf{a}(\theta_k) s_k(t) + \mathbf{n}(t) \in \mathbb{C}^N, \tag{1}$$

with *deterministic* signal $s_k(t) \in \mathbb{C}$, $\mathbf{a}(\theta_k) \in \mathbb{C}^N$ the steering vector of source $k \in \{1, \ldots, K\}$ at DoA θ_k given by[1]

$$\mathbf{a}(\theta_k) = [1,\ e^{\imath\theta_k}, \ldots, e^{\imath(N-1)\theta_k}]^{\mathsf{T}}/\sqrt{N} \in \mathbb{C}^N, \tag{2}$$

and complex circular Gaussian noise $\mathbf{n}(t) \in \mathbb{C}^N$ having i.i.d. $\mathcal{CN}(0,1)$ entries. This model rewrites in matrix form as

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}, \quad \mathbf{A} = [\mathbf{a}(\theta_1), \ldots, \mathbf{a}(\theta_K)] \in \mathbb{C}^{N \times K}, \tag{3}$$

with $\mathbf{X} = [\mathbf{x}(1), \ldots, \mathbf{x}(T)] \in \mathbb{C}^{N \times T}$ the matrix of received signals, $\mathbf{A} \in \mathbb{C}^{N \times K}$ the matrix of steering vectors, $\mathbf{S} = [\mathbf{s}(1), \ldots, \mathbf{s}(T)] \in \mathbb{C}^{K \times T}$ the matrix containing source signals, and random noise $\mathbf{N} = [\mathbf{n}(1), \ldots, \mathbf{n}(T)] \in \mathbb{C}^{N \times T}$.

[1]The normalization by $\sqrt{N}$ is made so that $\mathbf{a}(\theta_k)$ is of unit norm. Here, we use θ_k for the DoA in the Fourier space as in [5], which is related to the physical angle ϕ_k of the source wave via $\theta_k = \frac{2\pi d}{\lambda_0} \sin(\phi_k)$.

Under (3), subspace DoA methods such as MUSIC and ESPRIT are based the following observation on the *population* covariance $\mathbf{C} \in \mathbb{C}^{N\times N}$ of the received signal:

$$\begin{aligned}\mathbf{C} \equiv \mathbb{E}[\mathbf{X}\mathbf{X}^{\mathsf{H}}]/T &= \mathbf{A}\mathbf{S}\mathbf{S}^{\mathsf{H}}\mathbf{A}^{\mathsf{H}}/T + \mathbb{E}[\mathbf{N}\mathbf{N}^{\mathsf{H}}]/T \\ &= \mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}} + \mathbf{I}_N, \end{aligned} \tag{4}$$

for signal power matrix $\mathbf{P} = \mathbf{S}\mathbf{S}^{\mathsf{H}}/T \in \mathbb{C}^{K\times K}$. As such, the top-$K$ subspace (that corresponds to the largest K eigenvalues, also known as the "signal subspace") of the *population* covariance $\mathbf{C}$ is closely connected to the steering vectors $\mathbf{a}(\theta_k)$ and can be used for DoA estimation. In practice, the *population* covariance $\mathbf{C}$ in (4) is not accessible and one uses instead the sample covariance matrix (SCM) $\hat{\mathbf{C}}$ built from observations of T snapshots given by

$$\hat{\mathbf{C}} = \mathbf{X}\mathbf{X}^{\mathsf{H}}/T. \tag{5}$$

The ESPRIT method [2] then relies the following structure of rotational invariance: For steering matrix $\mathbf{A} \in \mathbb{C}^{N\times K}$ defined in (3) and $\mathbf{J}_1,\ \mathbf{J}_2 \in \mathbb{R}^{n\times N}$ two selection matrices that select n among N rows of $\mathbf{A}$ with distance $\Delta \geq 1$, i.e.,

$$\mathbf{J}_1^{\mathsf{T}} = [\mathbf{e}_\ell, \ldots, \mathbf{e}_{n+\ell-1}],\ \mathbf{J}_2^{\mathsf{T}} = [\mathbf{e}_{\ell+\Delta}, \ldots, \mathbf{e}_{n+\ell+\Delta-1}], \tag{6}$$

for $\mathbf{e}_i$ the canonical vector of $\mathbb{R}^N$ with $[\mathbf{e}_i]_j = \delta_{ij}$, one has

$$\mathbf{J}_1\mathbf{A}\operatorname{diag}\{e^{\imath\Delta\theta_k}\}_{k=1}^K = \mathbf{J}_2\mathbf{A}. \tag{7}$$

While the steering matrix $\mathbf{A}$ is unknown, it follows from (4) that the top-K subspace $\mathbf{U}_K \in \mathbb{C}^{N\times K}$ of $\mathbf{C}$ is the same as the subspace spanned by the columns of $\mathbf{A}\mathbf{P}^{-1/2}$, so that

$$\mathbf{U}_K = \mathbf{A}\mathbf{P}^{-1/2}\mathbf{M}, \tag{8}$$

for some $\mathbf{M} \in \mathbb{C}^{K\times K}$. Combing (7) with (8), the DoAs θ_k can be written as the angles of the complex eigenvalues of

$$\mathbf{\Phi} = (\mathbf{U}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\mathbf{U}_K)^{-1}\mathbf{U}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_2\mathbf{U}_K \equiv \mathbf{\Phi}_1^{-1}\mathbf{\Phi}_2, \tag{9}$$

assuming invertible $\mathbf{\Phi}_1 \equiv \mathbf{U}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\mathbf{U}_K \in \mathbb{C}^{K\times K}$. Then, ESPRIT proposes to estimate, when SCM $\hat{\mathbf{C}}$ is "close" to the population covariance $\mathbf{C}$, by replacing the population subspace $\mathbf{U}_K$ in (9) with the empirical estimate $\hat{\mathbf{U}}_K$ obtained from the SCM $\hat{\mathbf{C}}$. This leads to the ESPRIT DoA estimation procedure summarized in Algorithm 1.

B. Eigenspectral Inconsistency for Large-dimensional SCM

When the number of snapshots T is *much larger* than the array length N (i.e., as $T \to \infty$ with N fixed), the SCM $\hat{\mathbf{C}}$ in (5) is known to be a good estimate of population covariance $\mathbf{C}$, as a consequence of the strong law of large numbers. In the case of large array and/or limited sample with N, T of the same order of magnitude, $\hat{\mathbf{C}}$ is *not* a consistent estimator of $\mathbf{C}$ in a eigenspectral norm sense, and we should, a priori, *not* expect that the top subspace $\hat{\mathbf{U}}_K$ used in ESPRIT in Algorithm 1 is a good estimate of the true signal subspace.

In the following, we recall a few RMT results that provide precise eigenspectral characterization of the SCM in the large N, T regime, under the following assumptions.

Algorithm 1 ESPRIT DoA estimation [2]

Input: Received signal $\mathbf{X} \in \mathbb{C}^{N\times T}$, number of sources K.
Output: Estimated DoA angles $\hat{\theta}_k, k \in \{1, \ldots, K\}$.

1: Compute the SCM $\hat{\mathbf{C}} = \mathbf{X}\mathbf{X}^{\mathsf{H}}/T$ as in (5) to retrieve $\hat{\mathbf{U}}_K = [\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_K] \in \mathbb{C}^{N\times K}$ the estimated signal subspace composed of the top-K eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_K \in \mathbb{C}^N$ associated to the largest K eigenvalues of $\hat{\mathbf{C}}$;
2: Define two selection matrices $\mathbf{J}_1, \mathbf{J}_2 \in \mathbb{R}^{n\times N}$ as in (6) that both select n among N rows with a "distance" $\Delta \geq 1$;
3: Compute $\hat{\mathbf{\Phi}} = (\hat{\mathbf{U}}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\hat{\mathbf{U}}_K)^{-1}\hat{\mathbf{U}}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_2\hat{\mathbf{U}}_K \in \mathbb{C}^{K\times K}$, for invertible $\hat{\mathbf{U}}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\hat{\mathbf{U}}_K$, and then the *angles* of $\lambda_k(\hat{\mathbf{\Phi}})$, the kth (complex) eigenvalue of $\hat{\mathbf{\Phi}}$;
4: **return** $\hat{\theta}_k = \arg(\lambda_k(\hat{\mathbf{\Phi}}))/\Delta$.

Assumption 1 (Large array and limited sample regime)**.** *For N the array size, T the number of samples/snapshots, and n the size of selection matrices $\mathbf{J}_1, \mathbf{J}_2 \in \mathbb{R}^{n\times N}$ as defined in* (6)*, we have, as $T \to \infty$ that*

(i) $N/T \to c \in (0,\infty)$ and $n/N \to \tau \in (0,1)$, with the number of sources K fixed; and
(ii) the deterministic *signal matrix $\mathbf{S} \in \mathbb{C}^{K\times T}$ is such that $\mathbf{P} = \mathbf{S}\mathbf{S}^{\mathsf{H}}/T$ remains bounded as $T \to \infty$.*

For $\mathbf{A}$ the steering matrix as defined in (3), let the eigendecomposition of $\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}} \in \mathbb{C}^{N\times N}$ be

$$\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}} = \textstyle\sum_{k=1}^K \lambda_k(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}})\cdot\mathbf{u}_k(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}})\mathbf{u}_k^{\mathsf{H}}(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}}). \tag{10}$$

We assume that the eigenvalues of $\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}}$ are large enough to separate from the random noise in the following sense.

Assumption 2 (Subspace separation)**.** *Under the settings and notations of Assumption 1, we have, as $N, T \to \infty$, that the largest K eigenvalues $\lambda_k(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}})$ of $\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}}$ satisfy*

$$\lambda_1(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}}) \to \ell_1 > \ldots > \lambda_K(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}}) \to \ell_K > \sqrt{c}. \tag{11}$$

Under Assumptions 1 and 2, we have the following RMT result that precisely characterizes the SCM eigenspectral behavior in the large N, T regime, due to a sequence of remarkable previous efforts [10]–[13]. See also [3, Chapter 2] for a review.

Theorem 1 (Eigenspectral characterization of large SCM)**.** *Let Assumption 1 hold, we have, for $\mathbf{X} \in \mathbb{C}^{N\times T}$ as defined in* (3) *and as $N, T \to \infty$ with $N/T \to c \in (0,\infty)$ that, with probability one, the eigenvalue distribution of the SCM $\hat{\mathbf{C}} = \mathbf{X}\mathbf{X}^{\mathsf{H}}/T$ converges weakly to the Marčenko-Pastur law:*

$$\mu(dx) = (1 + c^{-1})^+\delta_0(x) + \frac{\sqrt{(x-E_-)^+(E_+-x)^+}\,dx}{2\pi c x},$$

with $E_\pm = (1 \pm \sqrt{c})^2$ and $(x)^+ = \max(x,0)$. Moreover, let Assumption 2 hold and let $\hat{\lambda}_1 > \ldots > \hat{\lambda}_N$ be the eigenvalues of $\hat{\mathbf{C}}$ listed in a decreasing order with corresponding eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_N \in \mathbb{C}^N$, we have

$$\hat{\lambda}_i \to \begin{cases} \bar{\lambda}_i = 1 + \ell_i + c\frac{1+\ell_i}{\ell_i} > E_+, & 1 \leq i \leq K, \\ E_+ = (1+\sqrt{c})^2, & i > K; \end{cases} \tag{12}$$

and for all deterministic vectors $\mathbf{a}, \mathbf{b} \in \mathbb{C}^N$ *of bounded norm,*

$$\mathbf{a}^{\mathsf{H}} \hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^{\mathsf{H}} \mathbf{b} - \frac{1 - c\ell_k^{-2}}{1 + c\ell_k^{-1}} \mathbf{a}^{\mathsf{H}} \mathbf{u}_k \mathbf{u}_k^{\mathsf{H}} \mathbf{b} \to 0, \; k \in \{1, \ldots, K\}, \quad (13)$$

almost surely as $N, T \to \infty$*, with* $\mathbf{u}_k \equiv \mathbf{u}_k(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}})$ *in* (10).

Theorem 1 states that for N, T both large and comparable, the top eigenvalues $\hat{\lambda}_k$ of the SCM $\hat{\mathbf{C}}$ (that are due to the "signal" $\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}}$ per (4)), instead of being close to those of its population counterpart $\mathbf{C} = \mathbf{I}_N + \mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}}$,

(i) are larger than the *population* eigenvalues $(1 + \ell_k)$, by a term that is proportional to the dimension ratio c; and
(ii) have their associated eigenvectors $\hat{\mathbf{u}}_k$ being "biased" estimate of the *population* eigenvectors $\mathbf{u}_k(\mathbf{A}\mathbf{P}\mathbf{A}^{\mathsf{H}})$, in the sense that for arbitrary deterministic $\mathbf{a}, \mathbf{b} \in \mathbb{C}^N$, the eigenspace $\hat{\mathbf{u}}_k \hat{\mathbf{u}}_k^{\mathsf{H}}$ is "biased" by a factor of $(1 - c\ell_k^{-2})/(1 + c\ell_k^{-1})$ as stated in (13).

Note that these large-dimensional correction terms (in the empirical eigenvalues or eigenvectors from their population counterparts) vanish in the limit of infinite snapshots as $c = \lim N/T \to 0$ *or* in the high signal-to-noise ratio (SNR) regime as $\ell_k \to \infty$.

III. Inconsistency of ESPRIT for Large Arrays

Built upon recent advances in RMT, we perform in this section an in-depth analysis of the classical ESPRIT method in Algorithm 1 for large arrays as $N, T \to \infty$ at the same pace. We show that in general classical ESPRIT provides *inconsistent* estimates of the DoAs in the case of widely spaced DoAs defined as follows.

Assumption 3 (Widely spaced DoAs)**.** *The DoAs* $\theta_1, \ldots, \theta_K$ *are fixed as* $N \to \infty$*. This corresponds DoAs having angular separation much larger than a beam-width* $2\pi/N$*.*

The case of widely spaced DoAs and large arrays as $N \to \infty$, corresponds to the case of (asymptotically) orthogonal steering vectors. This is discussed in the following remark.

Remark 1 (Steering matrix for widely spaced DoAs)**.** Under Assumptions 1 and 3, we have, as $N, n, T \to \infty$ at the same pace that $\mathbf{A}^{\mathsf{H}}\mathbf{A} \to \mathbf{I}_K$, $\mathbf{A}^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\mathbf{A} - \frac{n}{N}\mathbf{I}_K \to 0$, and

$$\mathbf{A}^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_2\mathbf{A} - \frac{n}{N} \operatorname{diag}\{e^{\imath\Delta\theta_i}\}_{i=1}^K \to 0. \quad (14)$$

This result will be exploited in our proof below.

A. Large-dimensional inconsistency of ESPRIT

Assume that $\hat{\mathbf{\Phi}}_1 \equiv \hat{\mathbf{U}}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\hat{\mathbf{U}}_K \in \mathbb{C}^{K \times K}$ is invertible (which happens with probability one in the large $n, N, T \to \infty$ limit), one has, for $\hat{\mathbf{\Phi}}$ defined in Algorithm 1 that

$$\hat{\mathbf{\Phi}} = (\hat{\mathbf{U}}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\hat{\mathbf{U}}_K)^{-1}\hat{\mathbf{U}}_K^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_2\hat{\mathbf{U}}_K \equiv \hat{\mathbf{\Phi}}_1^{-1}\hat{\mathbf{\Phi}}_2, \quad (15)$$

as empirical estimates of their population counterparts $\mathbf{\Phi}, \mathbf{\Phi}_1, \mathbf{\Phi}_2$ defined in (9), need be evaluated to assess the performance of ESPRIT.

In the following result, we provide a precise large-dimensional characterization of the classical ESPRIT method in the large array and limited sample regime.

Theorem 2 (Large-dimensional inconsistency of ESPRIT)**.** *Under Assumptions 1–3, let* $\hat{\theta}_k$ *denote the DoA estimates obtained from the ESPRIT method in Algorithm 1, we have,*

$$\hat{\theta}_k - \arg(\lambda_k(\bar{\mathbf{\Phi}}))/\Delta \to 0, \quad k \in \{1, \ldots, K\}, \quad (16)$$

almost surely as $N, T \to \infty$*, with* $\lambda_k(\bar{\mathbf{\Phi}})$ *the kth largest eigenvalue of* $\bar{\mathbf{\Phi}} = \bar{\mathbf{\Phi}}_1^{-1}\bar{\mathbf{\Phi}}_2$*,*

$$\begin{aligned} \bar{\mathbf{\Phi}}_1 &= \operatorname{diag}(\sqrt{\mathbf{g}})\mathbf{\Phi}_1 \operatorname{diag}(\sqrt{\mathbf{g}}) + \operatorname{diag}\{h_k\}_{k=1}^K, \\ \bar{\mathbf{\Phi}}_2 &= \operatorname{diag}(\sqrt{\mathbf{g}})\mathbf{\Phi}_2 \operatorname{diag}(\sqrt{\mathbf{g}}), \end{aligned} \quad (17)$$

with $\sqrt{\mathbf{g}} = [\sqrt{g_1}, \ldots, \sqrt{g_K}]^{\mathsf{T}} \in \mathbb{R}^K$ *and*

$$g_k \equiv \frac{1 - c\ell_k^{-2}}{1 + c\ell_k^{-1}} > 0, \quad h_k \equiv \frac{n}{N} \frac{c + c\ell_k^{-1}}{c + \ell_k} \geq 0. \quad (18)$$

Theorem 2 tells us that in the large N, T regime, the matrices $\hat{\mathbf{\Phi}}_1, \hat{\mathbf{\Phi}}_2$ used in ESPRIT obtained from the SCM $\hat{\mathbf{C}}$, due to the large-dimensional inconsistency of $\hat{\mathbf{C}}$ discussed in Section II-B, are "biased" from their population counterparts $\mathbf{\Phi}_1, \mathbf{\Phi}_2$ defined in (9). As a direct consequence of Theorem 2, we have, in the case of large arrays, that ESPRIT diverges from its original design discussed in Section II-A and should in general *not* be able to provide consistent DoA estimation.

In the following, we discuss scenarios where such large-dimensional inconsistency holds true (or not).

To start with, one may expect that in the limit of infinite snapshots and/or high SNR, the large-dimensional corrections in Theorem 1 vanish and, as a consequence, ESPRIT becomes consistent. This is true per the following remark.

Remark 2 (Limiting cases: infinite snapshots or high SNR)**.** In the limit of infinite snapshots as $c = \lim N/T \to 0$ *or* high SNR as $\ell_k \to \infty$, one has $g_k \to 0$ and $h_k \to 0$, so that $\bar{\mathbf{\Phi}} = \mathbf{\Phi}_1^{-1}\mathbf{\Phi}_2 = \mathbf{\Phi}$ and classical ESPRIT provides consistent DoA estimates in this setting.

Beyond the limiting cases of infinite snapshots or high SNR discussed in Remark 2, there exists scenarios in which classical ESPRIT is "lucky" enough so that the large-dimensional correction terms cancel out and leads to consistent DoA estimates. Below is an example of such special cases.

Remark 3 (Special case: Uncorrelated sources)**.** In the case of *uncorrelated* sources that $\mathbf{P} = \mathbf{S}\mathbf{S}^{\mathsf{H}}/T$ is (asymptotically) a diagonal matrix, we have, that the top-K population subspace $\mathbf{U}_K$ is approximately the same as that spanned by the steering vectors. And it follows from Remark 1 that $\mathbf{\Phi}_1 \simeq \mathbf{A}^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\mathbf{A} \simeq \frac{n}{N}\mathbf{I}_K$, $\mathbf{\Phi}_2 \simeq \frac{n}{N} \operatorname{diag}\{e^{\imath\Delta\theta_k}\}_{k=1}^K$, so that $\bar{\mathbf{\Phi}}$ defined in (17) writes

$$\bar{\mathbf{\Phi}} \simeq \operatorname{diag}(\sqrt{\mathbf{g}}) \operatorname{diag}\{e^{\imath\Delta\theta_k}\}_{k=1}^K \operatorname{diag}(\sqrt{\mathbf{g}}), \quad (19)$$

for $\mathbf{g}$ a real vector. As such, $\bar{\mathbf{\Phi}}$ has the same eigenvalues *angles* as $\mathbf{\Phi}$, so that by Theorem 2 we have $\hat{\theta}_k - \theta_k \to 0$ almost surely, and that the classical ESPRIT provides consistent DoA estimation in this setting.

In general, however, classical ESPRIT does *not* provide consistent DoA estimates. See the following example for a manifestation of this large-dimensional *inconsistency*.

Remark 4 (On correlated sources)**.** In the case of *correlated* sources where $\mathbf{P} = \mathbf{S}\mathbf{S}^{\mathsf{H}}/T \in \mathbb{C}^{K\times K}$ is no longer a diagonal matrix. Let $\mathbf{P} = \mathbf{U_P}\mathbf{L}\mathbf{U_P^{\mathsf{H}}}$ be its eigen-decomposition, we have $\boldsymbol{\Phi}_1 \simeq \mathbf{U_P^{\mathsf{H}}}\mathbf{A}^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_1\mathbf{A}\mathbf{U_P} \simeq \frac{n}{N}\mathbf{I}_K$ and $\boldsymbol{\Phi}_2 \simeq \frac{n}{N}\mathbf{U_P^{\mathsf{H}}}\operatorname{diag}\{e^{\imath\Delta\theta_i}\}_{i=1}^K\mathbf{U_P}$ by Remark 1, so that

$$\bar{\boldsymbol{\Phi}} \simeq \operatorname{diag}(\sqrt{\mathbf{g}})\mathbf{U_P^{\mathsf{H}}}\operatorname{diag}\{e^{\imath\Delta\theta_i}\}_{i=1}^k\mathbf{U_P}\operatorname{diag}(\sqrt{\mathbf{g}}). \tag{20}$$

As such, $\bar{\boldsymbol{\Phi}}$ has, in general, its eigenvalues different from those of $\boldsymbol{\Phi}$. This leads to inconsistent ESPRIT estimates, and can be checked, in the case of $K = 2$ sources with DoAs $\theta_1 \neq \theta_2 \in (-\pi/2, \pi/2)$, by showing that $\lambda = ae^{\Delta\theta_1}$ for any $a \in \mathbb{R}$ *cannot* be an eigenvalue of $\bar{\boldsymbol{\Phi}}$ *unless* $\mathbf{U_P} = \mathbf{I}_2$ (that is, when the two sources are *uncorrelated*). Precisely, it can be checked that $\lambda = ae^{\Delta\theta_1}$ is an eigenvalue of $\bar{\boldsymbol{\Phi}}$ if and only if $(g_1 - g_2)^2|[\mathbf{U_P}]_{1,1}|^2|[\mathbf{U_P}]_{1,2}|^2 = 0$. This, in the case of correlated sources (with $\mathbf{U_P} \neq \mathbf{I}_2$), contradicts with Assumption 2. The same conclusion can be similarly drawn for θ_2.

B. Proof of Theorem 2

Here, we provide a proof sketch of Theorem 2, and refer to an extended version of this paper for the detailed proof.

The major technical challenge in characterizing the large-dimensional behavior of ESPRIT is that, the corresponding DoA estimates, as (the angles of) the complex eigenvalues of the K-by-K random matrix $\hat{\boldsymbol{\Phi}}$ defined in (15), depend on the entries of two *strongly dependent* random matrices $\hat{\boldsymbol{\Phi}}_1$ and $\hat{\boldsymbol{\Phi}}_2$ in a highly non-trivial fashion. Additionally, the (i,j) (complex) entry of $\hat{\boldsymbol{\Phi}}_2$ in (15), for $i \neq j$, writes

$$[\hat{\boldsymbol{\Phi}}_2]_{ij} = \hat{\mathbf{u}}_i^{\mathsf{H}}\mathbf{J}_1^{\mathsf{H}}\mathbf{J}_2\hat{\mathbf{u}}_j = \sum_{m=\ell}^{n+\ell-1}\mathbf{e}_{m+\Delta}^{\mathsf{T}}\hat{\mathbf{u}}_i\hat{\mathbf{u}}_j^{\mathsf{H}}\mathbf{e}_m, \tag{21}$$

and cannot be handled using standard RMT techniques. Indeed, standard RMT and contour integration techniques *only* provides access to the (limit of) absolute value of such complex random variable (but not its angle, see, e.g., [3, Section 2.5]), making the complex eigenvalues of $\hat{\boldsymbol{\Phi}}$ inaccessible.

To address this challenge, we propose to assess the eigenvalues of a (random or deterministic) matrix by working on *all* (combinations of) its entries with their indices forming a cycle. This is described in the following result.

Theorem 3 (Eigenvalue approximation between two matrices)**.** *For two matrices $\mathbf{A}, \mathbf{B} \in \mathbb{C}^{K\times K}$, if for* any *$m$-node cycle of the index set $\{i_1, \ldots, i_m\} \subseteq \{1, \ldots, K\}$, the entries of $\mathbf{A}, \mathbf{B}$ satisfy, for $\varepsilon \in [0, 1)$ that*

$$|A_{i_1i_2}A_{i_2i_3}\ldots A_{i_mi_1} - B_{i_1i_2}B_{i_2i_3}\ldots B_{i_mi_1}| \leq \varepsilon, \tag{22}$$

then, the eigenvalues of $\mathbf{A}, \mathbf{B}$ satisfy

$$|\lambda_k(\mathbf{A}) - \lambda_k(\mathbf{B})| \leq \sqrt[K+1]{\varepsilon}, \quad k \in \{1, \ldots, K\}. \tag{23}$$

Proof of Theorem 3. It is known, e.g., from [14] that the characteristic polynomial of $\mathbf{A} \in \mathbb{C}^{K\times K}$ is given by

$$\det(\lambda\mathbf{I}_K - \mathbf{A}) = \lambda^K - S_1(\mathbf{A})\lambda^{K-1}\ldots + (-1)^K S_K(\mathbf{A}), \tag{24}$$

for $S_m(\mathbf{A})$, the sum of all m-by-m principal minors of $\mathbf{A}$, with $S_1(\mathbf{A}) = \operatorname{tr}(\mathbf{A})$ and $S_K(\mathbf{A}) = \det(\mathbf{A})$.

Consider now the m-by-m principal minors of $\mathbf{A}$ and $\mathbf{B}$, denoted $\mathbf{A}[\mathcal{J}_m]$ and $\mathbf{B}[\mathcal{J}_m]$, with ordered indices $\mathcal{J}_m : 1 \leq i_1 < \ldots < i_m \leq K$, we have, by definition of principal minor that $\mathbf{A}[\mathcal{J}_m] - \mathbf{B}[\mathcal{J}_m] = \sum^{m!}(-1)^{\tau(j_1,\ldots,j_m)}(A_{i_1j_1}\ldots A_{i_mj_m} - B_{i_1j_1}\ldots B_{i_mj_m})$, with $\tau(j_1,\ldots,j_m) = \sum_{a=1}^{m-1}\sum_{b=a+1}^{m} H(j_a - j_b)$ with $H(x)$ the Heaviside step function that is 1 if $x > 0$ and 0 otherwise.

It then follows from the uniqueness of decomposition of permutation into (pairwise) disjoint cycles (see [15]) and the inequality in (22) that $|A_{i_1j_1}\ldots A_{i_mj_m} - B_{i_1j_1}\ldots B_{i_mj_m}| \leq \sum_{q=1}^{p} c_q\varepsilon^q$, for some $1 \leq p \leq m$ and constant $c_q > 0$ that depends m. As such, one has $|S_m(\mathbf{A}) - S_m(\mathbf{B})| \leq \sum_{|\mathcal{J}_m|=m}^{\binom{K}{m}}|\mathbf{A}[\mathcal{J}_m] - \mathbf{B}[\mathcal{J}_m]| \leq C_m\varepsilon$ for some constant C_m. This, together with (24) and the continuity of polynomials (see [16]), concludes the proof of Theorem 3. ☐

With Theorem 3 at hand, it then suffices to apply standard RMT techniques to derive asymptotic approximations of products of the entries of $\hat{\boldsymbol{\Phi}}_1$ and $\hat{\boldsymbol{\Phi}}_2$ (that forms a cycle as in Theorem 3), and thus the conclusion of Theorem 2. We refer the readers to an extended version of this paper for the detailed calculation for Theorem 2.

IV. Consistent DoA Estimation with G-ESPRIT

We have seen in Theorem 2 that classical ESPRIT is, in general, *incapable* of providing consistent DoA estimates for large arrays. In this section, we introduce an improved approach: the generalized ESPRIT (G-ESPRIT) method, that fixes this large-dimensional inconsistency of classical ESPRIT.

The G-ESPRIT method is almost identity to classical ESPRIT, but with large-dimensional extra terms of the latter consistently estimated and removed. Precisely, it follows from Theorem 2 that the top subspace $\hat{\mathbf{U}}_K$ of the SCM $\hat{\mathbf{C}}$ leads to additional large-dimensional bias terms in $\bar{\boldsymbol{\Phi}}$ of the form g_k, h_k defined in (18). These quantities, for known and large N, n, T, can be empirically and consistently estimated from the SCM $\hat{\mathbf{C}}$ per the following result.

Lemma 1 (Consistent estimates of g_k and h_k)**.** *Under Assumptions 1 and 2, let $\hat{\lambda}_k$ be the kth largest eigenvalue of SCM $\hat{\mathbf{C}}$ as in Theorem 1 and g_k, h_k be defined in* (18)*, one has, for $k \in \{1, \ldots, K\}$ that $\hat{g}_k - g_k \to 0$ and $\hat{h}_k - h_k \to 0$ almost surely as $N, T \to \infty$ with*

$$\hat{g}_k = \frac{1 - \frac{N}{T}\hat{\ell}_k^{-2}}{1 + \frac{N}{T}\hat{\ell}_k^{-1}}, \quad \hat{h}_k = \frac{n}{T}\frac{1 + \hat{\ell}_k^{-1}}{\frac{N}{T} + \hat{\ell}_k}, \tag{25}$$

for $\hat{\ell}_k = \frac{1}{2}\left(\hat{\lambda}_k - 1 - \frac{N}{T} + \sqrt{(\hat{\lambda}_k - 1 - \frac{N}{T})^2 - \frac{4N}{T}}\right)$.

Lemma 1 provides consistent estimates of the bias terms in classical ESPRIT and leads to the following result, as well as the proposed G-ESPRIT DoA estimation procedure summarized in Algorithm 2.

Algorithm 2 The proposed G-ESPRIT DoA estimation

Input: Received signal $\mathbf{X} \in \mathbb{C}^{N \times T}$, number of sources K.
Output: Estimated DoA angles $\hat{\theta}_k, k \in \{1, \ldots, K\}$.

1: Compute the SCM $\hat{\mathbf{C}} = \mathbf{X}\mathbf{X}^{\mathsf{H}}/T$ as in (5) to retrieve $\hat{\mathbf{U}}_K = [\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_K] \in \mathbb{C}^{N \times K}$ the estimated signal subspace composed of the top-K eigenvectors $\hat{\mathbf{u}}_1, \ldots, \hat{\mathbf{u}}_K \in \mathbb{C}^N$ associated to the largest K eigenvalues of $\hat{\mathbf{C}}$;
2: Define two selection matrices $\mathbf{J}_1, \mathbf{J}_2 \in \mathbb{R}^{n \times N}$ as in (6) that both select n among N rows with a "distance" $\Delta \geq 1$;
3: Compute $\hat{\boldsymbol{\Phi}}_1, \hat{\boldsymbol{\Phi}}_2$ using $\hat{\mathbf{U}}_K$ and $\mathbf{J}_1, \mathbf{J}_2$ as in (15);
4: Compute $\tilde{\boldsymbol{\Phi}}$ as in Corollary 1 and then the *angles* of $\lambda_k(\tilde{\boldsymbol{\Phi}})$, the kth (complex) eigenvalue of $\tilde{\boldsymbol{\Phi}}$;
5: **return** $\hat{\theta}_k = \arg(\lambda_k(\tilde{\boldsymbol{\Phi}}))/\Delta$.

Corollary 1 (Consistent DoA estimation with G-ESPRIT)**.** *Define* $\lambda_k(\tilde{\boldsymbol{\Phi}})$ *the kth largest eigenvalue of* $\tilde{\boldsymbol{\Phi}} = \tilde{\boldsymbol{\Phi}}_1^{-1}\tilde{\boldsymbol{\Phi}}_2$ *with*

$$\tilde{\boldsymbol{\Phi}}_1 = \operatorname{diag}(\hat{\mathbf{g}}^{-1/2}) \left(\hat{\boldsymbol{\Phi}}_1 - \operatorname{diag}\{\hat{h}_k\}_{k=1}^K \right) \operatorname{diag}(\hat{\mathbf{g}}^{-1/2}),$$
$$\tilde{\boldsymbol{\Phi}}_2 = \operatorname{diag}(\hat{\mathbf{g}}^{-1/2}) \hat{\boldsymbol{\Phi}}_2 \operatorname{diag}(\hat{\mathbf{g}}^{-1/2}), \tag{26}$$

for $\hat{\mathbf{g}}^{-1/2} = [1/\sqrt{\hat{g}_1}, \ldots, 1/\sqrt{\hat{g}_K}]^{\mathsf{T}}$ *and* $\hat{g}_k, \hat{h}_k$ *as defined in* (25)*. Then, under the same settings and notations as in Theorem 2, we have, for* $k \in \{1, \ldots, K\}$ *that* $\arg(\lambda_k(\tilde{\boldsymbol{\Phi}}))/\Delta - \theta_k \to 0$ *almost surely as* $N, T \to \infty$*.*

V. Numerical Simulations

We provide numerical simulations in Figure 1 to support the asymptotic results derived in previous sections.

In the left plot of Figure 1, we consider the case of $K = 2$ correlated sources at DoA $\theta_1 = 0$ and $\theta_2 = \pi/4$, with $N = 400$, $T = 1\,000$, and $\mathbf{P} = \left(\begin{smallmatrix} 1.9953 & 0.7981 \\ 0.7981 & 1.9953 \end{smallmatrix}\right)$, $\ell = 1$, $n = N - 1$, $\Delta = 1$, We compare the DoAs estimates $\hat{\theta}$ from classical ESPRIT (as well as its theoretical characterization $\bar{\theta}$ given in Theorem 2), to $\tilde{\theta}$ those from the proposed G-ESPRIT method in Algorithm 2 and Corollary 1. We observe that (i) the theoretical analysis ($\bar{\theta}$) perfectly match the behavior of ESPRIT ($\hat{\theta}$, that is biased from the true DoAs θ); and (ii) the proposed G-ESPRIT method $\tilde{\theta}$ successfully removes this bias.

The right plot of Figure 1 illustrates the decreases in mean squared errors (MSEs) and variances of DoA estimates from both ESPRIT ($\hat{\theta}$) and G-ESPRIT ($\tilde{\theta}$), as the array length N increases, with a fixed ratio $N/T = 0.4$. One observes that (i) the classical ESPRIT provides *inconsistent* DoA estimates, with MSE much *larger than* the variance; and (ii) G-ESPRIT provides consistent estimates and, in addition, yields *smaller* variances than classical ESPRIT.

VI. Conclusion

In this paper, we exploit RMT to examine the performance of ESPRIT in the case of large arrays, revealing its tendency for *inconsistent* DoA estimates. Introducing the improved G-ESPRIT, we prove its asymptotic consistency in the same scenario. Numerical simulations confirm G-ESPRIT's consistent performance and reduced variance compared to ESPRIT. It would be of future interest to extend the RMT analysis here to characterize the second-order behavior of both ESPRIT and G-ESPRIT, as to assess quantitatively their performance gaps from the Cramér–Rao bound.

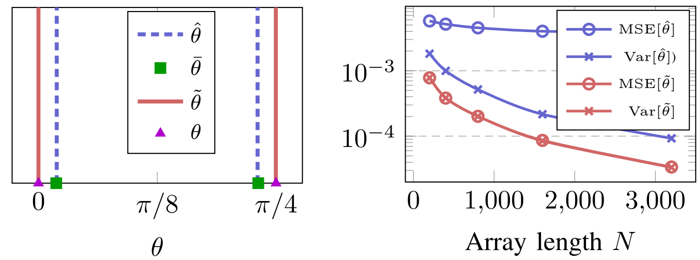

Fig. 1. **Left**: comparison between $\hat{\theta}$ from ESPRIT, $\bar{\theta}$ from $\bar{\boldsymbol{\Phi}}$ in Theorem 2, $\tilde{\theta}$ from $\tilde{\boldsymbol{\Phi}}$ of G-ESPRIT in Corollary 1, and true DoAs θ. **Right**: DoA estimation MSEs and variances of ESPRIT and the improved G-ESPRIT methods as the array length N increases. Results estimated over 100 independent runs.

References

[1] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
[2] A. Paulraj, R. Roy, and T. Kailath, "A subspace rotation approach to signal parameter estimation," *Proceedings of the IEEE*, vol. 74, no. 7, pp. 1044–1046, 1986.
[3] Romain Couillet and Zhenyu Liao, *Random Matrix Methods for Machine Learning*, Cambridge University Press, 2022.
[4] Debashis Paul and Alexander Aue, "Random matrix theory in statistics: A review," *Journal of Statistical Planning and Inference*, vol. 150, pp. 1–29, July 2014.
[5] Pascal Vallet, Xavier Mestre, and Philippe Loubaton, "Performance analysis of an improved MUSIC DoA estimator," *IEEE Transactions on Signal Processing*, vol. 63, no. 23, pp. 6407–6422, 2015.
[6] Xavier Mestre and Miguel Ángel Lagunas, "Modified subspace algorithms for DoA estimation with large arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 598–614, 2008.
[7] R. Roy, A. Paulraj, and T. Kailath, "Comparative performance of ESPRIT and MUSIC for direction-of-arrival estimation," in *ICASSP '87. IEEE International Conference on Acoustics, Speech, and Signal Processing*, Apr. 1987, vol. 12, pp. 2344–2347.
[8] T.B. Lavate, V.K. Kokate, and A.M. Sapkal, "Performance Analysis of MUSIC and ESPRIT DOA Estimation Algorithms for Adaptive Array Smart Antenna in Mobile Communication," in *2010 Second International Conference on Computer and Network Technology*, Apr. 2010, pp. 308–311.
[9] Pascal Vallet and Philippe Loubaton, "On the Performance of MUSIC With Toeplitz Rectification in the Context of Large Arrays," *IEEE Transactions on Signal Processing*, vol. 65, no. 22, pp. 5848–5859, Nov. 2017.
[10] Vladimir A Marcenko and Leonid Andreevich Pastur, "Distribution of eigenvalues for some sets of random matrices," *Mathematics of the USSR-Sbornik*, vol. 1, no. 4, pp. 457, 1967.
[11] Jack W. Silverstein and Zhidong Bai, "On the Empirical Distribution of Eigenvalues of a Class of Large Dimensional Random Matrices," *Journal of Multivariate Analysis*, vol. 54, no. 2, pp. 175–192, 1995.
[12] Jinho Baik and Jack W. Silverstein, "Eigenvalues of large sample covariance matrices of spiked population models," *Journal of Multivariate Analysis*, vol. 97, no. 6, pp. 1382–1408, 2006.
[13] Florent Benaych-Georges and Raj Rao Nadakuditi, "The eigenvalues and eigenvectors of finite, low rank perturbations of large random matrices," *Advances in Mathematics*, vol. 227, no. 1, pp. 494–521, 2011.
[14] Roger A Horn and Charles R Johnson, *Matrix analysis*, pp. 38–42, Cambridge university press, 2012.
[15] William Raymond Scott, *Group theory*, pp. 8–11, Courier Corporation, 2012.
[16] Melvyn B. Nathanson and David A. Ross, "Continuity of the roots of a polynomial," 2023.