Multidimensional Frequency Estimation with Finite Snapshots in the Presence of Identical Frequencies

Jun Liu, Student Member, IEEE, Xiangqian Liu, Member, IEEE, and Xiaoli Ma, Member, IEEE

Abstract—Recently an eigenvector-based algorithm has been developed for multidimensional frequency estimation with a single snapshot of data mixture. Unlike most existing algebraic approaches that estimate frequencies from eigenvalues, the eigenvector-based algorithm achieves automatic frequency pairing without joint diagonalization of multiple matrices, but it fails when there exist identical frequencies in certain dimensions because eigenvectors are not linearly independent anymore. In this paper, we develop an eigenvector-based algorithm for multidimensional frequency estimation with finite data snapshots. We also analyze the identifiability and performance of the proposed algorithm. It is shown that our algorithm offers the most relaxed statistical identifiability condition for multidimensional frequency estimation from finite snapshots. More important, we introduce complex weighting factors so that the algorithm is still operational when there exist identical frequencies in one or more dimensions. Furthermore, the weighting factors are optimized to minimize the mean-square errors of the frequency estimates. Simulation results show that the proposed algorithm offers competitive performance when compared to existing algebraic approaches but at lower complexity.

Index Terms—Frequency estimation, eigenvalue decomposition, multidimensional signal processing, identifiability, perturbation analysis

I. INTRODUCTION

In radar, signal processing, and communications, many parameter estimation problems can be reduced into two-dimensional (2-D) or in general N-dimensional (N-D) frequency estimation, such as 2-D direction estimation [26], joint frequency and 2-D angle estimation [25], joint angle and delay estimation [23], joint angle and frequency estimation [10], and multidimensional parameter estimation in wireless channel sounding [14], [21]. In these problems, it is desirable to estimate multiple N-D frequency components, each with N frequencies, from several snapshots of observation data. High resolution subspace-based parametric methods based on eigenvalue decomposition (EVD) have been well-studied for N-D frequency estimation, such as N-D MUSIC and N-D Unitary ESPRIT (see e.g., [5], [14], [17]).

One of the important problems for parametric frequency estimation methods is to determine the maximum number of resolvable frequency components for a given sample size in the noiseless case, which is referred to as the identifiability bound herein. The identifiability bound is of importance in situations where data samples come at a premium, for example, in spatial sampling for direction estimation using an antenna array, and in time sampling for time-varying parameter estimation in wireless channel sounding. Recently much progress has been made on the statistical identifiability of multidimensional frequency estimation from a single snapshot of data [9], [13], [15]. For example, in the 2-D case, the MDF algorithm [15] can uniquely estimate up to 0.25M1M2 2-D frequencies almost surely from a single snapshot of M1 × M2 data mixture, provided that the 2-D frequencies are drawn from a continuous distribution (hence the name of statistical identifiability). We have shown in [13] that the improved MDF (IMDF) algorithm can retrieve up to approximately 0.34M1M2 2-D frequencies for the same data size.

The aforementioned identifiability results concern the single snapshot case. Few results were found in the multiple snapshot case. When multiple snapshots are available, most existing algebraic approaches estimate frequencies from the sample covariance matrix. A sufficient large number of snapshots usually need to be collected in order to calculate the sample covariance matrix. For example, it is assumed that the number of snapshots is greater than the data size of one snapshot for the MUSIC algorithm in 1-D frequency estimation [22]. However, in many cases it is possible that only a few snapshots may be collected. For example, in multiple target tracking, only a few snapshots of data may be available for a fixed 2-D DOA if the targets move fast. In these circumstances, the study of the identification condition for finite snapshots becomes important. For instance, in [11], an identifiability bound is obtained for the estimation of multiple constant modulus signals from finite snapshots.

Another challenge of N-D frequency estimation is the frequency association or frequency pairing problem, which concerns how to associate the N frequencies of the same N-D frequency component. If N-D MUSIC is implemented by N-D grid search, no frequency association step is needed, but N-D grid search is very complex. Alternatively, 1-D frequency estimation can be performed in individual dimensions of the N-D MUSIC spectrum followed by a frequency pairing step [8], [17]. On the other hand, the Unitary ESPRIT algorithm uses an iterative Jacobi-type algorithm for approximate joint diagonalization to achieve automatic frequency pairing [5]. Recently, as an alternative to existing algebraic algorithms that...
estimate frequencies from eigenvalues, an eigenvector-based algorithm (the IMDF algorithm) is developed in [13] for N-D frequency estimation with a single snapshot of data mixture. The IMDF algorithm achieves automatic frequency pairing without using joint diagonalization since the frequencies of an N-D frequency component are estimated from the same column of a structural matrix, and thus the computational complexity is reduced. The transformation matrix between the structural matrix and the signal subspace is obtained from the eigenvectors of a matrix pencil [13]. However, the IMDF algorithm fails when there exist identical frequencies in the first dimension, because in this case, the corresponding eigenvectors are not linearly independent any more. Note that eigenvalue-based approaches do not have this drawback.

In this paper, we develop an eigenvector-based algorithm for N-D frequency estimation with multiple snapshots, in the presence of possible identical frequencies in one or more dimensions. Our approach is based on the IMDF algorithm, but it is not a trivial extension of the IMDF algorithm to the multiple snapshot case. Similar to IMDF, our proposed algorithm estimates frequencies from a structural matrix that is obtained from the eigenvectors. More important, our novel contribution in this algorithm is two-fold: most relaxed identifiability bound for N-D frequency estimation with finite snapshots, and optimal weighting factors to deal with identical frequencies in one or more dimensions.

Intuitively, the maximum number of resolvable frequency components should increase as the number of snapshots increases until it reaches a threshold. We prove that when the number of snapshots is smaller than a threshold (approximately $M/2$), the identifiability bound of our N-D frequency estimation algorithm is approximately $2T M / (1 + \sqrt{2T})^N$, where $T$ is the number of snapshots, and $M$ is the sample size of one snapshot. When the number of snapshots is greater than the threshold, both our approach and those based on sample covariance matrix give the same identifiability bound, which is approximately $M$.

To deal with possible identical frequencies in one or more dimensions, we introduce complex weighting factors in our eigenvector-based approach to ensure that the eigenvalues of the matrix pencil obtained from the signal subspace are distinct, so that eigenvectors are linearly independent. Similar factors have been adopted in the (eigenvalue-based) 2-D ESPRIT-type algorithm [19]. However, here we derive the theoretic mean-square errors (MSEs) of the frequency estimates, and we optimize the weighting factors by minimizing the MSEs, while the weighting factors in [19] are chosen randomly. Numerical simulation results demonstrate that the proposed approach is effective in improving the performance of eigenvector-based N-D frequency estimation in the presence of identical frequencies, and optimal weighting factors significantly outperform random weighting factors since randomly chosen factors can not guarantee that the MSEs are minimized in the noisy case.

The rest of the paper is organized as follows. The data model and several preliminary lemmas for identifiability argument are introduced in Section II. In Section III we present the eigenvector-based approach for N-D frequency estimation from finite data snapshots. Section IV gives the statistical identifiability bound of the proposed algorithm. The performance of proposed algorithm in noisy case is analyzed in V, where the theoretic MSEs are derived, and based on which, an approach to obtain optimal weighting factors are proposed in Section VI. Simulation results are discussed in Section VII and conclusions are drawn in Section VIII.

In the following, upper (lower) bold face letters are used for matrices (column vectors). $A^\dagger$, $A^T$, $A^H$, and $A^\odot$ denote the conjugate, transpose, Hermitian transpose, and pseudo-inverse of $A$, respectively. We will use $\otimes$ for the Kronecker product, $\circ$ for the Khatri-Rao (column-wise Kronecker) product,

$$[a_1 \ a_2 \ \ldots \ a_N] \odot [b_1 \ b_2 \ \cdots \ b_N] := [a_1 \otimes b_1 \ a_2 \otimes b_2 \ \cdots \ a_N \otimes b_N].$$

We also use $I_p$ for a $p \times p$ identity matrix, $0_{M \times N}$ for an $M \times N$ zero matrix, $D(a)$ for a diagonal matrix with $a$ as its diagonal, $A^{(m)}$ for a sub-matrix of $A$ formed by its first $m$ rows, $\| \|$ for the $l_2$ norm, $a_{i,j}$ or $[A]_{i,j}$ for the $(i,j)$-th element of $A$, $E(\cdot)$ for expectation, $C$ for the complex field, $\otimes$ for the set $(-\pi, \pi]$, and $R(\cdot)$ and $I(\cdot)$ for the real and imaginary parts respectively.

II. DATA MODEL

The data model for N-D frequency estimation can take a variety of forms. In this paper, we refer to a single snapshot N-D frequency mixture as an N-D array $\mathbf{X}$ with

$$x_{m_1,m_2,\ldots,m_N} = \sum_{f=1}^{F} \epsilon_f \prod_{n=1}^{N} e^{i \omega_{f,n} (m_n - 1)} + w_{m_1,m_2,\ldots,m_N},$$

(1)

where $m_n = 1, \ldots, M_n$, for $n = 1, \ldots, N$, and $M_n$ is the sample size of the $n$-th dimension. The total sample size is $M := \prod_{n=1}^{N} M_n$. In (1), the frequencies $\omega_{f,n} \in \mathbb{C}$, for $f = 1, \ldots, F$, $n = 1, \ldots, N$, and $w_{m_1,m_2,\ldots,m_N}$ is white Gaussian noise with variance $\sigma^2$. Similarly $T$ snapshots of N-D frequency data mixtures may be modeled as $T$ N-D arrays, $\mathbf{X}(t)$, with

$$x_{m_1,m_2,\ldots,m_N}(t) = \sum_{f=1}^{F} c_f(t) \prod_{n=1}^{N} e^{i \omega_{f,n} (m_n - 1)} + w_{m_1,m_2,\ldots,m_N}(t),$$

(2)

for $t = 1, \ldots, T$, where $t$ is the snapshot index, which can be a time index, or trial index in case of multiple trials of experiments. $T = 1$ corresponds to the single snapshot case. The frequency set $\{\omega_{f,n}\}_{n=1}^{N}$ is an N-D frequency component, and there are $F$ such components. The objective of N-D frequency estimation is to estimate $\{\omega_{f,n}\}_{n=1}^{N}$, for $f = 1, \ldots, F$, from $\mathbf{X}(t)$, $t = 1, \ldots, T$. Notice that the same data model for the case of multiple snapshots has been used in [5], [17].

In order to facilitate the presentation, we introduce the equivalent data models using the notation of Khatri-Rao product. Given (2), define the sample vector $\mathbf{x}(t)$ for the $t$-th
snapshot as

\[
x(t) = \begin{bmatrix}
x_{1,1,\ldots,1}(t) \\
x_{1,1,\ldots,2}(t) \\
\vdots \\
x_{1,1,\ldots,M_N}(t) \\
x_{1,1,\ldots,21}(t) \\
\vdots \\
x_{M_1,M_2,\ldots,M_N}(t)
\end{bmatrix}.
\]

Furthermore, define \( N \) Vandermonde matrices \( \mathbf{A}_n \in \mathbb{C}^{M_n \times F} \) with generators \( \{e^{j\omega_{f,n}}\}_{f=1}^F \) such that

\[
\mathbf{A}_n := [a_{1,n} \ a_{2,n} \ \cdots \ a_{F,n}],
\]

\[
a_{f,n} := \begin{bmatrix}
e^{j\omega_{f,1}} & \cdots & e^{j(M_n-1)\omega_{f,n}}
\end{bmatrix}^T, \quad n = 1, \ldots, N.
\]

It can be verified that

\[
x(t) = \mathbf{A}\mathbf{c}(t) + \mathbf{w}(t), \quad t = 1, \ldots, T, \tag{3}
\]

where \( \mathbf{w}(t) \) is the noise vector, and

\[
\mathbf{A} := \mathbf{A}_1 \odot \mathbf{A}_2 \odot \cdots \odot \mathbf{A}_N,
\]

\[
\mathbf{c}(t) := [c_1(t) \ c_2(t) \ \cdots \ c_F(t)]^T.
\]

Define

\[
\mathbf{X} := [x(1) \ x(2) \ \cdots \ x(T)] \in \mathbb{C}^{M \times T},
\]

\[
\mathbf{C} := [c(1) \ c(2) \ \cdots \ c(T)] \in \mathbb{C}^{F \times T},
\]

then the data model in (3) can be rewritten in matrix form as

\[
\mathbf{X} = \mathbf{A}\mathbf{C} + \mathbf{W}, \tag{4}
\]

where \( \mathbf{W} \) is the corresponding noise matrix.

We need the following lemmas later. Lemma 1 is used to prove Lemma 4, which is a key result used in our algorithm. Lemma 3 connects \( N \)-D data smoothing to the Khatri-Rao product of \( N \) Vandermonde matrices. Lemmas 1–3 are available in the literature, but reproduced here for convenience.

**Lemma 1:** Consider an analytic function \( h(\mathbf{x}) \) of vector with complex variable elements \( \mathbf{x} = [x_1, \ldots, x_N]^T \in \mathbb{C}^{N \times 1} \). If \( h \) is nontrivial in the sense that there exists \( x_0 \in \mathbb{C}^{N \times 1} \) such that \( h(x_0) \neq 0 \), then the zero set of \( h(x) \):

\[
\mathcal{Z} := \{ \mathbf{x} \in \mathbb{C}^{N \times 1} | h(\mathbf{x}) = 0 \}
\]

is of measure (Lebesgue measure in \( \mathbb{C}^{N \times 1} \)) zero.

**Proof:** See Lemma 2 of [9].

**Lemma 2:** Given \( N \) Vandermonde matrices \( \mathbf{A}_n \in \mathbb{C}^{M_n \times F} \) for \( n = 1, \ldots, N \geq 2 \), the rank of

\[
\mathbf{A} = \mathbf{A}_1 \odot \mathbf{A}_2 \odot \cdots \odot \mathbf{A}_N
\]

is \( \min(\prod_{n=1}^N M_n, F) \) almost surely, if the \( NF \) generators \( \{e^{j\omega_{f,n}}\}_{f=1}^F \) of \( \mathbf{A}_n, \ n = 1, \ldots, N \), are drawn from a continuous distribution with respect to the Lebesgue measure in \( \mathbb{C}^{NF \times 1} \).

**Proof:** See Theorem 2 of [9].

**Lemma 3:** Define a set of \( N \)-D selection matrices as

\[
\mathbf{J}_{\ell_1,\ell_2,\ldots,\ell_N} := \mathbf{J}_{\ell_1}^{K_1} \odot \mathbf{J}_{\ell_2}^{K_2} \odot \cdots \odot \mathbf{J}_{\ell_N}^{K_N},
\]

\[
\mathbf{J}_{\ell_n}^{K_n} := \begin{bmatrix}
0_{K_n \times (\ell_n-1)} \ I_{K_n} \ 0_{K_n \times (L_n-\ell_n)}
\end{bmatrix}, \tag{5}
\]

where \( \ell_n = 1, \ldots, L_n \), and \( K_n \) and \( L_n \) are positive integers satisfying \( K_n + L_n = M_n + 1 \) for \( n = 1, \ldots, N \). Further define an \( N \)-D smoothing operator for the snapshot vector in (3) as

\[
\mathbf{S}[\mathbf{x}(t)] := \begin{bmatrix}
\mathbf{J}_{1,1,\ldots,1}\mathbf{x}(t) \mathbf{J}_{1,1,\ldots,2}\mathbf{x}(t) \cdots \mathbf{J}_{1,1,\ldots,L_N}\mathbf{x}(t) \\
\mathbf{J}_{1,1,\ldots,2,1}\mathbf{x}(t) \cdots \mathbf{J}_{L_1,L_2,\ldots,L_N}\mathbf{x}(t)
\end{bmatrix}, \tag{6}
\]

then it can be verified that in the absence of noise

\[
\mathbf{X}\mathbf{S}(t) := \mathbf{S}[\mathbf{x}(t)]
\]

\[
= \left( \mathbf{A}_1^{(K_1)} \odot \mathbf{A}_2^{(K_2)} \odot \cdots \odot \mathbf{A}_N^{(K_N)} \right)
\]

\[
\mathbf{D}(\mathbf{c}(t)) \left( \mathbf{A}_1^{(L_1)} \odot \mathbf{A}_2^{(L_2)} \odot \cdots \odot \mathbf{A}_N^{(L_N)} \right)^T \tag{8}
\]

**Proof:** See Lemma 2 of [13].

Here we give an example of the smoothing operator in the 2-D case. Suppose that \( \mathbf{x} \) is a snapshot of 2-D observed data in (3), then the smoothed data matrix \( \mathbf{S}(\mathbf{x}) \) in (7) is equivalent to

\[
\mathbf{S}(\mathbf{x}) = \begin{bmatrix}
\mathbf{X}_1 & \mathbf{X}_2 & \cdots & \mathbf{X}_{L_1} \\
\mathbf{X}_2 & \mathbf{X}_3 & \cdots & \mathbf{X}_{L_1+1} \\
\vdots & \vdots & \ddots & \vdots \\
\mathbf{X}_{K_1} & \mathbf{X}_{K_1+1} & \cdots & \mathbf{X}_{M_1}
\end{bmatrix},
\]

where

\[
\mathbf{X}_k = \begin{bmatrix}
x_{k,1} & x_{k,2} & \cdots & x_{k,L_2} \\
x_{k,2} & x_{k,3} & \cdots & x_{k,L_2+1} \\
\vdots & \vdots & \ddots & \vdots \\
x_{k,K_2} & x_{k,K_2+1} & \cdots & x_{k,M_2}
\end{bmatrix}.
\]

**Lemma 3** claims that

\[
\mathbf{S}(\mathbf{x}) = \left( \mathbf{A}_1^{(K_1)} \odot \mathbf{A}_2^{(K_2)} \right) \mathbf{D}(\mathbf{c}) \left( \mathbf{A}_1^{(L_1)} \odot \mathbf{A}_2^{(L_2)} \right)^T.
\]

**Lemma 4:** Given \( N \) Vandermonde matrices \( \mathbf{A}_n \in \mathbb{C}^{M_n \times F} \), with generators \( \{e^{j\omega_{f,n}}\}_{f=1}^F \) for \( n = 1, \ldots, N \), and a complex matrix \( \mathbf{C} \in \mathbb{C}^{F \times T} \), define

\[
\mathbf{B} := \begin{bmatrix}
\mathbf{C}^T \\
\Pi_T \mathbf{C}^H \mathbf{D}(\mathbf{\beta})
\end{bmatrix}, \tag{9}
\]

where \( \Pi_T \) is a \( T \times T \) permutation matrix with ones on its anti-diagonal, and

\[
\mathbf{\beta} = \left[ e^{-j\beta_1} \ e^{-j\beta_2} \ \cdots \ e^{-j\beta_F} \right]^T,
\]

\[
\beta_f = \sum_{n=1}^N (M_n - 1)\omega_{f,n}. \tag{10}
\]

If the \( NF \) frequencies \( \{\omega_{f,n}\}_{f=1}^F \) are drawn from distributions that are continuous with respect to the Lebesgue measure in \( \mathbb{C}^{NF \times 1} \) and \( \mathbb{C}^{TF \times 1} \), respectively, then the rank of the matrix

\[
\tilde{\mathbf{H}} := \mathbf{A}_1^{(L_1)} \odot \mathbf{A}_2^{(L_2)} \cdots \odot \mathbf{A}_N^{(L_N)} \mathbf{B} \tag{11}
\]


\[
\text{is min}\left\{2T \prod_{n=1}^N L_n, F \right\} \text{ almost surely.}
\]

**Proof:** See Appendix A.
III. THE EIGENVECTOR-BASED ALGORITHM FOR N-D FREQUENCY ESTIMATION

In this section we present the algorithm for N-D frequency estimation from multiple snapshots. For simplicity of exposition, the algorithm is developed in the noiseless case. The identifiability of the algorithm is discussed in Section IV and the noise effect on the performance of the algorithm is analyzed in Section V.

Given (3) in the noiseless case, we can apply the smooth operator \( S \) defined in Lemma 3 to every snapshot \( x(t) \), and obtain

\[
X_S(t) := S[x(t)] = GD(c(t))H^T, \tag{12}
\]

where

\[
G := A_1^{(K_1)} \odot A_2^{(K_2)} \cdots \odot A_N^{(K_N)},
H := A_1^{(L_1)} \odot A_2^{(L_2)} \cdots \odot A_N^{(L_N)}.
\]

The positive integers \( K_n \) and \( L_n, \ n = 1, \cdots, N \), are chosen such that

\[
K_n + L_n = M_n + 1, \quad 1 \leq n \leq N. \tag{13}
\]

To further explore the data structure, we can perform the forward-backward smoothing on the data vector \( x(t) \) in (3). Define

\[
y(t) := \Pi_M x^*(t), \tag{14}
\]

where \( \Pi_M \) is an \( M \times M \) backward permutation matrix. It can be verified that

\[
y(t) = A\bar{c}(t),
\]

where

\[
\bar{c}(t) = [\bar{c}_1(t), \bar{c}_2(t), \cdots, \bar{c}_F(t)]^T \quad \text{with} \quad \bar{c}_f(t) = c_f(t)e^{-j\beta_f}, \quad \beta_f \text{ is defined in (10).}
\]

Applying the same technique to \( y(t) \) that we used to construct \( X_S(t) \) from \( x(t) \), we obtain

\[
Y_S(t) := S[y(t)] = GD(\bar{c}(t))H^T.
\]

Then we can collect all the smoothed data matrices to obtain

\[
\bar{X} := [X_S(1) X_S(2) \cdots X_S(T) Y_S(T)]
\]

\[
Y_S(T-1) \cdots Y_S(1)]. \tag{15}
\]

Define \( K := \prod_{n=1}^N K_n \) and \( L := \prod_{n=1}^N L_n \). The size of \( \bar{X} \) is \( K \times 2TL \). It can be verified that

\[
\bar{X} = G(H \odot B)^T = G\tilde{H}^T, \tag{16}
\]

where \( B \) and \( \tilde{H} \) are defined in (9) and (11), respectively. A key step of our algorithm is the construction of \( \bar{X} \) to ensure that it is of rank \( F \) almost surely. In (16), since \( G \) is the Khatri-Rao product of multiple Vandermonde matrices, invoking Lemma 2, \( G \) has full column rank almost surely if \( \prod_{n=1}^N K_n \geq F \). According to Lemma 4, if \( 2T \prod_{n=1}^N L_n \geq F \), \( H \) has full column rank almost surely. Based on the Sylvester’s inequality

\[
\text{rank}(G) + \text{rank}(H^T) - F \leq \text{rank}(G\tilde{H}^T),
\]

\[
\leq \min\{\text{rank}(G), \text{rank}(H^T)\},
\]

\( \bar{X} \) is of rank \( F \) almost surely. The singular value decomposition (SVD) of \( \bar{X} \) yields

\[
\bar{X} = U_s \Sigma_s V_s^H, \tag{17}
\]

where \( U_s \) has \( F \) columns that together span the column space of \( \bar{X} \). Since the same space is spanned by the columns of \( G \), there exists an \( F \times F \) nonsingular matrix \( T^{-1} \) such that

\[
U_s = G T^{-1}. \tag{18}
\]

Similar to the IMDF algorithm [13], once the signal subspace \( U_s \) is obtained, we can construct two matrices from \( U_s \) whose general eigenvalues are the exponentials of the first dimension, and then use the general eigenvectors to estimate N-D frequencies. However, as mentioned before, the IMDF algorithm fails when there exist identical frequencies in the first dimension since the eigenvectors are not linearly independent anymore. Furthermore, it has been shown in [13] that the performance of the IMDF algorithm degrades if there are close frequencies in the first dimension. To address this problem, in the following, we present a method to construct two matrices whose general eigenvalues are weighted sums of the N-D exponentials. The N-D frequencies are still resolved from the general eigenvectors.

We define two selection matrices \( J_1 \) and \( J_2 \) as

\[
J_1 := J_{1,1} \odot J_{1,2} \cdots \odot J_{1,N}, \quad J_2 := \sum_{n=1}^N \alpha_n \bar{J}_{2,n}, \tag{19}
\]

where

\[
J_{1,n} = [I_{K_n-1} 0_{(K_n-1)\times 1}], \quad J_{2,n} = [0_{(K_n-1)\times 1} I_{K_n-1}],
\bar{J}_{2,n} := (J_{1,1} \odot \cdots \odot J_{1,n-1}) \odot J_{2,n} \odot (J_{1,n+1} \odot \cdots \odot J_{1,N}).
\]

Here \( \{\alpha_n\}_{n=1}^N \) are complex weighting factors, which can be randomly chosen initially. As we will show in Section V, the MSEs of the frequency estimates are affected by these weighting factors in the noisy case. Section VI discusses how to optimize these weighting factors to minimize MSEs. Next, we obtain two equal-sized matrices \( U_1 \) and \( U_2 \) by

\[
U_1 := J_1 U_s, \quad U_2 := J_2 U_s. \tag{20}
\]

According to the property of Khatri-Rao product [1]: \((A \odot B)(C \odot D) = AC \odot BD\), it can be verified that

\[
U_1 = PT^{-1}, \quad U_2 = PD(\zeta)T^{-1}. \tag{21}
\]

where \( P = A_1^{(K_1-1)} \odot A_2^{(K_2-1)} \cdots \odot A_N^{(K_N-1)} \). It is clear that \( P \) has full column rank almost surely if \( F \leq \prod_{n=1}^N (K_n-1) \). In (21), we have

\[
\zeta := [\zeta_1, \zeta_2, \cdots, \zeta_F]^T, \quad \zeta_f = \sum_{n=1}^N \alpha_n e^{j\omega_{f,n}}.
\]

We can resolve \( T \) from the matrix pencil \( U_1 \) and \( U_2 \) via least-squares or total-least-squares approaches [20]. Here, we use the least-squares approach. \( T \) is retrieved from the following EVD up to column permutation and scaling ambiguity.

\[
U_1^T U_2 = TD(\zeta)T^{-1}. \tag{22}
\]
Clearly we can choose \( \{ \omega_n \}_{n=1}^N \) to ensure that the elements of \( \zeta \) are distinct even if there exist identical frequencies in one or more dimensions, but this is not guaranteed by randomly generated \( \{ \omega_n \}_{n=1}^N \). We will discuss how to choose the weighting factors in Section VI. Suppose that the EVD of \( U_j^* U_j \) gives \( T_{sp} = T \Delta \), where \( \Delta \) is a nonsingular diagonal column scaling matrix and \( \Delta \) is a permutation matrix. Once we obtain \( T_{sp} \), we can retrieve \( P \) up to column permutation and scaling ambiguity according to

\[
P_{sp} = U_j T_{sp} = P \Lambda \Delta. \tag{23}
\]

Notice that \( P \) is the Khatri-Rao product of \( N \) Vandermonde matrices, and there are \( F \) columns in \( P \). The \( N \) frequencies of the same \( N \)-D component appear in the same column of \( P \). In other words, for a fixed \( f \), \( \{ \omega_{f,n} \}_{n=1}^N \) appear in the same column of \( P \). Thanks to this structure, we can obtain \( F \)-\( N \)-D frequency components by dividing suitably chosen elements of the aforementioned columns of \( P_{sp} \). Therefore the column scaling and permutation will not have a material effect on the algorithm. For this reason, we drop subscript “\( sp \)” from now on as it is clear from the context. Suppose that \( \{ \omega_{f,n} \}_{n=1}^N \) appear in the \( f \)-th column of \( P \). Then, \( \exp^{j \omega_{f,n}} \) can be obtained by anyone of the following quotients

\[
\exp^{j \omega_{f,n}} = \frac{p_{k,f}}{p_{k-K'_n,f}}, \quad \text{mod} \ (k-1, K'_n-1) \geq K'_n, \tag{24}
\]

for \( f = 1, \ldots, F \), where \( 1 \leq k \leq K'_n \), \( p_{k,f} \) is the \((k,f)\)-th element of \( P \), and

\[
K'_n := \left\{ \prod_{p=n+1}^{N} (K_p - 1), \quad \begin{array}{l}
0 \leq n \leq N-1, \\
1, \quad n = N.
\end{array} \right.
\]

Notice that the frequencies are automatically paired because the frequencies \( \{ \omega_{f,n} \}_{n=1}^N \) of the same \( N \)-D component (the \( f \)-th component) are obtained from the same column of \( P \).

If the data observations are noisy as given in (3), applying the above algorithm we can obtain the estimate of \( P \) as \( \hat{P} \). In order to reduce the MSEs of frequency estimates, we use the average of all the quotients in (24) to obtain an estimate of the \( N \)-D exponential. In other words, \( \exp^{j \omega_{f,n}} \) is estimated by the following average

\[
\hat{\exp^{j \omega_{f,n}}} = \frac{1}{\mu_n} \sum_{k=1}^{K'_n} \frac{\hat{p}_{k,f}}{p_{k-K'_n,f}}, \quad n = 1, \ldots, N, \tag{25}
\]

where \( \mu_n = K'_n (K_n - 2) / (K_n - 1) \). The average is also the so-called “circular mean” in direction statistics [16]. Finally the frequency estimates are obtained by

\[
\hat{\omega}_{f,n} = \mathcal{I} \left( \ln \hat{\exp^{j \omega_{f,n}}} \right). \tag{26}
\]

After the frequency estimates are obtained, the amplitude matrix \( C \) can be obtained by solving (4) using a least-squares approach.

IV. STATISTICAL IDENTIFIABILITY

The identifiability bound provides a sufficient identification condition for the \( N \)-D frequency estimation problem. The higher the identifiability bound an algorithm can achieve, the more relaxed the sufficient condition is. We will show that our proposed algorithm offers a higher identifiability bound compared with existing ones. We summarize our main result on statistical identifiability for the proposed algorithm in the following theorem.

Theorem 1: Given \( T \) snapshots of sums of \( F \) \( N \)-D exponentials as in (2), in the absence of noise, the parameter set \( \{ \{ \omega_{f,n} \}_{n=1}^N, \{ c_f(t) \}_{t=1}^T \} \), \( f = 1, \ldots, F \), is almost surely uniquely identifiable by the weighted IMDF algorithm given in Section III, provided that

\[
F \leq \max \left[ \frac{\prod_{n=1}^{N} (K_n - 1), 2T \prod_{n=1}^{N} L_n}{K_n + L_n = M_n + 1, 1 \leq n \leq N} \right], \tag{27}
\]

where the \( NF \) frequencies \( \{ \omega_{f,n} \}_{n=1}^N, f = 1, \ldots, F \), and \( TF \) complex amplitudes \( \{ c_f(t) \}_{t=1}^T \), \( f = 1, \ldots, F \), are assumed to be drawn from distributions that are continuous with respect to the Lebesgue measure in \( \mathbb{C}^{NF \times 1} \) and \( \mathbb{C}^{TF \times 1} \), respectively.

Proof: In Section III we have shown that \( F \)-\( N \)-D frequency components can be uniquely recovered almost surely, provided that

\[
\prod_{n=1}^{N} K_n \geq F, \quad 2T \prod_{n=1}^{N} L_n \geq F, \quad \prod_{n=1}^{N} (K_n - 1) \geq F,
\]

where \( K_n \) and \( L_n \), \( n = 1, \ldots, N \), are positive integers subject to the constraint that

\[
K_n + L_n = M_n + 1, \quad 1 \leq K_n \leq M_n, \quad 1 \leq n \leq N.
\]

Since \( \prod_{n=1}^{N} (K_n - 1) \geq F \) implies \( \prod_{n=1}^{N} K_n \geq F \), if (27) holds, \( F \)-\( N \)-D frequency components are uniquely identifiable almost surely from (2). This concludes the proof. \( \blacksquare \)

It is difficult to find a closed form solution for the right hand side (RHS) of (27). But we can find its upper bound, which is given in Lemma 5. Defining the following function

\[
\Gamma_T(M_1, \ldots, M_N) := \max_{1 \leq K_n \leq M_n, 1 \leq n \leq N} \min \left( \prod_{n=1}^{N} K_n, 2T \prod_{n=1}^{N} L_n \right), \tag{28}
\]

then the RHS of (27) is \( \Gamma_T(M_1 - 1, \ldots, M_N - 1) \).

Lemma 5: If \( T \leq \prod_{n=1}^{N} (M_n - 1)/2 \), then

\[
\Gamma_T(M_1 - 1, M_2 - 1, \ldots, M_N - 1) \leq \left[ \frac{2TM}{(1 + \sqrt{2T})^N} \right]^{1/2},
\]

where \( M \) is defined in (1), and if \( T \geq \prod_{n=1}^{N} (M_n - 1)/2 \), then

\[
\Gamma_T(M_1 - 1, M_2 - 1, \ldots, M_N - 1) = \prod_{n=1}^{N} (M_n - 1).
\]

Proof: Define \( \rho_n = L_n/M_n, n = 1, \ldots, N \), and \( \rho = \sum_{n=1}^{N} \rho_n / N \). It is clear that \( \rho \in (0, 1] \). If \( K_n, n = 1, \ldots, N \)}
are all real numbers,
\[
\min \left( \prod_{n=1}^{N} (K_n - 1), 2T \prod_{n=1}^{N} L_n \right)
= \prod_{n=1}^{N} M_n \min \left( \prod_{n=1}^{N} (1 - \rho_n), 2T \prod_{n=1}^{N} \rho_n \right)
\leq M \min \left( (1 - \rho)^N, 2T \rho^N \right)
= \begin{cases} 
2T M, & 0 < \rho \leq \frac{1}{1 + \sqrt{2T}}, \\
M (1 - \rho)^N, & \frac{1}{1 + \sqrt{2T}} \leq \rho \leq 1,
\end{cases}
\leq \frac{2T M}{(1 + \sqrt{2T})^N}.
\]

The two equalities hold when \( L_n = \frac{M_n}{1 + \sqrt{2T}}, n = 1, \ldots, N \). Since \( L_n, n = 1, \ldots, N \), and \( \Gamma_T(M_1, \ldots, M_N) \) are all integers, the exact value of \( \frac{2T M}{(1 + \sqrt{2T})^N} \) may not be achieved, and thus
\[
\Gamma_T(M_1 - 1, \ldots, M_N - 1) \leq \frac{2T M}{(1 + \sqrt{2T})^N}.
\]

When \( T \geq \prod_{n=1}^{N} (M_n - 1)/2 \), since \( L_n \geq 1, M_n \geq K_n \) for \( n = 1, \ldots, N \), we have
\[
2T \prod_{n=1}^{N} L_n \geq \prod_{n=1}^{N} (M_n - 1) \prod_{n=1}^{N} L_n \geq \prod_{n=1}^{N} (M_n - 1) - \prod_{n=1}^{N} (K_n - 1).
\]

Therefore \( \Gamma_T(M_1 - 1, \ldots, M_N - 1) = \prod_{n=1}^{N} (M_n - 1) \) when \( T \geq \prod_{n=1}^{N} (M_n - 1)/2 \).  

It can be seen from Lemma 5 that when the number of snapshots \( T = 1 \), the identifiability bound is approximately \( \frac{2T M}{(1 + \sqrt{2T})^N} \) (which is approximately \( 0.34M_1M_2 \) in the 2-D case): when \( 1 \leq T \leq \prod_{n=1}^{N} (M_n - 1)/2 \), the identifiability bound is approximately \( \frac{2T M}{(1 + \sqrt{2T})^N} \), which increases as \( T \) increases; when \( T \geq \prod_{n=1}^{N} (M_n - 1)/2 \), the identifiability bound becomes a constant \( \prod_{n=1}^{N} (M_n - 1) \) (approximately the data sample size of one snapshot), and thus no data smoothing is necessary in this case from the perspective of maximizing identifiability bound since \( L_n = 1, K_n = M_n \), for \( n = 1, \ldots, N \).

Remark 1: In some applications, if it is known a priori that the frequencies in some dimension, say the \( i \)-th dimension, are distinct, then the statistical identifiability bound can be further relaxed. In this case, it is not necessary to use random weighting factors in (19) to ensure the eigenvalues of the matrix pencil distinct. In fact, we can construct a matrix pencil along the \( i \)-th dimension, the selection matrices in (19) can be replaced by
\[
J_1 := (I_{K_1} \otimes \cdots \otimes I_{K_{i-1}}) \otimes J_{1,i} \otimes (I_{K_{i+1}} \otimes \cdots \otimes I_{K_N}),
J_2 := (I_{K_1} \otimes \cdots \otimes I_{K_{i-1}}) \otimes J_{2,i} \otimes (I_{K_{i+1}} \otimes \cdots \otimes I_{K_N}).
\]

All the remaining steps are the same as those given in Section III. In this case, the identifiability condition (27) becomes
\[
F \leq \Gamma_T(M_1, \ldots, M_i - 1, \ldots, M_N).
\]

It can be shown that
\[
\Gamma_T(M_1, \ldots, M_i - 1, \ldots, M_N) \geq \Gamma_T(M_1 - 1, \ldots, M_i - 1, \ldots, M_N - 1),
\]
and thus the identifiability bound in (29) is higher than that in (27). When \( T = 1 \), the identifiability bound in (29) is that of the IMDF algorithm in [13].

If the frequencies in all \( N \) dimensions are distinct, we can construct a matrix pencil along the dimension with the largest dimension size.

Lemma 6: If \( M_i \geq M_j \), then
\[
\Gamma_T(M_1, \ldots, M_i - 1, \ldots, M_N) \geq \Gamma_T(M_1, \ldots, M_j - 1, \ldots, M_N).
\]

Proof: The proof is similar to that of Proposition 2 in [13].

Remark 2: We note that a similar data smoothing approach is also used in the Unitary ESPRIT algorithm [5]. However, it was not proved that the smoothed data matrix is of rank \( F \). Furthermore it was not pointed out in [5] that the \( N \)-D preprocessing step can be applied when \( T > 1 \). Although the identifiability bound was considered in a special 3-D frequency estimation case when \( M_1 \) and \( M_2 \) are fairly small and smoothing is only performed in the third dimension, the identifiability bound in a general case is not given in [5]. Actually, if \( N \)-D data smoothing is applied in the multiple snapshot case, it can be shown that the identifiability bound of the Unitary ESPRIT algorithm is \( \Gamma_T(M_1, \ldots, M_n - 1, \ldots, M_N) \), where \( M_n = \min_{1 \leq i < N} M_i \). From Lemma 6, we conclude the identifiability bound of the Unitary ESPRIT algorithm is less than or equal to that in (29), but is greater than that in (27). To our knowledge, (29) gives the most relaxed identifiability bound to date for \( N \)-D frequency estimation from finite snapshots.

As a special case, consider \( N = 2 \). In this case, the identifiability condition (27) becomes
\[
F \leq \max_{K_1 + L_1 = M_1, K_2 + L_2 = M_2} \min \{(K_1 - 1)K_2, 2TL_1L_2\}.
\]

The identifiability condition (29) becomes (assuming \( i = 1 \))
\[
F \leq \max_{K_1 + L_1 = M_1 + 1, K_2 + L_2 = M_2 + 1} \min \{(K_1 - 1)K_2, 2TL_1L_2\}.
\]

The identifiability condition of the Unitary ESPRIT algorithm [5] is
\[
F \leq \max_{K_1 + L_1 = M_1 + 1, K_2 + L_2 = M_2 + 1} \min \{(K_1 - 1)K_2, (K_1 - 1)K_2, 2TL_1L_2\}.
\]

In Fig. 1 (a), we plot the identifiability bound in (30) versus dimension size with \( M_1 = M_2 \). It can be seen that the identifiability bound generally increases with the increase of snapshot size and the number of snapshots. In Fig. 1 (b), we plot the identifiability bounds in (30), (31) and (32) versus the number of snapshots, respectively. It can be seen that when the number of snapshots \( T \) is less than a threshold (approximately \( M/2 \), as \( T \) increases, the bounds increase like stairs. When \( T \)}
is greater than the threshold, the bounds become approximately $M$. For example, for the identifiability condition given in (31), when $T \geq (M_1 - 1)M_2/2$, the bound becomes $M_1 - 1)M_2$.

V. PERFORMANCE ANALYSIS

In this section, we use perturbation analysis to derive the theoretical MSEs for the proposed algorithm in the noisy case. We need the following two assumptions for our analysis.

1) In the noiseless case, the nonzero singular values of $\tilde{X}$ in (17) are distinct, and the eigenvalues of $U_1^T U_2$ in (22) are distinct.

2) The eigenvectors in $T$ of (22) are normalized, and the permutation matrix $\Delta = I$ since permutation does not affect the MSEs.

Suppose that the perturbed data matrix is $\tilde{X} = X + \Delta \tilde{X}$, where $\Delta \tilde{X}$ is the perturbation term, which is related to the noise term in (4), $W$, such that (c.f., (15))

$$\Delta \tilde{X} = \left[ S[w(1)] \cdots S[w(T)] \right] - S[\Pi_M w^*(T)] \cdots S[\Pi_M w^*(1)] \right]$$

(33)

Let the SVD of $\tilde{X}$ in the noiseless case be

$$\tilde{X} = U_s \Sigma_s V_s^H + U_n \Sigma_n V_n^H,$$

where the vectors in $U_s$, associated with the $F$ non-zero singular values, span the signal subspace, while the vectors in $U_n$, associated with the zero singular values, span the orthogonal subspace of $U_s$. It is clear that $\Sigma_n = 0$. The perturbed data matrix is $\tilde{X} = X + \Delta \tilde{X}$, whose SVD is given by

$$\tilde{X} = \tilde{U}_s \tilde{\Sigma}_s \tilde{V}_s^H + \tilde{U}_n \tilde{\Sigma}_n \tilde{V}_n^H.$$  \hspace{1cm} (35)

The perturbed signal subspace is $\tilde{U}_s = U_s + \Delta U_s$. In Appendix B, we give a first-order approximation of $\Delta U_s$ as a function of $w(t)$, $t = 1, \ldots, T$. Suppose that the frequencies estimated from $\tilde{X}$ are $\tilde{\omega}_{f,n}$, which is written as $\tilde{\omega}_{f,n} = \omega_{f,n} + \Delta \omega_{f,n}$, for $f = 1, \ldots, F$ and $n = 1, \ldots, N$. In the following, we derive the theoretical MSEs of $\omega_{f,n}$ based on the result of Appendix B.

Given $\Delta U_s$, the perturbations of $U_1$ and $U_2$ in (20) are

$$\Delta U_1 = J_1 \Delta U_s, \quad \Delta U_2 = J_2 \Delta U_s.$$  \hspace{1cm} (36)

According to (21), we have

$$U_1 T D(\zeta) = U_2 T.$$  \hspace{1cm} (7)

Differentiating (37), we obtain

$$(\Delta U_1 \zeta_f - \Delta U_2) t_f + U_1 t_f \Delta \zeta_f = - (U_1 \zeta_f - U_2) \Delta t_f.$$  \hspace{1cm} (38)

Since $t_f$ is the $f$-th normalized eigenvector of $U_1^T U_2$, according to [24, Chapter 2], we can write $\Delta t_f$ as a linear combination of the eigenvectors $t_g$, $1 \leq g \leq F$, $g \neq f$, such that

$$\Delta t_f = \sum_{g=1, g \neq f}^F k_{g,f} t_g.$$  \hspace{1cm} (39)

From (21), we have

$$P_0^H U_2 = D(\zeta) P_0^H U_1.$$  \hspace{1cm} (40)

Suppose that $P_0^H = [s_1 s_2 \cdots s_F]^T$, then $s_0^H (U_1 \zeta_f - U_2) = 0^T.$

Also from (21), it can be verified that $s_0^H U_1 t_g = \delta_{g,g'}$ and $s_0^H U_2 t_g = \delta_{g,g'}$, where $\delta_{g,g'}$ is the Kronecker delta ($\delta_{g,g'} = 1$ when $g = g'$; $\delta_{g,g'} = 0$ when $g \neq g'$). If we multiply both sides of (39) by $s_0^H$, we can obtain

$$\Delta \zeta_f = - s_0^H (\Delta U_1 \zeta_f - \Delta U_2) t_f.$$  \hspace{1cm} (41)

We note that the above analysis is similar to the perturbation analysis of eigenvalues in ESPRIT-type algorithms [6, 10, 12, 18]. The expression of $\Delta \zeta_f$ as in (40) was also obtained in [6, 10, 12, 18]. However, here we are interested in the perturbation of eigenvectors. We multiply both sides of (39) by $s_0^H (g \neq f)$ and obtain

$$k_{g,f} = - s_0^H (\Delta U_1 \zeta_f - \Delta U_2) t_f,$$  \hspace{1cm} (42)

where $\Delta t_f = R_f \Delta U_s t_f$, where

$$R_f := \sum_{g=1, g \neq f}^F - t_g s_0^H (J_1 \zeta_f - J_2)\frac{\zeta_f - \zeta_g}{\zeta_f - \zeta_g} = - T D_f P_0^H (J_1 \zeta_f - J_2).$$  \hspace{1cm} (43)

Here $D_f$ is a diagonal matrix with $[D_f]_{g,g} = \frac{1}{\zeta_f - \zeta_g}$, for $g \neq f$ and $[D_f]_{g,g} = 0$. Since the $f$-th column of $P$ is obtained using $p_f = U_1 t_f$, we have

$$\Delta p_f = \Delta U_1 t_f + U_1 \Delta t_f = (J_1 + U_1 R_f) \Delta U_s t_f.$$  \hspace{1cm} (44)

Now we are ready to obtain the perturbation of $\omega_{f,n}$. Differentiating (26), we have

$$\Delta \omega_{f,n} = \frac{\Delta e^{j \omega_{f,n}}}{e^{j \omega_{f,n}}} = \frac{1}{\mu_n} \tau_{f,n}^T \Delta p_f.$$  \hspace{1cm} (45)

where $\tau_{f,n}$ is defined as

$$\tau_{f,n} = \frac{1}{\mu_n} \sum_{t=1}^T \left[ e^{j \omega_{f,n}} w(t) + e^{j \omega_{f,n}} T \right].$$  \hspace{1cm} (46)

Here $p_{1,f}$ is the $(1,f)$-th element of $P_0$, which is also the $f$-th diagonal element of $A$. Notice that the nonzero elements of $\tau_{f,n}$ are the reciprocals of corresponding $[p_{k,f}]$ with scaling ambiguity, and the zero elements are distributed regularly. Substituting (71) to (42) and (42) to (44), we have

$$\frac{\Delta e^{j \omega_{f,n}}}{e^{j \omega_{f,n}}} = \frac{1}{\mu_n} \sum_{t=1}^T \left[ e^{j \omega_{f,n}} w(t) + e^{j \omega_{f,n}} T \right].$$  \hspace{1cm} (47)
where \( t^T_f := [t_{f,1}, t_{f,2}, \ldots, t_{f,F}] \),
\[
E[f^T_{j,n}(t) = \tau^T_{f,n}(J_1 + U_1 R_f) \sum_{g=1}^{F} t_{f,g} \Phi_{g,t},
\]
and \( \Phi_{g,t} \) and \( \Phi_{g,t} \) are defined in (71). Since \( E[w(t)] = 0 \), it is ready to see
\[
E[\Delta \omega_{f,n}] = 0.
\] (46)
This means that the estimation is unbiased if only the first-order perturbation is considered. Using the following properties of complex Gaussian noise,
\[
E[\mathcal{R}(w(t)) \mathcal{R}(w^T(s))] = \frac{\sigma^2}{2} I_{M_1 M_2} \delta_{t,s},
\]
\[
E[I(w(t)) I(w^T(s))] = \frac{\sigma^2}{2} I_{M_1 M_2} \delta_{t,s},
\]
\[
E[\mathcal{R}(w(t)) \mathcal{R}(w^T(s))] = E[I(w(t)) I(w^T(s))] = 0,
\]
we can obtain the theorethic MSES for the frequency estimates as
\[
E[(\Delta \omega_{f,n})^2] = \frac{\sigma^2}{2}\rho_n^2 \sum_{t=1}^{T} \left[ \| \mathcal{R}(\xi^T_{f,n}(t) - \xi^T_{f,n}(t)) \|^2 + \| \mathcal{I}(\xi^T_{f,n}(t) + \xi^T_{f,n}(t)) \|^2 \right],
\] (47)
for \( n = 1, \ldots, N, \) and \( f = 1, \ldots, F \).

VI. Optimizing weighting factors to minimize MSES

Based on the MSES of estimation obtained in the previous section, we now seek to optimize the weighting factors \( \{\alpha_n\}_{n=1}^{N} \) to minimize the MSES and improve the performance of the proposed algorithm. We notice that minimizing (47) with respect to \( \{\alpha_n\}_{n=1}^{N} \) is too complex to be solved. However, the MSES can be regarded as asymptotical error variances when the noise power \( \sigma^2 \) approximates zero, and thus we can take its ratio with the Cramér-Rao Bound (CRB) as an equivalent performance measure, and then use inequalities to bound this ratio so that only entries related to \( \{\alpha_n\}_{n=1}^{N} \) need to be minimized.

We have derived the CRB of N-D frequency estimation from (2) using a method similar to that of [22]. The result is given as follows:
\[
\text{var}_{\text{CRB}}(\hat{\omega}_{f,n}) = \frac{\sigma^2 b(n-1)F+f}{2},
\] (48)
where \( b(n-1)F+f \) is the \((nF - F + f)\)-th element of following vector
\[
b = \text{Diag}\left\{ \frac{1}{N} \sum_{i=1}^{T} \mathcal{R}(C_f^H(t)Z^H(I_M - A(A^H A)^{-1} A^H) Z C_e(t))^{-1} \right\}
\]
where \( C_e(t) := I_N \otimes D(e(t)) \), and
\[
Z = [z_{1,1}, z_{2,1}, \ldots, z_{F,1}, z_{1,2}, \ldots, z_{F,N}] : M \times NF,
\]
\[
z_{f,n} = (\theta_{f,1} \otimes \ldots \otimes \theta_{f,n-1}) \otimes \tau_{f,n} \otimes (\theta_{f,n+1} \otimes \ldots \otimes \theta_{f,N})
\]
\[
\theta_{f,n} = [1, e^{j\omega_{n}^T}, \ldots, e^{j(M_n-1)\omega_{n}^T}] \quad \text{T} : M_n \times 1,
\]
\[
\theta_{f,n} = [0, e^{j\omega_{n}^T}, \ldots, j(M_n-1) e^{j(M_n-1)\omega_{n}^T}] \quad \text{T} : M_n \times 1.
\]
We then define the asymptotical efficiency of the algorithm on the estimation of \( \omega_{f,n} \) as
\[
\eta_{f,n} := \lim_{\sigma^2 \to 0} \frac{E[\Delta \omega_{f,n}^2]}{\text{var}_{\text{CRB}}(\hat{\omega}_{f,n})}.
\] (49)
We can bound \( \eta_{f,n} \) using properties of matrix norms such that
\[
\eta_{f,n} \leq \frac{1}{\rho_n^2 b(n-1)F+f} \| \tau_{f,n} \|^2 \left( \| J_1 \| + \| \nu_1 \| |R_f| \right)^2.
\]
\[
\sum_{t=1}^{T} \left( \| \tau_{f,n} \| \left( \| \Phi_{g,t} \| + \| \Phi_{g,t} \| \right)^2 + \sum_{g=1}^{F} \| \tau_{f,n} \| \left( \| \Phi_{g,t} \| - \| \Phi_{g,t} \| \right)^2 \right).
\]
In the RHS of the above inequality, only \( R_f \) depends on the weighting factors \( \{\alpha_n\}_{n=1}^{N} \). Furthermore, \( \| R_f \| \leq \| T \| \| D_f \| \left( \| J_1 \| + \| J_2 \| \right) \). Define \( \alpha := [\alpha_1, \ldots, \alpha_N]^T \) and \( \omega_f := [e^{j\omega_1^T}, \ldots, e^{j\omega_n^T}]^T \). Notice that \( \omega_f = \omega_f^T \) (see (21)). Ignoring all the entries that do not depend on \( \alpha \), we find the upper bound of the efficiency \( \eta_{f,n} \) is decided by
\[
\| D_f \| \left( \| J_1 \| + \| J_2 \| \right) = \sum_{g=1}^{N} \alpha_n (\| J_1 \| + \| J_2 \|) \]
\[
= \sum_{g=1}^{N} \alpha_n (\| J_1 \| + \| J_2 \|) \| (\omega_f^T - \omega_f^T) \| \alpha \| \| \omega_f^T \| \| \omega_f^T \| \alpha \|
\]
\[
\leq \frac{1}{NF} \sum_{f=1}^{N} \sum_{g=1}^{F} \eta_{f,n},
\] (52)
we propose to minimize the following cost function
\[
\gamma(\alpha) = \frac{1}{2} \sum_{f=1}^{F} \gamma_f(\alpha) = \frac{F-1}{2} \sum_{f=1}^{F} \frac{1}{\| \omega_f^T - \omega_f^T \| \alpha \| \omega_f^T \| \alpha \|}
\] (53)
An optimal \( \alpha \) can be obtained by
\[
\alpha_{\text{opt}} = \arg \min_{\alpha} \gamma(\alpha), \quad \text{subject to } \| \omega_f \| \leq 1.
\] (54)
In fact the optimization criterion (54) has a simple interpretation in term of the eigenvalue distribution of \( U_1^T U_2 \). We define the difference of eigenvalues as \( d_{f_1, f_2} := \zeta_{f_1} - \zeta_{f_2} = (\omega_{f_1}^T - \omega_{f_2}^T) \alpha \), for \( 1 \leq f_1, f_2 \leq F \). Notice that these differences appear as the denominators in the entries of \( \gamma(\alpha) \) (see (53)) and the diagonal elements of \( D_f \) (see (41)). If \( \alpha \) is approximately orthogonal to some \( \omega_{f_1}^T - \omega_{f_2}^T, d_{f_1, f_2} \) is close
to zero, then $\gamma(\alpha)$ will be very large, and the performance of the algorithm will degrade dramatically due to high MSE.

We further use an example to show the effect of $\alpha$ on $\eta$. Suppose that three 3-D frequency components are to be estimated from 3 snapshots of $6\times6\times6$ noisy data samples as given in (2). The 3-D frequency components are

\[
\begin{align*}
  f = 1 : & \quad (\omega_{1,1},\omega_{1,2},\omega_{1,3}) = (0.70\pi, 0.50\pi, 0.20\pi), \\
  f = 2 : & \quad (\omega_{2,1},\omega_{2,2},\omega_{2,3}) = (0.80\pi, 0.60\pi, 0.20\pi), \\
  f = 3 : & \quad (\omega_{3,1},\omega_{3,2},\omega_{3,3}) = (0.80\pi, 0.50\pi, 0.40\pi).
\end{align*}
\]

Fig. 2 depicts the logrithmic plot of $\eta$ as a function of $\alpha_1$ and $\alpha_2$, while $\alpha_3$ is fixed at $\frac{1}{\sqrt{3}}$. For simplicity of the plot, we also set $|\alpha_1| = \sqrt{\frac{1}{3}}$ and hence the $x$- and $y$-axes in Fig. 2 are the angles of $\alpha_1$ and $\alpha_2$ respectively.

As illustrated in Fig. 2, in most cases, $\alpha$ keeps the average efficiency $\eta$ as small as 1, but when $\alpha$ falls into some “bad regions”, the average efficiency becomes very large. A random selected $\alpha$ does not guarantee that $\eta$ does not fall into the “bad regions”. Therefore we propose to use (54) to optimize the choice of $\alpha$.

The optimization problem (54) is a so-called sum-of-ratios fractional programming problem, which is a difficult global optimization problem [3]. There is no efficient algorithm available to solve it to date. From Fig. 2 we notice that the “bad regions” are often regular and small, most choices of $\alpha$ are fairly good to obtain a small average efficiency. We can use grid search in the super-sphere $f$ estimated from 3 snapshots of $\omega_{i,j,k}$ to zero, then

\[
\begin{align*}
  f = 1 : & \quad (\omega_{1,1},\omega_{1,2},\omega_{1,3}) = (0.70\pi, 0.50\pi, 0.20\pi), \\
  f = 2 : & \quad (\omega_{2,1},\omega_{2,2},\omega_{2,3}) = (0.80\pi, 0.60\pi, 0.20\pi), \\
  f = 3 : & \quad (\omega_{3,1},\omega_{3,2},\omega_{3,3}) = (0.80\pi, 0.50\pi, 0.40\pi).
\end{align*}
\]

solve (54) by grid search first then refine it using a Newton method, referred to as “Grid+Newton”. These approaches are applied to estimate three 3-D frequency components from 3 snapshots of $6\times6\times6$ noisy data samples. The frequency components are given in (55)-(57). Fig. 3 (a) depicts the root mean-square error (RMSE) versus signal-to-noise ratio (SNR). The RMSE is obtained by averaging over all frequencies after only one iteration of Steps 3-4, except for the case indicated with “Random $\alpha$” where only Steps 1-2 are executed. The SNR is defined as (c.f., (4)).

\[
\text{SNR} = 10\log_{10} \frac{\|AC\|^2}{M_1M_2M_3T\sigma^2}.
\]

The corresponding CRB on standard deviation (STD) is also plotted.

It can be seen from Fig. 3 (a) that the three optimization methods provide similar performance, and all outperform the case with randomly chosen $\alpha$. It turns out one iteration of Steps 3-4 is sufficient, as demonstrated by Fig. 3 (b), where we plot the RMSE of frequency estimates versus the number of iterations of Steps 3-4. Zero iteration corresponds to the case with only randomly chosen $\alpha$. We observe that one execution of Steps 3 and 4 is sufficient as further iterations only provide negligible performance improvement if any. This is largely due to that “bad regions” only account for a very small percentage of the total area as shown in Fig. 2. We also note here our main aim is to avoid $\alpha$ falling into those “bad regions”, rather than finding an absolute optimal solution, therefore there is no convergence issue as there are many good choices of $\alpha$. Furthermore, it can be seen that “Minmax+Newton” and “Grid+Newton” are comparable, and both are slightly better than “Minmax”. Because “Grid+Newton” has a higher complexity than “Minmax+Newton”, we choose “Minmax+Newton” with one iteration as the proposed algorithm in Table I for optimizing the weighting factors, which is the algorithm used in the simulations of Section VII.

VII. SIMULATION RESULTS

In this section we present the Monte Carlo simulation results to demonstrate the performance (measured by RMSE) of the proposed algorithm, which is also compared to other N-D frequency estimation algorithms as well as the associated CRB.

A. 2-D Identical Frequency Estimation from Single Snapshot

In the first experiment, the proposed algorithm, MEMP [7], Unitary ESPRIT [5] and 2-D ESPRIT [19] are applied to estimate three 2-D frequency components from a $20\times20$ noisy data set. The amplitudes, $c_f(1)$ for $f = 1, \ldots, F$, are set to be one for this case. The three frequency pairs are

\[
\begin{align*}
  (\omega_{1,1},\omega_{2,1}) &= (0.55\pi, 0.20\pi), \\
  (\omega_{2,1},\omega_{2,2}) &= (0.60\pi, 0.20\pi), \\
  (\omega_{3,1},\omega_{3,2}) &= (0.60\pi, 0.25\pi).
\end{align*}
\]

Notice that there are identical frequencies in both dimensions, which is a case that the IMDF algorithm fails to deal with. Fig. 4 depicts the performance comparison. In Fig. 4 (a), we plot
the simulated RMSE of various algorithms and the average CRB on STD in the two dimensions. The RMSE results are averaged over all frequencies and obtained through 1000 realizations. In Fig. 4 (a), “Proposed algorithm” refers to the one with one iteration of Steps 3-4 using “Minmax+Newton” as given in Table I. For our proposed algorithm, the smoothing parameters \((K_n)_{n=1}^2; (L_n)_{n=1}^2\) are chosen such that the identifiability bound in (30) can be achieved. In other words, here we do not optimize the smoothing parameters with respect to RMSE performance, as similar analysis can be found in [13]. The parameters for other algorithms are chosen as follows according to (notations are also from) their respective references: (i) for the Unitary ESPRIT algorithm, \(M_{\text{sub1}} = 10, M_{\text{sub2}} = 10\); (ii) for the MEMP algorithm, \(K = 6\) and \(L = 6\), and exhaustive search is used for frequency pairing; (iii) for the 2-D ESPRIT algorithms, \(K = 6, L = 6\) and \(\beta = 8\). As shown in Fig. 4 (a), the proposed algorithm offers comparable performance as that of the Unitary ESPRIT algorithm, and outperforms 2-D ESPRIT and MEMP.

In Fig. 4 (b), we compare the optimized weighting factors to randomly chosen ones, where “Random \(\alpha\)” means zero iteration of Steps 3-4 in Table I. The theoretic RMSE is obtained from the square root of the average of (47), for \(n = 1, 2\) and \(F = 1, \ldots, F\) and \(t = 1, \ldots, T\), with an optimized \(\alpha\) by solving (54) using the true frequencies, which serves as a benchmark since in our algorithm \(\alpha\) is optimized when the true frequencies are unknown. It is clear that the proposed algorithm significantly outperforms the one with random weighting factors, and the simulated RMSE of the proposed algorithm matches well to the theoretic RMSE for moderate to high SNR range.

B. 2-D Close Frequency Estimation from Multiple Snapshots

In the second experiment, the proposed algorithm, Unitary ESPRIT and RARE are applied to estimate three 2-D frequency components from 10 snapshots of noisy data, each of size \(12 \times 12\), as given in (3). The amplitudes, \(c_f(t)\), for \(f = 1, \ldots, F\) and \(t = 1, \ldots, T\), are drawn from a complex Gaussian distribution. The three frequency pairs are

\[
(\omega_{1,1}, \omega_{1,2}) = (0.72\pi, 0.62\pi),
(\omega_{2,1}, \omega_{2,2}) = (0.74\pi, 0.58\pi),
(\omega_{3,1}, \omega_{3,2}) = (0.76\pi, 0.60\pi).
\]

Notice that frequencies are close to each other in both dimensions. Fig. 5 depicts the simulated RMSE of various algorithms, along with the corresponding CRB and the theoretic RMSE of the proposed algorithm. Multidimensional data smoothing is also performed for the Unitary ESPRIT algorithm and the RARE algorithm. It can be seen from Fig. 5 (a) that the proposed algorithm offers competitive performance when compared with the Unitary ESPRIT and RARE algorithms. Note that the proposed algorithm has lower complexity than these two algorithms, because the proposed algorithm does not require a frequency pairing step (the Unitary ESPRIT algorithm achieves automatic frequency pairing through iterative joint diagonalization). Fig. 5 (b) confirms again that optimized weighting factors outperform randomly chosen weighting factors, and the simulated RMSE of the proposed algorithm matches its theoretic RMSE at high SNR.

C. 3-D Identical Frequency Estimation from Multiple Snapshots

In the third experiment, the proposed algorithm and the Unitary ESPRIT algorithm are applied to estimate three 3-D frequency components from 10 snapshots of \(6 \times 6 \times 6\) noisy data samples. The frequency components are given in (55)-(57). Notice that there are identical frequencies in all dimensions. The amplitudes are drawn from a complex Gaussian distribution. Fig. 6 shows the performance comparisons. From Fig. 6, we notice the proposed algorithm also offer competitive performance in 3-D frequencies estimation compared to the Unitary ESPRIT algorithm. Because the pairing strategy of RARE algorithm is not applicable when all three dimensions have identical frequencies, we do not include RARE in this experiment. Notice that the simulated RMSE of the proposed algorithm matches its theoretic RMSE for moderate to high SNR range.

D. Complexity Comparison

Because it is difficult to analytically calculate the complexity order of the optimization step of the proposed algorithm, here we use the Performance API (PAPI) tool [2] of Matlab to count the number of floating point operations. We compare the complexity of our algorithm to that of \(N\)-D Unitary ESPRIT [5] whose analytical form of complexity order is also unavailable due to the iteration nature of the simultaneous Schur decomposition. Note that both algorithms obtain the signal subspace using SVD with a similar data smoothing step in the single snapshot case, therefore the cost of the frequency estimation and pairing steps determines the difference in complexity. In order to have a fair comparison, we apply the two algorithms to estimate 3-D frequencies from a data set of size \(M_1 \times M_2 \times M_3\) using the same smoothing parameters. We set \(M_1 = M_2 = M_3\). Figs. 7 (a) and 7 (b) plot the number of floating point operations of the two algorithms as a function of \(M_1\) for \(F = 3\) and \(F = 30\), respectively. It can be observed that the proposed algorithm has lower complexity than Unitary ESPRIT and the difference increases as the number of frequency components increases.

VIII. CONCLUSIONS

We have proposed an eigenvector-based algorithm for \(N\)-D frequency estimation from multiple data snapshots. We have analytically quantified the identifiability (in the noiseless case) and the performance (in the noisy case) of the proposed algorithm. It is shown that our algorithm offers the most relaxed statistical identifiability bound to date. It remains operational when there exist identical frequencies in one or more dimensions, due to the adoption of weighting factors. Furthermore, a low-complexity (one iteration) approach is developed to optimize the weighting factors by minimizing MSEs of frequency estimates. Simulation results show that the proposed algorithm offers better or competitive performance...
when compared to existing algebraic approaches for N-D frequency estimation, but at lower complexity since frequency estimates are automatically paired without multiple iterations of joint diagonalization (as in the Unitary ESPRIT) or requiring an extra pairing step (as in MEMP or RARE). Numerical examples also demonstrate that the optimized weighting factors significantly outperform randomly chosen weighting factors.

**APPENDIX A**

**Proof of Lemma 4**

*Proof:* Define $L := \prod_{p=1}^{N} L_p$. We only need to prove that the determinant of $\tilde{H}$ is nonzero when $2TL = F$. Since when $2TL < F$ or $2TL > F$, our proof still holds for the corresponding column-reduced or row-reduced square submatrix.

Suppose that $2TL = F$. The determinant of $\tilde{H}$, $\det(\tilde{H})$, is a polynomial of $(N + T)F$ variables $\{e^{i\omega_{j,n}}\}_{j=1}^{N}$, $\{e^{i\omega_{j,n}}\}_{j=1}^{N}$, $f = 1, \cdots, F$, and thus is analytic in $\mathbb{C}^{(T+N)F \times 1}$. According to Lemma 1, if we can find one point in $\mathbb{C}^{(T+N)F \times 1}$ such that $\det(\tilde{H}) \neq 0$, $\tilde{H}$ has full column rank almost surely. Define

$$L_n := \begin{cases} \prod_{p=n+1}^{N} L_p, & 0 \leq n \leq N - 1, \\ 1, & n = N. \end{cases}$$

(61)

If we choose the $(T + N)F$ variables such that

$$e^{i\omega_{f,n}} = c e^{i(\gamma_n^T + \nu_0)},$$

for $t = 1, \cdots, T$;

$$e^{i\omega_{j,n}} = e^{i\nu(Tn)},$$

for $n = 1, \cdots, N$,

where $f = 1, \cdots, F$, $c > 0$, $\nu_0 := 2\pi \frac{L_n}{2T}$, and

$$\nu_0 := \nu_0 \left[ \sum_{n=1}^{N} (M_n - 1) L_n + 1 + \frac{1}{2T} \right].$$

Then, it can be verified that in this case both $\mathbf{B}$ in (9) and $\mathbf{H}$ in (11) are Vandermonde matrices with generators equi-spaced on the unit circle, and thus $\det(H) \neq 0$. Therefore $\tilde{H}$ has full rank almost surely.

**APPENDIX B**

**Perturbation of the Signal Subspace**

Given (4), (15), (34), and (35), we derive the first-order perturbation of the signal subspace, $\Delta U_s$, when $\mathbf{X}$ is perturbed by $\mathbf{W}$, or equivalently, $\tilde{X}$ is perturbed by $\Delta \tilde{X}$. Suppose that the nonzero singular values of $\tilde{X}$ in (34) are $\lambda_1 > \lambda_2 > \cdots > \lambda_F$. For the $f$-th singular vector $u_f$ in $U_s$, we have

$$\tilde{X} \tilde{X}^H u_f = \lambda_f^2 u_f.$$  

(62)

Differentiating (62), we obtain

$$\Delta \tilde{X} \tilde{X}^H u_f + \tilde{X} \Delta \tilde{X}^H u_f + \Delta \tilde{X} \tilde{X}^H u_f = 2\lambda_f \Delta \lambda_f u_f + \lambda_f^2 \Delta u_f.$$  

(63)

According to [24, Chapter 2], if $\Delta u_f$ is approximately orthogonal to $u_f$, we can express the perturbation $\Delta u_f$ as a linear combination of $u_g$, where $g = 1, \ldots, K, g \neq f$. Therefore,

$$\Delta u_f = \sum_{g=1, g \neq f}^{K} q_{gf} u_g,$$

(64)

where $\{q_{gf}\}$ are the coefficients to be determined. Substituting it into (63), multiplying both sides from the left by $u^H_g$, $g \neq f$, and using $u^H_X \tilde{X}^H u_g = \lambda_g^2 \delta_{g,g'}$ and $u^H u_g = \delta_{g,g'}$, where $\delta_{g,g'}$ is the Kronecker delta ($\delta_{g,g'} = 1$ when $g = g'$; $\delta_{g,g'} = 0$ when $g \neq g'$), we have

$$q_{gf} = \frac{u^H_g \left( \Delta \tilde{X} \tilde{X}^H + \tilde{X} \Delta \tilde{X}^H \right) u_f}{\lambda_g^2 - \lambda_f^2},$$

(65)

Substituting (65) to (64) and using $\tilde{X}^H u_f = v_f \lambda_f$ and $U^H_n \tilde{X} = 0$, we obtain

$$\Delta u_f = U_n U^H_n \Delta \tilde{X} v_f \lambda_f^{-1} + U_s S_f U^H_s \Delta \tilde{X} v_f \lambda_f + U_s S_f \Sigma F^H \Delta \tilde{X} u_f,$$

(66)

where $v_f$ is the $f$-th column of $V_s$, and $S_f$ is a diagonal matrix of size $F \times F$ with the $(f,f)$-th element as 0 and $(g,g)$-th element as $\frac{1}{\lambda_g^2 - \lambda_f^2}$, for $g = 1, \ldots, F, g \neq f$.

Notice that the result in (66) is different from the first-order perturbation results of the signal subspace given in [12], where only the contribution of the noise subspace (i.e., the first term of the RHS of (66)) was taken into account. However, (66) is consistent with the results of the eigenvector perturbation analysis of [4], [10], [18].

Next we express $\Delta u_f$ as a function of $W$ by substituting (33) to (66). First, we split the vector $v_f$ into $2T$ equal-sized sub-vectors such that

$$v_f := [v^T_f, v^T_{f,2} \cdots v^T_{f,T}, \tilde{v}^T_f, \cdots \tilde{v}^T_{f,1}]$$

Then, it can be verified that in this case both $\mathbf{B}$ in (9) and $\mathbf{H}$ in (11) are Vandermonde matrices with generators equi-spaced on the unit circle, and thus $\det(H) \neq 0$. Therefore $\tilde{H}$ has full rank almost surely.
where \( E_f := U_n U_n^H \lambda_f^{-1} + U_s S_f U_s^H \lambda_f \). For the third term on the RHS of (66), we can divide the matrix \( U_s S_f \sum_s V_s^H \) into 27 blocks such that

\[
U_s S_f \sum_s V_s^H = \begin{bmatrix}
K_{f,1} & K_{f,2} & \cdots & K_{f,T} & \widetilde{K}_{f,T} & \cdots & \widetilde{K}_{f,1}
\end{bmatrix}.
\]

Then, it can be verified that

\[
U_s S_f \sum_s V_s^H \Delta X^H u_f = \sum_{t=1}^T K_{f,t} \mathcal{U}_f w(t)
+ \sum_{t=1}^T \widetilde{K}_{f,t} \mathcal{U}_f \Pi_M w(t), \tag{70}
\]

where

\[
\mathcal{U}_f := \begin{bmatrix}
u_f^T J_{1,1,\ldots,1} \\
u_f^T J_{1,1,\ldots,2} \\
\vdots \\
u_f^T J_{1,1,\ldots,L_N}
\end{bmatrix}.
\]

Summing (69) and (70), we can write (66) as

\[
\Delta u_f = \sum_{t=1}^T \Phi_{f,t} w(t) + \sum_{t=1}^T \widetilde{\Phi}_{f,t} w^*(t), \tag{71}
\]

where

\[
\Phi_{f,t} = E_f Q_{f,t} + \widetilde{K}_{f,t} \mathcal{U}_f \Pi_M,
\]

\[
\widetilde{\Phi}_{f,t} = E_f \widetilde{Q}_{f,t} \Pi_M + K_{f,t} \mathcal{U}_f.
\]

ACKNOWLEDGEMENT

The authors would like to thank an anonymous reviewer whose comments have helped to improve the perturbation analysis of the signal subspace.

REFERENCES


Xiangqian Liu (S’99-M’02) received the B.S. degree in Electrical Engineering from Beijing Institute of Technology, Beijing, China, in 1997, the M.S. degree in Electrical Engineering from the University of Virginia, Charlottesville, in 2000, and the Ph.D. degree in Electrical Engineering from the University of Minnesota, Minneapolis, in 2002. Since 2002, he has been an assistant professor of Electrical and Computer Engineering at University of Louisville. His current research interests are in the areas of wireless communications and signal processing, including multiuser detection, channel modeling and estimation, multi-antenna systems, spectral estimation, sensor array processing, as well as signal processing techniques for wireless sensor networks.

Xiaoli Ma (M’03) received the B.S. degree in Automatic Control from Tsinghua University, Beijing, China in 1998, the M.S. degree in Electrical Engineering from the University of Virginia in 2000, and the Ph.D. degree in Electrical Engineering from the University of Minnesota in 2003. From 2003 to 2005, she was an assistant professor of Electrical and Computer Engineering at Auburn University. Since 2006, she has been with the School of Electrical and Computer Engineering at Georgia Institute of Technology. Her research interests include transceiver designs and diversity techniques for wireless time- and frequency-selective channels, channel modeling, estimation and equalization, carrier frequency synchronization for OFDM systems, routing and cooperative designs for wireless networks.

Dr. Ma serves as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS, and is a member of the Signal Processing Theory and Methods Technical Committee of the IEEE Signal Processing Society.

### TABLE I

**AN IMPROVED EIGENVECTOR-BASED ALGORITHM USING OPTIMAL WEIGHTING FACTORS**

1) Given (3), follow (12)–(17) to obtain $U_s$.

2) Randomly select $\alpha$ subject to $|\alpha_n| = \sqrt{\frac{1}{M}}, n = 1, \ldots, N$, compute $\hat{\omega}_{f,n} |_{n=1}^{N}, f = 1, \ldots, F$, using (19)–(26).

3) Based on $\hat{\omega}_{f,n} |_{n=1}^{N}$, for $f = 1, \ldots, F$, obtain an updated $\alpha_{opt}$ by first solving (59) using SQP to get initials, then solving (54) using a Newton method.

4) Compute updated $\hat{\omega}_{f,n} |_{n=1}^{N}, 1 \leq f \leq F$, with $\alpha_{opt}$ using (19)–(26).

5) Iterate Steps 3 and 4 until frequency estimates converge (typically one execution of Steps 3-4 is sufficient).

Fig. 1. Identifiability (ID) bound of 2-D frequency estimation: (a) ID bound versus $M_1$ when $M_1 = M_2$; (b) ID bounds versus the number of snapshots where $M_1 = 10$ and $M_2 = 6$.

![Fig. 1](image)

Fig. 2. The average efficiency as a function of $\alpha_1$ and $\alpha_2$ ($|\alpha_1| = \sqrt{\frac{1}{3}}, |\alpha_2| = \sqrt{\frac{2}{3}},$ and $\alpha_3 = \sqrt{\frac{1}{3}}$).

![Fig. 2](image)
Fig. 3. (a) RMSE of different optimization methods versus SNR; (b) RMSE of different optimization methods versus the number of iterations of Steps 3-4.

Fig. 4. (a) Comparison of different algorithms for 2-D frequency estimation from single snapshot; (b) Comparison of optimized $\alpha$ and randomly chosen $\alpha$. 
Fig. 5. (a) Comparison of different algorithms for 2-D frequency estimation from multiple snapshots; (b) Comparison of optimized $\alpha$ and randomly chosen $\alpha$.

Fig. 6. (a) Comparison of different algorithms for 3-D frequency estimation from multiple snapshots; (b) Comparison of optimized $\alpha$ and randomly chosen $\alpha$. 
Fig. 7. The number of floating point operations versus $M_1$ when: (a) $F = 3$; (b) $F = 30$. 

Proposed Algorithm 
Unitary ESPRIT