

Finite Sample Identifiability of Multiple Constant Modulus Sources

Amir Leshem, *Member, IEEE*, Nicolas Petrochilos, and
Alle-Jan van der Veen, *Senior Member, IEEE*

Abstract—We prove that mixtures of continuous alphabet constant modulus sources can be identified with probability 1 with a finite number of samples (under noise-free conditions). This strengthens earlier results which only considered an infinite number of samples. The proof is based on the linearization technique of the analytical constant modulus algorithm (ACMA), together with a simple inductive argument. We then study the finite-alphabet case. In this case, we provide a subexponentially decaying upper bound on the probability of nonidentifiability for a finite number of samples. We show that under practical assumptions, this upper bound is tighter than the currently known bound. We then provide an improved exponentially decaying upper bound for the case of L -PSK signals (L is even).

Index Terms—Blind source separation, Chernoff bound, constant modulus signals, finite sample analysis, identifiability, large deviations, phase-shift keying (PSK).

I. INTRODUCTION

The constant modulus algorithm (CMA) is very popular for blind equalization [1], [2]. Similarly, the separation of constant modulus (CM) signals has attracted much attention in the signal processing literature. In [3], a direction-of-arrival (DOA) estimation based on the sequential separation of CM signals using the CM array is proposed. In [4], a maximum-likelihood approach is taken, also using a structured array manifold. A blind analytic solution based on the analytic CMA [5] demonstrated that good performance can be achieved with a relatively small number of samples. It was also recognized that the underlying CM cost function can also be used for the separation of non-Gaussian signals, and more specifically, finite-alphabet signals [6]. While practical algorithms do exist, the issue of identifiability was still open. Identifiability is an important issue, establishing that the only existing solutions in the noiseless case are the original source signals up to inherent indeterminacies of permutation and phase. Identifiability analysis has been mostly based on the expected value of the CM cost function, so that the results are only valid for infinitely many samples and ergodic scenarios. Not much is known about identifiability based on a *finite* number of samples. Previous results for the special case of harmonic retrieval in one and more dimensions were given in [7] and the references therein. For the separation of a linear mixture of d continuous CM sources, [5] conjectured that about $2d$ samples should be sufficient. The provided argument was

Manuscript received August 10, 2001; revised January 26, 2003. The work of A. Leshem was supported in part by the NOEMI project of the STW under Contract DEL77-4476. The material in this correspondence was presented in part at the IEEE Sensor Arrays and Multichannel Signal Processing 2002 Workshop, Washington DC, August 2002.

A. Leshem is with the School of Engineering, Bar-Ilan University, 52900 Ramat-Gan, Israel. He is also with Metalink Broadband Access, 60972, Yakum, Israel (e-mail: leshem@metalink.co.il).

N. Petrochilos is with the Department of ITS/Electrical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands. He is also with I3S-CNRS, Algorithmes, 06903 Sophia-Antipolis, France (e-mail: petro@ieee.org).

A.-J. van der Veen is with the Department of ITS/Electrical Engineering, Delft University of Technology, 2628 CD Delft, The Netherlands (e-mail: allejan@cas.et.tudelft.nl).

Communicated by V. V. Veeravalli, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2003.815791

unsatisfying and based on counting the number of equations and unknowns, ignoring possible indeterminacies. For binary signals (binary phase-shift keying (BPSK)), a sufficient condition for identifiability in [6] was based on the premise that all 2^{d-1} combinations of constellation points (up to sign) have been received. This means that an average of approximately $(d-1)2^{(d-1)}$ many samples is needed for BPSK signals and much more for higher constellations. Moreover, there is always a nonzero probability that any finite number of samples does not provide identifiability (e.g., if all inputs are identical). The proof in [6] does not generalize to continuous CM sources.

In this correspondence, we give a rigorous proof of identifiability of a mixture of d continuous or discrete complex CM sources, with finitely many samples. We use the linearization technique of [5], together with a simple inductive argument, to show that for continuous CM sources, $d(d-1)+1$ many samples suffice with probability 1 [5]. The analysis of the finite-alphabet case is harder because there is a nonzero probability that sample vectors are repeated. For sufficiently large N , we specify an upper bound on the probability that a data set with N samples is not yet identifiable. The probability decays exponentially.

The structure of the correspondence is as follows. The problem is formulated in Section II. Section III reduces the general problem of more channels than sources to the case of an equal number of channels and sources. A simple result characterizing linear transformations of the torus onto itself is given. Section IV contains the proof of the main theorem, where we discuss the finite-sample case. Identifiability is specified in terms of “persistently exciting signals,” which are studied in Section V. In Section V-B, we provide a simple bound for the finite-alphabet CM sources based on large deviations theory. Then, in Section V-C, we strengthen the result for L -PSK signals. In this case, we obtain that the probability of nonidentifiability decreases exponentially as a function of the number of samples and as L^{-N-1} as a function of the alphabet size for a given number of samples. Finally, in Section VI, we demonstrate some of the results by means of a simulation.

II. THE IDENTIFICATION PROBLEM

Consider an array with p sensors receiving d narrow-band constant modulus signals. Under standard assumptions for the array manifold, we can describe the received signal as an instantaneous linear combination of the source signals

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) \quad (1)$$

where

$\mathbf{x}(n) = [x_1(n), \dots, x_p(n)]^T$ is a $p \times 1$ vector of received signals at discrete time n (T denotes matrix transposition);

$\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_d]$, where \mathbf{a}_i is the array response vector toward the i th signal;

$\mathbf{s}(n) = [s_1(n), \dots, s_d(n)]^T$ is a $d \times 1$ vector of source signals at time n .

We further assume that all sources have constant modulus, i.e., for all n , $|s_i(n)| = 1$ ($i = 1, \dots, d$), and that \mathbf{A} has full column rank (this implies $p \geq d$).

In our problem, the array is assumed to be uncalibrated so that the array response vectors \mathbf{a}_i are unknown. Unequal source powers are absorbed in the mixing matrix. Phase offsets of the sources after demodulation are part of the s_i . Thus, we can write $s_i(n) = e^{j\phi_i(n)}$, where $\phi_i(n)$ is the unknown phase modulation for source i , and we define $\phi(n) = [\phi_1(n), \dots, \phi_d(n)]^T$ as the phase vector for all sources at time n . Note that this leads to the fundamental indeterminacy of phase exchange between a source and the corresponding column in the

mixing matrix. Furthermore, we can permute the sources and simultaneously permute the columns of \mathbf{A} . Thus, \mathbf{A} is determined only up to a permutation of its columns and a complex unit-modulus scaling of each column.

The identifiability problem asks for the number of samples needed in order to ensure (with probability 1) that in the *noiseless* case we have a unique solution up to the above indeterminacies.

III. IDENTIFIABILITY WITH INFINITELY MANY SAMPLES

Let $\mathbb{T} = \{z : |z| = 1\}$ be the complex unit circle, and let \mathbb{T}^d be the Cartesian product of d copies of \mathbb{T} , representing the collection of d -dimensional CM source vectors. Topologically, this collection is a d -dimensional torus embedded in a d -dimensional complex vector space \mathbb{C}^d .

We first characterize linear transformations \mathbf{G} mapping \mathbb{T}^d into itself. Consider the set \mathbb{G}

$$\mathbb{G} = \left\{ \mathbf{G} \in \mathbb{C}^{d \times d} \mid \mathbf{G} \text{ invertible; } \mathbf{s} \in \mathbb{T}^d \Rightarrow \mathbf{G}\mathbf{s} \in \mathbb{T}^d \right\}.$$

Lemma 1: Let $\mathbf{G} \in \mathbb{G}$. Then $\mathbf{G} = \mathbf{P}\mathbf{\Lambda}$, where \mathbf{P} is a permutation matrix and $\mathbf{\Lambda}$ a diagonal matrix with diagonal elements on the unit circle.

Proof: We will prove that each row of \mathbf{G} contains at most one nonzero element with magnitude 1. Let

$$\mathbf{g} = [g_1, \dots, g_d] = [r_1 e^{j\phi_1}, \dots, r_d e^{j\phi_d}]$$

be a row of \mathbf{G} where r_i is the magnitude of g_i . For each $\mathbf{s} \in \mathbb{T}^d$, we know that $|\mathbf{g}\mathbf{s}| = 1$. Choose \mathbf{s}_1 such that $\mathbf{s}_1 = [e^{-j\phi_1}, \dots, e^{-j\phi_d}]^T$. We obtain

$$\mathbf{g}\mathbf{s}_1 = r_1 + \sum_{i>1} r_i = 1 \quad (2)$$

since all r_i are nonnegative real numbers. Similarly, define \mathbf{s}_2 by $(\mathbf{s}_2)_1 = e^{-j\phi_1}$ and $(\mathbf{s}_2)_i = -e^{-j\phi_i}$ for $2 \leq i \leq d$. Then

$$\mathbf{g}\mathbf{s}_2 = r_1 - \sum_{i>1} r_i.$$

Since $|\mathbf{g}\mathbf{s}_2| = 1$ we have either

$$r_1 - \sum_{i>1} r_i = 1 \quad (3)$$

or

$$r_1 - \sum_{i>1} r_i = -1.$$

From (2) and (3), we obtain in the first case that $r_1 = 1$ and $\sum_{i>1} r_i = 0$ whereas in the second case $r_1 = 0$ and $\sum_{i>1} r_i = 1$. Proceeding inductively, we obtain that exactly one element of \mathbf{g} is nonzero with magnitude 1. Since all the rows of \mathbf{G} have this property and \mathbf{G} is invertible, it must be a permutation of a diagonal matrix with unit-modulus diagonal entries. \square

The identifiability theorem for infinite samples follows directly from the preceding lemma.

Theorem 2: Consider an infinite collection of vectors $\mathbf{s}(n) \in \mathbb{T}^d$, $n = 1, \dots, \infty$, and suppose that the collection is dense in \mathbb{T}^d . (For that to hold it is sufficient that $\mathbf{s}(n)$ is an ergodic process, e.g., when $\mathbf{s}(n)$ are independent and identically distributed (i.i.d.) with independent coordinates.) Suppose that we have available the observations $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)$, where $\mathbf{A} \in \mathbb{C}^{p \times d}$ is full column rank d . Then \mathbf{A} is

uniquely determined by the observations, up to a permutation and a unit-modulus complex scaling of the columns.

Proof: Suppose that there is another matrix $\mathbf{B} \in \mathbb{C}^{p \times d}$ and collection of source signals $\mathbf{z}(n) \in \mathbb{T}^d$ which generate the same data $\{\mathbf{x}(n)\}$.

The linear span of the collection $\{\mathbf{s}(n)\}$ is \mathbb{C}^d , so that the linear span of $\{\mathbf{x}(n)\}$ is a d -dimensional subspace in \mathbb{C}^p . Hence, \mathbf{B} is full column rank d . Since its column span must be the same as that of \mathbf{A} , $\mathbf{G} := \mathbf{B}^\dagger \mathbf{A} \in \mathbb{C}^{d \times d}$ is full rank. Moreover, since $\mathbf{s}(n)$ is dense in \mathbb{T}^d , it follows that for any $\mathbf{s} \in \mathbb{T}^d$, $\mathbf{G}\mathbf{s} = \mathbf{z} \in \mathbb{T}^d$. Hence, $\mathbf{G} \in \mathbb{G}$, and Lemma 1 claims that $\mathbf{G} = \mathbf{P}\mathbf{\Lambda}$, so that $\mathbf{A} = \mathbf{B}\mathbf{P}\mathbf{\Lambda}$. \square

The proof of the theorem shows that the infinite collection of vectors $\{\mathbf{s}(n)\}$ is only used to quickly deduce that $\mathbf{G} \in \mathbb{G}$. The question is whether this can be done using a finite set of vectors.

IV. IDENTIFIABILITY WITH FINITELY MANY SAMPLES

In this section, we derive a sufficient condition on the number of samples needed to guarantee identifiability with probability 1, for the case of constant modulus signals with continuous alphabet. Based on the discussion of the previous section we restrict ourselves to invertible linear transformations from \mathbb{T}^d to \mathbb{T}^d .

Consider a collection of N vectors

$$S = \{\mathbf{s}(n) \in \mathbb{T}^d, n = 1, \dots, N\}$$

and let

$$\mathbf{\Psi} = \begin{bmatrix} 1 & s_1(1)s_2^*(1) & s_1^*(1)s_2(1) & \cdots & s_d^*(1)s_{d-1}(1) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & s_1(N)s_2^*(N) & s_1^*(N)s_2(N) & \cdots & s_d^*(N)s_{d-1}(N) \end{bmatrix} \quad (4)$$

where $*$ denotes complex conjugate and $\mathbf{\Psi}$ has size $N \times d(d-1) + 1$. We call S “persistently exciting” if $\mathbf{\Psi}$ has full column rank. Note that this implies that $N \geq d(d-1) + 1$. It also implies that the constellation is complex (for BPSK constellations, columns of $\mathbf{\Psi}$ are repeated and a modified definition can be introduced).

Lemma 3: Let $N \geq d(d-1) + 1$, and let $\mathbf{s}(n) \in \mathbb{T}^d$, $n = 1, \dots, N$, be a persistently exciting collection. Consider an invertible linear transformation $\mathbf{G} \in \mathbb{C}^{d \times d}$ such that $\mathbf{G}\mathbf{s}(n) \in \mathbb{T}^d$, for $n = 1, \dots, N$. Then $\mathbf{G} = \mathbf{P}\mathbf{\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix with unit norm diagonal entries and \mathbf{P} is a permutation matrix.

Proof: Since \mathbf{G} is invertible, it is sufficient to prove that each row \mathbf{g} of \mathbf{G} contains exactly one nonzero element which is unit norm. Let $\mathbf{g} = [g_1, \dots, g_d]$, and let $y(n) = \mathbf{g}\mathbf{s}(n)$ be the corresponding entry of $\mathbf{G}\mathbf{s}(n)$.

Then for each $n \in \{1, \dots, N\}$, we have

$$\begin{aligned} y(n) &= \sum_{i=1}^d g_i s_i(n) \\ \Rightarrow |y(n)|^2 &= \left| \sum_{i=1}^d g_i s_i(n) \right|^2 \\ \Rightarrow 1 &= \sum_{1 \leq i, j \leq d} g_i g_j^* s_i(n) s_j^*(n). \end{aligned} \quad (5)$$

Denote $P_{ij} = g_i g_j^*$ and $P_T = \sum_{i=1}^d P_{ii}$. By linearizing (5) and considering all n , we obtain (as in [5])

$$\mathbf{\Psi}\mathbf{p} = \mathbf{1} \quad (6)$$

where

$$\begin{aligned} \mathbf{p} &= [P_T, P_{12}, P_{21}, \dots, P_{d,d-1}]^T \\ \mathbf{1} &= [1, \dots, 1]^T \end{aligned}$$

and Ψ is as in (4). A particular solution of (6) for \mathbf{p} is given by

$$\begin{cases} P_T = 1 \\ P_{ij} = 0, \quad \forall i \neq j. \end{cases} \quad (7)$$

Suppose that $g_j \neq 0$ for some j , then since $P_{ij} = g_i g_j^*$ we immediately obtain that $g_i = 0$ for all $i \neq j$. Since $P_T = 1$, $|g_j|^2 = 1$. Hence, each row \mathbf{g} of \mathbf{G} has precisely one nonnull entry, which is unit-modulus. It follows that $\mathbf{G} = \mathbf{P}\mathbf{A}$.

Since Ψ is full column rank, this is the only solution to (7). \square

Combining with Theorem 2 we obtain the following.

Theorem 4: Identifiability as in Theorem 2 already holds for a finite collection of source signals $\mathbf{s}(n)$, $n = 1, \dots, N$, where $N \geq d(d-1) + 1$ if this collection is persistently exciting.

V. PERSISTENCE OF EXCITATION

The remaining issue is to establish when a collection of vectors in \mathbb{T}^d is persistently exciting. As usual, this is hard to characterize in a deterministic setting. In a stochastic sense, any "sufficiently random" collection of $N \geq d(d-1) + 1$ complex vectors in \mathbb{T}^d is expected to be persistently exciting. Although this appears a reasonable argument, the interrelations of the elements of Ψ make it not completely evident that this is the case. Moreover, in the case of discrete-alphabet CM sources, e.g., quaternary phase-shift keying (QPSK), proofs are harder because pathological cases appear with positive probability. We first make a more explicit statement for continuous CM sources, and then consider the discrete-alphabet CM case.

A. The Continuous-Alphabet Case

Lemma 5: Let $\mathbf{s}(n)$, for $n = 1, \dots, N$, be a collection of continuous-alphabet i.i.d. complex vectors in \mathbb{T}^d with stochastically independent components. If $N \geq d(d-1) + 1$, then the matrix Ψ has full column rank with probability 1 (relative to the Lebesgue measure on \mathbb{T}^d !).

Proof: Given $N \geq d(d-1) + 1$ samples of $\mathbf{s}(n)$, assume toward contradiction that there exists a vector $\boldsymbol{\alpha} \neq \mathbf{0}$ such that $\Psi\boldsymbol{\alpha} = \mathbf{0}$, or equivalently, $\exists\{\alpha_0, \alpha_{ij}, 1 \leq i \neq j \leq d\}$, not all zeros such that for every $n = 1, \dots, N$, the next equation holds

$$\alpha_0 + \sum_{i \neq j} \alpha_{ij} s_i(n) s_j^*(n) = 0. \quad (8)$$

After multiplying every equation by $s_1^*(n)$, this becomes, for all $n = 1, \dots, N$

$$\sum_{j > 1} \alpha_{1j} s_j^*(n) + s_1^*(n) \left(\alpha_0 + \sum_{i \neq j > 1} \alpha_{ij} s_i(n) s_j^*(n) \right) + (s_1^*(n))^2 \sum_{i > 1} \alpha_{i1} s_i(n) = 0.$$

After taking the conjugate of this expression, we see that it is a set of N independent quadratic equations in $s_1(n)$

$$a(n) + b(n)s_1(n) + c(n)s_1^2(n) = 0.$$

Hence, one of the following holds: a) $s_1(n)$ is a function of $(s_2(n), \dots, s_d(n))$, which contradicts the independence assumption, and for $s_1(n)$ in a continuous alphabet is a zero-measure event, or b) the coefficients satisfy

$$a(n) = b(n) = c(n) = 0, \quad \forall n \in \{1, \dots, N\}.$$

Hence,

$$1) \quad a(n) = 0 \Rightarrow \sum_{j > 1} \alpha_{1j} s_j^*(n) = 0.$$

For each n , this is a linear condition on the $d-1$ coefficients $\{\alpha_{1j}\}$. Using $d-1$ independent samples suffices to derive that $\forall i, \alpha_{1i} = 0$.

2) Similarly, from the condition $c(n) = 0$ and using $d-1$ other independent samples, we obtain that $\forall i, \alpha_{i1} = 0$.

$$3) \quad b(n) = 0 \Rightarrow \alpha_0 + \sum_{i \neq j > 1} \alpha_{ij} s_i(n) s_j^*(n) = 0.$$

This condition is verified by applying inductively the same argument on $b(n)$ as we did on (8).

We thus obtain that all α_{ij} are equal to zero, and $\alpha_0 = 0$. Therefore, Ψ is full rank with probability one.

If $d = 1$, then we trivially need one sample to conclude that Ψ is full rank. Hence, the recursive application of the argument needs

$$2(d-1) + \dots + 2 + 1 = d(d-1) + 1$$

independent samples, and this number is sufficient with probability 1. \square

B. The Finite-Alphabet Case—Large Deviations Bound

For discrete-alphabet sources, we have to use a different approach. Let $\mathbf{s}(n)$, for $n = 1, \dots, N$, be a collection of zero mean i.i.d. complex vectors in \mathbb{T}^d with stochastically independent and circularly symmetric components, or more explicitly

$$\begin{aligned} \mathbb{E}(|s_i|^2) &= 1 \\ \mathbb{E}(s_i^2) &= 0 \\ \mathbb{E}(s_i s_j^*) &= 0, \quad i \neq j \\ \mathbb{E}(s_i^2 s_j^{*2}) &= 0, \quad i \neq j \\ \mathbb{E}(s_i s_j s_k^{*2}) &= 0, \quad i \neq j \neq k \\ \mathbb{E}(s_i s_j s_k^* s_l^*) &= 0, \quad i \neq j \neq k \neq l. \end{aligned} \quad (9)$$

Denote a generic n th row of Ψ by

$$\mathbf{v}(n) = [1, \quad s_1(n)s_2^*(n), \quad s_2(n)s_1^*(n), \dots]. \quad (10)$$

Then (omitting the index n) we have

$$\mathbf{v}^H(n)\mathbf{v}(n) = \begin{bmatrix} 1 & s_1 s_2^* & s_2 s_1^* & \dots \\ s_2 s_1^* & 1 & s_2^2 s_1^{*2} & \dots \\ s_1 s_2^* & s_1^2 s_2^{*2} & 1 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}.$$

With the assumptions (9), it follows that $\mathbb{E}(\mathbf{v}^H \mathbf{v}) = \mathbf{I}$. Note that $\frac{1}{N} \Psi^H \Psi \rightarrow \mathbb{E}(\mathbf{v}^H \mathbf{v})$ as $N \rightarrow \infty$. Hence, for sufficiently large N , Ψ must have full column rank. For continuous CM sources we already proved that $N \geq d(d-1) + 1$ is sufficient with probability 1. For discrete-alphabet sources it can happen that the same constellation vector is received multiple times and, hence, N might have to be larger.

We next quantify the probability that N samples of the array output are sufficient. We first provide a simple proof which gives subexponentially decreasing probability of nonidentifiability. Subsequently, in the next subsection, we provide a more accurate (but also more complex) analysis providing an exponentially decreasing upper bound on the probability of nonidentifiability. Let

$$\hat{\mathbf{R}}_N = \frac{1}{N} \Psi^H \Psi = \frac{1}{N} \sum_{n=1}^N \mathbf{v}(n)^H \mathbf{v}(n).$$

As we have shown, $\mathbb{E}(\hat{\mathbf{R}}_N) = \mathbf{I}$. We now analyze the rate of convergence of $\hat{\mathbf{R}}_N$ to \mathbf{I} and provide an upper bound on the probability that $\hat{\mathbf{R}}_N$ is singular. To that end we use the following consequence of Gershgorin's theorem.

Theorem 6 [8, p. 349]: Let $\mathbf{A} = [a_{ij}]$ be a Hermitian matrix. Assume that for all i , $|a_{ii}| > 0$, and that \mathbf{A} is diagonally dominated, i.e., for all i

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}|$$

then \mathbf{A} is strictly positive definite.

Assume that all off-diagonal elements of $\hat{\mathbf{R}}_N$ have magnitude less than $\frac{1}{d(d-1)+1}$. Then for all i

$$\sum_{j \neq i} (\hat{\mathbf{R}}_N)_{ij} < \frac{d(d-1)}{d(d-1)+1} < (\hat{\mathbf{R}}_N)_{ii} = 1 \quad (11)$$

and by Theorem 6 we can conclude that $\hat{\mathbf{R}}_N$ is strictly positive definite. It remains to compute a bound on the probability that all off-diagonal elements of $\hat{\mathbf{R}}_N$ have magnitude less than $\frac{1}{d(d-1)+1}$. This will provide a lower bound on the probability of persistence of excitation since as discussed above, if $\hat{\mathbf{R}}_N$ is nonsingular then Ψ is full rank.

To obtain the bound we use large deviation theory. The following theorem from Feller is instrumental in our analysis.

Theorem 7 [9]: Let x_n be a sequence of zero-mean i.i.d. random variables with variance σ^2 . Let

$$S_N = \frac{1}{\sigma\sqrt{N}} \sum_{n=1}^N x_n.$$

Assume that $t = o(\sqrt{N})$. Then for all $\varepsilon > 0$ and N sufficiently large

$$P(S_N > t) < e^{-\frac{1}{2}(1-\varepsilon)t^2}.$$

We will need the following results on the variances of products of independent complex circularly symmetric CM signals, which are not hard to derive [10]:

$$\begin{aligned} \text{var}(s_i s_j^*) &= \begin{cases} 0, & i = j \\ 1, & i \neq j \end{cases} \\ \text{var}(s_i s_j^* s_k s_l^*) &= \begin{cases} 0, & (i = j \wedge k = l) \text{ or } (i = l \wedge j = k) \\ 1, & \text{otherwise.} \end{cases} \end{aligned} \quad (12)$$

Using Theorem 7 we now prove the following lemma.

Lemma 8: For every $i \neq j$, for all $\varepsilon > 0$, and N sufficiently large

$$P\left((\hat{\mathbf{R}}_N)_{ij} > \frac{1}{d(d-1)+1}\right) < 2e^{-\frac{1}{2}(1-\varepsilon)\frac{N}{(d(d-1)+1)^2 \log \log N}}.$$

Proof: Let $\mathbf{v}(n)$ be defined as in (10). Fix $i \neq j$ and let $x_n = \mathbf{v}_i(n)\mathbf{v}_j(n)^*$. Using (12), x_n is a complex random variable with zero mean and variance $\sigma^2 = 1$. By definition

$$(\hat{\mathbf{R}}_N)_{ij} = \frac{1}{N} \sum_{n=1}^N x_n.$$

We would like to bound $P\left(\left|(\hat{\mathbf{R}}_N)_{ij}\right| > \frac{1}{d(d-1)+1}\right)$. Note that

$$\begin{aligned} P\left(\left|(\hat{\mathbf{R}}_N)_{ij}\right| > \frac{1}{d(d-1)+1}\right) &= P\left(\frac{1}{N} \left|\sum x_n\right| > \frac{1}{d(d-1)+1}\right) \\ &= P\left(\frac{1}{\sqrt{N}} \left|\sum x_n\right| > \frac{\sqrt{N}}{d(d-1)+1}\right) \\ &\leq P\left(\frac{1}{\sqrt{N}} \left|\sum x_n\right| > t\right) \end{aligned}$$

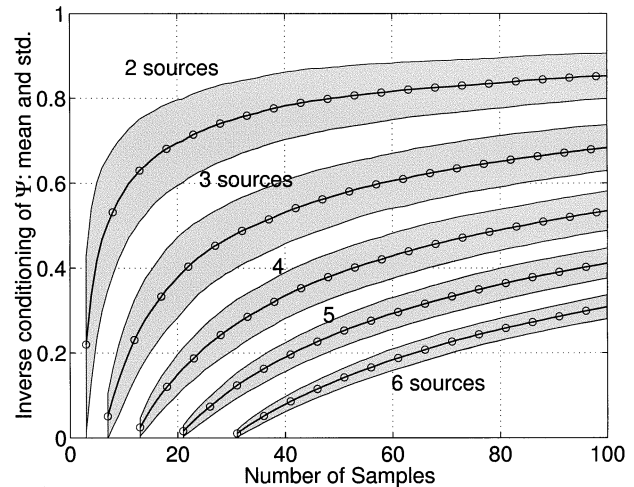


Fig. 1. (circles) Mean of the inverse condition number of Ψ , and (shaded area) its $\pm 1\sigma$ standard deviation interval for the continuous CM sources.

where

$$t = \sqrt{\frac{N}{(d(d-1)+1)^2 \log \log N}}$$

and $\log \log N$ is introduced to have $t = o(\sqrt{N})$. Applying Theorem 7 separately to the real and imaginary parts of $\hat{\mathbf{R}}_{ij}$ we obtain

$$P\left((\hat{\mathbf{R}}_N)_{ij} > \frac{1}{d(d-1)+1}\right) < 2e^{-\frac{1}{2}(1-\varepsilon)\frac{N}{(d(d-1)+1)^2 \log \log N}}$$

as claimed. \square

We now use Lemma 8 to bound the probability that $\hat{\mathbf{R}}_N$ is nonsingular. Note that since $\hat{\mathbf{R}}_N$ is Hermitian it is sufficient to obtain that all entries above the diagonal are sufficiently small. There are $\frac{1}{2}[d(d-1)+1][d(d-1)]$ entries, and since most entries are uncorrelated (although not independent) we obtain that the probability that all entries are smaller than $1/d(d-1)$, for N sufficiently large, is bounded by (but not equal to)

$$\left(1 - 2e^{-\frac{1}{2}(1-\varepsilon)\frac{N}{(d(d-1)+1)^2 \log \log N}}\right)^{\frac{1}{2}(d(d-1)+1)(d(d-1))}. \quad (13)$$

Since for any x such that $0 < x < 1$ we have $(1-x)^n > 1-nx$, we can bound (13) by

$$1 - d^4 e^{-\frac{1}{2}(1-\varepsilon)\frac{N}{(d(d-1)+1)^2 \log \log N}}. \quad (14)$$

In summary, the probability of having a data set that is not persistently exciting is asymptotically less than

$$d^4 e^{-\frac{1}{2}(1-\varepsilon)\frac{N}{(d(d-1)+1)^2 \log \log N}}$$

(for any $\varepsilon > 0$).

C. Chernoff Bound and Finite-Alphabet CM Signals

We now provide a more accurate bounding using the Chernoff bound on finite-alphabet L -PSK signals. This bound holds for all values of N . Furthermore, it also shows that for any fixed $N > d(d-1)$, increasing the alphabet size L decreases the probability of nonidentifiability at least as $\frac{1}{L^{N-1}}$. Our goal is to bound

$$P\left(\frac{1}{N} \sum_{n=1}^N \mathbf{v}_i(n)^H \mathbf{v}_j(n) > \frac{1}{d(d+1)}\right) \quad (15)$$

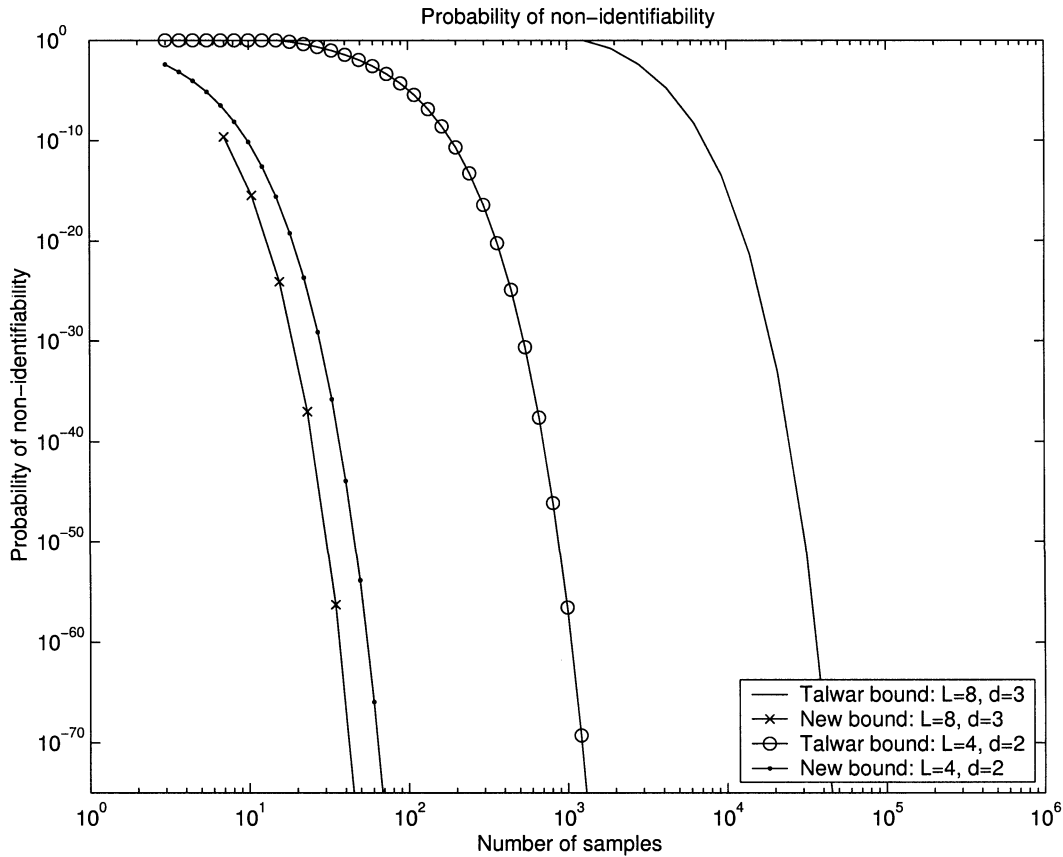


Fig. 2. Finite-alphabet CM sources: Upper bound on the probability that N source samples are not persistently exciting. \mathbf{d} sources and L -PSK constellations.

where $\mathbf{v}_i(n)$, $\mathbf{v}_j(n)$ are the i th and j th rows of Ψ as defined in (10). To that end, fix $i \neq j$. Let $x_n = \mathbf{v}_i(n)^H \mathbf{v}_j(n)$. For every n , x_n is uniformly distributed over the L th-order roots of unity (the roots of unity form a multiplicative group and the convolution of a uniform distribution on the group with any other distribution is uniform). Since in practical applications L is always even (and actually a power of 2), let $L = 2K$. Using the fact that L is even, we obtain that if a is a symbol also $-a$ is a symbol. Let the alphabet be

$$\mathbb{A} = \{a_1, -a_1, a_2, -a_2, \dots, a_K, -a_K\}.$$

We now have that

$$\sum_{n=1}^N x_n = \sum_{i=1}^K (n_i a_i - n_{-i} a_i) \quad (16)$$

where n_i is the number of occurrences of a_i and n_{-i} is the number of occurrences of $-a_i$, among x_1, \dots, x_N . Therefore, we can bound

$$\begin{aligned} P\left(\frac{1}{N} \left| \sum_{n=1}^N x_n \right| > k\right) &= P\left(\frac{1}{N} \left| \sum_{i=1}^K (n_i - n_{-i}) a_i \right| > k\right) \\ &\leq P\left(\frac{1}{N} \sum_{i=1}^K |n_i - n_{-i}| > k\right) \\ &= P\left(\sum_{i=1}^K |n_i - n_{-i}| > kN\right). \end{aligned} \quad (17)$$

Using the uniformity of the distribution we obtain that the preceding equation becomes

$$P\left(\frac{1}{N} \left| \sum_{n=1}^N x_n \right| > k\right) \leq \sum_{i=1}^K P(|n_i - n_{-i}| > k'N)$$

$$\begin{aligned} &\leq KP(|n_1 - n_{-1}| > k'N) \\ &= 2KP(n_1 - n_{-1} > k'N). \end{aligned} \quad (18)$$

where $k' = \frac{k}{K} = \frac{2k}{L}$. The inequality uses the fact that there must be at least one element greater than or equal to the mean, and the last inequality uses symmetry of the distribution. We now finish the bounding using the Chernoff bound [11]. Define a sequence of i.i.d. random variables y_i with distribution

$$y_i = \begin{cases} 1, & \text{with probability } \frac{1}{L} \\ -1, & \text{with probability } \frac{1}{L} \\ 0, & \text{with probability } \frac{L-2}{L} \end{cases}. \quad (19)$$

Then, for any $\nu \geq 0$, and for any k

$$P\left(\sum_{i=1}^N y_i > kN\right) \leq E e^{\nu(\sum_{i=1}^N y_i - kN)} = e^{-\nu kN} (E(e^{\nu y_i}))^N. \quad (20)$$

The parameter ν is used to obtain a tighter fit of the inequality. Using the distribution of y_i (19) we obtain

$$E(e^{\nu y_i}) = \frac{2}{L} \sinh(\nu). \quad (21)$$

Optimizing ν (see the Appendix) we obtain

$$\nu = \tanh^{-1}(k). \quad (22)$$

Substituting into (20) and simplifying we obtain

$$P\left(\sum_{i=1}^N y_i > kN\right) \leq \left(\frac{2}{L}\right)^N e^{-k \tanh^{-1}(k)N} \left(\frac{k}{\sqrt{1-k^2}}\right)^N. \quad (23)$$

Substituting $k' = \frac{2}{Ld(d+1)}$ and using (18) we obtain

$$P\left(\frac{1}{N} \sum_{n=1}^N \mathbf{v}_i^H(n) \mathbf{v}_j(n) < \frac{1}{d(d+1)}\right) \leq 2 \left(\frac{2}{L}\right)^{N-1} e^{-\frac{2N}{Ld(d+1)} \tanh^{-1}\left(\frac{2}{Ld(d+1)}\right)} \left(\frac{2}{L\sqrt{d^2(d+1)^2 - \frac{2}{L}}}\right)^N. \quad (24)$$

Similarly to (14), we now obtain a bound: the probability of identifiability of d sources using n vector samples taken from L -PSK i.i.d. source $P_{i,d}(L, d, N)$ satisfies the following inequality:

$$P_{i,d}(L, N, d) \geq 1 - 2d^d \left(\frac{2}{L}\right)^{N-1} e^{-\frac{2N}{Ld(d+1)} \tanh^{-1}\left(\frac{2}{Ld(d+1)}\right)} \left(\frac{2}{L\sqrt{d^2(d+1)^2 - \frac{2}{L}}}\right)^N. \quad (25)$$

This is better than the large deviation bound (14) since the dependence on N is exponential and not subexponential and is also valid for all values of N . Moreover, we can see that as the alphabet size is increased, the probability of nonidentifiability approaches 0 as $L^{-(N-1)}$.

VI. SIMULATIONS

To demonstrate the results of Lemma 5 for continuous CM sources, as well as the relevance of the result in the presence of noise we simulated 1000 independent runs with $N = 1, \dots, 100$ samples and $d = 2, \dots, 6$ sources. For each number of samples, we computed the rank and the conditioning of the Ψ matrix. For $N > d(d-1)$, the rank of Ψ was always equal to $d(d-1) + 1$, as predicted by Lemma 5.

Fig. 1 shows the inverse of the condition number of Ψ , for varying d and N . Interestingly, it is seen that Ψ can be rather ill-conditioned when N is close to its lower limit. Indeed, the lemma did not ensure the conditioning for Ψ . Nevertheless, with a few samples above the lower limit, this adverse situation improves significantly. In the limit, the conditioning converges to 1. This has important implication on the performance of least squares CMA (LS-CMA) algorithms, where a finite amount of data is reused. A further analysis of this problem appeared in [12], where the connection between the local minima of the CMA cost function and the local minima of its finite sample approximation was considered.

We also illustrate a comparison of the new upper bound on failure of identifiability (14) to the bound by Talwar [6], see Fig. 2. We can clearly see that the new bound is much better with orders of magnitudes less samples necessary for a given probability of identifiability.

VII. CONCLUSION

We presented a rigorous proof of a sufficient condition for the identifiability of mixtures of CM signals, based on finitely many samples. For continuous-CM sources, $N = d(d-1) + 1$ samples are sufficient with probability 1. For finite-alphabet cases, only an upper bound on the probability of nonidentifiability given alphabet size, number of sources, and number of samples could be derived. However, the new bound is much tighter than previously known bound and shows that probability of nonidentifiability goes down exponentially with sample size and polynomially with alphabet size.

APPENDIX

OPTIMIZING ν IN THE CHERNOFF BOUND

The optimal value of ν satisfies the equation (see, e.g., [11, p. 54])

$$E y_i e^{\nu y_i} - k e^{\nu y} = 0.$$

Using (19) we obtain

$$\frac{1}{L} (e^\nu - e^{-\nu}) - \frac{k}{L} (e^\nu + e^{-\nu}) = 0$$

simplifying we obtain

$$\tanh(\nu) = k.$$

Hence, the optimal ν is given by

$$\nu = \tanh^{-1}(k).$$

Using the equality

$$\cosh^2(\nu) - \sinh^2(\nu) = 1$$

we now obtain

$$k^2 = \frac{\sinh^2(\nu)}{1 + \sinh^2(\nu)}$$

and, therefore,

$$\sinh(\nu) = \frac{k}{\sqrt{1 - k^2}}.$$

ACKNOWLEDGMENT

The authors would like to thank the anonymous referees and the Associate Editor for asking for proof of the finite alphabet case, and Simon Litsyn for helpful discussion on Chernoff's bound.

REFERENCES

- [1] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COM-28, pp. 1867–1875, Nov. 1980.
- [2] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 459–471, Apr. 1983.
- [3] J. Shynk and R. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. Signal Processing*, vol. 44, pp. 652–660, Mar. 1996.
- [4] A. Leshem, "Maximum likelihood separation of constant modulus signals," *IEEE Trans. Signal Processing*, vol. 48, pp. 2948–2952, Oct. 2000.
- [5] A. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.
- [6] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. I. algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.
- [7] T. Jiang, N. Sidiropoulos, and J. ten Berge, "Almost-sure identifiability of multidimensional harmonic retrieval," *IEEE Trans. Signal Processing*, vol. 49, pp. 1849–1859, Sept. 2001.
- [8] R. Horn and C. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1994.
- [9] W. Feller, *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: Wiley, 1971.
- [10] A. van der Veen, "Statistical performance analysis of the algebraic constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 50, pp. 3083–3097, Dec. 2002.
- [11] J. Proakis, *Digital Communications*, 3rd ed. New York: McGraw-Hill, 1995.
- [12] A. Leshem and A.-J. van der Veen, "On the finite sample behavior of the constant modulus cost function," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP) 2000*, vol. 5, Istanbul, Turkey, June 2000, pp. 2537–2540.