ON THE NUMBER OF SAMPLES NEEDED TO IDENTIFY A MIXTURE OF FINITE ALPHABET CONSTANT MODULUS SOURCES

Amir Leshem[1,2,3] *and Alle-Jan van der Veen*[1]

[1]Delft University of Technology, Dept. ITS/Electrical Eng., 2628 CD Delft, The Netherlands
[2]Metalink Broadband Access, Yakum Business Park, 60972, Yakum, Israel
[3] School of Engineering, Bar Ilan University, 52900, Ramat-Gan, Israel
leshem@cas.et.tudelft.nl, allejan@cas.et.tudelft.nl

Constant-modulus algorithms try to separate linear mixtures of sources with modulus 1. We study the identifiability of this problem: how many samples are needed to ensure that in the noiseless case we have a unique solution? For finite-alphabet (L-PSK) sources, finite sample identifiability can hold only with a probability close to but not equal to 1. In a previous paper, we provided a sub-exponentialy decaying upper bound on the probability of non-identifiability. Here, we provide an improved exponentialy decaying upper bound, based on Chernoff bounds. We show that under practical assumptions, this upper bound is much tighter than previously known bounds.

1. INTRODUCTION

The constant modulus algorithm (CMA) [1] is very popular for blind equalization and for source separation of multiple constant modulus (CM) signals using antenna arrays. It was soon recognized that the underlying CM cost function can be used for the separation of non-Gaussian signals as well, and more specifically for finite alphabet signals. Although many CMAs are implemented as adaptive LMS-type algorithms, block algorithms such as the blind analytic algorithm 'ACMA' [2] demonstrate that good performance can be achieved with already a relatively small number of samples. With sufficiently good initialization, the same is seen in the block-iterative finite alphabet algorithms ILSP and ILSE [3].

While practical algorithms do exist, the issue of identifiability is still relevant. Identifiability is an important issue, establishing that the only 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.

For the separation of a linear mixture of d continuous CM sources, [2] conjectured that about $2d$ samples should be sufficient. The provided argument was unsatisfying and based on counting the number of equations and unknowns, ignoring possible indeterminacies. For binary signals (BPSK), a sufficient condition for identifiability in [3] 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 [3] does not generalize to continuous CM sources.

In this paper we give a rigorous proof of identifiability of a mixture of d discrete alphabet complex CM sources, with finitely many samples. First, we use the linearization technique of [2], together with a simple inductive argument, to show that for continuous CM sources, $d(d-1)+1$ many samples suffice with probability 1. 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. In a previous submission [4], we showed a sub-exponentialy decaying upper bound on the probability of non-identifiability. Here, we provide an improved exponentialy decaying upper bound.

2. PROBLEM DEFINITION

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,

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

where

- $\mathbf{x}(n) = [x_1(n), \cdots, x_p(n)]^T$ is a $p \times 1$ vector of received signals at discrete time n (^{T} denotes matrix transposition),
- $\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_d]$, where $\mathbf{a}_i$ is the array response vector towards the i-th signal,
- $\mathbf{s}(n) = [s_1(n), \cdots, s_d(n)]^T$ 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, $|s_i(n)| = 1$ $(i = 1, \cdots, d)$, 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 $\phi_i(n)$

Amir Leshem was partially supported by the NOEMI project of the STW under contract no. DEL77-4476.

is the unknown phase modulation for source i, and we define $\phi(n) = [\phi_1(n), \cdots, \phi_d(n)]^T$ 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.

3. 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 $\mathcal{C}^d$.

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

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

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

The proof of the lemma appears in [4]. Based on this, we have the following identifiability theorem for an *infinite* number of samples:

Theorem 2 *Consider an infinite collection of vectors $\mathbf{s}(n) \in \mathbb{T}^d$, $n = 1, \cdots, \infty$, and suppose that the collection is dense in $\mathbb{T}^d$. Suppose that we have available the observations $\mathbf{x}(n) = \mathbf{As}(n)$, 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.*

The proof of the theorem is in [4] and starts from the premise that if there exists another matrix $\mathbf{A}'$, then $\mathbf{G} := \mathbf{A}'^{\dagger}\mathbf{A}$ is such that $\mathbf{s}' = \mathbf{Gs} \in \mathbb{T}$. Since this holds for an infinite collection of vectors $\{\mathbf{s}(n)\}$, it follows that $\mathbf{G} \in \mathbb{G}$, and Lemma 1 gives the result.

The question is whether the same can be proved using a finite set of vectors.

4. IDENTIFIABILITY WITH FINITELY MANY SAMPLES

In this section, we derive a sufficient condition on the number of samples needed to guarantee identifiability. Based on the discussion of the previous section we can restrict ourselves to invertible linear transformations from $\mathbb{T}^d$ to $\mathbb{T}^d$.

Consider a collection of N vectors $\mathcal{S} = \{\mathbf{s}(n) \in \mathbb{T}^d, n = 1, \dots, N\}$, and let

$$\mathbf{\Psi} = \begin{bmatrix} 1 & s_1(1)s_2^*(1) & s_1^*(1)s_2(1) & \cdots & s_d^*(1)s_{d-1}(1) \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & s_1(N)s_2^*(N) & s_1^*(N)s_2(N) & \cdots & s_d^*(N)s_{d-1}(N) \end{bmatrix} \quad (2)$$

where $*$ denotes complex conjugate and $\mathbf{\Psi}$ has size $N \times d(d-1)+1$. Note that the n-th row of $\mathbf{\Psi}$ contains all cross products of the signals at time n, taking into account the fact that we already know that $|s_i|^2 = 1$ for all i.

We call $\mathcal{S}$ *"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 $\mathcal{S} = \{\mathbf{s}(n) \in \mathbb{T}^d,\ n = 1, \dots, N\}$ be a persistently exciting collection. Consider an invertible linear transformation $\mathbf{G} \in \mathbb{C}^{d\times d}$ such that $\mathbf{Gs}(n) \in \mathbb{T}^d$, for $n = 1, \dots, N$. Then $\mathbf{G} = \mathbf{P\Lambda}$, where $\mathbf{\Lambda}$ is a diagonal matrix with unit norm diagonal entries $\mathbf{\Lambda}$ and $\mathbf{P}$ is a permutation matrix.*

The proof is again in [4] and makes use of ideas in ACMA [2] to show that $\mathbf{Gs} \in \mathbb{T}^d \Rightarrow \mathbf{\Psi p} = \mathbf{1}$, with $\mathbf{p} = [1, 0, \cdots, 0]^T$ as an obvious solution. Persistence of excitation, by definition, implies that $\mathbf{p}$ is unique, which allows to derive that $\mathbf{G} = \mathbf{P\Lambda}$.

Combining with theorem 2 we obtain

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

5. 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 inter-relations of the elements of $\mathbf{\Psi}$ make it not completely evident that this is the case. Moreover, in the case of discrete alphabet CM sources, e.g. QPSK, proofs are harder because pathological cases appear with positive probability.

5.1. Large deviations bound for arbitrary discrete alphabets

Let $\mathbf{s}(n)$, for $n = 1, \dots, N$, be a collection of zero mean independent identically distributed complex vectors in $\mathbb{T}^d$ with stochastically independent and circularly symmetric components, or more explicitly,

$$\begin{array}{lll} \mathrm{E}(|s_i|^2) & = 1, & \\ \mathrm{E}(s_i^2) & = 0, & \\ \mathrm{E}(s_i s_j^*) & = 0, & i \neq j, \\ \mathrm{E}(s_i^2 s_j^{*2}) & = 0, & i \neq j, \\ \mathrm{E}(s_i s_j s_k^{*2}) & = 0, & i \neq j \neq k, \\ \mathrm{E}(s_i s_j s_k^* s_l^*) & = 0, & i \neq j \neq k \neq l. \end{array} \quad (3)$$

Denote a generic n'th row of $\mathbf{\Psi}$ by

$$\mathbf{v}(n) = [1, s_1(n)s_2^*(n), s_2(n)s_1^*(n), \cdots]. \quad (4)$$

Then (omitting the index n) we have

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

With the assumptions (3), it follows that $\mathrm{E}(\mathbf{v}^H\mathbf{v}) = \mathbf{I}$. Note that $\frac{1}{N}\mathbf{\Psi}^H\mathbf{\Psi} \to \mathrm{E}(\mathbf{v}^H\mathbf{v})$ as $N \to \infty$. Hence for sufficiently large N, $\mathbf{\Psi}$ must have full column rank. Following this argument, we conclude that, for continuous and "sufficiently random" CM sources, $N \geq d(d-1)+1$ is sufficient w.p. 1. However, 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 argument which leads to a sub-exponentially decreasing upper bound on the probability of non-identifiability. Subsequently, in the next subsection, we provide a more accurate (but also more complex) analysis providing an exponentially decreasing upper bound.

Let

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

As we have shown $\mathrm{E}(\hat{\mathbf{R}}_N) = \mathbf{I}$. 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 5 ([5, 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

$$\sum_{j\neq i}(\hat{\mathbf{R}}_N)_{ij} \;<\; \frac{d(d-1)}{d(d-1)+1} \;<\; (\hat{\mathbf{R}}_N)_{ii} \;=\; 1 \qquad (5)$$

and by theorem 5 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 non-singular then $\mathbf{\Psi}$ is full rank.

To obtain the first bound we use large deviation theory as in [6]. This leads to the following result:

Theorem 6 *The probability of having a data set that is not persistently exciting is asymptotically less than*

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

(for any $\epsilon > 0$).

The proof of the theorem is omitted but will appear in [7]. Note that this is an asymptotic result (large N), and that the dependence on N is sub-exponential.

5.2. 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 non-identifiability at least as $\frac{1}{L^{N-1}}$. Our goal is to bound

$$P\big((\hat{\mathbf{R}}_N)_{ij} > \frac{1}{d(d+1)}\big) \equiv P\big(\frac{1}{N}\sum_{n=1}^{N} v_i^*(n)v_j(n) > \frac{1}{d(d+1)}\big) \qquad (6)$$

($i \neq j$) where $v_i(n), v_j(n)$ are the i'th and j'th entry of $\mathbf{v}(n)$, the n-th row of $\mathbf{\Psi}$ as defined in (4). To that end fix $i \neq j$. Let $x_n = v_i(n)^* v_j(n)$. 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

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

We now have that

$$\sum_{n=1}^{N} x_n = \sum_{i=1}^{K}(n_i a_i - n_{-i}a_i) \qquad (7)$$

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

$$\begin{aligned} P\left(\tfrac{1}{N}\left|\textstyle\sum_{n=1}^{N} x_n\right| > k\right) &= P\left(\tfrac{1}{N}\left|\textstyle\sum_{i=1}^{K}(n_i - n_{-i})a_i\right| > k\right) \\ &\leq P\left(\tfrac{1}{N}\textstyle\sum_{i=1}^{K}|n_i - n_{-i}| > k\right) \\ &= P\left(\textstyle\sum_{i=1}^{K}|n_i - n_{-i}| > kN\right). \end{aligned} \qquad (8)$$

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

$$\begin{aligned} P\left(\tfrac{1}{N}\left|\textstyle\sum_{n=1}^{N} x_n\right| > k\right) &\leq \textstyle\sum_{i=1}^{K} P\left(|n_i - n_{-i}| > kN\right) \\ &\leq KP\left(|n_1 - n_{-1}| > kN\right) \\ &= 2KP(n_1 - n_{-1} > kN) \\ &= LP(n_1 - n_{-1} > kN) \end{aligned} \qquad (9)$$

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 symmetricity of the distribution. We now finish the bounding using the Chernoff bound [8]. Define a sequence of i.i.d. random variables y_i with distribution

$$y_i = \begin{cases} 1 & \text{with probability } \frac{1}{L} \\ -1 & \text{with probability } \frac{1}{L} \\ 0 & \text{with probability } \frac{L-2}{L} \end{cases} \qquad (10)$$

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

$$P\left(\sum_{i=1}^{N} y_i > kN\right) \leq Ee^{\nu\left(\sum_{i=1}^{N} y_i - kN\right)} = e^{-\nu kN}\left(E\left(e^{\nu y_i}\right)\right)^N \qquad (11)$$

The parameter ν is used to obtain a tighter fit of the inequality. Using the distribution of y_i (10) we obtain

$$E\left(e^{\nu y_1}\right) = \frac{2}{L}\sinh(\nu). \qquad (12)$$

Optimizing ν (see appendix) we obtain that the best choice is

$$\nu = \tanh^{-1}(k). \qquad (13)$$

Substituting into (11) and simplifying we find

$$P\left(\sum_{i=1}^{N} y_i > kN\right) \leq \left(\frac{2}{L}\right)^N e^{-k\tanh^{-1}(k)N}\left(\frac{k}{\sqrt{1-k^2}}\right)^N \qquad (14)$$

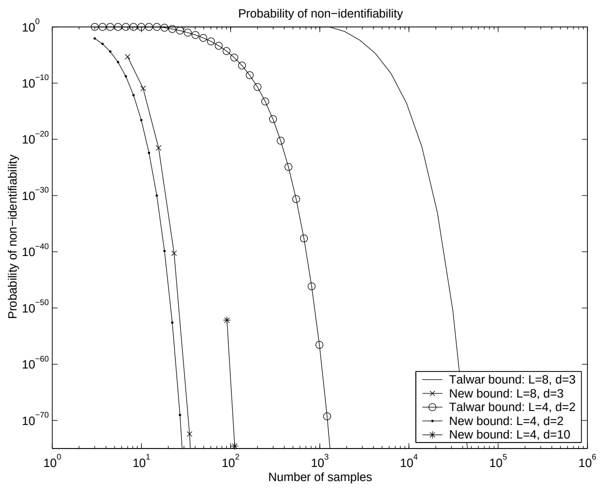

Figure 1. Finite alphabet CM sources: Upper bound on the probability that N source samples are not persistently exciting. d sources and L-PSK constellations.

Substituting $k = \frac{1}{d(d+1)}$ as in (6) and using equation (9) we obtain

$$\begin{aligned} & P\left((\hat{\mathbf{R}}_N)_{ij} > \tfrac{1}{d(d+1)}\right) \\ & = P\left(\tfrac{1}{N}\sum_{n=1}^{N} v_i(n)^* v_j(n) > \tfrac{1}{d(d+1)}\right) \\ & \leq LP\left(\sum_{i=1}^{N} y_i > kN\right) \\ & \leq 2\left(\tfrac{2}{L}\right)^{N-1} e^{-\frac{N}{d(d+1)}\tanh^{-1}\left(\frac{1}{d(d+1)}\right)} \left(\frac{1}{\sqrt{d^2(d+1)^2-1}}\right)^N \end{aligned} \quad (15)$$

The matrix $\hat{\mathbf{R}}_N$ is nonsingular if all entries $(\hat{\mathbf{R}}_N)_{ij}$ above the main diagonal are smaller than $1/(d(d+1))$. There are $\frac{1}{2}[d(d-1)+1][d(d-1)] \leq \frac{1}{2}d^4$ such entries. Finally using the fact that for any x such that $0 < x < 1$ we have $(1-x)^n > 1 - nx$, we can bound the probability of identifiability of d sources using N vector samples taken from L-PSK i.i.d. sources, $P_{id}(L, d, N)$ as follows:

Theorem 7 *Consider a linear mixture of N samples of d i.i.d. sources with L-PSK alphabet. The probability $P_{id}(L, d, N)$ that the data set is identifiable satisfies*

$$\begin{aligned} & P_{id}(L,d,N) \geq \\ & 1 - d^4\left(\tfrac{2}{L}\right)^{N-1} e^{-\frac{N}{d(d+1)}\tanh^{-1}\left(\frac{1}{d(d+1)}\right)} \left(\frac{1}{\sqrt{d^2(d+1)^2-1}}\right)^N \end{aligned} \quad (16)$$

This is better than the large deviation bound in theorem 6, since the dependence on N is exponential and not sub-exponential and is also valid for all values of N. Moreover we can see that as the alphabet size is increased the probability of non-identifiability approaches 0 as $L^{-(N-1)}$.

6. SIMULATIONS

We now illustrate a comparison of the new upper bound on identifiability, theorem 7, to the bound by Talwar [3], see figure 1. We can clearly see that the new bound is much better with orders of magnitudes less samples necessary for a given probability of identifiability.

7. CONCLUSION

We presented a rigorous proof of a sufficient condition for the identifiability of mixtures of finite alphabet CM signals, based on finitely many samples. For finite-alphabet cases, only an upper bound on the probability of identifiability given alphabet size, number of sources and number of samples could be derived. However the new bound is much tighter than previously known bounds.

8. ACKNOWLEDGEMENT

We would like to thank S. Litsyn for helpfull discussion on Chernoff's bound.

Appendix: Optimizing ν in the Chernoff bound

The optimal value of ν satifies the equation (see e.g., [8] page 54)

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

Using equation (10) we obtain

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

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 finally obtain

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

9. REFERENCES

[1] J. Treichler and B. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, pp. 459–471, Apr. 1983.

[2] A. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.

[3] 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.

[4] A. Leshem, N. Petrochilos, and A.-J. van der Veen, "Finite sample identifiability of multiple constant modulus signals," in *Proc. IEEE Workshop on Sensor array and multichannel signal processing*, Aug. 2002.

[5] R. Horn and C. Johnson, *Matrix Analysis.* Cambridge, 1994.

[6] W. Feller, *An introduction to probability theory and its applications, vol. 2, 2nd edition.* Wiley, 1971.

[7] A. Leshem, N. Petrochilos, and A.-J. van der Veen, "Finite sample identifiability of multiple constant modulus signals," *IEEE trans. on Information Theory*, Submitted 2001.

[8] J. Proakis, *Digital communications.* McGraw-Hill, 3rd ed., 1995.