Finite sample identifiability of multiple constant modulus sources

Nicolas PETROCHILOS[1,2], Amir LESHEM[1,3], Alle-Jan VAN DER VEEN[1]

[1]Circuits and Systems, Dept. of Electrical Engineering,
Delft University of Technology, Mekelweg 4, 2628CD Delft, The Netherlands

[2]I3S–CNRS, Algorithmes, 2000 route des Lucioles, BP.121, 06903 Sophia-Antipolis, France

[3]Metalink Broadband Access, Yakum Business Park, 60972, Israel

nicolas,leshem,allejan@cas.et.tudelft.nl

Résumé – Nous prouvons que les melanges de sources à modules constant peuvent etre identifiées avec une probabilitè 1 en un nombre fini d'echantillons. Ce resultat renforce des publications précedentes où l'on ne considerait que des echantillonages infini. La preuve est basée sur la technique de linearization de l'algorithme ACMA, suivi d'un argument recursif. Nous appliquons cette technique à l'identification un mélange de sources provenants d'un radar de surveillance secondaire.

Abstract – We prove that mixtures of constant modulus sources can be identified with probability 1 with a finite number of samples. 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, together with a simple inductive argument. We also discuss an application of this technique to the identification of mixtures of Secondary Surveillance Radar signals.

1 Introduction

The constant modulus algorithm (CMA) is very popular for blind equalization [1, 2] and blind separation of multiple Constant Modulus (CM) signals [3]. It was soon recognized that the CM cost function can be used also for the separation of non-Gaussian signals, and more specifically finite alphabet signals [4, 5].

The analysis of the CMA algorithm has been mostly studied in a statistical framework, based on the expected value of the CM cost function. Hence identifiability results are strictly speaking only valid for infinitely many samples and ergodic scenarios. For the finite alphabet case, an important result in [5] was a proof of identifiability when only finitely many samples are available. For the CM case, an unsatisfying argument in [3] argued that about $2d$ should be sufficient for identification of d sources. The argument was based on counting the equations and the number of unknowns.

In this paper we give a rigorous proof of identifiability of a mixture of d continuous CM sources with finitely many samples. We use the linearization technique of [3], together with a simple inductive argument, to show that $d^2 - d + 1$ many samples suffice with probability 1.

The problem is formulated in section 2. Section 3 contains the proof of the main theorem, and section 4 discusses an application to the separation of Secondary Surveillance Radar signals.

2 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,

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

where

$\mathbf{x}(t) = [x_1(t), \cdots, x_p(t)]^T$ is a $p \times 1$ vector of received signals at time t,

$\mathbf{A} = [\mathbf{a}_1, \cdots, \mathbf{a}_d]$, where $\mathbf{a}_i$ is the array response vector towards the i'th signal,

$\mathbf{s}(t) = [s_1(t), \cdots, s_d(t)]^T$ is a $d \times 1$ vector of source signals at time t. We further assume that all sources have constant modulus. This is represented by the assumption that for all t, $|s_i(t)| = 1$ $(i = 1, \cdots, d)$.

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 $s_i(t) = e^{j\phi_i(t)}$, where $\phi_i(t)$ is the unknown phase modulation for source i, and we define $\boldsymbol{\phi}(t) = [\phi_1(t), \cdots, \phi_d(t)]^T$ as the phase vector for all sources at time t. 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, the problem is determined only up to a permutation and a scaling with complex elements of unit modulus.

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

3 Main theorem

In this section, we derive an upper bound on the number of samples needed to guarantee unique identifiability (up to the known indeterminacies) of a constant modulus instantaneous mixture. We first show that given infinite data the above indeterminacies are the only ones. We then continue to derive finite sample results.

Let $\mathbb{T} = \{z : |z| = 1\}$ be the complex unit circle. Let $\mathbb{T}^d$ be the Cartesian product of d copies of $\mathbb{T}$. Note that $\mathbb{T}^d$ is the collection of d-tuples of CM signals. Define the range of a linear transformation $\mathbf{A}$ restricted to a set X as $\mathbf{A}(X) = \{\mathbf{A}(\mathbf{v}) \,|\, \mathbf{v} \in X\}$, and define $\mathbf{A}^\dagger$ as the pseudo inverse of $\mathbf{A}$.

Claim 3.1 *Assume that $\mathbf{A}, \mathbf{A}' \in \mathbb{C}_{p\times d}$ are two linear transformations of rank d (i.e., full column rank), such that $\mathbf{A}(\mathbb{T}^d) = \mathbf{A}'(\mathbb{T}^d)$. Then $\mathbf{G} = \mathbf{A}^\dagger \mathbf{A}'$ is a linear transformation mapping $\mathbb{T}^d$ onto $\mathbb{T}^d$.*

The proof of the claim is straightforward using the fact that both transformations are one-to-one from their domain ($\mathbb{C}^d$) onto their range.

Let $\mathcal{A}$ be the set of linear transformations from $\mathbb{C}^d$ to $\mathbb{C}^d$ which map $\mathbb{T}^d$ onto itself:

$$\mathcal{A} = \left\{ \mathbf{A} \in \mathbb{C}_{d\times d} | \mathbf{A} : \mathbb{T}^d \stackrel{onto}{\longrightarrow} \mathbb{T}^d \right\}.$$

The following theorem characterizes the linear transformations in $\mathcal{A}$.

Theorem 3.2 *Assume $\mathbf{A} \in \mathcal{A}$, then $\mathbf{A}$ is a product of a permutation and a diagonal matrix with diagonal elements on the unit circle.*

Proof Assume that $\mathbf{A}$ is as above. Then for every vector $\mathbf{v} \in \mathbb{T}^d$, $\mathbf{A}\mathbf{v} \in \mathbb{T}^d$. We will first prove that each row of $\mathbf{A}$ contains at most one non zero element with magnitude 1. Let

$$\mathbf{a} = [a_1 \ldots a_d] = \left[r_1 e^{j\phi_1} \ldots r_d e^{j\phi_d}\right]$$

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

$$\mathbf{s}_1 \mathbf{a} = r_1 + \sum_{k>1}^{d} r_k = 1 \tag{2}$$

since all r_k are positive real numbers. Similarly define $(\mathbf{s}_2)_1 = e^{-j\phi_1}$ and $(\mathbf{s}_2)_k = -e^{-j\phi_k}$ for $1 < k \le d$. Then

$$\mathbf{s}_2 \mathbf{a} = r_1 - \sum_{k>1}^{d} r_k \tag{3}$$

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

$$r_1 - \sum_{k>1}^{d} r_k = 1 \tag{4}$$

or

$$r_1 - \sum_{k>1}^{d} r_k = -1. \tag{5}$$

In the first case we obtain from (2) and (4) that $r_1 = 1$ and $\sum_{k>1} r_k = 0$ which is the desired result. In the latter case $r_1 = 0$ and $\sum_{k>1} r_k = 1$. Proceeding inductively we obtain that exactly one element of $\mathbf{a}$ is non-zero with magnitude 1. Since $\mathbf{A}$ is invertible (remember that $\mathbf{A}(\mathbb{T}^d) \cap \mathbb{T}^d$ contains a basis of $\mathbb{C}^d$) we obtain that $\mathbf{A}$ is a permutation of a diagonal matrix with diagonal elements in $\mathbb{T}$. □

We now show that infinitely many (i.i.d.) samples from a continuous alphabet which is supported on a dense set will suffice. By the independence assumption we are certain that we obtain a dense set of samples in $\mathbb{T}^d$ (otherwise there is an open set which we have missed, but this happens with probability zero). Since by continuity a linear transformation is determined by its values on a dense set we are finished.

A consequence of the above discussion is that it would be sufficient to characterize linear transformations mapping $\mathbb{T}^d$ into itself. This reduces the identifiability question into the solution of a specific set of quadratic equations, which we now analyze.

Theorem 3.3 *Let $N = d(d-1) + 1$. Let $\mathbf{s}(k)$, for $k = \{1, \ldots, N\}$, be i.i.d vectors (with independent components) in $\mathbb{T}^d$. Assume that there is a linear transformation $\mathbf{A}$ such that $\mathbf{y}(k) = \mathbf{A}\mathbf{s}(k)$, $\forall k = 1, \ldots, N$, with $\mathbf{y}(k) \in \mathbb{T}^d$, then with probability 1 $\mathbf{A} \in \mathcal{A}$, i.e., there is a diagonal matrix, with unit norm diagonal entries $\mathbf{\Lambda}$ and a permutation matrix $\mathbf{P}$ such that $\mathbf{A} = \mathbf{\Lambda}\mathbf{P}$.*

Proof It is sufficient to prove that each row of $\mathbf{A}$ contains exactly one non-zero element a_{ij} which is unit norm: $|a_{ij}| = 1$. Moreover since $\mathbf{s}(k)$ are random vectors, $\mathbf{A}$ will be invertible with probability 1. Hence we can consider separately each row of $\mathbf{A}$. Let $\mathbf{a}_m = [a_1, \ldots, a_d]$ be the m'th row of $\mathbf{A}$. Let $y(k) = \mathbf{y}_m(k)$ be the m'th element of $\mathbf{y}(k)$.

Then $\forall k \in \{1, \ldots, N\}$, we have:

$$\begin{aligned} y(k) &= \sum_{i=1}^{d} a_i s_i(k) \\ \Rightarrow \quad |y_i(k)|^2 &= |\sum_{i=1}^{d} a_i s_i(k)|^2 \\ \Rightarrow \quad 1 &= \sum_{1 \le i,j \le d} a_i a_j^* s_i(k) s_j^*(k) \end{aligned} \tag{6}$$

Denote $P_{ij} = a_i a_j^*$ and $P_T = \sum_{i=1}^{d} P_{ii}$. By linearizing (6) we obtain:

$$\mathbf{\Psi}\mathbf{p} = \mathbf{1} \tag{7}$$

where

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

$\mathbf{p} = \left[P_T, P_{12}, \ldots, P_{d(d-1)}\right]^T$, and $\mathbf{1} = [1, \ldots, 1]^T$. A non-trivial solution of (7) is:

$$\begin{cases} P_T = 1 \\ P_{ij} = 0, \quad \forall i \ne j \end{cases} \tag{8}$$

Since $P_{ij} = a_i a_j^*$ we immediately obtain that there is a j such that $a_i = 0$ for all $i \ne j$. Since $P_T = 1$ this also implies $|a_j|^2 = 1$ as needed. Two different rows cannot have their non-null element in the same column because

$\mathbf{A}$ would not be a bijection. Thus the absolute value of $\mathbf{A}$ is a permutation. Hence it is sufficient to prove that Ψ is full column rank (w.p.1), which is the content of the next lemma. □

Lemma 3.4 *Under conditions of theorem 3.3, the matrix Ψ has full column rank w.p. 1.*

(Note that if $N \le d(d-1)$, Ψ is definitely not full column rank because it is a wide matrix.)

Proof Given $N > d(d-1)$ samples of $\mathbf{s}(k)$, assume towards contradiction that there exists a vector $\boldsymbol{\alpha}$ such that $\Psi\boldsymbol{\alpha} = 0$, or equivalently $\exists\{\alpha_0, \alpha_{ij}, 1 \le i \ne j \le d\}$, not all zeros such that for every $t = 1, \cdots, N$, the next equation holds:

$$\alpha_0 + \sum_{i \ne j} \alpha_{ij} s_i(t) s_j^*(t) = 0 \tag{9}$$

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

$$\sum_{1<j} \alpha_{1j} s_j^*(t) \quad + s_1^*(t)\left(\alpha_0 + \sum_{1<i \ne j} \alpha_{ij} s_i(t) s_j^*(t)\right) + (s_1^*(t))^2 \sum_{1<i} \alpha_{i1} s_i(t) = 0$$

After taking the conjugate of this expression, we see that it is a set of N quadratic equations in s_1: $a(t) + b(t)s_1(t) + c(t)s_1^2(t) = 0$. Hence one of the following holds:
(a) $s_1(t)$ is a function of $(s_2(t), \cdots, s_d(t))$ which contradicts the independence assumption.
(b) The coefficients satisfy: $a(t) = b(t) = c(t) = 0$, $\forall t \in [1, N]$, hence:

1. $a(t) = 0 \Rightarrow \sum_{1<j} \alpha_{1j} s_j^*(t) = 0$. Since the $s_i(t)$ are i.i.d., it means that $\forall i$, $\alpha_{1i} = 0$.
2. Similarly, from $c(t) = 0$ we obtain $\forall i$, $\alpha_{i1} = 0$.
3. $b(t) = 0 \Rightarrow \alpha_0 + \sum_{1<i\ne j} \alpha_{ij} s_i(t) s_j^*(t) = 0$.

Applying inductively the same argument on $b(t)$ as we did on equation (9), we obtain that all α_{ij} are equal to zero. Therefore Ψ is full rank with probability one.

Note that the recursive application of the argument needs $2d + 2(d-1) + \cdots + 2 = d(d-1)$ many independent samples. Hence w.p.1 $d(d-1)$ samples are sufficient. □

4 Application to Secondary Surveillance Radar

Secondary Surveillance Radar (SSR) is a two-way radar, which behaves almost as a wireless communication system. One of the research topics in this field is the uplink reception on an array antenna. The uplink sources can be modeled as Zero/Constant Modulus (ZCM) signals: the complex signal is either 0, or has a modulus 1, with equal probability.

Using the same philosophy as in the preceding section, it is possible to derive how many samples would be sufficient to identify the mixing matrix of ZCM sources. The proof follows the same course, except that the matrix Ψ is $N \times (d^2(d+1)/2)$, and that the probability that Ψ is full column rank is not 1, but bounded by the probability described in the following theorem.

Theorem 4.1 *Let $\mathbf{s}(k)$, for $\{k = 1, \ldots, N\}$ be i.i.d ZCM d-vectors with independent components. Assume that there is a linear transformation $\mathbf{A}$ such that $\mathbf{y}(k) = \mathbf{A}\mathbf{s}(k)$, $\forall k = 1, \ldots, N$, where $\mathbf{y}(k)$ is ZCM d-vector as well. Then, with probability larger than $P(N)$, $\mathbf{A} \in \mathcal{A}$, i.e., $\mathbf{A}$ is identifiable up to a diagonal and a permutation. $P(N)$ is given by*

$$P(N) = P_\alpha(N)\left[\prod_{i=1}^{d-2} P_T^{(i)}(N-d)\right] p_f$$

where $P_\alpha(N) = \prod_{i=1}^{d}\left[1 - \left(1 - \frac{d-i+1}{2^{d-i+1}}\right)^{N-i+1}\right]$, *and*

$$P_T^{(i)}(N) = \sum_{N_i = l_{d-i+1}}^{N - L_{d-i+1}} P_C(N_i, N - L_{d-i+1}) \prod_{j=1}^{l_{d-i+1}} p_0(N_i - j + 1)$$

with

$$\begin{aligned} L_{d-i+1} &= \textstyle\sum_{j=1}^{i-1} l_{d-j+1}\,, \\ l_d &= d\left\lceil \frac{3d-1}{2} \right\rceil - 1\,, \\ P_C(N_a, N_b) &= \frac{1}{2^{N_a}} \frac{N_a!}{(N_a - N_b)! N_b!}\,, \\ p_f &= \textstyle\prod_{i=0}^{2} 1 - \left(\frac{3}{4}\right)^{N-d-i-L_3}\,, \\ p_o(N) &= 1 - 2^{-N}\,. \end{aligned}$$

Proof It is sufficient to prove that each row of $\mathbf{A}$ contains exactly one non-zero element a_{ij}, which is unit norm: $|a_{ij}| = 1$. Hence we can consider separately each row of $\mathbf{A}$. Let $\mathbf{a}_m = [a_1, \ldots, a_d]$ be the m'th row of $\mathbf{A}$. Let $y(k) = \mathbf{y}_m(k)$ be the m'th element of $\mathbf{y}(k)$. Then $\forall k \in \{1, \ldots, N\}$, we have:

$$y(k) = \sum_{i=1}^{d} a_i s_i(k) \tag{10}$$

The ZCM property can be written as: $\{y = 0$, or $|y| = 1\}$, which is equivalent to $y(yy^* - 1) = 0$, and $y = |y|^2 y$, which leads $\forall k \in \{1, \ldots, N\}$ to:

$$\begin{aligned} \sum_{l=1}^{d} a_l s_l(k) &= \sum_{i,j,k=1}^{d} a_i^* a_j a_k s_i^*(k) s_j(k) s_l(k) \\ &= \sum_{i=1}^{d} |a_i|^2 a_i |s_i(k)|^2 s_i(k) \\ &+ \sum_{\substack{1 \le i \\ 1 \le j < l \\ i \ne j \ne l}}^{d} 2 a_i^* a_j a_k s_i^*(k) s_j(k) s_l(k) \end{aligned} \tag{11}$$

Denote:

$$\begin{aligned} \mathbf{p}^T \;=\; & [a_1\left(|a_1|^2 - 1\right), \ldots, a_d\left(|a_d|^2 - 1\right), \\ & 2a_1^* a_1 a_2, \ldots, 2a_d^* a_d a_{d-1}, a_1^* a_2^2, \ldots, a_d^* a_{d-1}^2] \end{aligned}$$

by linearizing (11), and using $s_i^2 s_i^* = s_i$, we obtain:

$$\Psi\mathbf{p} = \mathbf{0} \tag{12}$$

where Ψ is a $N \times (d^2(d+1)/2)$ matrix, which k'th row, Ψ_k is equal to:

$$\begin{aligned} \Psi_k \;&= [s_1(k), s_2(k), \ldots, s_d(k), |s_1|^2(k) s_2(k), \ldots, \\ & |s_d|^2(k) s_{d-1}(k), s_1(k)^* s_2^2(k), \ldots, s_d^*(k) s_{d-1}^2(k)] \end{aligned}$$

Ψ is full rank with probability $P(N)$ (the proof is too long to be presented here, a sketch of it is placed at the end of this proof). Assuming that Ψ is full column rank, the only solution is $\mathbf{p} = \mathbf{0}$.

Since $\mathbf{A}$ is a bijection, there is at least an element a_i non-null. Then $\mathbf{p} = \mathbf{0}$ implies that:

$$\begin{cases} a_i(|a_i|^2 - 1) = 0 \\ a_i^* a_j^2 = 0, \quad \forall j \neq i \end{cases} \Rightarrow \begin{cases} |a_i| = 1 \\ a_j = 0, \quad \forall j \neq i \end{cases}$$

Two different rows cannot have their non-null element in the same column because $\mathbf{A}$ would not be a bijection. Thus the absolute value of $\mathbf{A}$ is a permutation, and the identifiability holds with probability $P(N)$. $\square$

Outlines of probability $P(N)$: The construction of the probability follows a similar path as lemma 3.4. We consider for each source s_i, the probability to receive a set of samples $\mathcal{S}_i$, with modulus 1. For this set $\mathcal{S}_i$, this source is a constant modulus source on which we use lemma 3.4.

5 Simulations

In order to check the validity of lemma 3.4 for continuous CM sources, we simulated 1000 independent runs with $N = 1, \cdots, 100$ samples and $d = 2, \cdots, 6$ 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 3.4.

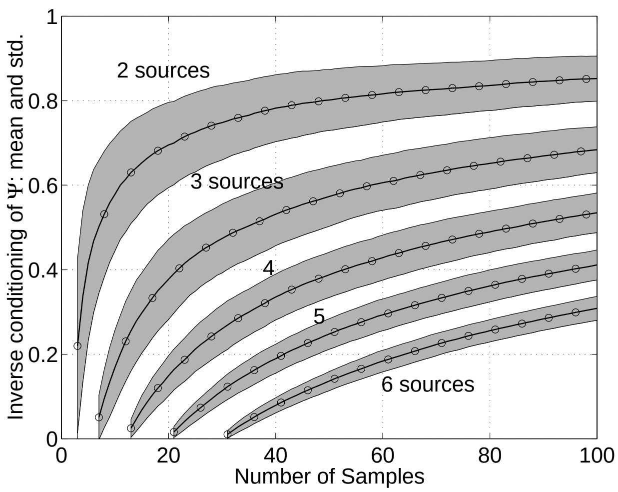

FIG. 1: *(circles) Average mean of the inverse condition number of Ψ, and (shaded area) its $\pm 1\sigma$ standard deviation interval for the continuous CM sources.*

Figure 1 shows the inverse of the condition number of Ψ, for varying d and N. Interestingly, it appears that Ψ is rather ill-conditioned when N is close to its lower limit. Indeed, the lemma did not insure the conditioning for Ψ. Nevertheless, one can note that with a few samples above the lower limit, this adverse situation improve significantly.

We present the result of theorem 4.1 in figure 2, where the complementary probability of identification is shown for several numbers of sources as a function of the number of samples.

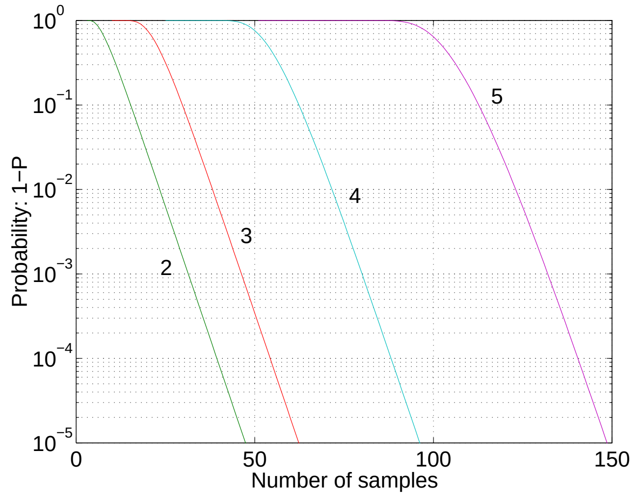

FIG. 2: *Complementary Probability of identification for* 2, 3, 4, *and* 5 *sources as a function of N.*

Figure 2 shows that the probability curves have the same negative slope, which is one "decade" for roughly 10 samples. Unfortunately, the initial point is of the order of d^3, which for a desired identifiability probability makes the required number of samples quite high.

6 Conclusion

In [3], the minimum number of samples needed for identifiability of the CM source separation problem was indicated, but the argument was based on counting the equations and the number of unknowns. Here we gave a rigorous proof of identifiability, but for a higher number of samples. An extension to the finite alphabet case is to be investigated, as this method has the potential to reduce the number of samples needed compared with [5].

References

[1] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Communications*, vol. 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. 31, pp. 459–471, Apr. 1983.

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

[4] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Proc. Letters*, vol. 1, pp. 29–31, Feb. 1994.

[5] S. Talwar and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. II. Performance analysis," *IEEE Tr. Signal Processing*, vol. 45, pp. 706–718, Mar. 1997.