Asymptotic Properties of the Algebraic Constant Modulus Algorithm

Alle-Jan van der Veen, *Member, IEEE*

***Abstract*—The algebraic constant modulus algorithm (ACMA) is a noniterative blind source separation algorithm. It computes jointly beamforming vectors for all constant modulus sources as the solution of a joint diagonalization problem. In this paper, we analyze its asymptotic properties and show that (unlike CMA) it converges to the Wiener beamformer when the number of samples or the signal-to-noise ratio (SNR) goes to infinity. We also sketch its connection to the related JADE algorithm and derive a version of ACMA that converges to a zero-forcing beamformer. This gives improved performance in applications that use the estimated mixing matrix, such as in direction finding.**

***Index Terms*—Array signal processing, blind beamforming, constant modulus algorithm, simultaneous diagonalization.**

I. INTRODUCTION

CONSTANT modulus algorithms (CMAs) enjoy widespread popularity as methods for blind source separation and equalization of communication signals. First derived as LMS-type adaptive equalizers by Godard [8] and Treichler *et al.* [24], [25], CMAs are straightforward to implement, robust, and computationally of modest complexity. Quite soon, the algorithms were also applied to blind beamforming (spatial source separation), which gave rise to the similar constant modulus array [21]. An extensive literature exists, but it will not be cited here; instead, we refer to the special issue of the PROCEEDINGS OF THE IEEE, October 1998, and, in particular, [10], [26], and references therein.

Despite its effectiveness and apparent simplicity, adaptive implementations of the CMA come along with several complicating factors that have never really been solved. In particular, convergence can be slow (order hundreds of samples) at an unpredictable speed depending on initialization, and the step size may have to be tuned to avoid stability problems. For the purpose of blind source separation, an additional complication is that only a single source is found at a time. To recover the other signals successively or in parallel, the previous solutions have to be removed from the data, or independence constraints must be introduced, with additional complications for the convergence [11], [14]–[16], [21], [23].

The algebraic constant modulus algorithm (ACMA) was introduced in [29] as an algebraic method for computing the complete collection of beamformers in one shot, as the solution of a generalized eigenvalue problem. Only a small batch of samples is needed (order number of sources squared), and the number of constant modulus signals can be detected as well. Convergence is not an issue. It has been successfully applied to real data in a variety of scenarios for up to six sources simultaneously [29].

The potential performance of the CMA receiver, i.e., the minima of the modulus error cost function to which the adaptive CMA tries to converge, has been studied in detail recently in a series of papers by Tong, Johnson, and others [7], [9], [18], [32], [33]. These papers provide quantitative evidence for the observation already made by Godard that the minima of the constant modulus cost function are often very near the (nonblind) Wiener receivers or linear minimum mean square error (LMMSE) receivers.

Although very promising, the performance of ACMA has not been studied so far, except empirically and with seemingly contradicting conclusions [17], [22]. In this paper, we make a start at a theoretical analysis by investigating the asymptotic properties of ACMA. The main result is that with Gaussian noise, ACMA converges *exactly* to the Wiener solution when the number of samples or the signal-to-noise ratio (SNR) goes to infinity.

The analysis is based on a reformulation of ACMA as a fourth-order statistics method. As such, it can be directly derived from the CMA cost function by replacing the nonlinear optimization by two steps: a linear one in which a subspace is found, followed by a nonlinear optimization restricted to this subspace. This reformulation shows that ACMA is closely related to the JADE algorithm by Cardoso and Souloumiac [4], which is a well-known blind beamforming algorithm for separating independent non-Gaussian sources. We sketch the relations between the two algorithms. This complements the known relations between JADE and the larger class of algebraic fourth-order cumulant-based separation techniques based on contrasts or cumulant matching [5], [6], [12], [19], [20], [30], [31], [34]; see [2] and [3] for an overview. An inspiring start to this analysis was found in [20] and [30], in which relations between several fourth-order source separation algorithms are investigated, including CMA and JADE. In these papers, the algorithms are placed in a common framework of least squares matching of fourth-order cumulants, where the beamformer after a prewhitening step is constrained to be unitary. The essential role played by this prewhitening step (in fact, the prewhitening suggested in [20] is inaccurate) is not noted in [20] and [30] . Indeed, it will be shown here that the precise choice of the prewhitening is crucial for the asymptotic convergence of ACMA to the Wiener receiver and of JADE to a zero-forcing receiver.

Wiener receivers are attractive because they maximize the output signal-to-interference-plus-noise ratio (SINR). However,

Manuscript received April 13, 2000; revised April 30, 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Alex C. Kot.

The author is with the Department of Electrical Engineering/DIMES, Delft University of Technology, Delft, The Netherlands (e-mail: allejan@cas.et.tudelft.nl).

Publisher Item Identifier S 1053-587X(01)05854-8.

for some applications such as direction finding, a zero-forcing beamformer is preferred because its inverse provides an unbiased estimate of the mixing matrix from which the directions can be estimated. For this case, we construct a slightly modified version of ACMA that asymptotically in the number of samples converges to a point close to the zero-forcing solution. All algorithms (CMA, ACMA, ZF-ACMA, and JADE) are subsequently tested in simulations and compared with the Wiener and zero-forcing receivers.

Outline: Section II defines the problem, and Section III provides a compact presentation of the original ACMA. We subsequently look at the connection to the CMA cost function (Section IV), the noise-free properties (Section V), and the asymptotic properties of the algorithm in noise (Section VI), from which it follows that ACMA converges to the Wiener solution. We also derive a version of ACMA that approximately converges to the zero-forcing solution (Section VII) and compare CMA, ACMA, ZF-ACMA, and JADE in simulations (Section VIII).

Notation: We adopt the following notation:

$\overline{\cdot}$	Complex conjugation.
${}^{\mathrm{T}}$	Matrix transpose.
${}^{\mathrm{H}}$	Matrix complex conjugate transpose.
$\dagger$	Matrix pseudo-inverse (Moore–Penrose inverse).
$\underline{\cdot}$	Prewhitened data.
$\mathbf{0}$	Vector of all 0s.
$\mathbf{1}$	Vector of all 1s.
$\mathrm{E}(\cdot)$	Mathematical expectation operator.
vec$(\mathbf{A})$	Stacking of the columns of $\mathbf{A}$ into a vector.
$\otimes$	Kronecker product.
$\circ$	Khatri–Rao product (column-wise Kronecker product):

$$\mathbf{A}\circ\mathbf{B} := [\mathbf{a}_1\otimes\mathbf{b}_1 \ \ \mathbf{a}_2\otimes\mathbf{b}_2 \ \ \cdots].$$

Notable properties are, for matrices $\mathbf{A}, \mathbf{B}, \ldots$ and vectors $\mathbf{a}, \mathbf{b}$ of compatible sizes

$$\mathrm{vec}(\mathbf{a}\mathbf{b}^{\mathrm{H}}) = \overline{\mathbf{b}}\otimes\mathbf{a} \tag{1}$$

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = \mathbf{AC}\otimes\mathbf{BD} \tag{2}$$

$$\mathrm{vec}(\mathbf{ABC}) = \left(\mathbf{C}^{\mathrm{T}}\otimes\mathbf{A}\right)\mathrm{vec}(\mathbf{B}) \tag{3}$$

$$\mathrm{vec}(\mathbf{A}\mathrm{diag}(\mathbf{b})\mathbf{C}) = \left(\mathbf{C}^{\mathrm{T}}\circ\mathbf{A}\right)\mathbf{b}. \tag{4}$$

II. Data Model and Preliminaries

Consider d independent sources, transmitting complex-valued signals $s_i(t)$ with constant modulus waveforms ($|s_i(t)| = 1$) in a wireless scenario. The signals are received by an array of M antennas, demodulated to baseband and sampled with period T. We stack the resulting outputs $x_i(kT)$ into vectors $\mathbf{x}_k = \mathbf{x}(kT)$ and collect N samples in a matrix $\mathbf{X}$: $M\times N$. Assuming that the sources are sufficiently narrowband in comparison to the delay spread of the multipath channel, this leads to the well-known data model

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k \quad\Leftrightarrow\quad \mathbf{X} = \mathbf{AS}. \tag{5}$$

$\mathbf{A} = [\mathbf{a}_1 \ \cdots \ \mathbf{a}_d] \in \mathbb{C}^{M\times d}$ is the array response matrix. The rows of $\mathbf{S}\in\mathbb{C}^{d\times N}$ contain the samples of the source signals. Both $\mathbf{A}$ and $\mathbf{S}$ are unknown, and the objective is, given $\mathbf{X}$, to find a factorization $\mathbf{X} = \hat{\mathbf{A}}\hat{\mathbf{S}}$ such that $|\hat{\mathbf{S}}_{ij}| = 1$. If the problem is identifiable, then $\mathbf{S}$ is recovered up to the usual indeterminacies of arbitrary scalings and ordering of its rows. Alternatively, and more conveniently, we try to find a beamforming matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_d] \in \mathbb{C}^{M\times d}$ of full row rank d such that $\hat{\mathbf{S}} = \mathbf{W}^{\mathrm{H}}\mathbf{X}$.

In the presence of additive noise, we write $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k$, or

$$\mathbf{X} = \mathbf{AS} + \mathbf{N}. \tag{6}$$

Assumptions: Model assumptions used in the analysis are summarized as follows.

1) $M\geq d$, $N\geq d^2$.
2) $\mathbf{A}$ has full column rank d.
3) The signals are assumed to be random, independent identically distributed (i.i.d.), zero mean, circularly symmetric, with modulus equal to 1. Note that this rules out BPSK (± 1) sources.
4) The noise is assumed to be additive white, zero mean, circularly symmetric, complex Gaussian distributed with covariance $\mathbf{R}_n = \mathrm{E}(\mathbf{n}\mathbf{n}^{\mathrm{H}}) = \sigma^2\mathbf{I}$ and independent from the sources.
5) The noise-free problem is considered essentially identifiable.

Identifiability: We will assume that the problem is *essentially identifiable*, i.e., that for a given matrix $\mathbf{X}$ of size $M\times N$, we can find a factorization $\mathbf{X} = \hat{\mathbf{A}}\hat{\mathbf{S}}$ ($|\hat{\mathbf{S}}_{ij}| = 1$), which is unique up to the above-mentioned phase-and-ordering indeterminacies. Despite extensive research on CMA, minimal conditions that guarantee this identifiability for finite N are not completely known, nor will they be studied in this paper. ACMA requires $N\geq d^2$ and sufficiently exciting signals. By counting the number of equations and unknowns (a not completely convincing argument), it was motivated in [29] that identifiability is expected in general already for $N\geq 2d$ and sufficiently exciting signals.

Wiener and Zero-Forcing Beamformers: We will compare the outcome of ACMA to two beamformers that assume known source data $\mathbf{S}$, namely, the Wiener beamformer and the zero-forcing beamformer. In a deterministic context, the Wiener beamformer based on sample data is derived as the solution to the LMMSE problem

$$\begin{aligned}\hat{\mathbf{W}} &= \arg\min_{\mathbf{W}} \left\|\mathbf{W}^{\mathrm{H}}\mathbf{X} - \mathbf{S}\right\|_{\mathrm{F}}^2 = (\mathbf{S}\mathbf{X}^{\dagger})^{\mathrm{H}} \\ &= \left(\frac{1}{N}\mathbf{X}\mathbf{X}^{\mathrm{H}}\right)^{-1}\frac{1}{N}\mathbf{X}\mathbf{S}^{\mathrm{H}}.\end{aligned} \tag{7}$$

As $N\rightarrow\infty$, the Wiener beamformer converges to

$$\hat{\mathbf{W}} = \mathbf{R}_x^{-1}\mathbf{A}. \tag{8}$$

Likewise, the zero-forcing beamformer based on sample data is simply a left-inverse of a least-squares estimate of $\mathbf{A}$ or $\hat{\mathbf{W}}^{\mathrm{H}} = \hat{\mathbf{A}}^{\dagger}$, where

$$\hat{\mathbf{A}} = \arg\min_{\mathbf{A}} \|\mathbf{X} - \mathbf{AS}\|_{\mathrm{F}}^2 = \mathbf{XS}^{\dagger} = \frac{1}{N}\mathbf{XS}^{\mathrm{H}}\left(\frac{1}{N}\mathbf{SS}^{\mathrm{H}}\right)^{-1}. \tag{9}$$

As $N \to \infty$, we have that $\hat{\mathbf{A}} \to \mathbf{A}$ and that the zero-forcing beamformer converges to

$$\hat{\mathbf{W}} = \mathbf{A}^{\dagger\mathrm{H}}. \tag{10}$$

III. Derivation of the ACMA

A. Algorithm Outline

We summarize the derivation of the basic ACMA algorithm for the noiseless case (see also [29]). The objective is to find all independent beamforming vectors $\mathbf{w}$ that reconstruct a signal with a constant modulus, i.e.,

$$\mathbf{w}^{\mathrm{H}}\mathbf{x}_k = \hat{s}_k, \qquad \text{such that} \quad |\hat{s}_k|^2 = 1 \; (k = 1, \ldots, N).$$

Let $\mathbf{x}_k$ be the kth column of $\mathbf{X}$. By substitution, we find

$$\mathbf{w}^{\mathrm{H}}(\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}})\mathbf{w} = 1, \qquad k = 1, \ldots, N. \tag{11}$$

Note that $\mathbf{w}^{\mathrm{H}}(\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}})\mathbf{w} = (\overline{\mathbf{x}}_k \otimes \mathbf{x}_k)^{\mathrm{H}}(\overline{\mathbf{w}} \otimes \mathbf{w})$. We can stack the rows $(\overline{\mathbf{x}}_k \otimes \mathbf{x}_k)^{\mathrm{H}}$ of the data into a matrix $\mathbf{P} := [\overline{\mathbf{X}} \circ \mathbf{X}]^{\mathrm{H}}$ (size $N \times d^2$). Then, (11) is equivalent to finding all $\mathbf{w}$ that satisfy

$$\mathbf{Py} = \mathbf{1}, \qquad \mathbf{y} = \overline{\mathbf{w}} \otimes \mathbf{w}.$$

This is a linear system of equations, subject to a quadratic constraint. The linear system is overdetermined once $N \geq d^2$, and we will assume that this is the case.

In general outline, the ACMA technique solves this problem using the following steps.

1) *First, Solve the Linear System* $\mathbf{Py} = \mathbf{1}$. Note that there are at least d independent solutions to the linear system, namely, $\overline{\mathbf{w}}_i \otimes \mathbf{w}_i$ ($i = 1, \ldots, d$). In addition, however, a linear combination of these solutions

$$\mathbf{y} = \lambda_1(\overline{\mathbf{w}}_1 \otimes \mathbf{w}_1) + \cdots + \lambda_d(\overline{\mathbf{w}}_d \otimes \mathbf{w}_d)$$

(scaled such that $\sum \lambda_i = 1$) will also solve $\mathbf{Py} = \mathbf{1}$.

To find a basis of solutions, let $\mathbf{Q}$ be any unitary matrix such that $\mathbf{Q1} = \sqrt{N}\left[\begin{smallmatrix}1\\ \mathbf{0}\end{smallmatrix}\right]$, for example, found by computing a QR factorization of $[\mathbf{1} \quad \mathbf{P}]$. Apply $\mathbf{Q}$ to $[\mathbf{1} \quad \mathbf{P}]$ and partition the result as

$$\mathbf{Q}[\mathbf{1} \quad \mathbf{P}] =: \sqrt{N}\begin{bmatrix} 1 & \hat{\mathbf{p}}^{\mathrm{H}} \\ \mathbf{0} & \mathbf{G} \end{bmatrix}. \tag{12}$$

Then

$$\mathbf{Py} = \mathbf{1} \quad \Leftrightarrow \quad \mathbf{Q}[\mathbf{1} \quad \mathbf{P}]\begin{bmatrix} -1 \\ \mathbf{y} \end{bmatrix} = \mathbf{0}$$
$$\Leftrightarrow \quad \begin{cases} \hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = 1 \\ \mathbf{Gy} = \mathbf{0}. \end{cases} \tag{13}$$

The condition $\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = 1$ is momentarily dropped since it can always be satisfied by a scaling of $\mathbf{y}$ [cf. step 3) below]. All solutions to the condition $\mathbf{Gy} = \mathbf{0}$ are found from a basis $\{\mathbf{y}_i\}$ of the null space of the matrix $\mathbf{G}$, which is conveniently obtained from an SVD of $\mathbf{G}$. Generically (after prefiltering, see below), there are precisely d solutions.

2) *Decouple*: Find a basis $\{\overline{\mathbf{w}}_1 \otimes \mathbf{w}_1, \ldots, \overline{\mathbf{w}}_d \otimes \mathbf{w}_d\}$ of structured vectors that span the same linear subspace as $\{\mathbf{y}_1, \ldots, \mathbf{y}_d\}$. This can be formulated as a subspace fitting problem

$$\mathbf{W} = \arg\min_{\mathbf{W},\mathbf{M}} \left\|\mathbf{Y} - \left(\overline{\mathbf{W}} \circ \mathbf{W}\right)\mathbf{M}\right\|_{\mathrm{F}}^2$$

where $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_d]$, and $\mathbf{M}$ is a full-rank $d \times d$ matrix that relates the two bases of the subspace. Alternatively, we can formulate this as a joint diagonalization problem since

$$\begin{aligned}\left\|\mathbf{Y} - \left(\overline{\mathbf{W}} \circ \mathbf{W}\right)\mathbf{M}\right\|_{\mathrm{F}}^2 &= \sum_i \left\|\mathbf{y}_i - \left(\overline{\mathbf{W}} \circ \mathbf{W}\right)\mathbf{m}_i\right\|_{\mathrm{F}}^2 \\ &= \sum_i \left\|\mathbf{Y}_i - \mathbf{W}\mathbf{\Lambda}_i\mathbf{W}^{\mathrm{H}}\right\|_{\mathrm{F}}^2\end{aligned}$$

where $\mathbf{m}_i$ is the ith column of $\mathbf{M}$, $\mathbf{\Lambda}_i = \mathrm{diag}(\mathbf{m}_i)$ is the diagonal matrix constructed from this vector, and $\mathbf{Y}_i$ is the matrix obtained by unstacking $\mathbf{y}_i$ such that $\mathrm{vec}(\mathbf{Y}_i) = \mathbf{y}_i$; we have also used (4). The latter equation shows that all $\mathbf{Y}_i$ can be diagonalized by the same matrix $\mathbf{W}$. The resulting joint diagonalization problem is a generalization of the standard eigenvalue decomposition problem and can be solved [29]. An overview and comparison of techniques for this is found in [12].

3) In Step 1), we implemented the condition $\mathbf{Gy} = \mathbf{0}$ but dropped the scaling condition $\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = 1$ and, thus, lost the correct scaling of the $\mathbf{w}_i$. Rather than constraining $\mathbf{M}$ or the $\mathbf{\Lambda}_i$, this is more easily fixed by scaling each solution such that the average output power

$$\begin{aligned}\frac{1}{N}\sum_{k=1}^{N} |(s_i)_k|^2 &= \frac{1}{N}\sum_{k=1}^{N} \mathbf{w}_i^{\mathrm{H}}\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}\mathbf{w}_i \\ &= \mathbf{w}_i^{\mathrm{H}}\left(\frac{1}{N}\sum_{k=1}^{N}\mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}\right)\mathbf{w}_i\end{aligned} \tag{14}$$

is equal to 1.

In the noise-free case and with $N \geq d^2$, this algorithm produces the exact separating beamformer $\mathbf{W} = \mathbf{A}^{\dagger\mathrm{H}}$.

By squaring (12), we obtain explicit expressions for $\hat{\mathbf{p}}$ and a matrix $\hat{\mathbf{C}} := \mathbf{G}^{\mathrm{H}}\mathbf{G}$ that will be useful later:

$$\begin{aligned}\hat{\mathbf{p}} &= \frac{1}{N}\mathbf{P}^{\mathrm{H}}\mathbf{1} = \frac{1}{N}\sum \overline{\mathbf{x}}_k \otimes \mathbf{x}_k \\ \hat{\mathbf{C}} &:= \mathbf{G}^{\mathrm{H}}\mathbf{G} = \frac{1}{N}\mathbf{P}^{\mathrm{H}}\mathbf{P} - \hat{\mathbf{p}}\hat{\mathbf{p}}^{\mathrm{H}} \\ &= \frac{1}{N}\sum (\overline{\mathbf{x}}_k \otimes \mathbf{x}_k)(\overline{\mathbf{x}}_k \otimes \mathbf{x}_k)^{\mathrm{H}} \\ &\quad - \left[\frac{1}{N}\sum \overline{\mathbf{x}}_k \otimes \mathbf{x}_k\right]\left[\frac{1}{N}\sum \overline{\mathbf{x}}_k \otimes \mathbf{x}_k\right]^{\mathrm{H}}.\end{aligned} \tag{15}$$

The former expression shows that

$$\hat{\mathbf{p}} = \frac{1}{N}\sum \overline{\mathbf{x}}_k \otimes \mathbf{x}_k = \operatorname{vec}\left(\frac{1}{N}\sum \mathbf{x}_k\mathbf{x}_k^{\mathrm{H}}\right) = \operatorname{vec}(\hat{\mathbf{R}}_x)$$

where $\hat{\mathbf{R}}_x$ is the sample covariance matrix of the data. Thus (for $\mathbf{y} = \overline{\mathbf{w}} \otimes \mathbf{w}$)

$$\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = \mathbf{w}^{\mathrm{H}}\hat{\mathbf{R}}_x\mathbf{w} \tag{16}$$

and we see that the condition $\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = 1$ in (13) is implemented by step 3) of the algorithm outline, where the average output power of each beamformer is fixed to 1.

The significance of $\hat{\mathbf{C}}$ is two fold. First, its null space is evidently the same as that of $\mathbf{G}$; hence, we can obtain the basis $\{\mathbf{y}_i\}$ from an eigenvalue decomposition of $\hat{\mathbf{C}}$. Numerically, this is not advisable, but an analysis of this null space is much easier done for $\hat{\mathbf{C}}$ since it has fixed size and converges to a matrix $\mathbf{C}$ as $N \to \infty$. Second, as a closer inspection of equation (15) shows, $(1/N)\mathbf{C}$ has an important interpretation as being the *covariance matrix of the sample data covariance* $\hat{\mathbf{R}}_x$ (and $\hat{\mathbf{C}}$ is a sample estimate of $\mathbf{C}$).

B. Whitening and Rank Reduction

A crucial aspect of the above technique is that the basis $\{\mathbf{y}_i\}$ should not contain other components than the desired $\{\overline{\mathbf{w}}_i \otimes \mathbf{w}_i\}$; otherwise, we cannot pose the problem as a joint diagonalization. For this, it is essential that there are precisely d linearly independent solutions to $\mathbf{P}\mathbf{y} = \mathbf{1}$ and no additional spurious solutions. However, additional solutions exist if $\mathbf{X}$ is rank deficient, e.g., because the number of sensors is larger than the number of sources ($\mathbf{A}$ tall). This is simply treated by a prefiltering operation that reduces the number of rows of $\mathbf{X}$ from M to d, as we discuss here.

The underscore (_) is used to denote prefiltered variables. Thus, let $\underline{\mathbf{X}} := \mathbf{F}^{\mathrm{H}}\mathbf{X}$, where $\mathbf{F}$: $M \times d$. Then

$$\underline{\mathbf{X}} = \underline{\mathbf{A}}\mathbf{S} + \underline{\mathbf{N}}, \qquad \text{where} \quad \underline{\mathbf{A}} := \mathbf{F}^{\mathrm{H}}\mathbf{A},\ \underline{\mathbf{N}} := \mathbf{F}^{\mathrm{H}}\mathbf{N}.$$

$\underline{\mathbf{X}}$ has only d channels, and $\underline{\mathbf{A}}$: $d \times d$ is square. The blind beamforming problem is now replaced by finding a separating beamforming matrix $\mathbf{T}$: $d \times d$ with columns $\mathbf{t}_i$, acting on $\underline{\mathbf{X}}$ (see Fig. 1). After $\mathbf{T}$ has been found, the beamforming matrix on the original data will be $\mathbf{W} = \mathbf{F}\mathbf{T}$.

Assume that the noise is white i.i.d. with covariance matrix $\mathbf{R}_n = \sigma^2\mathbf{I}$. We can choose $\mathbf{F}$ such that the resulting data matrix $\underline{\mathbf{X}}$ is white, as follows. Let $\hat{\mathbf{R}}_x = (1/N)\mathbf{X}\mathbf{X}^{\mathrm{H}}$ be the noisy sample data covariance matrix, with eigenvalue decomposition

$$\hat{\mathbf{R}}_x = \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}^2\hat{\mathbf{U}}^{\mathrm{H}} = \begin{bmatrix}\hat{\mathbf{U}}_s & \hat{\mathbf{U}}_n\end{bmatrix}\begin{bmatrix}\hat{\mathbf{\Sigma}}_s^2 & \\ & \hat{\mathbf{\Sigma}}_n^2\end{bmatrix}\begin{bmatrix}\hat{\mathbf{U}}_s^{\mathrm{H}} \\ \hat{\mathbf{U}}_n^{\mathrm{H}}\end{bmatrix}. \tag{17}$$

Here, $\hat{\mathbf{U}}$ is $M \times M$ unitary, and $\hat{\mathbf{\Sigma}}^2$ is $M \times M$ diagonal ($\hat{\mathbf{\Sigma}}$ contains the singular values of $\mathbf{X}/\sqrt{N}$). The d largest eigenvalues are collected into a diagonal matrix $\hat{\mathbf{\Sigma}}_s^2$ and the corresponding d eigenvectors into $\hat{\mathbf{U}}_s$ (they span the "signal subspace"). In this notation, define $\mathbf{F}$ as

$$\mathbf{F} = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s^{-1}. \tag{18}$$

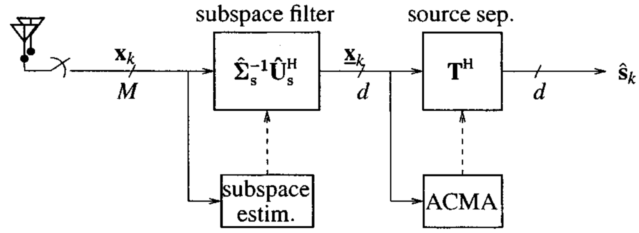

Fig. 1. Blind beamforming prefiltering structure.

Given $\mathbf{X}$, compute beamformers $\mathbf{W}$ and $\hat{\mathbf{S}} = \mathbf{W}^{\mathrm{H}}\mathbf{X}$:

1. SVD: $\mathbf{X} =: \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}$

   Prefiltering: $\underline{\mathbf{X}} := \hat{\mathbf{\Sigma}}_s^{-1}\hat{\mathbf{U}}_s^{\mathrm{H}}\mathbf{X} = \hat{\mathbf{V}}_s$

   $$\underline{\hat{\mathbf{C}}} := \frac{1}{N}\sum(\overline{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k)(\overline{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k)^{\mathrm{H}} + -\left[\frac{1}{N}\sum \overline{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k\right]\left[\frac{1}{N}\sum \overline{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k\right]^{\mathrm{H}}.$$

   EVD of $\underline{\hat{\mathbf{C}}}$, let $\{\mathbf{y}_i\}$ be the d least dominant eigenvectors.
2. $\mathbf{Y}_i = \operatorname{vec}^{-1}\mathbf{y}_i \ \ (i = 1, \cdots, d)$

   Find $\mathbf{T}$ to jointly diagonalize $\mathbf{Y}_i$ as $\mathbf{Y}_i = \mathbf{T}\mathbf{\Lambda}_i\mathbf{T}^{\mathrm{H}}$
3. Scale each column of $\mathbf{T}$ to norm 1.

   Set $\mathbf{W} = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s^{-1}\mathbf{T}$ and $\hat{\mathbf{S}} = \mathbf{T}^{\mathrm{H}}\underline{\mathbf{X}}$

Fig. 2. Summary of ACMA.

This prewhitening is such that $\underline{\hat{\mathbf{R}}}_x := (1/N)\underline{\mathbf{X}}\underline{\mathbf{X}}^{\mathrm{H}}$ is unity: $\underline{\hat{\mathbf{R}}}_x = \mathbf{I}$, and it also reduces the dimension of $\mathbf{X}$ from M rows to d rows. After prewhitening, we can continue with the algorithm, as outlined before.

The resulting algorithm is summarized in Fig. 2. In comparison with the outline, an additional ingredient is the prefiltering, for which an SVD of the data matrix $\mathbf{X}$ is needed. The prefiltering is primarily used to reduce the dimension. The preference for a prefilter that *whitens* the data covariance matrix follows from an analysis of the algorithm in the presence of noise, as done in the next sections.

IV. Formulation as an Optimization Problem

The ACMA procedure outlined in the previous section was derived for the noiseless case. With noise, the same algorithm is used unchanged, but obviously, the resulting beamformers will be noise-perturbed as well. The analysis of their properties is facilitated if we write these as the solutions of an optimization problem. This will also point out the correspondence to CMA.

The CMA cost function is usually defined as [24]

$$\mathbf{w} = \arg\min_{\mathbf{w}} \mathrm{E}\left(|\hat{s}_k|^2 - 1\right)^2, \qquad \hat{s}_k = \mathbf{w}^{\mathrm{H}}\mathbf{x}_k. \tag{19}$$

Given a finite batch of N data samples, we cannot solve (19). Therefore, we pose a corresponding least squares problem

$$\mathbf{w} = \arg\min_{\mathbf{w}} \frac{1}{N}\sum\left(|\hat{s}_k|^2 - 1\right)^2, \qquad \hat{s}_k = \mathbf{w}^{\mathrm{H}}\mathbf{x}_k. \tag{20}$$

We refer to this as the CMA(2,2) problem in this paper; its solution coincides with that of (19) as $N \to \infty$. Introducing $\mathbf{y} = \overline{\mathbf{w}} \otimes \mathbf{w}$ and using the factorization in (12), we derive that

$$\frac{1}{N}\sum \left(|\hat{s}_k|^2 - 1\right)^2 = \frac{1}{N}\sum \left[(\overline{\mathbf{x}}_k \otimes \mathbf{x}_k)^{\mathrm{H}}(\overline{\mathbf{w}} \otimes \mathbf{w}) - 1\right]^2 = \frac{1}{N}\|\mathbf{P}\mathbf{y} - \mathbf{1}\|^2 = \left|\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} - 1\right|^2 + \|\mathbf{G}\mathbf{y}\|^2.$$

Let $\hat{\mathbf{y}}$ be the (structured) minimizer of this expression, and define $\beta = \hat{\mathbf{p}}^{\mathrm{H}}\hat{\mathbf{y}}$. Equation (16) shows it is the output power of the beamformer corresponding to $\hat{\mathbf{y}}$, and hence, $\beta > 0$. Regarding β as some known fixed constant, we can add a condition that $\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = \beta$ to the optimization problem without changing the outcome:

$$\begin{aligned}\hat{\mathbf{y}} &= \arg\min_{\substack{\mathbf{y}=\overline{\mathbf{w}}\otimes\mathbf{w}\\ \hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y}=\beta}} \left|\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} - 1\right|^2 + \|\mathbf{G}\mathbf{y}\|^2 \\ &= \arg\min_{\substack{\mathbf{y}=\overline{\mathbf{w}}\otimes\mathbf{w}\\ \hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y}=\beta}} |\beta - 1|^2 + \|\mathbf{G}\mathbf{y}\|^2 \\ &= \arg\min_{\substack{\mathbf{y}=\overline{\mathbf{w}}\otimes\mathbf{w}\\ \hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y}=\beta}} \|\mathbf{G}\mathbf{y}\|^2.\end{aligned}$$

Since β is real and positive, replacing β by 1 will only scale the solution $\hat{\mathbf{y}}$ to $(1/\beta)\hat{\mathbf{y}}$ and does not affect the fact that it has a Kronecker structure. The scaled condition $\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = 1$ in turn motivates in a natural way the choice of a prewhitening filter $\mathbf{F} = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s^{-1}$, as given in (18). Indeed, we derived in (16) that $\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = \mathbf{w}^{\mathrm{H}}\hat{\mathbf{R}}_x\mathbf{w}$. If we change variables to $\underline{\mathbf{x}} = \hat{\mathbf{\Sigma}}_s^{-1}\hat{\mathbf{U}}_s^{\mathrm{H}}\mathbf{x}$ and $\mathbf{w} = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s^{-1}\mathbf{t}$, then $\underline{\hat{\mathbf{R}}}_x = \mathbf{I}$, and

$$\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = \mathbf{w}^{\mathrm{H}}\hat{\mathbf{R}}_x\mathbf{w} = \mathbf{t}^{\mathrm{H}}\mathbf{t} = \|\mathbf{t}\|^2.$$

Moreover, $(\overline{\mathbf{t}}\otimes\mathbf{t})^{\mathrm{H}}(\overline{\mathbf{t}}\otimes\mathbf{t}) = \overline{\mathbf{t}}^{\,\mathrm{H}}\overline{\mathbf{t}}\otimes\mathbf{t}^{\mathrm{H}}\mathbf{t} = \|\mathbf{t}\|^4$. It thus follows that $\hat{\mathbf{p}}^{\mathrm{H}}\mathbf{y} = 1 \Leftrightarrow \|\overline{\mathbf{t}}\otimes\mathbf{t}\| = 1$. Hence, up to a scaling that is not important,[1] the CMA(2,2) optimization problem is equivalent to solving

$$\mathbf{t} = \arg\min_{\substack{\mathbf{y}=\overline{\mathbf{t}}\otimes\mathbf{t}\\ \|\mathbf{y}\|=1}} \|\underline{\mathbf{G}}\mathbf{y}\|^2 = \arg\min_{\substack{\mathbf{y}=\overline{\mathbf{t}}\otimes\mathbf{t}\\ \|\mathbf{y}\|=1}} \mathbf{y}^{\mathrm{H}}\underline{\hat{\mathbf{C}}}\mathbf{y}. \tag{21}$$

and setting $\mathbf{w} = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s^{-1}\mathbf{t}$.

At this point, ACMA and CMA(2,2) diverge in two distinct but closely related directions.

— CMA(2,2) numerically optimizes the minimization problem in (21) and find d independent solutions. The solutions will be unit-norm vectors $\mathbf{y}$ that have the required Kronecker structure and minimize $\|\underline{\mathbf{G}}\mathbf{y}\|^2$. With noise, the solutions will not exactly be in the approximate nullspace of $\underline{\mathbf{G}}$ since this space will not admit the Kronecker structure.

— ACMA is making a twist on this problem. Instead of solving for the true minimum, it first finds an orthonormal basis $\mathbf{Y} = [\mathbf{y}_1, \ldots, \mathbf{y}_d]$ for the d-dimensional approximate nullspace of $\underline{\mathbf{G}}$ (or $\underline{\hat{\mathbf{C}}}$)

$$\mathbf{Y} = \arg\min_{\mathbf{Y}^{\mathrm{H}}\mathbf{Y}=\mathbf{I}} \|\underline{\mathbf{G}}\mathbf{Y}\|_{\mathrm{F}}^2 = \arg\min_{\mathbf{Y}^{\mathrm{H}}\mathbf{Y}=\mathbf{I}} \sum \mathbf{y}_i^{\mathrm{H}}\underline{\hat{\mathbf{C}}}\mathbf{y}_i$$

whose solution is the set of d least dominant eigenvectors of $\underline{\hat{\mathbf{C}}}$. It then looks for unit-norm vectors in this subspace that best fit the required structure $\mathbf{y} = \overline{\mathbf{t}} \otimes \mathbf{t}$ by considering

$$\mathbf{T} = \arg\min_{\mathbf{T},\mathbf{M}} \left\|\mathbf{Y} - (\overline{\mathbf{T}} \circ \mathbf{T})\mathbf{M}\right\|_{\mathrm{F}}^2.$$

We thus see that ACMA and CMA(2,2) are closely related, provided we whiten the data using the noisy data covariance matrix $\hat{\mathbf{R}}_x$. As we will show in Section VI, the two-step approach taken by ACMA makes it converge to the Wiener solution in (8) as $N \to \infty$, whereas CMA(2,2) is known to be close but generally unequal to the Wiener solution [9], [32].

V. Analysis of the Noise-Free Case

The analysis of ACMA in the noisefree case can be limited to an analysis of the solutions of $\hat{\mathbf{C}}_0\mathbf{y} = \mathbf{0}$, where $\hat{\mathbf{C}}_0 \equiv \hat{\mathbf{C}}$ is as defined in (15), and for future distinction, the "0" in $\hat{\mathbf{C}}_0$ is introduced to indicate that there is no noise. If the solutions $\{\mathbf{y}_i\}$ span the same subspace as spanned by $\{\overline{\mathbf{w}}_i \otimes \mathbf{w}_i;\ i = 1, \ldots, d\}$, then the joint diagonalization step is able to separate an arbitrary basis of the null space into its rank-1 components, and we recover the true beamformers.

With $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k$, we obtain from (15) that

$$\hat{\mathbf{C}}_0 = \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right]\hat{\mathbf{C}}_s\left[\overline{\mathbf{A}} \otimes \mathbf{A}\right]^{\mathrm{H}} \tag{22}$$

where

$$\hat{\mathbf{C}}_s := \frac{1}{N}\sum(\overline{\mathbf{s}}_k \otimes \mathbf{s}_k)(\overline{\mathbf{s}}_k \otimes \mathbf{s}_k)^{\mathrm{H}} - \frac{1}{N}\left[\sum \overline{\mathbf{s}}_k \otimes \mathbf{s}_k\right]\frac{1}{N}\left[\sum \overline{\mathbf{s}}_k \otimes \mathbf{s}_k\right]^{\mathrm{H}}. \tag{23}$$

$\hat{\mathbf{C}}_0$ is positive semidefinite because it is constructed as $\hat{\mathbf{C}}_0 = \mathbf{G}^{\mathrm{H}}\mathbf{G}$. Hence, the null space of $\hat{\mathbf{C}}_0$ has two components: the null space of $[\overline{\mathbf{A}}\otimes\mathbf{A}]^{\mathrm{H}}$ plus vectors $\mathbf{y}$ such that $[\overline{\mathbf{A}}\otimes\mathbf{A}]^{\mathrm{H}}\mathbf{y}$ is a vector in the null space of $\hat{\mathbf{C}}_s$. The purpose of prefiltering with dimension reduction is to remove the former solutions beforehand by working with $\underline{\hat{\mathbf{C}}}_0 = [\underline{\overline{\mathbf{A}}} \otimes \underline{\mathbf{A}}]\hat{\mathbf{C}}_s[\underline{\overline{\mathbf{A}}} \otimes \underline{\mathbf{A}}]^{\mathrm{H}}$, where $\underline{\mathbf{A}}$ is a square full-rank matrix. In that case, $\underline{\overline{\mathbf{A}}} \otimes \underline{\mathbf{A}}$ is also square full rank, with an empty null space. Thus, the interesting part is the analysis of the null space of $\hat{\mathbf{C}}_s$, which is only dependent on the signals and not on their mixing.

For the sake of exposition, we specialize $\hat{\mathbf{C}}_s$ for the case of two CM signals $s_1(k)$ and $s_2(k)$. Define

$$\rho := \frac{1}{N}\sum s_1(k)\overline{s}_2(k) \qquad a := 1 - |\rho|^2$$

$$q := \frac{1}{N}\sum [s_1(k)]^2[\overline{s}_2(k)]^2 \qquad b := q - \rho^2.$$

[1] as well as he fact that (if $M > d$) the prewhitening also involves a dimension reduction: This will force $\mathbf{w} = \hat{\mathbf{U}}_s\hat{\mathbf{\Sigma}}_s^{-1}\mathbf{t}$ to lie in the dominant column span of $\mathbf{X}$. We ignore this effect here.

Then (suppressing the time index)

$$\begin{aligned}\hat{\mathbf{C}}_s &= \frac{1}{N}\sum_k \begin{bmatrix}\bar{s}_1 s_1\\ \bar{s}_1 s_2\\ \bar{s}_2 s_1\\ \bar{s}_2 s_2\end{bmatrix}[s_1\bar{s}_1 \quad s_1\bar{s}_2 \quad s_2\bar{s}_1 \quad s_2\bar{s}_2]\\
&\quad - \frac{1}{N}\sum_k \begin{bmatrix}\bar{s}_1 s_1\\ \bar{s}_1 s_2\\ \bar{s}_2 s_1\\ \bar{s}_2 s_2\end{bmatrix}\frac{1}{N}\sum_k [s_1\bar{s}_1 \quad s_1\bar{s}_2 \quad s_2\bar{s}_1 \quad s_2\bar{s}_2]\\
&= \begin{bmatrix}1 & \rho & \bar{\rho} & 1\\ \bar{\rho} & 1 & \bar{q} & \bar{\rho}\\ \rho & q & 1 & \rho\\ 1 & \rho & \bar{\rho} & 1\end{bmatrix} - \begin{bmatrix}1\\ \bar{\rho}\\ \rho\\ 1\end{bmatrix}[1 \quad \rho \quad \bar{\rho} \quad 1]\\
&= \begin{bmatrix}0 & 0 & 0 & 0\\ 0 & a & \bar{b} & 0\\ 0 & b & a & 0\\ 0 & 0 & 0 & 0\end{bmatrix}. \end{aligned} \tag{24}$$

We immediately see that $\hat{\mathbf{C}}_s$ has null space vectors

$$\begin{bmatrix}1\\0\\0\\0\end{bmatrix}, \qquad \begin{bmatrix}0\\0\\0\\1\end{bmatrix}. \tag{25}$$

These are the desired null space vectors. The remaining 2×2 submatrix in the center of (24) is hopefully nonsingular. If the sources are independent and circularly symmetric, then asymptotically (in N) $q \to 0$ and $\rho \to 0$ so that $a \to 1$ and $b \to 0$. Thus, for a sufficiently large number of samples, it is clear that the submatrix is nonsingular with probability 1. Singularity occurs almost surely only with BPSK-type signals (for which $a = b$) [29], and for this case, a modified algorithm called RACMA has to be used to avoid the additional solutions [27].

The null space of $\hat{\mathbf{C}}_0$ contains vectors $\mathbf{y}$ for which $[\overline{\mathbf{A}} \otimes \mathbf{A}]^{\mathrm{H}}\mathbf{y}$ is a vector in the null space of $\hat{\mathbf{C}}_s$, i.e., either vector in (25). Assuming that $\mathbf{A}$ has full column rank, $\overline{\mathbf{A}} \otimes \mathbf{A}$ also has full column rank. Let $\mathbf{W} = [\mathbf{w}_1 \quad \mathbf{w}_2]$ be a separating beamformer such that $\mathbf{W}^{\mathrm{H}}\mathbf{A} = \mathbf{I}$; then

$$[\overline{\mathbf{A}} \otimes \mathbf{A}]^{\mathrm{H}}[\overline{\mathbf{W}} \otimes \mathbf{W}] = \overline{\mathbf{A}}^{\mathrm{H}}\overline{\mathbf{W}} \otimes \mathbf{A}^{\mathrm{H}}\mathbf{W} = \mathbf{I} \otimes \mathbf{I} = \mathbf{I}$$

from which we see that

$$[\overline{\mathbf{A}} \otimes \mathbf{A}]^{\mathrm{H}}(\overline{\mathbf{w}}_1 \otimes \mathbf{w}_1) = \begin{bmatrix}1\\0\\0\\0\end{bmatrix}$$

$$[\overline{\mathbf{A}} \otimes \mathbf{A}]^{\mathrm{H}}(\overline{\mathbf{w}}_2 \otimes \mathbf{w}_2) = \begin{bmatrix}0\\0\\0\\1\end{bmatrix}.$$

The solutions to $\mathbf{y}^{\mathrm{H}}\hat{\mathbf{C}}_0\mathbf{y} = 0$ are thus spanned by a basis of the null space of $[\overline{\mathbf{A}} \otimes \mathbf{A}]^{\mathrm{H}}$ (removed by prefiltering with dimension reduction) plus linear combinations of the desired solutions

$$\overline{\mathbf{w}}_1 \otimes \mathbf{w}_1, \qquad \overline{\mathbf{w}}_2 \otimes \mathbf{w}_2.$$

If only the desired solutions are present in the null space of $\hat{\mathbf{C}}_0$, then the joint diagonalization step can find them from an arbitrary basis of this subspace.

The above analysis easily generalizes to more than two signals. A key property that is valid for any number of signals and explicitly used by the algorithm is the fact that certain columns (and rows) of $\hat{\mathbf{C}}_s$ are identically zero. This property comes from $|s_k|^2 = 1$ alone and follows *by construction* for any number of samples. We do not have to wait for asymptotic convergence of the cross terms to zero. Many other blind source separation techniques require stochastic independence and rely on this. This aspect is the key to the good small-sample performance that can be achieved with constant modulus signals.

VI. Asymptotic Behavior of ACMA with Noise

An analysis of the asymptotic behavior of ACMA in noise will reveal the close connections of this method with other blind source separation methods based on fourth-order moments. Assume that we compute $\hat{\mathbf{C}}$ in the same way as in (15). As $N \to \infty$, $\hat{\mathbf{C}}$ converges to

$$\begin{aligned}\mathbf{C} := &\mathrm{E}\left[\overline{\mathbf{x}}_k \otimes \mathbf{x}_k)(\overline{\mathbf{x}}_k \otimes \mathbf{x}_k)^{\mathrm{H}}\right]\\ &- \mathrm{E}\left[\sum \overline{\mathbf{x}}_k \otimes \mathbf{x}_k\right]\mathrm{E}\left[\sum \overline{\mathbf{x}}_k \otimes \mathbf{x}_k\right]^{\mathrm{H}}.\end{aligned} \tag{26}$$

We will first analyze the structure of $\mathbf{C}$ in terms of the data model $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k$.

A. Cumulants

The asymptotic analysis requires the introduction of fourth-order cumulants. For a zero-mean stochastic vector $\mathbf{x}$ with components x_i, define the fourth-order cumulants

$$\begin{aligned}\kappa_{i,k}^{j,l} &:= \mathrm{cum}(x_i, \bar{x}_j, x_k, \bar{x}_l)\\ &:= \mathrm{E}(x_i\bar{x}_j x_k\bar{x}_l) - \mathrm{E}(x_i\bar{x}_j)\mathrm{E}(x_k\bar{x}_l)\\ &\quad - \mathrm{E}(x_i\bar{x}_l)\mathrm{E}(x_k\bar{x}_j) - \mathrm{E}(x_i x_k)\mathrm{E}(\bar{x}_j\bar{x}_l)\end{aligned}$$

where $i, j, k, l = 1, \ldots, d$, and d is the dimension of $\mathbf{x}$. We assume circularly symmetric sources (hence non-BPSK) so that the last term vanishes. If we collect the $\kappa_{i,k}^{j,l}$ into a matrix $\mathbf{K}_x$ with entries $\mathbf{K}_{i+jd,\, l+kd} = \kappa_{i,k}^{j,l}$, then

$$\begin{aligned}\mathbf{K}_x = &\mathrm{E}\left[(\overline{\mathbf{x}} \otimes \mathbf{x})(\overline{\mathbf{x}} \otimes \mathbf{x})^{\mathrm{H}}\right] - \mathrm{E}[\overline{\mathbf{x}} \otimes \mathbf{x}]\mathrm{E}[\overline{\mathbf{x}} \otimes \mathbf{x}]^{\mathrm{H}}\\ &- \mathrm{E}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right]^{\mathrm{T}} \otimes \mathrm{E}\left[\mathbf{x}\mathbf{x}^{\mathrm{H}}\right].\end{aligned}$$

Note that $\mathrm{E}[\mathbf{x}\mathbf{x}^{\mathrm{H}}] = \mathbf{R}_x$, $\mathrm{E}[\overline{\mathbf{x}} \otimes \mathbf{x}] = \mathrm{vec}(\mathbf{R}_x)$. Compared with (26), it is seen that

$$\mathbf{C} = \mathbf{K}_x + \mathbf{R}_x^{\mathrm{T}} \otimes \mathbf{R}_x. \tag{27}$$

Cumulants have several well-known nice properties, such as multilinearity, addivity, and the fact that Gaussian signals have zero cumulants. For our model $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k$, define $\mathbf{s}_k = \sum s_i(k)\mathbf{e}_i$, where $\mathbf{e}_i$ is the ith unit coordinate vector. Let us also define the auto-cumulants

$$\kappa_i = \text{cum}(s_i, \bar{s}_i, s_i, \bar{s}_i).$$

Assuming independent signals, additivity implies

$$\mathbf{K}_s = \sum_{i=1}^{d} \kappa_i [\mathbf{e}_i \otimes \mathbf{e}_i] [\mathbf{e}_i \otimes \mathbf{e}_i]^{\mathrm{T}}.$$

Circularly symmetric CM signals have autocumulants $\kappa_i = -1$. Further assuming independent Gaussian noise ($\mathbf{K}_n = 0$), we obtain

$$\begin{aligned}\mathbf{K}_x &= \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right] \mathbf{K}_s \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right]^{\mathrm{H}} + \mathbf{K}_n \\ &= \left[\overline{\mathbf{A}} \circ \mathbf{A}\right] (-\mathbf{I}) \left[\overline{\mathbf{A}} \circ \mathbf{A}\right]^{\mathrm{H}}. \end{aligned} \tag{28}$$

Using these properties, we can derive that *without noise* (or $\mathbf{R}_x = \mathbf{A}\mathbf{A}^{\mathrm{H}}$), the CMA(2,2) or ACMA criterion matrix $\hat{\mathbf{C}} = \hat{\mathbf{C}}_0$ converges asymptotically in N to

$$\begin{aligned}\mathbf{C}_0 &= \mathbf{K}_x + \mathbf{R}_x^{\mathrm{T}} \otimes \mathbf{R}_x \\ &= \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right] \mathbf{K}_s \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right]^{\mathrm{H}} + \overline{\mathbf{A}}\,\overline{\mathbf{A}}^{\mathrm{H}} \otimes \mathbf{A}\mathbf{A}^{\mathrm{H}} \\ &= \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right] (\mathbf{K}_s + \mathbf{I}) \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right]^{\mathrm{H}} \\ &= \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right] \mathbf{C}_s \left[\overline{\mathbf{A}} \otimes \mathbf{A}\right]^{\mathrm{H}}. \end{aligned} \tag{29}$$

Note that $\mathbf{C}_s = \mathbf{K}_s + \mathbf{I}$ is diagonal, with zero entries at the location of the source autocumulants, and "1" entries elsewhere on the diagonal. Like in the finite sample case, the null space of $\mathbf{C}_s$ is given by $\{\mathbf{e}_i \otimes \mathbf{e}_i\}$, and hence, the null space of $\mathbf{C}$ by $\{\overline{\mathbf{w}}_i \otimes \mathbf{w}_i\}$, plus the null space of $[\overline{\mathbf{A}} \otimes \mathbf{A}]^{\mathrm{H}}$ (this is removed by prefiltering).

With noise (or $\mathbf{R}_x = \mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{R}_n$), $\hat{\mathbf{C}}$ converges asymptotically in N to

$$\begin{aligned}\mathbf{C} &= \mathbf{K}_x + \mathbf{R}_x^{\mathrm{T}} \otimes \mathbf{R}_x \\ &= \mathbf{K}_x + \left(\mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{R}_n\right)^{\mathrm{T}} \otimes \left(\mathbf{A}\mathbf{A}^{\mathrm{H}} + \mathbf{R}_n\right) \\ &= \mathbf{C}_0 + \mathbf{E} + \mathbf{C}_n \end{aligned}$$

where $\mathbf{C}_0$ is given in (29), and

$$\mathbf{E} := \overline{\mathbf{A}}\,\overline{\mathbf{A}}^{\mathrm{H}} \otimes \mathbf{R}_n + \mathbf{R}_n^{\mathrm{T}} \otimes \mathbf{A}\mathbf{A}^{\mathrm{H}}, \qquad \mathbf{C}_n := \mathbf{R}_n^{\mathrm{T}} \otimes \mathbf{R}_n. \tag{30}$$

Thus, the noise contributes a second-order and a fourth-order term to the ACMA criterion matrix $\mathbf{C}$, even if the noise has zero fourth-order cumulants. If we do not correct for it and proceed as in the noise-free case, this will result in a certain bias at the output of the beamformer. As we show next, this bias is precisely such that ACMA converges to the Wiener solution.

B. Asymptotic Analysis of ACMA

In the analysis of ACMA, we also have to take the effect of the initial prewhitening step into account. Recall that this step is $\underline{\mathbf{X}} = \mathbf{F}^{\mathrm{H}}\mathbf{X}$ such that $\hat{\underline{\mathbf{R}}}_x = \mathbf{F}^{\mathrm{H}}\hat{\mathbf{R}}_x\mathbf{F} = \mathbf{I}$. Introducing this into the expression for $\mathbf{C}$ in (27) and using (28), we obtain that asymptotically ($N \to \infty$), $\hat{\underline{\mathbf{C}}}$ converges to

$$\begin{aligned}\underline{\mathbf{C}} &= (\mathbf{F} \otimes \mathbf{F})^{\mathrm{H}} \mathbf{C} (\mathbf{F} \otimes \mathbf{F}) \\ &= (\mathbf{F} \otimes \mathbf{F})^{\mathrm{H}} \mathbf{K}_x (\mathbf{F} \otimes \mathbf{F}) + \mathbf{I} \otimes \mathbf{I} \\ &= \left[\overline{\underline{\mathbf{A}}} \otimes \underline{\mathbf{A}}\right] \mathbf{K}_s \left[\overline{\underline{\mathbf{A}}} \otimes \underline{\mathbf{A}}\right]^{\mathrm{H}} + \mathbf{I} \\ &= \left[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}\right] (-\mathbf{I}) \left[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}\right]^{\mathrm{H}} + \mathbf{I}. \end{aligned} \tag{31}$$

Consequently, the CMA(2,2) cost function (21) becomes as $N \to \infty$

$$\begin{aligned} &\underset{\mathbf{y}=\bar{\mathbf{t}}\otimes\mathbf{t},\, \|\mathbf{y}\|=1}{\arg\min}\ \mathbf{y}^{\mathrm{H}} \underline{\mathbf{C}} \mathbf{y} \\ &= \underset{\mathbf{y}=\bar{\mathbf{t}}\otimes\mathbf{t},\, \|\mathbf{y}\|=1}{\arg\min}\ \mathbf{y}^{\mathrm{H}} \left\{ \left[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}\right] (-\mathbf{I}) \left[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}\right]^{\mathrm{H}} + \mathbf{I} \right\} \mathbf{y} \\ &= \underset{\mathbf{y}=\bar{\mathbf{t}}\otimes\mathbf{t},\, \|\mathbf{y}\|=1}{\arg\max}\ \mathbf{y}^{\mathrm{H}} \left\{ \left[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}\right] \left[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}\right]^{\mathrm{H}} \right\} \mathbf{y}. \end{aligned} \tag{32}$$

Unlike CMA(2,2), ACMA does not optimize (32) directly but solves the unstructured problem first. Indeed, it looks for an unconstrained d-dimensional basis $\{\mathbf{y}_i\}$ of the null space of $\underline{\mathbf{C}}$ or, equivalently, d dominant eigenvectors of $[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}]\,[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}]^{\mathrm{H}}$. Since this is a rank-$d$ matrix, we have that the d dominant eigenvectors together span the same subspace as the column span of $[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}]$; hence, asymptotically ($N \to \infty$)

$$\begin{aligned} &\text{span}\{\mathbf{y}_1, \ldots, \mathbf{y}_d\} \\ &\quad = \text{span}\left[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}\right] = \text{span}\{\overline{\underline{\mathbf{a}}}_1 \otimes \underline{\mathbf{a}}_1, \ldots, \overline{\underline{\mathbf{a}}}_d \otimes \underline{\mathbf{a}}_d\}. \end{aligned}$$

As a second step, the joint diagonalization procedure is used to replace the unstructured basis by one that has the required Kronecker product structure, i.e., d independent vectors of the form $\overline{\mathbf{t}} \otimes \mathbf{t}$ within this column span. From the above equation, we see that the unique solution is $\overline{\mathbf{t}}_i \otimes \mathbf{t}_i = \overline{\underline{\mathbf{a}}}_i \otimes \underline{\mathbf{a}}_i$ (up to a scaling to make $\mathbf{t}_i$ have unit norm), and thus

$$\mathbf{t}_i = \underline{\mathbf{a}}_i, \qquad i = 1, \ldots, d.$$

Hence, the beamformer on the whitened problem is equal to the whitened direction vector (a matched spatial filter). If we go back to the resulting beamformer on the original (unwhitened) data matrix $\mathbf{X}$, we find (for $i = 1, \ldots, d$)

$$\mathbf{t}_i = \underline{\mathbf{a}}_i = \mathbf{F}^{\mathrm{H}}\mathbf{a}_i \quad \Rightarrow \quad \mathbf{w}_i = \mathbf{F}\mathbf{t}_i = \mathbf{F}\mathbf{F}^{\mathrm{H}}\mathbf{a}_i = \mathbf{R}_x^{-1}\mathbf{a}_i \tag{33}$$

since $\mathbf{F} = \mathbf{U}_s\mathbf{\Sigma}_s^{-1}$, $\mathbf{R}_x^{-1} = \mathbf{U}\mathbf{\Sigma}^{-2}\mathbf{U}^{\mathrm{H}} = \mathbf{U}_s\mathbf{\Sigma}_s^{-2}\mathbf{U}_s^{\mathrm{H}} + \sigma^{-2}\mathbf{U}_n\mathbf{U}_n^{\mathrm{H}}$, and $\mathbf{U}_n^{\mathrm{H}}\mathbf{a}_i = \mathbf{0}$. We have just shown that as $N \to \infty$, the beamformers provided by ACMA converge to the Wiener receivers (8). In general, this is a very attractive property.

Does this two-step procedure solve the CMA(2,2) optimization problem (32)? This is not likely since in this asymptotic case, ACMA finds its structured solutions only inside the subspace spanned by the columns of $[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}]$. A solution to CMA(2,2) is expected to be close to a dominant eigenvector of $[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}]\,[\overline{\underline{\mathbf{A}}} \circ \underline{\mathbf{A}}]^{\mathrm{H}}$, but it is not restricted to be inside the subspace. Thus, if the eigenvectors are not equal to $\{\overline{\underline{\mathbf{a}}}_i \otimes \underline{\mathbf{a}}_i\}$, the CMA(2,2) optimal solution might be different. This happens if the columns of $\underline{\mathbf{A}}$ are not orthogonal, but there are only two situations where the columns of $\underline{\mathbf{A}}$ are precisely orthogonal: if there is no noise, or (assuming white Gaussian noise) if the columns of the unwhitened $\mathbf{A}$ are orthogonal. This is a rather special case, approximately true if the sources are well separated, and the number of sensors is large. Thus, CMA(2,2) does in general not lead to the Wiener solution. This result matches that in the equalization context [9]. Obviously, if the noise is small, then the discrepancy will be small as well.

C. Connection to JADE

JADE [4] is a widely used algorithm for the blind separation of independent non-Gaussian sources in white Gaussian noise. It is based on the construction of the fourth-order cumulant matrix $\mathbf{K}_x$ in (28) but uses a different prefiltering strategy, namely, $\mathbf{F} = \hat{\mathbf{U}}_s(\hat{\mathbf{\Sigma}}_s^2 - \sigma^2\mathbf{I})^{-1/2}$, where $\hat{\mathbf{U}}_s$ and $\hat{\mathbf{\Sigma}}_s$ are estimated from the eigenvalue decomposition of $\hat{\mathbf{R}}_x$. The prefiltering leads to $\underline{\underline{\mathbf{X}}} = \mathbf{F}^{\mathrm{H}}\mathbf{X} = \underline{\underline{\mathbf{A}}}\mathbf{S} + \underline{\underline{\mathbf{N}}}$, where $\underline{\underline{\mathbf{A}}} = \mathbf{F}^{\mathrm{H}}\mathbf{A}$. This choice is motivated by the fact that as $N \to \infty$, $\mathbf{F}$ converges to $\mathbf{F} = \mathbf{U}_A\mathbf{\Sigma}_A^{-1}$, where $\mathbf{U}_A$ and $\mathbf{\Sigma}_A$ are minimal-size factors of the SVD of $\mathbf{A} = \mathbf{U}_A\mathbf{\Sigma}_A\mathbf{V}_A$, and thus

$$\underline{\underline{\mathbf{A}}} = \mathbf{\Sigma}_A^{-1}\mathbf{U}_A^{\mathrm{H}}\mathbf{A} = \mathbf{V}_A$$

which is a unitary matrix. Asymptotically, the sample fourth-order cumulant matrix $\underline{\underline{\hat{\mathbf{K}}}}_x$ converges to

$$\underline{\underline{\mathbf{K}}}_x = \left[\underline{\underline{\bar{\mathbf{A}}}} \circ \underline{\underline{\mathbf{A}}}\right] (-\mathbf{I}) \left[\underline{\underline{\bar{\mathbf{A}}}} \circ \underline{\underline{\mathbf{A}}}\right]^{\mathrm{H}}.$$

JADE computes a basis of the dominant column span of $\underline{\underline{\hat{\mathbf{K}}}}_x$, which in the asymptotic situation spans the same subspace as

$$\left\{\underline{\underline{\bar{\mathbf{a}}}}_i \otimes \underline{\underline{\mathbf{a}}}_i; \qquad i = 1, \ldots, d\right\}$$

Like ACMA, it then performs a joint diagonalization to identify the vectors $\underline{\underline{\mathbf{a}}}_i$. After correcting for the prefiltering, we find

$$\mathbf{T} = \underline{\underline{\mathbf{A}}} = \mathbf{V}_A \quad \Rightarrow \quad \mathbf{W} = \mathbf{F}\mathbf{T} = \mathbf{U}_A\mathbf{\Sigma}_A^{-1}\mathbf{V}_A = \mathbf{A}^{\dagger\mathrm{H}}.$$

Hence, this strategy leads asymptotically to the zero-forcing beamformer [cf. (10)], as well as the true $\mathbf{A}$-matrix.

Apart from different prefiltering, the asymptotic equations of JADE and ACMA look rather similar. JADE searches for eigenvectors corresponding to nonzero eigenvalues given by the nonzero entries of $\mathbf{K}_s$, which, here, are equal to -1, whereas ACMA looks for the null space vectors generated by the zero entries of $\mathbf{K}_s + \mathbf{I}$. The result is the same.

However, the finite-sample properties are quite different. In the absence of noise, the null space information of $\hat{\mathbf{C}}$ in ACMA is exact by construction, and hence, the algorithm produces the exact separating beamformers. The dominant column span of $\underline{\underline{\hat{\mathbf{K}}}}_x$ used in JADE is not exact since the signal sources do not decorrelate exactly in finite samples: $\hat{\mathbf{K}}_s$ is a full matrix. Thus, keeping the number of samples fixed, the performance of JADE saturates as SNR $\to \infty$.

Furthermore, in the proposed implementation in [4], JADE explicitly uses the fact that (with the $\mathbf{\Sigma}_A^{-1}\mathbf{U}_A^{\mathrm{H}}$-prefiltering) $\underline{\underline{\mathbf{A}}} = \mathbf{V}_A$ and is hence unitary. It thus forces the joint diagonalization to produce a unitary matrix. A finite-sample problem is that $\hat{\mathbf{R}}_x$ does not reveal yet the true $\mathbf{U}_A$ and $\mathbf{\Sigma}_A$, and the restriction might make the results less accurate. This problem was noted in [1], where optimal combinations of second- and fourth-order statistics are presented.

In summary, we can say that JADE and ACMA are quite similar but differ in the following points.

- The prefiltering scheme ACMA, such that as $N \to \infty$, converges to a Wiener solution and JADE to a zero-forcing beamformer.
- JADE explicitly relies on stochastic independence of sources, whereas ACMA explicitly relies on the CM property. This leads to different finite sample behavior.
- JADE, as in [4], forces the unitarity of $\underline{\underline{\mathbf{A}}}$, which leads to saturation of the performance for large SNRs and finite number of samples.

Given data $\mathbf{X}$, estimate a zero-forcing beamformer $\mathbf{W}$

1. SVD: $\mathbf{X} =: \hat{\mathbf{U}}\hat{\mathbf{\Sigma}}\hat{\mathbf{V}}$
   Prefiltering: $\underline{\mathbf{X}} := \hat{\mathbf{\Sigma}}_s^{-1}\hat{\mathbf{U}}_s^{\mathrm{H}}\mathbf{X}$
   $\underline{\hat{\mathbf{C}}} := \frac{1}{N}\sum(\bar{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k)(\bar{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k)^{\mathrm{H}} + {}$
   $\quad - \left[\frac{1}{N}\sum \bar{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k\right]\left[\frac{1}{N}\sum \bar{\underline{\mathbf{x}}}_k \otimes \underline{\mathbf{x}}_k\right]^{\mathrm{H}}.$
   $\underline{\hat{\mathbf{E}}} := \mathbf{I} \otimes \hat{\mathbf{\Sigma}}_s^{-2} + \hat{\mathbf{\Sigma}}_s^{-2} \otimes \mathbf{I}$
   GEVD of $(\underline{\hat{\mathbf{C}}}, \underline{\hat{\mathbf{E}}})$, let $\{\mathbf{y}_i\}$ be the d least dominant eigenvectors.
2. Continue as in the usual ACMA, step 2 (figure 2)

Fig. 3. ZF-ACMA.

VII. Zero-Forcing ACMA

We have seen before, in (30), that as $N \to \infty$, $\hat{\mathbf{C}} \to \mathbf{C} = \mathbf{C}_0 + \mathbf{E} + \mathbf{C}_n$, where $\mathbf{C}_0$ is the noise-free part, and $\mathbf{E}$ and $\mathbf{C}_n$ represent noise bias terms that cause ACMA to converge to the Wiener solution $\mathbf{W} = \mathbf{R}_x^{-1}\mathbf{A}$. If an unbiased estimate of $\mathbf{A}$ is desired (e.g., for direction estimation), then we cannot simply invert $\mathbf{W}$, as is usually done. We could map $\mathbf{W}$ to an unbiased estimate of $\mathbf{A}$ via premultiplication by $\hat{\mathbf{R}}_x$, but the finite-sample properties of this appear not to be very good. Here, we look at an alternative, based on estimating and removing the noise terms from $\hat{\mathbf{C}}$, to obtain an estimate of $\mathbf{C}_0$. This technique was first presented in [28].

Let us assume that we know the noise covariance $\mathbf{R}_n$. We cannot know $\mathbf{E}$ since it depends on noise-free data, but we can construct

$$\hat{\mathbf{E}} := \hat{\mathbf{R}}_x^{\mathrm{T}} \otimes \mathbf{R}_n + \mathbf{R}_n^{\mathrm{T}} \otimes \hat{\mathbf{R}}_x. \tag{34}$$

When $N \to \infty$

$$\hat{\mathbf{E}} \quad \to \quad \overline{\mathbf{A}}\,\overline{\mathbf{A}}^{\mathrm{H}} \otimes \mathbf{R}_n + \mathbf{R}_n^{\mathrm{T}} \otimes \mathbf{A}\mathbf{A}^{\mathrm{H}} + 2\mathbf{R}_n^{\mathrm{T}} \otimes \mathbf{R}_n$$

so that

$$\hat{\mathbf{C}} - \hat{\mathbf{E}} + \mathbf{R}_n^{\mathrm{T}} \otimes \mathbf{R}_n \quad \to \quad \mathbf{C}_0$$

is an asymptotically unbiased estimate of $\mathbf{C}_0$. If we can assume that $\|\mathbf{R}_n\|_{\mathrm{F}}^2 \ll \|\mathbf{A}\mathbf{A}^{\mathrm{H}}\|_{\mathrm{F}}^2$, i.e., the SNR is sufficiently large, then we can ignore $\mathbf{R}_n^{\mathrm{T}} \otimes \mathbf{R}_n$ compared with $\hat{\mathbf{E}}$ and use $\hat{\mathbf{C}} - \hat{\mathbf{E}}$ to estimate $\mathbf{C}_0$.

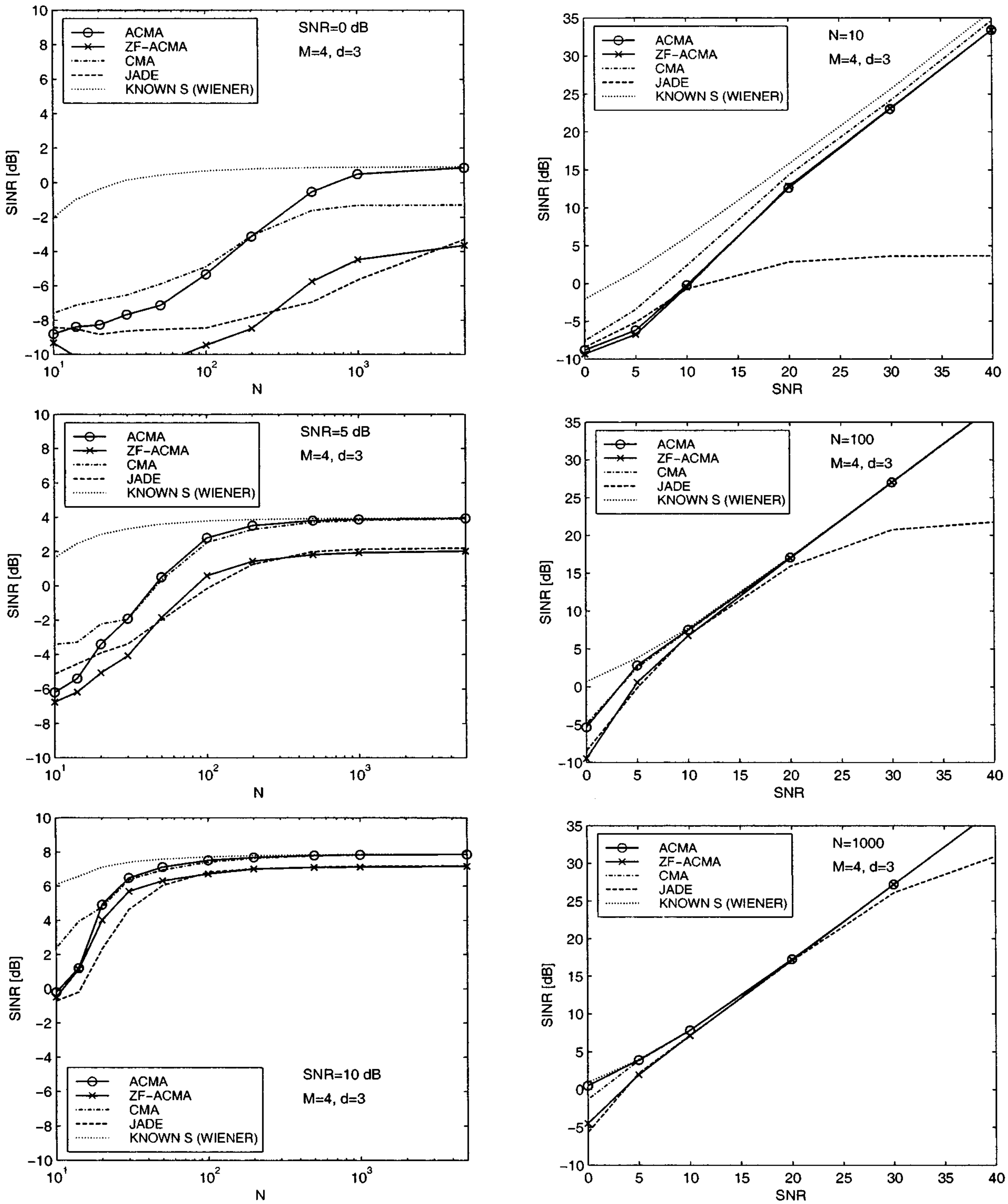

Fig. 4. SINR performance of ACMA, ZF-ACMA, CMA, and JADE.

Let us now assume that the noise is white with covariance $\mathbf{R}_n = \sigma^2\mathbf{I}$ but that the noise power σ^2 is unknown. Redefining $\hat{\mathbf{E}}$ as

$$\hat{\mathbf{E}} := \hat{\mathbf{R}}_x^{\mathrm{T}} \otimes \mathbf{I} + \mathbf{I} \otimes \hat{\mathbf{R}}_x$$

it follows that we have available the data matrices $\hat{\mathbf{C}}$ and $\hat{\mathbf{E}}$, satisfying the approximate model (ignoring fourth-order terms)

$$\hat{\mathbf{C}} - \sigma^2\hat{\mathbf{E}} \simeq \mathbf{C}_0.$$

Since $\mathbf{C}_0$ is rank deficient with a kernel of dimension d, we can estimate σ^2 as the (average of the) smallest d eigenvalues of the matrix pencil $(\hat{\mathbf{C}}, \hat{\mathbf{E}})$, corresponding to the generalized eigenvalue equation

$$\left(\hat{\mathbf{C}} - \lambda\hat{\mathbf{E}}\right)\mathbf{y} = \mathbf{0}.$$

An estimate of the basis $\{\mathbf{y}_i\}$ of the kernel of $\mathbf{C}_0$ is given by the corresponding eigenvectors. At this point, we can continue with

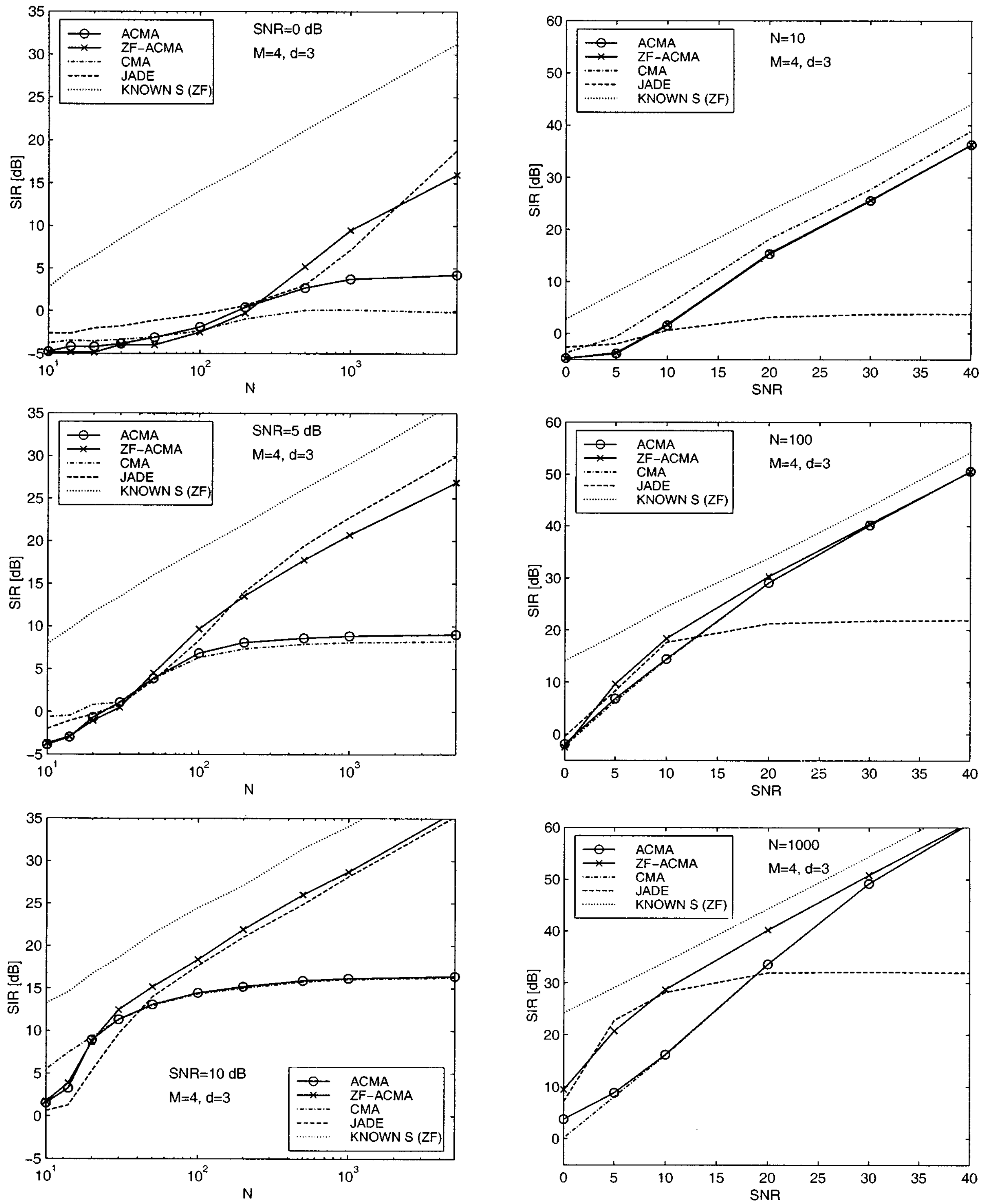

Fig. 5. SIR performance of ACMA, ZF-ACMA, CMA, and JADE, as function of N and SNR.

the joint diagonalization and recover the beamforming matrix $\mathbf{W}$. Asymptotically as both $N \to \infty$ and SNR $\to \infty$, we obtain $\mathbf{W} \to \mathbf{A}^{\dagger \mathrm{H}}$.

The algorithm is called ZF-ACMA.[2] As in ACMA, a dimension-reducing prefiltering $\mathbf{F}$ is necessary. If we take the same prewhitening prefilter as in ACMA, then after whitening, $\hat{\underline{\mathbf{R}}}_x = \mathbf{I}$, and $\underline{\mathbf{R}}_n = \sigma^2 \hat{\mathbf{\Sigma}}_s^{-2}$. Thus, $\hat{\underline{\mathbf{E}}} = \mathbf{I} \otimes \hat{\mathbf{\Sigma}}_s^{-2} + \hat{\mathbf{\Sigma}}_s^{-2} \otimes \mathbf{I}$ is diagonal and constructed without much additional effort. The resulting algorithm is summarized in Fig. 3.

[2]It was first presented as W-ACMA in [28].

VIII. Simulations

Some performance results are shown in Figs. 4–6. In the simulations, we took a uniform linear array with $M = 4$ antennas spaced at half wavelengths, and $d = 3$ equal-power constant-modulus sources with directions $-10°$, $0°$, $20°$ respectively. We compare the performance of ACMA, ZF-ACMA, JADE, and CMA.

ACMA as used here is the original algorithm as presented in [29], which is almost as the algorithm presented here, except that the joint diagonalization is implemented as a joint Schur decomposition, with perhaps slightly different results. The CMA used for reference is obtained as the numerically determined optimum of the deterministic CMA(2,2) cost function (20), for d independent beamformers, using a gradient technique initialized by the sample data Wiener receiver $\hat{\mathbf{W}} = (\mathbf{S}\mathbf{X}^{\dagger})^{\mathrm{H}}$ with known $\mathbf{S}$. Note that this is not a practical algorithm; it describes the best performance of CMA for a block of N data samples and may not be achieved by the usual sample-adaptive algorithm.

In Fig. 4, we vary the number of samples N and the SNR. The performance measure is the residual signal-to-interference-plus-noise ratio (SINR) at the output of the beamformers. We only consider the SINR of the worst output channel and find a permutation $\mathbf{\Pi}$ that maximizes this. Specifically, the SINR is defined here as

$$\begin{aligned}\mathrm{sinr}(\mathbf{a}, \mathbf{w}) &:= \frac{\mathbf{w}^{\mathrm{H}}\left(\mathbf{a}\mathbf{a}^{\mathrm{H}}\right)\mathbf{w}}{\mathbf{w}^{\mathrm{H}}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}} - \mathbf{a}\mathbf{a}^{\mathrm{H}} + \sigma^2\mathbf{I}\right)\mathbf{w}} \\ \mathrm{SINR}(\mathbf{A}, \mathbf{W}) &:= [\mathrm{sinr}(\mathbf{a}_1, \mathbf{w}_1) \quad \cdots \quad \mathrm{sinr}(\mathbf{a}_d, \mathbf{w}_d)] \\ \mathrm{SINR} &:= \max_{\mathbf{\Pi}} \min \mathrm{SINR}(\mathbf{A}, \mathbf{W\Pi}).\end{aligned}$$

The reference performance is that of a Wiener receiver based on sample data with known $\mathbf{S}$, i.e., $\hat{\mathbf{W}} = (\mathbf{S}\mathbf{X}^{\dagger})^{\mathrm{H}}$, as in (7). As seen from the left column of the figure, ACMA converges asymptotically (in N) to the Wiener beamformer. CMA is known theoretically not to reach this performance, but it is seen that for positive SNR, the performance is almost identical to that of ACMA. The right column of the figure shows that the SINR performance of JADE saturates as function of SNR (as predicted). CMA, ACMA, and ZF-ACMA do not have this problem.

Fig. 5 shows the signal-to-interference ratio (SIR) performance, which is defined similarly as

$$\begin{aligned}\mathrm{sir}(\mathbf{a}, \mathbf{w}) &:= \frac{\mathbf{w}^{\mathrm{H}}(\mathbf{a}\mathbf{a}^{\mathrm{H}})\mathbf{w}}{\mathbf{w}^{\mathrm{H}}(\mathbf{A}\mathbf{A}^{\mathrm{H}} - \mathbf{a}\mathbf{a}^{\mathrm{H}})\mathbf{w}} \\ \mathrm{SIR}(\mathbf{A}, \mathbf{W}) &:= [\mathrm{sir}(\mathbf{a}_1, \mathbf{w}_1) \quad \cdots \quad \mathrm{sir}(\mathbf{a}_d, \mathbf{w}_d)] \\ \mathrm{SIR} &:= \max_{\mathbf{\Pi}} \min \mathrm{SIR}(\mathbf{A}, \mathbf{W\Pi}).\end{aligned}$$

This indicates how well the computed beamforming matrix $\mathbf{W}$ is an inverse of $\mathbf{A}$, up to an arbitrary permutation. The reference performance is that of a zero-forcing (ZF) beamformer based on sample data with known $\mathbf{S}$, as given in (9).

It is seen that the SIR performance of ACMA saturates as function of N (for finite SNR) because it converges to the Wiener solution, and hence, it is biased. The whitening in ZF-ACMA removes this saturation so that it can converge to a few decibels below the ZF solution. As for the SINR, the SIR performance of JADE saturates as function of SNR.

If our objective is direction of arrival estimation, then we can first estimate $\mathbf{W}$ using ACMA or ZF-ACMA, compute $\hat{\mathbf{A}} = \mathbf{W}^{\dagger\mathrm{H}}$, and estimate the directions from the individual columns of $\hat{\mathbf{A}}$. This technique was proposed in [13]. Fig. 6 shows a test of this, in a scenario with three equal powered sources with directions $[-10,\, 0,\, 20]$, for varying N, and an SNR of 10 dB. The graph shows the root mean squared error of the DOA estimate of the first source and the Cramer–Rao bound (CRB) for this model [13]. It is seen that the estimate from ACMA is biased so that its performance saturates as $N \to \infty$, whereas the estimates from ZF-ACMA and JADE are asymptotically error free. ZF-ACMA has a small advantage over JADE, which is to be expected since more information on the sources is used.

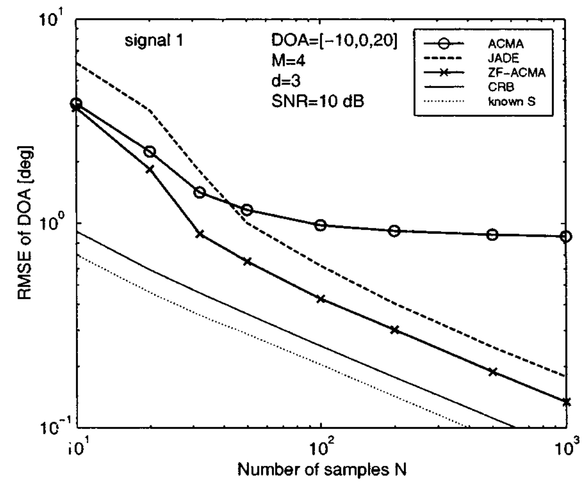

Fig. 6. DOA estimation performance for varying N.

IX. Concluding Remarks

We have shown that ACMA converges to the Wiener solution (in samples or SNR), whereas the minima of the CMA(2,2) cost function only have this property if there is no noise or the mixing matrix is orthogonal. However, for positive SNR, the differences in SINR performance are rather insignificant.

Furthermore, we have derived a modification (ZF-ACMA) which is close to the zero-forcing solution if the noise power is small (say SNR better than 10 dB). We have made a performance comparison with the related JADE algorithm, which separates independent non-Gaussian sources based on their nonzero kurtosis. The conclusion is not unequivocal because JADE converges to a zero-forcing beamformer asymptotically in the number of samples but not in SNR. In the simulation example, we saw that for $N > 20$ samples and SNR > 10 dB, ZF-ACMA has the best SIR performance, and ACMA has the best SINR performance.

In a future submission, we will consider in detail theoretical expressions for the finite sample performance of ACMA, in particular, expressions that predict the covariance of $\mathbf{W}$ and the resulting SINR as function of N, σ^2, and $\mathbf{A}$.

References

[1] J.-F. Cardoso, S. Bose, and B. Friedlander, "On optimal source separation based on second and fourth order cumulants," in *Proc. 8th IEEE Signal Process. Workshop Stat. Signal Array Process.*, Corfu, Greece, 1996, pp. 198–201.

[2] J. F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.

[3] J. F. Cardoso and P. Comon, "Independent component analysis, a survey of some algebraic methods," in *Proc. IEEE ISCAS*, 1996, pp. 93–96.
[4] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng. F*, vol. 140, pp. 362–370, Dec. 1993.
[5] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 36, pp. 287–314, Apr. 1994.
[6] ——, "Contrasts for multichannel blind deconvolution," *IEEE Signal Processing Lett.*, vol. 3, pp. 209–211, July 1996.
[7] I. Fijalkow, C. E. Manlove, and C. R. Johnson, "Adaptive fractionally spaced blind CMA equalization: Excess MSE," *IEEE Trans. Signal Processing*, vol. 46, pp. 227–231, Jan. 1998.
[8] D. N. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," *IEEE Trans. Commun.*, vol. COMM-28, pp. 1867–1875, Nov. 1980.
[9] M. Gu and L. Tong, "Geometrical characterizations of constant modulus receivers," *IEEE Trans. Signal Processing*, vol. 47, pp. 2745–2756, Oct. 1999.
[10] R. Johnson, P. Schniter, T. J. Endres, J. D. Behm, D. R. Brown, and R. A. Casas, "Blind equalization using the constant modulus criterion: A review," *Proc. IEEE*, vol. 86, pp. 1927–1950, Oct. 1998.
[11] A. V. Keerthi, A. Mathur, and J. J. Shynk, "Misadjustment and tracking analysis of the constant modulus array," *IEEE Trans. Signal Processing*, vol. 46, pp. 51–58, Jan. 1998.
[12] L. De Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, Katholieke Univ. Leuven, Leuven, Belgium, 1997.
[13] A. Leshem and A. J. van der Veen, "Direction of arrival estimation for constant modulus signals," *IEEE Trans. Signal Processing*, vol. 47, pp. 3125–3129, Nov. 1999.
[14] A. Mathur, A. V. Keerthi, J. J. Shynk, and R. P. Gooch, "Convergence properties of the multistage constant modulus array for correlated sources," *IEEE Trans. Signal Processing*, vol. 45, pp. 280–286, Jan. 1997.
[15] T. Nguyen and Z. Ding, "Blind CMA beamforming for narrowband signals with multipath arrivals," *Int. J. Adaptive Contr. Signal Process.*, vol. 12, pp. 157–172, Mar. 1998.
[16] C. B. Papadias and A. J. Paulraj, "A constant modulus algorithm for multiuser signal separation in presence of delay spread using antenna arrays," *IEEE Signal Processing Lett.*, vol. 4, pp. 178–181, June 1997.
[17] C. W. Reed and K. Yao, "Performance of blind beamforming algorithms," in *Proc. Ninth IEEE Signal Process. Workshop Stat. Signal Array Process.*, 1998, pp. 256–259.
[18] P. A. Regalia, "On the equivalence between the Godard and Shalvi–Weinstein schemes of blind equalization," *Signal Process.*, vol. 73, pp. 185–190, Feb. 1999.
[19] S. Shamsunder and G. Giannakis, "Modeling of non-Gaussian array data using cumulants: DOA estimation of more sources with less sensors," *Signal Process.*, vol. 30, pp. 279–297, Feb. 1993.
[20] J. Sheinvald, "On blind beamforming for multiple non-Gaussian signals and the constant-modulus algorithm," *IEEE Trans. Signal Processing*, vol. 46, pp. 1878–1885, July 1998.
[21] J. J. Shynk and R. P. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. Signal Processing*, vol. 44, pp. 652–660, Mar. 1996.
[22] A. L. Swindlehurst, M. J. Goris, and B. Ottersten, "Some experiments with array data collected in actual urban and suburban environments," in *Proc. IEEE Workshop Signal Process. Advances Wireless Commun.*, Paris, France, Apr. 1997, pp. 301–304.
[23] A. Touzni, I. Fijalkow, M. Larimore, and J. R. Treichler, "A globally convergent approach for blind MIMO adaptive deconvolution," in *Proc. IEEE ICASSP*, vol. 4, 1998, pp. 2385–2388.
[24] J. R. Treichler and B. G. 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.
[25] J. R. Treichler and M. G. Larimore, "New processing techniques based on constant modulus adaptive algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 420–431, Apr. 1985.
[26] J. R. Treichler, M. G. Larimore, and J. C. Harp, "Practical blind demodulators for high-order QAM signals," *Proc. IEEE*, vol. 86, pp. 1907–1926, Oct. 1998.
[27] A. J. van der Veen, "Analytical method for blind binary signal separation," *IEEE Trans. Signal Processing*, vol. 45, pp. 1078–1082, Apr. 1997.
[28] ——, "Weighted ACMA," in *Proc. IEEE ICASSP*, Phoenix, AZ, Mar. 1999.
[29] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.
[30] M. Wax and Y. Anu, "A least-squares approach to blind beamforming," *IEEE Trans. Signal Processing*, vol. 47, pp. 231–234, Jan. 1999.
[31] M. Wax and J. Sheinvald, "A least-squares approach to joint diagonalization," *IEEE Signal Processing Lett.*, vol. 4, pp. 52–53, Feb. 1997.
[32] H. H. Zeng, L. Tong, and C. R. Johnson, "Relationships between the constant modulus and Wiener receivers," *IEEE Trans. Inform. Theory*, vol. 44, pp. 1523–1538, July 1998.
[33] ——, "An analysis of constant modulus receivers," *IEEE Trans. Signal Processing*, vol. 47, pp. 2990–2999, Nov. 1999.
[34] J. Zhu, X. R. Cao, and Z. Ding, "An algebraic principle for blind separation of white non-Gaussian sources," *Signal Process.*, vol. 76, pp. 105–115, 1999.

Alle-Jan van der Veen (S'87–M'94) was born in The Netherlands in 1966. He graduated (cum laude) from the Department of Electrical Engineering, Delft University of Technology, Delft, The Netherlands, in 1988 and received the Ph.D. degree (cum laude) from the same institute in 1993.

Throughout 1994, he was a postdoctoral scholar at Stanford University, Stanford, CA, in the Scientific Computing/Computational Mathematics group and in the Information Systems Lab. At present, he is an Associate Professor with the Signal Processing group of DIMES, Delft University of Technology. His research interests are in the general area of system theory applied to signal processing and, in particular, algebraic methods for array signal processing.

Dr. Van der Veen received a 1994 and a 1997 IEEE Signal Processing Society Young Author paper award. Currently, he is an Associate Editor for IEEE TRANSACTIONS ON SIGNAL PROCESSING, and vice-chairman of the IEEE SPS SPCOM Technical Committee.