

LARGE SAMPLE PERFORMANCE ANALYSIS OF ACMA

Alle-Jan van der Veen

Delft University of Technology, Department of Electrical Engineering/DIMES
Mekelweg 4, 2628 CD Delft, The Netherlands

The ‘‘Algebraic Constant Modulus Algorithm’’ (ACMA) is a non-iterative block algorithm for blind separation of constant modulus sources. We previously showed that, unlike CMA, it asymptotically converges to the (non-blind) Wiener receiver. In this paper, we present a finite sample statistical performance analysis. This can be used to predict the SINR performance, as well as the deviation from the Wiener receivers. The theoretical performance is illustrated by numerical simulations and shows a good match.

1. INTRODUCTION

In this paper we study the performance of ACMA (‘‘Analytical Constant Modulus Algorithm’’), proposed in [1]. ACMA is a non-recursive blind source separation algorithm for constant modulus signals. It is a batch algorithm that under noise-free conditions can compute exact separating beamformers for all sources at the same time, using only a small number of samples. Although it has been derived as a deterministic method, it is closely related to JADE and other fourth-order statistics based source separation techniques.

We could recently show that (unlike CMA), ACMA beamformers converge asymptotically in the number of samples to the (non-blind) Wiener receivers [2]. Here, we will extend the analysis by deriving the large *finite* sample performance of a block of N samples. For this we need the statistics of the eigenvectors of a fourth order covariance matrix with non-Gaussian sources.

2. DATA MODEL

We consider a linear data model of the form

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

where $\mathbf{x}_k \in \mathbb{C}^M$ is the data vector received by an array of M sensors at time k , $\mathbf{s}_k \in \mathbb{C}^d$ is the source vector at time k , and $\mathbf{n}_k \in \mathbb{C}^M$ an additive noise vector. $\mathbf{A} = [\mathbf{a}_1 \cdots \mathbf{a}_d]$ represents an $M \times d$ complex-valued instantaneous mixing matrix (or array response matrix). The sources are constant modulus (CM), i.e. each entry s_l of \mathbf{s} satisfies $|s_l| = 1$.

We collect N samples in a matrix $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] : M \times N$. Similarly defining $\mathbf{S} : d \times N$ and $\mathbf{N} : M \times N$, we obtain

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}. \quad (2)$$

\mathbf{A} , \mathbf{S} and \mathbf{N} are unknown. The objective is to reconstruct \mathbf{S} using linear beamforming, i.e., to find a beamforming matrix $\mathbf{W} = [\mathbf{w}_1, \dots, \mathbf{w}_d] \in \mathbb{C}^{M \times d}$ of full row rank d such that $\hat{\mathbf{S}} = \mathbf{W}^H \mathbf{X}$ approximates \mathbf{S} . Since \mathbf{S} is unknown, the criterion for this is that $\hat{\mathbf{S}}$ should be as close to a CM matrix as possible, i.e., we aim to make $|\hat{S}_{ik}| = |\mathbf{w}_i^H \mathbf{x}_k| = 1 \forall i, k$. If this is the case, then $\hat{\mathbf{S}}$ is equal to \mathbf{S} up to unknown permutations and unit-norm scalings of its rows. With noise, we can obviously recover the sources only approximatively.

We work under the following assumptions:

1. $N \geq d^2$. \mathbf{A} has full rank d , and $M \geq d$. To avoid complications in the analysis, we assume $M = d$.
2. The sources are statistically independent constant modulus sources, circularly symmetric, with covariance $\mathbf{R}_s := \mathbf{E}(\mathbf{s}\mathbf{s}^H) = \mathbf{I}$.
3. The noise is additive white Gaussian, zero mean, circularly symmetric, independent from the sources, with covariance $\mathbf{R}_n := \mathbf{E}(\mathbf{n}\mathbf{n}^H) = \sigma^2 \mathbf{I}$.

Notation Overbar ($\bar{\cdot}$) denotes complex conjugation, $^\top$ is the matrix transpose, H the matrix complex conjugate transpose, † the matrix pseudo-inverse (Moore-Penrose inverse). \mathbf{I} (or \mathbf{I}_p) is the ($p \times p$) identity matrix; \mathbf{e}_i is its i -th column. $\mathbf{0}$ and $\mathbf{1}$ are vectors with all entries equal to 0 and 1, respectively. $\text{vec}(\mathbf{A})$ is a stacking of the columns of a matrix \mathbf{A} into a vector. For a vector, $\text{diag}(\mathbf{v})$ is a diagonal matrix with the entries of \mathbf{v} on the diagonal. \odot is the Schur-Hadamard (entry-wise) matrix product, \otimes is the Kronecker product, \circ is the Khatri-Rao product, which is a column-wise Kronecker product. $\mathbf{E}(\cdot)$ denotes the expectation operator.

For a matrix-valued stochastic variable $\hat{\mathbf{R}}$, define its covariance matrix $\text{cov}\{\hat{\mathbf{R}}\} = \mathbf{E}\{\text{vec}(\hat{\mathbf{R}} - \mathbf{E}(\hat{\mathbf{R}}))[\text{vec}(\hat{\mathbf{R}} - \mathbf{E}(\hat{\mathbf{R}}))]^H\}$.

For a zero mean random vector $\mathbf{x} = [x_i]$, define the fourth order cumulant matrix

$$\mathbf{K}_\mathbf{x} = \mathbf{E}(\bar{\mathbf{x}} \otimes \mathbf{x})(\bar{\mathbf{x}} \otimes \mathbf{x})^H - \mathbf{E}(\bar{\mathbf{x}} \otimes \mathbf{x})\mathbf{E}(\bar{\mathbf{x}} \otimes \mathbf{x})^H - \mathbf{E}(\bar{\mathbf{x}}\bar{\mathbf{x}}^H) \otimes \mathbf{E}(\mathbf{x}\mathbf{x}^H) - \mathbf{E}(\bar{\mathbf{x}} \otimes \mathbf{1})(\mathbf{1} \otimes \mathbf{x})^H \odot \mathbf{E}(\mathbf{1} \otimes \mathbf{x})(\bar{\mathbf{x}} \otimes \mathbf{1})^H.$$

For circularly symmetric variables, the last term vanishes.

3. FORMULATION OF THE ALGORITHM

In brief outline, ACMA consists of two main steps: a prewhitening operation, and the algorithm proper. Define the data covariance matrix and its sample estimate

$$\mathbf{R}_\mathbf{x} := \mathbf{E}\{\mathbf{x}\mathbf{x}^H\}, \quad \hat{\mathbf{R}}_\mathbf{x} := \frac{1}{N} \sum \mathbf{x}_k \mathbf{x}_k^H.$$

Assuming that $M = d$ for simplicity of the analysis, the prewhitening filter transforms the data to

$$\underline{\mathbf{X}} := \hat{\mathbf{R}}_\mathbf{x}^{-1/2} \mathbf{X} =: \underline{\mathbf{A}}\mathbf{S} + \underline{\mathbf{N}}$$

where the underscore indicates the prewhitening. Note that $\underline{\hat{\mathbf{R}}}_\mathbf{x} = \mathbf{I}$.

Given the N data samples $[\mathbf{x}_k]$, the purpose of a beamforming vector \mathbf{w} is to recover one of the sources as $\hat{s}_k = \mathbf{w}^H \mathbf{x}_k$. One technique for estimating such a beamformer is by minimizing the deterministic CMA(2,2) cost function, $\hat{\mathbf{w}} = \arg\min_{\mathbf{w}} \frac{1}{N} \sum (|\mathbf{w}^H \mathbf{x}_k|^2 - 1)^2$. Define

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

In [2], we have derived that CMA(2,2) is equivalent to (up to a scaling of \mathbf{w} which is not of interest to its performance)

$$\hat{\mathbf{w}} = \hat{\mathbf{R}}_{\mathbf{x}}^{-1/2} \hat{\mathbf{t}}, \quad \hat{\mathbf{t}} = \underset{\substack{\mathbf{y} = \bar{\mathbf{T}} \otimes \mathbf{t} \\ \|\mathbf{y}\| = 1}}{\operatorname{argmin}} \mathbf{y}^H \hat{\mathbf{C}}_{\mathbf{x}} \mathbf{y}, \quad (3)$$

ACMA is obtained as a two-step approach to the latter minimization problem [2]:

1. Find an orthonormal basis $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_d]$ of independent minimizers of $\mathbf{y}^H \hat{\mathbf{C}}_{\mathbf{x}} \mathbf{y}$, i.e., the eigenvectors corresponding to the d smallest eigenvalues of $\hat{\mathbf{C}}_{\mathbf{x}}$.
2. Find a basis $\{\tilde{\mathbf{t}}_1 \otimes \hat{\mathbf{t}}_1, \dots, \tilde{\mathbf{t}}_d \otimes \hat{\mathbf{t}}_d\}$ that spans the same linear subspace as $\{\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_d\}$, and with $\|\tilde{\mathbf{t}}_i\| = 1$, i.e., solve

$$\hat{\mathbf{T}} = \min_{\mathbf{T}, \mathbf{M}} \|\hat{\mathbf{Y}} - (\bar{\mathbf{T}} \circ \mathbf{T}) \mathbf{M}\|_F^2, \quad (4)$$

subject to the constraint $\operatorname{diag}(\mathbf{T}^H \mathbf{T}) = \mathbf{I}$.

It was shown in [2] that $\hat{\mathbf{T}}$ converges asymptotically in N to a matrix $\mathbf{T} = \underline{\mathbf{A}}_0$, where $\underline{\mathbf{A}}_0$ is equal to $\underline{\mathbf{A}}$ except for a scaling and permutation of its columns. In the non-whitened domain, $\hat{\mathbf{W}} = \hat{\mathbf{R}}_{\mathbf{x}}^{-1/2} \hat{\mathbf{T}}$ converges asymptotically to $\mathbf{W} = \mathbf{R}_{\mathbf{x}}^{-1} \mathbf{A}_0$, the Wiener receiver (except for the scaling and the permutation).

A performance analysis is now possible, and follows in outline the analysis of the MUSIC and WSF DOA estimators [3], but extended to fourth order statistics of non-Gaussian sources. The following limitations are introduced to keep the derivations tractable.

1. N is sufficiently large, and we neglect terms of order N^{-2} over terms of order N^{-1} . The noise power σ^2 is sufficiently small and we neglect σ^4 over σ^2 .
2. We assume that the whitening step is based on the *true* covariance matrix $\mathbf{R}_{\mathbf{x}}$. (This is accurate for $M = d$.)
3. We assume that the exact solution to (4) is computed.

4. COVARIANCE OF $\hat{\mathbf{C}}_{\mathbf{x}}$

In this and the next sections, we drop for convenience the underscore from the notation since all variables are based on whitened data. Our objective in this section is to find a compact approximate expression for the covariance of $\hat{\mathbf{C}}_{\mathbf{x}}$, denoted by $\boldsymbol{\Omega}_{\mathbf{x}}$. Define

$$\mathbf{C}_{\mathbf{x}} = \mathbb{E}\{(\bar{\mathbf{x}}_k \otimes \mathbf{x}_k)(\bar{\mathbf{x}}_k \otimes \mathbf{x}_k)^H\} - \mathbb{E}\{\bar{\mathbf{x}}_k \otimes \mathbf{x}_k\} \mathbb{E}\{\bar{\mathbf{x}}_k \otimes \mathbf{x}_k\}^H.$$

Using properties of cumulants, we can show that [2]

$$\mathbf{C}_{\mathbf{x}} = -[\bar{\mathbf{A}} \circ \mathbf{A}][\bar{\mathbf{A}} \circ \mathbf{A}]^H + \bar{\mathbf{R}}_{\mathbf{x}} \otimes \mathbf{R}_{\mathbf{x}} = -[\bar{\mathbf{A}} \circ \mathbf{A}][\bar{\mathbf{A}} \circ \mathbf{A}]^H + \mathbf{I}. \quad (5)$$

Furthermore, a straightforward derivation shows that

$$\operatorname{cov}\{\hat{\mathbf{R}}_{\mathbf{x}}\} = \frac{1}{N} \mathbf{C}_{\mathbf{x}}. \quad (6)$$

Thus, $\mathbf{C}_{\mathbf{x}}$ is the covariance of $\hat{\mathbf{R}}_{\mathbf{x}}$, and $\hat{\mathbf{C}}_{\mathbf{x}}$ is a (biased) sample estimate of it. A second interpretation of $\mathbf{C}_{\mathbf{x}}$ is obtained by defining a “data” sequence

$$\mathbf{g}_k := \bar{\mathbf{x}}_k \otimes \mathbf{x}_k - \mathbb{E}\{\bar{\mathbf{x}}_k \otimes \mathbf{x}_k\}, \quad k = 1, \dots, N, \quad (7)$$

and considering its covariance and sample covariance

$$\mathbf{R}_{\mathbf{g}} := \mathbb{E}\{\mathbf{g}_k \mathbf{g}_k^H\}, \quad \hat{\mathbf{R}}_{\mathbf{g}} := \frac{1}{N} \sum \mathbf{g}_k \mathbf{g}_k^H.$$

It is straightforward to show that

$$\mathbb{E}\{\hat{\mathbf{R}}_{\mathbf{g}}\} = \mathbf{R}_{\mathbf{g}} = \mathbf{C}_{\mathbf{x}}, \quad \hat{\mathbf{R}}_{\mathbf{g}} = \hat{\mathbf{C}}_{\mathbf{x}}(1 + \mathcal{O}(\frac{1}{N})).$$

Thus, $\mathbf{C}_{\mathbf{x}}$ is the covariance of \mathbf{g}_k , and $\hat{\mathbf{R}}_{\mathbf{g}}$ is an unbiased sample estimate of it; in first order approximation it has the same properties as the biased estimate $\hat{\mathbf{C}}_{\mathbf{x}}$. Similar to (6), it follows that $\operatorname{cov}\{\hat{\mathbf{R}}_{\mathbf{g}}\} = \frac{1}{N} \mathbf{C}_{\mathbf{g}}$ where

$$\mathbf{C}_{\mathbf{g}} := \mathbb{E}\{(\bar{\mathbf{g}} \otimes \mathbf{g})(\bar{\mathbf{g}} \otimes \mathbf{g})^H\} - \mathbb{E}\{\bar{\mathbf{g}} \otimes \mathbf{g}\} \mathbb{E}\{\bar{\mathbf{g}} \otimes \mathbf{g}\}^H. \quad (8)$$

In summary, we can prove

Theorem 1. $\boldsymbol{\Omega}_{\mathbf{x}} := \operatorname{cov}\{\hat{\mathbf{C}}_{\mathbf{x}}\} = \frac{1}{N} \mathbf{C}_{\mathbf{g}} + \mathcal{O}(\frac{1}{N^2})$.

It remains to find a compact description of $\mathbf{C}_{\mathbf{g}}$ in terms of our data model. Inserting the model $\mathbf{x}_k = \mathbf{A} \mathbf{s}_k + \mathbf{n}_k$ in the definition of \mathbf{g}_k , we obtain

$$\mathbf{g}_k = \mathbf{A}_{\mathbf{c}} \mathbf{c}_k + \mathbf{n}_k$$

where

$$\begin{aligned} \mathbf{c} &:= \sum_{i \neq j} \mathbf{e}'_{ij} \bar{s}_j s_i = [\bar{s}_1 s_2, \dots, \bar{s}_1 s_d, \bar{s}_2 s_1, \bar{s}_2 s_3, \dots]^T \\ \mathbf{A}_{\mathbf{c}} &:= [\bar{\mathbf{a}}_j \otimes \mathbf{a}_i]_{i \neq j} \\ \mathbf{n} &:= \bar{\mathbf{n}} \otimes \mathbf{n} - \bar{\mathbf{R}}_{\mathbf{n}} + \bar{\mathbf{A}} \bar{\mathbf{s}} \otimes \mathbf{n} + \bar{\mathbf{n}} \otimes \mathbf{A} \mathbf{s} \end{aligned}$$

where $\mathbf{e}'_{ij} = \operatorname{vec}'(\mathbf{e}_i \mathbf{e}_j^H)$, and $\operatorname{vec}'(\cdot)$ is a vectoring operator which skips the main diagonal. The vector \mathbf{c} is CM (with certain dependencies among its entries). Likewise, the matrix $\mathbf{A}_{\mathbf{c}}$ skips the $\bar{\mathbf{a}}_i \otimes \mathbf{a}_i$ columns of $\bar{\mathbf{A}} \otimes \mathbf{A}$.

The model $\mathbf{g}_k = \mathbf{A}_{\mathbf{c}} \mathbf{c}_k + \mathbf{n}_k$ has several properties that are similar to that of $\mathbf{x}_k = \mathbf{A} \mathbf{s}_k + \mathbf{n}_k$. However, \mathbf{c} and \mathbf{n} are not independent (only uncorrelated), not circularly symmetric, and $\mathbf{K}_{\mathbf{n}} \neq \mathbf{0}$. A good approximation for $\mathbf{C}_{\mathbf{g}}$ taking into account all terms up to $\mathcal{O}(\sigma^2)$, is given as

Theorem 2. $\mathbf{C}_{\mathbf{g}} \approx [\bar{\mathbf{A}}_{\mathbf{c}} \otimes \mathbf{A}_{\mathbf{c}}] \mathbf{K}_{\mathbf{c}}' [\bar{\mathbf{A}}_{\mathbf{c}} \otimes \mathbf{A}_{\mathbf{c}}]^H + \bar{\mathbf{R}}_{\mathbf{g}} \otimes \mathbf{R}_{\mathbf{g}} + \mathbf{E} + \mathbf{E}^H$ where

$$\begin{aligned} \mathbf{E} &= [\mathbf{A} \otimes \bar{\mathbf{R}}_{\mathbf{n}}^{1/2} \otimes \mathbf{A}_{\mathbf{c}}] \mathbf{E}_1 [\bar{\mathbf{A}}_{\mathbf{c}} \otimes \bar{\mathbf{R}}_{\mathbf{n}}^{1/2} \otimes \mathbf{A}]^H \\ &\quad + [\mathbf{R}_{\mathbf{n}}^{1/2} \otimes \bar{\mathbf{A}} \otimes \mathbf{A}_{\mathbf{c}}] \mathbf{E}_2 [\bar{\mathbf{A}}_{\mathbf{c}} \otimes \bar{\mathbf{A}} \otimes \mathbf{R}_{\mathbf{n}}^{1/2}]^H \\ \mathbf{K}_{\mathbf{c}}' &= \mathbf{K}_{\mathbf{c}} + \sum_{i \neq j} \sum_{k \neq l} (\mathbf{e}'_{ij} \otimes \mathbf{e}'_{kl})(\mathbf{e}'_{ik} \otimes \mathbf{e}'_{jl})^H \\ \mathbf{K}_{\mathbf{c}} &= -[\sum_{i \neq j} (\mathbf{e}'_{ij} \otimes \mathbf{e}'_{ij})(\mathbf{e}'_{ij} \otimes \mathbf{e}'_{ij})^H + (\mathbf{e}'_{ji} \otimes \mathbf{e}'_{ij})(\mathbf{e}'_{ji} \otimes \mathbf{e}'_{ij})^H \\ &\quad + (\mathbf{e}'_{ij} \otimes \mathbf{e}'_{ij})(\mathbf{e}'_{ji} \otimes \mathbf{e}'_{ji})^H] \\ \mathbf{E}_1 &= \sum_i \sum_{j \neq i} \sum_{k \neq l} \mathbf{e}_{ji}^H \otimes \mathbf{e}_k \otimes \mathbf{I}_d \otimes \mathbf{e}_j^H \otimes \mathbf{e}'_{ik} \\ &\quad + \mathbf{e}_{ij}^H \otimes \mathbf{e}_j \otimes \mathbf{I}_d \otimes \mathbf{e}_k^H \otimes \mathbf{e}'_{ki} + \mathbf{e}_{ij}^H \otimes \mathbf{e}_k \otimes \mathbf{I}_d \otimes \mathbf{e}_k^H \otimes \mathbf{e}'_{ji} (1 - \delta_k^j) \\ \mathbf{E}_2 &= \sum_i \sum_{j \neq i} \sum_{k \neq l} \mathbf{e}_{ji}^H \otimes \mathbf{e}_k^H \otimes \mathbf{I}_d \otimes \mathbf{e}_j \otimes \mathbf{e}'_{ik} \\ &\quad + \mathbf{e}_{ij}^H \otimes \mathbf{e}_j^H \otimes \mathbf{I}_d \otimes \mathbf{e}_k \otimes \mathbf{e}'_{ki} + \mathbf{e}_{ij}^H \otimes \mathbf{e}_k^H \otimes \mathbf{I}_d \otimes \mathbf{e}_k \otimes \mathbf{e}'_{ji} (1 - \delta_k^j). \end{aligned}$$

(All indices range over $1, \dots, d$. Note, the latter matrices are data independent and simply collections of ‘1’ entries.)

PROOF Omitted.

It can be shown experimentally that the term $\bar{\mathbf{R}}_{\mathbf{g}} \otimes \mathbf{R}_{\mathbf{g}}$ is the dominant term, so that

$$\mathbf{C}_{\mathbf{g}} \approx \bar{\mathbf{C}}_{\mathbf{x}} \otimes \mathbf{C}_{\mathbf{x}} \quad (9)$$

is a good approximation. This is the same as regarding \mathbf{c} and \mathbf{n} as Gaussian vectors with independent entries. Making this approximation would lead to particularly simple results in the eigenvector perturbation study and subsequent steps, as we basically can apply the theory in Viberg [3].

5. EIGENVECTOR PERTURBATION

In this section we consider the statistical properties of the eigenvectors of $\hat{\mathbf{C}}_{\mathbf{x}}$, a fourth order sample covariance matrix based on nonGaussian signals. We first give a general derivation and then specialize to the case at hand. The generalization is needed because most existing derivations consider Gaussian sources.

For a covariance matrix \mathbf{R} with unbiased sample estimate $\hat{\mathbf{R}}$ based on N samples of a (not necessarily Gaussian) vector process, consider the eigenvalue decompositions $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, $\hat{\mathbf{R}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^H$. If we elaborate on the equality

$$\hat{\mathbf{R}} - \mathbf{R} = (\hat{\mathbf{U}} - \mathbf{U})\mathbf{\Lambda}\mathbf{U}^H - \hat{\mathbf{R}}(\hat{\mathbf{U}} - \mathbf{U})\mathbf{U}^H + \hat{\mathbf{U}}(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda})\mathbf{U}^H$$

and assume that we partition the eigenvalue decomposition of \mathbf{R} as

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^H, \quad (10)$$

where the eigenvalues in $\mathbf{\Lambda}_s$ are distinct and unequal to any eigenvalue in $\mathbf{\Lambda}_n$, then we can derive directly that in first order

$$\text{vec}(\mathbf{P}_n\hat{\mathbf{U}}_s) \approx [\mathbf{I} \otimes \mathbf{U}_n][\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]^{-1}[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H \text{vec}(\hat{\mathbf{R}} - \mathbf{R})$$

where $\mathbf{P}_n = \mathbf{U}_n\mathbf{U}_n^H$. From the latter we can immediately find an expression for the covariance of the ‘‘signal’’ eigenvectors projected into the ‘‘noise’’ subspace:

Lemma 3. *Let $\hat{\mathbf{R}}$ be a sample covariance matrix converging to \mathbf{R} , and assume that \mathbf{R} has eigenvalue decomposition (10) where the entries in $\mathbf{\Lambda}_s$ are distinct and unequal to any entry in $\mathbf{\Lambda}_n$. Then*

$$\text{cov}\{\mathbf{P}_n\hat{\mathbf{U}}_s\} = [\mathbf{I} \otimes \mathbf{U}_n][\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]^{-1}[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H \cdot \text{cov}\{\hat{\mathbf{R}}\} \cdot [\bar{\mathbf{U}}_s \otimes \mathbf{U}_n][\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]^{-1}[\mathbf{I} \otimes \mathbf{U}_n]^H + o(N^{-1}). \quad (11)$$

Essentially the same result appears in [4], but written as summations and with a more indirect proof.

We now specialize to our situation. We have

$$\mathbf{R} \leftrightarrow \mathbf{R}_g = \mathbf{C}_x \\ \text{cov}\{\hat{\mathbf{R}}\} \leftrightarrow \mathbf{\Omega}_x = \text{cov}\{\hat{\mathbf{C}}_x\} = \frac{1}{N}\mathbf{C}_g + o(N^{-2}).$$

Introduce the eigenvalue decomposition of \mathbf{C}_x as

$$\mathbf{C}_x = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = \mathbf{U}_s\mathbf{\Lambda}_s\mathbf{U}_s^H + \mathbf{U}_n\mathbf{\Lambda}_n\mathbf{U}_n^H \quad (12)$$

where $\mathbf{\Lambda}_s$ collects the d smallest eigenvalues of \mathbf{C}_x . Likewise, \mathbf{U}_s is a basis for the approximate null space of \mathbf{C}_x . Also introduce the singular value decomposition

$$\mathbf{\Lambda} := \bar{\mathbf{\Lambda}} \circ \mathbf{\Lambda} = \mathbf{U}_A \mathbf{\Sigma}_A \mathbf{V}_A, \quad (13)$$

where \mathbf{U}_A has d orthonormal columns, $\mathbf{\Sigma}_A = \text{diag}[\sigma_k]$ is a $d \times d$ diagonal matrix, and \mathbf{V}_A is $d \times d$ unitary. Let \mathbf{U}_A^\perp be the orthogonal complement of \mathbf{U}_A . It follows from (5) that the eigenvalue decomposition of \mathbf{C}_x is given by

$$\mathbf{C}_x = [\mathbf{U}_A \quad \mathbf{U}_A^\perp] \begin{bmatrix} \mathbf{I} - \mathbf{\Sigma}_A^2 & \\ & \mathbf{I} \end{bmatrix} [\mathbf{U}_A \quad \mathbf{U}_A^\perp]^H. \quad (14)$$

In view of the partitioning in (12) we set $\mathbf{U}_s = \mathbf{U}_A$, $\mathbf{\Lambda}_s = \mathbf{I} - \mathbf{\Sigma}_A^2$, and $\mathbf{\Lambda}_n = \mathbf{I}$. Inserting this in (11), we obtain

Theorem 4. $\text{cov}\{\mathbf{P}_n^\perp\hat{\mathbf{U}}_s\} = \frac{1}{N}\mathbf{C}_u + o(N^{-1})$, where

$$\mathbf{C}_u := [\mathbf{\Sigma}_A^{-2}\bar{\mathbf{U}}_A^H \otimes \mathbf{P}_A^\perp] \mathbf{C}_g [\bar{\mathbf{U}}_A \mathbf{\Sigma}_A^{-2} \otimes \mathbf{P}_A^\perp].$$

Significant simplifications are possible if we allow the approximation of \mathbf{C}_g in (9).

6. SUBSPACE FITTING

6.1. Cost function

The next item in the analysis is the subspace fitting problem in (4). We can follow in outline the performance analysis technique described in [3]. Some notational changes are necessary.

In equation (4), we computed a $d \times d$ separating beamforming matrix $\hat{\mathbf{T}}$ (in the whitened domain), with columns constrained to have unit norm. W.l.o.g., we can further constrain the first nonzero entry of each column to be positive real. Let $\mathbf{A}(\boldsymbol{\theta})$ be a minimal parametrization of such matrices. The true mixing matrix can then be written as $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}_0)\mathbf{B}$, where \mathbf{B} is a diagonal scaling matrix which is unidentifiable by the subspace fitting. We assume that the true parameter vector $\boldsymbol{\theta}_0$ is uniquely identifiable and that $\mathbf{A}(\boldsymbol{\theta})$ is continuously differentiable around $\boldsymbol{\theta}_0$. We proved in [2] that as $N \rightarrow \infty$, $\hat{\mathbf{T}}$ converges to $\mathbf{A}_0 \equiv \mathbf{A}(\boldsymbol{\theta}_0)$, and thus we can write $\hat{\mathbf{T}} = \mathbf{A}(\hat{\boldsymbol{\theta}})$. In this notation, equation (4) becomes

$$\mathbf{A}(\hat{\boldsymbol{\theta}}) = \underset{\mathbf{A}(\boldsymbol{\theta}), \mathbf{M}}{\text{argmin}} \|\hat{\mathbf{U}}_s - \mathbf{A}(\boldsymbol{\theta})\mathbf{M}\|_F^2, \quad \mathbf{A}(\boldsymbol{\theta}) := \bar{\mathbf{A}}(\boldsymbol{\theta}) \circ \mathbf{A}(\boldsymbol{\theta}).$$

As usual, the problem is separable, and the optimum for \mathbf{M} given $\mathbf{A}(\boldsymbol{\theta})$ is $\mathbf{A}(\boldsymbol{\theta})^\dagger \hat{\mathbf{U}}_s$. Eliminating \mathbf{M} , we obtain

$$\mathbf{A}(\boldsymbol{\theta}) = \underset{\mathbf{A}(\boldsymbol{\theta})}{\text{argmin}} \|\mathbf{P}_{\mathbf{A}(\boldsymbol{\theta})}^\perp \hat{\mathbf{U}}_s\|_F^2$$

where $\mathbf{P}_{\mathbf{A}(\boldsymbol{\theta})}^\perp = \mathbf{I} - \mathbf{A}(\boldsymbol{\theta})\mathbf{A}(\boldsymbol{\theta})^\dagger$. Hence we will consider the minimization of the cost function

$$J(\boldsymbol{\theta}) = \|\mathbf{P}_{\mathbf{A}(\boldsymbol{\theta})}^\perp \hat{\mathbf{U}}_s\|_F^2 = \text{vec}(\mathbf{P}_{\mathbf{A}(\boldsymbol{\theta})}^\perp \hat{\mathbf{U}}_s)^H \text{vec}(\mathbf{P}_{\mathbf{A}(\boldsymbol{\theta})}^\perp \hat{\mathbf{U}}_s) \quad (15)$$

(This can be generalized to a *weighted* norm as usual.)

6.2. Covariance of $\hat{\boldsymbol{\theta}}$

Choose a specific parametrization of $\mathbf{A}(\boldsymbol{\theta})$. Since the columns of $\mathbf{A}(\boldsymbol{\theta})$ are not coupled, we can write $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_1), \dots, \mathbf{a}(\boldsymbol{\theta}_d)]$, where $\mathbf{a}(\boldsymbol{\theta}_i)$ is a parametrization of a unit-norm vector with real non-negative first entry, which requires $p := 2(d-1)$ real-valued parameters per vector. Denote θ_{ij} the i -th parameter of $\boldsymbol{\theta}_j$, and define the derivative matrix

$$\mathbf{D} = \left[\frac{\partial \mathbf{a}_1}{\partial \theta_{11}}, \frac{\partial \mathbf{a}_1}{\partial \theta_{21}}, \dots, \frac{\partial \mathbf{a}_2}{\partial \theta_{12}}, \dots \right] (\boldsymbol{\theta}_0). \quad (16)$$

Theorem 5. *Let $\mathbf{A}_0 := \bar{\mathbf{A}}(\boldsymbol{\theta}_0) \circ \mathbf{A}(\boldsymbol{\theta}_0)$, $\mathbf{A}_e := \mathbf{A}(\boldsymbol{\theta}_0) \otimes \mathbf{1}_p^T$,*

$$\begin{aligned} \mathbf{D} &= \bar{\mathbf{A}}_e \circ \mathbf{D} + \bar{\mathbf{D}} \circ \mathbf{A}_e \\ \mathbf{M} &:= (\mathbf{A}_0^\dagger \mathbf{U}_A)^H \otimes \mathbf{I}_p^T \\ \mathbf{C}_u &:= [\mathbf{\Sigma}_A^{-2} \bar{\mathbf{U}}_A^H \otimes \mathbf{P}_A^\perp] \mathbf{C}_g [\bar{\mathbf{U}}_A \mathbf{\Sigma}_A^{-2} \otimes \mathbf{P}_A^\perp] \\ \mathbf{Q} &:= 4[\mathbf{M} \circ \mathbf{P}_A^\perp \mathbf{D}]^H \mathbf{C}_u [\mathbf{M} \circ \mathbf{P}_A^\perp \mathbf{D}] \\ \mathbf{H} &:= 2[\mathbf{M} \circ \mathbf{P}_A^\perp \mathbf{D}]^H [\mathbf{M} \circ \mathbf{P}_A^\perp \mathbf{D}], \end{aligned}$$

where \mathbf{U}_A and $\mathbf{\Sigma}_A$ are defined in (13). For large N , the covariance of $\hat{\boldsymbol{\theta}}$ that minimizes the subspace fitting problem (15) is in first order approximation

$$\mathbf{R}_\theta := \text{cov}\{\hat{\boldsymbol{\theta}}\} = \frac{1}{N} \mathbf{H}^{-1} \mathbf{Q} \mathbf{H}^{-1}.$$

PROOF Omitted; along the lines of [3].

6.3. Covariance of \mathbf{T}

It remains to map the previous result to an expression for the covariance of the beamforming vectors. With some abuse of notation, let $\mathbf{t} = \text{vec}(\mathbf{T})$, where $\mathbf{T} = \mathbf{A}(\boldsymbol{\theta}_0)$, and let $\hat{\mathbf{t}} = \text{vec}(\hat{\mathbf{T}}) = \text{vec}(\mathbf{A}(\hat{\boldsymbol{\theta}}))$. Then, for small perturbations, $\hat{\mathbf{t}} = \mathbf{t} + \sum_{\eta} \frac{\partial \mathbf{t}}{\partial \theta_{\eta}} (\hat{\theta}_{\eta} - \theta_{\eta})$, so that $\hat{\mathbf{t}}$ has covariance

$$\begin{aligned} \mathbf{R}_{\hat{\mathbf{t}}} &= \begin{bmatrix} \frac{\partial \mathbf{t}}{\partial \theta_{11}} & \frac{\partial \mathbf{t}}{\partial \theta_{21}} & \dots \end{bmatrix} \mathbf{R}_{\boldsymbol{\theta}} \begin{bmatrix} \frac{\partial \mathbf{t}}{\partial \theta_{11}} & \frac{\partial \mathbf{t}}{\partial \theta_{21}} & \dots \end{bmatrix}^H \\ &= [(\mathbf{I}_d \otimes \mathbf{I}_p) \circ \mathbf{D}] \mathbf{R}_{\boldsymbol{\theta}} [(\mathbf{I}_d \otimes \mathbf{I}_p) \circ \mathbf{D}]^H, \end{aligned} \quad (17)$$

where \mathbf{D} was defined in (16). The covariance of a beamformer \mathbf{t}_j is the jj -th subblock of size $p \times p$ of $\mathbf{R}_{\hat{\mathbf{t}}}$.

6.4. SINR performance

To allow a better interpretation of the performance of the beamformers, we derive a mapping of $\mathbf{R}_{\hat{\mathbf{t}}}$ to the inverse SINR, or the INSR (interference plus noise to signal ratio), defined for a beamforming vector \mathbf{t} and array response vector \mathbf{a} of the corresponding source as (recall that $\mathbf{R}_{\mathbf{x}} = \mathbf{I}$)

$$\text{INSR}(\mathbf{t}) := \frac{\mathbf{t}^H (\mathbf{I} - \mathbf{a}\mathbf{a}^H) \mathbf{t}}{\mathbf{t}^H \mathbf{a}\mathbf{a}^H \mathbf{t}}.$$

The optimal solution that minimizes the INSR is $\mathbf{t} = \alpha \mathbf{a}$ (for an arbitrary nonzero scaling α). Consider a perturbation: $\hat{\mathbf{t}} = \mathbf{t} + \mathbf{d}$ where $\mathbf{t} = \alpha \mathbf{a}$. Then

$$\text{INSR}(\hat{\mathbf{t}}) \approx \frac{1}{\mathbf{a}^H \mathbf{a}} \left(1 - \mathbf{a}^H \mathbf{a} + \frac{\mathbf{d}^H \mathbf{P}_{\mathbf{a}} \mathbf{d}}{\mathbf{t}^H \mathbf{t}} \right), \quad (18)$$

where the approximation is good if $\mathbf{d}^H \mathbf{P}_{\mathbf{a}} \mathbf{d} \ll \mathbf{t}^H \mathbf{t}$. Let $\boldsymbol{\Delta} := \frac{\mathbf{E}(\mathbf{d}\mathbf{d}^H)}{\mathbf{t}^H \mathbf{t}}$ be a normalized (scale-invariant) definition of the covariance of $\hat{\mathbf{t}}$. Then in the above approximation

$$\mathbf{E}\{\text{INSR}(\hat{\mathbf{t}})\} = \frac{1 - \mathbf{a}^H \mathbf{a}}{\mathbf{a}^H \mathbf{a}} + \frac{\text{tr}(\mathbf{P}_{\mathbf{a}} \boldsymbol{\Delta})}{\mathbf{a}^H \mathbf{a}}. \quad (19)$$

The first term represents the asymptotic performance of the Wiener beamformer ($\hat{\mathbf{t}} = \mathbf{a}$ with $\boldsymbol{\Delta} = 0$). The second term is the excess INSR due to the deviation of $\hat{\mathbf{t}}$ from the optimum. We can simply plug in the estimates of $\mathbf{R}_{\hat{\mathbf{t}}}$ from equation (17) in place of $\boldsymbol{\Delta}$ to obtain the INSR corresponding to the ACMA beamformers.

For comparison, we consider the Wiener beamformer estimated from finite samples and known \mathbf{S} , or $\hat{\mathbf{T}}_W = (\mathbf{X}\mathbf{X}^H)^{-1} \mathbf{X}\mathbf{S}^H$. Let $\hat{\mathbf{t}}_W$ be one of the columns of $\hat{\mathbf{T}}_W$, and \mathbf{a} the corresponding column of \mathbf{A} . The normalized covariance of $\hat{\mathbf{t}}_W$ is derived as

$$\boldsymbol{\Delta}_W = \frac{\text{cov}(\hat{\mathbf{t}}_W - \mathbf{a})}{\mathbf{a}^H \mathbf{a}} = \frac{1}{N} \frac{1 - \mathbf{a}^H \mathbf{a}}{\mathbf{a}^H \mathbf{a}} \mathbf{I} + \mathcal{O}\left(\frac{1}{N^2}\right),$$

so that for the expected INSR of the finite-sample Wiener we find in first order approximation

$$\mathbf{E}\{\text{INSR}(\hat{\mathbf{t}}_W)\} = \frac{1 - \mathbf{a}^H \mathbf{a}}{\mathbf{a}^H \mathbf{a}} + \frac{d-1}{N} \cdot \frac{1 - \mathbf{a}^H \mathbf{a}}{(\mathbf{a}^H \mathbf{a})^2}. \quad (20)$$

7. SIMULATIONS

Figure 1 shows performance plots of the first source for a simulation with $d = 3$ sources, $M = 3$ antennas in a uniform linear array, source powers $\mathbf{B} = \text{diag}(1, 1.2, 0.9)$, and source angles $\boldsymbol{\alpha} = [0, \alpha, -\alpha]$, for varying N and SNR. The figure shows the excess INSR relative to the INSR of the asymptotic Wiener beamformer,

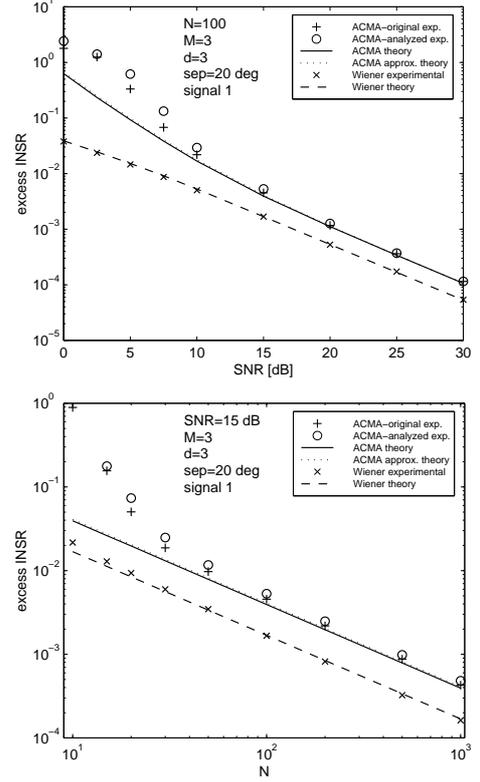


Figure 1. Finite sample INSR in excess of the asymptotic INSR of the Wiener beamformer.

evaluated for source 1 (i.e. the second terms in (19) and (20)). The experimental results show with '+' the outcome of the original ACMA algorithm of [1], and with 'o' the algorithm as analyzed here, i.e., with prewhitening based on the true covariance matrix $\mathbf{R}_{\mathbf{x}}$, and using Gauss-Newton optimization to solve the subspace fitting step. The dotted line is the approximation resulting from (9), which is indeed very good. As is seen from the figures, the theoretical curves are a good prediction of the actual performance once $N > 30$, SNR > 5 dB. The small difference in performance between the original algorithm and the analyzed algorithm is caused by the different prewhitening. Not shown in the figures are the results for *weighted* subspace fitting: these turned out to be virtually identical to the unweighted results.

References

- [1] A.J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.
- [2] A.J. van der Veen, "Asymptotic properties of the Algebraic Constant Modulus Algorithm," *IEEE Trans. Signal Processing*, vol. 49, Aug. 2001.
- [3] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Proc.*, vol. 39, pp. 1110–1121, May 1991.
- [4] N. Yuen and B. Friedlander, "Asymptotic performance analysis of ESPRIT, Higher-Order ESPRIT, and Virtual ESPRIT algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 2537–2550, Oct. 1996.