Statistical Performance Analysis of the Algebraic Constant Modulus Algorithm

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

***Abstract*—This paper presents a large sample analysis of the covariance of the beamformers computed by the analytical constant modulus algorithm (ACMA) method for blindly separating constant modulus sources. This can be used to predict the signal-to-interference plus noise ratio (SINR) performance of these beamformers, as well as their deviation from the (nonblind) Wiener receivers to which they asymptotically converge. The analysis is based on viewing ACMA as a subspace fitting optimization, where the subspace is spanned by the eigenvectors of a fourth-order covariance matrix. The theoretical performance is illustrated by numerical simulations and shows a good match.**

***Index Terms*—Blind beamforming, constant modulus algorithms (CMAs), eigenvector perturbation, performance analysis.**

I. Introduction

The last decade has seen a large interest in blind source separation techniques. An important part thereof is claimed by constant-modulus algorithms. Not only is the constant-modulus property directly applicable to many communication scenarios, it is also very robust in practice, can be applied to nonconstant modulus communication signals, and can often provide better separation performance than algorithms based on channel properties, such as direction finding [9]. In contrast to the overwhelming number of algorithms that have been proposed is the low number of algorithms whose performance has been studied in more detail.

In this paper we will study the performance of the analytical constant modulus algorithm (ACMA), which was proposed in [18]. ACMA is a nonrecursive 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. It has good performance in noise and fits several applications: not only blind source separation, but direction finding [21], [9] and frequency estimation [11] as well. In communication scenarios, it provides an excellent starting point for more optimal nonlinear receivers, such as ILSE [15], and it can be extended to handle convolutive channels [19]. Although it has been derived as a deterministic method, it is closely related to JADE [5] and other fourth-order statistics-based source separation techniques.

We have shown that in the presence of noise, the ACMA beamformers converge asymptotically in the number of samples to the (nonblind) Wiener receivers [16]. This is unlike CMA, whose asymptotic solutions are known to be close to but not coinciding with the Wiener receivers [7], [24], [25]. Here, we will continue the analysis of ACMA by deriving the large *finite* sample performance of a block of N samples. Apart from the theoretical interest, this gives answers to practical design issues, e.g., the choice of the data block size, the required SNR, the effect of fading and channel conditioning, and the tradeoff between training-based versus blind source separation. Finite sample performance of CMA is still terra incognita, with only a small start made in [10].

Our approach will be to derive the statistical properties and, in particular, the covariance of the beamformers computed by ACMA. For a better understanding, this is then mapped in terms of scale-invariant parameters such as the resulting signal-to-interference plus noise ratio (SINR) at the output.

For purpose of the analysis, ACMA is written as the solution to a subspace fitting problem, where the subspace is spanned by the eigenvectors of a fourth-order covariance matrix. The performance can thus be derived along the lines of the analysis of MUSIC or WSF [20]. Complicating factors in the analysis are that we need the statistics of a fourth-order covariance matrix, which gives rise to eighth-order statistics, and that the usual eigenvector analysis results for Gaussian sources are not applicable here.

Several other papers that analyze the performance of fourth-order source separation algorithms provide useful ingredients, e.g., a fourth-order MUSIC DOA algorithm is analyzed in [4]. Similarly, [13] contains expressions for the covariance of fourth-order cumulants. In our case, we can be more explicit because constant-modulus sources have kurtosis -1 and known sixth- and eighth-order statistics. A second difference is that the algorithm is not based on the cumulant matrix but on a different fourth-order covariance matrix, in which the influence of the Gaussian noise is nonzero. An expression for the covariance of eigenvectors in the non-Gaussian case has been presented in [23] in the form of a summation over six indices. Here, we derive a more compact expression from first principles.

The outline of the paper is as follows. The data model is formulated in Section II and the ACMA in Section III. Subsequently, Section IV provides an approximate expression for the statistics of the fourth-order covariance matrix used by ACMA, Section V the statistics of the eigenvectors of this matrix, and Section VI of the beamformers resulting from the subspace fitting step, mapped to an expression for the resulting SINR at the output of the beamformer. Finally, Section VII compares the theoretical performance to the experimental performance in simulations.

Manuscript received March 15, 2001; revised July 11, 2002. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dimitrios Hatzinakos.

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

Digital Object Identifier 10.1109/TSP.2002.805502

II. Data Model

A. Problem Statement

We consider a linear data model of the form

$$\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k \tag{1}$$

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 $\mathbf{n}_k \in \mathbb{C}^M$ is an additive noise vector. $\mathbf{A} = [\mathbf{a}_1 \ \ldots \ \mathbf{a}_d]$ 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_i of $\mathbf{s}$ satisfies $|s_i| = 1$.

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

$$\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}. \tag{2}$$

$\mathbf{A}$, $\mathbf{S}$, and $\mathbf{N}$ are unknown. The objective is to reconstruct $\mathbf{S}$ using linear beamforming, i.e., to find a beamforming matrix $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_d] \in \mathbb{C}^{M \times d}$ 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 $|\hat{\mathbf{S}}_{ik}| = |\mathbf{w}_i^H\mathbf{x}_k| = 1 \ \forall i, k$. 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) $\mathbf{A}$ has full rank d, and $M \geq d$ (so that $\mathbf{A}$ has a left inverse). The analysis will in fact only consider the case $M = d$.
2) $N \geq d^2$ (this is required by the algorithm).
3) The sources are statistically independent constant modulus sources, circularly symmetric, with covariance $\mathbf{R}_s := E(\mathbf{s}\mathbf{s}^H) = \mathbf{I}$.
4) The noise is additive white Gaussian, zero mean, circularly symmetric, independent from the sources, with covariance $\mathbf{R}_n := E(\mathbf{n}\mathbf{n}^H) = \sigma^2\mathbf{I}$.

B. Notation

An overbar, i.e., $\bar{\ }$ denotes complex conjugation, ^{T} is the matrix transpose, ^{H} is the matrix complex conjugate transpose, and † is the matrix pseudo-inverse (Moore–Penrose inverse). $\mathbf{I}$ (or $\mathbf{I}_p$) is the $(p \times p)$ identity matrix, and $\mathbf{e}_i$ is its ith column. The operator Re$(\cdot)$ selects the real part of its argument. $\mathbf{0}$ and $\mathbf{1}$ are vectors for which all entries are equal to 0 and 1, respectively.

vec$(\mathbf{A})$ is a stacking of the columns of a matrix $\mathbf{A}$ into a vector. For a vector, diag$(\mathbf{v})$ is a diagonal matrix with the entries of $\mathbf{v}$ on the diagonal. For a matrix, vecdiag$(\mathbf{A})$ is a vector consisting of the diagonal entries of $\mathbf{A}$. $\odot$ is the Schur–Hadamard (entry-wise) matrix product, $\otimes$ is the Kronecker product, and $\circ$ is the Khatri-Rao product, which is a column-wise Kronecker product:

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

Notable properties are (for matrices and vectors of compatible sizes)

$$\text{vec}(\mathbf{a}\mathbf{b}^H) = \bar{\mathbf{b}} \otimes \mathbf{a} \tag{3}$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = \mathbf{AC} \otimes \mathbf{BD} \tag{4}$$

$$(\mathbf{A} \circ \mathbf{B})^H(\mathbf{C} \circ \mathbf{D}) = \mathbf{A}^H\mathbf{C} \odot \mathbf{B}^H\mathbf{D} \tag{5}$$

$$(\mathbf{a}^H \otimes \mathbf{B})\mathbf{C} = \mathbf{a}^H \otimes \mathbf{BC} \tag{6}$$

$$(\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \circ \mathbf{D}) = \mathbf{AC} \circ \mathbf{BD} \tag{7}$$

$$\text{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\text{vec}(\mathbf{B}) \tag{8}$$

$$\text{vec}(\mathbf{A}\,\text{diag}(\mathbf{b})\,\mathbf{C}) = (\mathbf{C}^T \circ \mathbf{A})\mathbf{b} \tag{9}$$

$$\begin{aligned}[\mathbf{a} \otimes \mathbf{b}][\mathbf{c} \otimes \mathbf{d}]^H &= \mathbf{a} \otimes \mathbf{b}\mathbf{c}^H \otimes \mathbf{d}^H \\ &= \mathbf{c}^H \otimes \mathbf{a}\mathbf{d}^H \otimes \mathbf{b}.\end{aligned} \tag{10}$$

$E(\cdot)$ denotes the expectation operator. For a matrix-valued stochastic variable $\hat{\mathbf{R}}$, define its covariance matrix

$$\text{cov}\{\hat{\mathbf{R}}\} = E\{[\text{vec}(\hat{\mathbf{R}} - E(\hat{\mathbf{R}}))][\text{vec}(\hat{\mathbf{R}} - E(\hat{\mathbf{R}}))]^H\}.$$

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

$$\mathbf{K}_{\mathbf{x}} = \sum_{abcd} (\mathbf{e}_b \otimes \mathbf{e}_a)(\mathbf{e}_c \otimes \mathbf{e}_d)^H \,\text{cum}(x_a, \bar{x}_b, x_c, \bar{x}_d) \tag{11}$$

where

$$\begin{aligned}\text{cum}(x_a, \bar{x}_b, x_c, \bar{x}_d) = &E(x_a\bar{x}_b x_c\bar{x}_d) - E(x_a\bar{x}_b)E(x_c\bar{x}_d) \\ &- E(x_a\bar{x}_d)E(\bar{x}_b x_c) - E(x_a x_c)E(\bar{x}_b\bar{x}_d).\end{aligned}$$

This can be written compactly as

$$\begin{aligned}\mathbf{K}_{\mathbf{x}} = &E(\bar{\mathbf{x}} \otimes \mathbf{x})(\bar{\mathbf{x}} \otimes \mathbf{x})^H - E(\bar{\mathbf{x}} \otimes \mathbf{x})E(\bar{\mathbf{x}} \otimes \mathbf{x})^H \\ &- E(\bar{\mathbf{x}}\bar{\mathbf{x}}^H) \otimes E(\mathbf{x}\mathbf{x}^H) \\ &- E(\bar{\mathbf{x}} \otimes \mathbf{1})(\mathbf{1} \otimes \mathbf{x})^H \odot E(\mathbf{1} \otimes \mathbf{x})(\bar{\mathbf{x}} \otimes \mathbf{1})^H.\end{aligned} \tag{12}$$

For circularly symmetric variables, the last term vanishes.

III. Formulation of the Algorithm

In this section, we present a brief outline of the ACMA in a form that admits its analysis. ACMA consists of two main steps: a prewhitening operation and the algorithm proper (see Fig. 1). We discuss each in turn.

A. Prewhitening

The main purpose of the prewhitening filter is to reduce the data vector dimension from M channels to d, which is the number of sources. This is necessary in the noise-free case to avoid the existence of nullspace beamformers $\mathbf{w}_0$ such that $\mathbf{w}_0^H\mathbf{X} = \mathbf{0}$ since these can be added to any solution $\mathbf{w}$ without changing the outcome and, hence, create a nonuniqueness. In the presence of noise, this dimension reduction gives improved performance. A second purpose of the prewhitening is to whiten the data covariance matrix. Although the algorithm will work without this aspect, it was shown in [16] that the whitening causes the beamformer to converge asymptotically in N to the Wiener beamformer, which is a very desireable feature.

Define the data covariance matrix and its sample estimate

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

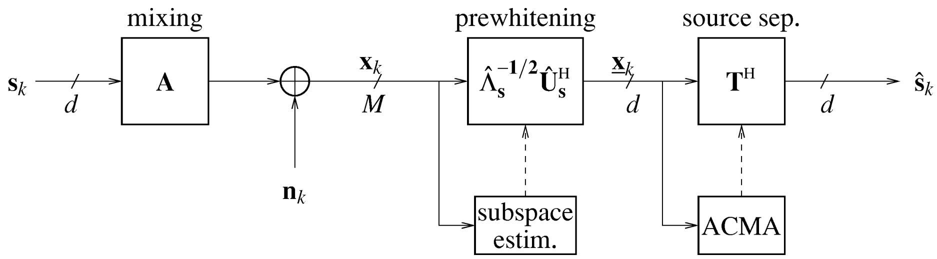

Fig. 1. Data model and beamforming structure.

Since $\mathbf{R_s} = \mathbf{I}$ and $\mathbf{R_n} = \sigma^2\mathbf{I}$, $\mathbf{R_x}$ has as a model

$$\mathbf{R_x} = \mathbf{A}\mathbf{A}^H + \sigma^2\mathbf{I}.$$

We now introduce the eigenvalue decomposition

$$\mathbf{R_x} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = [\,\mathbf{U}_s \quad \mathbf{U}_n\,]\begin{bmatrix}\mathbf{\Lambda}_s & \\ & \mathbf{\Lambda}_n\end{bmatrix}\begin{bmatrix}\mathbf{U}_s^H \\ \mathbf{U}_n^H\end{bmatrix}.$$

$\mathbf{U}$ is a unitary $M \times M$ matrix, and $\mathbf{\Lambda}$ a diagonal matrix whose diagonal contains the eigenvalues in descending order. $\mathbf{U}$ and $\mathbf{\Lambda}$ are partitioned such that the largest d eigenvalues are in $\mathbf{\Lambda}_s$ (the "signal" eigenvalues) and the remaining $M - d$ in $\mathbf{\Lambda}_n$ (the "noise" eigenvalues). Note that the latter are equal to σ^2. Likewise, we can introduce the corresponding sample eigenvectors and eigenvalues from the decomposition of $\hat{\mathbf{R}}_\mathbf{x}$

$$\hat{\mathbf{R}}_\mathbf{x} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^H = \hat{\mathbf{U}}_s\hat{\mathbf{\Lambda}}_s\hat{\mathbf{U}}_s^H + \hat{\mathbf{U}}_n\hat{\mathbf{\Lambda}}_n\hat{\mathbf{U}}_n^H.$$

The prewhitening filter that the algorithm uses is defined by

$$\underline{\mathbf{X}} = [\hat{\mathbf{\Lambda}}_s^{-1/2}\hat{\mathbf{U}}_s^H]\mathbf{X}$$

where the underscore indicates the prewhitening. Note that $\underline{\mathbf{X}}$ has d rows and that $\underline{\hat{\mathbf{R}}}_\mathbf{x} = \mathbf{I}$. The data model in the whitened domain is

$$\underline{\mathbf{X}} = \underline{\mathbf{A}}\mathbf{S} + \underline{\mathbf{N}}$$

where $\underline{\mathbf{A}} = [\hat{\mathbf{\Lambda}}_s^{-1/2}\hat{\mathbf{U}}_s^H]\mathbf{A}$ has size $d \times d$ and is invertible.

B. ACMA Outline

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

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \frac{1}{N}\sum(|\mathbf{w}^H\mathbf{x}_k|^2 - 1)^2.$$

Define

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

and define $\hat{\underline{\mathbf{C}}}\mathbf{x}$ similarly, but based on the whitened data. In [16], 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)

$$\hat{\mathbf{w}} = \underset{\substack{\mathbf{y}=\bar{\mathbf{w}}\otimes\mathbf{w} \\ \mathbf{w}^H\hat{\mathbf{R}}_\mathbf{x}\mathbf{w}=1}}{\operatorname{argmin}} \mathbf{y}^H\hat{\mathbf{C}}_\mathbf{x}\mathbf{y}. \tag{13}$$

If we ignore the effect of dimension reduction, this is equivalent to finding a beamformer $\hat{\mathbf{t}}$ in the whitened domain

$$\hat{\mathbf{t}} = \underset{\substack{\mathbf{y}=\bar{\mathbf{t}}\otimes\mathbf{t} \\ \|\mathbf{y}\|=1}}{\operatorname{argmin}} \mathbf{y}^H\hat{\underline{\mathbf{C}}}_\mathbf{x}\mathbf{y} \tag{14}$$

followed by setting $\hat{\mathbf{w}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}_s^{-1/2}\hat{\mathbf{t}}$. The dimension reduction forces $\hat{\mathbf{w}}$ to lie in the dominant column span of $\mathbf{X}$.

ACMA is obtained as a two-step approach to the latter minimization problem (in the whitened domain) [16]:

1) Find an orthonormal basis $\hat{\mathbf{Y}} = [\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_d]$ of independent minimizers of $\mathbf{y}^H\hat{\underline{\mathbf{C}}}_\mathbf{x}\mathbf{y}$, i.e.,

$$\hat{\mathbf{Y}} = \underset{\mathbf{Y}^H\mathbf{Y}=\mathbf{I}}{\operatorname{argmin}} \operatorname{tr}(\mathbf{Y}^H\hat{\underline{\mathbf{C}}}_\mathbf{x}\mathbf{Y}). \tag{15}$$

The solutions $\hat{\mathbf{y}}_i$ are the eigenvectors corresponding to the d smallest eigenvalues of $\hat{\underline{\mathbf{C}}}_\mathbf{x}$.

2) Find a basis $\{\bar{\hat{\mathbf{t}}}_1 \otimes \hat{\mathbf{t}}_1, \ldots, \bar{\hat{\mathbf{t}}}_d \otimes \hat{\mathbf{t}}_d\}$ that spans the same linear subspace as $\{\hat{\mathbf{y}}_1, \ldots, \hat{\mathbf{y}}_d\}$ and with $\|\hat{\mathbf{t}}_i\| = 1$, i.e., solve

$$\hat{\mathbf{T}} = \min_{\mathbf{T},\mathbf{M}} \|\hat{\mathbf{Y}} - (\bar{\mathbf{T}} \circ \mathbf{T})\mathbf{M}\|_F^2 \tag{16}$$

subject to the constraint $\operatorname{diag}(\mathbf{T}^H\mathbf{T}) = \mathbf{I}$. $\mathbf{M}$ is a $d \times d$ invertible matrix that relates the two bases.

By using (9) of Kronecker products, the second step can also be be written as a joint diagonalization problem

$$\hat{\mathbf{T}} = \min_{\mathbf{T},\{\mathbf{\Lambda}_i\}} \sum \|\hat{\mathbf{Y}}_i - \mathbf{T}\mathbf{\Lambda}_i\mathbf{T}^H\|_F^2$$

where $\hat{\mathbf{Y}}_i = \operatorname{vec}^{-1}(\hat{\mathbf{y}}_i)$, and $\mathbf{\Lambda}_i$ is a diagonal matrix whose diagonal is equal to the ith column of $\mathbf{M}$. The original ACMA paper introduced a Jacobi iteration to (approximately) solve the latter problem. This can then be used as an initial point for a Gauss–Newton algorithm to solve (16) exactly, if so desired [17].

It was shown in [16] 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. Transforming back to the nonwhitened domain, we obtain that $\hat{\mathbf{W}} = \hat{\mathbf{U}}_s\hat{\mathbf{\Lambda}}_s^{-1/2}\hat{\mathbf{T}}$ converges asymptotically to $\mathbf{W} = \mathbf{R}_\mathbf{x}^{-1}\mathbf{A}_0$, which is the Wiener receiver (except for the scaling and the permutation, which are of no consequence to the usual scale-independent performance criteria).

A performance analysis is now possible and consists of the following steps.

1) Find the statistics (covariance) of $\hat{\underline{\mathbf{C}}}_\mathbf{x}$; see Section IV.
2) Find the covariance of the eigenvectors of $\hat{\underline{\mathbf{C}}}_\mathbf{x}$; see Section V.

3) Find perturbation results for the subspace fitting step; see Section VI.

The steps in this outline are identical to the performance analysis of the MUSIC and WSF DOA estimators [20]. However, in that case, the statistics were that of a second-order covariance matrix of Gaussian variables. Here, we need to extend these results to fourth-order non-Gaussian statistics.

The following limitations to the analysis are introduced to keep the derivations tractable.

1) We assume that N is sufficiently large so that we can neglect terms of order N^{-2} compared with terms of order N^{-1}. Similarly, we will assume that the noise power σ^2 is sufficiently small so that we can neglect σ^4 over σ^2.
2) We assume that the prewhitening step is based on the eigenvalue decomposition of the *true* covariance matrix $\mathbf{R_x} = \mathbf{U\Lambda U}^H$.
3) Instead of the joint diagonalization using Jacobi iterations, we assume that the exact solution to (16) is computed by the algorithm, e.g., via a Gauss–Newton iteration.

Thus, we will analyze a slightly different algorithm than the original ACMA. If $M = d$, then it can be shown that the difference in prewhitening using $\mathbf{R_x}$ instead of $\hat{\mathbf{R}}_\mathbf{x}$ is negligible for N sufficiently large. However, if $M > d$, the dimension reduction in the prewhitening step will introduce an additional, complicated effect that is not incorporated in the analysis. Since this form of prewhitening is quite common, a detailed study of this is of independent interest and, hence, is omitted here. Thus, our analysis is valid only for $M = d$.

IV. Variance 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. The whitening is of no consequence for the results in this section: We will not use the fact that $\mathbf{R_x} = \mathbf{I}$.

Our objective is to find an expression for the covariance of $\hat{\mathbf{C}}_\mathbf{x}$, which is denoted by $\mathbf{\Omega_x}$. Since $\hat{\mathbf{C}}_\mathbf{x}$ contains fourth-order moments, its covariance involves eighth-order statistics. A precise description has very many terms and does not give additional insight (cf. [4] and [14, App.A] for the covariance of fourth-order *cumulant* matrices). Here, we derive a compact approximation.

A. Model for $\hat{\mathbf{C}}_\mathbf{X}$

Define

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

Then, $\mathbf{C_x}$ has model

$$\mathbf{C_x} = \mathbf{K_x} + \bar{\mathbf{R}}_\mathbf{x} \otimes \mathbf{R_x}$$

where $\mathbf{R_x} = \mathbf{AA}^H + \mathbf{R_n}$, and $\mathbf{K_x}$ denotes the fourth-order cumulant matrix with entries $\mathrm{cum}(x_k, \bar{x}_\ell, x_m, \bar{x}_n)$ as defined in (12) and satisfies the model

$$\begin{aligned}\mathbf{K_x} &= [\bar{\mathbf{A}} \otimes \mathbf{A}]\mathbf{K_s}[\bar{\mathbf{A}} \otimes \mathbf{A}]^H + \mathbf{K_n} \\ &= [\bar{\mathbf{A}} \circ \mathbf{A}][\bar{\mathbf{A}} \circ \mathbf{A}]^H\end{aligned}$$

since $\mathbf{K_s} = -[\mathbf{I}_d \circ \mathbf{I}_d][\mathbf{I}_d \circ \mathbf{I}_d]^H$ and $\mathbf{K_n} = 0$; we also used (7).

$\mathbf{C_x}$ has an interpretation as the covariance of $\hat{\mathbf{R}}_\mathbf{x}$. Indeed

$$\begin{aligned}\mathrm{cov}\{\hat{\mathbf{R}}_\mathbf{x}\} &= E\{[\mathrm{vec}(\hat{\mathbf{R}}_\mathbf{x}) - \mathrm{vec}(\mathbf{R_x})][\mathrm{vec}(\hat{\mathbf{R}}_\mathbf{x}) - \mathrm{vec}(\mathbf{R_x})]^H\} \\ &= E\left\{\left[\frac{1}{N}\sum \bar{\mathbf{x}} \otimes \mathbf{x} - E(\bar{\mathbf{x}} \otimes \mathbf{x})\right] \cdot \right. \\ &\qquad \left. \left[\frac{1}{N}\sum \bar{\mathbf{x}} \otimes \mathbf{x} - E(\bar{\mathbf{x}} \otimes \mathbf{x})\right]^H\right\} \\ &= \frac{1}{N}\left[E\{(\bar{\mathbf{x}} \otimes \mathbf{x})(\bar{\mathbf{x}} \otimes \mathbf{x})^H\}\right. \\ &\qquad \left. - E\{\bar{\mathbf{x}} \otimes \mathbf{x}\}E\{\bar{\mathbf{x}} \otimes \mathbf{x}\}^H\right] \\ &= \frac{1}{N}\mathbf{C_x}. \end{aligned} \tag{17}$$

Thus, $1/N\mathbf{C_x}$ is the covariance of $\hat{\mathbf{R}}_\mathbf{x}$, and $\hat{\mathbf{C}}_\mathbf{x}$ provides a sample estimate of this covariance. However, it is a biased estimate since some simple but tedious manipulations show that

$$\begin{aligned}E\{\hat{\mathbf{C}}_\mathbf{x}\} &= E\left\{\frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x})(\bar{\mathbf{x}} \otimes \mathbf{x})^H \right. \\ &\qquad \left. - \frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x})\frac{1}{N}\left(\sum \bar{\mathbf{x}} \otimes \mathbf{x}\right)^H\right\} \\ &= \left(1 - \frac{1}{N}\right)\mathbf{C_x}.\end{aligned}$$

This is entirely analogous to the bias that usually occurs in an estimated variance when the mean is also estimated from the data.

A second interpretation for $\mathbf{C_x}$ is obtained by defining a "data" sequence

$$\mathbf{g}_k := \bar{\mathbf{x}}_k \otimes \mathbf{x}_k - E\{\bar{\mathbf{x}}_k \otimes \mathbf{x}_k\}, \quad k = 1, \ldots, N \tag{18}$$

and considering its covariance and sample covariance

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

Lemma 1:

$$\begin{aligned}\mathbf{R_g} &= \mathbf{C_x} \\ E\{\hat{\mathbf{R}}_\mathbf{g}\} &= \mathbf{R_g} = \mathbf{C_x} \\ \hat{\mathbf{R}}_\mathbf{g} &= \hat{\mathbf{C}}_\mathbf{x}\left(1 + \mathcal{O}(N^{-1})\right).\end{aligned}$$

Proof: The first two properties are straightforward. To prove the third property, note that

$$\begin{aligned}\hat{\mathbf{R}}_\mathbf{g} &= \frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x} - E\{\bar{\mathbf{x}} \otimes \mathbf{x}\})(\bar{\mathbf{x}} \otimes \mathbf{x} - E\{\bar{\mathbf{x}} \otimes \mathbf{x}\})^H \\ &= \left[\frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x})(\bar{\mathbf{x}} \otimes \mathbf{x})^H \right. \\ &\qquad \left. - \frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x})\,\frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x})^H\right] \\ &\qquad + \left[\frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x}) - E\{\bar{\mathbf{x}} \otimes \mathbf{x}\}\right] \\ &\qquad \times \left[\frac{1}{N}\sum(\bar{\mathbf{x}} \otimes \mathbf{x}) - E\{\bar{\mathbf{x}} \otimes \mathbf{x}\}\right]^H.\end{aligned}$$

The first term in brackets is recognized as $\hat{\mathbf{C}}_\mathbf{x}$. For the second term, note that $[(1/N)\sum(\bar{\mathbf{x}} \otimes \mathbf{x}) - E\{\bar{\mathbf{x}} \otimes \mathbf{x}\}\ ]$ has covariance $1/N\mathbf{C_x}$ [see (17)] so that $\hat{\mathbf{R}}_\mathbf{g} - \hat{\mathbf{C}}_\mathbf{x} = \mathcal{O}(1/N\mathbf{C_x}) = \mathcal{O}(1/N\hat{\mathbf{C}}_\mathbf{x})$. □

Thus, $\mathbf{C_x}$ 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 (17), it follows that $\mathrm{cov}\{\hat{\mathbf{R}}_\mathbf{g}\} = 1/N\mathbf{C_g}$, where

$$\mathbf{C_g} := E\{(\bar{\mathbf{g}} \otimes \mathbf{g})(\bar{\mathbf{g}} \otimes \mathbf{g})^H\} - E\{\bar{\mathbf{g}} \otimes \mathbf{g}\}E\{\bar{\mathbf{g}} \otimes \mathbf{g}\}^H. \quad (19)$$

In summary, we have proven the following theorem.

Theorem 2: $\mathbf{\Omega_x} := \mathrm{cov}\{\hat{\mathbf{C}}_\mathbf{x}\} = 1/N\mathbf{C_g} + \mathcal{O}(1/N^2)$.

A compact description of $\mathbf{C_g}$ in terms of our data model is derived next.

B. Approximate Expression for $\mathbf{C_g}$

Inserting the model $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k$ in the definition of $\mathbf{g}_k$, we obtain

$$\mathbf{g}_k = (\bar{\mathbf{A}} \otimes \mathbf{A})\underset{\sim}{\mathbf{s}}_k + \underset{\sim}{\mathbf{n}}_k, \quad k = 1, \ldots, N$$

where

$$\begin{aligned}\underset{\sim}{\mathbf{s}} &:= \bar{\mathbf{s}} \otimes \mathbf{s} - \mathrm{vec}(\mathbf{I}) = \sum_{i \neq j} \mathbf{e}_{ij}\bar{s}_j s_i \\ \underset{\sim}{\mathbf{n}} &:= \bar{\mathbf{n}} \otimes \mathbf{n} - \mathbf{R_n} + \bar{\mathbf{A}}\bar{\mathbf{s}} \otimes \mathbf{n} + \bar{\mathbf{n}} \otimes \mathbf{A}\mathbf{s}.\end{aligned}$$

Here, $\mathbf{e}_{ij} := \mathbf{e}_j \otimes \mathbf{e}_i = \mathrm{vec}(\mathbf{e}_i\mathbf{e}_j^H)$ is a vector with only a single nonzero entry. The vector $\underset{\sim}{\mathbf{s}}$ has entries that are either zero or constant modulus (CM). We can conveniently drop the zero entries by defining a truncated vector $\mathbf{e}'_{ij} = \mathrm{vec}'(\mathbf{e}_i\mathbf{e}_j^H)$, where $\mathrm{vec}'(\cdot)$ is a vectoring operator that skips the main diagonal. We thus obtain a model

$$\mathbf{g}_k = \mathbf{A_c}\mathbf{c}_k + \underset{\sim}{\mathbf{n}}_k$$

where

$$\begin{aligned}\mathbf{c} &:= \sum_{i \neq j} \mathbf{e}'_{ij}\bar{s}_j s_i \\ &= [\bar{s}_1 s_2, \ldots, \bar{s}_1 s_d, \bar{s}_2 s_1, \bar{s}_2 s_3, \ldots]^T \\ \mathbf{A_c} &:= [\bar{\mathbf{a}}_j \otimes \mathbf{a}_i]_{i \neq j} \\ &= [\bar{\mathbf{a}}_1 \otimes \mathbf{a}_2, \ldots, \bar{\mathbf{a}}_1 \otimes \mathbf{a}_d, \bar{\mathbf{a}}_2 \otimes \mathbf{a}_1, \bar{\mathbf{a}}_2 \otimes \mathbf{a}_3, \ldots].\end{aligned}$$

The vector $\mathbf{c}$ is CM (with certain dependencies among its entries). Likewise, the matrix $\mathbf{A_c}$ skips the $\bar{\mathbf{a}}_i \otimes \mathbf{a}_i$ columns of $\bar{\mathbf{A}} \otimes \mathbf{A}$.

The model $\mathbf{g}_k = \mathbf{A_c}\mathbf{c}_k + \underset{\sim}{\mathbf{n}}_k$ has several properties that are similar to that of $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k$. In particular, since

$$E(\mathbf{c}) = \mathbf{0}, \qquad E(\underset{\sim}{\mathbf{n}}) = \mathbf{0}, \qquad E(\mathbf{c}\underset{\sim}{\mathbf{n}}^H) = \mathbf{0}$$

we have that

$$\mathbf{R_g} = \mathbf{A_c}\mathbf{R_c}\mathbf{A_c}^H + \mathbf{R}_{\underset{\sim}{\mathbf{n}}}$$

where

$$\begin{aligned}\mathbf{R_c} &= E(\mathbf{c}\mathbf{c}^H) = \mathbf{I} \\ \mathbf{R}_{\underset{\sim}{\mathbf{n}}} &= E(\underset{\sim}{\mathbf{n}}\underset{\sim}{\mathbf{n}}^H) = \bar{\mathbf{R}}_\mathbf{n} \otimes \mathbf{R_n} + \bar{\mathbf{A}}\bar{\mathbf{A}}^H \otimes \mathbf{R_n} + \bar{\mathbf{R}}_\mathbf{n} \otimes \mathbf{A}\mathbf{A}^H.\end{aligned}$$

For the model of $\mathbf{x}_k$, we know that $\mathbf{C_x} = \mathbf{K_x} + \bar{\mathbf{R}}_\mathbf{x} \otimes \mathbf{R_x}$, where $\mathbf{R_x} = \mathbf{A}\mathbf{A}^H + \mathbf{R_n}$, and $\mathbf{K_x} = [\bar{\mathbf{A}} \otimes \mathbf{A}]\mathbf{K_s}[\bar{\mathbf{A}} \otimes \mathbf{A}]^H$. A first idea is that analogously, $\mathbf{C_g}$ defined in (19) can be written as $\mathbf{C_g} \approx \mathbf{K_g} + \bar{\mathbf{R}}_\mathbf{g} \otimes \mathbf{R_g}$ with $\mathbf{K_g} \approx [\bar{\mathbf{A}}_\mathbf{c} \otimes \mathbf{A_c}]\mathbf{K_c}[\bar{\mathbf{A}}_\mathbf{c} \otimes \mathbf{A_c}]^H$. Since $\mathbf{c}$ and $\underset{\sim}{\mathbf{n}}$ are not independent (only uncorrelated) and not circularly symmetric, and $\mathbf{K}_{\underset{\sim}{\mathbf{n}}} \neq \mathbf{0}$, this is not true with equality. Certain cross-terms are ignored. In addiiton, $\mathbf{K_c}$ does not have a simple diagonal structure because the entries of $\mathbf{c}$ are related. We now set out to find a more accurate description of $\mathbf{C_g}$, only taking into account the terms up to $\mathcal{O}(\sigma^2)$.

Theorem 3: $\mathbf{C_g} = [\bar{\mathbf{A}}_\mathbf{c} \otimes \mathbf{A_c}]\mathbf{K}'_\mathbf{c}[\bar{\mathbf{A}}_\mathbf{c} \otimes \mathbf{A_c}]^H + \bar{\mathbf{R}}_\mathbf{g} \otimes \mathbf{R_g} + \mathbf{E} + \mathbf{E}^H + \mathcal{O}(\sigma^4)$, where $\mathbf{E} = [\mathbf{A} \otimes \bar{\mathbf{R}}_\mathbf{n}^{1/2} \otimes \mathbf{A_c}]\mathbf{E}_1[\bar{\mathbf{A}}_\mathbf{c} \otimes \bar{\mathbf{R}}_\mathbf{n}^{1/2} \otimes \mathbf{A}]^H + [\mathbf{R}_\mathbf{n}^{1/2} \otimes \bar{\mathbf{A}} \otimes \mathbf{A_c}]\mathbf{E}_2[\bar{\mathbf{A}}_\mathbf{c} \otimes \bar{\mathbf{A}} \otimes \mathbf{R}_\mathbf{n}^{1/2}]^H$, and

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

(All indices range over $1, \ldots, d$. Note that the latter matrices are data independent and represent simply collections of "1" entries.)

Proof: See Appendix A.

Fig. 2 gives an overview of the structure of the model for $\mathbf{C_g}$ and its components, for $d = 3$, $\mathbf{A} = \mathbf{I}$, and $\mathbf{R_n} = \sigma^2\mathbf{I}$. It is seen that the dominant term is $\bar{\mathbf{R}}_\mathbf{g} \otimes \mathbf{R_g}$ but that there are many other terms that are caused by the dependencies among the entries of $\mathbf{c}$ and $\underset{\sim}{\mathbf{n}}$.

The preceding model for $\mathbf{C_g}$ gives a good description of the covariance of $\hat{\mathbf{C}}_\mathbf{x}$ for reasonably large N and signal-to-noise ratio (SNR). Note that if $\mathbf{c}$ and $\underset{\sim}{\mathbf{n}}$ are simply regarded as Gaussian vectors with independent entries, we obtain

$$\mathbf{C_g} \approx \bar{\mathbf{C}}_\mathbf{x} \otimes \mathbf{C_x}. \quad (20)$$

Making this perhaps crude approximation would lead to particularly simple results in the eigenvector perturbation study and

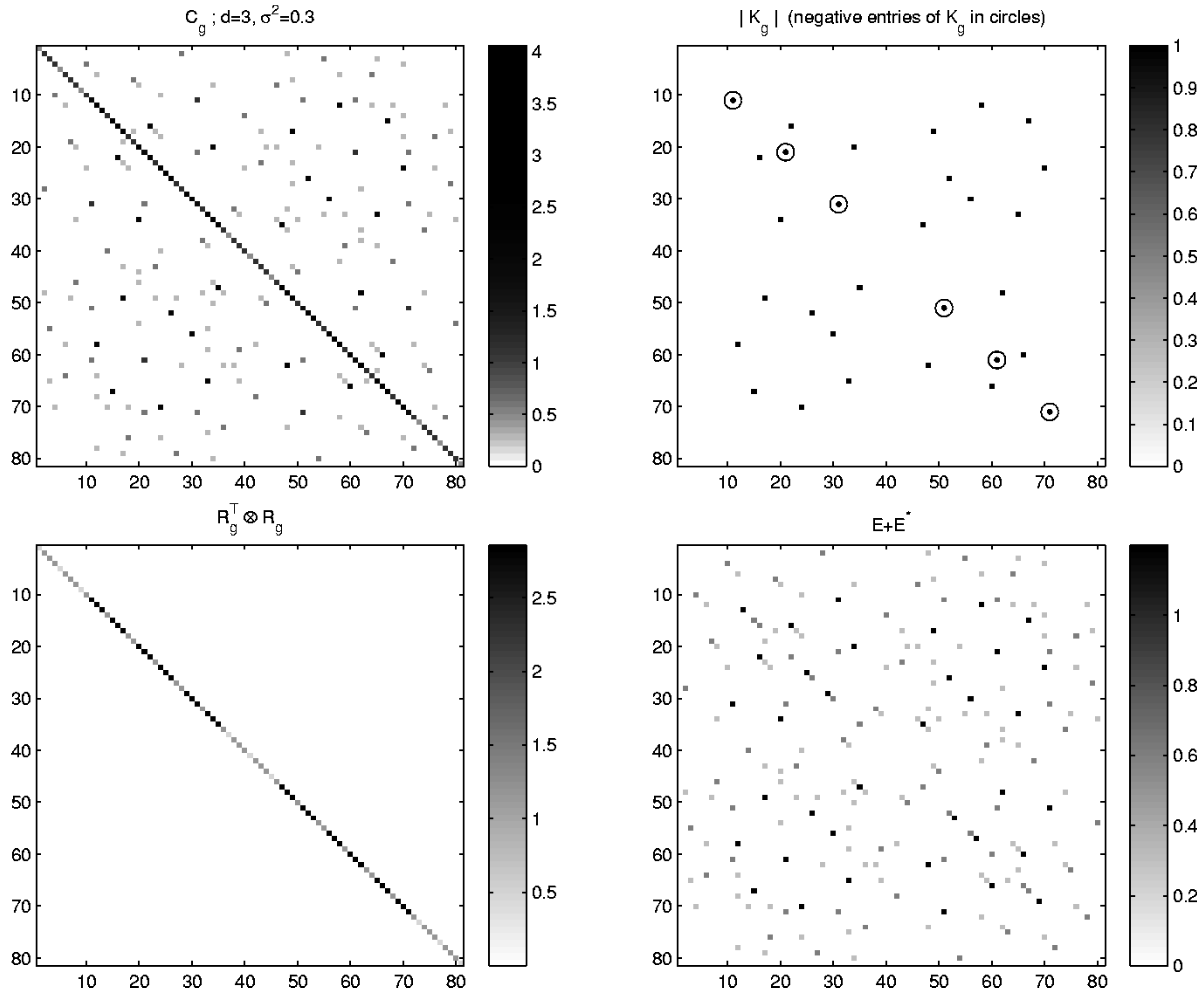

Fig. 2. Structure of $\mathbf{C}_g$ and its components. $\mathbf{A} = \mathbf{I}$, $\mathbf{R}_n = \sigma^2\mathbf{I}$, $d = 3$, and $\sigma^2 = 0.3$.

subsequent steps as we basically can apply the theory in Viberg [20].

V. Eigenvector Perturbation

All variables in this section are based on whitened data, and we drop the underscore from the notation. The results in Section V-B depend on the whitening.

In this section, we consider the statistical properties of the eigenvectors of $\hat{\mathbf{C}}_{\mathbf{x}}$, which is a fourth-order sample covariance matrix based on non-Gaussian signals. We first give a general derivation and then specialize to the case at hand. The generalization is needed because existing derivations typically consider Gaussian sources, e.g., standard results for a second-order covariance matrix of a Gaussian signal have been derived in [1], and extended in [8] for complex Gaussian signals. Results for deterministic signals in Gaussian noise can be found in [12]. A general result appears in [23], based on the (real-valued) perturbation analysis in Wilkinson [22]. The derivation in what follows leads to essentially the same result but written more compactly in tensor notation and with a self-contained proof.

A. General

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 $\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, $\hat{\mathbf{R}} = \hat{\mathbf{U}}\hat{\mathbf{\Lambda}}\hat{\mathbf{U}}^H$. Note that

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

If the eigenvalues are distinct and the phase ambiguity of the eigenvectors is resolved in some default manner (see [6]), then $\hat{\mathbf{R}} \to \mathbf{R}$ implies $\hat{\mathbf{U}} \to \mathbf{U}$ and $\hat{\mathbf{\Lambda}} \to \mathbf{\Lambda}$. Thus, in first-order approximation

$$\begin{aligned}
\hat{\mathbf{R}} - \mathbf{R} &= (\hat{\mathbf{U}} - \mathbf{U})\mathbf{\Lambda}\mathbf{U}^H - \mathbf{R}(\hat{\mathbf{U}} - \mathbf{U})\mathbf{U}^H \\
&\quad + \mathbf{U}(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda})\mathbf{U}^H \\
&\Leftrightarrow \\
\mathrm{vec}(\hat{\mathbf{R}} - \mathbf{R}) &= [\bar{\mathbf{U}} \otimes \mathbf{I}][\mathbf{\Lambda} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{R}]\mathrm{vec}(\hat{\mathbf{U}} - \mathbf{U}) \\
&\quad + [\bar{\mathbf{U}} \circ \mathbf{U}]\mathrm{vecdiag}(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}) \\
&= [\bar{\mathbf{U}} \otimes \mathbf{U}][\mathbf{\Lambda} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}][\mathbf{I} \otimes \mathbf{U}^H]\mathrm{vec}(\hat{\mathbf{U}} - \mathbf{U}) \\
&\quad + [\bar{\mathbf{U}} \circ \mathbf{U}]\mathrm{vecdiag}(\hat{\mathbf{\Lambda}} - \mathbf{\Lambda}). \qquad (21)
\end{aligned}$$

Assume now that we partition the eigenvalue decomposition of $\mathbf{R}$ as

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H = [\,\mathbf{U}_s \quad \mathbf{U}_n\,] \begin{bmatrix} \mathbf{\Lambda}_s & \\ & \mathbf{\Lambda}_n \end{bmatrix} [\,\mathbf{U}_s \quad \mathbf{U}_n\,]^H \qquad (22)$$

where (in this section) the partitioning is arbitrary, as long as the eigenvalues in $\mathbf{\Lambda}_s$ are distinct and unequal to any eigenvalue in

$\mathbf{\Lambda}_n$. We are interested in the perturbations of the estimate of $\mathbf{U}_s$ in directions orthogonal to this subspace.

Note that $[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H[\bar{\mathbf{U}} \circ \mathbf{U}] = ([\mathbf{I} \quad 0] \circ [0 \quad \mathbf{I}]) = 0$. Thus, premultiplying (21) with $[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H$ removes the second term, giving

$$\begin{aligned}
&[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H \mathrm{vec}(\hat{\mathbf{R}} - \mathbf{R})\\
&=([\mathbf{I} \quad 0] \otimes [0 \quad \mathbf{I}])[\mathbf{\Lambda} \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}][\mathbf{I} \otimes \mathbf{U}^H]\mathrm{vec}(\hat{\mathbf{U}} - \mathbf{U})\\
&=[\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]([\mathbf{I} \quad 0] \otimes \mathbf{U}_n^H)\mathrm{vec}(\hat{\mathbf{U}} - \mathbf{U})\\
&\Leftrightarrow\\
&[\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]^{-1}[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H \mathrm{vec}(\hat{\mathbf{R}} - \mathbf{R})\\
&=([\mathbf{I} \quad 0] \otimes \mathbf{U}_n^H)\mathrm{vec}(\hat{\mathbf{U}} - \mathbf{U})\\
&\Rightarrow [\mathbf{I} \otimes \mathbf{U}_n][\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]^{-1}[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H \mathrm{vec}(\hat{\mathbf{R}} - \mathbf{R})\\
&=([\mathbf{I} \quad 0] \otimes \mathbf{P}_n)\mathrm{vec}(\hat{\mathbf{U}} - \mathbf{U})\\
&=\mathrm{vec}(\mathbf{P}_n \hat{\mathbf{U}}_s)
\end{aligned}$$

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

Lemma 4: Let $\hat{\mathbf{R}}$ be a sample covariance matrix converging to $\mathbf{R}$, and assume that $\mathbf{R}$ has eigenvalue decomposition (22), where the entries in $\mathbf{\Lambda}_s$ are distinct and unequal to any entry in $\mathbf{\Lambda}_n$. Then

$$\begin{aligned}
&\mathrm{cov}\{\mathbf{P}_n \hat{\mathbf{U}}_s\}\\
&=[\mathbf{I} \otimes \mathbf{U}_n][\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]^{-1}[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H \cdot \mathrm{cov}\{\hat{\mathbf{R}}\}\\
&\quad \cdot [\bar{\mathbf{U}}_s \otimes \mathbf{U}_n][\mathbf{\Lambda}_s \otimes \mathbf{I} - \mathbf{I} \otimes \mathbf{\Lambda}_n]^{-1}[\mathbf{I} \otimes \mathbf{U}_n]^H + o(N^{-1}).
\end{aligned} \tag{23}$$

Here are two remarks as an aside. If $\mathbf{R}$ is the covariance matrix of a Gaussian signal vector, then $\mathrm{cov}\{\hat{\mathbf{R}}\} = (1/N)\bar{\mathbf{R}} \otimes \mathbf{R}$, and the usual eigenvector perturbation result [8] follows. A second observation is that if $\mathbf{R}$ is the covariance due to a data model $\mathbf{x}_k = \mathbf{A}\mathbf{s}_k + \mathbf{n}_k$, where the sources can have any fourth-order statistics but the noise is Gaussian, then $\mathrm{cov}\{\hat{\mathbf{R}}\} = 1/N(\bar{\mathbf{R}} \otimes \mathbf{R} + [\bar{\mathbf{A}} \otimes \mathbf{A}]\mathbf{K_s}[\bar{\mathbf{A}} \otimes \mathbf{A}]^H)$. Suppose the noise is white, and we partition the eigenvalues such that $\mathbf{\Lambda}_s$ contains the d largest eigenvalues. Then, $\mathbf{U}_n \perp \mathbf{A}$; hence, $[\bar{\mathbf{U}}_s \otimes \mathbf{U}_n]^H[\bar{\mathbf{A}} \otimes \mathbf{A}] = 0$. The term in $\mathrm{cov}\{\hat{\mathbf{R}}\}$ contributed by $\mathbf{K_s}$ drops out. Hence, in first-order approximation, *the projected eigenvector statistics* $\mathrm{cov}\{\mathbf{P}_n\hat{\mathbf{U}}_s\}$ *are independent of the higher order source statistics and only depend on* $\mathbf{R}$. A consequence is that the performance of subspace-based algorithms such as MUSIC, ESPRIT, or WSF do not depend on whether the sources are Gaussian or not. This corroborates results in [3] and [12].

B. Application

We now specialize to our situation. We have

$$\begin{aligned}
\mathbf{R} \leftrightarrow \mathbf{R_g} &= \mathbf{C_x}\\
\mathrm{cov}\{\hat{\mathbf{R}}\} \leftrightarrow \mathbf{\Omega_x} &= \mathrm{cov}\{\hat{\mathbf{C}}_\mathbf{x}\} = \frac{1}{N}\mathbf{C_g} + O(N^{-2}).
\end{aligned}$$

Introduce the eigenvalue decomposition of $\mathbf{C_x}$ as

$$\begin{aligned}
\mathbf{C_x} &= \underset{\sim}{\mathbf{U}}\underset{\sim}{\mathbf{\Lambda}}\underset{\sim}{\mathbf{U}}^H\\
&= \begin{bmatrix} \underset{\sim}{\mathbf{U}}_s & \underset{\sim}{\mathbf{U}}_n \end{bmatrix} \begin{bmatrix} \underset{\sim}{\mathbf{\Lambda}}_s & \\ & \underset{\sim}{\mathbf{\Lambda}}_n \end{bmatrix} \begin{bmatrix} \underset{\sim}{\mathbf{U}}_s & \underset{\sim}{\mathbf{U}}_n \end{bmatrix}^H
\end{aligned} \tag{24}$$

where $\underset{\sim}{\mathbf{\Lambda}}_s$ collects the d *smallest* eigenvalues of $\mathbf{C_x}$. Likewise, $\underset{\sim}{\mathbf{U}}_s$ is a basis for the approximate null space of $\mathbf{C_x}$. The model for $\mathbf{C_x}$, in the whitened domain, is

$$\begin{aligned}
\mathbf{C_x} &= -[\bar{\mathbf{A}} \circ \mathbf{A}][\bar{\mathbf{A}} \circ \mathbf{A}]^H + \bar{\mathbf{R}}_\mathbf{x} \otimes \mathbf{R_x}\\
&= -[\bar{\mathbf{A}} \circ \mathbf{A}][\bar{\mathbf{A}} \circ \mathbf{A}]^H + \mathbf{I}.
\end{aligned} \tag{25}$$

Introduce the singular value decomposition

$$\underset{\sim}{\mathbf{A}} = \bar{\mathbf{A}} \circ \mathbf{A} = \underset{\sim}{\mathbf{U}}_\mathbf{A}\underset{\sim}{\mathbf{\Sigma}}_\mathbf{A}\underset{\sim}{\mathbf{V}}_\mathbf{A} \tag{26}$$

where $\underset{\sim}{\mathbf{U}}_\mathbf{A}$ has d orthonormal columns, $\underset{\sim}{\mathbf{\Sigma}}_\mathbf{A} = \mathrm{diag}[\sigma_k]$ is a $d \times d$ diagonal matrix, and $\underset{\sim}{\mathbf{V}}_\mathbf{A}$ is $d \times d$ unitary. Let $\underset{\sim}{\mathbf{U}}_\mathbf{A}^\perp$ be the orthogonal complement of $\underset{\sim}{\mathbf{U}}_\mathbf{A}$. It follows from (25) that the eigenvalue decomposition of $\mathbf{C_x}$ is given by

$$\mathbf{C_x} = \begin{bmatrix} \underset{\sim}{\mathbf{U}}_\mathbf{A} & \underset{\sim}{\mathbf{U}}_\mathbf{A}^\perp \end{bmatrix} \begin{bmatrix} \mathbf{I} - \underset{\sim}{\mathbf{\Sigma}}_A^2 & \\ & \mathbf{I} \end{bmatrix} \begin{bmatrix} \underset{\sim}{\mathbf{U}}_\mathbf{A} & \underset{\sim}{\mathbf{U}}_\mathbf{A}^\perp \end{bmatrix}^H. \tag{27}$$

In view of the partitioning in (24), we set the "signal" subspace $\underset{\sim}{\mathbf{U}}_s = \underset{\sim}{\mathbf{U}}_\mathbf{A}$ and $\underset{\sim}{\mathbf{\Lambda}}_s = \mathbf{I} - \underset{\sim}{\mathbf{\Sigma}}_A^2$. The "noise" eigenvalues are all equal and given by $\underset{\sim}{\mathbf{\Lambda}}_n = \mathbf{I}$.

Inserting this in (23), we obtain the needed expression for the eigenvector perturbation.

Theorem 5: The covariance of the estimated signal subspace eigenvectors $\hat{\underset{\sim}{\mathbf{U}}}_s$ of $\hat{\mathbf{C}}_\mathbf{x}$ outside the true signal subspace spanned by $\underset{\sim}{\mathbf{U}}_\mathbf{A}$ is

$$\mathrm{cov}\{\underset{\sim}{\mathbf{P}}_\mathbf{A}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s\} = \frac{1}{N}\mathbf{C_u} + o(N^{-1})$$

where

$$\mathbf{C_u} := [\underset{\sim}{\mathbf{\Sigma}}_\mathbf{A}^{-2}\bar{\underset{\sim}{\mathbf{U}}}_\mathbf{A}^H \otimes \underset{\sim}{\mathbf{P}}_\mathbf{A}^\perp]\mathbf{C_g}[\bar{\underset{\sim}{\mathbf{U}}}_\mathbf{A}\underset{\sim}{\mathbf{\Sigma}}_\mathbf{A}^{-2} \otimes \bar{\underset{\sim}{\mathbf{P}}}_\mathbf{A}^\perp].$$

An approximate model for $\mathbf{C_g}$ was shown previously in Theorem 3. Note that the higher order terms in this case do *not* drop out since $\bar{\underset{\sim}{\mathbf{P}}}_\mathbf{A}^\perp \mathbf{A_c} \neq 0$.

VI. Subspace Fitting

All variables in this section are based on whitened data, and we drop the underscore from the notation.

A. Cost Function

For the analysis of ACMA, we assume that the joint diagonalization step is implemented as the solution to the subspace fitting problem in (16). This allows us to follow in outline the performance analysis technique described in [20].[1] Some notational changes are necessary.

In (16), we described the subspace fitting problem as the computation of a $d \times d$ separating beamforming matrix $\hat{\mathbf{T}}$ (in the whitened domain). The columns of $\hat{\mathbf{T}}$ were constrained to have unit norm, and we can further constrain the first nonzero entry of each column to be positive real. With some abuse of notation, 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 that is unidentifiable by the subspace fitting. We assume that the true parameter vector $\boldsymbol{\theta}_0$ is uniquely

[1]See also [2], which contains many statistical properties related to subspace fitting but, unfortunately, only considers the real domain. In particular, it gives proofs for semi-definite weightings.

identifiable and that $\mathbf{A}(\boldsymbol{\theta})$ is continuously differentiable around $\boldsymbol{\theta}_0$.

We proved in [16] that as $N \to \infty$, $\hat{\mathbf{T}}$ converges to $\mathbf{A}_0 \equiv \mathbf{A}(\boldsymbol{\theta}_0)$, and thus, we can regard $\hat{\mathbf{T}}$ as an estimate of $\mathbf{A}_0$ and write $\hat{\mathbf{T}} = \mathbf{A}(\hat{\boldsymbol{\theta}})$. In this notation, (16) becomes

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

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

$$\mathbf{A}(\boldsymbol{\theta}) = \underset{\mathbf{A}(\boldsymbol{\theta})}{\operatorname{argmin}} \|\underset{\sim}{\mathbf{P}}^\perp_{\mathbf{A}(\boldsymbol{\theta})} \hat{\underset{\sim}{\mathbf{U}}}_s\|_F^2$$

where $\underset{\sim}{\mathbf{P}}^\perp_{\mathbf{A}(\boldsymbol{\theta})} = \mathbf{I} - \underset{\sim}{\mathbf{A}}(\boldsymbol{\theta})\underset{\sim}{\mathbf{A}}(\boldsymbol{\theta})^\dagger$. It is further customary to solve the latter subspace fitting problem in a more general *weighted* norm, namely, $\|\mathbf{X}\|^2_{\boldsymbol{\Gamma}} := \mathrm{vec}(\mathbf{X})^H \boldsymbol{\Gamma} \mathrm{vec}(\mathbf{X})$, where $\boldsymbol{\Gamma}$ is a positive definite weighting matrix that can be used to minimize the estimator variance. Hence, we will consider the minimization of the cost function

$$J(\boldsymbol{\theta}) = \|\underset{\sim}{\mathbf{P}}^\perp_{\mathbf{A}(\boldsymbol{\theta})} \hat{\underset{\sim}{\mathbf{U}}}_s\|^2_{\boldsymbol{\Gamma}} = \mathrm{vec}(\underset{\sim}{\mathbf{P}}^\perp_{\mathbf{A}(\boldsymbol{\theta})} \hat{\underset{\sim}{\mathbf{U}}}_s)^H \boldsymbol{\Gamma} \mathrm{vec}(\underset{\sim}{\mathbf{P}}^\perp_{\mathbf{A}(\boldsymbol{\theta})} \hat{\underset{\sim}{\mathbf{U}}}_s) \tag{28}$$

where $\boldsymbol{\Gamma}$ is positive (semi-)definite.

B. Variance of $\boldsymbol{\theta}$

Denote the gradient of $J(\boldsymbol{\theta})$ by the vector $\mathbf{J}'(\boldsymbol{\theta}) = [\partial J / \partial \theta_i(\boldsymbol{\theta})]$ and the second derivative (Hessian) by the matrix $\mathbf{J}''(\boldsymbol{\theta}) = [\partial^2 J / \partial\theta_i \partial\theta_j(\boldsymbol{\theta})]$, and define

$$\begin{aligned} \mathbf{Q} &:= \lim_{N\to\infty} NE\{\mathbf{J}'(\boldsymbol{\theta}_0)\,\mathbf{J}'(\boldsymbol{\theta}_0)^H\} \\ \mathbf{H} &:= \lim_{N\to\infty} \mathbf{J}''(\boldsymbol{\theta}_0). \end{aligned} \tag{29}$$

Following [20], note that at the minimum of the cost function, $\mathbf{J}'(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. Since $\hat{\boldsymbol{\theta}}$ is strongly consistent, a first-order Taylor expansion of $\mathbf{J}'(\boldsymbol{\theta})$ around $\boldsymbol{\theta}_0$ then leads to $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \approx -\mathbf{H}^{-1}\mathbf{J}'(\boldsymbol{\theta}_0)$, which gives a description of the variance of $\hat{\boldsymbol{\theta}}$, as follows.

Lemma 6: (Viz. [2] and [20]) $\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is asymptotically normal distributed, with zero mean and covariance $\mathbf{R}_{\boldsymbol{\theta}} = \mathbf{H}^{-1}\mathbf{Q}\mathbf{H}^{-1}$, where $\mathbf{Q}$ and $\mathbf{H}$ are defined in (29).

In view of this lemma, explicit expressions for $\mathbf{Q}$ and $\mathbf{H}$, which are a function of the specific choice for the parametrization of $\mathbf{A}(\boldsymbol{\theta})$, remain to be found. Since the columns of $\mathbf{A}(\boldsymbol{\theta})$ are not coupled, we can write $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\boldsymbol{\theta}_1), \ldots, \mathbf{a}(\boldsymbol{\theta}_d)]$, where $\mathbf{a}(\boldsymbol{\theta}_i)$ is a parametrization of a unit-norm vector with real nonnegative first entry. A possible parametrization is given in Appendix B.[2] Suppose that the number of (real-valued) parameters per vector is p. For future notational convenience, we arrange the parameters of $\mathbf{A}(\boldsymbol{\theta})$ in two equivalent ways:

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_d] : p \times d, \qquad \boldsymbol{\theta} = \mathrm{vec}(\boldsymbol{\Theta}).$$

The entries of $\boldsymbol{\Theta}$ are denoted by θ_{ij} ($i = 1, \ldots, p$, $j = 1, \ldots d$).

[2]If more information is known about $\mathbf{a}(\boldsymbol{\theta})$, e.g., an array manifold structure, then this can easily be taken into account at this point.

The derivatives of $\mathbf{A}(\boldsymbol{\theta})$ and $\underset{\sim}{\mathbf{A}}(\boldsymbol{\theta})$, which are evaluated at $\boldsymbol{\theta}_0$, are collected in vectors and matrices defined as

$$\begin{aligned} \mathbf{D} &= \left[\frac{\partial \mathbf{a}_1}{\partial\theta_{11}}, \frac{\partial \mathbf{a}_1}{\partial\theta_{21}}, \ldots, \frac{\partial \mathbf{a}_2}{\partial\theta_{12}}, \ldots\right](\boldsymbol{\theta}_0) \\ \underset{\sim}{\mathbf{D}} &= \left[\frac{\partial \underset{\sim}{\mathbf{a}}_1}{\partial\theta_{11}}, \frac{\partial \underset{\sim}{\mathbf{a}}_1}{\partial\theta_{21}}, \ldots, \frac{\partial \underset{\sim}{\mathbf{a}}_2}{\partial\theta_{12}}, \ldots\right](\boldsymbol{\theta}_0). \end{aligned} \tag{30}$$

Since $\underset{\sim}{\mathbf{A}}(\boldsymbol{\theta}) = [\bar{\mathbf{a}}(\boldsymbol{\theta}_1) \otimes \mathbf{a}(\boldsymbol{\theta}_1), \ldots, \bar{\mathbf{a}}(\boldsymbol{\theta}_d) \otimes \mathbf{a}(\boldsymbol{\theta}_d)]$, we obtain

$$\underset{\sim}{\mathbf{D}} = \bar{\mathbf{A}}_e \circ \mathbf{D} + \bar{\mathbf{D}} \circ \mathbf{A}_e, \qquad \mathbf{A}_e := \mathbf{A}(\boldsymbol{\theta}_0) \otimes \mathbf{1}_p^T. \tag{31}$$

Theorem 7: Let $\underset{\sim}{\mathbf{A}}_0 := \bar{\mathbf{A}}(\boldsymbol{\theta}_0) \circ \mathbf{A}(\boldsymbol{\theta}_0)$

$$\begin{aligned} \mathbf{M} &:= (\underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}})^H \otimes \mathbf{1}_p^T \\ \mathbf{C}_{\mathbf{u}} &:= \left[\underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^{-2} \bar{\underset{\sim}{\mathbf{U}}}_{\mathbf{A}}^H \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp\right] \mathbf{C}_{\mathbf{g}} \left[\bar{\underset{\sim}{\mathbf{U}}}_{\mathbf{A}} \underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^{-2} \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp\right] \end{aligned}$$

where $\underset{\sim}{\mathbf{U}}_{\mathbf{A}}$, and $\underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}$ are defined in (26). Then

$$\begin{aligned} \mathbf{Q} &= 4\left[\mathbf{M} \circ \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{D}}\right]^H \boldsymbol{\Gamma}\mathbf{C}_{\mathbf{u}}\boldsymbol{\Gamma} \left[\mathbf{M} \circ \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{D}}\right] \\ \mathbf{H} &= 2\left[\mathbf{M} \circ \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{D}}\right]^H \boldsymbol{\Gamma} \left[\mathbf{M} \circ \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{D}}\right] \end{aligned}$$

where $\underset{\sim}{\mathbf{D}}$ is defined in (31) so that for large N, the variance of $\hat{\boldsymbol{\theta}}$ that minimizes the subspace fitting problem (28) is, in first-order approximation

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

Proof: See Appendix C.

The covariance depends on the choice of weighting $\boldsymbol{\Gamma}$, and this weighting can be used to minimize the parameter covariance. Under some technical conditions outlined in [2] that have to do with identifiability in case $\boldsymbol{\Gamma}$ is singular, it is known that

$$\mathbf{R}_{\boldsymbol{\theta}}^{\boldsymbol{\Gamma}} \geq \mathbf{R}_{\boldsymbol{\theta}}^{\mathrm{opt}} := \frac{1}{N}\left(\left[\mathbf{M} \circ \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{D}}\right]^H \mathbf{C}_{\mathbf{u}}^\dagger \left[\mathbf{M} \circ \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{D}}\right]\right)^{-1} \tag{32}$$

and that the lower bound on the covariance $\mathbf{R}_{\boldsymbol{\theta}}^{\mathrm{opt}}$ is achieved by the weighting $\boldsymbol{\Gamma}^{\mathrm{opt}} = \mathbf{C}_{\mathbf{u}}^\dagger$.

A useful suboptimal weight $\boldsymbol{\Gamma}^s$ is obtained by ignoring the non-Gaussian part in $\mathbf{C}_{\mathbf{g}}$ [(20): $\mathbf{C}_{\mathbf{g}} \approx \bar{\mathbf{C}}_{\mathbf{x}} \otimes \mathbf{C}_{\mathbf{x}}$] and using the decomposition $\mathbf{C}_{\mathbf{x}} = \underset{\sim}{\mathbf{U}}\underset{\sim}{\boldsymbol{\Lambda}}\underset{\sim}{\mathbf{U}}^H = \underset{\sim}{\mathbf{U}}_{\mathbf{A}}(\mathbf{I} - \underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^2)\underset{\sim}{\mathbf{U}}_{\mathbf{A}}^H + \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp$

$$\begin{aligned} \mathbf{C}_{\mathbf{u}} &\approx \left[\underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^{-2} \bar{\underset{\sim}{\mathbf{U}}}_{\mathbf{A}}^H \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp\right] (\bar{\underset{\sim}{\mathbf{U}}}\underset{\sim}{\boldsymbol{\Lambda}}\bar{\underset{\sim}{\mathbf{U}}}^H \otimes \underset{\sim}{\mathbf{U}}\underset{\sim}{\boldsymbol{\Lambda}}\underset{\sim}{\mathbf{U}}^H) \\ &\quad \cdot \left[\bar{\underset{\sim}{\mathbf{U}}}_{\mathbf{A}} \underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^{-2} \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp\right] \\ &= \underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^{-2}(\mathbf{I} - \underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^2)\underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^{-2} \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \\ \Rightarrow \quad \boldsymbol{\Gamma}^s &:= \underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^4 (\mathbf{I} - \underset{\sim}{\boldsymbol{\Sigma}}_{\mathbf{A}}^2)^{-1} \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp. \end{aligned} \tag{33}$$

It is straightforward to obtain a consistent estimate of this weight from the eigenvalue decomposition of $\hat{\mathbf{C}}_{\mathbf{x}}$ [see (27)]. In particular, compute $\hat{\mathbf{C}}_{\mathbf{x}} = \hat{\underset{\sim}{\mathbf{U}}}\hat{\underset{\sim}{\mathbf{\Lambda}}}\hat{\underset{\sim}{\mathbf{U}}}^H = \hat{\underset{\sim}{\mathbf{U}}}_s\hat{\underset{\sim}{\mathbf{\Lambda}}}_s\hat{\underset{\sim}{\mathbf{U}}}_s^H + \hat{\underset{\sim}{\mathbf{U}}}_n\hat{\underset{\sim}{\mathbf{\Lambda}}}_n\hat{\underset{\sim}{\mathbf{U}}}_n^H$. Then, $\hat{\underset{\sim}{\mathbf{\Sigma}}}_{\mathbf{A}} = (\mathbf{I} - \hat{\underset{\sim}{\mathbf{\Lambda}}}_s)^{1/2}$ so that

$$\hat{\mathbf{\Gamma}}^s = (\mathbf{I} - \hat{\underset{\sim}{\mathbf{\Lambda}}}_s)^2\hat{\underset{\sim}{\mathbf{\Lambda}}}_s^{-1} \otimes \mathbf{I}.$$

Note that we could replace the projection $\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}$ by $\mathbf{I}$ since this does not alter the result ($\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}$ can be absorbed in vec$\{\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s\}$ in the cost function). With this weight, the cost function (28) becomes

$$\begin{aligned}\hat{J}^s(\boldsymbol{\theta}) =& \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}(\boldsymbol{\theta})}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s)^H\left[(\mathbf{I} - \hat{\underset{\sim}{\mathbf{\Lambda}}}_s)^2\hat{\underset{\sim}{\mathbf{\Lambda}}}_s^{-1} \otimes \mathbf{I}\right]\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}(\boldsymbol{\theta})}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s)\\ =& \|\underset{\sim}{\mathbf{P}}_{\mathbf{A}(\boldsymbol{\theta})}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s(\mathbf{I} - \hat{\underset{\sim}{\mathbf{\Lambda}}}_s)\hat{\underset{\sim}{\mathbf{\Lambda}}}_s^{-1/2}\|_F^2.\end{aligned}$$

Thus, this suboptimal weight involves only a scaling of each column of the estimated null space basis by a function of the corresponding eigenvalues. This is reminiscent of the result in [20] for the WSF technique.

With the approximation (33), we can obtain a more compact expression for $\mathbf{R}_{\boldsymbol{\theta}}^{\text{opt}}$. Inserting in (32), using property (5), and writing $\mathbf{A} = \mathbf{A}_0\mathbf{B}$, we find

$$\begin{aligned}(\mathbf{R}_{\boldsymbol{\theta}}^{\text{opt}})^{-1}&\\ \approx&\frac{1}{N}[\mathbf{M}\circ\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{D}}]^H(\underset{\sim}{\mathbf{\Sigma}}_{\mathbf{A}}^4(\mathbf{I}-\underset{\sim}{\mathbf{\Sigma}}_{\mathbf{A}}^2)^{-1}\otimes\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp})\\ &\cdot[\mathbf{M}\circ\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{D}}]\\ =&\frac{1}{N}[\mathbf{M}^H\underset{\sim}{\mathbf{\Sigma}}_{\mathbf{A}}^4(\mathbf{I}-\underset{\sim}{\mathbf{\Sigma}}_{\mathbf{A}}^2)^{-1}\mathbf{M}]\odot[\underset{\sim}{\mathbf{D}}^H\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{D}}]\\ =&\frac{1}{N}[\underset{\sim}{\mathbf{A}}_0^{\dagger}\underset{\sim}{\mathbf{U}}_{\mathbf{A}}\underset{\sim}{\mathbf{\Sigma}}_{\mathbf{A}}^4(\mathbf{I}-\underset{\sim}{\mathbf{\Sigma}}_{\mathbf{A}}^2)^{-1}\underset{\sim}{\mathbf{U}}_{\mathbf{A}}^H\underset{\sim}{\mathbf{A}}_0^{\dagger H}\otimes\mathbf{1}\mathbf{1}^T]\\ &\odot[\underset{\sim}{\mathbf{D}}^H\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{D}}]\\ =&\frac{1}{N}[|\mathbf{B}|^2(\underset{\sim}{\mathbf{A}}^{\dagger}\underset{\sim}{\mathbf{A}}^{\dagger H}-\mathbf{I})^{-1}|\mathbf{B}|^2\otimes\mathbf{1}\mathbf{1}^T]\odot[\underset{\sim}{\mathbf{D}}^H\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{D}}].\end{aligned}$$

C. *Covariance of* $\mathbf{t}_j$

We still need to map the previous result to an expression for the covariance of the beamforming vectors. Let $\mathbf{t}_j = \underset{\sim}{\mathbf{a}}_j(\boldsymbol{\theta}_0)$ be the optimal beamformer for the jth source, and let $\hat{\mathbf{t}}_j = \text{vec}(\mathbf{a}_j(\hat{\boldsymbol{\theta}}))$. Then, for small perturbations

$$\hat{\mathbf{t}}_j = \mathbf{t}_j + \sum_{\eta}\frac{\partial\mathbf{t}_j}{\partial\theta_\eta}(\hat{\boldsymbol{\theta}}_\eta - \boldsymbol{\theta}_\eta)$$

so that $\hat{\mathbf{t}}_j$ has covariance

$$\begin{aligned}\mathbf{R}_{t_j} =& E\{(\hat{\mathbf{t}}_j - \mathbf{t}_j)(\hat{\mathbf{t}}_j - \mathbf{t}_j)^H\}\\ =& \sum_{\eta,\xi}\frac{\partial\mathbf{t}_j}{\partial\boldsymbol{\theta}_\eta}E\left\{(\hat{\boldsymbol{\theta}}_\eta - \boldsymbol{\theta}_\eta)(\hat{\boldsymbol{\theta}}_\xi - \boldsymbol{\theta}_\xi)^H\right\}\frac{\partial\mathbf{t}_j^H}{\partial\boldsymbol{\theta}_\xi}.\end{aligned}$$

The derivatives are to be evaluated at the true value of the parameters, where $\mathbf{t}_j = \mathbf{a}_j(\boldsymbol{\theta}_0)$. Since the columns are parametrized independently, the derivatives are only nonzero for $\partial\mathbf{t}_j/\partial\theta_\eta = \partial\mathbf{a}_j/\partial\theta_{ij}(\boldsymbol{\theta}_0)$. It follows that

$$\mathbf{R}_{t_j} = \mathbf{D}_j[\mathbf{R}_{\boldsymbol{\theta}}]_{jj}\mathbf{D}_j^H, \quad [\mathbf{R}_{\boldsymbol{\theta}}]_{jj} := [\mathbf{e}_j \otimes \mathbf{I}_p]^H\mathbf{R}_{\boldsymbol{\theta}}[\mathbf{e}_j \otimes \mathbf{I}_p] \tag{34}$$

where $\mathbf{D}_j = [\partial\mathbf{a}_j/\partial\theta_{1j}(\boldsymbol{\theta}_0), \ldots, \partial\mathbf{a}_j/\partial\theta_{pj}(\boldsymbol{\theta}_0)]$ is a submatrix of $\mathbf{D}$, as defined in (30), and $[\mathbf{R}_{\boldsymbol{\theta}}]_{jj}$ is the jj-th subblock of size $p \times p$ of $\mathbf{R}_{\boldsymbol{\theta}}$.

D. *SINR Performance*

To allow interpretation of the performance of the beamformers, a plot of a scale-independent parameter such as the output SINR is more informative than the description of their covariance. This also allows a comparison of the performance to that of the optimal nonblind (Wiener) receiver. In this section, we derive a mapping of $\mathbf{R}_{t_j}$ to the inverse SINR or the interference plus noise-to-signal ratio (INSR)[3] 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), i.e., the matched beamformer or Wiener beamformer in the whitened domain. Consider a perturbation $\hat{\mathbf{t}} = \mathbf{t} + \mathbf{d}$, where $\mathbf{t} = \alpha\mathbf{a}$. Then

$$\begin{aligned}\text{INSR}(\hat{\mathbf{t}}) =& \frac{(\mathbf{t}+\mathbf{d})^H(\mathbf{I}-\mathbf{a}\mathbf{a}^H)(\mathbf{t}+\mathbf{d})}{(\mathbf{t}+\mathbf{d})^H\mathbf{a}\mathbf{a}^H(\mathbf{t}+\mathbf{d})}\\ =& \frac{1}{\mathbf{a}^H\mathbf{a}}\left(1 - \mathbf{a}^H\mathbf{a}\right.\\ &\left.+\frac{\mathbf{d}^H\mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{d}}{\mathbf{t}^H\mathbf{t}+\mathbf{t}^H\mathbf{d}+\mathbf{d}^H\mathbf{t}+\mathbf{d}^H\mathbf{P}_{\mathbf{a}}\mathbf{d}}\right)\\ \approx& \frac{1}{\mathbf{a}^H\mathbf{a}}\left(1 - \mathbf{a}^H\mathbf{a} + \frac{\mathbf{d}^H\mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{d}}{\mathbf{t}^H\mathbf{t}}\right)\end{aligned}$$

where the approximation is good if $\mathbf{d}^H\mathbf{P}_{\mathbf{a}}\mathbf{d} \ll \mathbf{t}^H\mathbf{t}$ (since $\mathbf{P}_{\mathbf{a}} = \mathbf{P}_{\mathbf{t}}$, this also implies $|\mathbf{d}^H\mathbf{t}| \ll \mathbf{t}^H\mathbf{t}$).

Let $\mathbf{\Delta} := E(\mathbf{d}\mathbf{d}^H)/\mathbf{t}^H\mathbf{t}$ be a normalized (scale-invariant) definition of the covariance of $\hat{\mathbf{t}}$. Then, in the above approximation

$$E\{\text{INSR}(\hat{\mathbf{t}})\} = \frac{1 - \mathbf{a}^H\mathbf{a}}{\mathbf{a}^H\mathbf{a}} + \frac{\text{tr}(\mathbf{P}_{\mathbf{a}}^{\perp}\mathbf{\Delta})}{\mathbf{a}^H\mathbf{a}}. \tag{35}$$

The first term represents the asymptotic performance of the Wiener beamformer ($\hat{\mathbf{t}} = \mathbf{a}$ with $\mathbf{\Delta} = 0$). 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 $\mathbf{R}_{t_j}$ from (34) in place of $\mathbf{\Delta}$ (since $\mathbf{t}_j$ is normalized) 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

[3]This parameter is chosen because it admits a good approximation in subsequent derivations.

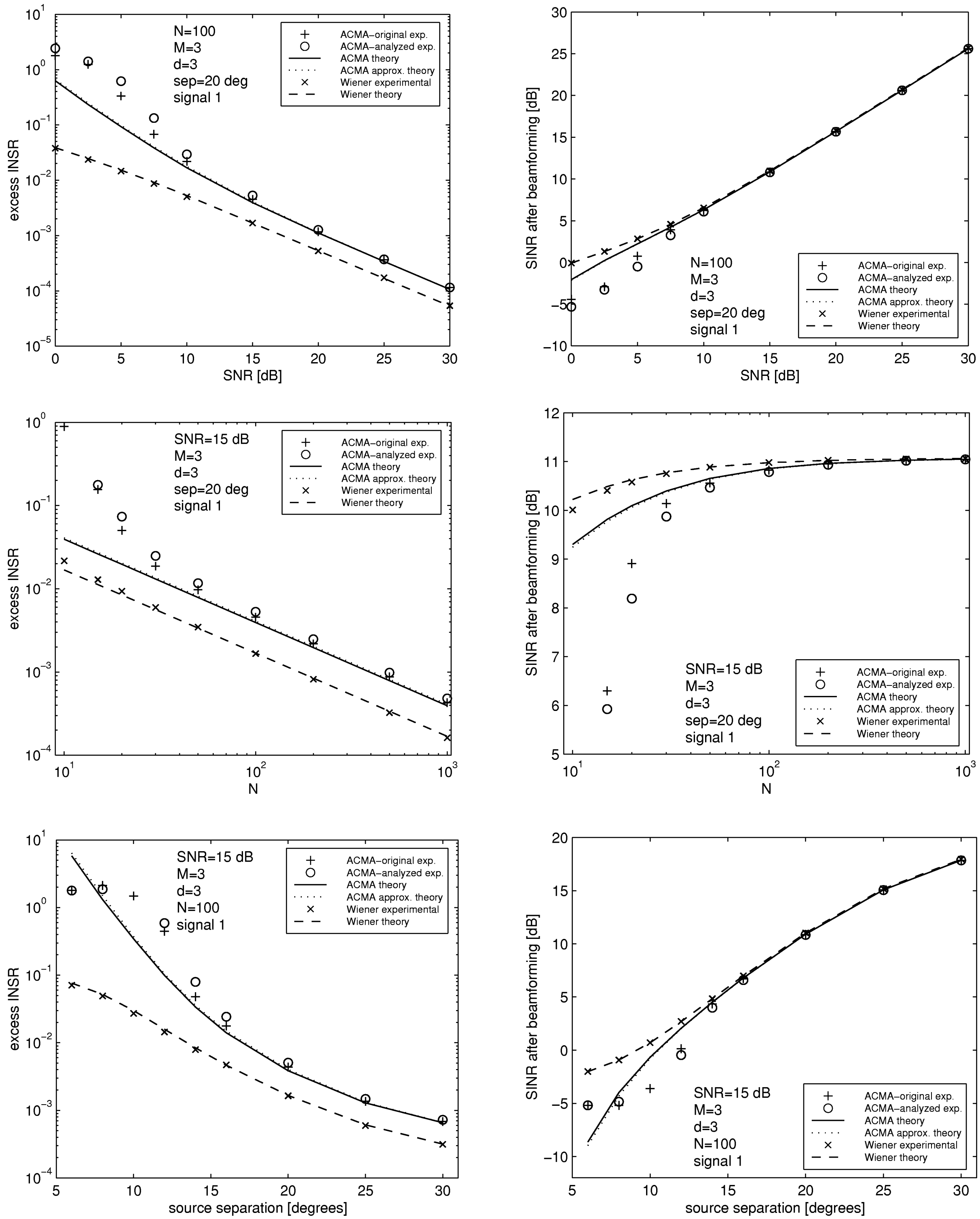

Fig. 3. Theoretical and experimental performance for source 1. Left column: finite sample excess INSR relative to the asymptotic INSR of the Wiener beamformer. Right column: SINR performance.

$\hat{\mathbf{T}}_W$, and $\mathbf{a}$ the corresponding column of $\mathbf{A}$. The normalized covariance of $\hat{\mathbf{t}}_W$ is derived in Appendix E as

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

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

$$E\{\text{INSR}(\hat{\mathbf{t}}_W)\} = \frac{1-\mathbf{a}^H\mathbf{a}}{\mathbf{a}^H\mathbf{a}} + \frac{d-1}{N}\cdot\frac{1-\mathbf{a}^H\mathbf{a}}{(\mathbf{a}^H\mathbf{a})^2}. \tag{36}$$

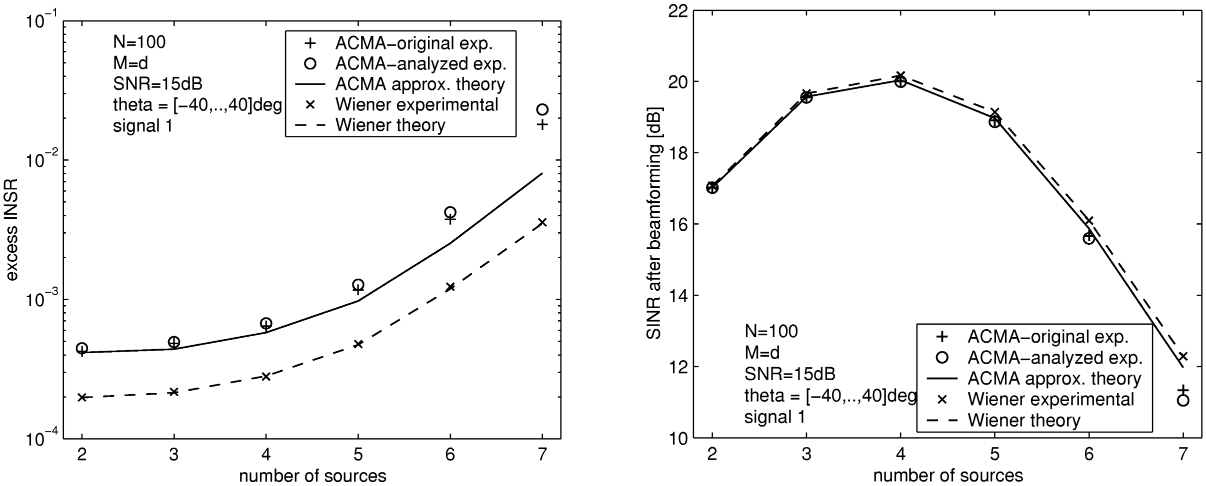

Fig. 4. Same as Fig. 3 for varying number of sources.

VII. Simulations

Some simulations are now presented to compare the derived theoretical performance expressions with the experimental performance. We use the data model $\mathbf{X} = \mathbf{A}(\boldsymbol{\alpha})\mathbf{B}\mathbf{S} + \mathbf{N}$, where $\mathbf{A}(\boldsymbol{\alpha}) = [\mathbf{a}(\alpha_1)\ldots\mathbf{a}(\alpha_d)]$, and $\mathbf{a}(\boldsymbol{\alpha})$ is the response of a uniform linear array with half-wavelength spacing of its M elements. $\mathbf{B}$ is a diagonal scaling matrix containing the source powers. We make sure that sources do not have identical powers because otherwise, the source eigenvalues coincide.[4] $\mathbf{S}$ is a random constant modulus matrix with independent entries, and $\mathbf{N}$ is white Gaussian noise with independent entries. The noise power σ^2 is set according to the desired SNR, where SNR is defined with respect to the first antenna and the first user.

Fig. 3 shows performance plots of the first source for a scenario with $d = 3$ sources, $M = 3$ antennas, source powers $\mathbf{B} = \text{diag}(1, 1.2, 0.9)$, and source angles $\boldsymbol{\alpha} = [0, \alpha, -\alpha]$ for varying SNR, number of samples N, and source separation α, respectively. The left column of Fig. 3 shows the excess INSR relative to the INSR of the asymptotic Wiener beamformer, evaluated for source 1.[5] The right column shows the resulting SINR. The experimental results show (with a "$+$") the outcome of the original ACMA algorithm of [18] and (with a "$\circ$") 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 (initialized by the same Jacobi iteration as used for the original algorithm). As is seen from the figures, for three sources and three antennas, the theoretical curves are a good prediction of the actual performance once $N > 30$ and SNR > 5 dB, and the separation is more than 10°. Below these values, there is some deviation, partly because of the approximations in the model and partly because the algorithm starts to break down (the gap between the "signal" and "noise" eigenvalues of $\hat{\mathbf{C}}_{\mathbf{x}}$ is too small). The small difference in performance between the original algorithm and the analyzed algorithm is caused by the different prewhitening. The changes due to extending the Jacobi iterations by a Gauss–Newton optimization step are negligible.

A second conclusion is that the "Gaussian" approximation (20) of $\mathbf{C}_{\mathbf{g}}$ is good enough to use since the dotted curves are almost indistinguishable from the full model. Moreover, although there is about a factor of 2 difference in excess INSR between ACMA and Wiener, the difference in actual SINR performance is very small in the region where the theoretical curves are valid. The results of the weighted version of the subspace-fitting step are not shown in the figures. Both in the theoretical model and in the experiments, it was found that there is no visible performance improvement in applying the weighting.

For further insight, Fig. 4 shows the SINR performance plots for a varying number of sources, which are evenly spread in the interval $[-40^\circ, 40^\circ]$, $N = 100$ samples, SNR$= 15$ dB, and as many antennas as sources. The performance of the algorithm varies because the number of sources that fall within the (varying) beamwidth determines the conditioning of $\mathbf{A}$ and $\underset{\sim}{\mathbf{A}}$, but it is seen that the performance of the Wiener beamformer varies in the same way. The accuracy of the theoretical performance prediction is quite good, provided that $N \gg d^2$ and that the resulting SINR performance is positive.

VIII. Concluding Remarks

We have derived theoretical models for the performance of the ACMA beamformers. By describing ACMA as an eigenvector decomposition followed by a subspace-fitting step, the analysis could follow the lines of the analysis of WSF, except that extensions were needed to take into account that the eigenvectors are obtained from a fourth-order covariance matrix and that the model is not Gaussian. The performance model turns out to be already quite accurate for a small number of samples (in the order of 30 for three sources) and for reasonably positive SNR ($\geq$ 8 dB) and conditioning of the problem. The analysis was limited to $M = d$, i.e., equal number of sources and antennas. The case $M > d$ requires a more detailed analysis of the prewhitening step that is deferred here.

[4]This is not relevant for the algorithm but has to be avoided for the analysis to be valid; cf. Lemma 4 .

[5]For the experimental curves, it is the mean of $\text{INSR}(\hat{\mathbf{w}}) - \text{INSR}(\mathbf{a})$ based on 5000 Monte Carlo runs, and for the theoretical curves, it is evaluated as $\text{tr}(\mathbf{P}_{\mathbf{a}}^{\perp}\boldsymbol{\Delta})(\mathbf{a}^H\mathbf{a})^{-1}$.

APPENDIX A
PROOF OF THEOREM 3

In the derivation, we will compute with $\underset{\sim}{\mathbf{s}}$ instead of $(\bar{\mathbf{A}} \otimes \mathbf{A})\underset{\sim}{\mathbf{s}}$ to simplify the notation. We will use the following and similar straightforwardly verified equations, where $\simeq$ denotes an approximation of order $\mathcal{O}(\sigma^4)$:

$$\begin{aligned}
E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{n}})(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{n}})^H &\simeq 0\\
E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{n}})\mathbf{E}(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{n}})^H &\simeq 0\\
E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{n}})(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})^H &\simeq 0\\
E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{s}})^H &= 0\\
E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{n}})^H &\simeq E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{s}})\mathbf{E}(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{n}})^H\\
E(\underset{\sim}{\bar{\mathbf{s}}}\underset{\sim}{\mathbf{s}}^H \otimes \underset{\sim}{\mathbf{n}}\underset{\sim}{\mathbf{n}}^H) &\simeq E(\underset{\sim}{\bar{\mathbf{s}}}\underset{\sim}{\mathbf{s}}^H) \otimes \mathbf{E}(\underset{\sim}{\mathbf{n}}\underset{\sim}{\mathbf{n}}^H).
\end{aligned}$$

Then, a somewhat tedious derivation shows that

$$\begin{aligned}
\mathbf{C_g} =& E\{[(\underset{\sim}{\bar{\mathbf{s}}} + \underset{\sim}{\bar{\mathbf{n}}}) \otimes (\underset{\sim}{\mathbf{s}} + \underset{\sim}{\mathbf{n}})][(\underset{\sim}{\bar{\mathbf{s}}} + \underset{\sim}{\bar{\mathbf{n}}}) \otimes (\underset{\sim}{\mathbf{s}} + \underset{\sim}{\mathbf{n}})]^H\}\\
&- E[(\underset{\sim}{\bar{\mathbf{s}}} + \underset{\sim}{\bar{\mathbf{n}}}) \otimes (\underset{\sim}{\mathbf{s}} + \underset{\sim}{\mathbf{n}})]E[(\underset{\sim}{\bar{\mathbf{s}}} + \underset{\sim}{\bar{\mathbf{n}}}) \otimes (\underset{\sim}{\mathbf{s}} + \underset{\sim}{\mathbf{n}})]^H\\
\simeq& E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{s}})^H - E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{s}})E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{s}})^H\\
&+ E(\underset{\sim}{\bar{\mathbf{s}}}\underset{\sim}{\mathbf{s}}^H) \otimes E(\underset{\sim}{\mathbf{n}}\underset{\sim}{\mathbf{n}}^H) + E(\underset{\sim}{\bar{\mathbf{n}}}\underset{\sim}{\mathbf{s}}^H) \otimes E(\underset{\sim}{\mathbf{s}}\underset{\sim}{\mathbf{s}}^H)\\
&+ E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{n}})^H + E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{n}})(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})^H.
\end{aligned}$$

Introducing cumulants of $\underset{\sim}{\mathbf{s}}$ according to (12) and adding a term $E(\underset{\sim}{\bar{\mathbf{n}}}\underset{\sim}{\bar{\mathbf{n}}}^H) \otimes E(\underset{\sim}{\bar{\mathbf{n}}}\underset{\sim}{\bar{\mathbf{n}}}^H)$ of order $\mathcal{O}(\sigma^4)$ gives

$$\begin{aligned}
\mathbf{C_g} \simeq& \mathbf{K}_{\underset{\sim}{\mathbf{s}}} + E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \mathbf{1})(\mathbf{1} \otimes \underset{\sim}{\mathbf{s}})^H \odot E(\mathbf{1} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \mathbf{1})^H\\
&+ \bar{\mathbf{R}}_\mathbf{g} \otimes \mathbf{R_g}\\
&+ E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{n}})^H + E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{n}})(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\bar{\mathbf{s}}})^H. \qquad (37)
\end{aligned}$$

Explicit expressions for these terms, using the CM distribution of $\mathbf{s}$, the independence of $\mathbf{s}$ and $\mathbf{n}$, and the fact that we only take second-order terms in $\mathbf{n}$ into account, are derived next.

In particular, for CM signals, we can write, denoting by $\oplus$ the binary "or" operator and δ_i^j the Kronecker delta

$$\begin{aligned}
E(s_i\bar{s}_j) =& \delta_i^j\\
E(s_i\bar{s}_j s_k\bar{s}_l) =& \delta_{ik}^{jl} := \delta_i^j\delta_k^l \oplus \delta_i^l\delta_k^j\\
E(s_i\bar{s}_j s_k\bar{s}_l s_m\bar{s}_n) =& \delta_{ikm}^{jln} := \delta_i^j\delta_{km}^{ln} \oplus \delta_i^l\delta_{km}^{jn} \oplus \delta_i^n\delta_{km}^{jl}.
\end{aligned}$$

Using $\underset{\sim}{\mathbf{s}} = \sum_{i\neq j} \mathbf{e}_{ij}s_i\bar{s}_j$, we obtain

$$\begin{aligned}
E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \mathbf{1})(\mathbf{1} \otimes \underset{\sim}{\mathbf{s}})^H &= \sum_{i\neq j}\sum_{k\neq l}(\mathbf{e}_{ij} \otimes \mathbf{1})(\mathbf{1} \otimes \mathbf{e}_{kl})^H\delta_{ik}^{jl}\\
&= \sum_{i\neq j}(\mathbf{e}_{ij} \otimes \mathbf{1})(\mathbf{1} \otimes \mathbf{e}_{ji})^H
\end{aligned}$$

and similarly

$$E(\mathbf{1} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \mathbf{1})^H = \sum_{k\neq l}(\mathbf{1} \otimes \mathbf{e}_{kl})(\mathbf{e}_{lk} \otimes \mathbf{1})^H$$

so that

$$\begin{aligned}
&E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \mathbf{1})(\mathbf{1} \otimes \underset{\sim}{\mathbf{s}})^H \odot E(\mathbf{1} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \mathbf{1})^H\\
&\qquad = \sum_{i\neq j}\sum_{k\neq l}(\mathbf{e}_{ij} \otimes \mathbf{e}_{kl})(\mathbf{e}_{lk} \otimes \mathbf{e}_{ji})^H.
\end{aligned}$$

To find an expression for $\mathbf{K}_{\underset{\sim}{\mathbf{s}}}$, insert $\underset{\sim}{\mathbf{s}} = \sum_{i\neq j} \mathbf{e}_{ij}s_i\bar{s}_j$ into the general expression (11) to obtain

$$\begin{aligned}
\mathbf{K}_{\underset{\sim}{\mathbf{s}}} = \sum_{i\neq j}\sum_{k\neq l}\sum_{m\neq n}\sum_{p\neq q}(\mathbf{e}_{kl} \otimes \mathbf{e}_{ij})(\mathbf{e}_{mn} \otimes \mathbf{e}_{pq})^H\cdot\\
\cdot\text{cum}(s_i\bar{s}_j, \bar{s}_k s_l, s_m\bar{s}_n, \bar{s}_p s_q).
\end{aligned}$$

Note that for CM signals, cum$(\cdot)$ is nonzero $(= -1)$ only for combinations of the form cum$(s, \bar{s}, s, \bar{s})$, cum$(s, s, \bar{s}, \bar{s})$, or cum$(s, \bar{s}, \bar{s}, s)$. For $\mathbf{K}_{\underset{\sim}{\mathbf{s}}}$, this means that there is a response only for

$$\begin{aligned}
i =& k = m = p \wedge j = l = n = q\\
\text{or } i =& l = n = p \wedge j = k = m = q\\
\text{or } i =& k = n = q \wedge j = l = m = p
\end{aligned}$$

so that we find the equation at the bottom of the page.

For the last two terms in (37), first apply (10), which gives

$$E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{n}})^H = E(\underset{\sim}{\bar{\mathbf{s}}}^H \otimes \underset{\sim}{\bar{\mathbf{n}}}\underset{\sim}{\mathbf{n}}^H \otimes \underset{\sim}{\mathbf{s}})$$

with

$$\begin{aligned}
\underset{\sim}{\bar{\mathbf{n}}}\underset{\sim}{\mathbf{n}}^H \simeq& \overline{(\bar{\mathbf{s}} \otimes \mathbf{n} + \bar{\mathbf{n}} \otimes \mathbf{s})}(\bar{\mathbf{s}} \otimes \mathbf{n} + \bar{\mathbf{n}} \otimes \mathbf{s})^H\\
\simeq& [\mathbf{s} \otimes \bar{\mathbf{n}}][\bar{\mathbf{n}} \otimes \mathbf{s}]^H + [\mathbf{n} \otimes \bar{\mathbf{s}}][\bar{\mathbf{s}} \otimes \mathbf{n}]^H\\
=& \mathbf{s} \otimes \bar{\mathbf{n}}\bar{\mathbf{n}}^H \otimes \mathbf{s}^H + \bar{\mathbf{s}}^H \otimes \mathbf{n}\mathbf{n}^H \otimes \bar{\mathbf{s}}
\end{aligned}$$

so that we have equation at the bottom of the page. $E(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{n}})(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})^H$ is equal to the conjugate transpose of this expres-

---

$$\mathbf{K}_{\underset{\sim}{\mathbf{s}}} = -\sum_{i\neq j}(\mathbf{e}_{ij} \otimes \mathbf{e}_{ij})(\mathbf{e}_{ij} \otimes \mathbf{e}_{ij})^H + (\mathbf{e}_{ji} \otimes \mathbf{e}_{ij})(\mathbf{e}_{ji} \otimes \mathbf{e}_{ij})^H + (\mathbf{e}_{ij} \otimes \mathbf{e}_{ij})(\mathbf{e}_{ji} \otimes \mathbf{e}_{ji})^H.$$

$$\begin{aligned}
E(\underset{\sim}{\bar{\mathbf{n}}} \otimes \underset{\sim}{\mathbf{s}})(\underset{\sim}{\bar{\mathbf{s}}} \otimes \underset{\sim}{\mathbf{n}})^H \simeq& E\{\underset{\sim}{\bar{\mathbf{s}}}^H \otimes (\mathbf{s} \otimes \bar{\mathbf{R}}_\mathbf{n} \otimes \mathbf{s}^H + \bar{\mathbf{s}}^H \otimes \mathbf{R_n} \otimes \bar{\mathbf{s}}) \otimes \underset{\sim}{\mathbf{s}}\}\\
=& \sum_{i\neq j}\sum_{kl}\sum_{m\neq n}\mathbf{e}_{ij}^H \otimes (\mathbf{e}_k \otimes \bar{\mathbf{R}}_\mathbf{n} \otimes \mathbf{e}_l^H + \mathbf{e}_k^H \otimes \mathbf{R_n} \otimes \mathbf{e}_l) \otimes \mathbf{e}_{mn}\delta_{ikm}^{jln}.
\end{aligned}$$

sion. Subsequently, note that δ_{ikm}^{jln} can be combined with the condition $i \neq j$ and $m \neq n$ so that it effectively reduces to $\delta_i^l \delta_k^n \delta_m^j \oplus \delta_i^n \delta_k^j \delta_m^l \oplus \delta_i^n \delta_k^l \delta_m^j$. This shows that the summation is really over only three indices, with the other three fixed in various ways, and leads to the claimed result. □

APPENDIX B
PARAMETRIZATION OF A UNIT-NORM VECTOR

A minimal parametrization of a unit-norm vector $\mathbf{a}$ with d complex entries and real non-negative first entry is provided by a sequence of Givens rotations:

$$\mathbf{a} = \mathbf{\Phi}\mathbf{R}_1(\alpha_1)\mathbf{R}_2(\alpha_2)\ldots\mathbf{R}_{d-1}(\alpha_{d-1})\mathbf{e}_1$$

where

$$\begin{aligned}\mathbf{\Phi} =& \text{diag}\left[1, \quad e^{j\phi_1}, \quad \ldots, \quad e^{j\phi_{d-1}}\right], \qquad 0 \leq \phi_i < 2\pi \\ \mathbf{R}_i(\alpha) =& \begin{bmatrix} c & & -s & \\ & \mathbf{I}_{i-1} & & \\ s & & c & \\ & & & \mathbf{I}_{d-1-i}\end{bmatrix} \\ c =& \cos(\alpha),\ s = \sin(\alpha), -\frac{\pi}{2} \leq \alpha < \frac{\pi}{2}.\end{aligned}$$

The parameter vector for $\mathbf{a} = \mathbf{a}(\boldsymbol{\theta})$ is

$$\boldsymbol{\theta} = \left[\alpha_1, \quad \ldots, \quad \alpha_{d-1}, \quad \phi_1, \ldots, \phi_{d-1}\right]^T.$$

We will also need the derivative of $\mathbf{a}(\boldsymbol{\theta})$ to each of the $p = 2(d-1)$ parameters:

$$\begin{aligned}\mathbf{d}_k(\boldsymbol{\theta}) :=& \frac{\partial \mathbf{a}}{\partial \theta_k}(\boldsymbol{\theta}) \\ =& \begin{cases} \frac{\partial \mathbf{a}}{\partial \alpha_i} & i = k, \qquad\qquad 1 \leq k \leq d-1, \\ \frac{\partial \mathbf{a}}{\partial \phi_i} & i = k-d+1, \quad d \leq k \leq 2(d-1)\end{cases}\end{aligned}$$

where

$$\begin{aligned}\frac{\partial \mathbf{a}}{\partial \alpha_i} =& \mathbf{\Phi}\mathbf{R}_1(\alpha_1)\ldots\mathbf{R}_{i-1}(\alpha_{i-1})\mathbf{R}_i'(\alpha_i)\mathbf{R}_{i+1}(\alpha_{i+1})\cdot \\ & \ldots\mathbf{R}_{d-1}(\alpha_{d-1})\mathbf{e}_1 \\ \frac{\partial \mathbf{a}}{\partial \phi_i} =& j\mathbf{e}_{i+1}\mathbf{e}_{i+1}^H\mathbf{a}, \\ \mathbf{R}_i'(\alpha) =& \begin{bmatrix} -s & & -c & \\ & \mathbf{0}_{i-1} & & \\ c & & -s & \\ & & & \mathbf{0}_{d-1-i}\end{bmatrix}.\end{aligned}$$

Further note that

$$\frac{\partial \underset{\sim}{\mathbf{a}}}{\partial \theta_i} = \frac{\partial}{\partial \theta_i}(\bar{\mathbf{a}} \otimes \mathbf{a}) = \bar{\mathbf{a}} \otimes \mathbf{d}_i + \bar{\mathbf{d}}_i \otimes \mathbf{a}.$$

In some cases, other parametrizations of $\mathbf{a}$ are possible, e.g., in terms of directions of arrival. This leads to similar definitions of $\mathbf{d}_i$ but with $p = 1$ or $p = 2$ parameters.

APPENDIX C
PROOF OF THEOREM 7

In view of Lemma 6, $\mathbf{Q}$ and $\mathbf{H}$ as defined there remains to be computed. Let η be the index of one of the parameters θ_{ij} in $\boldsymbol{\theta}$, and let $\underset{\sim}{\mathbf{A}}_\eta := (\partial \underset{\sim}{\mathbf{A}}(\boldsymbol{\theta})/\partial \theta_\eta)(\boldsymbol{\theta}_0)$. The derivative of $\underset{\sim}{\mathbf{P}}_{\mathbf{A}(\boldsymbol{\theta})}^\perp$ evaluated at $\boldsymbol{\theta}_0$ is [20]

$$\underset{\sim}{\mathbf{P}}_\eta := \frac{\partial \underset{\sim}{\mathbf{P}}_{\mathbf{A}(\boldsymbol{\theta})}^\perp}{\partial \theta_\eta}(\boldsymbol{\theta}_0) = -\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{A}}_\eta \underset{\sim}{\mathbf{A}}_0^\dagger - (\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{A}}_\eta \underset{\sim}{\mathbf{A}}_0^\dagger)^H.$$

Thus, the derivative of the cost function to θ_η evaluated at $\boldsymbol{\theta}_0$ is

$$\begin{aligned}J_\eta =& \text{vec}(\underset{\sim}{\mathbf{P}}_\eta \hat{\underset{\sim}{\mathbf{U}}}_s)^H \mathbf{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s) \\ & + \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s)^H \mathbf{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_\eta \hat{\underset{\sim}{\mathbf{U}}}_s) \\ =& -\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{A}}_\eta \underset{\sim}{\mathbf{A}}_0^\dagger \hat{\underset{\sim}{\mathbf{U}}}_s + \underset{\sim}{\mathbf{A}}_0^{\dagger H} \underset{\sim}{\mathbf{A}}_\eta^H \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s)^H \\ & \times \mathbf{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s) - [*]^H\end{aligned}$$

where $[*]^H$ stands for the Hermitian transpose of the previous terms in the expression. For large N, $\hat{\underset{\sim}{\mathbf{U}}}_s \to \underset{\sim}{\mathbf{U}}_{\mathbf{A}}$, and $\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s \to \mathbf{0}$, giving in first-order approximation

$$J_\eta = -2\text{Re}\left\{\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{A}}_\eta \underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}})^H \mathbf{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s)\right\}. \quad (38)$$

Without loss of generality, we can assume that $\mathbf{\Gamma}$ is restricted such that the expression in braces is real by itself (see Appendix D) so that we can drop the Re{} operator. From theorem 5, we know that

$$NE\{\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s(\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \hat{\underset{\sim}{\mathbf{U}}}_s))^H\} = \mathbf{C}_{\mathbf{u}} + o(1)$$

and we find

$$\begin{aligned}Q_{\eta\xi} :=& \lim_{N\to\infty} NE\{J_\eta J_\xi^H\} \\ =& 4\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{A}}_\eta \underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}})^H \mathbf{\Gamma} \mathbf{C}_{\mathbf{u}} \mathbf{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp \underset{\sim}{\mathbf{A}}_\xi \underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}}) \\ =& 4\text{vec}(\underset{\sim}{\mathbf{A}}_\eta)^H [(\underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}})^T \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp]^H \mathbf{\Gamma} \mathbf{C}_{\mathbf{u}} \mathbf{\Gamma} \cdot \\ & \cdot [(\underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}})^T \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp] \text{vec}(\underset{\sim}{\mathbf{A}}_\xi).\end{aligned}$$

Letting $\mathbf{Q} = [Q_{\eta\xi}]$, then

$$\mathbf{Q} = 4\underset{\sim}{\mathbf{D}}_e^H [(\underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}})^T \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp]^H \mathbf{\Gamma} \mathbf{C}_{\mathbf{u}} \mathbf{\Gamma} [(\underset{\sim}{\mathbf{A}}_0^\dagger \underset{\sim}{\mathbf{U}}_{\mathbf{A}})^T \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^\perp] \underset{\sim}{\mathbf{D}}_e \quad (39)$$

where

$$\begin{aligned}\underset{\sim}{\mathbf{D}}_e :=& [\text{vec}(\underset{\sim}{\mathbf{A}}_\eta)]_{\eta=1,\ldots,pd} \\ =& \left[\frac{\partial \text{vec}(\underset{\sim}{\mathbf{A}}(\boldsymbol{\theta}))}{\partial \theta_{11}}, \frac{\partial \text{vec}(\underset{\sim}{\mathbf{A}}(\boldsymbol{\theta}))}{\partial \theta_{21}}, \ldots\right](\theta_0).\end{aligned}$$

Compared with the definition of $\underset{\sim}{\mathbf{D}}$, we see that $\underset{\sim}{\mathbf{D}}_e$ merely augments each column of $\underset{\sim}{\mathbf{D}}$ with many zero entries since parameter θ_{ij} affects only column j of $\underset{\sim}{\mathbf{A}}(\boldsymbol{\theta})$. Thus

$$\begin{aligned}\underset{\sim}{\mathbf{D}}_e =& [\mathbf{e}_1 \cdots \mathbf{e}_1 \,|\, \mathbf{e}_2 \cdots \mathbf{e}_2 \,|\, \cdots \,|\, \mathbf{e}_d \cdots \mathbf{e}_d] \circ \underset{\sim}{\mathbf{D}} \\ =& (\mathbf{I}_d \otimes \mathbf{1}_p^T) \circ \underset{\sim}{\mathbf{D}}\end{aligned}$$

and after substitution of this in (39), we obtain the claimed expression for $\mathbf{Q}$.

The expression for $\mathbf{H}$ follows likewise. Let $J_{\eta\xi}$ be the second derivative of $J(\boldsymbol{\theta})$ evaluated at $\boldsymbol{\theta}_0$. Then

$$\begin{aligned} J_{\eta\xi} =& -\text{vec}(\underset{\sim}{\mathbf{P}}_{\eta\xi}\hat{\underset{\sim}{\mathbf{U}}}_s)^H \boldsymbol{\Gamma} \underline{\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s)} \\ &+ \text{vec}(\underset{\sim}{\mathbf{P}}_{\eta}\hat{\underset{\sim}{\mathbf{U}}}_s)^H \boldsymbol{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\xi}\hat{\underset{\sim}{\mathbf{U}}}_s) \\ &+ \text{vec}(\underset{\sim}{\mathbf{P}}_{\xi}\hat{\underset{\sim}{\mathbf{U}}}_s)^H \boldsymbol{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\eta}\hat{\underset{\sim}{\mathbf{U}}}_s) \\ &- \underline{\text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s)^H} \boldsymbol{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\eta\xi}\hat{\underset{\sim}{\mathbf{U}}}_s). \end{aligned}$$

For $N \to \infty$, $\underset{\sim}{\mathbf{U}}_s \to \underset{\sim}{\mathbf{U}}_{\mathbf{A}}$, the underlined terms drop out, and we obtain

$$\begin{aligned} \lim_{N\to\infty} J_{\eta\xi} =& \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{A}}_{\eta}\underset{\sim}{\mathbf{A}}_0^{\dagger}\underset{\sim}{\mathbf{U}}_{\mathbf{A}})^H \boldsymbol{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{A}}_{\xi}\underset{\sim}{\mathbf{A}}_0^{\dagger}\underset{\sim}{\mathbf{U}}_{\mathbf{A}}) + [*]^H \\ =& 2\text{Re}\left\{ \text{vec}(\underset{\sim}{\mathbf{A}}_{\eta})^H [(\underset{\sim}{\mathbf{A}}_0^{\dagger}\underset{\sim}{\mathbf{U}}_{\mathbf{A}})^T \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}]^H \boldsymbol{\Gamma} \right. \\ & \left. \times [(\underset{\sim}{\mathbf{A}}_0^{\dagger}\underset{\sim}{\mathbf{U}}_{\mathbf{A}})^T \otimes \underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}] \cdot \text{vec}(\underset{\sim}{\mathbf{A}}_{\xi}) \right\}. \end{aligned}$$

Again, the term in braces is real by itself, giving the claimed expression for $\mathbf{H} = \lim_{N\to\infty}[J_{\eta\xi}]$. (Note also that $\underset{\sim}{\mathbf{A}}_0^{\dagger}\underset{\sim}{\mathbf{U}}_{\mathbf{A}}$ is real.) □

APPENDIX D
REAL PROCESSING

Since $\bar{\mathbf{x}} \otimes \mathbf{x} = \text{vec}(\mathbf{x}\mathbf{x}^H)$, the entries of this vector have a certain Hermitian symmetry property. It follows that there exists a unitary matrix $\mathbf{Z}$ such that $\mathbf{Z}(\bar{\mathbf{x}} \otimes \mathbf{x})$ is real for any $\mathbf{x}$. Consequently, all derived matrices can be mapped to real: $\mathbf{Z}\mathbf{C}_{\mathbf{x}}\mathbf{Z}^H$, $\mathbf{Z}\hat{\underset{\sim}{\mathbf{U}}}_s$, $\mathbf{Z}\underset{\sim}{\mathbf{A}}$, and $\mathbf{Z}\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\mathbf{Z}^H$ are all real.

In the subspace fitting cost function (28), we can without loss of generality restrict $\boldsymbol{\Gamma}$ to be of the form

$$\boldsymbol{\Gamma} = [\mathbf{I} \otimes \mathbf{Z}^H]\, \boldsymbol{\Gamma}_r\, [\mathbf{I} \otimes \mathbf{Z}] \tag{40}$$

where $\boldsymbol{\Gamma}_r = \boldsymbol{\Gamma}_r^T$ is real and symmetric. Indeed, in general, we can write any Hermitian $\boldsymbol{\Gamma}$ as $\boldsymbol{\Gamma} = [\mathbf{I} \otimes \mathbf{Z}^H]\,(\boldsymbol{\Gamma}_r + j\boldsymbol{\Gamma}_c)\,[\mathbf{I} \otimes \mathbf{Z}]$, where the complex part represented by $\boldsymbol{\Gamma}_c$ is skew-symmetric ($\boldsymbol{\Gamma}_c^T = -\boldsymbol{\Gamma}_c$). In the cost function

$$\begin{aligned} J =& \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s)^H \boldsymbol{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s) \\ =& \text{vec}(\mathbf{Z}\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\mathbf{Z}^H \cdot \mathbf{Z}\hat{\underset{\sim}{\mathbf{U}}}_s)^H (\boldsymbol{\Gamma}_r + j\boldsymbol{\Gamma}_c) \text{vec}(\mathbf{Z}\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\mathbf{Z}^H \cdot \mathbf{Z}\hat{\underset{\sim}{\mathbf{U}}}_s) \\ =& v_r^T \boldsymbol{\Gamma}_r + j v_r^T \boldsymbol{\Gamma}_c v_r \end{aligned}$$

where $\mathbf{v}_r := \text{vec}(\mathbf{Z}\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\mathbf{Z}^H \cdot \mathbf{Z}\hat{\underset{\sim}{\mathbf{U}}}_s)$ is real. It follows that $\mathbf{v}_r^T \boldsymbol{\Gamma}_c \mathbf{v}_r = 0$ for any (skew-symmetric) $\boldsymbol{\Gamma}_c$ so that this component has no influence on the cost. Hence, we can assume that $\boldsymbol{\Gamma}$ is of the form (40). After this, it is straightforward to show that the expression in braces in (38), viz.

$$\begin{aligned} & \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\underset{\sim}{\mathbf{A}}_{\eta}\underset{\sim}{\mathbf{A}}_0^{\dagger}\hat{\underset{\sim}{\mathbf{U}}}_s)^H \boldsymbol{\Gamma} \text{vec}(\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\hat{\underset{\sim}{\mathbf{U}}}_s) \\ &= \text{vec}(\mathbf{Z}\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\mathbf{Z}^H \cdot \mathbf{Z}\underset{\sim}{\mathbf{A}}_{\eta} \cdot \underset{\sim}{\mathbf{A}}_0^{\dagger}\mathbf{Z}^H \cdot \mathbf{Z}\hat{\underset{\sim}{\mathbf{U}}}_s)^H [\mathbf{I} \otimes \mathbf{Z}] \boldsymbol{\Gamma} [\mathbf{I} \otimes \mathbf{Z}^H] \cdot \\ &\quad \cdot \text{vec}(\mathbf{Z}\underset{\sim}{\mathbf{P}}_{\mathbf{A}}^{\perp}\mathbf{Z}^H \cdot \mathbf{Z}\hat{\underset{\sim}{\mathbf{U}}}_s) \end{aligned}$$

is real by itself.

APPENDIX E
FINITE SAMPLE WIENER BEAMFORMER COVARIANCE

The large but finite sample performance of the Wiener beamformer computed from known $\mathbf{X}$ and $\mathbf{S}$ is derived from a first-order perturbation analysis as follows. Define (in the whitened domain)

$$\begin{aligned} \frac{1}{N}\mathbf{X}\mathbf{X}^H &=: \mathbf{I} + \mathbf{E}_x \\ \frac{1}{N}\mathbf{S}\mathbf{S}^H &=: \mathbf{I} + \mathbf{E}_s \\ \frac{1}{N}\mathbf{N}\mathbf{N}^H &=: \mathbf{R}_n + \mathbf{E}_n. \end{aligned} \tag{41}$$

Then

$$\begin{aligned} \text{cov}(\mathbf{E}_x) &= \frac{1}{N}\mathbf{I} \\ \text{cov}(\mathbf{E}_s) &= \frac{1}{N}\mathbf{I} \\ \text{cov}(\mathbf{E}_n) &= \frac{1}{N}\bar{\mathbf{R}}_n \otimes \mathbf{R}_n \\ \text{cov}\left(\frac{1}{N}\mathbf{N}\mathbf{S}^H\right) &= \frac{1}{N}\mathbf{I} \otimes \mathbf{R}_n \\ \text{cov}\left(\frac{1}{N}\mathbf{S}\mathbf{N}^H\right) &= \frac{1}{N}\bar{\mathbf{R}}_n \otimes \mathbf{I}. \end{aligned} \tag{42}$$

Note that $(1/N\mathbf{X}\mathbf{X}^H)^{-1} = \mathbf{I} - \mathbf{E}_x + \mathcal{O}(N^{-2})$ and that $\mathbf{E}_x$ is not independent from $\mathbf{E}_s$ etc., but is related via the model $\mathbf{X} = \mathbf{A}\mathbf{S} + \mathbf{N}$ as

$$\begin{aligned} \mathbf{E}_x =& \frac{1}{N}\mathbf{X}\mathbf{X}^H - \mathbf{I} \\ =& \mathbf{A}\frac{1}{N}\mathbf{S}\mathbf{S}^H\mathbf{A}^H + \frac{1}{N}\mathbf{N}\mathbf{N}^H \\ &+ \mathbf{A}\frac{1}{N}\mathbf{S}\mathbf{N}^H + \frac{1}{N}\mathbf{N}\mathbf{S}^H\mathbf{A}^H - \mathbf{I} \\ =& \mathbf{A}\mathbf{E}_s\mathbf{A}^H + \mathbf{E}_n + \mathbf{A}\frac{1}{N}\mathbf{S}\mathbf{N}^H + \frac{1}{N}\mathbf{N}\mathbf{S}^H\mathbf{A}^H. \end{aligned} \tag{43}$$

The Wiener beamformer (in the whitened domain) can thus be approximated to $\mathcal{O}(N^{-2})$ as

$$\begin{aligned} \hat{\mathbf{T}} &= \left(\frac{1}{N}\mathbf{X}\mathbf{X}^H\right)^{-1} \frac{1}{N}\mathbf{X}\mathbf{S}^H \\ &\simeq (\mathbf{I} - \mathbf{E}_x)\left(\mathbf{A}(\mathbf{I} + \mathbf{E}_s) + \frac{1}{N}\mathbf{N}\mathbf{S}^H\right) \\ &\simeq \mathbf{A} + \mathbf{A}\mathbf{E}_s + \frac{1}{N}\mathbf{N}\mathbf{S}^H - \mathbf{E}_x\mathbf{A}. \end{aligned}$$

Inserting (43), we obtain

$$\begin{aligned} \hat{\mathbf{T}} - \mathbf{A} =& \mathbf{A}\mathbf{E}_s - \mathbf{E}_x\mathbf{A} + \frac{1}{N}\mathbf{N}\mathbf{S}^H \\ =& \mathbf{A}\mathbf{E}_s(\mathbf{I} - \mathbf{A}^H\mathbf{A}) - \mathbf{E}_n\mathbf{A} \\ &- \mathbf{A}\left(\frac{1}{N}\mathbf{S}\mathbf{N}^H\right)\mathbf{A} + \left(\frac{1}{N}\mathbf{N}\mathbf{S}^H\right)(\mathbf{I} - \mathbf{A}^H\mathbf{A}) \end{aligned}$$

and using (41)

$$\begin{aligned} & N\text{cov}(\hat{\mathbf{T}} - \mathbf{A}) \\ &= [(\mathbf{I} - \mathbf{A}^H\mathbf{A})^T \otimes \mathbf{A}][(\mathbf{I} - \mathbf{A}^H\mathbf{A})^T \otimes \mathbf{A}]^H \\ &\quad + [\mathbf{A}^T \otimes \mathbf{I}][\bar{\mathbf{R}}_n \otimes \mathbf{R}_n][\mathbf{A}^T \otimes \mathbf{I}]^H \\ &\quad + [\mathbf{A}^T \otimes \mathbf{A}][\bar{\mathbf{R}}_n \otimes \mathbf{I}][\mathbf{A}^T \otimes \mathbf{A}]^H \\ &\quad + [(\mathbf{I} - \mathbf{A}^H\mathbf{A})^T \otimes \mathbf{I}][\mathbf{I} \otimes \mathbf{R}_n][(\mathbf{I} - \mathbf{A}^H\mathbf{A})^T \otimes \mathbf{I}]^H. \end{aligned}$$

Specializing to the covariance of the jth column $\hat{\mathbf{t}}_j$ of $\hat{\mathbf{T}}$, we finally find, after some straightforward manipulations (using $\mathbf{A}\mathbf{A}^H + \mathbf{R_n} = \mathbf{I}$)

$$\begin{aligned}\text{cov}(\hat{\mathbf{t}}_j - \mathbf{a}_j) =& [\mathbf{e}_j^T \otimes \mathbf{I}]\text{cov}(\hat{\mathbf{T}} - \mathbf{A})[\mathbf{e}_j \otimes \mathbf{I}]\\ =& \frac{1}{N}(1 - \mathbf{a}_j^H \mathbf{a}_j)\mathbf{I}.\end{aligned}$$

REFERENCES

[1] T. W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Stat.*, vol. 34, pp. 122–148, 1963.
[2] J.-F. Cardoso and E. Moulines, "Invariance of subspace based estimators," *IEEE Trans. Signal Processing*, vol. 48, pp. 2495–2505, Sept. 2000.
[3] J. F. Cardoso and E. Moulines, "A robustness property of DOA estimators based on covariance," *IEEE Trans. Signal Processing*, vol. 42, pp. 3285–3287, Nov. 1994.
[4] ——, "Performance analysis of direction-finding algorithms based on fourth-order cumulants," *IEEE Trans. Signal Processing*, vol. 43, pp. 214–224, Jan. 1995.
[5] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng. F, Radar Signal Process.*, vol. 140, pp. 362–370, Dec. 1993.
[6] B. Friedlander and A. J. Weiss, "On the second-order statistics of the eigenvectors of sample covariance matrices," *IEEE Trans. Signal Processing*, vol. 46, pp. 3136–3139, Nov. 1998.
[7] M. Gu and L. Tong, "Geometrical characterizations of constant modulus receivers," *IEEE Trans. Signal Processing*, vol. 47, pp. 2745–2756, Oct. 1999.
[8] M. Kaveh and A. J. Barabell, "The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 331–341, Apr. 1986.
[9] 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.
[10] ——, "On the finite sample behavior of the constant modulus cost function," *Proc. IEEE ICASSP*, vol. 5, pp. 2537–2540, June 2000.
[11] O. J. Micka and A. J. Weiss, "Estimating frequencies of exponentials in noise using joint diagonalization," *IEEE Trans. Signal Processing*, vol. 47, pp. 341–348, Feb. 1999.
[12] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, pp. 590–600, Mar. 1992.
[13] B. Porat and B. Friedlander, "Blind equalization of digital communication channels using high-order moments," *IEEE Trans. Signal Processing*, vol. 39, pp. 522–526, Feb. 1991.
[14] ——, "Direction finding algorithms based on high-order statistics," *IEEE Trans. Signal Processing*, vol. 39, pp. 2016–2024, Sept. 1991.
[15] S. Talwar, A. Paulraj, and M. Viberg, "Blind separation of synchronous co-channel digital signals using an antenna array. Part II. Performance analysis," *IEEE Trans. Signal Processing*, vol. 45, pp. 706–718, Mar. 1997.
[16] A. J. van der Veen, "Asymptotic properties of the algebraic constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 49, pp. 1796–1807, Aug. 2001.
[17] ——, "Joint diagonalization via subspace fitting techniques," in *Proc. IEEE ICASSP*, May 2001.
[18] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.
[19] A. J. van der Veen and A. Trindade, "Combining blind equalization with constant modulus properties," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Oct. 2000.
[20] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, pp. 1110–1121, May 1991.
[21] A. J. Weiss and B. Friedlander, "Array processing using joint diagonalization," *Signal Process.*, vol. 50, pp. 205–222, May 1996.
[22] J. H. Wilkinson, *The Algebraic Eigenvalue Problem*. Oxford, U.K.: Clarendon, 1965.
[23] 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.
[24] 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.
[25] ——, "An analysis of constant modulus receivers," *IEEE Trans. Signal Processing*, vol. 47, pp. 2990–2999, Nov. 1999.

Alle-Jan van der Veen (S'87–M'94–SM'02) 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 Laboratory. At present, he is a Full 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 was the recipient of the 1994 and 1997 IEEE SPS Young Author paper awards and was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1998 to 2001. He is currently Chairman of the IEEE SPS SPCOM Technical Committee and Editor-in-Chief of the IEEE SIGNAL PROCESSING LETTERS.