MIMO instantaneous blind identification based on second-order temporal structure

Citation for published version (APA):

Laar, van de, J., Moonen, M., & Sommen, P. C. W. (2008). MIMO instantaneous blind identification based on second-order temporal structure. *IEEE Transactions on Signal Processing*, 56(9), 4354-4364.
<https://doi.org/10.1109/TSP.2008.926105>

DOI:

[10.1109/TSP.2008.926105](https://doi.org/10.1109/TSP.2008.926105)

Document status and date:

Published: 01/01/2008

Document Version:

Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.tue.nl/taverne

Take down policy

If you believe that this document breaches copyright please contact us at:

openaccess@tue.nl

providing details and we will investigate your claim.

MIMO Instantaneous Blind Identification Based on Second-Order Temporal Structure

Jakob van de Laar, Marc Moonen, *Fellow, IEEE*, and Piet C. W. Sommen, *Senior Member, IEEE*

Abstract—Blind identification is a crucial subtask in signal processing problems such as blind signal separation (BSS) and direction-of-arrival (DOA) estimation. This paper presents a procedure for *multiple-input multiple-output instantaneous blind identification* based on second-order temporal properties of the signals, such as coloredness and nonstationarity. The procedure consists of two stages. First, based on assumptions on the second-order temporal structure (SOTS) of the source and noise signals, and using subspace techniques, the problem is reformulated in a particular way such that each column of the unknown mixing matrix satisfies a system of multivariate homogeneous polynomial equations. Then, this nonlinear system of equations is solved by means of a so-called homotopy method, which provides a general tool for solving (possibly nonexact) systems of nonlinear equations by smoothly deforming the known solutions of a simple start system into the desired solutions of the target system. Our blind identification procedure allows to estimate the mixing matrix for scenarios with more sources than sensors without resorting to sparsity assumptions, something that is often believed to be impossible when using only second-order statistics. In addition, since our algorithm does not require *any* assumption on the mixing matrix, also mixing matrices that are rank-deficient or even have identical columns can be identified. Finally, we give examples and performance results for speech source signals.

Index Terms—Blind identification/separation, homogeneous system, homotopy, second-order statistics, temporal structure.

I. INTRODUCTION

IN this paper, we consider the so-called *multiple-input multiple-output (MIMO) instantaneous blind identification (MIBI)* problem. In this problem, a number of mutually statistically independent source signals are mixed by a MIMO instantaneous mixing system and only the mixed signals are available, i.e., both the mixing system and the original source signals are unknown [9], [12]. The goal of MIBI is to recover the instantaneous MIMO mixing system, or its parameters such as in the case of DOA estimation, from the observed mixtures

Fig. 1. MIMO instantaneous blind identification setup.

of the source signals only. Fig. 1 shows the MIBI problem setup for S source and D sensor signals. The source, sensor and additive noise signals are denoted by $s_1[n], \dots, s_S[n]$, $x_1[n], \dots, x_D[n]$ and $\nu_1[n], \dots, \nu_D[n]$ respectively, and are assumed to be real-valued. The instantaneous mixing system is modeled by a real-valued matrix $\mathbf{A}$ of size $D \times S$. A problem closely related to MIBI is *instantaneous blind signal separation (IBSS)* [9], [12], where the goal is to separate mutually statistically independent source signals from their observed instantaneous mixtures only. Contrary to MIBI, the main interest in IBSS is in the source signals instead of the mixing system. In fact, once MIBI has been performed, the source signals can be recovered (approximately) by applying the (pseudo-)inverse of the estimated mixing system to the observed mixtures. In this paper, the main focus is on MIBI, while IBSS is considered as an application.

Many researchers have investigated the use of second-order statistics (SOS) for IBSS [2], [6], [8], [13], [16], [17], [22], [23], [28]. The majority of the available algorithms is based on the (generalized) eigenvalue decomposition or joint approximate diagonalization of two or more sensor correlation matrices of the form $\mathbf{R}^x[n, k] \triangleq E\{\mathbf{x}[n]\mathbf{x}[n-k]^T\}$ with $\mathbf{x}[n]$ the observation vector. For example, see AMUSE (Algorithm for Multiple Unknown Signals Extraction) [22], [23], SOBI (Second-Order Blind Identification) [2], SOBI-RO (Robust SOBI with Robust Orthogonalization) [3], [8], [9], SONS (Second-Order Nonstationary Source Separation) [6], [7], [9]; see also [13] and [16]. All those algorithms employ sensor correlation matrices containing the sensor correlation values for different lags and/or times arranged in the same conventional manner shown above. In our work we arrange the available sensor correlation values in a particular fashion that allows a different and natural formulation of the problem, as well as the estimation of more columns than sensors. Many possible applications exist for MIBI and IBSS, see e.g., [9], [12] and the references therein. Examples of (parameterized) MIBI can be found in source localization problems, which are crucial to many sensor array systems such as

Manuscript received August 17, 2007; revised May 15, 2008. First published May 28, 2008; last published August 13, 2008 (projected). The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Athanasios P. Liavas.

J. van de Laar is with the Digital Signal Processing Group, Philips Research Laboratories, 5656 AE Eindhoven, The Netherlands (e-mail: Jakob.van.de.Laar@philips.com; jakob.vd.laar@casema.nl).

M. Moonen is with the Electrical Engineering Department, Katholieke Universiteit Leuven, B-3001 Leuven-Herverlee, Belgium (e-mail: Marc.Moonen@esat.kuleuven.be).

P. C. W. Sommen is with the Electrical Engineering Department, Technische Universiteit Eindhoven, 5600 MB Eindhoven, The Netherlands (e-mail: p.c.w.sommen@tue.nl).

Color versions of one or more of the figures in this paper are available at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2008.926105

radar and sonar. Examples of IBSS can be found in the field of biomedical engineering, where the goal is e.g., to reveal independent sources in biological signals like EEGs or ECGs. Other examples can be found in the separation of speech and communication signals, images, etc. Although many practical problems can be described more adequately by more complex MIMO blind identification models such as convolutive and/or nonlinear models, MIBI often serves as a good starting point, e.g., for a frequency-domain approach in the convolutive case and provides enhanced insight.

This work is a continuation and elaboration of our previous work presented in [24]–[27]. In [24], a practical algorithm based on second-order statistics and the generalized eigenvalue decomposition (GEVD) was given for the real-valued square MIBI case with $D = S$. The underlying concepts were generalized in [25] to the exploitation of the temporal structure in the data of some arbitrary fixed order. In [26], we briefly touched upon the more general case with complex-valued mixing system and source signals, arbitrary order statistics, and arbitrary conjugation patterns, but the main focus was on the development of the so-called TIME-MUSIC algorithm for DOA estimation. In [27], we first used a homotopy method for estimating the mixing matrix from the system of equations derived for the real-valued SOS case.

In this paper, we present various results for real-valued MIBI based on exploiting the second-order temporal structure (SOTS) in the data, and again use a homotopy method for estimating the columns of the mixing matrix. Based on a few natural assumptions about the SOTS of the source and noise signals that are easily satisfied in many practical signal scenarios, we show that if this structure is exploited in a specific way the MIBI problem can be “projected onto” the problem of solving a well-structured system of homogeneous polynomial equations of degree two. By “projected,” we mean that the system of equations follows from our MIBI problem definition and formulation, but not necessarily the other way around. The physically plausible assumptions underlying our problem formulation primarily serve to ensure that sufficient temporal structure is present in the source signals, that the source and noise signals are mutually unrelated, and that the noise signals have a simpler temporal structure than the source signals. We will show how this latter assumption allows one to annihilate the influence of additive sensor noise (apart from errors in the estimation of the statistics) on the estimation of the mixing system by considering the sensor correlation functions on a so-called *noise-free region of support* (ROS), which only contains time–lag pairs, times, or lags, for which the sensor correlation functions are noise free. This principle is a major advantage of exploiting the temporal structure in the data. The condition on the required temporal structure of the source signals is formulated in terms of the linear independence of the source autocorrelation functions on the noise-free ROS. This is equivalent to saying that the source power spectral density functions are linearly independent, i.e., our method exploits the spectral diversity of the source signals. The MIBI method presented in this paper allows to estimate the mixing matrix for several *underdetermined* mixing scenarios with more sources than sensors without resorting to sparsity assumptions, something that is often believed to be impossible when using only second-order

statistics. Only a few methods exist that can blindly estimate the mixing matrix in underdetermined MIBI scenarios. Most of them assume that the source signals are sparse (possibly in some transform domain) [4], [14], [19]. Another category of methods that do not require the sparsity assumption is based on the algebraic decomposition of a higher order tensor, see e.g., the methods by Cardoso [5], de Lathauwer [11], and Comon and Mourrain [10].

The outline of this paper is as follows. Firstly, the structure and assumptions of the MIBI model are explained in Sections II and III respectively. Then, the derivation of the system of homogeneous polynomial equations is presented in Section IV. Next, in Section V, we present a homotopy method for solving the system of equations. In Section VI, the theory is applied to several examples and performance results are presented. Finally, conclusions are discussed in Section VII. It should be noted that the analysis presented in this paper also applies to complex-valued systems and signals with any possible way of conjugating the arguments of the employed correlation functions.

II. MIBI MODEL STRUCTURE

A block diagram of the MIBI problem setup is shown in Fig. 1. S mutually statistically independent but otherwise unknown source signals are mixed by an unknown MIMO linear instantaneous mixing system, which is represented by the matrix $\mathbf{A}$, and D sensor signals $x_1[n], \dots, x_D[n]$ corrupted by D additive noise signals $\nu_1[n], \dots, \nu_D[n]$ are observed. For convenience and without loss of generality it is assumed that all signals are zero-mean. Mathematically, the MIBI observation model can be written as follows:

$$\mathbf{x}[n] = \sum_{j=1}^S \mathbf{a}^j s_j[n] + \boldsymbol{\nu}[n] = \mathbf{A}\mathbf{s}[n] + \boldsymbol{\nu}[n] \quad \forall n \in \mathbb{Z} \quad (\text{II.1})$$

where

$$\mathbf{x}[n] \triangleq \begin{bmatrix} x_1[n] \\ \vdots \\ x_D[n] \end{bmatrix}, \mathbf{s}[n] \triangleq \begin{bmatrix} s_1[n] \\ \vdots \\ s_S[n] \end{bmatrix}, \boldsymbol{\nu}[n] \triangleq \begin{bmatrix} \nu_1[n] \\ \vdots \\ \nu_D[n] \end{bmatrix}, \mathbf{a}^j \triangleq \begin{bmatrix} a_1^j \\ \vdots \\ a_D^j \end{bmatrix}$$

are real-valued column vectors of sensor signals, source signals, additive noise signals, and mixing elements respectively. The vectors $\mathbf{x}[n]$, $\boldsymbol{\nu}[n]$, and $\mathbf{a}^1, \dots, \mathbf{a}^S$ are length- D column vectors, whereas the vector $\mathbf{s}[n]$ is a length- S column vector. The unknown mixing matrix $\mathbf{A}$ is of size $D \times S$ and can be written in terms of its columns as $\mathbf{A} = [\mathbf{a}^1 \cdots \mathbf{a}^S]$. Sometimes we refer to the columns of the mixing matrix as the *array response vectors*. Subscript and superscript indexes are used to index the components of a column and row vector respectively. Furthermore, the symbol n denotes discrete time. According to (II.1), the i th sensor signal $x_i[n]$ is given by:

$$x_i[n] = \sum_{j=1}^S a_i^j s_j[n] + \nu_i[n] \quad \forall n \in \mathbb{Z} \quad \forall 1 \leq i \leq D \quad (\text{II.2})$$

where a_i^j denotes the instantaneous transfer coefficient from the j th source to the i th sensor, $s_j[n]$ the j th source signal at discrete time n , and $\nu_i[n]$ the i th noise signal at discrete time n .

From (II.1) it follows directly that two indeterminacies are inherent to the MIBI model [9], [12], viz. the norms and order of the mixing matrix columns and the source signals cannot be resolved. This means that the columns and source signals can only be recovered up to a scaling factor and a permutation. Taking into account these indeterminacies, the goal of MIBI is to recover the columns of the mixing system in some arbitrary order and with arbitrary nonzero norms.

III. MIBI MODEL ASSUMPTIONS

In order to be able to exploit the SOTS in the data, several assumptions are made that mainly serve to ensure that sufficient (second-order) temporal structure is present in the source signals and that the noise signals have a “simpler” temporal structure than the source signals. The assumptions are formulated in terms of correlation functions that we will define now. Let $v_i[n]$ be the i th component of a length- G time dependent random vector $\mathbf{v}[n]$ at time index n . Then, for all $1 \leq i_1, i_2 \leq G$ the correlation function of the i_1 th and i_2 th elements of $\mathbf{v}[n]$ at time n and lag k is defined as follows:

$$r_{i_1 i_2}^v[n, k] \triangleq E \{v_{i_1}[n]v_{i_2}[n - k]\} \quad \forall n, k \in \mathbb{Z}.$$

Likewise, for all $1 \leq i \leq G$ and $1 \leq j \leq P$, the two-dimensional (cross-)correlation function of the i th component of a length- G time dependent random vector $\mathbf{v}[n]$ and the j th component of another length- P random vector $\mathbf{w}[n]$ at time n and lag k is denoted by $r_{ij}^{vw}[n, k]$ and defined as

$$r_{ij}^{vw}[n, k] \triangleq E \{v_i[n]w_j[n - k]\} \quad \forall n, k \in \mathbb{Z}.$$

In the formulation of the assumptions underlying our MIBI method (see AS1)–AS4) further in this section) we use the concept of a so-called *noise-free ROS*. We define the noise-free ROS Ω as a set in the domain of time-lag index pairs (n, k) , times n , or lags k , on which no noise enters the computation of the sensor correlation functions. Suppose that we have specified *a priori* such a noise-free ROS by a set of $N \geq S$ time pairs as follows:

$$\Omega = \{(n_1, k_1), \dots, (n_N, k_N)\} \quad (\text{III.1})$$

where (n_q, k_q) is the q th time-lag pair of Ω . The time-lag pairs can be chosen in such a way that the coloredness and/or non-stationarity of the source signals is exploited. This requires some *a priori* knowledge about the signal structure. For example, if the source signals are colored and the sensor noise is temporally white (with any spatial correlation structure), then a suitable noise-free ROS consists of a set of nonzero lags and the time-dependence can be omitted, e.g., $\Omega = \{1, 2, 3, \dots\}$, where the elements are lags (hence, the time-lag pairs in (III.1) degenerate to lags). If the source signals are temporally white, but non- or quasi-stationary, and there is no noise, then a suitable noise-free ROS consists of a set of time or block indexes with zero lags (hence, the time-lag pairs in (III.1) degenerate to time or block indices). In general, we have a scenario in between the previous two. For example, if the source signals are both colored and non- or quasi-stationary, and the sensor noise is temporally white with any spatial structure, then a suitable noise-free ROS

consists of a set of pairs, each of which contains a time or block index and a nonzero lag index. See Section VI for examples with speech signals. Note that the prior knowledge mentioned in the examples above in general is not stronger than the prior knowledge adopted in other signal processing applications, where for example it is often assumed that the sensor noise is (spatially and temporally) white and Gaussian. We are now in a position to formulate the assumptions underlying the MIBI method presented in the sequel as follows:

AS1) the source signals have zero cross-correlation on the noise-free ROS Ω :

$$r_{j_1 j_2}^s[n, k] = 0 \quad \forall 1 \leq j_1 \neq j_2 \leq S;$$

AS2) the source autocorrelation functions are linearly independent on the noise-free ROS Ω :

$$\sum_{j=1}^S \xi^j r_{jj}^s[n, k] = 0 \implies \xi^j = 0 \quad \forall 1 \leq j \leq S;$$

AS3) the noise signals have zero auto- and cross-correlation functions on the noise-free ROS Ω :

$$r_{i_1 i_2}^v[n, k] = 0 \quad \forall 1 \leq i_1, i_2 \leq D;$$

AS4) the cross-correlation functions between the source and noise signals are zero on the noise-free ROS Ω :

$$r_{ij}^{vs}[n, k] = r_{ji}^{sv}[n, k] = 0 \quad \forall 1 \leq i \leq D, 1 \leq j \leq S.$$

First, AS1) ensures that the source signals are mutually uncorrelated. Second, AS2) ensures that sufficient temporal structure is present in the source signals by specifying precisely how the auto-correlation structures of the source signals should differ from each other. Equivalently, this assumption requires that the source power spectral density functions, or functions that are defined similarly from parts of the source auto-correlation functions, are linearly independent. Note that AS2) requires the source auto-correlation matrix (see (IV.2)) to be full rank. As a matter of fact, many algorithms based on SOS work by virtue of AS2) without being explicitly mentioned so by the developers. Usually, it is merely stated that the source auto-correlation functions have to be different, which is often an inaccurate and insufficient condition. Third, AS3) states that the noise correlation functions are zero on the considered ROS. Finally, AS4) ensures that the source and noise signals are mutually uncorrelated. Essentially, AS3) and AS4) define the noise-free ROS Ω , or, equivalently, its complement. Note that Ω should contain at least S elements because otherwise AS1)–AS4) cannot be satisfied. Together, AS1)–AS4) imply that the noise signals have a simpler temporal structure than the source signals. Apart from errors in the estimation of the statistics, the influence of additive sensor noise on the estimation of the mixing system can be annihilated by considering the sensor correlation functions only on the noise-free ROS. This principle is a major advantage of exploiting the temporal structure in the data.

It should be noted that we have made *no assumptions on the mixing matrix $\mathbf{A}$* . The identification method based on

AS1)–AS4) that we will develop in the sequel can also identify mixing matrices that are rank-deficient or even have identical columns. Most IBSS methods not only require that the number of sensors is larger than the number of sources, but also that the mixing matrix is full rank. As far as robustness and generality are concerned, this is a significant advantage w.r.t. to other methods. In addition, note also that no assumptions are required on the probability density functions of the noise and source signals. Hence, arbitrarily distributed signals can be dealt with as long as assumptions AS1)–AS4) are satisfied.

IV. FORMULATING MIBI AS THE PROBLEM OF SOLVING A SYSTEM OF HOMOGENEOUS POLYNOMIAL EQUATIONS

Using the assumptions made in the previous section, in this section we will show that the array response vectors $\mathbf{a}^1, \dots, \mathbf{a}^S$ satisfy a well-structured system of D -variate homogeneous polynomial equations of degree two, thereby “projecting” the MIBI problem onto a mathematical problem, the solution of which yields estimates of the array response vectors. In the course of our derivation we highlight the algebraic structure of the problem formulation.

A. Sensor Correlation Matrix Definition and Structure

We start our derivation by expressing the sensor correlation functions in terms of the mixing matrix elements and source auto-correlation functions. Then, several sensor correlation function values are arranged in a matrix in a particular fashion. Using AS1)–AS4), it follows that the sensor correlation functions can be expressed on Ω in terms of the mixing matrix elements and source auto-correlation functions as

$$r_{i_1 i_2}^x[n, k] = \sum_{j=1}^S a_{i_1}^j a_{i_2}^j r_{jj}^s[n, k] \quad \forall (n, k) \in \Omega, \quad 1 \leq i_1, i_2 \leq D.$$

We now stack all those correlation values in the $(D)^2 \times N$ sensor correlation matrix $\mathbf{C}_D^x$ that is defined as [see the equation shown at the bottom of the page], with the q th column given by

$$\mathbf{r}_D^x[n_q, k_q] \triangleq E \{ \mathbf{x}[n_q] \otimes \mathbf{x}[n_q - k_q] \} \quad (\text{IV.1})$$

where $\otimes$ denotes the Kronecker product, and the time-lag pairs are given in (III.1). Likewise, the $S \times N$ source autocorrelation matrix $\mathbf{C}_S^s$ is defined as follows:

$$\begin{aligned} \mathbf{C}_S^s &\triangleq [\mathbf{r}_S^s[n_1, k_1] \cdots \mathbf{r}_S^s[n_N, k_N]] \\ &= \begin{bmatrix} r_1^s[n_1, k_1] & r_1^s[n_2, k_2] & \cdots & r_1^s[n_N, k_N] \\ r_2^s[n_1, k_1] & r_2^s[n_2, k_2] & \cdots & r_2^s[n_N, k_N] \\ \vdots & \vdots & \ddots & \vdots \\ r_S^s[n_1, k_1] & r_S^s[n_2, k_2] & \cdots & r_S^s[n_N, k_N] \end{bmatrix} \end{aligned} \quad (\text{IV.2})$$

with the q th column given by

$$\mathbf{r}_S^s[n_q, k_q] \triangleq E \{ \mathbf{s}[n_q] \odot \mathbf{s}[n_q - k_q] \} \quad (\text{IV.3})$$

where $\odot$ denotes the elementwise product. The linear space spanned by the rows of the source autocorrelation matrix $\mathbf{C}_S^s$ is called the *source subspace* and its dimension d_S^s equals the rank of the source auto-correlation matrix. From the equations above, it follows immediately that

$$\boxed{\mathbf{C}_D^x = \mathbf{A}_{D, \diamond} \mathbf{C}_S^s} \quad (\text{IV.4})$$

where $\mathbf{A}_{D, \diamond}$ is the so-called second-order *Khatri–Rao product* of $\mathbf{A}$, which is defined as

$$\mathbf{A}_{D, \diamond} \triangleq [\mathbf{a}^1 \otimes \mathbf{a}^1 \cdots \mathbf{a}^S \otimes \mathbf{a}^S]. \quad (\text{IV.5})$$

B. Dimensions of Signal and Source Subspaces

We now show how the rank d_D^x of the sensor correlation matrix $\mathbf{C}_D^x$ depends on the mixing matrix $\mathbf{A}$ and the source signal properties. Firstly, AS2) implies that the source autocorrelation matrix $\mathbf{C}_S^s$ has full rank. Hence, the source subspace dimension d_S^s equals the number of sources

$$d_S^s \triangleq \text{rank}(\mathbf{C}_S^s) = S. \quad (\text{IV.6})$$

Using linear algebra, it then follows from (IV.4) that

$$\boxed{d_D^x \triangleq \text{rank}(\mathbf{C}_D^x) = \text{rank}(\mathbf{A}_{D, \diamond})}. \quad (\text{IV.7})$$

$$\mathbf{C}_D^x \triangleq [\mathbf{r}_D^x[n_1, k_1] \cdots \mathbf{r}_D^x[n_N, k_N]] = \begin{bmatrix} r_{11}^x[n_1, k_1] & r_{11}^x[n_2, k_2] & \cdots & r_{11}^x[n_N, k_N] \\ r_{12}^x[n_1, k_1] & r_{12}^x[n_2, k_2] & \cdots & r_{12}^x[n_N, k_N] \\ \vdots & \vdots & \ddots & \vdots \\ r_{D, D-1}^x[n_1, k_1] & r_{D, D-1}^x[n_2, k_2] & \cdots & r_{D, D-1}^x[n_N, k_N] \\ r_{DD}^x[n_1, k_1] & r_{DD}^x[n_2, k_2] & \cdots & r_{DD}^x[n_N, k_N] \end{bmatrix}.$$

The number of unique rows of $\mathbf{A}_{D,\diamond}$ equals $(1/2)D(D+1)$. In general, if the mixing matrix is full rank and if $S \leq (1/2)D(D+1)$, then $(\mathbf{A}_{D,\diamond}) = S$, which implies that $d_D^x = S$. Hence, if the number of sources is unknown, it can be estimated as the effective rank of the subspace matrix. The rank of the Khatri–Rao product matrix has been studied in several works; see, e.g., [20] and [29].

C. Deriving the System of Homogeneous Polynomial Equations

If the number of rows of the sensor correlation matrix $\mathbf{C}_D^x$ is larger than the dimension d_D^x of the subspace spanned by its rows, then $\mathbf{C}_D^x$ has a nonzero left null space $\mathcal{N}_l(\mathbf{C}_D^x)$. Let Φ be a matrix such that its rows form a basis for $\mathcal{N}_l(\mathbf{C}_D^x)$, i.e., $\Phi \mathbf{C}_D^x = \mathbf{0}$. The matrix Φ can be determined directly from the singular value decomposition (SVD) of $\mathbf{C}_D^x$. The maximum number of linearly independent rows of Φ equals $\dim(\mathcal{N}_l(\mathbf{C}_D^x)) = (D)^2 - d_D^x$. Substituting (IV.4) into the equation $\Phi \mathbf{C}_D^x = \mathbf{0}$, and using the fact that $\mathbf{C}_S^s$ has full rank due to AS2), it follows immediately that

$$\Phi \mathbf{A}_{D,\diamond} = \mathbf{0}. \quad (\text{IV.8})$$

This system describes the relation between the *unknown* coefficients of the mixing matrix $\mathbf{A}$ and the *known* coefficients of the matrix Φ . Let $\tilde{\varphi}_q$ be the q th row of Φ , let $\mathbf{z} \in \mathbb{C}_D$ be a vector of variables with the same size as a column of $\mathbf{A}$, and define the functions $f_{D,q}(\mathbf{z})$ with $1 \leq q \leq (D)^2 - d_D^x$ as

$$f_{D,q}(\mathbf{z}) \triangleq \tilde{\varphi}_q \cdot (\mathbf{z} \otimes \mathbf{z}) = \sum_{i_1=1}^D \sum_{i_2=1}^D \varphi_q^{i_1 i_2} z_{i_1} z_{i_2}. \quad (\text{IV.9})$$

Then, (IV.8) states that all columns $\mathbf{a}^1, \dots, \mathbf{a}^S$ of $\mathbf{A}$ satisfy

$$f_{D,q}(\mathbf{a}^j) = 0 \quad \forall 1 \leq q \leq (D)^2 - d_D^x, \quad 1 \leq j \leq S.$$

Hence, at this point the MIBI problem has been “projected” onto the problem of solving the following system of equations for the columns of the mixing matrix $\mathbf{A}$:

$$\{f_{D,q}(\mathbf{z}) = 0\}_{1 \leq q \leq (D)^2 - d_D^x}. \quad (\text{IV.10})$$

By “projected,” we mean that the system of equations follows from our MIBI problem definition and formulation, but not necessarily the other way around. See Algorithm 1 for a summary of the results that we have developed so far. Note that, as opposed to many other approaches such as AMUSE and SOBI, the mixing matrix columns are determined directly by properly combining information from different times and/or lags without

resorting to the two-stage procedure of whitening followed by identification of a rotation matrix.

Algorithm 1: $D \times S$ MIBI based on SOTS

- 1: Estimate sensor correlation functions on noise-free ROS.
- 2: Arrange these values in sensor correlation matrix $\mathbf{C}_D^x$.
- 3: Compute SVD of $\mathbf{C}_D^x$ and split result into signal and noise subspace parts as follows:

$$\mathbf{C}_D^x = \mathbf{U} \Sigma \mathbf{V}^T = \mathbf{U}^s \Sigma_s^s \mathbf{V}_s^T + \mathbf{U}^\nu \Sigma_\nu^\nu \mathbf{V}_\nu^T.$$

- 4: Let Φ be a matrix with $(D)^2 - d_D^x$ rows and $(D)^2$ columns, whose rows span the left null space of $\mathbf{C}_D^x$:

$$\Phi \triangleq (\mathbf{U}^\nu)^T.$$

- 5: With each row $\tilde{\varphi}_q$ of Φ associate a D -variate homogeneous polynomial of degree two:

$$f_{D,q}(\mathbf{z}) \triangleq \tilde{\varphi}_q \cdot (\mathbf{z} \otimes \mathbf{z}) = \sum_{i_1=1}^D \sum_{i_2=1}^D \varphi_q^{i_1 i_2} z_{i_1} z_{i_2}.$$

- 6: The following system remains to be solved for the columns of the mixing matrix (see Algorithm 2):

$$\{f_{D,q}(\mathbf{z}) = 0\}_{1 \leq q \leq (D)^2 - d_D^x}.$$

D. Properties and Structure of Functions and System

All functions in system (IV.10) have the same specific form. First, from the definition in (IV.9), it is clear that each function $f_{D,q}(\mathbf{z}) = f_{D,q}(z_1, \dots, z_D)$ is a D -variate polynomial function containing only product terms of degree two. This implies that each function $f_{D,q}(\mathbf{z})$ is *homogeneous of degree two*, meaning that

$$f_{D,q}(\eta \mathbf{z}) = (\eta)^2 f_{D,q}(\mathbf{z}) \quad \forall \eta \in \mathbb{C} \quad \forall \mathbf{z} \in \mathbb{C}_D. \quad (\text{IV.11})$$

Hence, if $\mathbf{v}$ is a solution of (IV.10), then so is $\eta \mathbf{v}$ for all $\eta \in \mathbb{C}$. This is reminiscent of the scaling indeterminacy inherent to MIBI (see also Section II) and implies that the norms of the solutions of system (IV.10) can be chosen arbitrarily. Algebraically, the zero contour level of each function $f_{D,q}(z_1, \dots, z_D)$ defines a quadratic *cone* in the D -dimensional Euclidian space. Hence, geometrically, solving (IV.10) is equivalent to finding the one-dimensional intersections between $(D)^2 - d_D^x$ quadratic cones that are embedded in a D -dimensional Euclidian space, where ideally each intersection corresponds to a column of the mixing matrix. See Section VI for examples. The Kronecker product $\mathbf{z} \otimes \mathbf{z}$ in the definition of $f_{D,q}(\mathbf{z})$ in (IV.9) is a vector of length $(D)^2$ containing only $(1/2)D(D+1)$ different products. Combining the coefficients of $\tilde{\varphi}_q$ in (IV.9) corresponding

to equal products, and also combining the corresponding rows of the sensor correlation matrix yields a nonredundant system of $(1/2)D(D+1) - d_D^x$ equations containing functions of the form

$$f_{D,q}(\mathbf{z}) \triangleq \sum_{i_1=1}^D \sum_{i_2=i_1}^D \alpha_q^{i_1 i_2} z_{i_1} z_{i_2} \quad (\text{IV.12})$$

where for each $q = 1, \dots, (1/2)D(D+1) - d_D^x$ the coefficients follow from the SVD of the corresponding reduced sensor correlation matrix.

E. Maximum Number of Identifiable Columns

We will now derive heuristically the maximum number of columns of the mixing matrix $\mathbf{A}$ that can be identified with our approach, i.e., which are determined uniquely by the system of equations that we have derived above. Generally, that is in the non-degenerate case, the zero contour level of a D -variate homogenous function is a $(D-1)$ -dimensional surface embedded in a D -dimensional Euclidian space. From algebra, geometry, and intuition, it is known that in a D -dimensional Euclidian space at least $D-1$ surfaces of degree $D-1$ are required in order to uniquely define one-dimensional intersections. For example, in three-dimensional space two surfaces of dimension two generally define a one-dimensional intersection curve. Algebraically, the aforesaid means that in general the dimension of the solution set of a system of $D-1$ homogeneous polynomials in D variables has dimension one. Hence, since at least $D-1$ equations are required to uniquely define the one-dimensional solution sets corresponding to the columns of $\mathbf{A}$, and we know that there are $(1/2)D(D+1) - d_D^x$ unique equations in our system, we find the constraint $(1/2)D(D+1) - d_D^x \geq D-1$. As we have seen in Section IV-B, generally, if the mixing matrix is full rank and if $S \leq (1/2)D(D+1)$, then $d_D^x = S$. Hence, it follows that (an upper bound on) the maximum number of columns that can be identified with D sensors equals

$$S_{\max} = \frac{1}{2}D(D+1) - (D-1). \quad (\text{IV.13})$$

For example, under the assumption that the employed sensor correlations can be estimated with sufficient accuracy, it follows from (IV.13) that with 2, 3, 4, 5 and 6 sensors, we can identify 2, 4, 7, 11 and 16 columns, respectively, with second-order statistics only and without resorting to sparsity assumptions. Extensive simulations have confirmed that (IV.13) indeed gives the true maximum number of columns that can be identified (see also Section VI-B).

V. SOLVING THE SYSTEM OF POLYNOMIAL EQUATIONS BY MEANS OF A HOMOTOPY METHOD

In this section we summarize the main ideas behind the so-called homotopy method that provides a deterministic means for solving a system of nonlinear equations and can

also deal with non-exact equations. The resulting algorithm is summarized in Algorithm 2 and will be employed in the next section for solving systems of the form (IV.10). Homotopy methods are based on the so-called path following or continuation techniques. Excellent discussions can be found in several articles and books; see, e.g., [1], [15], [18], and [21]. In the context of homotopy methods, the system of equations to be solved is commonly referred to as the *target system*. In this paper, we denote a general target system by $\mathbf{p}(\mathbf{z}) = \mathbf{0}$ and assume that $\mathbf{p} : \mathbb{C}_D \rightarrow \mathbb{C}_P$. Furthermore, its solution set is denoted by $\mathcal{P}$, i.e., $\mathcal{P} \triangleq \{\mathbf{z}_p \in \mathbb{C}_D | \mathbf{p}(\mathbf{z}_p) = \mathbf{0}\}$. The target systems we consider here consist of bivariate homogeneous polynomial equations of degree two together with a unit-norm constraint. The rationale behind homotopy methods is to deform the known solutions of a simple *start system* into the *a priori* unknown solutions of the *target system*, where the start system should be compatible with the target system as much as possible in the sense that it possesses the same or a similar structure and has at least the same number of solutions. For our current problem, this primarily means that it should also consist of bivariate homogeneous polynomial equations of degree two; see (V.5) for an example. We denote a general start system by $\mathbf{g}(\mathbf{z}) = \mathbf{0}$, where $\mathbf{g} : \mathbb{C}_D \rightarrow \mathbb{C}_P$, and its solution set by $\mathcal{G}$, i.e., $\mathcal{G} \triangleq \{\mathbf{z}_g \in \mathbb{C}_D | \mathbf{g}(\mathbf{z}_g) = \mathbf{0}\}$. The start and target systems are embedded in a family of systems, the *homotopy*, which we define as follows:

$$\mathbf{h}(\mathbf{z}, \lambda) = \gamma_g(\lambda - \lambda_e)\mathbf{g}(\mathbf{z}) + \gamma_p(\lambda - \lambda_0)\mathbf{p}(\mathbf{z}) \quad \forall \lambda \in \mathcal{C} \quad (\text{V.1})$$

where $\lambda \in \mathbb{C}$ is called the so-called *continuation parameter* and $\mathcal{C}$ is a certain predefined curve in the complex plane that is to be traversed by λ . The curve $\mathcal{C}$ has a starting point λ_0 that is different from its end point λ_e and it can be chosen in several manners [21]. Here, we use the line segment $\mathcal{C} \triangleq \{s+j\lambda | 0 \leq s \leq 1\}$ with starting point $\lambda_0 = 0$ and end point $\lambda_e = 1 + j$. The constants $\gamma_g \in \mathbb{C}$ and $\gamma_p \in \mathbb{C}$ are randomly chosen fixed constants whose purpose will be explained soon. For each solution $\mathbf{z}_g$ of the start system $\mathbf{g}(\mathbf{z})$, numerical path following or continuation methods attempt to trace the path $(\mathbf{z}, \lambda) \in \mathbf{h}^{-1}(\mathbf{0})$ defined implicitly by $\mathbf{h}(\mathbf{z}, \lambda) = \mathbf{0}$ from the starting point $(\mathbf{z}_g, \lambda_0)$, which is a solution of the start system, to a solution point $(\mathbf{z}_p, \lambda_e)$ of the target system. As is explained in [18] and [21], for example, the constants $\gamma_g \in \mathbb{C}$ and $\gamma_p \in \mathbb{C}$ are randomly chosen fixed constants that serve to avoid singularities and crossings along the different paths, i.e., they make the problem well-conditioned. If the homotopy system is deformed in small steps from the start system into the target system, at each iteration the new system is close to the previous system and under some mild conditions its solutions deviate only slightly from those of the previous system. Hence, each solution of the new system can be found by a local iterative optimization method that uses the corresponding solution of the previous system as its initial solution. For each solution of the start system, these steps are repeated until the target system is reached. In theory, under some mild uniqueness and smoothness assumptions each path converges to a geometrically isolated solution, especially for polynomial systems [1], [18].

Mathematically, the main problem to be solved by a homotopy method is to find the curve $(\mathbf{z}(\lambda), \lambda)$ that solves the following system as λ is moved from λ_0 to λ_e along $\mathcal{C}$:

$$\mathbf{h}(\mathbf{z}(\lambda), \lambda) = \mathbf{0} \quad \forall \lambda \in \mathcal{C}. \quad (\text{V.2})$$

This can be done by iteratively deforming the system and finding the solutions of the new system by means of the so-called *predictor* and *corrector* steps that can be derived from (V.2) (and (V.1)). Given a certain point on the path, we first consider a single iteration of the method. Suppose that we have a solution $\mathbf{z}(\lambda)$ of $\mathbf{h}(\mathbf{z}(\lambda), \lambda) = \mathbf{0}$ for a certain known value of λ . Then, this solution is an approximate solution of the slightly deformed system $\mathbf{h}(\mathbf{z}(\lambda + \Delta\lambda), \lambda + \Delta\lambda) = \mathbf{0}$, where $\Delta\lambda$ is a small increment in λ . The solution of the deformed system can be found in two steps. First, in the *predictor* step an approximation of the solution $\mathbf{z}(\lambda + \Delta\lambda)$ is predicted, e.g., by using an Euler step (see Algorithm 2). Second, the predicted approximate solution is corrected (possibly repeatedly) to lie “exactly” on the solution curve by applying a corrector step, which can be any local zero finding method such as Newton’s method (see Algorithm 2) [1]. Performing these two steps for each solution of the start system and each iteration from $\lambda = \lambda_0$ to $\lambda = \lambda_e$, we finally end up at the solutions of the target system. We can derive an Euler predictor and a Newton corrector as follows by considering a local model of the homotopy function via its first-order Taylor expansion in both $\mathbf{z}$ and λ :

$$\mathbf{h}(\mathbf{z} + \Delta\mathbf{z}, \lambda + \Delta\lambda) \approx \mathbf{h}(\mathbf{z}, \lambda) + \nabla_{\mathbf{z}}\mathbf{h}(\mathbf{z}, \lambda)\Delta\mathbf{z} + \partial_{\lambda}\mathbf{h}(\mathbf{z}, \lambda)\Delta\lambda$$

where $\nabla_{\mathbf{z}}\mathbf{h}$ is the Jacobian matrix of $\mathbf{h}(\mathbf{z}, \lambda)$ w.r.t. $\mathbf{z}$ and $\partial_{\lambda}\mathbf{h}$ is the partial derivative of $\mathbf{h}(\mathbf{z}, \lambda)$ w.r.t. λ . First, if we have a point $(\mathbf{z}, \lambda)$ on or near the path, we can predict a new approximate solution $\mathbf{z} + \Delta\mathbf{z}$ of $\mathbf{z}$ at $\lambda + \Delta\lambda$ by setting $\mathbf{h}(\mathbf{z}, \lambda)$ and $\mathbf{h}(\mathbf{z} + \Delta\mathbf{z}, \lambda + \Delta\lambda)$ to zero and solving for $\Delta\mathbf{z}$ to obtain the following Euler predictor step:

$$\Delta\mathbf{z} \approx -[\nabla_{\mathbf{z}}\mathbf{h}(\mathbf{z}, \lambda)]^{\dagger} \partial_{\lambda}\mathbf{h}(\mathbf{z}, \lambda)\Delta\lambda \quad (\text{V.3})$$

where the notation $[\cdot]^{\dagger}$ indicates the Moore–Penrose inverse. Second, if we have a point $(\mathbf{z}, \lambda)$ near the path for which $\mathbf{h}(\mathbf{z}, \lambda)$ is not (yet) sufficiently small, we can correct to a new approximate solution $\mathbf{z} + \Delta\mathbf{z}$ of $\mathbf{z}$ at λ by setting $\Delta\lambda$ and $\mathbf{h}(\mathbf{z} + \Delta\mathbf{z}, \lambda + \Delta\lambda)$ to zero and solving for $\Delta\mathbf{z}$ to obtain the following Newton corrector step:

$$\Delta\mathbf{z} \approx -[\nabla_{\mathbf{z}}\mathbf{h}(\mathbf{z}, \lambda)]^{\dagger} \mathbf{h}(\mathbf{z}, \lambda). \quad (\text{V.4})$$

The key ingredients of the resulting homotopy algorithm are summarized in Algorithm 2. The Euler predictor and Newton corrector steps given by (V.3) and (V.4), respectively, are used in the third step. Note that for any vector functions $\mathbf{g}(\mathbf{z})$ and $\mathbf{p}(\mathbf{z})$ the Jacobian matrix $\nabla_{\mathbf{z}}\mathbf{h}(\mathbf{z}, \lambda)$ and the partial derivative $\partial_{\lambda}\mathbf{h}(\mathbf{z}, \lambda)$ can be computed from (V.1). At each fixed value of λ , we stop applying the corrector steps either if the Newton correction becomes smaller than a certain tolerance $NwtTol$ (set to 10^{-8} in all our simulations), or when a certain maximum number $MaxNwtIt$ of Newton corrections has been applied (set to 3). Useful choices for γ_g and γ_p are given by $\gamma_g = \exp(j\theta_g)$ and $\gamma_p = \exp(j\theta_p)$ with θ_g and θ_p picked at

random from a uniform distribution on the interval $[0, 2\pi]$ [21]. Furthermore, from extensive simulations it turns out that only a few steps along the path $\mathcal{C}$ are sufficient. In the examples presented in Section VI, we will always use only 11 points. Hence, the stepsize $\Delta\lambda$ of λ along the path equals $\Delta\lambda = (1/10)(1+j)$. In all simulations, we employ the same basic start system

$$\begin{cases} \{g_q(\mathbf{z}) \triangleq (z_q)^2 - (z_D)^2 = 0\}_{1 \leq q \leq D-1} \\ g_D(\mathbf{z}) \triangleq \left(\sum_{i=1}^D (z_i)^2\right) - 1 = 0 \end{cases} \quad (\text{V.5})$$

where $g_q(\mathbf{z})$ is the q th function in $\mathbf{g}(\mathbf{z})$ and z_q is the q th variable in $\mathbf{z}$. For example, for $D = 3$ the functions in $\mathbf{g}(\mathbf{z})$ are given by $g_1(\mathbf{z}) \triangleq (z_1)^2 - (z_3)^2$, $g_2(\mathbf{z}) \triangleq (z_2)^2 - (z_3)^2$, and $g_3(\mathbf{z}) \triangleq (z_1)^2 + (z_2)^2 + (z_3)^2 - 1$. Note that the functions in the first part of the start system have the same homogeneous structure as the functions in the target systems we want to solve, and that the last function represents the unit norm constraint for real-valued vectors. The start system is simple and can directly be solved for its solutions: $(1/\sqrt{D})[\pm 1, \pm 1, \dots, \pm 1, 1]^T$.

Algorithm 2: Overview of Homotopy Continuation Method

1: Initialize $\gamma_g, \gamma_p, \Delta\lambda = (1/10)(1+j)$, $NwtTol = 10^{-8}$, and $MaxNwtIt = 3$.

2: Compute the solutions of the start system $\mathbf{g}(\mathbf{z}) = \mathbf{0}$ and store them in a set $\mathcal{G}$.

3: For each solution $\mathbf{z}_g \in \mathcal{G}$ of $\mathbf{h}(\mathbf{z}, \lambda_0) = \mathbf{0} \equiv \mathbf{g}(\mathbf{z}) = \mathbf{0}$, trace the curve $(\mathbf{z}(\lambda), \lambda) \in \mathbf{h}^{-1}(\mathbf{0})$ defined implicitly by $\mathbf{h}(\mathbf{z}(\lambda), \lambda) = \mathbf{0}$ from the starting point $(\mathbf{z}_g, \lambda_0)$ to a solution point $(\mathbf{z}_p, \lambda_e)$ of $\mathbf{h}(\mathbf{z}, \lambda_e) = \mathbf{0} \equiv \mathbf{p}(\mathbf{z}) = \mathbf{0}$:

for all $\mathbf{z}_g \in \mathcal{G}$ **do**

$\mathbf{z} := \mathbf{z}_g$;

for $\lambda = 0 \rightarrow (1+j) - \Delta\lambda$ **do**

% Euler predictor step:

$d\mathbf{z}/d\lambda = -[\nabla_{\mathbf{z}}\mathbf{h}]^{\dagger} \partial_{\lambda}\mathbf{h}$;

$\mathbf{z} := \mathbf{z} + \Delta\lambda(d\mathbf{z}/d\lambda)$;

% Newton corrector steps:

$\lambda := \lambda + \Delta\lambda$;

for $m = 1 \rightarrow MaxNwtIt$ **do**

$NwtCorr := -[\nabla_{\mathbf{z}}\mathbf{h}]^{\dagger} \mathbf{h}$;

$\mathbf{z} := \mathbf{z} + NwtCorr$;

if $norm(NwtCorr) < NwtTol$ **then**

Leave corrector loop;

end if

end for

end for

Store solution $\mathbf{z}_p = \mathbf{z}$ in set $\mathcal{P}$;

end for

Return $\mathcal{P}$;

VI. EXAMPLES WITH SPEECH AND THREE SENSORS

In this section, we present examples with speech source signals and three sensors, and compare the results to the well-known SOBI [2] and SOBI-RO [3], [8], [9] algorithms. For convenience and without any loss of generality, we assume that the columns of the mixing matrix $\mathbf{A}$ have unit Euclidian norm. We use the following performance index for measuring the IBSS performance of the various algorithms for the square mixing matrix case (see also [9]):

$$\text{PI} = (D)^2 \left(\sum_{i=1}^D \left\{ \left(\frac{\sum_{j=1}^D |t_{ij}^j|}{\max_k |t_{ik}^j|} - 1 \right) + \left(\frac{\sum_{j=1}^D |t_{ij}^i|}{\max_k |t_{jk}^i|} - 1 \right) \right\} \right)^{-1}$$

where t_{pq}^q is the element in the p th row and q th column of the total transfer matrix $\mathbf{T} \triangleq \hat{\mathbf{A}}^{-1} \mathbf{A}$ from the source to the output signals with $\hat{\mathbf{A}}$ an estimate of the mixing matrix $\mathbf{A}$. In practice, a value of PI larger than 30 indicates a rather good performance. The signal-to-noise ratio (SNR) is defined as the ratio of the variance of the noise-free part of the sensor observation vector to the variance of the noise vector, i.e.,

$$\text{SNR} \triangleq \frac{E\{\|\mathbf{A}\mathbf{s}[n]\|^2\}}{E\{\|\boldsymbol{\nu}[n]\|^2\}} = \frac{S(\sigma^s)^2}{D(\sigma^\nu)^2} \quad (\text{VI.1})$$

where $(\sigma^s)^2$ and $(\sigma^\nu)^2$ are the variances of the source and noise signals respectively. First, in Sections VI-A and VI-B, we apply our method to two examples with three and four speech source signals respectively and visualize the procedure and results. Then, in Section VI-C we compare the performance of our method to that of SOBI and SOBI-RO. The speech source signals are sampled at 8 kHz, consist of 10 000 samples (1250 ms), and are normalized to unit variance $(\sigma^s)^2 = 1$. The noise signals $\nu_1[n]$, $\nu_2[n]$ and $\nu_3[n]$ are mutually statistically independent white Gaussian noise sequences. In the examples presented in Sections VI-A and VI-B, and also in Section VI-C, we exploit both the nonstationarity and coloredness by using sensor correlation values for different times and different lags. To do so, the signal sequences are partitioned into five disjoint blocks consisting of 2000 samples (speech signals generally have significantly changed after 250 ms), and for each block, the one-dimensional sensor correlation functions are computed for lags 1, 2, 3, 4, and 5. Lag zero is omitted because the corresponding correlation values are noise-contaminated. Hence, in total for each sensor correlation function 25 values are estimated and employed, i.e., the employed noise-free ROS in the domain of block-lag pairs is given by

$$\Omega = \{(1,1), (1,2), \dots, (1,5), \dots, (5,1), (5,2), \dots, (5,5)\}$$

where the first index in each pair represents the block index and the second the lag index. For each block and lag the sensor correlation values are estimated by averaging products of the form $x_{i1}[n]x_{i2}[n-k]$ over the block length. For additional comparison, in Section VI-C we also compute the performance index as a function of the SNR when only stationarity is employed and only the first 2000 samples (250 ms) are used.

Fig. 2. Source and sensor signals for 3×3 example.

Fig. 3. Zero contour spherical ellipses of $f_1(z_1, z_2, z_3)$, $f_2(z_1, z_2, z_3)$ and $f_3(z_1, z_2, z_3)$ for 3×3 example.

A. Three Mixtures of Three Speech Signals

The sensor signals are obtained from (II.1) with

$$\mathbf{A} = \begin{bmatrix} 0.6749 & 0.4082 & 0.8083 \\ 0.5808 & -0.8165 & 0.1155 \\ 0.4552 & -0.4082 & -0.5774 \end{bmatrix} \quad (\text{VI.2})$$

and noise variance $(\sigma^\nu)^2 = 4$. Hence, the SNR defined in (VI.1) equals -6 dB, which is quite bad. Fig. 2 shows the source signals $s_1[n]$, $s_2[n]$ and $s_3[n]$ at the left side, and the three noise-contaminated mixtures $x_1[n]$, $x_2[n]$ and $x_3[n]$ at the right side. As can be seen, the signals of interest are completely buried in the sensor noise. Using only the sensor data, the functions in the system to be solved are obtained by means of Algorithm 1. From the results in Section IV-D, it follows that there are $(1/2)D(D+1) - d_D^2 = 3$ unique functions of the form (IV.12) in the system, i.e., we can write the system as follows: $\{f_1(z_1, z_2, z_3) = 0, f_2(z_1, z_2, z_3) = 0, f_3(z_1, z_2, z_3) = 0\}$. The surfaces describing the zero contour levels are quadric cones in three-dimensional space. In Fig. 3 we have depicted the intersections of each cone with the unit sphere in different colors/grey shades. These intersections are a kind of ‘ellipses

Fig. 4. Source and noise-free estimated source signals for 3×3 example.

on the sphere', which we will call 'spherical ellipses'. Note that each intersection consists of two parts that are point-symmetric with respect to the origin. The unit length black arrows represent the ideal columns of $\mathbf{A}$, whereas the gray arrows are the opposites of the black ones. The magenta prolonged arrows of length $3/2$ and the large dots indicate the columns of the estimate $\hat{\mathbf{A}}$ obtained by Algorithm 2. The figure shows that the columns of $\hat{\mathbf{A}}$ are aligned approximately with those of $\mathbf{A}$ and also that they point approximately to the points where three spherical ellipses intersect each other. Hence, the array response vectors are determined uniquely and correctly by the intersections induced by the system of equations. Because our purpose is to demonstrate MIBI and IBSS, and not noise reduction, we apply the inverse of $\hat{\mathbf{A}}$ to the noise-free sensor signals to recover the source signals (recall however, that $\hat{\mathbf{A}}$ has been estimated from the noisy sensor data), i.e., $\mathbf{y}[n] = \hat{\mathbf{A}}^{-1}(\mathbf{x}[n] - \mathbf{v}[n])$. Fig. 4 shows the source signals at the left side, and the noise-free estimated source signals at the right side. Evidently, the signals are well-separated. This also follows from the total transfer matrix from the source to the output signals, which is given by

$$\mathbf{T} \triangleq \hat{\mathbf{A}}^{-1} \mathbf{A} = \begin{bmatrix} 1.0009 & -0.0479 & -0.0145 \\ -0.0155 & -0.0093 & -0.9857 \\ -0.0186 & -0.9771 & -0.0410 \end{bmatrix}.$$

Evidently, this total transfer matrix approximately equals the product of a permutation and a diagonal matrix. This is also clear from Fig. 4. The performance index PI for $\mathbf{T}$ equals 30, which indicates quite a good performance for such a bad SNR (see also Section VI-C). In addition, listening to the separated signals confirmed that the separation was successful.

B. Three Mixtures of Four Speech Signals

In this example we mix four speech signals according to (II.1) with the 3×4 mixing matrix $\mathbf{A}$ obtained by appending one column, viz. $[-0.1690, -0.5071, 0.8452]^T$, to the matrix given by (VI.2). Again the noise variance $(\sigma^v)^2 = 4$ and thus the SNR now equals -4.8 dB. It follows from the results in Section IV-D that now there are only $(1/2)D(D+1) - d_D^2 = 2$ unique trivariate homogeneous polynomial functions in the system. As in Fig. 3, in Fig. 5 we have depicted the intersections of each

Fig. 5. Zero contour spherical ellipses of $f_1(z_1, z_2, z_3)$ and $f_2(z_1, z_2, z_3)$ for 3×4 example.

zero contour level cone with the unit sphere. Using Algorithm 2 yields the following estimate of $\mathbf{A}$:

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.5952 & -0.1107 & -0.4176 & -0.8050 \\ 0.5981 & -0.6115 & 0.8171 & -0.1778 \\ 0.5367 & 0.7834 & 0.3975 & 0.5660 \end{bmatrix}.$$

The columns of $\hat{\mathbf{A}}$ are indicated in Fig. 5 by the prolonged arrows and large dots, and again the columns of $\mathbf{A}$ are indicated by unit length black and gray arrows. Let θ_j be the angle between the j th column of $\mathbf{A}$ and its estimate. Then, $\theta_1 = 6.6^\circ$, $\theta_2 = 0.8^\circ$, $\theta_3 = 3.6^\circ$, and $\theta_4 = 7.7^\circ$. Again, we see that the estimated columns approximately equal the ideal ones and are determined uniquely by the two intersecting cones. The deviations from the true columns are only due to errors in the estimation of the statistics. If more and/or better conditioned data is used, the estimates will converge to the true columns. It should also be realized that the more source signals are mixed into the same number of sensors, the more the sensor signals become correlated and hence the more data is required for reliable estimation of the mixing matrix.

We have seen that solving a system consisting of trivariate homogeneous polynomial equations of degree two is equivalent to finding the intersections between the corresponding two-dimensional quadric cones embedded in a three-dimensional Euclidean space. Because such cones can intersect in at most four directions, the maximum number of columns that can be identified with three sensors equals four, which also follows from (IV.13) and is equal to the maximum number of different intersections between two (spherical) ellipses.

C. Performance Comparison

In this section, we compare our method to SOBI and SOBI-RO, which also employ the SOTS in the data and have comparable computational complexities. Using (II.1) with $\mathbf{A}$ given by (VI.2), we compute the performance index PI for each of the algorithms as a function of the SNR for two cases.

Fig. 6. Performance when using five blocks of 2000 samples and five lags for MIBI, and one block of 10 000 samples and 25 lags for SOBI and SOBI-RO.

Fig. 7. Performance when using one block of 2000 samples and 25 lags for all methods.

In the first experiment, for MIBI we exploit both the non-stationarity and coloredness by using sensor correlation values for the different times and lags described at the beginning of Section VI. Hence, we use five blocks of 2000 samples with five lags each. In the SOBI and SOBI-RO algorithms we use the same amount of data and 25 lags. The results are depicted in Fig. 6. It can be seen that SOBI performs worst for all SNRs and that above an SNR of 6-dB SOBI-RO and our MIBI method approximately deliver the same very good performance. We also see that especially for very low SNR values our method significantly outperforms the other methods. In the second experiment, we only exploit the stationarity by using sensor correlation values computed over the first 2000 samples for 25 lags. In this case, all algorithms use exactly the same correlation values as their input, thereby providing a completely fair comparison. The results of this experiment are shown in Fig. 7. Now, the new method clearly outperforms the others over the whole SNR range (except for very low values). Because of the relatively short block length, the estimates of the correlation values deviate significantly from their true values. Clearly, the MIBI homotopy algorithm can cope the best with this difficult scenario.

VII. CONCLUSION

We have presented a procedure for exploiting SOTS in the source signals such as coloredness and/or nonstationarity for performing MIMO Instantaneous blind identification (MIBI). The MIBI problem has been reformulated in such a way that each column of the unknown mixing matrix satisfies a system of multivariate homogeneous polynomial equations that is subsequently solved by means of a homotopy method. Our main natural assumption regarding the SOTS of the source signals is that the source auto-correlation functions are linearly independent. Equivalently, this assumption requires that the source power spectral density functions are linearly independent. We have shown that apart from errors in the estimation of the statistics the influence of additive sensor noise on the estimation of the mixing system can be annihilated by considering the sensor correlation functions only on a so-called *noise-free ROS*, which exists if the noise signals have a simpler temporal structure than the source signals. This principle is a major advantage of exploiting temporal structure. Furthermore, if the number of sources is unknown, it can be estimated as the effective rank of the sensor correlation matrix. Our blind identification procedure allows to estimate the mixing matrix for scenarios with more sources than sensors without resorting to sparsity assumptions, something that is often believed to be impossible when using only second-order statistics. In addition, our algorithm does not require *any* assumption on the mixing matrix and can identify mixing matrices that are rank-deficient or even have identical columns. As far as robustness and generality are concerned, this is a significant advantage w.r.t. to other methods. Finally, using speech source signals, we have illustrated various examples and performance results. We have shown that especially for bad SNRs our method outperforms the well-known SOBI and SOBI-RO methods.

ACKNOWLEDGMENT

The authors would like to thank Dr. E. Habets very much for proofreading this manuscript and for his valuable comments.

REFERENCES

- [1] E. Allgower and K. Georg, *Introduction to Numerical Continuation Methods*, ser. Classics in Applied Mathematics. Philadelphia, PA: SIAM, 2003.
- [2] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique based on second order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.
- [3] A. Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context," *Electron. Lett.*, vol. 36, no. 24, pp. 2050–2053, 2000.
- [4] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Process., Elsevier*, vol. 81, no. 11, pp. 2353–2362, 2001.
- [5] J.-F. Cardoso, "Super-symmetric decomposition of the fourth-order cumulant tensor. Blind identification of more sources than sensors," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 1991)*, Toronto, ON, Canada, 1991, vol. 5, pp. 3109–3112.
- [6] S. Choi and A. Cichocki, "Blind separation of nonstationary and temporally correlated sources from noisy mixtures," in *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP 2000)*, Sydney, Australia, Dec. 2000, pp. 405–414.
- [7] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," *Electron. Lett.*, vol. 36, no. 9, pp. 848–849, Apr. 2000.
- [8] S. Choi, A. Cichocki, and A. Belouchrani, "Second order nonstationary source separation," *J. VLSI Signal Process.*, vol. 32, no. 1–2, pp. 93–104, Aug. 2002.

- [9] A. Cichocki and S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: Wiley, 2002.
- [10] P. Comon and B. Mourrain, "Decomposition of quantics in sums of powers of linear forms," *Signal Process., Elsevier*, vol. 53, no. 2, pp. 93–107, Sep. 1996.
- [11] L. de Lathauwer, "Signal processing based on multilinear algebra," Ph.D. dissertation, Katholieke Universiteit Leuven, Leuven, Belgium, 1997.
- [12] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, ser. Wiley Series on Adaptive and Learning Systems for Signal Processing, S. Haykin, Ed. New York: Wiley, 2001.
- [13] M. Joho and H. Mathis, "Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation," in *Proc. IEEE Sensor Array Multichannel Signal Processing Workshop (SAM 2002)*, Washington DC, Aug. 2002, pp. 273–277.
- [14] M. Lewicki and T. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, 2000.
- [15] T. Li, "Numerical solution of multivariate polynomial systems by homotopy continuation methods," *Acta Numerica*, no. 6, pp. 399–436, 1997.
- [16] U. Lindgren and A.-J. van der Veen, "Source separation based on second order statistics—An algebraic approach," in *Proc. IEEE SP Workshop Statistical Signal Array Processing*, Corfu, Greece, Jun. 1996, pp. 324–327.
- [17] L. Molgedey and H. Schuster, "Separation of independent signals using time-delayed correlations," *Phys. Rev. Lett.*, vol. 72, no. 23, pp. 3634–3637, Jun. 1994.
- [18] A. Morgan, *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [19] R. Olsson and L. Hansen, "Blind separation of more sources than sensors in convolutive mixtures," presented at the IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 2006), Toulouse, France, May 2006.
- [20] N. D. Sidiropoulos and R. Bro, "On the uniqueness of multilinear decomposition of N -way arrays," *J. Chemometrics*, vol. 14, pp. 229–239, 2000.
- [21] A. Sommese and C. Wampler, *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*. Singapore: World Scientific, 2005.
- [22] L. Tong, R. Liu, V. Soon, and Y. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, no. 5, pp. 499–509, May 1991.
- [23] L. Tong, V. Soon, Y. Huang, and R. Liu, "AMUSE: A new blind identification algorithm," in *Proc. IEEE Int. Symp. Circuits Systems (ISCAS 1990)*, New Orleans, LA, 1990, vol. 3, pp. 1784–1787.
- [24] J. van de Laar, "Instantaneous blind source separation based on the exploitation of temporal correlations and nonstationarity," in *Proc. 14th Annual Workshop Circuits, Systems, Signal Processing (ProRISC 2003)*, Veldhoven, The Netherlands, Nov. 2003, pp. 391–398.
- [25] J. van de Laar, "On MIMO instantaneous blind identification based on the exploitation of the time structure of signals using arbitrary-order cumulants," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP)*, Montreal, QC, Canada, May 2004, pp. 549–552.
- [26] J. van de Laar, "TIME-MUSIC DOA estimation based on the exploitation of some arbitrary-order temporal structure in the data," in *Proc. 3rd IEEE Sensor Array Multichannel Signal Processing Workshop (SAM)*, Barcelona, Spain, Jul. 2004, pp. 308–312.
- [27] J. van de Laar, "MIMO instantaneous blind identification based on second order temporal structure and homotopy method," in *Proc. 9th Int. Workshop Acoustic Echo Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sep. 2005, pp. 241–244.
- [28] S. van Gerven and D. van Compernelle, "On the use of decorrelation in scalar signal separation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing (ICASSP 1994)*, Adelaide, Australia, 1994, pp. 57–60.
- [29] M. Vanderveen, A.-J. van der Veen, and A. Paulraj, "Estimation of multipath parameters in wireless communications," *IEEE Trans. Signal Process.*, vol. 46, no. 3, pp. 682–690, Mar. 1998.

blind source separation.

wireless communications, DSL, and audio signal processing.

all dealing with different basic and advanced signal processing topics. His main field of research is in adaptive array signal processing, with applications in acoustic communication systems.

Dr. Sommen is Editor of EURASIP's *Journal of Applied Signal Processing*, for which he served as Guest Editor of a Special Issue on Signal Processing for Acoustic Communication Systems in 2003. He is a member of the ProRISC board, Vice-President of the IEEE Benelux Signal Processing Chapter, and Officer of the Administrative Committee of EURASIP.

Jakob van de Laar received the M.Sc. degree in electrical engineering (*cum laude*) and the Ph.D. degree from the Eindhoven University of Technology (EUT), Eindhoven, The Netherlands, in 2000 and 2007 respectively.

Since 2005, he has been with Philips Research Laboratories, Eindhoven, The Netherlands, where he currently is employed as a Senior Scientist and is working on advanced signal processing topics. His main interests are in the fields of acoustic telecommunication, adaptive array signal processing, and

Marc Moonen (M'94–SM'06–F'07) received the electrical engineering degree and the Ph.D. degree in applied sciences from Katholieke Universiteit Leuven (K.U. Leuven), Belgium, in 1986 and 1990, respectively.

Since 2004, he has been a Full Professor in the Electrical Engineering Department of Katholieke Universiteit Leuven, where he is currently heading a research team of 16 Ph.D. candidates and post-docs, working in the area of numerical algorithms and signal processing for digital communications,

Dr. Moonen was Chairman of the IEEE Benelux Signal Processing Chapter from 1998 to 2002 and has been an AdCom Member of the European Association for Signal, Speech and Image Processing (EURASIP) since 2000 and a member of the IEEE Signal Processing Society Technical Committee on Signal Processing for Communications. He served as Editor-in-Chief for the *EURASIP Journal on Applied Signal Processing* from 2003 to 2005 and has been a member of the Editorial Board of IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS I from 2002 to 2003 and *IEEE Signal Processing Magazine* from 2003 to 2005. He is currently a member of the Editorial Board of *Integration*, the *VLSI Journal*, the *EURASIP Journal on Applied Signal Processing*, the *EURASIP Journal on Wireless Communications and Networking*, and *Signal Processing*.

Piet C. W. Sommen (SM'03) received the Ingenieur degree in electrical engineering from the Delft University of Technology, Delft, The Netherlands, in 1981 and the Ph.D. degree from Eindhoven University of Technology (EUT), Eindhoven, The Netherlands, in 1992.

From 1981 to 1989, he was with Philips Research Laboratories, Eindhoven, The Netherlands, and since 1989, he has been with the Faculty of Electrical Engineering, EUT, where he is currently an Associate Professor. He is involved in internal and external courses,