A Prewhitening-Induced Bound on the Identification Error in Independent Component Analysis

Lieven De Lathauwer, *Member, IEEE*, Bart De Moor, *Fellow, IEEE*, and Joos Vandewalle, *Fellow, IEEE*

***Abstract*—In this paper, we derive a prewhitening-induced lowerbound on the Frobenius norm of the difference between the true mixing matrix and its estimate in independent component analysis. This bound applies to all algorithms that employ a prewhitening. Our analysis allows one to assess the contribution to the overall error of the partial estimation errors on the components of the singular value decomposition of the mixing matrix. The bound indicates the performance that can theoretically be achieved. It is actually reached for sufficiently high signal-to-noise ratios by good algorithms. This is illustrated by means of a numerical experiment. A small-error analysis allows to express the bound on the average precision in terms of the second-order statistics of the estimator of the signal covariance.**

***Index Terms*—Eigenvalue decomposition (EVD), higher order statistics (HOS), independent component analysis (ICA), principal component analysis.**

I. Introduction

LET us use the following notation for the basic independent component analysis (ICA) model:

$$Y = \mathbf{M}X + N = \tilde{Y} + N \tag{1}$$

in which the observation vector $Y \in \mathbb{C}^J$, the noise vector $N \in \mathbb{C}^J$, and the source vector $X \in \mathbb{C}^I$ are zero-mean stochastic vectors, with $I \leqslant J$. The mixing matrix $\mathbf{M} \in \mathbb{C}^{J \times I}$ is assumed to be full column rank. The vector $\tilde{Y}$ is the signal part of the observations. Signal and noise are uncorrelated. The goal is to exploit the assumed mutual statistical independence of the source components to estimate the mixing matrix and/or the source signals from realizations of Y.

Many ICA algorithms are prewhitening based. For instance, the algebraic algorithms presented in [3], [6], [9], [14] belong to this class. An eigenvalue decomposition (EVD) of the observed covariance, or a singular-value decomposition (SVD) of the data matrix, allows one to estimate the number of sources and to decorrelate them. The remaining rotational degrees of freedom are fixed by resorting to the higher order statistics (HOS) of the observations. (An alternative is, e.g., to exploit the structure of the spatial covariance matrices for different time lags if the sources are temporally correlated [3]. Our paper applies to all algorithms consisting of a prewhitening followed by the determination of a unitary matrix.) Because higher order cumulants are asymptotically insensitive to additive Gaussian noise [24], the prewhitening step has the disadvantage w.r.t. the higher order step that its partial results are directly affected by this noise. The error introduced at this stage may not be compensated by the higher order step (this will be further explained later; see also [7]). It may actually introduce an upper-bound to the performance of the ICA algorithm. This observation has led to the development of higher order only ICA procedures [4], [5], [10], [11] and soft whitening techniques [17], [26], [28], [30]. Of course, in many ICA algorithms the second- and higher order statistical information is combined in a more implicit way [1], [2].

If one wishes to evaluate the quality of the outcome of an ICA, in fact two points of view are possible. One may wonder how well the estimated sources have been separated. Natural criteria for this *separation* quality are the signal-to-interference ratio (SIR), or the signal-to-interference-plus-noise ratio (SINR). On the other hand, in several applications, the goal is the estimation of the mixing matrix, rather than the separation of the sources. Here, it would be natural to evaluate the *identification* accuracy in terms of the Frobenius norm of the difference between the estimated and the true mixing matrix. The latter aspect is omitted in most papers introducing new ICA algorithms, because of the intuitive link between the quality of separation and identification. Nevertheless, it is clear that it is unnatural to quantify the identification accuracy in terms of SIR or SINR. Moreover, large variations of the separation index may correspond to small variations of the identification index, and vice-versa. (This will be illustrated in Section V.) Hence, it is preferable to analyze each aspect in its proper way. The distinction between the two viewpoints is more established in the processing of convolutive mixtures/channels, where it is reflected by the terminology: blind identification versus blind deconvolution. The goal of blind identification is the estimation of the channel coefficients, while the goal of blind deconvolution is the estimation of the inputs.

The identification problem is important in, for instance, wireless communications, where directions of arrival (DOA) may be computed from the estimated mixing matrix [22]. In seismology and geophysics, the mixing coefficients may be related to physical parameters one wishes to estimate [8], [23]. A third application is the extraction of the fetal electrocardiogram from cutaneous potential recordings [15], where the mixing

Manuscript received October 21, 2003; revised July 29, 2004. This work is supported in part by the Research Council K.U.Leuven under Grant GOA-Mefisto-666 and Grant GOA-AMBioRICS, in part by the Flemish Government under F.W.O. Project G.0240.99, F.W.O. research communities ICCoS, ANMMM, and Tournesol Project T2004.13, and in part by the Belgian Federal Science Policy Office under IUAP P5/22. This paper was recommended by Associate Editor W. X. Zheng.

L. De Lathauwer is with the Research Group ETIS, UMR 8051, Centre National de la Recherche Scientifique (C.N.R.S.), F 95014 Cergy-Pontoise Cedex, France (e-mail: delathau@ensea.fr).

B. De Moor and J. Vandewalle are with the Group SCD-SISTA , E.E. Dept. (ESAT), Katholieke Universiteit Leuven, B-3001 Leuven (Heverlee), Belgium (e-mail: demoor@esat.kuleuven.ac.be; vdwalle@esat.kuleuven.ac.be).

Digital Object Identifier 10.1109/TCSI.2004.843061

vectors indicate how strongly the different electrodes capture each source signal. From this information, better measurement positions might be deduced. We mention that the positioning of the electrodes is one of the most crucial factors for the success of the method. A fourth example comes from operating response analysis in vibro-acoustics [25]. Here, the animation of "independent operating field shapes" (IOFS) may show the contribution of the distinct mutually statistically independent contributions to the vibration problem [12]. The IOFS are a graphical representation, as a function of frequency, of the mixing vectors.

As far as the prewhitening-induced performance bound is concerned, [7] focuses on the quality of separation and an appropriate performance bound in terms of the intersymbol interference is derived. Our paper is the "identification" counterpart of [7]. However, the methodology to derive an appropriate bound is completely different.

In the next section, we will have a closer look at the concept of prewhitening, which will allow us to describe the goals of this paper in some more detail. Section III contains the core result of this paper. In this section, we will derive a deterministic prewhitening-induced lowerbound on the Frobenius norm of the difference between the true mixing matrix and its estimate. In Section IV we will interpret this result in a statistical context and conduct a small-error analysis. The derivations are based on the perturbation analysis theorems given in the Appendix . To our knowledge, the second-order perturbation analysis results are new. Section V illustrates our study by means of some simulations.

A. Notation

Scalars are denoted by lower case letters $(a, b, \ldots; \alpha, \beta, \ldots)$, vectors are written as capitals $(A, B, \ldots)$ (italic shaped) and matrices correspond to bold-face capitals $(\mathbf{A}, \mathbf{B}, \ldots)$. This notation is consistently used for lower-order parts of a given structure. For example, the entry with row index i and column index j in a matrix $\mathbf{A}$, i.e., $(\mathbf{A})_{ij}$, is symbolized by a_{ij}. We have made one exception to this rule: as we frequently use the characters i and j in the meaning of indices, I and J are reserved to denote the index upper bounds. The ith column vector of a matrix $\mathbf{A}$ is denoted as A_i, i.e., $\mathbf{A} = [A_1 A_2 \ldots]$. By $\cdot^T$ we denote the transpose and by $\cdot^H$ the complex conjugated transpose. A_i^H is the complex conjugated transpose of the ith column of $\mathbf{A}$ (and not the ith column of $\mathbf{A}^H$). $\cdot^\dagger$ denotes the Moore–Penrose pseudo-inverse. $\mathtt{Re}(\cdot)$ is the real part of a complex number. $\langle \mathbf{A}, \mathbf{B} \rangle = \sum_{ij} a_{ij}^* b_{ij}$ is the inner product of matrices $\mathbf{A}$ and $\mathbf{B}$. $E\{\cdot\}$ denotes the statistical expectation and $\mathrm{var}\{\cdot\}$ the variance.

II. Prewhitening-Based ICA

Let us first briefly repeat the general scheme of a prewhitening-based ICA, thereby introducing some notations.

Let us write the covariance matrix of a random vector Z as $\mathbf{C}_Z$. Then we have from (1) the following relationship between the covariance matrices of the signals under consideration

$$\mathbf{C}_Y = \mathbf{M} \cdot \mathbf{C}_X \cdot \mathbf{M}^H + \mathbf{C}_N = \mathbf{C}_{\tilde{Y}} + \mathbf{C}_N. \tag{2}$$

It is well known that in ICA the columns of the mixing matrix can only be found up to a scaling factor and permutation. We assume that $\mathbf{M}$ corresponds to unit-variance sources. We have then

$$\mathbf{C}_Y = \mathbf{M} \cdot \mathbf{M}^H + \mathbf{C}_N. \tag{3}$$

From this equation, it is clear that the mixing matrix may be found, up to a multiplicative unitary factor, as a square root of the covariance of the signal part of the observations. The most common way to determine such a square root is by computation of the EVD of $\mathbf{C}_{\tilde{Y}}$. Let the SVD of the mixing matrix be given by

$$\mathbf{M} = \mathbf{E}\mathbf{D}\mathbf{Q}^H \tag{4}$$

in which $\mathbf{E} \in \mathbb{C}^{J \times I}$ has mutually orthonormal columns, $\mathbf{D} \in \mathbb{R}^{I \times I}$ is positive diagonal, and $\mathbf{Q} \in \mathbb{C}^{I \times I}$ is unitary. Then, we have

$$\mathbf{C}_{\tilde{Y}} = \mathbf{E}\mathbf{D}^2\mathbf{E}^H \tag{5}$$

from which $\mathbf{E}$ and $\mathbf{D}$ may be found. The unitary factor $\mathbf{Q}$ cannot be determined from the second-order statistics of the data. Hence, we have to resort to their HOS (assuming only that the sources are independent and that at most one of them is Gaussian). However, the higher order equivalent of (2), expressed in terms of the unknown $\mathbf{Q}$, is overdetermined. This allows to distinguish between different solution strategies, depending on how $\mathbf{Q}$ is estimated from the HOS of the data (see [14], [20] and the references therein).

In practice, we have to work with an estimate $\hat{\mathbf{C}}_{\tilde{Y}}$ of the covariance of $\tilde{Y}$. Let us for instance consider a more-sensors-than-sources setup subject to white noise. Call $\hat{\sigma}_N^2$ an estimate of the noise variance on each data channel. This estimate can be obtained as the mean of the $I - J$ smallest eigenvalues of $\hat{\mathbf{C}}_Y$. Then $\hat{\mathbf{C}}_{\tilde{Y}}$ can be obtained from the sample observations covariance $\hat{\mathbf{C}}_Y$ by subtracting $\hat{\sigma}_N^2$ from the J largest eigenvalues and setting the $I - J$ smallest eigenvalues equal to zero. If no estimate of the noise covariance is available, the estimate $\hat{\mathbf{C}}_{\tilde{Y}}$ is taken equal to $\hat{\mathbf{C}}_Y$ itself.

One then considers the EVD

$$\hat{\mathbf{C}}_{\tilde{Y}} = \hat{\mathbf{E}}\hat{\mathbf{D}}^2\hat{\mathbf{E}}^H \tag{6}$$

in which $\hat{\mathbf{E}} \in \mathbb{C}^{J \times I}$, having mutually orthonormal columns, is an estimate of $\mathbf{E}$ and the positive diagonal matrix $\hat{\mathbf{D}} \in \mathbb{R}^{I \times I}$ is an estimate of $\mathbf{D}$. (Note that we assume that the number of sources is estimated correctly. If this is not the case, the mixing matrix and its estimate have different dimensions, such that the Frobenius norm of their difference is not defined. Quantification of how close matrices of different dimensions are, and evaluation of the obtained accuracy, in the case of an incorrect estimation of I, is outside the scope of this paper.) An estimate $\hat{\mathbf{M}}$ of the mixing matrix is subsequently obtained as

$$\hat{\mathbf{M}} = \hat{\mathbf{E}}\hat{\mathbf{D}}\hat{\mathbf{Q}}^H \tag{7}$$

in which the unitary matrix $\hat{\mathbf{Q}} \in \mathbb{C}^{I \times I}$ is an estimate of $\mathbf{Q}$, obtained from the sample HOS of the observations.

The goal of this paper is to explain that errors in the estimation of $\mathbf{E}$ and $\mathbf{D}$, due to an imperfect estimation of $\mathbf{C}_{\tilde{Y}}$, may imply a bound on the overall accuracy with which the mixing matrix can be estimated. Our error measure is the Frobenius norm $\|\mathbf{M}-\hat{\mathbf{M}}\|$, in which the columns of both matrices are normalized such that they correspond to unit-variance sources. This particular normalization convention has the advantage that the squared Frobenius norm of a mixing vector reflects the "energy" of the corresponding independent component in the dataset. We will explain in which way the estimation errors on $\mathbf{E}$ and $\mathbf{D}$ contribute to the overall bound. Instead of deriving $\hat{\mathbf{M}}$ for one particular algorithm and looking how well it compares to $\mathbf{M}$, we will address the problem in an algorithm-independent way. This means that we will derive a bound on $\|\mathbf{M}-\hat{\mathbf{M}}\|$, regardless of how well $\hat{\mathbf{Q}}$ is chosen. (Of course, the bound depends on the values taken by $\hat{\mathbf{E}}$ and $\hat{\mathbf{D}}$, but the prewhitening is not really specific for a particular ICA algorithm. As indicated above, algorithms are considered different when a different approach is followed for the estimation of $\mathbf{Q}$.) This bound is the ultimate performance that can be obtained by an ICA algorithm, given a certain prewhitening. The actual performance of a particular algorithm can then be assessed by comparing its results with this ultimate bound. So the results of this paper can be used to analyze the performance of (the $\mathbf{Q}$ step of) a prewhitening-based ICA algorithm. Also, the theorem allows one to judge whether the prewhitening step is too critical in a typical problem setup. Based on this knowledge, one may choose to use a higher order only algorithm or a method in which second- and higher order information are exploited in a more balanced way. The results can also be used to verify whether a nonprewhitening based algorithm indeed goes beyond the bound.

III. Bound on Identification Error

First, we mention the following lemma.

Lemma 1: Let the SVDs of the matrices $\mathbf{A} \in \mathbb{C}^{I\times J}$, $\mathbf{B} \in \mathbb{C}^{J\times I}$ and product $\mathbf{AB}$, with $I \leqslant J$, be given by

$$\mathbf{A} = \mathbf{U_A S_A V_A}^H \quad (8)$$

$$\mathbf{B} = \mathbf{U_B S_B V_B}^H \quad (9)$$

$$\mathbf{AB} = \mathbf{U_{AB} S_{AB} V_{AB}}^H. \quad (10)$$

Let the respective singular values be given by $\sigma_i(\mathbf{A}) \geqslant 0$, $\sigma_i(\mathbf{B}) \geqslant 0$ and $\sigma_i(\mathbf{AB}) \geqslant 0$ $(1 \leqslant i \leqslant I)$. Then, we have

$$\sum_i \sigma_i(\mathbf{AB}) \leqslant \sum_i \sigma_i(\mathbf{A})\sigma_i(\mathbf{B}). \quad (11)$$

The equality sign holds iff $\mathbf{V_A}^H\mathbf{U_B}$ is a diagonal matrix containing only unit-norm scalars, in the case that all the singular values of $\mathbf{A}$, and all the singular values of $\mathbf{B}$, are mutually different. In the case that singular values of $\mathbf{A}$ (or $\mathbf{B}$) are equal, the equality sign still holds if $\mathbf{V_A}^H\mathbf{U_B}$ is equal to a diagonal matrix that contains only unit-norm scalars up to unitary transformation of the corresponding rows (columns).

Proof: [19, pp. 176–177]. ■

Using the notation introduced in the previous section, we now have the following theorem.

Theorem 2: The quality of the mixing matrix estimate is bounded by the quality of the prewhitening in the following way:

$$\|\mathbf{M}-\hat{\mathbf{M}}\|^2 \geqslant \sum_i (d_{ii}^2 + \hat{d}_{ii}^2 - 2s_{ii}) \quad (12)$$

$$\geqslant \sum_i (d_{ii} - \hat{d}_{ii})^2 \quad (13)$$

$$\geqslant 0 \quad (14)$$

in which s_{ii} is the ith singular value of $\hat{\mathbf{C}}_{\tilde{Y}}^{H/2} \cdot \mathbf{C}_{\tilde{Y}}^{1/2}$, involving arbitrary square roots of $\mathbf{C}_{\tilde{Y}}$ and $\hat{\mathbf{C}}_{\tilde{Y}}$. The right-hand side of (12) defines the minimal error given a prewhitening based on the estimate $\hat{\mathbf{C}}_{\tilde{Y}}$. The difference between the left-hand side and the right-hand side of (12) depends entirely on the choice of the unitary factor $\hat{\mathbf{Q}}$. There always exists a unitary matrix $\hat{\mathbf{Q}}$ for which the error exactly reduces to the expression on the right-hand side of (12). This bound is linked to the errors on the estimates $\hat{\mathbf{E}}$ and $\hat{\mathbf{D}}$ in the following way. The second inequality vanishes iff the eigenvectors are correctly estimated. The third inequality vanishes if the eigenvalues of $\mathbf{C}_{\tilde{Y}}$ are correctly estimated.

It has to be specified what is meant by a "correct estimation of the eigenvectors" in the previous paragraph. If the eigenvalues of $\mathbf{C}_{\tilde{Y}}$ are mutually different and so are all the eigenvalues of $\hat{\mathbf{C}}_{\tilde{Y}}$, then a correct estimation of the eigenvectors obviously means that the estimated and the true eigenvectors are equal up to multiplication by a unit-modulus scalar. If some eigenvalues of $\mathbf{C}_{\tilde{Y}}$ are equal and/or if some eigenvalues of $\hat{\mathbf{C}}_{\tilde{Y}}$ are equal, then there is an indeterminacy in some eigenspace(s). In this case, we call the estimation correct when $\hat{\mathbf{E}}\mathbf{W_1} = \mathbf{E}\mathbf{W_2}$, in which $\mathbf{W_1}$ and $\mathbf{W_2}$ are block-diagonal matrices containing unitary blocks on the positions where the corresponding eigenvalues are equal.

Proof: The minimization of $\|\mathbf{M}-\hat{\mathbf{M}}\|^2$ in terms of $\hat{\mathbf{Q}}$ is a unitary Procrustes problem ([18, p. 582]). We have

$$\|\mathbf{M}-\hat{\mathbf{M}}\|^2 = \|\mathbf{M}\|^2 + \|\hat{\mathbf{M}}\|^2 - 2\mathrm{Re}(\langle \mathbf{M}, \hat{\mathbf{M}}\rangle) \quad (15)$$

$$= \|\mathbf{D}\|^2 + \|\hat{\mathbf{D}}\|^2 - 2\mathrm{Re}(\langle \hat{\mathbf{D}}\hat{\mathbf{E}}^H\mathbf{E}\mathbf{D}, \tilde{\mathbf{Q}}\rangle) \quad (16)$$

in which $\tilde{\mathbf{Q}} \stackrel{\text{def}}{=} \hat{\mathbf{Q}}^H\mathbf{Q}$. If the SVD of $\hat{\mathbf{D}}\hat{\mathbf{E}}^H\mathbf{E}\mathbf{D}$ is given by $\mathbf{USV}^H$, then the optimal $\tilde{\mathbf{Q}}$ takes the form of $\mathbf{UV}^H$. Since $\mathbf{C}_{\tilde{Y}} = \mathbf{MM}^H$ and $\hat{\mathbf{C}}_{\tilde{Y}} = \hat{\mathbf{M}}\hat{\mathbf{M}}^H$, $\mathbf{ED}$ and $\hat{\mathbf{E}}\hat{\mathbf{D}}$ are square roots of $\mathbf{C}_{\tilde{Y}}$ and $\hat{\mathbf{C}}_{\tilde{Y}}$, respectively. Multiplying $\mathbf{ED}$ and/or $\hat{\mathbf{E}}\hat{\mathbf{D}}$ from the right by a unitary factor just leads to a unitary transformation of the optimal $\tilde{\mathbf{Q}}$ but does not change the bound. This proves the first inequality.

The second inequality is equivalent to

$$\sum_i s_{ii} \leqslant \sum_i d_{ii}\hat{d}_{ii}. \quad (17)$$

The latter inequality follows from Lemma 1, in which $\mathbf{A} = \hat{\mathbf{D}}\hat{\mathbf{E}}^H$ and $\mathbf{B} = \mathbf{ED}$. ■

This theorem tells us that, no matter how accurate the higher order step of an ICA algorithm is, one can never do better than stated by (12). Consequently, if one wishes to assess the accuracy of the (higher order step of the) ICA procedure (given the subresults obtained by the prewhitening), then one should not simply evaluate how large the Frobenius norm of the difference between the true mixing matrix and its estimate is (taking a zero error as reference) but examine how close the Frobenius norm is to the bound specified in (12).

The theorem further allows one to assess to what extent the bound is caused by inaccuracies in the estimation of the eigenvectors of $\mathbf{C}_{\tilde{Y}}$, or by inaccuracies in the estimation of its eigenvalues. If the eigenvectors are exactly estimated, then the bound reduces to the lower value specified in (13). So the difference between bounds (12) and (13) is due to errors in the estimation of the eigenvectors, given the estimates of the eigenvalues. Equation (13) by itself shows the part of the overall bound that is due to errors in the estimation of the eigenvalues. If one starts from perfect estimates of both the eigenvectors and the eigenvalues, then of course it is theoretically possible to find a perfect estimate of the mixing matrix. This is made explicit in (14). When the additive noise is spatially white and $I = J$, the estimated eigenvectors asymptotically (for the number of samples going to infinity) approach the true eigenvectors and only the eigenvalue estimates are biased. In this case, the error bound approaches (13).

Note that bound (12) is sharp by definition. The proof demonstrates that there always exists a unitary matrix such that the bound is exactly reached. Also, if the eigenvectors of $\mathbf{C}_{\tilde{Y}}$ are exactly known, then there always exists a unitary matrix such that bound (13) is reached. It is trivial to say that, when both the eigenvectors and the eigenvalues of $\mathbf{C}_{\tilde{Y}}$ are known, exact estimation of the mixing matrix is theoretically possible.

It is well known that eigenvectors corresponding to eigenvalues that are close, are ill conditioned [21]. In other words, principal components associated to eigenvalues that are close, are hard to estimate. Nevertheless, this does not pose a major problem for the global ICA. The reason is that, if the eigenvectors are only known up to some rotation, in this case, the rotation can be absorbed by $\mathbf{Q}$ and the error can more or less be compensated by the higher order step. This fact is reflected by the theorem. Indeed, if eigenvalues of $\mathbf{C}_{\tilde{Y}}$ are the same and the corresponding eigenspace is accurately estimated, then the corresponding singular values of $\hat{\mathbf{C}}_{\tilde{Y}}^{H/2} \cdot \mathbf{C}_{\tilde{Y}}^{1/2}$ are given by $d_{ii}\hat{d}_{ii}$. The corresponding part of (12) then reduces to the form of (13).

The philosophy behind Theorem 2 is different from the one behind a Cramer–Rao type bound. A Cramer–Rao bound allows one to assess the optimal average performance of an unbiased estimator. In contrast, Theorem 2 not only applies on the average, but also to each individual ICA run. Moreover, it is not algorithm-specific.

IV. Statistical Considerations and Small-Error Analysis

Because (12)–(14) apply to each individual ICA run, they also hold on the average. We obtain the following statistical bounds:

$$E\{\|\mathbf{M}-\hat{\mathbf{M}}\|^2\} \geqslant \sum_i E\{(d_{ii}^2 + \hat{d}_{ii}^2 - 2s_{ii})\} \quad (18)$$

$$\geqslant \sum_i E\{(d_{ii} - \hat{d}_{ii})^2\} \quad (19)$$

$$\geqslant 0 \quad (20)$$

in which one averages over all runs.

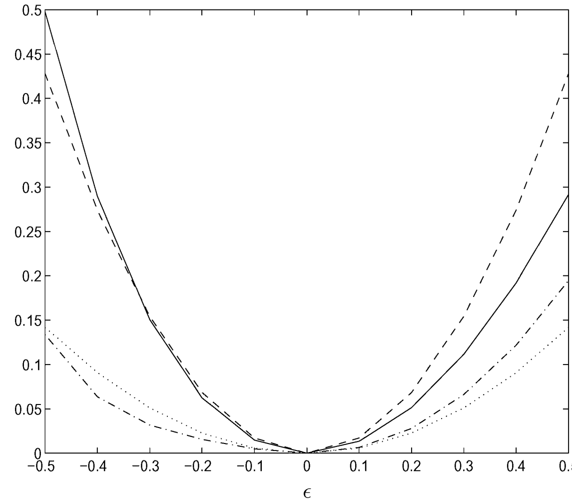

Fig. 1. Bounds on $\|\mathbf{M} - \hat{\mathbf{M}}\|^2$ in Example 1. Exact curves given by (12) (solid) and (13) (dash-dotted), and small-error approximations, following from (21) (dashed) and (22) (dotted).

More explicit relations between the average identification accuracy and the statistics of the estimator $\hat{\mathbf{C}}_{\tilde{Y}}$ of $\mathbf{C}_{\tilde{Y}}$ can be derived when we assume that the error on the estimated covariance is small. For convenience, we assume that $\mathbf{C}_{\tilde{Y}}$ is full rank ($I = J$) and that all its eigenvalues are distinct. For small errors, we have the following theorem.

Theorem 3: Let $\hat{\mathbf{C}}_{\tilde{Y}} = \mathbf{C}_{\tilde{Y}} + \epsilon\, \mathbf{C}$, in which $\mathbf{C}$ characterizes the estimation error. In first order, the Frobenius norm of the difference between the true and the estimated mixing matrix is bounded as follows:

$$\|\mathbf{M}-\hat{\mathbf{M}}\| \geqslant \frac{|\epsilon|}{\sqrt{2}} \left(\sum_{ij} \frac{|\tilde{c}_{ij}|^2}{d_{ii}^2 + d_{jj}^2} \right)^{1/2} \quad (21)$$

$$\geqslant \frac{|\epsilon|}{2} \left(\sum_i \frac{\tilde{c}_{ii}^2}{d_{ii}^2} \right)^{1/2} \quad (22)$$

$$\geqslant 0 \quad (23)$$

in which $\tilde{\mathbf{C}} = \mathbf{E}^H \cdot \mathbf{C} \cdot \mathbf{E}$.

The proof is given in the Appendix .

Example 1: Consider $\hat{\mathbf{C}}_{\tilde{Y}}(\epsilon) = \mathbf{C}_{\tilde{Y}} + \epsilon\, \mathbf{C}$, with

$$\mathbf{C}_{\tilde{Y}} = \begin{pmatrix} 3 & -i \\ i & 3 \end{pmatrix} \quad \mathbf{C} = \begin{pmatrix} 3.6 & 0.4 - 1.6i \\ 0.4 + 1.6i & -0.8 \end{pmatrix}$$

for ϵ varying between -0.5 and 0.5. In Fig. 1 we have plotted the exact bounds (12) and (13) on $\|\mathbf{M} - \hat{\mathbf{M}}\|^2$, together with their small-error approximations, following from (21) and (22).

Hence, for small errors, the expected identification accuracy, for a finite dataset generated in accordance with (1) can be bounded by a function of $E\{|\tilde{c}_{ij}|^2\}$. In other words, the bound can be written as a function of the second-order statistics of the estimator $\hat{\mathbf{C}}_{\tilde{Y}}$. If this estimator is unbiased (due to noise compensation, as discussed in Section II), we obtain

$$E\{\|\mathbf{M}-\hat{\mathbf{M}}\|^2\} \geqslant \frac{1}{2} \sum_{ij} \frac{\mathrm{var}\{(\mathbf{E}^H \hat{\mathbf{C}}_{\tilde{Y}} \mathbf{E})_{ij}\}}{d_{ii}^2 + d_{jj}^2} \quad (24)$$

$$\geqslant \frac{1}{4} \sum_i \frac{\mathrm{var}\{(\mathbf{E}^H \hat{\mathbf{C}}_{\tilde{Y}} \mathbf{E})_{ii}\}}{d_{ii}^2} \quad (25)$$

$$\geqslant 0. \quad (26)$$

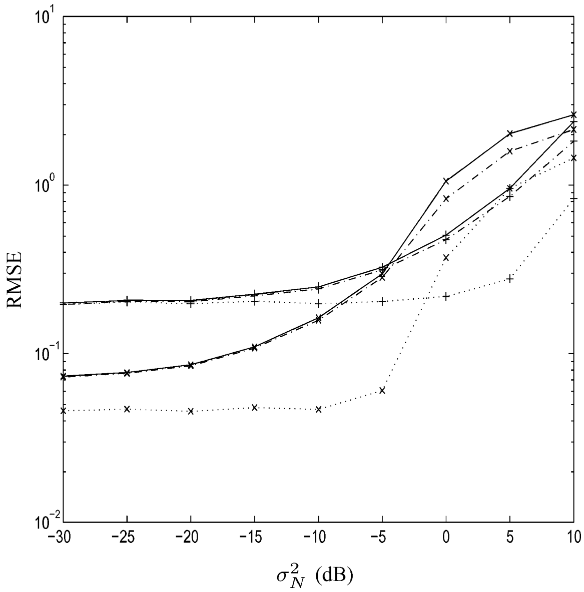

Fig. 2. RMSE between the true mixing matrix and its estimate. Effect of the SNR on the quality of the reconstruction. $\phi_1 = 0$. Solid: the achieved performance. Dash-dotted: error lowerbound (18). Dotted: error lowerbound (19). "x"-curves: $\phi_2 = 0.02$. "+"-curves: $\phi_2 = 0.1$ (in this case the mixing vectors are mutually orthogonal).

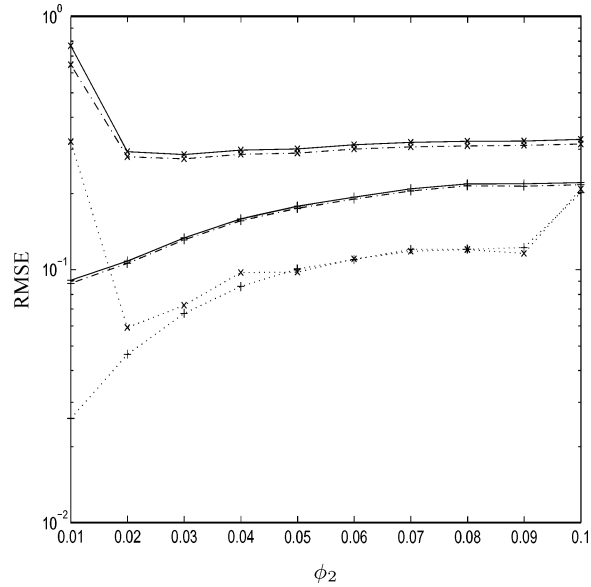

Fig. 3. RMSE between the true mixing matrix and its estimate. Effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. Solid: the achieved performance. Dash-dotted: error lowerbound (18). Dotted: error lowerbound (19). "x"-curves: $\mathrm{SNR} = 5$ dB. "+"-curves: $\mathrm{SNR} = 15$ dB.

V. Simulations

Usually a text introducing an ICA algorithm, or proposing ICA as the solution to a typical problem, illustrates the performance of the technique by means of some simulations in which the true independent components are known. Our results can help in the evaluation of the method and the interpretation of the data. In this section we illustrate our results by means of a numerical experiment. The emphasis is on explaining the principles of this paper, rather than on investigating how well different ICA algorithms approach the bound in different scenarios. We start from the experimental setup in the simulations of [6], which is particularly instructive.

We consider two zero-mean complex-valued source signals, uniformly distributed over the unit circle. Both signals impinge on a linear $\lambda/2$ equispaced array of 10 unit-gain omnidirectional sensors in the far field of the emitters. Under this assumption, the theoretical values of the elements of the mixing matrix are given by $m_{pq} = \sigma_q e^{2j\pi p \phi_q}$, where ϕ_q equals the electrical angle of source q. The noise is Gaussian, with power σ_N^2. We set the data length $T = 100$ and the angle $\phi_1 = 0$.

For the ICA, we used the efficient algorithm described in [6], which is known to be asymptotically equivalent to the methods derived in [9] and [16]. In Figs. 2 and 3 we plot the root mean-square error (RMSE), $\sqrt{E\{\|\mathbf{M} - \hat{\mathbf{M}}\|^2\}}$, between the true mixing matrix and its estimate. We assume that the columns of the estimate are optimally ordered. The dash-dotted and dotted lines give the two error lowerbounds (18) and (19). All curves are obtained by averaging over 500 Monte Carlo simulations. The variance of the displayed results is small: the worst value of the variance divided by the squared mean is for each curve in the order of magnitude of $1e-4$ to $1e-3$.

Note that for moderate to high signal-to-noise ratios (SNRs), bound (18) is actually reached. This means that, in this region, algorithm [6] performs as well as can be expected. Additional improvement is not possible, given the errors introduced by the prewhitening. Furthermore, Theorem 2 can be used in the following sense. The curves corresponding to $\phi_2 = 0.02$ in Fig. 2 show that here some improvement is still possible for SNRs lower than 5 dB. When comparing another prewhitening-based algorithm to the algorithm of [6], its quality should now be assessed in how closely it approaches the dash-dotted line. If in some way one has prior knowledge of the eigenvectors of $\mathbf{C}^{\tilde{Y}}$, then the ultimate performance becomes that indicated by the dotted line. In other words, the difference between the dotted and the dash-dotted lines shows how much is lost by not knowing the eigenvectors of $\mathbf{C}_{\tilde{Y}}$ in advance. The theorem thus allows to attribute the bound on the achievable performance to contributions by the two aspects of the prewhitening stage. Namely, the difference between the dash-dotted and the dotted lines is due to a misfit of the eigenvectors (given a certain estimate of the eigenvalues), and the value indicated by the dotted lines is due to the inaccuracy of the eigenvalue estimates.

Let us compare the analysis of the performance in terms of the RMSE to an analysis of the performance in terms of the interference-to-signal ratio (ISR) and the interference plus noise-to-signal ratio (INSR) of the source estimates $\hat{X}$. These estimates are obtained from the observations Y by premultiplication with a matrix $\mathbf{W}^H$, following some beamforming strategy [29]

$$\hat{X} = \mathbf{W}^H Y. \tag{27}$$

The average ISR and INSR are then defined as follows:

$$\mathrm{ISR}_{ij} \stackrel{\mathrm{def}}{=} E\left\{ \frac{\sigma_{x_j}^2 |W_i^H M_j|^2}{\sigma_{x_i}^2 |W_i^H M_i|^2} \right\} \tag{28}$$

$$\mathrm{INSR}_i \stackrel{\mathrm{def}}{=} E\left\{ \frac{W_i^H (\mathbf{C}_Y - \sigma_{x_i}^2 M_i M_i^H) W_i}{\sigma_{x_i}^2 |W_i^H M_i|^2} \right\} \tag{29}$$

in which $\sigma_{x_i}^2$ is the variance of the ith source.

The INSR is minimized by a Minimum Variance Distortionless Response (MVDR) filter, given by

$$\mathbf{W} = \mathbf{C}_Y^{\dagger} \mathbf{M}. \tag{30}$$

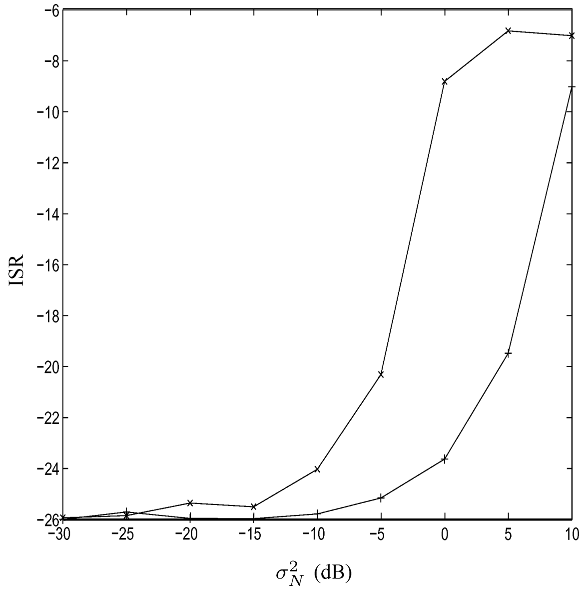

Fig. 4. Mean ISR of the LCMV-beamformer; effect of the SNR on the quality of separation. $\phi_1 = 0$. “x”-curves: $\phi_2 = 0.02$. “+”-curves: $\phi_2 = 0.1$.

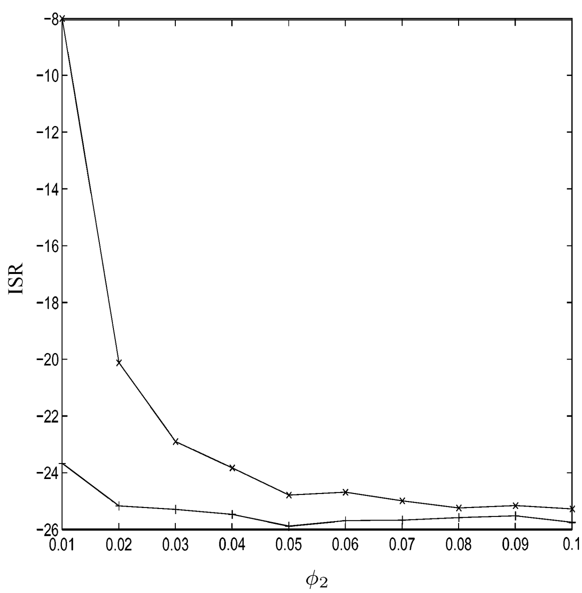

Fig. 5. Mean ISR of the LCMV-beamformer; effect of the difference in DOA ($\phi_1 = 0$) on the quality of separation. “x”-curves: $\mathrm{SNR} = 5$ dB. “+”-curves: $\mathrm{SNR} = 15$ dB.

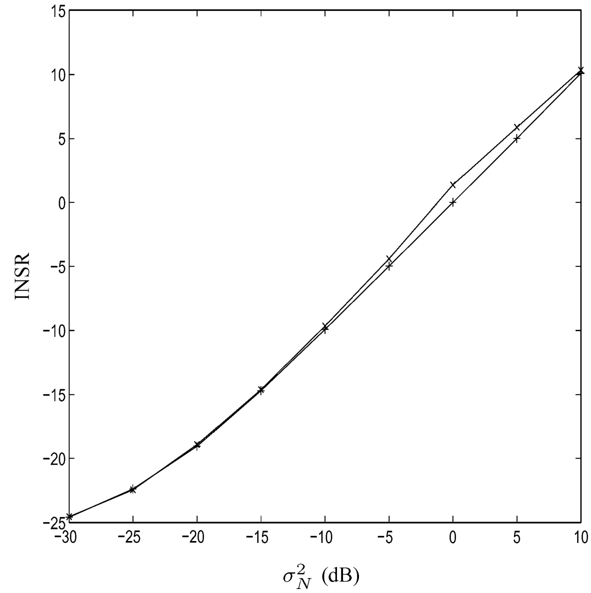

Fig. 6. Mean INSR of the MVDR-beamformer; effect of the SNR on the quality of the reconstruction. $\phi_1 = 0$. “x”-curves: $\phi_2 = 0.02$. “+”-curves: $\phi_2 = 0.1$.

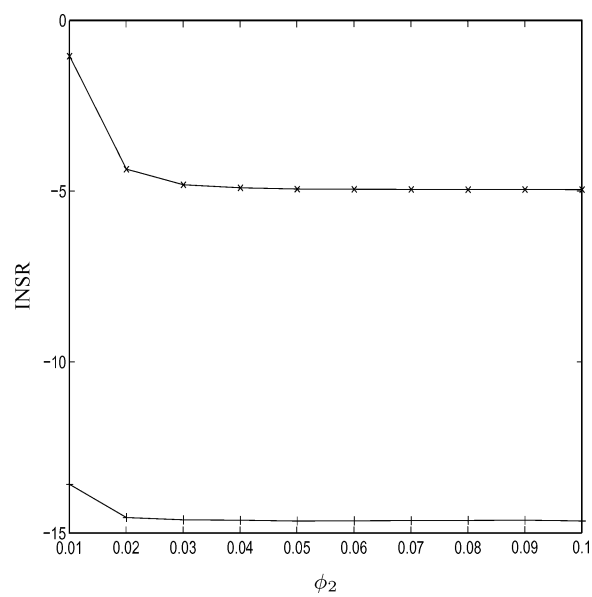

Fig. 7. Mean INSR of the MVDR-beamformer; effect of the difference in DOA ($\phi_1 = 0$) on the quality of the reconstruction. “x”-curves: $\mathrm{SNR} = 5$ dB. “+”-curves: $\mathrm{SNR} = 15$ dB.

On the other hand, the mutual interference of the sources can be cancelled by implementing a linear constrained minimum variance (LCMV) filter

$$\mathbf{W} = \mathbf{C}_{\tilde{Y}}^{\dagger}\mathbf{M}. \tag{31}$$

Of course, in the simulations, these filters have been approximated using sample statistics and the estimate of the mixing matrix.

A comparison of Figs. 2 and 3 to Figs. 4–7 shows the following. Although the general picture is more or less the same, there are some important differences. The INSR is monotonically increasing over the full SNR interval that we considered. On the other hand, the RMSE is almost constant over a substantial part of this interval. In Fig. 3 the RMSE is slowly increasing for values of ϕ_2 ranging from 0.02 to 0.1, while we see in Figs. 5 and 7 that the ISR is actually decreasing and the INSR is constant. For high SNR values, the estimation of the mixing matrix was more accurate when the parameter ϕ_2 was set equal to 0.02, rather than 0.1. In terms of the ISR, the opposite holds true. We conclude that the accuracy with which the mixing matrix has been estimated, should be examined using proper criteria, such as the Frobenius norm proposed in this paper. Inferring conclusions on the identification accuracy from the ISR and INSR curves should be avoided. The theorems of this paper can be used to see the performance that is actually achieved in the perspective of what is theoretically possible.

VI. Conclusion

Errors introduced in the prewhitening step of ICA algorithms cannot be compensated by the higher order step. In this paper, we have derived the theoretical bound on the accuracy with which the mixing matrix can be estimated. Our error measure is the Frobenius norm of the difference between the true mixing matrix and its estimate, which is natural if one is interested in the identification accuracy of the ICA algorithm. Our analysis allows one to assess the contribution to the overall error of the errors occurring in the estimation of the eigenmatrix of the covariance of the signal part of the observations, in the estimation of its eigenvalues and in the estimation of the unitary factor from the HOS of the data. The analysis was carried out in a way that is conceptually independent of the specific mechanisms of particular ICA procedures. We have performed a small-error analysis of the bound. A statistical interpretation of the result showed in which way the bound on the average performance is related to the autocorrelation of the estimator of the signal covariance matrix.

Appendix
Perturbation Analysis

Theorem 3:

Proof: Equations (21)–(23) are the result of a perturbation analysis of (12)–(14). The eigenvalues $\hat{d}_{ii}^2$ follow from Theorem 5, the square root $\hat{\mathbf{C}}_{\tilde{Y}}^{1/2}$ follows from Theorem 4 and the singular values s_{ii} follow from Theorem 6. These theorems are given further in this Appendix.

Let us first derive (21) from (12). For $\mathbf{C}_{\tilde{Y}}^{1/2}$ and $\hat{\mathbf{C}}_{\tilde{Y}}^{1/2}$, we take the Hermitian square root of $\mathbf{C}_{\tilde{Y}}$ and $\hat{\mathbf{C}}_{\tilde{Y}}$, respectively. Equation (12) can be expanded as

$$\|\mathbf{M}-\hat{\mathbf{M}}\|^2 \geqslant \mathrm{trace}\left(\mathbf{D}^2\right) + \mathrm{trace}\left(\mathbf{D}^2+\epsilon\,\mathbf{E}^H\mathbf{C}\mathbf{E}\right) + -2\,\mathrm{trace}\Big[\mathbf{D}^2+\frac{\epsilon}{2}\mathbf{E}^H\mathbf{C}\mathbf{E}+\frac{\epsilon^2}{2}\Big(\mathbf{E}^H\mathbf{H_1}\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{E}\boldsymbol{\Phi} + \mathbf{E}^H\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{H_1}\mathbf{E}\boldsymbol{\Psi}+\mathbf{E}^H\mathbf{H_2}\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{E} + \mathbf{E}^H\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{H_2}\mathbf{E}\Big)+O(\epsilon^3)\Big] \tag{32}$$

in which $\mathbf{H_1}$ and $\mathbf{H_2}$ satisfy the Lyapunov equations [19]

$$\mathbf{C}_{\tilde{Y}}^{1/2}\cdot\mathbf{H_1}+\mathbf{H_1}\cdot\mathbf{C}_{\tilde{Y}}^{1/2}=\mathbf{C} \tag{33}$$
$$\mathbf{C}_{\tilde{Y}}^{1/2}\cdot\mathbf{H_2}+\mathbf{H_2}\cdot\mathbf{C}_{\tilde{Y}}^{1/2}=-\mathbf{H_1}^2 \tag{34}$$

and in which $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ are skew-Hermitian matrices of which the entries are given by

$$\phi_{ij}=\frac{(d_{ii}g_{ij}+d_{jj}g_{ji})}{(d_{jj}^2-d_{ii}^2)} \tag{35}$$
$$\psi_{ij}=\frac{(d_{jj}g_{ij}+d_{ii}g_{ji})}{(d_{jj}^2-d_{ii}^2)} \tag{36}$$

with $\mathbf{G}=\mathbf{E}^H\mathbf{H_1}\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{E}$. Note that the expansion of $\sum_i \hat{d}_{ii}^2$ contains neither second nor higher order terms because

$$\begin{aligned}\sum_i \hat{d}_{ii}^2 &= \mathrm{trace}(\hat{\mathbf{C}}_{\tilde{Y}})\\ &= \mathrm{trace}(\mathbf{C}_{\tilde{Y}}+\epsilon\,\mathbf{C})\\ &= \mathrm{trace}(\mathbf{D}^2+\epsilon\,\mathbf{E}^H\mathbf{C}\mathbf{E}).\end{aligned}$$

It can easily be verified that the zeroth- and first-order terms in (32) cancel out.

Taking into account definitions (35) and (36) of $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$, we obtain

$$\begin{aligned}&\mathrm{trace}\left[\mathbf{E}^H\mathbf{H_1}\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{E}\boldsymbol{\Phi}+\mathbf{E}^H\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{H_1}\mathbf{E}\boldsymbol{\Psi}\right]\\ &\quad=\sum_{i<j}\frac{|(\mathbf{E}^H\mathbf{H_1}\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{E})_{ij}-(\mathbf{E}^H\mathbf{H_1}\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{E})_{ji}^*|^2}{d_{ii}^2+d_{jj}^2}\\ &\quad=\sum_{i<j}\frac{|(\tilde{\mathbf{H}}_\mathbf{1}\mathbf{D})_{ij}-(\tilde{\mathbf{H}}_\mathbf{1}\mathbf{D})_{ji}^*|^2}{d_{ii}^2+d_{jj}^2}\end{aligned} \tag{37}$$

in which $\tilde{\mathbf{H}}_\mathbf{1}=\mathbf{E}^H\mathbf{H_1}\mathbf{E}$. On the other hand, taking into account that the trace is invariant under similarity transformations and invoking (34), we obtain

$$\mathrm{trace}\left[\mathbf{E}^H\mathbf{H_2}\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{E}+\mathbf{E}^H\mathbf{C}_{\tilde{Y}}^{1/2}\mathbf{H_2}\mathbf{E}\right]=-\|\mathbf{H_1}\|^2=-\|\tilde{\mathbf{H}}_\mathbf{1}\|^2. \tag{38}$$

From (33) we have that $(\tilde{\mathbf{H}}_\mathbf{1})_{ij}=\tilde{c}_{ij}/(d_{ii}+d_{jj})$. Substituting this expression in (37) and (38) yields (21).

Now let us turn to (13) and (22). According to Theorem 5, we have

$$\frac{\hat{d}_{ii}^2}{d_{ii}^2}=1+\epsilon\,\frac{\tilde{c}_{ii}}{d_{ii}^2}+O(\epsilon^2).$$

Hence,

$$\hat{d}_{ii}=d_{ii}\left(1+\epsilon\,\frac{\tilde{c}_{ii}}{2d_{ii}^2}+O(\epsilon^2)\right).$$

Substituting this expression in (13) yields (22). ■

Theorem 4: Let $\mathbf{A_0}\in\mathbb{C}^{J\times J}$ be Hermitian positive definite and let $\mathbf{H_0}$ be its Hermitian square root, i.e., the polar decomposition of any square root of $\mathbf{A_0}$ is given by $\mathbf{H_0}\cdot\mathbf{Q}$, with $\mathbf{Q}$ unitary. Consider the perturbation $\mathbf{A}=\mathbf{A_0}+\epsilon\,\mathbf{A_1}+\epsilon^2\,\mathbf{A_2}$, with $\mathbf{A_1}$ and $\mathbf{A_2}$ Hermitian. Then the Hermitian square root of $\mathbf{A}$ is given by

$$\mathbf{H}=\mathbf{H_0}+\epsilon\,\mathbf{H_1}+\epsilon^2\,\mathbf{H_2}+O(\epsilon^3)$$

in which $\mathbf{H_1}$ and $\mathbf{H_2}$ are the solutions of the Lyapunov equations [19]

$$\mathbf{H_0}\cdot\mathbf{H_1}+\mathbf{H_1}\cdot\mathbf{H_0}=\mathbf{A_1} \tag{39}$$
$$\mathbf{H_0}\cdot\mathbf{H_2}+\mathbf{H_2}\cdot\mathbf{H_0}=\mathbf{A_2}-\mathbf{H_1}^2. \tag{40}$$

Proof: We have

$$\begin{aligned}&\mathbf{A_0}+\epsilon\mathbf{A_1}+\epsilon^2\mathbf{A_2}\\ &\quad=\mathbf{H}\cdot\mathbf{H}^H=\mathbf{A_0}+\epsilon(\mathbf{H_0}\cdot\mathbf{H_1}^H+\mathbf{H_1}\cdot\mathbf{H_0})\\ &\qquad+\epsilon^2(\mathbf{H_0}\cdot\mathbf{H_2}^H+\mathbf{H_2}\cdot\mathbf{H_0}+\mathbf{H_1}\cdot\mathbf{H_1}^H)+O(\epsilon^3).\end{aligned} \tag{41}$$

By equating the terms in ϵ and ϵ^2 in (41) and dropping the Hermitian transpose of $\mathbf{H_1}$ and $\mathbf{H_2}$, we obtain the Lyapunov

(39)–(40). Under the conditions of the theorem, these equations have a unique Hermitian solution. ∎

Theorem 5: Let $\mathbf{A_0} \in \mathbb{C}^{J \times J}$ be Hermitian, with EVD $\mathbf{A_0} = \mathbf{X_0} \cdot \mathbf{\Lambda_0} \cdot \mathbf{X_0}^H$. Consider the perturbation $\mathbf{A} = \mathbf{A_0} + \epsilon\, \mathbf{A_1}$, with $\mathbf{A_1}$ Hermitian. Let the matrix of eigenvalues of $\mathbf{A}$ be given by $\mathbf{\Lambda}$. Then, we have

$$\mathbf{\Lambda} = \mathbf{\Lambda_0} + \epsilon \operatorname{diag}(\mathbf{X_0}^H \cdot \mathbf{A_1} \cdot \mathbf{X_0}) + O(\epsilon^2). \qquad (42)$$

Proof: See [21]. ∎

Theorem 6: Let $\mathbf{A_0} \in \mathbb{C}^{J \times J}$, with SVD $\mathbf{A_0} = \mathbf{U_0} \cdot \mathbf{\Sigma_0} \cdot \mathbf{V_0}^H$, and let the singular values $\sigma_j^{(0)}$ of $\mathbf{A_0}$ be distinct. Consider the perturbation $\mathbf{A} = \mathbf{A_0} + \epsilon\, \mathbf{A_1} + \epsilon^2\, \mathbf{A_2}$. Let the SVD of $\mathbf{A}$ be given by $\mathbf{A} = \mathbf{U} \cdot \mathbf{\Sigma} \cdot \mathbf{V}^H$, with

$$\begin{aligned}\mathbf{U} &= \mathbf{U_0} + \epsilon\, \mathbf{U_1} + \epsilon^2\, \mathbf{U_2} + O(\epsilon^3)\\ \mathbf{\Sigma} &= \mathbf{\Sigma_0} + \epsilon\, \mathbf{\Sigma_1} + \epsilon^2\, \mathbf{\Sigma_2} + O(\epsilon^3)\\ \mathbf{V} &= \mathbf{V_0} + \epsilon\, \mathbf{V_1} + \epsilon^2\, \mathbf{V_2} + O(\epsilon^3).\end{aligned}$$

Then, we have

$$\mathbf{\Sigma_1} = \frac{1}{2} \operatorname{diag}(\mathbf{U_0}^H \cdot \mathbf{A_1} \cdot \mathbf{V_0} + \mathbf{V_0}^H \cdot \mathbf{A_1}^H \cdot \mathbf{U_0}) \qquad (43)$$

$$\begin{aligned}\mathbf{\Sigma_2} = \frac{1}{2} \operatorname{diag}(&\mathbf{U_0}^H \cdot \mathbf{A_1} \cdot \mathbf{V_1} + \mathbf{V_0}^H \cdot \mathbf{A_1}^H \cdot \mathbf{U_1}\\ &+ \mathbf{U_0}^H \cdot \mathbf{A_2} \cdot \mathbf{V_0} + \mathbf{V_0}^H \cdot \mathbf{A_2}^H \cdot \mathbf{U_0})\end{aligned} \qquad (44)$$

$$\mathbf{U_1} = \mathbf{U_0} \cdot \mathbf{\Psi} \qquad (45)$$

$$\mathbf{V_1} = \mathbf{V_0} \cdot \mathbf{\Phi} \qquad (46)$$

in which $\mathbf{\Psi}$ and $\mathbf{\Phi}$ are skew-Hermitian matrices of which the entries are given by

$$\psi_{ij} = \frac{(\sigma_j^{(0)} g_{ij} + \sigma_i^{(0)} g_{ji})}{(\sigma_j^{(0)^2} - \sigma_i^{(0)^2})} \qquad (47)$$

$$\phi_{ij} = \frac{(\sigma_i^{(0)} g_{ij} + \sigma_j^{(0)} g_{ji})}{(\sigma_j^{(0)^2} - \sigma_i^{(0)^2})} \qquad (48)$$

where $\mathbf{G} = \mathbf{U_0}^H \cdot \mathbf{A_1} \cdot \mathbf{V_0}$.

Proof: Consider

$$\begin{aligned}&(\mathbf{A_0} + \epsilon \mathbf{A_1} + \epsilon^2 \mathbf{A_2})(\mathbf{V_0} + \epsilon \mathbf{V_1} + \epsilon^2 \mathbf{V_2})\\ &= (\mathbf{U_0} + \epsilon \mathbf{U_1} + \epsilon^2 \mathbf{U_2})(\mathbf{\Sigma_0} + \epsilon \mathbf{\Sigma_1} + \epsilon^2 \mathbf{\Sigma_2}) + O(\epsilon^3)\end{aligned} \qquad (49)$$

$$\begin{aligned}&(\mathbf{A_0}^H + \epsilon \mathbf{A_1}^H + \epsilon^2 \mathbf{A_2}^H)(\mathbf{U_0} + \epsilon \mathbf{U_1} + \epsilon^2 \mathbf{U_2})\\ &= (\mathbf{V_0} + \epsilon \mathbf{V_1} + \epsilon^2 \mathbf{V_2})(\mathbf{\Sigma_0} + \epsilon \mathbf{\Sigma_1} + \epsilon^2 \mathbf{\Sigma_2}) + O(\epsilon^3).\end{aligned} \qquad (50)$$

Equations (43) and (45)–(48) are obtained by equating the terms in ϵ in (49)–(50) [27].

Equation (44) is obtained by equating the terms in ϵ^2 in (49)–(50) and multiplying them from the left by $\mathbf{U_0}^H$ and $\mathbf{V_0}^H$, respectively. ∎

REFERENCES

[1] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 1996, vol. 8, pp. 757–763.

[2] A. J. Bell and T. J. Sejnowski, "An information maximization approach to blind separation and blind deconvolution," *Neur. Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.

[3] A. Belouchrani *et al.*, "A blind source separation technique using second order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[4] 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'91)*, vol. 5, Toronto, Canada, 1991, pp. 3109–3112.

[5] ——, "A tetradic decomposition of 4th-order tensors. Application to the source separation problem," in *SVD and Signal Processing, III. Algorithms, Applications and Architectures*, B. De Moor and M. Moonen, Eds. Amsterdam: Elsevier, 1995, pp. 375–382.

[6] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Proc. Inst. Elect. Eng.-F*, vol. 140, no. 6, pp. 362–370, 1993.

[7] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. 7th Eur. Signal Processing Conf. (EUSIPCO'94)*, vol. 2, Edinburgh, U.K., Sep. 13–16, 1994, pp. 776–779.

[8] W. H. Cho and T. W. Spencer, "Estimation of polarization and slowness in mixed wavefields," *Geophys.*, vol. 57, no. 6, pp. 805–814, 1992.

[9] P. Comon, "Independent component analysis, A new concept?," *Signal Process.*, vol. 36, no. 3, pp. 287–314, 1994.

[10] P. Comon and B. Mourrain, "Decomposition of quantics in sums of powers of linear forms," *Signal Process.*, vol. 53, no. 2–3, pp. 93–108, Sep. 1996.

[11] L. De Lathauwer, B. De Moor, and J. Vandewalle, "Independent component analysis based on higher order statistics only," in *Proc. 8th IEEE Signal Processing Workshop on Statistical Signal and Array Processing (SSAP'96)*, Corfu, Greece, Jun. 24–26, 1996, pp. 356–359.

[12] L. De Lathauwer, "Signal processing based on multilinear algebra," Ph.D., E.E. Dept., Katholieke Universiteit Leuven , Leuven, Belgium, 1997.

[13] L. De Lathauwer and J. Vandewalle, "A residual bound for the mixing matrix in ICA," in *Proc. 9th Eur. Signal Processing Conf. (EUSIPCO'98)*, Rhodos, Greece, Sep. 8–11, 1998, pp. 2065–2068.

[14] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An introduction to independent component analysis," *J. Chemometr.*, vol. 14, no. 3, pp. 123–149, 2000.

[15] ——, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 567–572, May 2000.

[16] ——, "Independent component analysis and (simultaneous) third-order tensor diagonalization," *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2262–2271, Oct. 2001.

[17] L. De Lathauwer and B. De Moor, "On the blind separation of noncircular sources," in *Proc. 11th Eur. Signal Processing Conf. (EUSIPCO'02)*, vol. 2, Toulouse, France, Sep. 3–6, 2002, pp. 99–102.

[18] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[19] R. A. Horn and C. R. Johnson, *Topics in Matrix Analysis*. New York: Cambridge Univ. Press, 1999.

[20] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.

[21] T. Kato, *A Short Introduction to Perturbation Theory for Linear Operators*. New York: Springer-Verlag, 1982.

[22] H. Krim and M. Viberg, "Two decades of array signal processing research. The parametric approach," *IEEE Signal Processing Mag.*, vol. 13, no. 7, pp. 67–94, Jul. 1996.

[23] K. Nagano, "Blind detection of an elliptically polarized wave in three-component seismic measurement," in *Proc. 11th Euro. Signal Processing Conf. (EUSIPCO'02)*, vol. 3, Toulouse, France, Sep. 3–6, 2002, pp. 11–14.

[24] C. L. Nikias and A. P. Petropulu, *Higher Order Spectra Analysis. A Nonlinear Signal Processing Framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[25] D. Otte, "Development and evaluation of singular value analysis methodologies for studying multivariate noise and vibration problems," Ph.D. thesis, Mech. Dept., Katholieke Universiteit Leuven , Leuven, Belgium, 1994.

[26] D.-T. Pham, "Joint approximate diagonalization of positive definite matrices," *SIAM J. Matrix Anal. Appl.*, vol. 22, no. 4, pp. 1136–1152, 2001.

[27] G. W. Stewart, "Perturbation theory for the singular value decomposition," in *SVD and Signal Processing, II: Algorithms, Analysis and Applications*, R. J. Vaccaro, Ed. New York: Elsevier, 1991, pp. 99–109.

[28] A.-J. van der Veen, "Joint diagonalization via subspace fitting techniques," in *Proc. Int. Conf. Acoustics, Speech and Signal Processing (ICASSP'01)*, Salt Lake City, UT, May 2001.

[29] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE ASSP Mag.*, vol. 5, pp. 4–24, Apr. 1988.

[30] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.

Lieven De Lathauwer (M'04) was born in Aalst, Belgium, on November 10, 1969. He received the Master's degree in electro-mechanical engineering and the doctoral degree in applied sciences from the Katholieke Universiteit Leuven (K.U.Leuven), Leuven, Belgium, in 1992 and 1997, respectively.

His Ph.D. thesis concerned signal processing based on multilinear algebra. He is currently with the French Centre National de la Recherche Scientifique (C.N.R.S.), Cergy-Pontoise Cedex, France. His research interests include linear and multilinear algebra, statistical signal and array processing, higher order statistics, independent component analysis, identification, blind identification, and equalization.

Bart De Moor (F'04) received the Master's and the Ph.D. degrees in electrical engineering from the Katholieke Universiteit Leuven (K.U.Leuven), Leuven, Belgium, in 1983 and 1988, respectively.

He was a Visiting Research Associate at Stanford University, Stanford, CA, from 1988 to 1990 . Currently, he is a Full Professor in the Department of Electrical Engineering (ESAT), K.U.Leuven. His research interests are in numerical linear algebra and optimization, system theory, control and identification, quantum information theory, data mining, information retrieval, and bio-informatics, in which he (co)authored more than 400 papers and three books. From 1991 to 1999, he was the Chief Advisor on Science and Technology of several ministers of the Belgian Federal and the Flanders Regional Governments. He is in the board of three spin-off companies, of the Flemish Interuniversity Institute for Biotechnology, the Study Center for Nuclear Energy and several other scientific and cultural organizations.

Dr. De Moor has won him several scientific awards including the Leybold–Heraeus Prize (1986), Leslie Fox Prize in 1989, Guillemin–Cauer Best Paper Award of the IEEE TRANSACTION ON CIRCUITS AND SYSTEMS in 1990, the Laureate of the Belgian Royal Academy of Sciences (1992), the bi-annual Siemens Award (1994), the Best Paper Award of Automatica (IFAC) in 1996, and the IEEE Signal Processing Society Best Paper Award in 1999. Since 2002 he also makes regular television appearances in the Science Show 'Hoe?Zo!' on national television in Belgium. Full biographical details can be found at www.esat.kuleuven.ac.be/~demoor.

Joos Vandewalle (F'92) was born in Kortrijk, Belgium, in August 1948. He received the doctoral degree in applied sciences and the special doctoral degree from the Katholieke Universiteit Leuven (K.U.Leuven), Leuven, Belgium, in 1976 and 1984, respectively.

He was a Postdoctoral Researcher from 1976 to 1978 and a Visiting Assistant Professor from 1978 to 1979 in the Electrical Engineering and Computer Science Department, University of California, Berkeley. Since 1979, he has been with the Electrical Engineering Department, K.U.Leuven, where he has been a Full Professor since 1986. From August 1996 till August 1999, he was head of the Department of Electrical Engineering (ESAT) and from August 1999 till July 2002, he was Vice-Dean of the Faculty of Engineering. Since August 2003, he is Head of the Department of Electrical Engineering. He is also head of the Research Group SCD. He teaches courses in linear algebra, linear and nonlinear system and circuit theory, signal processing and neural networks. His main research interests are in system theory and its applications in circuit theory, signal processing, cryptography and neural networks. His recent research interests are in nonlinear methods (support vector machines, multilinear algebra) for data processing. He has (co)authored more than 200 international journal papers in these areas. He is the co author of four books and coeditor of five books. He is a member of the editorial board of the *International Journal of Circuit Theory and Its Applications*, *Neurocomputing*, *Neural Networks*, and the *Journal of Circuits Systems and Computers*.

He received several best paper awards and research awards. From 1989 till 1991, he was an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS. He was Deputy Editor-in-chief of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: FUNDAMENTAL THEORY AND APPLICATIONS from January 2002 till December 2003. He was Program Chairman of the International Symposium on Circuits and Systems 2000 in Geneva, and for the International Joint Conference on Cellular Neural Networks (IJCNN) 2004 in Budapest. From 1991 to 1992, he held the Francqui chair on Artificial Neural Networks at the University of Liège and in 2001 to 2002, he held this chair on Advanced Data Processing techniques at the Free University of Brussels. He is also Fellow of the Institute of Electrical Engineers, U.K., and a member of the Academia Europaea and of the Belgian Academy of sciences.