Independent Component Analysis and Blind Signal Separation ICA 2000, June 19–22, 2000, Helsinki, Finland, pp. 81–86

OVERDETERMINED BLIND SOURCE SEPARATION: USING MORE SENSORS THAN SOURCE SIGNALS IN A NOISY MIXTURE

Marcel Joho[1], *Heinz Mathis*[1], *and Russell H. Lambert*[2]

[1]Signal and Information Processing Laboratory,
Swiss Federal Institute of Technology Zurich, Switzerland
[2]Pivotal Technologies, 70 S. Lake Ave., Suite 900, Pasadena, CA 91101, USA
joho@isi.ee.ethz.ch, *mathis@isi.ee.ethz.ch*, *russ@pivotaltech.com*

ABSTRACT

This paper addresses the blind source separation problem for the case where more sensors than source signals are available. A noisy-sensor model is assumed. The proposed algorithm comprises two stages, where the first stage consists of a principal component analysis (PCA) and the second one of an independent component analysis (ICA). The purpose of the PCA stage is to increase the input SNR of the succeeding ICA stage and to reduce the sensor dimensionality. The ICA stage is used to separate the remaining mixture into its independent components. A simulation example demonstrates the performance of the algorithm proposed.

1. INTRODUCTION

1.1. Problem description

Blind source separation (BSS) is a problem posed by many applications related to acoustics or communications. Usually the BSS problem is analyzed for the case where there are just as many sensors as source signals. Furthermore, ideal sensors are usually assumed, which have no additive sensor noise. Only little work has been done on the analysis of algorithms in the case of noisy sensors [1, 2, 3, 4]. Usually one hopes that the sensor noise is low enough so as not to influence the performance of the BSS algorithm considerably. This paper concentrates on the case where a low SNR is present at the sensors, and shows that one possible way to enhance the performance of the separation is to use more sensors than source signals.

This situation is referred to as the *overdetermined blind source separation* problem and sometimes also as the undercomplete bases problem. It is overdetermined in the sense that if the sources are of interest, more observations than necessary for the reconstruction of the original signals are available. However, if referring to the Linear Algebra system of finding the separation matrix given the mixing matrix, the term underdetermined is used, since we have more unknowns than equations.

We divide the task at hand into two stages. Starting with M input sensors, the first stage performs a singular-value decomposition producing M_s virtual sensors ($M_\mathrm{s} < M$), which still contain a noisy mixture of the source signals, but with a higher SNR than the true sensors. The remaining $M - M_\mathrm{s}$ virtual sensors contain a mixture of the sensor noises and are discarded in the second stage. The second stage consists of an ordinary blind source separation algorithm for the $M_\mathrm{s} \times M_\mathrm{s}$ problem.

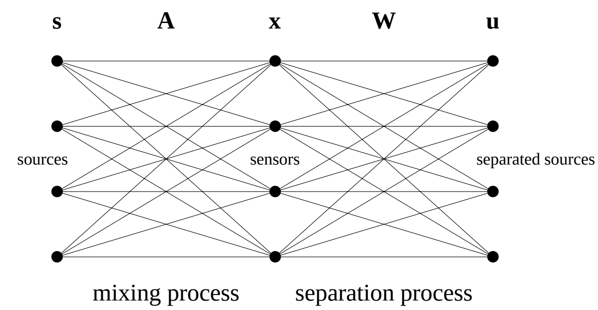

Figure 1: Blind source separation setup ($M_\mathrm{s} = M = 4$).

1.2. Notation

The notation used throughout this paper is the following: Vectors are written in lower case, matrices in upper case. Matrix and vector transpose, complex conjugation and Hermitian transpose are denoted by $(.)^T$, $(.)^*$, and $(.)^H = ((.)^*)^T$, respectively. The sample index is denoted by t. The identity matrix is denoted by $\mathbf{I}$, a vector or a matrix containing only zeros by $\mathbf{0}$. $E\{.\}$ denotes the expectation operator. Vector or matrix dimensions are given in superscript. The Frobenius norm of a matrix is denoted by $\|.\|_F$.

2. OVERDETERMINED BLIND SOURCE SEPARATION

2.1. Problem description

The mixing process is described as

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

where $\mathbf{s}_t = (s_1, \ldots, s_{M_\mathrm{s}})_t^T$ contains the samples of the unknown source signals at time t, $\mathbf{x}_t = (x_1, \ldots, x_M)_t^T$ the samples of the M sensor signals at sample time t, $\mathbf{n}_t$ the samples of the sensor noise at time t, and $\mathbf{A}^{M \times M_\mathrm{s}} = [\mathbf{a}_1 \cdots \mathbf{a}_M]^T$ is the unknown mixing matrix. In the overdetermined case we have more sensors than source signals ($M > M_\mathrm{s}$).

Solving the blind source separation problem means to find a separation matrix $\mathbf{W}^{M_\mathrm{s} \times M}$ such that the output of the separation process

$$\mathbf{u}_t = \mathbf{W}\mathbf{x}_t = \mathbf{W}(\mathbf{A}\mathbf{s}_t + \mathbf{n}_t) = \mathbf{P}\mathbf{s}_t + \mathbf{W}\mathbf{n}_t \tag{2}$$

retrieves waveform-preserving estimates of the unknown source signals by using only the time series of the sensor signals $\mathbf{x}_t$ for

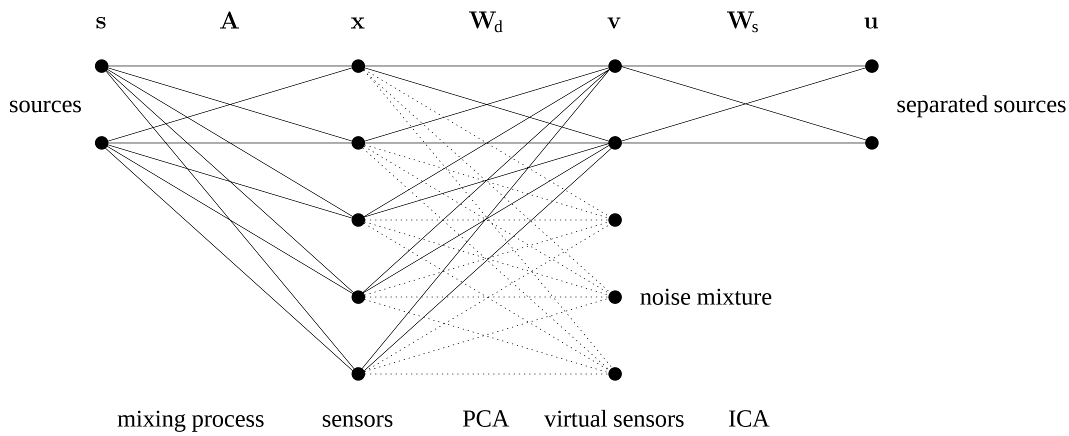

Figure 2: Overdetermined blind source separation setup using the proposed two-stage algorithm ($M_\mathrm{s} = 2$, $M = 5$).

$t = 1, 2, \ldots$. $\mathbf{P}$ denotes the total transfer matrix of the global system.

2.2. Assumptions

In addition to the problem proposed above, we make the following assumptions:

- $\mathcal{A}1$ Time-invariant mixing matrix $\mathbf{A}$.
- $\mathcal{A}2$ $\mathbf{A}$ has full rank M_s.
- $\mathcal{A}3$ Source signals $s_m, m = 1, \ldots, M_\mathrm{s}$, are mutually independent and iid.
- $\mathcal{A}4$ All source signals s_m but possibly one are non-Gaussian.
- $\mathcal{A}5$ All source signals are unknown and have the same power σ_s^2.
- $\mathcal{A}6$ There are more sensors than source signals $(M > M_\mathrm{s})$.
- $\mathcal{A}7$ All sensors have the same noise characteristics. The sensor noise is additive white Gaussian noise with power σ_n^2. The sensor noise of the sensors is mutually independent.
- $\mathcal{A}8$ The source signals and the sensor noise are mutually independent.

As a consequence, $\mathcal{A}3$ and $\mathcal{A}5$ imply

$$\mathbf{R_{ss}} \triangleq E\left\{\mathbf{s}_t \mathbf{s}_t^H\right\} = \sigma_\mathrm{s}^2\, \mathbf{I}_{M_\mathrm{s}} \tag{3}$$

and $\mathcal{A}7$ implies

$$\mathbf{R_{nn}} \triangleq E\left\{\mathbf{n}_t \mathbf{n}_t^H\right\} = \sigma_\mathrm{n}^2\, \mathbf{I}_M\,. \tag{4}$$

3. PROPOSED ALGORITHM

The proposed algorithm has two stages. The first stage is based on a *principal component analysis* (PCA) algorithm where we project the M sensor signals onto an M_s-dimensional signal-plus-noise subspace and an $M - M_\mathrm{s}$ dimensional noise subspace. The second stage performs an *independent component analysis* (ICA) of the signal-plus-noise space to obtain the estimates of the source signals.

3.1. First stage: Principal Component Analysis (PCA)

First we want to decorrelate the sensor signals. Decorrelation is a necessary but not a sufficient condition for independence. To this end we transform the sensor signals by a unitary transformation matrix $\mathbf{W}_\mathrm{d}$

$$\mathbf{v}_t = \mathbf{W}_\mathrm{d} \mathbf{x}_t \tag{5}$$

such that $\mathbf{R_{vv}} \triangleq E\left\{\mathbf{v}_t \mathbf{v}_t^H\right\}$ becomes a diagonal matrix.

By applying the SVD (*singular-value decomposition*) on $\mathbf{A}$, we have

$$\mathbf{A} = \mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^H = \mathbf{U}\begin{bmatrix}\widetilde{\mathbf{\Sigma}} \\ \mathbf{0}\end{bmatrix}\mathbf{V}^H \tag{6}$$

where $\mathbf{U}^{M\times M}$ and $\mathbf{V}^{M_\mathrm{s}\times M_\mathrm{s}}$ are unitary matrices, and $\mathbf{\Sigma}^{M\times M_\mathrm{s}}$ and $\widetilde{\mathbf{\Sigma}}^{M_\mathrm{s}\times M_\mathrm{s}}$ are diagonal matrices which contain the singular values $\sigma_m(\mathbf{A})$ of $\mathbf{A}$

$$\widetilde{\mathbf{\Sigma}} = \mathrm{diag}\left(\sigma_1, \ldots, \sigma_{M_\mathrm{s}}\right) \tag{7}$$

with

$$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{M_\mathrm{s}} > 0 \tag{8}$$

where the last inequality comes from the assumption $\mathcal{A}2$.

The SVD of the input correlation matrix $\mathbf{R_{xx}}$ gives with (3) and (4)

$$\begin{aligned}
\mathbf{R_{xx}} &\triangleq E\left\{\mathbf{x}_t \mathbf{x}_t^H\right\} && (9)\\
&= \mathbf{A}\,\mathbf{R_{ss}}\,\mathbf{A}^H + \mathbf{R_{nn}} && (10)\\
&= \sigma_\mathrm{s}^2 \mathbf{A}\mathbf{A}^H + \sigma_\mathrm{n}^2\,\mathbf{I}_M && (11)\\
&= \sigma_\mathrm{s}^2\,\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{V}^H\,\mathbf{V}\,\mathbf{\Sigma}^T\,\mathbf{U}^H + \sigma_\mathrm{n}^2\,\mathbf{I}_M && (12)\\
&= \sigma_\mathrm{s}^2\,\mathbf{U}\,\mathbf{\Sigma}\,\mathbf{\Sigma}^T\,\mathbf{U}^H + \sigma_\mathrm{n}^2\,\mathbf{I}_M && (13)\\
&= \sigma_\mathrm{s}^2\,\mathbf{U}\,\overline{\mathbf{\Sigma}}^2\,\mathbf{U}^H + \sigma_\mathrm{n}^2\,\mathbf{U}\,\mathbf{U}^H && (14)\\
&= \mathbf{U}\left(\sigma_\mathrm{s}^2\,\overline{\mathbf{\Sigma}}^2 + \sigma_\mathrm{n}^2\,\mathbf{I}_M\right)\mathbf{U}^H && (15)\\
&= \mathbf{U}\,\mathbf{\Sigma}_{\mathbf{xx}}^2\,\mathbf{U}^H && (16)
\end{aligned}$$

with

$$\overline{\mathbf{\Sigma}}^{M\times M} = \mathrm{diag}\left(\sigma_1, \dots, \sigma_{M_\mathrm{s}}, 0, \dots, 0\right) \tag{17}$$

$$\mathbf{\Sigma}^2_{\mathbf{xx}} = \mathrm{diag}\left(\sigma^2_{\mathrm{x}_1}, \dots, \sigma^2_{\mathrm{x}_{M_\mathrm{s}}}, \sigma^2_{\mathrm{x}_{M_\mathrm{s}+1}}, \dots, \sigma^2_{\mathrm{x}_M}\right) \tag{18}$$

$$= \mathrm{diag}\left(\sigma_1^2\sigma_\mathrm{s}^2 + \sigma_\mathrm{n}^2, \dots, \sigma^2_{M_\mathrm{s}}\sigma_\mathrm{s}^2 + \sigma_\mathrm{n}^2, \sigma_\mathrm{n}^2, \dots, \sigma_\mathrm{n}^2\right) . \tag{19}$$

The first term of (15) contains the contribution of the source signals, while the second one contains the contribution of the sensor noise. By choosing

$$\mathbf{W}_\mathrm{d} = \begin{bmatrix} \mathbf{W}_{\mathrm{d}1}^{M_\mathrm{s}\times M} \\ \mathbf{W}_{\mathrm{d}2}^{(M-M_\mathrm{s})\times M} \end{bmatrix} = \mathbf{U}^H = \mathbf{U}^{-1} \tag{20}$$

we obtain with (5) and (15) the correlation matrix

$$\mathbf{R}_{\mathbf{vv}} \triangleq E\left\{\mathbf{v}_t\mathbf{v}_t^H\right\} \tag{21}$$

$$= \mathbf{W}_\mathrm{d}\,\mathbf{R}_{\mathbf{xx}}\,\mathbf{W}_\mathrm{d}^H \tag{22}$$

$$= \sigma_\mathrm{s}^2\,\overline{\mathbf{\Sigma}}^2 + \sigma_\mathrm{n}^2\,\mathbf{I}_M = \mathbf{\Sigma}^2_{\mathbf{xx}} . \tag{23}$$

Therefore $\mathbf{R}_{\mathbf{vv}}$ becomes a diagonal matrix containing the singular values of $\mathbf{R}_{\mathbf{xx}}$ in descending order with $\sigma^2_{\mathrm{x}_1} \geq \dots \geq \sigma^2_{\mathrm{x}_{M_\mathrm{s}}} > \sigma^2_{\mathrm{x}_{M_\mathrm{s}+1}} = \dots = \sigma^2_{\mathrm{x}_M} = \sigma^2_\mathrm{n}$. Since $\mathbf{R}_{\mathbf{vv}}$ is a diagonal matrix, the signals in $\mathbf{v}$ are mutually uncorrelated, but not necessarily independent. Furthermore, the signals in $\mathbf{v}$ are ordered by their powers and since $\mathbf{W}_\mathrm{d}$ is a unitary matrix, the first M_s elements of $\mathbf{v}$ contain all the signal power of the source signals received at the sensors x_m, $m = 1, \dots, M$, and the remaining $M - M_\mathrm{s}$ elements of $\mathbf{v}$ contain only a mixture of the sensor noise. Assuming the knowledge of the number of source signals M_s, we can partition the vector $\mathbf{v}$ into its M_s *principal* and $M - M_\mathrm{s}$ *minor components*

$$\mathbf{v} = \begin{pmatrix} \mathbf{v}_\mathrm{s}^{M_\mathrm{s}} \\ \mathbf{v}_\mathrm{n}^{M-M_\mathrm{s}} \end{pmatrix} . \tag{24}$$

By doing so, we can now use $\mathbf{v}_\mathrm{s}$ for the second stage and discard $\mathbf{v}_\mathrm{n}$, as the elements of $\mathbf{v}_\mathrm{n}$ do not contain any components of the source signals. We only use the principal components for further processing. This procedure corresponds to a *principal component analysis* (PCA). Note that there are as many principal components as unknown source signals. Furthermore, the principal components are mutually uncorrelated. We therefore use

$$\mathbf{v}_\mathrm{s} = \mathbf{W}_{\mathrm{d}1}\,\mathbf{x} \tag{25}$$

to generate M_s *virtual sensors* from the M true sensors.

3.2. Second stage: Independent Component Analysis (ICA)

In the second stage we use an ICA algorithm to find a separation matrix $\mathbf{W}_\mathrm{s}$ such that

$$\mathbf{u}_t = \mathbf{W}_{\mathrm{s}_t}\mathbf{v}_{\mathrm{s}t} \tag{26}$$

is an estimate of $\mathbf{s}_t$ up to scaling and permutation of the elements. Perfect separation occurs if the output signals of $\mathbf{u}$ are mutually independent. The joint density of $\mathbf{u}$ is then a product of the marginal densities of u_m [5]

$$p_{\mathbf{u}(\mathbf{u})} = \prod_{m=1}^{M_\mathrm{s}} p_{u_m}(u_m) \tag{27}$$

because independence of the source signals is assumed by $\mathcal{A}3$, i.e. $p_{\mathbf{s}(\mathbf{s})} = \prod_{m=1}^{M_\mathrm{s}} p_{s_m}(s_m)$.

By using a PCA stage we have reduced the overdetermined $M \times M_\mathrm{s}$ BSS problem to an $M_\mathrm{s} \times M_\mathrm{s}$ BSS problem with additive sensor noise. We can now use any known BSS algorithm for the regular ($M_\mathrm{s} \times M_\mathrm{s}$) case, e.g., the natural gradient learning algorithm proposed in [6]

$$\mathbf{W}_{\mathrm{s}_{t+1}} = \mathbf{W}_{\mathrm{s}_t} + \mu\left[\mathbf{I}_{M_\mathrm{s}} - \mathbf{y}_t\mathbf{u}_t^H\right]\mathbf{W}_{\mathrm{s}_t} \tag{28}$$

the EASI algorithm proposed in [7]

$$\mathbf{W}_{\mathrm{s}_{t+1}} = \mathbf{W}_{\mathrm{s}_t} + \mu\left[\mathbf{I}_{M_\mathrm{s}} - \mathbf{u}_t\mathbf{u}_t^H + \mathbf{u}_t\mathbf{y}_t^H - \mathbf{y}_t\mathbf{u}_t^H\right]\mathbf{W}_{\mathrm{s}_t} \tag{29}$$

the Infomax algorithm [8]

$$\mathbf{W}_{\mathrm{s}_{t+1}} = \mathbf{W}_{\mathrm{s}_t} + \mu\left[\mathbf{W}_{\mathrm{s}_t}^{-H} - \mathbf{y}_t\mathbf{v}_{\mathrm{s}t}^H\right] \tag{30}$$

or a blind stochastic gradient algorithm, e.g., multichannel blind LMS (MBLMS) [9],

$$\mathbf{W}_{\mathrm{s}_{t+1}} = \mathbf{W}_{\mathrm{s}_t} + \mu\left(\mathbf{u}_t - \mathbf{y}_t\right)\mathbf{v}_{\mathrm{s}t}^H \tag{31}$$

where $\mathbf{y}_t = \mathbf{g}(\mathbf{u}_t)$ and $\mathbf{g}(.)$ is a nonlinear function known as the *Bussgang nonlinearity* or the *score function*, which depends on the pdf of the source signals [10, 11]. Alternatively, a batch algorithm can be used for the separation, see [12] for instance, which explicitly uses higher-order cummulants.

3.3. Combining both stages

Both stages are now combined as shown in Fig. 2. With (25) and (26) we obtain

$$\mathbf{W}_t^{M_\mathrm{s}\times M} = \mathbf{W}_{\mathrm{s}_t}\mathbf{W}_{\mathrm{d}1} . \tag{32}$$

The first stage, i.e. $\mathbf{W}_{\mathrm{d}1}$, can be seen as a preprocessing step, which decorrelates the sensor signals and reduces the input dimension. From $\mathcal{A}1$ and $\mathcal{A}5$ we know that $\mathbf{R}_{\mathbf{xx}}$ is time invariant, and therefore $\mathbf{W}_{\mathrm{d}1}$ does not have to be adapted. However, $\mathbf{W}_\mathrm{s}$ is computed by an iterative learning algorithm. With (2) we obtain the total transfer matrix of the global system

$$\mathbf{P}_t = \mathbf{W}_{\mathrm{s}_t}\mathbf{W}_{\mathrm{d}1}\mathbf{A} \tag{33}$$

which should become close to a scaled permutation matrix to attain a good signal separation. Of course $\mathbf{P}$ is available in a simulation environment only, as the mixing matrix $\mathbf{A}$ is unknown by assumption.

3.4. Minimum-norm solution

In the noiseless case, the underdetermined system equation $\mathbf{P} = \mathbf{W}\mathbf{A} = \mathbf{I}_{M_\mathrm{s}}$ is fulfilled by an infinite number of possible separation matrices $\mathbf{W}$, e.g., all matrices of the form $\mathbf{W} = \mathbf{V}[\widetilde{\mathbf{\Sigma}}^{-1}\ \mathcal{E}]\mathbf{U}^H$, where $\mathcal{E}$ is an arbitrary $M_\mathrm{s} \times (M - M_\mathrm{s})$ matrix. An additional constraint is necessary to make the separation matrix unique. One possibility is to constrain $\mathbf{W}$ to have the minimal possible Frobenius norm. Such a solution is referred to as the *minimum-norm* (MN) solution [13]. Because of the unitary invariance property of the Frobenius norm [14] we have $\|\mathbf{W}\|_F^2 = \|\,[\widetilde{\mathbf{\Sigma}}^{-1}\ \mathcal{E}\,]\,\|_F^2 =$

$\|\widetilde{\mathbf{\Sigma}}^{-1}\|_F^2 + \|\boldsymbol{\mathcal{E}}\|_F^2$. Thus, $\|\mathbf{W}\|_F$ is minimal with $\boldsymbol{\mathcal{E}} \equiv \mathbf{0}$, and the minimum-norm solution becomes

$$\mathbf{W}^{\mathrm{MN}} = \mathbf{V}[\widetilde{\mathbf{\Sigma}}^{-1}\ \mathbf{0}]\mathbf{U}^H = \mathbf{A}^{\#} \tag{34}$$

$$= \mathbf{V}\,\widetilde{\mathbf{\Sigma}}^{-1}\mathbf{W}_{\mathrm{d1}} \tag{35}$$

where $\mathbf{A}^{\#}$ denotes the Moore-Penrose pseudoinverse of $\mathbf{A}$, see also [15]. Since $\mathbf{W}_{\mathrm{d1}}$ is contained in $\mathbf{W}^{\mathrm{MN}}$, see (35), the proposed *PCA preprocessing stage is part of the minimum-norm solution*. Furthermore, from (32) we can define

$$\mathbf{W}_{\mathrm{s}}^{\mathrm{MN}} = \mathbf{V}\widetilde{\mathbf{\Sigma}}^{-1}. \tag{36}$$

3.5. Minimum mean-squared error solution

Using the mean-squared error (MSE) as the cost function, we can write down the cost to minimize as

$$J_{\mathrm{MSE}} = E\left\{\|\mathbf{s}_t - \mathbf{u}_t\|_2^2\right\}. \tag{37}$$

Using (2) and setting the cost derivative with respect to the separation matrix $\mathbf{W}$ to zero, we arrive after some simplifications at

$$\frac{\partial J_{\mathrm{MSE}}}{\partial \mathbf{W}} = E\left\{(-\mathbf{s}\mathbf{s}^H\mathbf{A}^H + \mathbf{W}\mathbf{A}\mathbf{s}\mathbf{s}^H\mathbf{A}^H + \mathbf{W}\mathbf{n}\mathbf{n}^H)\right\} = 0. \tag{38}$$

Solving Eq. (38) for the separation matrix $\mathbf{W}$ yields the *minimum mean-squared error* (MMSE) solution

$$\mathbf{W}^{\mathrm{MMSE}} = \mathbf{R}_{\mathbf{ss}}\mathbf{A}^H(\mathbf{A}\mathbf{R}_{\mathbf{ss}}\mathbf{A}^H + \mathbf{R}_{\mathbf{nn}})^{-1}. \tag{39}$$

Inserting $\mathbf{R}_{\mathbf{ss}}$ and $\mathbf{R}_{\mathbf{nn}}$ from (3) and (4), respectively, (39) can then be written as

$$\mathbf{W}^{\mathrm{MMSE}} = \mathbf{A}^H(\mathbf{A}\mathbf{A}^H + \frac{\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{s}}^2}\mathbf{I}_M)^{-1}. \tag{40}$$

By inserting (6) into (40) and after a longer calculation, we obtain

$$\mathbf{W}^{\mathrm{MMSE}} = \mathbf{V}\,\widetilde{\widetilde{\mathbf{\Sigma}}}^{-1}\mathbf{W}_{\mathrm{d1}} \tag{41}$$

where

$$\widetilde{\widetilde{\mathbf{\Sigma}}} = \mathrm{diag}\left(\frac{\sigma_1^2 + \frac{\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{s}}^2}}{\sigma_1}, \ldots, \frac{\sigma_{M_{\mathrm{s}}}^2 + \frac{\sigma_{\mathrm{n}}^2}{\sigma_{\mathrm{s}}^2}}{\sigma_{M_{\mathrm{s}}}}\right). \tag{42}$$

We now recognize, that if σ_{n} goes toward zero, $\widetilde{\widetilde{\mathbf{\Sigma}}}$ converges toward $\widetilde{\mathbf{\Sigma}}$, and therefore $\mathbf{W}^{\mathrm{MMSE}}$ approaches $\mathbf{W}^{\mathrm{MN}}$. Since $\mathbf{W}_{\mathrm{d1}}$ is also contained in $\mathbf{W}^{\mathrm{MMSE}}$, we can clearly see that the proposed *PCA preprocessing stage is part of the MMSE solution* $\mathbf{W}^{\mathrm{MMSE}}$. Furthermore, from (32) we can define

$$\mathbf{W}_{\mathrm{s}}^{\mathrm{MMSE}} = \mathbf{V}\,\widetilde{\widetilde{\mathbf{\Sigma}}}^{-1}. \tag{43}$$

3.6. Estimation of the number of sources

Until now we have assumed that we know the number of unknown source signals M_{s} which are involved in the mixing process. However, often M_{s} is unknown and therefore has to be estimated. Since the input correlation matrix $\mathbf{R}_{\mathbf{xx}}$ or $\mathbf{R}_{\mathbf{vv}}$ have M_{s} dominant singular values and $M - M_{\mathrm{s}}$ minor singular values which have all the same value σ_{n}^2, see also (19), the number of source signals involved in the mixing process can be evaluated by choosing a threshold just above σ_{n}^2. Moreover, if the number of active source signals M_{s} changes, it will be visible also from analyzing the distribution of the singular values. Alternatively, an information-theoretic criterion can be used to estimate the number of source signals [16].

4. SIMULATION

In the following, we give a simulation example to analyze the behavior of the algorithm proposed.

4.1. Performance measurement

In order to describe the performance of the algorithm proposed, we use the *signal-to-noise ratio* (SNR), the *signal-to-interference ratio* (SIR), the *signal-to-interference-plus-noise ratio* (SINR), and the convergence rate as the criteria. We define the following performance measurements:

SNR (x_m)	SNR of input sensor signal x_m,
SNR $(\mathbf{x})$	total input SNR,
SNR (v_m)	SNR of virtual sensor signal v_m,
SNR $(\mathbf{v}_{\mathrm{s}})$	total SNR of virtual sensors $\mathbf{v}_{\mathrm{s}}$,
SNR (u_m)	SNR of output signal u_m,
SNR $(\mathbf{u})$	total output SNR,
SIR (u_m)	SIR of output signal u_m,
SIR $(\mathbf{u})$	total output SIR,
SINR (u_m)	SINR of output signal u_m,
SINR $(\mathbf{u})$	total output SINR.

4.2. Simulation

In the simulation we have $M_{\mathrm{s}} = 5$ source signals, each being a 4-QAM signal. We compare the behavior of the proposed algorithm with $M = 5$, 10, and 20 sensors. The complex mixing matrices are set up as $\mathbf{A}^{5\times 5} = \mathbf{A}_1$, $\mathbf{A}^{10\times 5} = \left[\mathbf{A}_1^T\ \mathbf{A}_2^T\right]^T$, and $\mathbf{A}^{20\times 5} = \left[\mathbf{A}_1^T\ \mathbf{A}_2^T\ \mathbf{A}_3^T\ \mathbf{A}_4^T\right]^T$, where each $\mathbf{A}_n^{5\times 5}$, $n = 1, \ldots, 4$, is a random complex submatrix with $\|\mathbf{A}_n\|_2 = \sigma_1(\mathbf{A}_n) = 1$, condition number $\chi(\mathbf{A}_n) = \sigma_1(\mathbf{A}_n)/\sigma_{M_{\mathrm{s}}}(\mathbf{A}_n) = 10$ and logarithmically distributed singular values. The respective condition numbers are $\chi(\mathbf{A}^{5\times 5}) = 10$, $\chi(\mathbf{A}^{10\times 5}) = 3.3$, and $\chi(\mathbf{A}^{20\times 5}) = 2.3$. For the update of $\mathbf{W}_{\mathrm{s}}$ we use a block algorithm with (28), block length $L = 100$, step size $\mu = 0.15$, and the Bussgang nonlinearity $\mathbf{g}(\mathbf{u})$ as $g_m(u_m) = u_m|u_m|^2$.

The simulation is set up with $\sigma_s^2/\sigma_n^2 = 15$ dB. Fig. 3 (top) shows SNR (x_m), the resulting SNR's at the true sensors. Fig. 3 (middle) shows SNR (v_m), the SNR's at the virtual sensors after the PCA stage. Fig. 3 (bottom) shows SNR (u_m), the SNR's of the output signals after convergence, i.e. $\mathbf{W} = \mathbf{W}_\infty$. The constellation diagram of u_m after convergence is given in Fig. 5. The learning curves of SINR $(\mathbf{u})$, SIR $(\mathbf{u})$ and SNR $(\mathbf{u})$ are shown in Fig. 4. The input SNR, output SNR, SIR, and SINR after convergence are given in Table 1 together with the values for the minimum-norm

Table 1: SNR's with $M_s = 5$ source signals and $\sigma_s^2/\sigma_n^2 = 15$ dB

System		PCA	ICA: $\mathbf{W}_\infty$			Minimum norm: $\mathbf{W}^{\text{MN}}$			MMSE: $\mathbf{W}^{\text{MMSE}}$		
M	SNR $(\mathbf{x})$	SNR $(\mathbf{v}_s)$	SNR $(\mathbf{u})$	SIR $(\mathbf{u})$	SINR $(\mathbf{u})$	SNR $(\mathbf{u})$	SIR $(\mathbf{u})$	SINR $(\mathbf{u})$	SNR $(\mathbf{u})$	SIR $(\mathbf{u})$	SINR $(\mathbf{u})$
5	9.5 dB	9.5 dB	3.2 dB	11.9 dB	2.6 dB	0.3 dB	∞	0.3 dB	5.0 dB	8.4 dB	3.3 dB
10	9.5 dB	12.6 dB	10.0 dB	24.0 dB	9.8 dB	9.7 dB	∞	9.7 dB	10.1 dB	25.2 dB	9.9 dB
20	9.5 dB	15.6 dB	14.1 dB	24.6 dB	13.8 dB	14.0 dB	∞	14.0 dB	14.1 dB	33.5 dB	14.0 dB

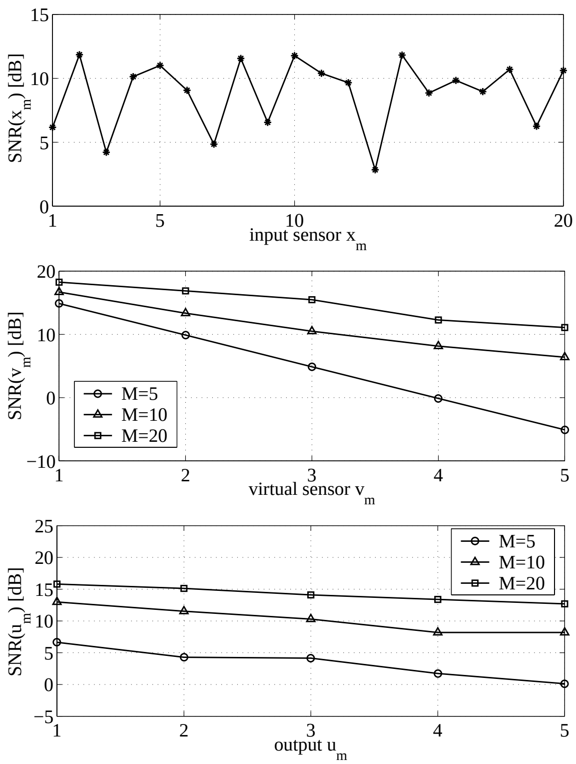

Figure 3: SNR's after convergence for $M = 5$,10, and 20: (top) SNR (x_m), (middle) SNR (v_m), (bottom) SNR (u_m). The output signals u_m are sorted by their SNR.

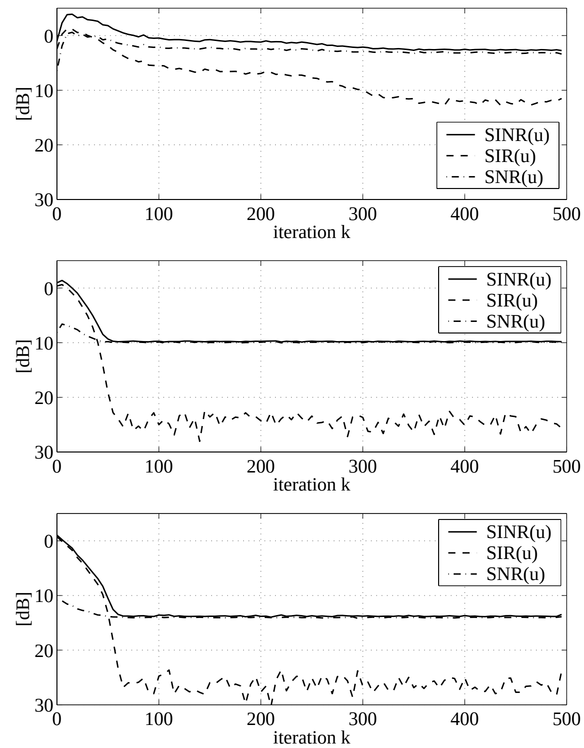

Figure 4: Convergence behavior of SNR $(\mathbf{u})$, SIR $(\mathbf{u})$, and SINR $(\mathbf{u})$ for $M_s = 5$: (top) $M = 5$, (middle) $M = 10$, (bottom) $M = 20$.

solution $\mathbf{W}^{\text{MN}}$ and the MMSE solution $\mathbf{W}^{\text{MMSE}}$ defined in (34) and (41), respectively.

We see that for five sensors, both the separation quality and the convergence time are very poor. Doubling the number of sensors, i.e. $M = 10$, improves the situation considerably. The main difference between 10 and 20 sensors is the higher output SNR, whereas the convergence rate and the separation quality differ only marginally. The improvement of SNR $(\mathbf{u})$ by using more sensors also stems from the fact that $\chi(\mathbf{A})$ is smaller in the case where more sensors are used, as the output SNR improvement is more than just 3 dB for each doubling of the number of sensors. A small singular value σ_m results in a small SNR (v_m), a noisy virtual sensor v_m. Since the blind algorithm under consideration steers all output signals u_m to have equal power, the noisy signal v_m is strongly amplified and therefore causes noisy output signals.

From Table 1 we see that the minimum-norm solution $\mathbf{W}^{\text{MN}}$ always forces perfect separation (SIR $= \infty$), irrespective of the

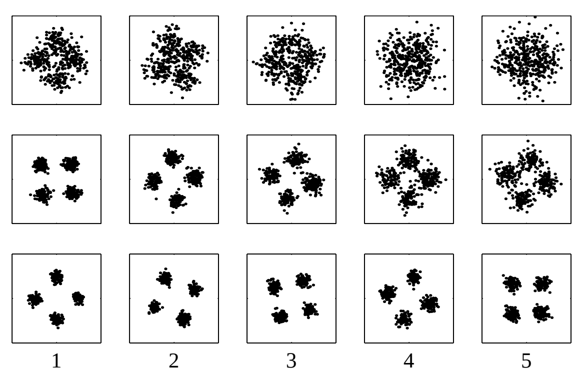

Figure 5: Constellation plot after convergence with $M_s = 5$ 4-QAM signals: (top) $M = 5$, (middle) $M = 10$, (bottom) $M = 20$. The output signals u_m are sorted by their SNR.

resulting output SNR. This leads to a poor SINR for five sensors, caused by the low output SNR. We conclude from this that the minimum-norm solution is not the preferred solution for low SNR $(\mathbf{v}_s)$, especially in a communication system, where minimizing the SINR is of major interest. The MMSE solution $\mathbf{W}^{\text{MMSE}}$ always achieves the highest SINR and SNR, however, typically there is only a small difference to $\mathbf{W}_\infty$. Note that for the adaptive algorithm a further increase of the final SIR can be achieved by reducing the step size μ. Furthermore, from Fig. 4 and Table 1 we see that the final output SINR is mainly limited by the output SNR and not by the output SIR. The sensor noise is the limiting factor for the quality of the output signals.

5. CONCLUSIONS

A two-stage algorithm to solve the overdetermined blind source separation problem with noisy sensors has been proposed. In a preprocessing step, a PCA divides the input space into a signal-plus-noise space and a noise space. The signals in the signal-plus-noise space are then propagated to a subsequent ICA stage. The proposed preprocessing stage is an implicit part of the minimum-norm solution and also of the MMSE solution. Furthermore, the advantage of using more sensors than source signals has been demonstrated by a simulation example, which show that a faster convergence rate as well as a higher steady-state SINR can be achieved.

6. REFERENCES

[1] A. Belouchrani and K. Abed-Meraim, "Constant modulus blind source separation technique: A new approach," in *Int. Symp. Signal Processing App.*, Gold Coast, Australia, August 25–30 1996, pp. 232–235.

[2] L. Zhang, S.-I. Amari, and A. Cichocki, "Natural gradient approach to blind separation of over- and undercomplete mixtures," in *Proc. ICA*, Aussois, France, Jan. 11–15, 1999, pp. 455–460.

[3] L.-Q. Zhang, A. Cichocki, and S. Amari, "Natural gradient algorithm for blind separation of overdetermined mixture with additive noise," *IEEE Signal Processing Lett.*, pp. 293–295, Nov. 1999.

[4] S.-I. Amari, "Natural gradient learning for over- and undercomplete bases in ICA," *Neural Computation*, vol. 11, no. 8, pp. 1875–1883, Nov. 1999.

[5] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 3rd edition, 1991.

[6] S.-I. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.

[7] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[8] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[9] R. H. Lambert and C. L. Nikias, "Blind deconvolution of multipath mixtures," in *Unsupervised Adaptive Filtering*, S. Haykin, Ed. 2000, vol. 1, John Wiley & Sons.

[10] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[11] T.-W. Lee, *Independent Component Analysis*, Kluwer Academic Publishers, 1998.

[12] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[13] S. Haykin, *Adaptive Filter Theory*, Prentice Hall, 3rd edition, 1996.

[14] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, 1990.

[15] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[16] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 33, pp. 387–392, Apr. 1985.

[17] W. Kellermann, "Echoes and noise with seamless acoustic man-machine interfaces – the challenge persists," in *Proc. IWAENC*, Pocono Manor, Pennsylvania, USA, Sept. 27–30 1999, pp. 1–7.

[18] S. C. Douglas, A. Cichocki, and S. Amari, "Bias removal technique for blind source separation with noisy measurements," *Electron. Lett.*, vol. 34, no. 14, pp. 1379–1380, July 1998.

[19] R. H. Lambert, "Difficulty measures and figures of merit for source separation," in *Proc. ICA*, Aussois, France, Jan. 11–15, 1999, pp. 133–138.

[20] M. Joho and H. Mathis, "Performance comparison of combined blind/non-blind source separation algorithms," in *Proc. ICA*, Aussois, France, Jan. 11–15, 1999, pp. 139–142.

[21] M. Joho, H. Mathis, and G. S. Moschytz, "An FFT-based algorithm for multichannel blind deconvolution," in *Proc. ISCAS*, Orlando, FL, May 30 – June 2, 1999, pp. III–203–206.

[22] P. Comon, C. Jutten, and J. Hérault, "Blind separation of sources, part II: Problems statement," *Signal Processing*, vol. 24, pp. 11–20, July 1991.

[23] S.-I. Amari and A. Cichocki, "Adaptive blind signal processing—neural network appraoches," *Proc. IEEE*, vol. 86, no. 10, pp. 2026–2048, Oct. 1998.