# A Blind Source Separation Technique Using Second-Order Statistics

Adel Belouchrani, *Member, IEEE*, Karim Abed-Meraim, Jean-François Cardoso, *Member, IEEE*, and Eric Moulines, *Member, IEEE*

***Abstract*—Separation of sources consists of recovering a set of signals of which only instantaneous linear mixtures are observed. In many situations, no *a priori* information on the mixing matrix is available: The linear mixture should be "blindly" processed. This typically occurs in narrowband array processing applications when the array manifold is unknown or distorted.**

**This paper introduces a new source separation technique exploiting the time coherence of the source signals. In contrast with other previously reported techniques, the proposed approach relies only on stationary second-order statistics that are based on a joint diagonalization of a set of covariance matrices. Asymptotic performance analysis of this method is carried out; some numerical simulations are provided to illustrate the effectiveness of the proposed method.**

## I. Introduction

In many situations of practical interest, one has to process multidimensional observations of the form

$$\mathbf{x}(t) = \mathbf{y}(t) + \mathbf{n}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{1}$$

i.e., $\mathbf{x}(t)$ is a noisy instantaneous linear mixture of source signals. This model is commonplace in the field of narrowband array processing. In this context, vector $\mathbf{s}(t) = [s_1(t), \cdots, s_n(t)]^T$ contains the signals emitted by $n$ narrowband sources, vector $\mathbf{y}(t) = [y_1(t), \cdots, y_n(t)]^T$ contains the array output sampled at time $t$, and matrix $\mathbf{A}$ is the transfer function between sources and sensors. In the following, it is referred to as the "array matrix" or the "mixing matrix."

Most array processing techniques rely on the modeling of $\mathbf{A}$: Each column of $\mathbf{A}$ is assumed to depend on a small number of parameters. This information may be provided either by physical modeling (for example, when the array geometry is known and when the sources are in the far field of the array) or by direct array calibration. In many circumstances, however, this information is not available or not reliable.

Blind source separation consists in identifying $\mathbf{A}$ and/or retrieving the source signals without resorting to any *a priori* information about mixing matrix $\mathbf{A}$; it exploits only the information carried by the received signals themselves, hence, the term *blind*. Performance of such blind techniques is, by nature, essentially unaffected by potential errors in the propagation model or in array calibration (this is obviously not the case of parametric array processing technique; see, for example, [1] and [2]). Of course, the lack of information on the structure of $\mathbf{A}$ must be compensated by some additional assumptions on source signals.

For non-Gaussian independent sources, the first approach traces back to the pioneering adaptive algorithm of Jutten–Hrault [3] (see also [4]–[7]). Batch algorithms, which are based mainly on higher order cumulants, were developed later; see, for instance, [8]–[11]. These algorithms exploit only the marginal distribution of the observations. Thus, they are suitable even when source signals are temporally independent. Otherwise, other approaches can be developed based on temporal correlations. Since these are second-order statistics, they are expected to be more robust in adverse signal to noise ratios.

For cyclostationary emitters, like those encountered in digital or analog communication systems, a sound approach consists of exploiting spectral redundancy at the cyclic frequency of the sources of interest, as proposed in [12]. However, these methods crucially rely on the assumption that the different sources have different cyclostationary features [13]. In addition, when the cyclic frequencies are not known in advance, they must be estimated.

A different context is considered herein: stationary sources with different spectral contents. It has already been shown that blind identification is feasible based on spatial covariance matrices [14]–[17]. These matrices (see below) show a simple structure that allows straightforward blind identification procedures based on eigendecomposition. In this paper, we introduce a blind identification technique based on a joint diagonalization of *several* covariance matrices. Robustness is significantly increased at low additional cost by processing such a matrix set rather than a unique matrix as in [14].

The paper is organized as follows. In Section II, the problem of blind source separation is stated together with the relevant hypothesis. Section III presents a second-order blind identification technique based on "joint diagonalization" of a set of spatial covariance matrices; an efficient Jacobi-like algorithm for solving this problem is described in Appendix A. In Section IV, a closed-form expression of the asymptotic

Manuscript received November 29, 1994; revised June 18, 1996. The associate editor coordinating the review of this paper and approving it for publication was Prof. Roger S. Cheng.

A. Belouchrani is with the Department of Electrical and Computer Engineering, Villanova University, Villanova, PA 19085 USA.

K. Abed-Meraim is with Department of Electrical and Electronic Engineering, University of Melbourne, Victoria 3052, Australia.

J.-F. Cardoso and E. Moulines are with the Ecole Nationale Supérieure des Télécommunications, Département Signal, 75634 Paris Cedex 13, France.

Publisher Item Identifier S 1053-587X(97)01178-1.

performance of the proposed method is derived. Numerical simulation illustrating the validity of this method are presented in Section V.

## II. Problem Formulation

### A. Assumptions

We start by specifying the signal. It is assumed that the source signal vector $\mathbf{s}(t)$ is either H1) a deterministic ergodic sequence or H2) a stationary multivariate process with

H1)

$$\lim_{T\to\infty} T^{-1} \sum_{t=1,\,T} \mathbf{s}(t+\tau)\mathbf{s}(t)^*$$
$$\stackrel{\text{def}}{=} E[\mathbf{s}(t+\tau)\mathbf{s}(t)^*] \tag{2}$$
$$= \text{diag}\,[\rho_1(\tau), \cdots, \rho_n(\tau)] \tag{3}$$

H2)

$$E[\mathbf{s}(t+\tau)\mathbf{s}(t)^*] = \text{diag}\,[\rho_1(\tau), \cdots, \rho_n(\tau)] \tag{4}$$

where superscript * denotes the conjugate transpose of a vector, and diag $[\cdot]$ is the diagonal matrix formed with the elements of its vector valued argument. For convenience, the same notation $E$ is used for the deterministic averaging operation under hypothesis H1) and for ensemble averaging under H2). This convention holds throughout. Assumptions H1) or H2) mean that the component processes $\mathbf{s}_i(t)$, $1 \le i \le n$ are mutually uncorrelated, and $\rho_i(\tau) = E[\mathbf{s}_i(t+\tau)\mathbf{s}_i^*(t)]$ denotes the autocovariance of $\mathbf{s}_i(t)$.

The additive noise $\mathbf{n}(t)$ is modeled as a stationary, temporally white, zero-mean complex random process independent of the source signals. For simplicity, we also require $\mathbf{n}(t)$ to be spatially white, i.e.,

$$E[\mathbf{n}(t+\tau)\mathbf{n}^*(t)] = \sigma^2\delta(\tau)\mathbf{I} \tag{5}$$

where $\delta(\tau)$ is the Kronecker delta, and $I$ denotes the identity matrix. The assumption of spatially white noise is not crucial: The method presented below may be extended to the case of an unknown noise covariance matrix (see Section III-A).

The $m \times n$ complex matrix $\mathbf{A}$ is assumed to have full column rank but is otherwise unknown. In contrast with traditional parametric methods, no specific array geometry or sensor characteristics are assumed, i.e., the array manifold is unknown.

Under the above assumptions, the covariance matrices of the array output take the following structure:

$$\mathbf{R}(0) = E[\mathbf{x}(t)\mathbf{x}^*(t)] = \mathbf{A}\mathbf{R}_s(0)\mathbf{A}^H + \sigma^2\mathbf{I} \tag{6}$$
$$\mathbf{R}(\tau) = E[\mathbf{x}(t+\tau)\mathbf{x}^*(t)] = \mathbf{A}\mathbf{R}_s(\tau)\mathbf{A}^H \qquad \tau \neq 0 \tag{7}$$

where superscript $H$ denotes the complex conjugate transpose of a matrix. The aim of blind source separation is to identify the mixture matrix and/or to recover the source signals from the array output $\mathbf{x}(t)$ *without any a priori knowledge of the array manifold*. The potential benefit of such a *blind* approach is that source separation is essentially unaffected by errors in the propagation model or in array calibration.

### B. Blind Identifiability

Before proceeding, it is important to specify the notion of blind identification. In the blind context, a full identification of the mixture matrix $\mathbf{A}$ is impossible because the exchange of a fixed scalar factor between a given source signal and the corresponding column of $\mathbf{A}$ does not affect the observations, as is shown by the following relation:

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) = \sum_{p=1}^{n} \frac{\mathbf{a}_p}{\alpha_p}\alpha_p s_p(t) + \mathbf{n}(t) \tag{8}$$

where $\alpha_p$ is an arbitrary complex factor, and $\mathbf{a}_p$ denotes the $p$th column of $\mathbf{A}$.

Advantage can be taken of this indeterminacy by assuming, *without any loss of generality*, that the source signals have unit variance so that the dynamic range of the sources is accounted for by the magnitude of the corresponding columns of $\mathbf{A}$. This normalization *convention* turns out to be convenient in the sequel; it does not affect the performance results presented below. Since the sources are assumed to be uncorrelated, we have

$$\mathbf{R}_s(0) = \mathbf{I} \qquad \text{so that } \mathbf{R}_y(0) \stackrel{\text{def}}{=} E[\mathbf{y}(t)\mathbf{y}^*(t)] = \mathbf{A}\mathbf{A}^H. \tag{9}$$

This normalization still leaves undetermined the ordering and the phases of the columns of $\mathbf{A}$. The following definition is then in order:

*Definition 1:* Two matrices $\mathbf{M}$ and $\mathbf{N}$ are said to be *essentially equal* if there exists a matrix $\mathbf{P}$ such that $\mathbf{M} = \mathbf{N}\mathbf{P}$, where $\mathbf{P}$ has exactly one nonzero entry in each row and column, where these entries have unit modulus. This is denoted $\mathbf{M} \doteq \mathbf{N}$.

In this paper, blind identification of $\mathbf{A}$ is understood as the determination of a matrix essentially equal to $\mathbf{A}$. Of course, these indeterminacies do not impede source separation: If the mixture matrix $\mathbf{A}$ is estimated up to permutation and phase shifts, it is still possible to determine the source signals up to the corresponding fixed permutation and phase shifts (blind identifiability is discussed at length in [18]).

## III. A Second-Order Identification Approach

### A. Whitening

The first step of our procedure consists of whitening the signal part $\mathbf{y}(t)$ of the observation. This is achieved by applying to $\mathbf{y}(t)$ a *whitening matrix* $\mathbf{W}$, i.e., a $n \times m$ matrix verifying:

$$E[\mathbf{W}\mathbf{y}(t)\mathbf{y}(t)^*\mathbf{W}^H] = \mathbf{W}\mathbf{R}_y(0)\mathbf{W}^H = \mathbf{W}\mathbf{A}\mathbf{A}^H\mathbf{W}^H = \mathbf{I}. \tag{10}$$

Equation (10) shows that if $\mathbf{W}$ is a whitening matrix, then $\mathbf{W}\mathbf{A}$ is a $n \times n$ unitary matrix. It follows that for any whitening matrix $\mathbf{W}$, there exists a $n \times n$ unitary matrix $\mathbf{U}$ such that $\mathbf{W}\mathbf{A} = \mathbf{U}$. As a consequence, matrix $\mathbf{A}$ can be factored as

$$\mathbf{A} = \mathbf{W}^{\#}\mathbf{U} \tag{11}$$

where superscript $\#$ denotes the Moore–Penrose pseudoinverse. We note that this whitening procedure reduces the determination of the $m \times n$ mixture matrix $\mathbf{A}$ to that of a unitary $n \times n$ matrix $\mathbf{U}$. The whitened process $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$ still obeys a linear model:

$$\mathbf{z}(t) \overset{\text{def}}{=} \mathbf{W}\mathbf{x}(t) = \mathbf{W}[\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)] = \mathbf{U}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t) \quad (12)$$

where the signal part of the whitened process now is a “unitary mixture” of the source signals. Note that all information contained in the covariance is “exhausted” after the whitening procedure in the sense that changing $\mathbf{U}$ in (12) to any other unitary matrix leaves the covariance of $\mathbf{z}(t)$ unchanged. Note also that besides whitening, the signal part of the observations (multiplication by a whitening matrix $\mathbf{W}$) reduces the array output to a $n$-dimensional vector.

Since we have, from (6) and (9), $\mathbf{A}\mathbf{A}^H = \mathbf{R}(0) - \sigma^2\mathbf{I}$, (10) shows that a whitening matrix $\mathbf{W}$ can be determined from the array output covariance $\mathbf{R}(0)$, provided the noise covariance matrix is known or can be estimated. A whitening matrix may also be determined from a linear combination of a set of covariance matrices taken at nonzero time lags, as suggested in [14]. In any case, as shown by (11), finding a whitening matrix still leaves undetermined a unitary factor in $\mathbf{A}$. This “missing factor” $\mathbf{U}$ can be determined from higher order statistics as investigated in [8]–[10]. Exploiting the time dependence structure (hypothesis H1 or H2), it may be also retrieved from covariance matrices at non zero lags, as explained below.

### *B. Determining the Unitary Factor*

Consider the spatially whitened covariance matrices $\underline{\mathbf{R}}(\tau)$ defined as

$$\forall \tau \neq 0 \quad \underline{\mathbf{R}}(\tau) = \mathbf{W}\mathbf{R}(\tau)\mathbf{W}^H. \quad (13)$$

These $n \times n$ complex matrices are nothing but the covariance matrices of the process $\mathbf{z}(t)$. By (7) and (11), we obtain the key relation

$$\forall \tau \neq 0 \quad \underline{\mathbf{R}}(\tau) = \mathbf{U}\mathbf{R}_s(\tau)\mathbf{U}^H. \quad (14)$$

Since $\mathbf{U}$ is unitary and $\mathbf{R}_s(\tau)$ is diagonal, the latter means that we have the following property.

*Property:* Any whitened covariance matrix is diagonalized by the unitary transform $\mathbf{U}$.

As a consequence, the unitary factor $\mathbf{U}$ may be obtained as a unitary diagonalizing matrix of a whitened covariance matrix $\underline{\mathbf{R}}(\tau)$ for some lag $\tau$. More formally, we have the following theorem.

*Theorem 1—First Uniqueness Condition:* Let $\tau$ be a nonzero time lag and $\mathbf{V}$ be a unitary matrix such that

$$\mathbf{V}^H\underline{\mathbf{R}}(\tau)\mathbf{V} = \text{diag}\,[d_1, \cdots, d_n] \quad (15)$$
$$\forall 1 \leq i \neq j \leq n \quad \rho_i(\tau) \neq \rho_j(\tau). \quad (16)$$

Then, we have the following:

- $\mathbf{V}$ is essentially equal to $\mathbf{U}$: $\mathbf{V} \doteq \mathbf{U}$.
- A permutation $\sigma$ on $\{1, \cdots, n\}$ exists such that $[\rho_1(\tau), \cdots, \rho_n(\tau)] = [d_{\sigma(1)}, \cdots, d_{\sigma(n)}]$.

This property is a direct consequence of the spectral theorem for normal matrices (see for example [19, Theorem 2.5.4]). We recall that an $n \times n$ matrix $\mathbf{M}$ is said to be *normal* if $\mathbf{M}\mathbf{M}^H = \mathbf{M}^H\mathbf{M}$. The spectral theorem states that a normal matrix $\mathbf{M}$ is *unitarily diagonalizable*, i.e., there exists a unitary matrix $\mathbf{U}$ and a diagonal matrix $\mathbf{D}$ such that $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{U}^H$. In our setting, the existence of the unitary matrix $\mathbf{V}$ (15) for any time lag $\tau$ is guaranteed by (14). In contrast, the existence of a time lag $\tau \neq 0$ such that (16) holds is not trivial and cannot be checked *a priori*. Note that the indeterminacies (phase shifts and permutations) in the diagonalization of a normal matrix correspond precisely to those encountered in the blind source separation problem. Thus, the diagonalization of $\underline{\mathbf{R}}(\tau)$ for a delay $\tau$ yields the relevant parameters if $\underline{\mathbf{R}}(\tau)$ has *distinct eigenvalues.* This identification scheme may be found in slightly different forms in [14] and in [15].

True indeterminacy arises in the case of degenerate eigenvalues. It does not seem possible to determine *a priori* a time lag $\tau$ such that the eigenvalues of $\underline{\mathbf{R}}(\tau)$ are distinct. Of course, if the source signals have different spectral shapes, eigenvalue degeneracy is unlikely, but the problem is not purely academic because it is to be expected that when an eigenvalue of $\underline{\mathbf{R}}(\tau)$ comes close to degeneracy, the robustness of determining $\mathbf{U}$ from an eigendecomposition is seriously affected.

The situation is more favorable if we consider simultaneous diagonalization of a set $\{\underline{\mathbf{R}}(\tau_i) | i = 1, \cdots, K\}$ of $K$ whitened covariance matrices.

*Theorem 2—Second Uniqueness Condition:* Let $\tau_1, \tau_2, \cdots, j\tau_K$ be $K$ nonzero time lags, and let $\mathbf{V}$ be a unitary matrix such that

$$\forall 1 \leq k \leq K \quad \mathbf{V}^H\underline{\mathbf{R}}(\tau_k)\mathbf{V} = \text{diag}\,[d_1(k), \cdots, d_n(k)] \quad (17)$$
$$\forall 1 \leq i \neq j \leq n \quad \exists k, 1 \leq k \leq K \quad d_i(k) \neq d_j(k). \quad (18)$$

Then, we have the following:

- $\mathbf{V}$ is essentially equal to $\mathbf{U}$: $\mathbf{V} \doteq \mathbf{U}$.
- A permutation $\sigma$ on $\{1, \cdots, n\}$ exists such that

$$[\rho_1(\tau_k), \cdots, \rho_n(\tau_k)] = [d_{\sigma(1)}(k), \cdots, d_{\sigma(n)}(k)]$$
$$1 \leq k \leq K.$$

This is a consequence of the essential uniqueness of joint diagonalization: see Theorem 3 below. Again, the existence of a unitary matrix $\mathbf{V}$ that simultaneously diagonalizes the set of covariance matrices $[\underline{\mathbf{R}}(\tau_1), \cdots, \underline{\mathbf{R}}(\tau_K)]$ is guaranteed for any choice of time lags thanks to (14). Even though (18) is much weaker than (16), it is not necessarily true; in particular, in the trivial case where the sources show identical normalized spectra, the mixing matrix $\mathbf{A}$ cannot be identified by resorting to Theorem 2. Conversely, when the source signals have different normalized spectra, it is always possible to find a set of time lags $\tau_1, \cdots, \tau_K$ such that (18) is met. This corresponds to the second-order identifiability condition found in [18].

The main point of our contribution is to consider the *joint diagonalization* of several covariance matrices. This approach is intended to reduce the probability that an unfortunate choice of time lag $\tau$ results in unidentifiability of $\mathbf{U}$ from $\underline{\mathbf{R}}(\tau)$; more importantly, this approach generally increases the statistical efficiency of the procedure by inferring the value of $\mathbf{U}$ from a larger set of statistics.

### *C. Joint Diagonalization*

In numerical analysis, the "off" of an $n \times n$ matrix $\mathbf{M}$ with entries $M_{ij}$ is defined as

$$\mathbf{off}\,(\mathbf{M}) \stackrel{\text{def}}{=} \sum_{1 \le i \ne j \le n} |M_{ij}|^2 \tag{19}$$

and the unitary diagonalization of a matrix $\mathbf{M}$ is equivalent to zeroing $\mathbf{off}\,(\mathbf{V}^H\mathbf{M}\mathbf{V})$ by some unitary matrix $\mathbf{V}$. As recalled above, the spectral theorem states that only normal matrices can be unitarily diagonalized. In addition, if a matrix $\mathbf{M}$ is in the form $\mathbf{M} = \mathbf{U}\mathbf{D}\mathbf{U}^H$, where $\mathbf{U}$ is unitary and $\mathbf{D}$ is diagonal with distinct diagonal elements, then it may be unitarily diagonalized only by unitary matrices that are essentially equal to $\mathbf{U}$, that is, if $\mathbf{off}\,(\mathbf{V}^H\mathbf{M}\mathbf{V}) = 0$, then $\mathbf{V} \doteq \mathbf{U}$.

Consider a set $\mathcal{M} = \{\mathbf{M}_I, \cdots, \mathbf{M}_K\}$ of $K$ matrices of size $n \times n$. The "joint diagonality" (JD) criterion is defined, for any $n \times n$ matrix $\mathbf{V}$, as the following nonnegative function of $\mathbf{V}$:

$$\mathcal{C}(\mathcal{M}, \mathbf{V}) \stackrel{\text{def}}{=} \sum_{k=1, K} \mathbf{off}\,(\mathbf{V}^H\mathbf{M}_k\mathbf{V}). \tag{20}$$

A unitary matrix is said to be a *joint diagonalizer* of the set $\mathcal{M}$ if it minimizes the JD criterion (20) over the set of all unitary matrices.

Let us first consider the case where each matrix in the set $\mathcal{M}$ is in the form $\mathbf{M}_k = \mathbf{U}\mathbf{D}_k\mathbf{U}^H$, where $\mathbf{D}_k$ is a diagonal matrix. Then clearly, $\mathcal{C}(\mathcal{M}, \mathbf{U}) = 0$, and this is the global minimum of the JD criterion (20) since $\mathcal{C}(\mathcal{M}, \mathbf{V}) \ge 0$ for any matrix $\mathbf{V}$. Thus, if each matrix in the set $\mathcal{M}$ can be unitarily diagonalized by $\mathbf{U}$, then according to our definition, matrix $\mathbf{U}$ is a joint diagonalizer of $\mathcal{M}$. This is of little interest; we are more interested in the uniqueness of a joint diagonalizer. We have the following

*Theorem 3—Essential Uniqueness of Joint Diagonalization:* Let $\mathcal{M} = \{\mathbf{M}_1, \cdots, \mathbf{M}_K\}$ be a set of $K$ matrices where, for $1 \le k \le K$, matrix $\mathbf{M}_k$ is in the form $\mathbf{M}_k = \mathbf{U}\mathbf{D}_k\mathbf{U}^H$ with $\mathbf{U}$ a unitary matrix, and $\mathbf{D}_k = \text{diag}\,[d_1(k), \cdots, d_n(k)]$. Any joint diagonalizer of $\mathcal{M}$ is essentially equal to $\mathbf{U}$ if and only if

$$\forall\, 1 \le i \ne j \le n \quad \exists\, k,\ 1 \le k \le K \quad d_i(k) \ne d_j(k). \tag{21}$$

The essential uniqueness condition (21) is of course much weaker than the requirement that each matrix in $\mathcal{M}$ is uniquely unitarily diagonalizable. In particular, it is easy to construct examples where each matrix in $\mathcal{M}$ has a degenerate eigenvalue spectrum but such that the joint diagonalizer of $\mathcal{M}$ is nonetheless essentially unique. The proof of Theorem 3 is given in Section VII-B.

An important feature of our definition of joint diagonalization is that it is *not* required that the matrix set under consideration can be exactly simultaneously diagonalized by a single unitary matrix. As a matter of fact, it is not even required that the matrices in the set are *individually* unitarily diagonalizable. This is because we do not require that the "off" of all the matrices are cancelled by a unitary transform; a joint diagonalizer is just a minimizer of the JD criterion. If the matrices in $\mathcal{M}$ are not in the form considered in Theorem 3, the JD criterion cannot be zeroed, and the matrices can only be approximately jointly diagonalized. Hence, an (approximate) joint diagonalizer defines a kind of an "average eigen-structure." This is particularly convenient for statistical inference where the structural information is to be extracted from sample statistics: Even though the true covariance matrices considered above can be exactly simultaneously diagonalized, their sample counterparts cannot because of the estimation errors. Hence, rather than exactly diagonalizing a single covariance matrix, the approximate joint diagonalization allows the information contained in a set of covariance matrices to be integrated in a single unitary matrix.

Another important feature of the (possibly approximate) joint diagonalization is the existence of a numerically efficient algorithm for its computation. This algorithm is a generalization of the Jacobi technique for the exact diagonalization of a single Hermitian matrix [20]. This technique consists of computing the unitary diagonalizer as a product of Givens rotations. It turns out that the Givens rotation parameters can be simply computed even when the matrices to be jointly diagonalized do not show any symmetry property. This is particularly convenient for processing sample covariance matrices that have no reason to be exactly normal. The extension of the Jacobi technique to approximate joint diagonalization is described in Section VII-A.

### *D. Implementation of the SOBI Algorithm*

Based on the previous sections, we can introduce a second-order blind identification (SOBI) algorithm. SOBI is defined by the following implementation:

1) Estimate the sample covariance $\hat{\mathbf{R}}(0)$ from $T$ data samples. Denote by $\lambda_1, \cdots, \lambda_n$ the $n$ largest eigenvalues and $\mathbf{h}_1, \cdots, \mathbf{h}_n$ the corresponding eigenvectors of $\hat{\mathbf{R}}(0)$.
2) Under the white noise assumption, an estimate $\hat{\sigma}^2$ of the noise variance is the average of the $m - n$ smallest eigenvalues of $\hat{\mathbf{R}}(0)$. The whitened signals are $\mathbf{z}(t) = [z_1(t), \cdots, z_n(t)]^T$, which are computed by $z_i(t) = (\lambda_i - \hat{\sigma}^2)^{-(1/2)}\mathbf{h}_i^*\mathbf{x}(t)$ for $1 \le i \le n$. This is equivalent to forming a whitening matrix by
$$\hat{\mathbf{W}} = [(\lambda_1 - \hat{\sigma}^2)^{-(1/2)}\mathbf{h}_1, \cdots, (\lambda_n - \hat{\sigma}^2)^{-(1/2)}\mathbf{h}_n]^H.$$
3) Form sample estimates $\hat{\underline{\mathbf{R}}}(\tau)$ by computing the sample covariance matrices of $\mathbf{z}(t)$ for a fixed set of time lags $\tau \in \{\tau_j \,|\, j = 1, \cdots, K\}$.
4) A unitary matrix $\hat{\mathbf{U}}$ is then obtained as joint diagonalizer of the set $\{\hat{\underline{\mathbf{R}}}(\tau_j) \,|\, j = 1, \cdots, K\}$.
5) The source signals are estimated as $\hat{\mathbf{s}}(t) = \hat{\mathbf{U}}^H\hat{\mathbf{W}}\mathbf{x}(t)$, and/or the mixing matrix $\mathbf{A}$ is estimated as $\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#}\hat{\mathbf{U}}$.

## IV. Asymptotic Performance Analysis

In this section, an asymptotic performance analysis of the proposed method is carried out. To ease the derivations, we make the following additional assumptions.

H1') Each source signal $s_i(t)$ is a circular stationary Gaussian process: $E[s_i(t+\tau)s_i(t)] = 0$ for $1 \le i \le n$ and for any time lag $\tau$.

H2') The source signals $s_i(t)$ are mutually independent and are independent of the noise $\mathbf{n}(t)$.

H3) The source signals are short range dependent, in the sense that $\sum_{\tau \in \mathbb{Z}} |\tau \rho_i(\tau)| < \infty$.

Hypothesis H3) is an extremely mild condition that is verified, for example, by all AR or ARMA processes. To get rid of phase and permutation indeterminacies, we shall assume that they are fixed in such a way that the matrix estimator $\hat{\mathbf{A}}$ is close to the true mixture matrix $\mathbf{A}$ rather than to some other matrix essentially equal to $\mathbf{A}$. In addition, the covariance matrices are computed at time lags $\tau_1, \cdots, \tau_K$ such that the uniqueness condition of Theorem 2 is verified.

### A. Performance Index

Rather than estimating the variance of the coefficients of the mixing matrix, it is more relevant to source separation to compute an index that quantifies the performance in terms of interference rejection, as follows. Assume that at each time instant $t$ an estimate of the vector of source signals is computed by applying to the received signal $\mathbf{x}(t)$ the pseudoinverse of the estimated mixture matrix, i.e.,

$$\hat{\mathbf{s}}(t) = \hat{\mathbf{A}}^{\#}\mathbf{x}(t) = \hat{\mathbf{A}}\mathbf{s}(t) + \hat{\mathbf{A}}^{\#}\mathbf{n}(t) \tag{22}$$

where $\hat{\mathbf{A}}^{\#}$ is given by $\hat{\mathbf{A}}^{\#} = \hat{\mathbf{U}}^H\hat{\mathbf{W}}$. We stress that in general, this procedure is not optimal for recovering the source signals based on an estimate $\hat{\mathbf{A}}$. For large enough sample size $T$, matrix $\hat{\mathbf{A}}$ should be close to the true mixing matrix $\mathbf{A}$ so that $\hat{\mathbf{A}}^{\#}\mathbf{A}$ is close to the identity matrix. The performance index used in the sequel is the interference to signal ratio (ISR), which is defined as

$$\mathcal{I}_{pq} = E|(\hat{\mathbf{A}}^{\#}\mathbf{A})_{pq}|^2. \tag{23}$$

This actually defines an ISR because, by our normalization convention (9), we have $\mathcal{I}_{pp} \simeq 1$ for large enough $T$. Thus, $\mathcal{I}_{pq}$ measures the ratio of the power of the interference of the $q$th source to the power of the $p$th source signal estimated as in (22). As a measure of the overall quality of the separation, we also define a global rejection level:

$$\mathcal{I}_{perf} \stackrel{\text{def}}{=} \sum_{q \ne p} \mathcal{I}_{pq}. \tag{24}$$

### B. Outline of Performance Analysis

The asymptotic variance of the estimates of $\mathbf{A}$ is expected to decrease as $1/T$ thanks to the short range dependence of the observed process (assumption H3). Thus, the leading term of $\mathcal{I}_{pq}$ is of order $T^{-1}$. The purpose of this section is to give its closed-form expression. Detailed computations are not reported herein due to lack of space. We rather outline the computation below and give further details in the Appendix.

Asymptotic performance is obtained along the following lines. We note that matrix estimate $\hat{\mathbf{A}}$ is a "function" of the sample covariance matrices $[\hat{\mathbf{R}}(0), \hat{\mathbf{R}}(\tau_1), \cdots, \hat{\mathbf{R}}(\tau_k)]$ of the observed signal $\mathbf{x}(t)$. The computation can then proceed in two steps; first, we express the asymptotic moments of the sample covariance matrices (see Lemma 1); second, we compute the leading term (in the sample covariance matrices) in the Taylor expansion of $|(\hat{\mathbf{A}}^{\#}\mathbf{A})_{pq}|^2$ (see Lemma 2). The final result is obtained by combining these two expressions.

*Lemma 1:* Under conditions H1', H2', and H3 and for any matrices $\mathbf{M}$ and $\mathbf{N}$ in $\mathbb{C}^{m \times m}$

$$\begin{aligned}
&\lim_{T\to\infty} TE\,\mathrm{Tr}\,\{\delta\mathbf{R}(\tau_k)\mathbf{M}\,\delta\mathbf{R}(\tau_l)\mathbf{N}\}\\
&\quad = \sum_{\tau\in\mathbb{Z}} \mathrm{Tr}\,\{\mathbf{R}(\tau_k+\tau)\mathbf{N}\}\,\mathrm{Tr}\,\{\mathbf{R}(\tau_l-\tau)\mathbf{M}\}\\
&\lim_{T\to\infty} TE\,\mathrm{Tr}\,\{\delta\mathbf{R}(\tau_k)\mathbf{M}\}\,\mathrm{Tr}\,\{\delta\mathbf{R}(\tau_l)\mathbf{N}\}\\
&\quad = \sum_{\tau\in\mathbb{Z}} \mathrm{Tr}\,\{\mathbf{M}\mathbf{R}(\tau_k+\tau)\mathbf{N}\mathbf{R}(\tau_l-\tau)\}\\
&\lim_{T\to\infty} TE\,\|\delta\mathbf{R}(\tau_k)\|^3 = 0
\end{aligned}$$

where $\delta\mathbf{R}(\tau_k) = \hat{\mathbf{R}}(\tau_k) - \mathbf{R}(\tau_k)$, and $\mathrm{Tr}\,\{\mathbf{M}\}$ denotes the trace of matrix $\mathbf{M}$.

*Proof:* See [21].

*Lemma 2:* The Taylor expansion of $|(\hat{\mathbf{A}}^{\#}\mathbf{A})_{pq}|^2$ is given for $p \ne q$ by

$$\begin{aligned}
|(\hat{\mathbf{A}}^{\#}\mathbf{A})_{pq}|^2 &= |\alpha_{pq}(0)C_{pq}(0)|^2 + \sum_{1\le|k|\le K} \alpha_{pq}(0)\alpha_{pq}(k)\\
&\quad \cdot [C_{qp}(0)C_{pq}(k) + C_{pq}(0)C_{qp}(k)]\\
&\quad + \sum_{1\le|k|,\,|l|\le K} \alpha_{pq}(k)\alpha_{pq}(l)C_{pq}(k)C_{qp}(l)\\
&\quad + O\left[\sum_{k=0}^{K} \|\delta\mathbf{R}(\tau_k)\|^3\right]
\end{aligned}$$

with

$$\begin{aligned}
\rho_r &= [\rho_r(\tau_1), \cdots, \rho_r(\tau_K)]^T\\
\alpha_{pq}(0) &= 1 + \frac{|\rho_p|^2 - |\rho_q|^2}{|\rho_p - \rho_q|^2}\\
\alpha_{pq}(k) &= \frac{\rho_p^*(\tau_k) - \rho_q^*(\tau_k)}{|\rho_p - \rho_q|^2} \qquad \text{for } k \ne 0\\
\mathbf{C}(0) &= -\frac{1}{2}\,\mathbf{A}^{\#}\delta\mathbf{R}(0)\mathbf{A}^{\#^H} + \frac{\mathrm{Tr}\,[\Pi\delta\mathbf{R}(0)]}{2(m-n)}\,(\mathbf{A}^H\mathbf{A})^{-1}\\
\mathbf{C}(k) &= \tfrac{1}{2}\mathbf{A}^{\#}\delta\mathbf{R}(\tau_k)\mathbf{A}^{\#^H} \qquad \text{for } k \ne 0
\end{aligned}$$

where $\Pi$ denotes the orthogonal projector on the noise subspace (i.e., the subspace orthogonal to the range of matrix $\mathbf{A}$), and $C_{pq}(.)$ are the $pq$th element of the matrix $\mathbf{C}(.)$.

*Proof:* See Section VII-C.

According to Lemma 2, the expectation of $|(\hat{\mathbf{A}}^{\#}\mathbf{A})_{pq}|^2$ can be computed from the expectations of $|C_{pq}(0)|^2$, $C_{qp}(0)C_{pq}(k)$, and $C_{qp}(l)C_{pq}(k)$. Lemma 1 reduces this computation to simple algebra, yielding

$$\begin{aligned}
E[|C_{pq}(0)|^2] &= \frac{1}{4T}\left[D_{pq}(0) + \sigma^2(J_{pp}+J_{qq})\right.\\
&\quad \left. + \sigma^4\left(\frac{|J_{pq}|^2}{m-n} + J_{pp}J_{qq}\right)\right]\\
E[C_{qp}(0)C_{pq}(k)] &= -\frac{1}{4T}\,\{D_{pq}(k)\\
&\quad + \sigma^2[J_{pp}\rho_q(k) + J_{qq}\rho_p(k)]\}\\
E[C_{pq}(k)C_{qp}(l)] &= \frac{1}{4T}\,\{D_{pq}(l+k) + \sigma^2[J_{pp}\rho_q(k+l)\\
&\quad + J_{qq}\rho_p(k+l)] + \delta(k+l)\sigma^4 J_{qq}J_{pp}\}
\end{aligned}$$

where we have set

$$D_{pq}(k) = \int_{-1/2}^{1/2} f_p(\lambda) f_q(\lambda) \exp(2i\pi\lambda\tau_k)\, d\lambda$$
$$J_{pq} = (\mathbf{A}^H\mathbf{A})^{-1}_{pq}$$

and where $f_p$ denotes the spectral density of the $p$th source signal. Using the above, the ISR is asymptotically given by

$$\mathcal{I}_{pq} = \mathcal{I}^0_{pq} + \sigma^2\mathcal{I}^1_{pq} + \sigma^4\mathcal{I}^2_{pq} \tag{25}$$

where the coefficients of the expansion are

$$\begin{aligned}\mathcal{I}^0_{pq} = \frac{1}{4T}\Bigg[&\alpha_{pq}(0)^2 D_{pq}(0) - 2\alpha_{pq}(0) \\ &\cdot \sum_{1\le|k|\le K} \alpha_{pq}(k) D_{qp}(k) \\ &+ \sum_{1\le|k|,\,|l|\le K} \alpha_{pq}(k)\alpha_{pq}(l) D_{pq}(k+l)\Bigg]\end{aligned}$$

$$\begin{aligned}\mathcal{I}^1_{pq} = \frac{1}{4T}\Bigg\{&\Bigg[\alpha_{pq}(0)^2 - 2\alpha_{pq}(0) \sum_{1\le|k|\le K} \alpha_{pq}(k)\rho_q(k) \\ &+ \sum_{1\le|k|,\,|l|\le K} \alpha_{pq}(k)\alpha_{pq}(l)\rho_q(k+l)\Bigg] J_{pp} \\ &+ \Bigg[\alpha_{pq}(0)^2 - 2\alpha_{pq}(0) \sum_{1\le|k|\le K} \alpha_{pq}(k)\rho_p(k) \\ &+ \sum_{1\le|k|,\,|l|\le K} \alpha_{pq}(k)\alpha_{pq}(l)\rho_p(k+l)\Bigg] J_{qq}\Bigg\}\end{aligned}$$

$$\begin{aligned}\mathcal{I}^2_{pq} = \frac{1}{4T}\Bigg\{&\frac{\alpha_{pq}(0)^2 |J_{pq}|^2}{m-n} \\ &+ J_{pp}J_{qq}\left[\alpha_{pq}(0)^2 + \frac{2}{|\rho_p - \rho_q|^2}\right]\Bigg\}.\end{aligned}$$

### C. Discussion

For high signal-to-noise ratio, (25) of the ISR is dominated by the first term $\mathcal{I}^0_{pq}$. This term shows two important features.

- $\mathcal{I}^0_{pq}$ is proportional to the spectral overlap of sources $p$ and $q$. If the sources $p$ and $q$ have no spectral overlap (i.e., their frequency supports are disjoint: $f_p(\lambda)f_q(\lambda) = 0$ for all $\lambda$), the corresponding ISR given by $\mathcal{I}_{pq}$ vanishes at first order. More generally, the ISR in the high SNR limit is proportional to the spectral overlap (this effect is illustrated in the next section).
- $\mathcal{I}^0_{pq}$ is independent of the mixing matrix. In the array processing context, it means that performance in terms of interference rejection is unaffected (surprisingly enough) by the array geometry and, in particular, by the number of sensors. The performance depends solely on the spectral overlap of the source signals. This (maybe surprising) phenomenon has been investigated in a more general context in [22].

In the above algorithm, the covariance matrices involved in the joint diagonalization criterion (20) are uniformly weighted.

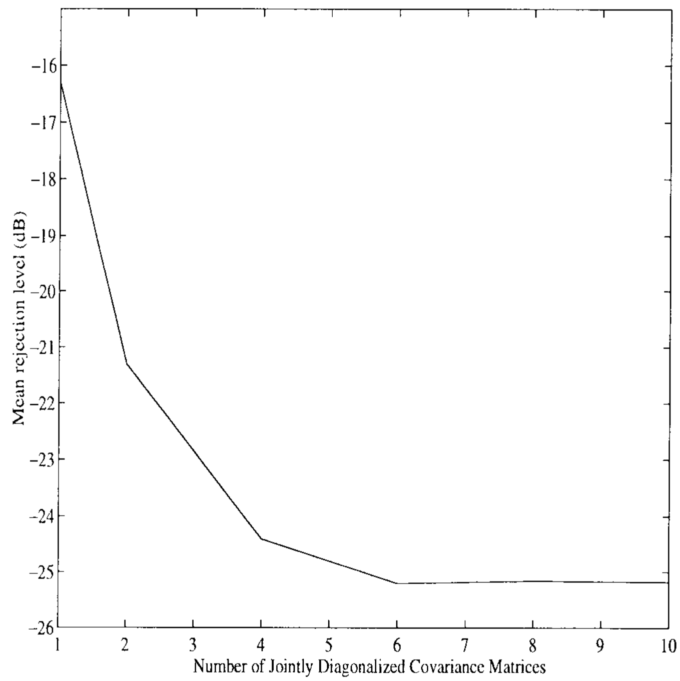



Fig. 1. Performance versus number of joint diagonalized covariance matrices.

Note that the JD criterion could be generalized by weighting each "off" term by an appropriate factor. Optimal weighting could, at least theoretically, be obtained by extending the previous derivations. This point is left to further study.

## V. Performance Evaluation

This section investigates the performance of the SOBI algorithm by computer simulations. The validity of the asymptotic performance analysis is also assessed.

### A. Numerical Simulations

In the simulated environment, a five-element uniform linear array with half wavelength sensor spacing receives two signals in the presence of stationary complex white noise. The two sources are unit variance, complex circular Gaussian with different but overlapping spectra. The sources arrive from different directions $\phi_1 = 12°$ and $\phi_2 = 13°$ (the particular structure of the array manifold is, of course, not exploited by the SOBI algorithm). The snapshot size is $T = 1000$ samples; the mean overall rejection level is estimated by averaging 300 independent trials.

*Example 1:* The source signals are generated by filtering a complex circular white Gaussian processes by an AR model of order 1 with coefficient $a_1 = \rho_1 \exp(j\theta_1)$ and $a_2 = \rho_2 \exp(j\theta_2)$. The time lags implicitly involved are $\tau_1, \cdots, \tau_K$, where $\tau_i$ is $i$ times the time unit.

In Fig. 1, the rejection level $\mathcal{I}_{perf}$ is plotted in decibels as a function of the number of the jointly diagonalized covariance matrices for SNR $= 10$ dB. The modulus of the AR coefficients of the two sources is $\rho_1 = \rho_2 = 0.85$; the angles are, respectively, equal to $\theta_1 = 0.5$ and $\theta_2 = 0.55$; we are dealing here with sources presenting a large spectral overlap.

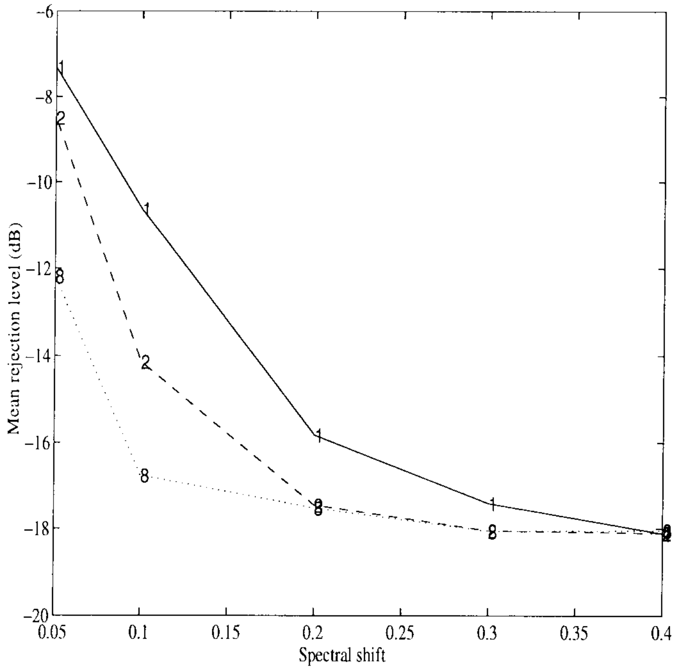



Fig. 2. Performance versus the "spectral shift" $\delta\theta$.

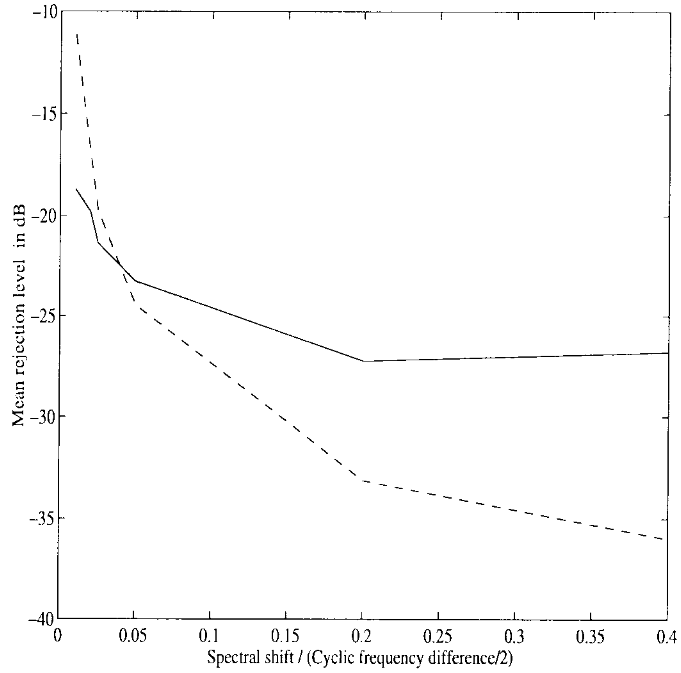



Fig. 3. Performance versus "spectral shift": SNR = 10 dB. Dashed line: SOBI method; solid line: the SCORE method.

Fig. 1 shows a performance gain reaching 9 dB by diagonalizing six matrices rather than only one. We have found experimentally that the most significant gain in performance are obtained in difficult environments: poor SNR, small spectral difference, ill-conditioned mixture matrix, etc.

In Fig. 2, the noise level is kept constant at 5 dB, and $\theta_1 = 0.5$. We let $\theta_2$ vary as $\theta_2 = \theta_1 + \delta\theta$. On the plot, the curves are labeled with the number of covariance matrices used in the identification. The plot shows the rejection level $\mathcal{I}_{perf}$ in decibels plotted as against the "spectral shift" $\delta\theta$. The plot evidences a significant increase in rejection performance by including two or eight covariance matrices in the joint diagonalization criterion.

*Example 2:* In this example, we compare the performance of the SOBI algorithm with the self-coherence restoral (SCORE) algorithm presented in the paper by Agee *et al.* [12]. In contrast with SOBI, the SCORE method assumes that the source signals are cyclostationary with different cyclic frequencies.

In this experiment, the first source is a first-order autoregressive Gaussian process ($\rho = 0.85, \theta = 0$) modulated by a complex exponential with normalized frequency $\alpha_1 = 0.3$ (the signal is thus cyclostationary with cyclic frequency $2\alpha_1$). The second source is also a first-order autoregressive Gaussian process ($\rho = 0.85$, $\theta = 0$) modulated by a complex exponential with normalized frequency $\alpha_2 = \alpha_1 + \delta\alpha$. Herein, the SOBI algorithm is used by jointly diagonalizing four covariance matrices.

The performance measure used to judge the quality of the processor output signal is the mean rejection level as defined in Section IV-A. In Fig. 3, the noise level is kept constant at $-10$ dB, and the mean rejection level is plotted in decibels as a function of the spectral shift $\delta\alpha$, which is also half the difference between the two cyclic frequencies $2\alpha_1$ and $2\alpha_2$.

It is seen in Fig. 3 that the SCORE method is less sensitive than SOBI to small values of $\delta\alpha$. In contrast, for large spectral shift $\delta\alpha$, the SOBI algorithm allows a performance gain reaching 10 dB.

In Fig. 4, spectral shift $\delta\alpha$ is kept constant at 0.4. The noise power is varied between $-25$ and 0 dB. The plot shows the mean rejection level in decibels as a function of the noise power $\sigma^2$. This figure demonstrates that in the case of large spectral shift, the SOBI method shows a performance gain of 10 dB compared with the SCORE algorithm.

Of course, it would be wrong to claim that the SOBI method yields consistently better results than the SCORE method. We only want to claim that in the situations where the sources are "sufficiently" separated in the stationary frequency domain, the SOBI algorithm yields acceptable results.

As a final note, we want to stress that the spectral separation of the sources is essential for the SOBI method; it is not required by the SCORE algorithm (or the further refinements of it [13], [23]), which is able to separate signals with a complete spectral overlap, provided they show different cyclostationary features.

### *B. Experimental Validation of the Asymptotic Performance Analysis*

In this section, a series of experiments to assess the domain of validity of the first-order performance approximation (25). The same settings than in Example 1 are used with the exception of the directions of arrivals, which are now $\phi_1 = 10°$ and $\phi_2 = 30°$.

The identification is performed using three covariance matrices, i.e., $\mathbf{R}(1)$, $\mathbf{R}(2)$, and $\mathbf{R}(3)$, and the overall rejection level is evaluated over 500 independent runs.

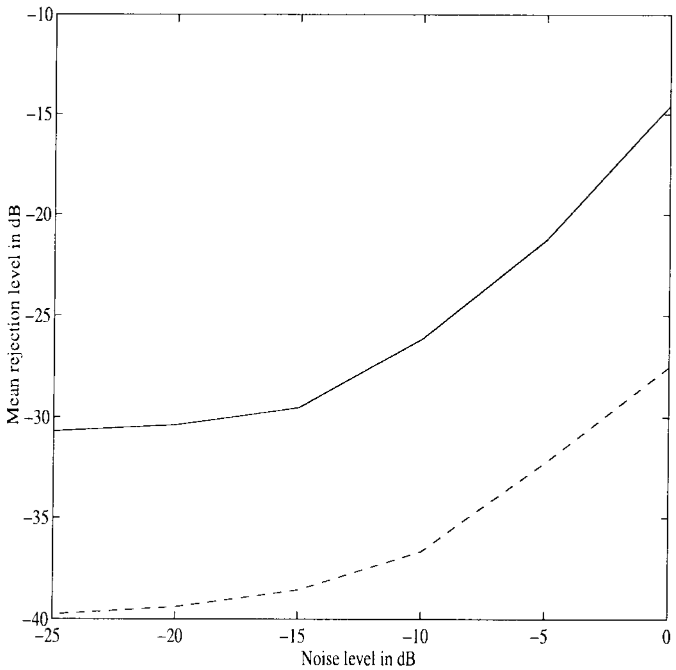



Fig. 4. Performance versus $\sigma^2$. Dashed line: the SOBI method; solid line: the SCORE method.

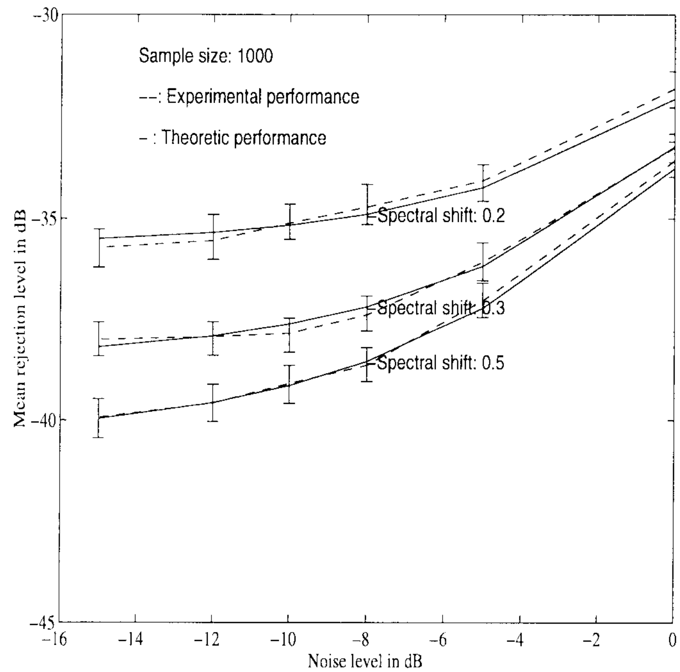



Fig. 5. Performance validation versus $\sigma^2$. Vertical lines indicate the 90% confidence intervals as obtained by bootstrapping the percentiles.

In Fig. 5, the rejection level $\mathcal{I}_{perf}$ is plotted in decibels as a function of the noise power $\sigma^2$ (which is also expressed in decibels). The sources are first-order autoregressive with parameters $\rho_1 = \rho_2 = 0.85$ and angle $\theta_1 = 0.5$, and $\theta_2 = \theta_1 + \delta\theta$. On the plots, the curves are labeled with the spectral shift $\delta\theta$. We note that the approximation is better at high SNR and for large spectral shift. This means that the asymptotic conditions are reached more quickly in this range of parameters.

In Fig. 6, the rejection level $\mathcal{I}_{perf}$ is plotted in decibels against sample size. On the plots, the curves are labeled as the function of the noise power $\sigma^2$ in decibels. This figure shows that the asymptotic closed-form expressions of the rejection levels are pertinent from a snapshot length of about 100 samples. This means that asymptotic conditions are reached even for small data block size.

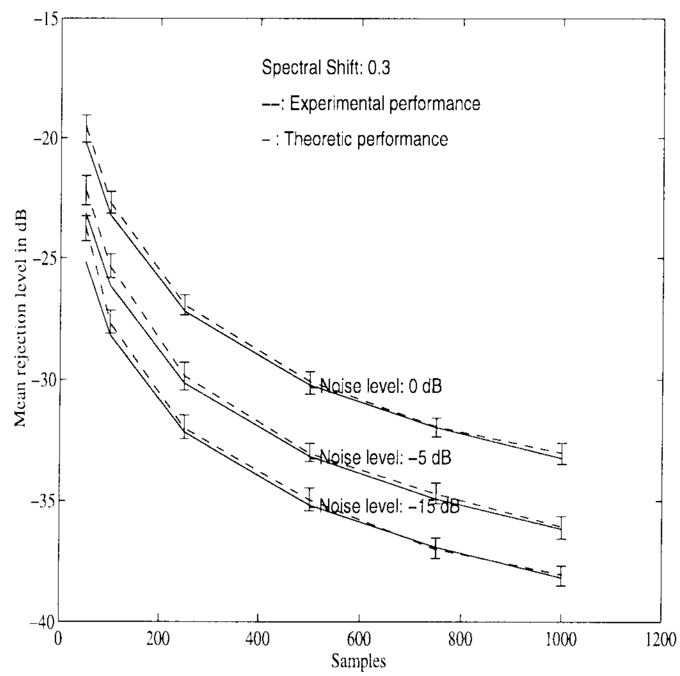



Fig. 6. Performance validation versus samples. Vertical lines indicate the 90% confidence intervals as obtained by bootstrapping the percentiles.

## VI. Conclusion

This paper presents a new blind source separation technique for temporally correlated sources. It is based on the "joint diagonalization" of an arbitrary set of covariance matrices. This method shows a number of attractive features:

i) It relies only on second-order statistics of the received signals.
ii) It allows—in contrast to higher order cumulant techniques—the separation of Gaussian sources.
iii) The use of several covariance matrices (in contrast with the previous proposal by [14]) makes the algorithm more robust: For practical purposes, it makes very unlikely indeterminacies.

Numerical experiments show the benefit of exploiting several covariance matrices in difficult contexts (low SNR, sources with little spectral difference). The main steps of the computation of the asymptotic performance analysis are also given.

## Appendix A
## A Joint Approximate Diagonalization Algorithm

The Jacobi technique [20] for diagonalizing a unique Hermitian matrix is extended for the joint approximate diagonalization of a set of normal matrices. The proposed method consists of minimizing the JD criterion (20) by successive Givens rotations, which leads to solving the same problem for $K$ $2 \times 2$ matrices:

$$\mathbf{H}_k = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \tag{26}$$

for $k = 1, \cdots, K$. A unitary matrix $\mathbf{V}$ is sought such that $\mathbf{H}'_k = \mathbf{V}^H \mathbf{H}_k \mathbf{V}$ ($k = 1, \cdots, K$) minimizes the criterion (20). The unitary transformation $\mathbf{V}$ is parameterized by a complex Givens rotation:

$$\mathbf{V} = \begin{bmatrix} \cos\theta & e^{j\phi}\sin\theta \\ -e^{-j\phi}\sin\theta & \cos\theta \end{bmatrix}. \tag{27}$$

Denoting by $a'_k$, $b'_k$, $c'_k$, and $d'_k$ the coefficients of $\mathbf{H}'_k$, optimization of (20) amounts to finding $\theta$ and $\phi$ such that $\sum_k |a'_k|^2 + |d'_k|^2$ is maximized. Noticing that $2(|a'_k|^2 + |a'_k|^2) = |a'_k - d'_k|^2 + |a'_k + d'_k|^2$ and that the trace $a'_k + d'_k$ is invariant in a unitary transformation, optimization of criterion (20) is equivalent at each Givens step to the maximization of $\mathbf{Q}$:

$$\mathbf{Q} \stackrel{\text{def}}{=} \sum_k |a'_k - d'_k|^2. \tag{28}$$

It is easily checked that

$$\begin{aligned} a'_k - d'_k = &(a_k - d_k)\cos 2\theta - (b_k + c_k)\sin 2\theta\cos\phi \\ &- j(c_k - b_k)\sin 2\theta \sin\phi \end{aligned} \tag{29}$$

for $k = 1, \cdots, K$. Then, by defining the vectors

$$\mathbf{u}^T \stackrel{\text{def}}{=} [a'_1 - d'_1, \cdots, a'_K - d'_K] \tag{30}$$

$$\mathbf{v}^T \stackrel{\text{def}}{=} [\cos 2\theta, -\sin 2\theta\cos\phi, -\sin 2\theta\sin\phi] \tag{31}$$

$$\mathbf{g}_k^T \stackrel{\text{def}}{=} [a_k - d_k, b_k + c_k, j(c_k - b_k)] \tag{32}$$

the $K$ equations (29) may be written in the form $\mathbf{u} = \mathbf{G}\mathbf{v}$, where $\mathbf{G}^T \stackrel{\text{def}}{=} [\mathbf{g}_1, \cdots, \mathbf{g}_K]$, so that $\mathbf{Q}$ also reads

$$\mathbf{Q} = \mathbf{u}^H\mathbf{u} = \mathbf{v}^T\mathbf{G}^H\mathbf{G}\mathbf{v} = \mathbf{v}^T \operatorname{Re}(\mathbf{G}^H\mathbf{G})\mathbf{v} \tag{33}$$

where we have used the fact that $\mathbf{G}^H\mathbf{G}$ being Hermitian by construction, its imaginary part is antisymmetric and hence contributes nothing to the above quadratic form. The last step is to recognize that the particular parameterization (31) of $\mathbf{v}$ is equivalent to the condition $\mathbf{v}^T\mathbf{v} = 1$. Maximizing a quadratic form under the unit norm constraint of its argument is classically obtained by taking $\mathbf{v}$ to be the eigenvector of $\operatorname{Re}(\mathbf{G}^H\mathbf{G})$ associated with the largest eigenvalue. Recall that this is a real $3 \times 3$ symmetric matrix: The analytic expressions of the parameters of the Givens rotation are simply derived from the coordinates of the eigenvector. The reader may check that setting $K = 1$ and $\mathbf{H}_1$ Hermitian, the above boils down to the standard Jacobi procedure. In addition, note that the main cost in this kind of technique is the update under Givens rotations of the various matrices involved in the diagonalization. This makes it clear that the cost of the proposed procedure is similar to $K$ times the diagonalization of a single matrix.

## APPENDIX B
## PROOF OF THEOREM 3

The sufficiency of (21) is established by proving that any linear combination (with at least two nonzero factors) of the vectors $\mathbf{u}_i$, $i = 1, \cdots, n$ cannot be a common eigenvector of the matrices $\mathbf{M}_k$, $k = 1, \cdots, K$:

Let $\mathbf{v} = \sum_{1 \le i \le n} \alpha_i \mathbf{u}_i$ be a common eigenvector of the matrices $\mathbf{M}_k$, $k = 1, \cdots, K$, and assume, for example, that $\alpha_1 \neq 0$. According to (21), for any index $i$, $1 < i \le n$, there exists an index $k$ such that $d_1(k) \neq d_i(k)$. For this index $k$, we have by hypotheses

$$\mathbf{M}_k\mathbf{v} = \lambda_k\mathbf{v} = \sum_{j=1}^n \lambda_k \alpha_j \mathbf{u}_j$$

and

$$\mathbf{M}_k\mathbf{v} = \sum_{j=1}^n \alpha_j \mathbf{M}_k \mathbf{u}_j = \sum_{j=1}^n \alpha_j d_j(k)\mathbf{u}_j.$$

By identification, we have $\alpha_j[d_j(k) - \lambda_k] = 0$ for $1 \le j \le n$. Since $\alpha_1 \neq 0$ and $d_1(k) \neq d_i(k)$, this leads to $\lambda_k = d_1(k)$ and $\alpha_i = 0$. Q.E.D.

Next, we establish the necessity of (21). Assume that there exists a pair $(i, j)$ such that $d_i(k) = d_j(k)$ for $k = 1, \cdots, K$. Then, any linear combination of the vectors $\mathbf{u}_i$ and $\mathbf{u}_j$ is a common eigenvector of the matrices $\mathbf{M}_k$, $k = 1, \cdots, K$. Q.E.D.

## APPENDIX C
## PROOF OF LEMMA 2

In this section, a sketch of the proof for Lemma 2 is presented. Giving a full proof is a tedious and lengthy exercise, which goes far beyond the scope of this paper (it can be obtained by request to the author). A part of the proof is based on the result on the perturbation of the joint diagonalization obtained in [24]. For brevity, this result is admitted.

The square modulus $|\hat{I}_{pq}|^2$ is expressed as

$$|\hat{I}_{pq}|^2 = |(\hat{\mathbf{U}}^H\hat{\mathbf{W}}\mathbf{A})_{pq}|^2. \tag{34}$$

We decompose the matrix $\hat{\mathbf{W}}\mathbf{A}$ under its polar form

$$\hat{\mathbf{V}}\hat{\mathbf{H}} = \hat{\mathbf{W}}\mathbf{A} \tag{35}$$

where $\hat{\mathbf{V}}$ is a unitary matrix, and $\hat{\mathbf{H}}$ is a nonnegative semi-defined Hermitian matrix; matrix $\hat{\mathbf{H}}$ verifies $\hat{\mathbf{H}}^2 = \mathbf{A}^H\hat{\mathbf{W}}^H\hat{\mathbf{W}}\mathbf{A}$ (see [19, Theorem 7.3.2, p. 412]). According to the convention outlined in Section III-A, matrix $\hat{\mathbf{H}}$ is expected to be close to the identity matrix; let $\delta\mathbf{H} = \hat{\mathbf{H}} - \mathbf{I}$ denote the estimation error of the Hermitian part of $\hat{\mathbf{W}}\mathbf{A}$. Using standard perturbation calculus (see, for example, [25]), it can be shown that

$$\begin{aligned} \delta\mathbf{H} \simeq &-\tfrac{1}{2}\mathbf{A}^{\#}\delta\mathbf{R}(0)\mathbf{A}^{\#H} \\ &+ \frac{1}{2(m-n)} \operatorname{Tr}[\Pi\delta\mathbf{R}(0)](\mathbf{A}^H\mathbf{A})^{-1} + o[\delta\mathbf{R}(0)]. \end{aligned} \tag{36}$$

From the polar decomposition (35), the whitened covariance matrices can be similarly approximated at the first order, for all $k \neq 0$, as

$$\underline{\hat{\mathbf{R}}}(\tau_k) = \hat{\mathbf{W}}[\mathbf{A}\mathbf{R}_s(k)\mathbf{A}^H + \hat{\mathbf{R}}(\tau_k) - \mathbf{R}(\tau_k)]\hat{\mathbf{W}}^H = \hat{\mathbf{V}}[\hat{\mathbf{H}}\mathbf{R}_s(k)\hat{\mathbf{H}} + \hat{\mathbf{V}}^H\hat{\mathbf{W}}\delta\mathbf{R}(\tau_k)\hat{\mathbf{W}}^H\hat{\mathbf{V}}]\hat{\mathbf{V}}^H. \quad (37)$$

The joint diagonalization criterion aims at searching the unitary matrix that minimizes the "off" of a set of matrices, which, here, is the whitened covariance matrix $\underline{\hat{\mathbf{R}}}(\tau_k)$. It is not difficult to guess (though actually difficult to prove in mathematical terms due to the indeterminacies inherent to these kinds of problems; see a discussion in [22] and [24]) that if the set of matrices entering in the JD are multiplied by a *common* unitary matrix, then the result of the JD will simply be multiplied by this common matrix. Formally, let $\mathbf{N}_1, \cdots, \mathbf{N}_p$ be arbitrary matrices and $\mathbf{U}$ an arbitrary unitary matrix; then, JD $\{\mathbf{U}\mathbf{N}_1\mathbf{U}^H, \cdots, \mathbf{U}\mathbf{N}_p\mathbf{U}^H\} = \mathbf{U}$JD $\{\mathbf{N}_1, \cdots, \mathbf{N}_p\}$. Applying this result in our situation, it comes from (37) that the unitary matrix $\hat{\mathbf{U}}$, resulting from the JD of the set of whitened covariance matrices $\underline{\hat{\mathbf{R}}}(1), \cdots, \underline{\hat{\mathbf{R}}}(K)$, can be decomposed as

$$\hat{\mathbf{U}} = \hat{\mathbf{V}}\hat{\mathbf{U}}_0$$

where the matrix $\hat{\mathbf{U}}_0$ minimizes the JD criterion for the matrices:

$$\begin{aligned}\mathbf{M}_k &\stackrel{\text{def}}{=} \hat{\mathbf{H}}\mathbf{R}_s(k)\hat{\mathbf{H}} + \hat{\mathbf{V}}^H\hat{\mathbf{W}}\delta\mathbf{R}(\tau_k)\hat{\mathbf{W}}^H\hat{\mathbf{V}}, \\ &\quad 1 \le k \le K \\ &= \mathbf{R}_s(k) + \mathbf{R}_s(k)\delta\mathbf{H} + \delta\mathbf{H}\mathbf{R}_s(k) \\ &\quad + \hat{\mathbf{V}}^H\hat{\mathbf{W}}\delta\mathbf{R}(\tau_k)\hat{\mathbf{W}}^H\hat{\mathbf{V}} + o[\delta\mathbf{R}(\tau_k)] \\ &= \mathbf{R}_s(k) + \mathbf{R}_s(k)\delta\mathbf{H} + \delta\mathbf{H}\mathbf{R}_s(k) \\ &\quad + \mathbf{A}^{\#}\delta\mathbf{R}(\tau_k)\mathbf{A}^{\#^H} + o[\delta\mathbf{R}(\tau_k)] \\ &= \mathbf{R}_s(k) + \xi_k + o[\delta\mathbf{R}(\tau_k)]\end{aligned}$$

where $\xi_k \stackrel{\text{def}}{=} \mathbf{R}_s(k)\delta\mathbf{H} + \delta\mathbf{H}\mathbf{R}_s(k) + \mathbf{A}^{\#}\delta\mathbf{R}(\tau_k)\mathbf{A}^{\#^H}$. Hence, (34) can be written as

$$|\hat{I}_{pq}|^2 = |(\hat{\mathbf{U}}_0^H\hat{\mathbf{H}})_{pq}|^2.$$

As shown in [24], the unitary matrix $\hat{\mathbf{U}}_0$ is given at first order by

$$\begin{aligned}\hat{\mathbf{U}}_0 &= \mathbf{I} + \delta\mathbf{U}_0 \\ \delta\mathbf{U}_0 &= \tfrac{1}{2}\sum_{r\neq s}\sum_{k=1}^{K}[\alpha_{rs}(k)\Pi_r\xi_k\Pi_s + \alpha_{rs}^*(k)\Pi_r\xi_k^H\Pi_s], \\ &\quad (\delta\mathbf{U}_0^H = -\delta\mathbf{U}_0) \qquad (38)\end{aligned}$$

where $\Pi_r = \mathbf{e}_r\mathbf{e}_r^*$ is the orthogonal projector on the $r$th vector column $\mathbf{e}_r$ of the identity matrix $\mathbf{I}_n$. The performance index becomes

$$\begin{aligned}|\hat{I}_{pq}|^2 &= |(\mathbf{I} - \delta\mathbf{U}_0)(\mathbf{I} + \delta\mathbf{H})|_{pq}^2 \\ &\simeq |\delta\mathbf{H} - \delta\mathbf{U}_0|_{pq}^2 \qquad \text{for } p \neq q. \qquad (39)\end{aligned}$$

Including (36) and (38) in (39) leads to the Taylor expansion of Lemma 2.

## ACKNOWLEDGMENT

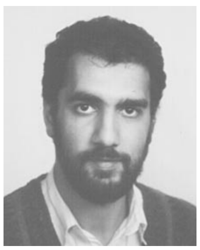
The authors would like to express this sincere thanks to Prof. P. Duhamel for his useful comments on an early version of this paper.

## REFERENCES

[1] F. Li and R. Vaccaro, "Sensitivity analysis of DOA estimation algorithms to sensor errors," *IEEE Trans. Aerospace Electron. Syst.,* vol. 28, pp. 708–717, July 1992.
[2] M. Viberg and A. L. Swindelhurst, "Analysis of the combined effects of finite samples and model errors on array processing performance," in *Proc. ICASSP*, 1993, vol. 4, pp. 372–375.
[3] C. Jutten and J. Hérault, "Blind separation of sources: An adaptive algorithm based on neuromimetic architecture," *Signal Processing,* vol. 24, pp. 1–10, 1991.
[4] A. Dinç and Y. Bar-Ness, "Bootstrap: A fast blind adaptive signal separator," in *Proc. ICASSP,* Mar. 1992, vol. 2, pp. 325–328.
[5] E. Moreau and O. Macchi, "A one stage self-adaptive algorithm for source separation," in *Proc. ICASSP,* Adelaide, Australia, 1994.
[6] B. Laheld and J.-F. Cardoso, "Adaptive source separation without prewhitening," in *Proc. EUSIPCO,* Edinburgh, Scotland, Sept. 1994, pp. 183–186.
[7] S. V. Gerven and D. Van Compernolle, "On the use of decorrelation in scalar signal separation," in *Proc. ICASSP,* Adelaide., Australia, 1994, pp. 57–60.
[8] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for nonGaussian signals," *Proc. Inst. Elec. Eng.,* pt. F, vol. 140, no. 6, pp. 362–370, 1993.
[9] P. Comon, "Independent component analysis, a new concept?," *Signal Processing,* vol. 36, pp. 287–314, 1994.
[10] M. Gaeta and J.-L. Lacoume, "Source separation without *a priori* knowledge: The maximum likelihood solution," in *Proc. EUSIPCO,* 1990, pp. 621–624.
[11] D. T. Pham, P. Garat, and C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach," in *Proc. EUSIPCO,* 1992, pp. 771–774.
[12] B. G. Agee, S. V. Schell, and W. A. Gardner, "Spectra self-coherence restore: A new approach to blind adaptive signal extraction using antenna arrays," *Proc. IEEE,* vol. 78, pp. 753–766, Apr. 1990.
[13] S. V. Schell and W. A. Gardner, "Maximum likelihood and common factor analysis-based blind adaptive spatial filtering for cyclostationary signals," in *Proc. ICASSP,* 1993, vol. 4, pp. 292–295.
[14] L. Tong, V. C. Soon, R. Liu, and Y. Huang, "AMUSE: A new blind identification algorithm," in *Proc. ISCAS,* New Orleans, LA, 1990.
[15] L. Féty, "Méthodes de traitement d'antenne adaptées aux radio-communications," Thése de docteur-ingénieur de l'ENST, June 1988.
[16] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, and E. Moulines, "Second-order blind separation of correlated sources," in *Proc. Int. Conf. Digital Signal Processing,* Cyprus, 1993, pp. 346–351.
[17] D. Pham and P. Garat, "Séparation aveugle de sources temporellement corrélées," in *Proc. GRETSI,* 1993, pp. 317–320.
[18] L. Tong, R. Liu, and Y. H. V. C. Soon, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.,* vol. 38, pp. 499–509, May 1991.
[19] R. Horn and C. Johnson, *Matrix Analysis.* Cambridge, U.K.: Cambridge Univ. Press, 1985.
[20] G. H. Golub and C. F. V. Loan, *Matrix Computations.* Baltimore, MD: Johns Hopkins Univ. Press, 1989.
[21] M. Rosenblatt, *Stationary Processes and Random Fields.* Boston, MA: Birkhauser, 1985.
[22] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO,* 1994, pp. 776–779.
[23] W. A. Gardner, S. V. Schell, and P. A. Murphy, "Multiplication of cellular radio capacity by blind adaptive spatial filtering," in *Proc. IEEE Int. Conf. Selected Topics Wireless Commun.,* June 1992, pp. 102–106.
[24] J.-F. Cardoso, "Perturbation of joint diagonalizers," Telecom Paris, Signal Dept., Tech. Rep. 94D023, 1994.
[25] A. Souloumiac, "Utilization des statistiques d'ordre supérieur en traitement d'antenne," Thèse de doctoral, Télécom Paris, Feb. 1993.

**Adel Belouchrani** (M'96) was born in Algiers, Algeria, on May 5, 1967. He received the State Engineering degree from the National Polytechnic School of Algiers, Algeria, in 1991, the M.Sc. degree in signal processing from Institut National Polytechnique de Grenoble (INPG), France, in 1992, and the Ph.D. degree in the field of signal and image processing from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1995.

He has been a visiting scholar at the Electrical Engineering and Computer Sciences Department of the University of California at Berkeley from 1995 to 1996, working on fast adaptive blind equalization and carrier phase tracking. He is currently with the Department of Electrical and Computer Engineering of Villanova University, Villanova, PA, as research associate. His research interests are in digital communications and statistical signal processing, including (blind) array processing and performance analysis.



**Karim Abed-Meraim** was born in Algiers, Algeria, in 1967. He received the M.Sc. degree from Ecole Polytechnique in 1990 and the Ph.D degree from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1995 in signal processing.

He currently is with the Department of Electrical Engineering, Melbourne University, Melbourne, Australia, as a research fellow. His research interests are in statistical signal processing and include system identification, (blind) array processing, and performance analysis.

**Jean-Francois Cardoso** (M'91) was born in 1958 in Tunis, Tunisia. He received the Agrégation de Physique degree from the École Normale Supérieure de Saint-Cloud, France, in 1981 and the Doctorat de Physique degree from the University of Paris, France, in 1984.

He currently is with the Centre National de la Recherche Scientifique (CNRS) and works in the "Signal" department of École Nationale Supérieure des Télécommunications (ENST), Paris, France. He is one of the coordinators of ISIS, which is the CNRS research group on signal and image processing. His research interests include statistical signal processing, with emphasis on (blind) array processing, and performance analysis.

Dr. Cardoso has been a member of the SSAP Technical Committee of the IEEE Signal Processing Society since 1995.

**Eric Moulines** (M'91) was born in Bordeaux, France, in 1963. He received the M.Sc. degree from Ecole Polytechnique in 1984 and the Ph.D. degree from Ecole Nationale Supérieure des Télécommunications, Paris, France, in signal processing in 1990.

From 1986 until 1990, he was a member of the technical staff at CNET, working on signal processing applied to low-bit rate speech coding and text-to-speech synthesis. Since 1990, he has been with Ecole Nationale Superieure des Télécommunications, where he is presently a Professor. His teaching and research interests include statistical and digital signal processing. Currently, he is engaged in research in various aspects of statistical signal processing including, among others, single and multichannel ARMA filtering and modeling, narrowband array processing, characterization, and estimation of point processes.

Dr. Moulines is on the editorial board of *Speech Communication* and is a member of the IEEE Speech Committee.