ASYMPTOTIC PERFORMANCE OF SECOND ORDER BLIND SEPARATION

K. Abed-Meraim, A. Belouchrani, J. F. Cardoso, E. Moulines
Télécom Paris, 46 rue Barrault, 75634 Paris Cedex 13
email: abedmer, adel, cardoso, moulines@sig.enst.fr

ABSTRACT
This communication deals with the problem of blind separation of an instantaneous linear mixture of mutually uncorrelated sources. A second order source separation technique exploiting the time coherence of the source signals is considered. Asymptotic performance analysis of the proposed method is performed. Several numerical simulations are presented to demonstrate the effectiveness of the proposed method and to validate the theoretical expression of the asymptotic performance index.

1. INTRODUCTION

When an array of $n$ sensors samples the fields radiated by $m$ narrow band sources its output is classically modeled as a $n$-dimensional multivariate process, whose components consist in an (unknown) instantaneous linear mixture of the $m$ source signals, corrupted by an additive noise. Source separation may be obtained by first identifying the directional vectors associated to each source, and then by projecting the array signal onto the estimated vectors. This is a standard program in array processing, but blind source separation proposal is to perform identification without resorting to the knowledge of the array manifold. All the source separation proposal so far are based on the crucial assumption of mutual statistical independence of the source signals. This strong but plausible assumption seems to be the price for ignoring the array manifold.

Various solutions have been proposed to the blind source separation problem. When the source signals are temporally uncorrelated, it has been recognized that the problem cannot be solved using only second-order information. One has then to resort to higher order statistics as in [1, 2, 3] or to non-linear spatial adaptive filters [4, 5].

On the other hand, if the source signals are temporally correlated, blind identification is possible based on spatial correlation matrices [6].

In this paper, we develop an approach based on an approximate joint-diagonalization of a set of correlation matrices. Robustness is significantly increased [7], at low additional cost, by processing such a matrix set rather than a unique matrix as in [6].

2. SIGNAL MODEL AND STATISTICS

Assume that $m$ narrow band signals impinge on an array of $n > m$ sensors. The measured array output is a weighted superposition of the signals, corrupted by additive noise, i.e.

$$x(t) = y(t) + n(t) = As(t) + n(t)$$

where $s(t) = [s_1(t), \ldots, s_m(t)]$ is the $m \times 1$ complex source signal vector (containing the complex envelope of the emitter signals), $n(t) = [n_1(t), \ldots, n_n(t)]$ is the $n \times 1$ complex noise vector and $A$ is the $n \times m$ mixture matrix. The source signal vector $s(t)$ is assumed to be a multivariate stationary complex stochastic process. The component processes $s_i(t), 1 \leq i \leq m$ are assumed to be mutually uncorrelated, with zero mean and second moment:

$$E \left[ s(t + \tau)s^\dagger(t) \right] = 0$$

$$S(\tau) = E \left[ s(t + \tau)s^\dagger(t) \right] = \text{diag}[\rho_1(\tau), \ldots, \rho_m(\tau)]$$

where $\rho_i(\tau) = E(s_i(t + \tau)s_i^\dagger(t))$ denotes the autocorrelation of the $s_i(t)$ 1. The additive noise $n(t)$ is modeled as a stationary, temporally white, zero-mean complex random process. For simplicity, we also require $n(t)$ to be circularly distributed and spatially white, i.e.

$$E \left[ n(t + \tau)n^\dagger(t) \right] = 0$$

$$E \left[ \text{diag}(n(t + \tau)n^\dagger(t)) \right] = \sigma^2 I$$

where $\delta_r$ is the Kronecker delta. The assumption of spatial white noise is not a restriction if the noise covariance is known (up to a unknown scalar factor $\sigma^2$). The noise is assumed to be independent with the signal wavefront. The $n \times 1$ complex matrix $A$ is assumed to have full rank but otherwise unknown. In contrast with traditional parametric method 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 familiar structure:

$$R(0) = E(x(t)x^\dagger(t)) = AS(0)A^* + \sigma^2 I$$

$$R(\tau) = E(x(t + \tau)x^\dagger(t)) = AS(\tau)A^*$$

The goal of blind source separation is to identify the mixture matrix and/or recover the source signals from the array output $x(t)$ without any a priori knowledge on 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.

3. SECOND ORDER IDENTIFICATION APPROACH

Indeterminations: Complete blind identification of matrix $A$ is impossible in the blind context, because the exchange of a fixed scalar between the source signal and the corresponding column of $A$ leaves the observations unaffected. We take advantage of this indetermination to assume, without any loss of generality, that the emitter signals have unit-variance, i.e.

$$S(0) = I$$

$1$ the superscripts $\dagger$ and $^*$ denote respectively the transpose and the complex conjugate transpose and $\text{diag}[\cdot]$ is the operator that transforms a $m \times 1$ vector in a $m \times m$ matrix containing the vector coordinates as diagonal element.

To appear in Proc. ICASSP-94

© IEEE 1994
Note that this normalization leaves the phase of the column of $\Lambda$ undetermined. Also note that the numbering of the signals is immaterial. It follows that the best that can be done is to determine $\Lambda$ up to a permutation and phase shifts of its columns. The crucial point is that these undeterminations do not impede source separation. If $\Lambda$ is determined up to permutation and phase shifts, it still allows to determine the source signals up to the corresponding fixed permutation and phase shifts.

**Whitening matrices:** The first step of the proposed estimation procedure consists in whitening the signal part $y(t)$ of the observation. This is achieved using a whitening matrix $W$, i.e., a $m \times m$ matrix such that

$$I = E((Wy(t))(Wy(t))^*).$$

Pinning the source signal covariance $R_\theta(0) = E(y(t)y(t)^*) = \Lambda \Lambda^*$ in the latter expression, shows that

$$I = (WA)(WA)^*$$

so that $WA$ is a $m \times m$ unitary matrix. For any whitening matrix $W$, it thus exists a $m \times m$ unitary matrix $U$ such that

$$WA = U$$

where the superscript $\#$ denotes the Moore-Penrose pseudoinverse: $W^* = W^* (WW^*)^{-1}$. Since the source signal covariance $R_\theta(0)$ may be expressed as $R_\theta(0) = \Lambda \Lambda^* = R(0) - \sigma^2 I$, a whitening matrix $W$ can be determined from the array output covariance $R(0)$. A whitening matrix may also be determined from a linear combination of a set of correlation matrices taken at non zero time lags, as suggested in [8]. In any case, as shown by (7), finding a whitening matrix still leaves an undetermined unitary factor in $A$. This unitary matrix can be determined from other statistics. Several authors have considered using 4th-order cumulants to determine the ”missing rotation” [see 1, 2, 3 for instance] when the source signals are non Gaussian. When this is not the case, an alternate solution is possible by resorting to the spatial correlation matrices (4) which may carry the missing information if the source signals are temporally correlated.

2nd-order identification procedure: Now consider the whitened correlation matrices $\tilde{R}(\tau)$ defined as

$$\forall \tau \neq 0 \quad \tilde{R}(\tau) = W R(\tau) W^*.$$  

These $m \times m$ complex matrices are the correlation matrices of the spatially whitened process $W X(t)$. Pinning the definition (4) into (7), it comes:

$$\forall \tau \neq 0 \quad \tilde{R}(\tau) = (WA) S(\tau)(WA)^* = U S(\tau) U^*.$$  

Since the matrix $U$ is unitary and the matrix $S(\tau)$ is diagonal, the latter equation shows that any whitened correlation matrix is diagonal in the basis of the columns of the matrix $U$ (the eigenvalues of $\tilde{R}(\tau)$ being the diagonal entries of $S(\tau)$).

If, for $\tau \neq 0$ the diagonal elements of $S(\tau)$ are all distinct, the missing unitary matrix $U$ may be ‘uniquely’ (i.e. up to permutation and phase shifts) retrieved by computing the eigendecomposition $\tilde{R}(\tau)$. Indetermination occurs in the case of degenerate eigenvalues, i.e. when $S_i(\tau) = S_j(\tau)$, $i \neq j$. It does not seem possible to a priori determine some value for the delay $\tau$ such that the diagonal entries of $S(\tau)$ are all distinct. Of course, if the source signals have different spectral shapes, such a kind of eigenvalue degeneracy is unlikely, but the problem is not purely academic because it is to be expected that when some eigenvalues of $\tilde{R}(\tau)$ comes close to degeneracy, the robustness of determining $U$ from eigendecomposition of a single whitened correlation matrix is seriously impaired.

The situation is more favorable when considering simultaneous diagonalization of a set $\{R(\tau) [i = 1, \ldots, K]\}$ of $K$ whitened correlation matrices. This set is simultaneously unitarily diagonalizable by the unitary matrix $U$ as in (9).

The matrix $U$ is unique (to a permutation matrix and phase factors) if, and only if, for any pair $(t, j)$ of sources there exists a lag $\tau_i$ such that $p_i(\tau_i) \neq p_j(\tau_i)$. Of course, the simultaneous diagonalization holds only for the exact statistics; empirical statistics may only be approximately simultaneously diagonalized under the same unitary transform. This calls for the definition to an approximate simultaneous diagonalization.

**Joint approximate diagonalization:** we begin here with some relevant definitions. A $m \times m$ matrix $M$ is said to be normal if $M^* = \Lambda M^*$, that is if $M$ commute with its Hermitian adjoint. Any normal matrix $M$ is unitarily diagonalizable, i.e. it exists a unitary matrix $U$ and a diagonal matrix $D$, such that $M = U^D U^*$. This property is often referred to as the spectral theorem for normal matrices, see for example [9]. Theorem 2.5.4. Let $V$ be a $m \times m$ unitary matrix, and define $M(V) = V^* M V = \{M(V)_{ij}\}_{i,j \leq m}$. Note that $M(U) = D$. From standard algebra, we may write:

$$\sum_{i,j=1}^m |M_{ij}|^2 = \text{Tr} M M^* = \text{Tr} M(V) M(V)^*$$

$$= \sum_{i=1}^m |M(V)_{ii}|^2 + \sum_{i \neq j} |M(V)_{ij}|^2$$

Applying these relations to $V = U$, it comes that $\text{Tr} M M^* = \sum_{i=1}^m |d_i|^2$. Thus, for any unitary matrix $V$, $\sum_{i=1}^m |M(V)_{ii}|^2 \leq \sum_{i=1}^m |M(U)_{ii}|^2$: diagonalization is equivalent to maximizing under unitary transform the sum of the squared moduli of the matrix $M(V) = V^* M V$.

**Lemma 1** Let $M$ be a $m \times m$ normal matrix. $U$ is a diagonalizing unitary matrix for $M$, if, and only if it maximizes

$$C(M, V) = \sum_i |M(V)_{ii}|^2$$

over the set $\mathcal{U}_m$ of $m \times m$ unitary matrices $V$.

Equivalently, searching for $U$ boils down to search for the unitary transform that minimizes the squared-moduli of the off-diagonal terms of $M(V)$.

If $\mathcal{M} = \{M_1, \ldots, M_K\}$ is a commuting family of normal matrices (i.e., $M_i M_j = M_j M_i$, for each pair of indices $(i, j)$), then $\mathcal{M}$ is simultaneously unitarily diagonalizable; that is, there exists a single unitary matrix $U$ that transforms each matrix in $\mathcal{M}$ into a diagonal matrix. Similar to lemma 1, the matrix $U$ maximizes, over the set of $m \times m$ unitary matrices $V$, the following criterion:

$$C(U, \mathcal{M}) = \sum_{k=1}^K C(M_k, V)$$

where $C(M, V)$ is defined by (11).

We quite naturally define the joint approximate diagonalization of a set $\mathcal{M} = \{M_k[k = 1, \ldots, K] \}$ of $K$ arbitrary $m \times m$ matrices as the maximization of the criterion 12. The set of unitary matrix $\mathcal{U}_m$ being compact for the topology induced by the scalar product $(A, B) = \text{Tr}(AB^*)$ and the function $C(\cdot, V)$ being continuous $\mathcal{U}_m \to \mathbb{R}$, the function $C(U, \mathcal{M})$ admits a global minimum, and this minimum
is reached in $U_m$. An efficient joint approximate diagonalization algorithm can be found in [10, 7] which is a generalization of the Jacobi technique for the exact diagonalization of a single hermitian matrix [11].

Identification procedure: We now have at hand all the necessary ingredients to derive the main identification procedure; it comprises the following steps

- From the eigendecomposition of the sample estimate of the autocorrelation matrix $\hat{R}(0)$, estimate (i) the variance of the noise $\hat{\sigma}^2$ (by averaging the eigenvalues associated with the $n - m$ smallest eigenvalues) (ii) a whitening matrix $\hat{W}$ (by computing a square-root of the pseudo-inverse of $\hat{R}(0) - \hat{\sigma}^2 I$),
- Determine the unitary matrix $\hat{U}$ by maximizing the joint approximate diagonalization criterion for a specific set of whitened correlation matrices $\hat{\Lambda}$,
- Obtain an estimate of the mixture matrix $\hat{A}$ as $\hat{A} = \hat{W}^T \hat{U}$.

4. ASYMPTOTIC PERFORMANCE.

The performance may be characterized in terms of signal rejection. After blind identification of the mixture matrix $\hat{A}$, the estimated source signals may be obtained as $\hat{s}(t) = \hat{\Lambda}^T x(t) = \hat{\Lambda}^T A s(t) + \hat{\Lambda}^T n(t)$ where $\hat{\Lambda}^T$ is the pseudo-inverse of the estimated matrix $\hat{A}$ of the mixture matrix.

The matrix $\hat{P}$ defined by $\hat{P} = \hat{\Lambda}^T \hat{A}$ should be close to some matrix $P$ with only one zero phase term in each row and each column (phase and permutation indetermination).

For convenience, we assume that $\hat{P}$ is close to diagonal rather than to some other permutation matrix. The $p$-th estimated source signal is

$$\hat{s}_p(t) = \sum_{q=1}^n \hat{P}_{pq} s_q(t) + (\Lambda^T n(t))_p$$

The power of the $q$-th source signal residual (interference) in the $p$-th estimated source signal is $I_{pq} = E[|\hat{P}_{pq}|^2]$ (since the sources have unit power, this quantity is nothing but the interference to signal ratio for the $q$-th and $p$-th source). As a global measure of performance, we use the overall rejection level defined as the sum of all the interferences

$$I_{\text{overall}} = \sum_{q \neq p} E[|\hat{P}_{pq}|^2] = \sum_{q \neq p} I_{pq}$$

In the case of gaussian noise and gaussian source signals, we have derived closed form expressions of the rejection index at the limit of large snapshots, in terms of the coefficients of the mixture matrix and of the spectral interference of the sources. Details of the calculation will be published elsewhere.

$$I_{pq} = E_n + E_c + E_s$$

where $E_n$, $E_c$ and $E_s$ are given by:

$$E_n = \frac{1}{4T} \left[ 1 + \frac{|\rho_p|^2 - |\rho_q|^2}{|\rho_p - \rho_q|^2} I(0) + \sigma (J_p + J_q) \right]$$

$$+ \sigma^2 \left( \frac{Tr(W^T \Pi_p W W^T \Pi_q)}{m - n} + J_q J_q \right)$$

$$E_c = \frac{1}{T} \sum_{k=1}^N \Re \left\{ (I + \frac{|\rho_p|^2 - |\rho_q|^2}{|\rho_p - \rho_q|^2}) (\rho_p^* - \rho_q^*)(k) \right\}$$

$$E_s = \frac{1}{2T} \sum_{l=1}^N \Re \left\{ \frac{(\rho_p^* - \rho_q^*)(l)(\rho_p - \rho_q)^\dagger(l + k)}{|\rho_p - \rho_q|^4} I(l + k) \right\} + (l - k) + \sigma (J_p \rho_k + l + \rho_q(l - k)) + \sigma$$

$$+ (J_q \rho_k + l + \rho_q(l - k)) + \frac{1}{2T} \sigma^2 \left( J_p J_q \right)^T$$

with

$$I(k) = \int \frac{1}{2\pi} \hat{F}_p(\lambda) F_q(\lambda) \exp(2\pi i k \lambda) d\lambda$$

$$J_p = Tr(W W^T \Pi_p)$$

$$\rho_p = [\rho_p(1), ..., \rho_p(N)]^T$$

$\Pi_p = u_p u_p^*$ is the $p$-th column vector of the unitary matrix $U = W A$.

From above expressions we note that:

- For high signal to noise ratio (i.e., when the noise variance is small $\sigma \rightarrow 0$), $I_{pq}$ is proportional to the spectral interference of sources $p$ and $q$. If the sources $p$ and $q$ have no spectral interference (i.e., their frequency supports are disjoint), the sources are perfectly separated as $T \rightarrow \infty$.

- Performance of the method decreases when the spectral interference between sources becomes important, i.e. $\rho_p \rightarrow \rho_q$. This is illustrated in (fig 1).

- For high enough signal to noise ratio, interference errors are independent of array geometry and thereby of the number of sensors. They depend only on the spectral interference of the source signals.

5. VALIDATION.

Extensive numerical simulations have been performed (i) to demonstrate the performance of the algorithm (ii) to validate performance expression (15). In the simulation below, $m = 2$ Gaussian sources are impinging on a half-wavelength uniform linear array of $n = 4$ sensors. The sources are assumed to be in the far-field of the array: the direction-of-arrival are respectively set to $\phi_1 = 0.1$ and $\phi_2 = 0.5$, so that the $(p, q)$-th entry of the mixture matrix $A$ is $\exp(j \pi \phi_p \phi_q)$ (note that the informations on the array manifold are not exploited by our algorithm). The source signals have unit variance and each one is generated by filtering a complex circular white Gaussian processes by a complex autoregressive (AR) model of order 1 with coefficient $\rho = 0.85 e^{i \theta}$. The number of correlation matrices taken into consideration is $K = 3$ correlation matrices (i.e. $R(1)$, $R(2)$ and $R(3)$); the overall rejection level is evaluated over 500 independent realizations.

In figure 1, the rejection level $I_{\text{overall}}$ is plotted in dB against $\sigma$ in dB. The figure is for $\theta_1 = 0.5$, $\theta_2 = \theta_1 + \theta$ and data block size of 1000 samples. On the plots, the curves are labeled with the spectral shift $\theta$. We note that the fitting between empirical and theoretical performance values is better at high SNR and large spectral shift. This means that the asymptotic performance is reached faster in this range of parameters. In figure 2, the rejection level $I_{\text{overall}}$ is plotted in dB against sample number. The figure is for a spectral shift $\theta = 0.2$. On the plots, the curves are labeled with the noise level $\sigma$ in dB. This figure shows that the asymptotic closed form expressions of the rejection levels is pertinent from snapshot length of about 50 samples.
This means that asymptotic conditions are still reached for short data block size.

Fig. 1 Performance vs noise level $\sigma$.

Fig. 2 Performance vs sample number $T$.

6. CONCLUSION

In this paper, the problem of blind separation of linear spatial mixture based on second order statistics has been investigated. A solution based on the joint diagonalization of a set of whitened correlation matrices has been proposed, along with an efficient algorithm to solve it. A closed form expression for the performance criterion of the method has been developed. Numerical simulations have been performed to evidence the usefulness of the method and to support our theoretical claims.

REFERENCES


