Second order blind separation of temporally correlated sources.

A. Belouchrani, K. Abed-Meraim, J.F.Cardoso, E.Moulines

Télécom Paris, CNRS URA 820, GdR TdSI
46 rue Barrault, 75634 Paris Cedex 13
email: cardoso@sig.enst.fr

Abstract

Blind separation of sources consists in recovering a set of statistically independent signals whose only mixtures are observed. Such instantaneous mixtures occur in narrow band array data which can then be processed without knowing the array manifold (blindness). This paper introduces a new source separation technique exploiting the possible time coherence of the source signals. In contrast to other previously reported techniques, the proposed approach relies only on second-order statistics, being based on a 'joint diagonalization' of correlation matrices. The effectiveness of the method in difficult contexts is illustrated by numerical simulations.

1 Introduction: Blind source separation.

The so-called 'source separation' problem consists in the identification of the independent components in a random vector. The term 'source' is generic but hints at the main application of source separation which is in the field of array processing. When an array of n sensors samples the fields radiated by m narrow band sources its output is classically modeled as a random vector made of m one-dimensional components, possibly corrupted by additive noise. Source separation may be obtained by first identifying the directional vectors associated to each of these components and then by (obliquely) 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. Hence, blind source separation is essentially unaffected by errors in the propagation model or in array calibration. All the source separation proposed 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 white, 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 correlated, blind identification is possible based on spatial correlation matrices [6]. These matrices (see below) show a simple structure which allows for straightforward blind identification procedures based on eigen-decomposition.

In this paper, we introduce an original blind identification technique, based on joint diagonalization of a set of correlation matrices. Robustness is significantly increased, at low additional cost, by processing such a matrix set rather than a unique matrix as in [6]. The paper is organized as follows. The data model and assumptions, the statistics to be used and the problem to be solved are stated the first section. Next section gives identifiability conditions and describes the proposed solution. The following section describes an efficient implementation of our approach.

2 Blind separation with correlation matrices.

2.1 Model and statistics.

Consider an array of n sensors receiving signals from m narrow band sources. The array output denoted $x(t)$ is a $n \times 1$ complex random vector. Corrupted by independent additive noise denoted $n(t)$, it is classically modelled as:

$$x(t) = y(t) + n(t) = As(t) + n(t) \tag{1}$$

where $s(t)$ is a m complex vector whose p-th component denoted $s_p(t)$ is the signal emitted by the p-th source. The $n \times m$ complex matrix A is assumed to have full rank but otherwise unknown. The source signals are second order stationary, mutually uncorrelated processes. We denote

$$E(s_i(t)s_j{}^*(t')) = \rho_i(t-t')\ \delta_{ij} \quad \forall t, t' \tag{2}$$

The additive noise n(t) is assumed to be spatially and temporally white and uncorrelated with the source signals. Under these assumptions, the correlation matrices take a simple structure :

$$R(\tau) \stackrel{\text{def}}{=} E\left(x(t+\tau)x^*(t)\right) = AS(\tau)A^H + \delta(\tau)\sigma I \tag{3}$$

where $S(\tau) \stackrel{\text{def}}{=} E\left(s(t+\tau)s^*(t)\right)$ is the correlation matrix of the source signals; by (2), it is a diagonal matrix with entries $\rho_i(\tau)$. In (3) as in the following, a vector is transconjugated by the $*$ superscript and a matrix by the H superscript.

2.2 Indeterminations.

Complete blind identification of matrix A is impossible in the blind context because the exchange of a fixed scalar factor between a source signal and the corresponding column of A leaves the observations unaffected. We take advantage of this fundamental indetermination to assume, without any loss of generality, that each signal has unit variance, *i.e.* $S(0) = I$. Note that this normalization leaves the phase of each column of A undetermined. Also note that the *numbering* of the signals is immaterial. It follows that the best that can be done based only on observations is to determine A up to a permutation and phase shifts of its columns. The crucial point is that these indeterminations do not impede source separation. If A is estimated up to permutation and phase shifts, it still allows to determine the source signals up to the corresponding fixed permutation and phase shifts.

Let W denote a $m \times n$ complex matrix such that $Wy(t)$ is spatially white. Such a matrix is called a *whitening matrix*. Note that besides whitening the signal part of the observations, multiplication by a whitening matrix W also reduces the array output to a m-dimensional vector. A whitening matrix verifies, by definition

$$I = E((Wy(t))(Wy(t))^*) = WAS(0)A^H W^H = (WA)(WA)^H \tag{4}$$

showing that if W is a whitening matrix, then WA is a $m \times m$ unitary matrix. It follows that for any whitening matrix W, there exists a $m \times m$ unitary U such that

$$WA = U \quad \text{or} \quad A = W^{\#}U \tag{5}$$

where the superscript $\#$ denotes the pseudoinverse.

Since $AS(0)A^H = R(0) - \sigma I$, equations (4) shows that 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 [7]. In any case, as shown by (5), 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 (3) which may carry the missing information if the source signals are temporally correlated.

2.3 A 2nd-order identification principle.

Let W denote a whitening matrix and consider the whitened correlation matrices $\underline{R}(\tau)$ defined as

$$\forall \tau \neq 0 \quad \underline{R}(\tau) = WR(\tau)W^H. \tag{6}$$

These $m \times m$ complex matrices are nothing but the correlation matrices of the process $Wx(t)$. By (3) and (5), we get the key relation:

$$\forall \tau \neq 0 \quad \underline{R}(\tau) = US(\tau)U^H. \tag{7}$$

Since U is unitary and $S(\tau)$ is diagonal, the latter just means that any whitened correlation matrix is diagonalized by the unitary transform U.

A simple blind identification principle then is the diagonalization of whitened correlation matrices. This idea may be found in slightly different forms in [5] and in [6]. Let us recall that while eigenvalues are uniquely defined, this is not the case of eigenvectors. For distinct eigenvalues, eigenvectors are determined, if normalized to unit norm, up to phase shifts and ordering. Since the normalized eigenvectors are found in the columns of U, we find the same indetermination in eigendecomposition (or diagonalization) as in the source separation problem. the diagonalization of some $\underline{R}(\tau)$ provides us with the relevant parameters if it has distinct eigenvalues.

True indetermination arises in the case of degenerate eigenvalues. It does not seem possible to *a priori* determine some delay value τ such that the eigenvalues of $\underline{R}(\tau)$ 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 some eigenvalues of $\underline{R}(\tau)$ comes close to degeneracy, the robustness of determining U from an eigendecomposition is seriously affected.

The situation is more favorable if we consider simultaneous diagonalization of a set $\{\underline{R}(\tau_i)|i = 1, \cdots, p\}$ of p whitened correlation matrices. This set is simultaneously diagonalized by U as in (7). The matrix U is essentially unique (in the sense discussed above) if and only if for any pair (i, j) of sources, there exists at least one lag τ_l in $(\tau_1, \cdots, \tau_p)$ such that $\rho_i(\tau_l) \neq \rho_j(\tau_l)$. We omit the straightforward proof of this statement. This makes clear that sources cannot be separated if they have identical normalized spectra.

It is the main point of this contribution to consider the *joint diagonalization* of several correlation matrices. This approach is intended to reduce the likelihood that the diagonalizing matrix is not uniquely determined; maybe more importantly it is intended to the increase statistical efficiency by basing the estimation of U on a larger set of statistics. This program is completed in the following by defining 'joint diagonalization' and by describing an efficient algorithm to this end.

2.4 Joint diagonalization

Of course, *exact* joint diagonalization of arbitrary matrices is generally impossible. We now define *approximate* joint diagonalization of a set of arbitrary matrices. We start by noting that the diagonalization of a single $n \times n$ matrix hermitian M may be understood as minimizing under unitary transform the sum of the squared moduli of all the off-diagonal terms. By constancy of the norm in unitary transforms, this is also equivalent to minimizing the criterion

$$C(M, V) \stackrel{\text{def}}{=} -\sum_i |v_i^* M v_i|^2 \tag{8}$$

over the set of unitary matrices $V = [v_1, \cdots, v_n]$. We quite naturally define the joint approximate diagonalization of a set $\mathbf{M} = \{M_k|k = 1..K\}$ of K arbitrary $n \times n$ matrices as the minimization of the criterion:

$$C(V) \stackrel{\text{def}}{=} -\sum_k C(M_k, V) = -\sum_{ki} |v_i^* M_k v_i|^2 \tag{9}$$

under the same unitary constraint. In this definition, the matrix set $\mathbf{M}$ is not restricted to contain hermitian (or more generally normal) matrices. Definition (9) applies to any set of arbitrary square matrices. A joint approximate diagonalization algorithm is given in appendix: it is a generalization of the Jacobi technique [8] for the exact diagonalization of a single hermitian matrix. As such, its computational cost is (not surprisingly) essentially equal to K times the cost of diagonalizing a single hermitian matrix by the Jacobi technique.

3 An algorithm for the blind separation of sources.

3.1 Implementation.

We have implemented the following algorithm. First the sample covariance $\hat{R}(0)$ is formed from T data samples and diagonalized as

$$\hat{R}(0) = L\Delta L^H \tag{10}$$

where $L = [l_1, \cdots, l_n]$ and $\Delta = diag(\lambda_1, \cdots, \lambda_n)$ with $\lambda_i \geq \lambda_j$ for $i > j$. The number of sources may be estimated from the spectrum Δ; this procedure is not specifically addressed in this paper. An estimate $\hat{\sigma}$ of the noise power is formed by averaging the $n - m$ smallest eigenvalues in Δ. A whitening matrix $\hat{W}$ is then computed as:

$$\hat{W} = \Delta' L'^H \text{ where } \Delta' = diag((\lambda_1 - \hat{\sigma})^{-\frac{1}{2}}, \cdots, (\lambda_m - \hat{\sigma})^{-\frac{1}{2}})) \text{ and } L' = [l_1, \cdots, l_m] \tag{11}$$

The data set is whitened into $\hat{W}x(t)$, which allows to form *in dimension* m the sample estimates $\underline{\hat{R}}(\tau)$ of the whitened correlation matrices for $\tau \in \{\tau_j | j = 1, \cdots, p\}$, a fixed set of time lags. A unitary matrix $\hat{U}$ is then obtained in the joint diagonalization (see appendix) of the set $\{\underline{\hat{R}}(\tau_j) | j = 1, \cdots, p\}$. The source signals are estimated as $\hat{s}(t) = \hat{U}^H \hat{W} x(t)$. If desired, an estimate of A is $\hat{A} = \hat{W}^{\#}\hat{U}$.

3.2 Performance.

The performance is sensibly characterized in terms of signal rejection. After blind identification, the estimated source signals are $\hat{s}(t) = \hat{A}^{\#}x(t) = \hat{A}^{\#}As(t) + \hat{A}^{\#}n(t)$ where $\hat{A}^{\#}$ is the pseudo inverse of the estimated matrix $\hat{A}$ of the mixture matrix, obtained as $\hat{A}^{\#} = \hat{U}^{\#}\hat{W}$. The matrix $\hat{P}$ defined by $\hat{P} = \hat{A}^{\#}A$ should be close to some matrix P with only one zero phase term in each row and each column (phase and permutation indetermination). The p-th estimated source signal is :

$$\hat{s}_p(t) = \sum_{q=1,n} \hat{P}_{pq} s_q(t) \tag{12}$$

and contains the q-th source signal at level $|\hat{P}_{pq}|^2$ since source signals are normalized to have unit power. A measure of the global quality of a separating matrix then is the overall rejection level:

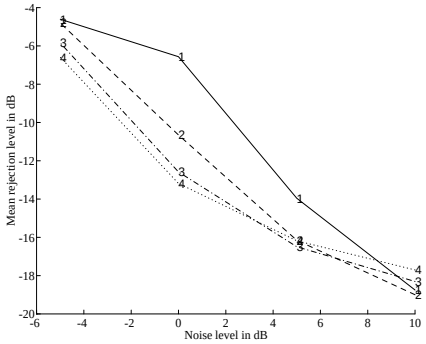
$$I_{perf} \stackrel{\text{def}}{=} \sum_{q \neq p} |\hat{P}_{pq}|^2 \tag{13}$$

where we have assumed for convenience that $\hat{P}$ is close to diagonal rather than to some other permutation matrix.

We ran simulations with two gaussian sources and an array of 5 sensors. The sources are 'located' at $\phi_1 = 0$ and $\phi_2 = 0.4$, in the sense that the q-th coordinate of the p-th column of A is $exp(j\pi q\phi_p)$. The additive gaussian noise has covariance $R_n = \sigma I$. The source signals have unit variance and each one is generated by filtering a complex circular white Gaussian processes by an AR1 model with coefficient $a_p = 0.85e^{j\theta_p}$. Statistics are evaluated on data blocks of 500 samples and the overall rejection level is evaluated over 100 realizations. On the plots, the curves are labeled with the number of correlation matrices used in the identification. When p correlation matrices are jointly diagonalized, the times lags implicitly involved are $\tau_1, \cdots, \tau_p$, where τ_i is i times the time unit.

In figure 1, the rejection level I_{perf} plotted in dB against σ^{-1} in dB. The figure is for $\theta_1 = 0.5$ and $\theta_2 = 0.55$: we are dealing here with a rather small spectral difference. At low SNR, i.e. for $\sigma = 5$ dB, the rejection level is close to -6 db meaning that the method is essentially useless in this range of parameters. Lowering the noise level to 0 dB, joint diagonalization of 3 or 4 matrices is seen to yield a convincing -13 dB overall rejection, which shows the benefit of using several correlation matrices. At high SNR ($\sigma = -10dB$), the diagonalization of a single correlation matrix appears sufficient to reach a satisfactory rejection of -18 dB.

In figure 2, the noise level is kept constant at 0 dB and $\theta_1 = 0.5$. We let θ_2 vary as $\theta_2 = \theta_1 + \delta\theta$. The plot shows the rejection level I_{perf}in dB plotted against the 'spectral shift' $\delta\theta$. The plot shows that there is a significant increase in performance for small spectral difference by including 3 or 4 correlations matrices in the joint diagonalization criterion.



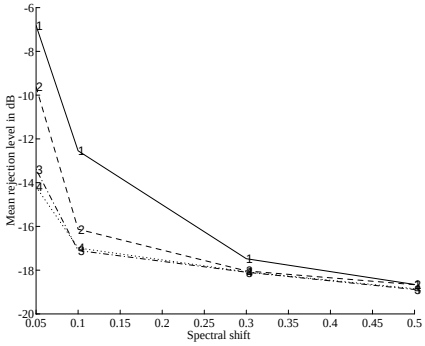


Fig. 1 (left) Performance vs σ^{-1}. Fig. 2 (right) Performance vs the 'spectral shift' $\delta\theta$.

Conclusion

Blind identification of spatial mixtures (source separation) is the processing of narrow band array data without knowing the array manifold. In presence of correlated sources, it admits straightforward solutions based on the diagonalization of some correlation matrix. By defining 'joint approximate diagonalization', an arbitrary set of correlation matrices can be used in the identification. The benefits are twofold: decreasing the likelihood of the cases of indetermination and increasing the statistical efficiency. This was illustrated by numerical experiments which show the necessity of exploiting several correlation matrices in difficult contexts (low SNR, sources with little spectral difference). An efficient algorithm for joint diagonalization, inspired from the Jacobi technique, was described. Asymptotic performance analysis of the proposed technique is under way and will be presented in a forthcoming communication.

Appendix: A joint approximate diagonalization algorithm

We generalize [9] the Jacobi technique for diagonalizing a unique hermitian matrix to the problem of joint approximate diagonalization of a set of arbitrary matrices. The proposed method consists in minimizing the diagonalization criterion (9) by successive Givens rotations, which leads to solve the same problem for K 2×2 matrices:

$$H_k = \begin{bmatrix} a_k & b_k \\ c_k & d_k \end{bmatrix} \tag{14}$$

for $k = 1, \ldots, K$. A unitary matrix V is sought such that $H'_k = V^H H_k V$ $(k = 1, \ldots, K)$ minimizes the criterion (9). The unitary transformation V is parameterized by a complex Givens rotation:

$$V = \begin{bmatrix} cos\theta & e^{j\phi} sin\theta \\ -e^{-j\phi} sin\theta & cos\theta \end{bmatrix}, \tag{15}$$

Denoting a'_k, b'_k, c'_k and d'_k the coefficients of H'_k, optimization of (9) amounts to finding θ and ϕ 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 (9) is equivalent at each Givens step to the maximization of Q:

$$Q \stackrel{\text{def}}{=} \sum_k |a'_k - d'_k|^2 \tag{16}$$

It is easily found that

$$a'_k - d'_k = (a_k - d_k) cos2\theta - (b_k + c_k) sin2\theta cos\phi - j(c_k - b_k) sin2\theta sin\phi \tag{17}$$

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

$$u^T \stackrel{\text{def}}{=} [a'_1 - d'_1, \cdots, a'_K - d'_K] \tag{18}$$

$$v^T \stackrel{\text{def}}{=} [cos2\theta, -sin2\theta cos\phi, -sin2\theta sin\phi] \tag{19}$$

$$g_k^T \stackrel{\text{def}}{=} [a_k - d_k, b_k + c_k, j(c_k - b_k)] \tag{20}$$

the K equations (17) may be written in the form $u = Gv$ where $G^T \stackrel{\text{def}}{=} [g_1, \cdots, g_K]$ so that Q also reads

$$Q = u^H u = v^T G^H G v = v^T \, Real(G^H G) \, v \tag{21}$$

where we have used that, $G^H G$ being hermitian by construction, its imaginary part is anti-symmetric, hence contributing nothing to the above quadratic form. The last step is to recognize that the particular parameterization (19) of v is equivalent to the condition $v^T v = 1$. Maximizing a quadratic form under the unit norm constraint of its argument is classically obtained by taking v to be the eigen-vector of $Re(G^H G)$ associated to the largest eigenvalue. Recall that this is a real 3×3 symmetric matrix: the analytic expressions of the parameters of the Givens rotation are simply derived from the coordinates of the eigen-vector. More details are in [3]. The reader may check that setting $K = 1$ and H_1 hermitian, the above boils down to the standard Jacobi procedure. Also 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 is why the proposed procedure offers a cost similar to K times the diagonalization of a single matrix.

References

[1] P. Comon, "Independent component analysis," in *Proc. Int. Workshop on Higher-Order Stat., Chamrousse, France*, pp. 111–120, 1991.

[2] M. Gaeta and J.-L. Lacoume, "Source separation without a priori knowledge: the maximum likelihood solution," in *Proc. EUSIPCO*, pp. 621–624, 1990.

[3] J.-F. Cardoso and A. Souloumiac, "An efficient technique for blind separation of complex sources," in *Proc. IEEE SP Workshop on Higher-Order Stat., Lake Tahoe, USA*, 1993. To appear.

[4] C. Jutten and J. Hérault, "Independent component analysis versus PCA," in *Proc. EUSIPCO*, pp. 643–646, 1988.

[5] L. Féty, "Méthodes de traitement d'antenne adaptées aux radio-communications," *Thèse de doctorat. Télécom Paris*, June 1988.

[6] L. Tong, R. Liu, and Y. H. V.C. Soon, "Indeterminacy and identifiability of blind identification," *IEEE Tr. on CS*, vol. 38, pp. 499–509, May 1991.

[7] L. Tong, Y. Inouye, and R. Liu, "A finite-step global convergence algorithm for the cumulant-based parameter estimation of multichannel moving average processes'," in *Proc. ICASSP*, pp. 3445–3448, 1991.

[8] G. H. Golub and C. F. V. Loan, *Matrix computations*. The Johns Hopkins University Press, 1989.

[9] A. Souloumiac and J.-F. Cardoso, "Comparaison de methodes de separation de sources," in *Proc. GRETSI, Juan les Pins,France*, 1991.