Separation of sources using high-order cumulants

Pierre Comon

Separation of sources using higher-order cumulants[1]

Pierre Comon

*THOMSON SINTRA, BP 53,
F-06801 France*

ABSTRACT

The problem is to recover stochastic processes from an unknown stationary linear transform. Our contribution is two-fold. First we focus on instantaneous mixtures: observation $\mathbf{e}(t)$ is assumed to write as a regular linear transform of the sources, $\mathbf{x}(t)$, as $\mathbf{e}(t)=B_0\mathbf{x}(t)$. The only assumption requested is that the sources $x_i(t)$ are mutually independent, and no additional knowledge upon their statistics is necessary provided they are not normal. Extensions to convolutional mixing are then pointed out, namely cases where $\mathbf{e}(t)=A(t)*\mathbf{x}(t)$ where $A(t)$ has a rational transfer function. Sensitive improvements to the algorithm of Giannakis et al for MA identification are included. Multivariate ARMA identification can be split into three successive estimation problems: AR identification, monic MA identification, and estimation of B_0 in last position.

Keywords: Antenna array processing, Noise reduction, Independent component analysis, Direction of arrival, MA identification, Deconvolution.

1. INTRODUCTION

Historical remarks

To our knowledge, this problem has been first investigated in 1986 by C. Jutten in his PhD thesis, that has been reproduced in a concise form in [8]. The solution proposed was using a set of linear neurons totally interconnected to each other, that were working in an unsupervised self-learning manner. For a few years, the behavior of the algorithm has not been understood. After the first analysis initiated in the PhD thesis of Fetty [5], a more complete explanation attempt has been proposed in [4]. These investigations have shown that the neuro-algorithm is able to separate the sources if they have non-normal and even probability distributions. However, for some pathological cases, due to a strong dependence on the initial value given to the unknown transform, the algorithm may converge extremely slowly, or may fail to give the correct solution, or even fail to converge at all. Since 1987, another interesting solution

[1]: Supported by D.R.E.T. Paris.

was proposed by Cardoso, which is based on the eigendecomposition of a matrix of particular fourth-order moments [1]. This technique was able to recover the sources when they had different non-normal probability densities. The author improved this recently by using more completely the information: in [2], eigenmatrices of the fourth-order moments tensor are considered. This second version of Cardoso's algorithm would perform better, for any set of non-normal probability distributions. Lastly Lacoume and Ruiz propose an exhaustive search of the absolute minima of a response function built from the sum of squares of cumulants [11]. This seems costly and does not fully take advantage of the polynomial form of cumulants in the present problem.

In this paper, the problem will be stated along the same lines as in [3], namely by using fourth-order cumulants. The principle of the solution is different from the above-mentioned methods. The use of cumulants (instead of moments) allows both to achieve a better conditioning of equations and to alleviate the notations, in particular in MA identification.

Applications include noise reduction when signal and noise distributions are unknown, direct estimation of direction of arrival when wavefronts are distorted, and more generally, parameter estimation.

Outline

Necessary and sufficient conditions for recovering the sources after instantaneous mixing are stated in section 2, and the proposed algorithm is able to perform the separation for any kind of non-normal probability densities. The algorithm is iterative but converges extremely fast. Sections 3 to 6 deal with convolutional mixing. The backbone of the method is based on splitting ARMA identification into three successive identification problems. First, the AR part is known to be identifiable only from covariance lags, but may be performed by using higher-order cumulants as well (section 6). Second, the MA part is identified from fourth-order cumulants by a technique directly inspired from the works of Giannakis and Mendel [6], and somewhat more accurate because of its least-squares fitting (especially for short data records). And third, identification of the instantaneous transform may be performed as described in section 2 (see figure 1). Our discussion will follow the reverse order, i.e., will start with (the most difficult) B_0 identification, and will end up with (the easiest) AR identification.

Notations

Consider the p×1 ARMA model:

$$\mathbf{e}(t) + \sum_{j=1}^{r} A_j\, \mathbf{e}(t\text{-}j) = \sum_{k=0}^{q} B_k\, \mathbf{x}(t\text{-}k), \tag{1}$$

where all coefficients B_k and A_j are p×p matrices, and $\mathbf{x}(t)$ is a white non-gaussian stationary noise with unit covariance matrix. Random vectors are boldface for convenience.

In the first section (section 2), we focus on the instantaneous model:

$$e(t) = B_0\, x(t), \tag{2}$$

where B_0 is any regular pxp matrix and x(t) is a (possibly white) process, whereas the general MA model is considered in section 5.

2. INSTANTANEOUS MIXTURE

2.1. Inherent undetermination in the problem

The goal is to find a pxp matrix, F, such that the process

$$s(t) = F\, e(t) \tag{3}$$

has mutually independent components. Hopefully s(t) will approximate the source process x(t), in a way that we define below. It has been shown in [3] that among the p^2 unknown parameters defining the transform F:

- p of them are not identifiable,
- p(p-1)/2 are identifiable only with the help of higher-order statistics (larger than 2),
- p(p-1)/2 parameters are identifiable from the second-order moments.

More precisely, if F is a solution to the problem, then so is the matrix F',

$$F' = \Lambda\, P\, F, \tag{4}$$

where Λ is a regular diagonal matrix, and P is a permutation. If we arbitrarily impose the output to have a unit covariance matrix, then we fix the p unidentifiable parameters of Λ, and the undetermination that remains is of the form

$$F' = \Delta\, P\, F, \tag{5}$$

where Δ is a diagonal matrix formed of ± 1's [3]. There exist other ways of fixing p entries of F, for instance set the diagonal entries to one, but we shall not mention them anymore in the sequel. We assume from now on that $\mathrm{cov}\{s(t)\}=I$.

Thus the free parameters of F may be defined as

$$F = Q\, L\, \Lambda, \tag{6}$$

where $L\Lambda$ is a lower triangular matrix (L being lower triangular with unit diagonal entries), and Q is orthogonal. The estimation of F can be carried out in two steps:

a) Compute the lower triangular matrix $L\Lambda$ by imposing that $\tilde{e}(t)=L\Lambda e(t)$ has a unit covariance matrix. This may be done adaptively without explicitly computing the Cholesky factor of $\mathrm{cov}\{e(t)\}$, as described in [3] for instance.

b) Compute the orthogonal matrix Q by resorting to cumulants of order higher than 2 (Sections 2.3 and 2.4 show how to do it).

The complete processing scheme is sketched on figure 1. Since step a) does not raise any particular problem, we shall assume without loss of generality that step a) has been already performed and that the observations are spatially uncorrelated (i.e. we assume that we actually observe $\tilde{e}(t)$ instead of $e(t)$).

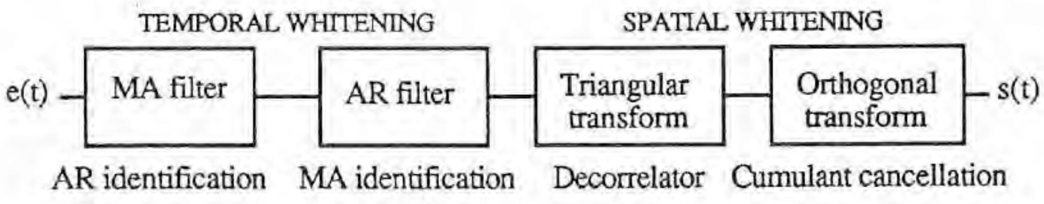

Figure 1

2.2. Input-output cumulant relations

2.2.1. Obtaining of a system of polynomials

The output cumulants of order n, denoted abstractly $\mathrm{cum}^{(n)}(s)$, are linear functions of input cumulants $\mathrm{cum}^{(n)}(e)$, and homogeneous polynomials of degree n in the unknowns (the entries of F). In order for the outputs $s_j(t)$ to be independent, it is necessary and sufficient that the joint characteristic function $\phi(s)$ separate into the product of marginal characteristic functions, $\phi(s_j)$. In other words, by taking the logarithm and then the expansion, this implies that all the cross cumulants of any order are null:

$$\mathrm{cum}^{(\sigma(1)+\ldots+\sigma(p))}(s_1^{\sigma(1)},\ldots,\ s_p^{\sigma(p)}) = 0. \tag{7}$$

If we restrict our attention to a finite order n, then the system obtained (7) is formed of

$$N(n,p) = C_{n+p-1}^{p-1} - p$$

polynomial equations of degree n. So for fixed n and as p increases, the number of unknowns grows as p^2 whereas the number of equations of degree n grows as p^n. The system (7) alone is overdetermined for $n>2$ and $p\geq 2$. The conclusion we can derive from this is that it is not realistic to solve such a system for large values of p, even for moderate n. If $n=4$ for example, we have $N=3$ polynomials of one variable for $p=2$, but already $N=65$ polynomials each of 10 variables for $p=5$. In practice, we are forced to utilize a limited number of equations of low order, and this will be broadly sufficient provided they are *well-conditioned*.

2.2.2. Solution of systems of polynomial equations

Solving a polynomial system of equations is obviously equivalent to finding the Greatest Common Divisor (GCD) to all polynomials. Unfortunately, due to measurement and coding errors, the GCD of two polynomials with real (or complex) coefficients turns out to be always one in practice, if computed in a standard way. Thus, it is necessary to look for an approximate GCD.

It is possible to define a *quasi-GCD*, D(X), of two polynomials of a single variable, A(X) and B(X), by minimizing a norm $\|U(X)A(X)+V(X)B(X)-D(X)\|$ instead of defining an exact GCD C(X) as satisfying C(X)=U(X)A(X)+V(X)B(X). So, the backbone of the GCD computation is still the Bezout identity. Of course, this principle can be extended to any finite number of polynomials. Unfortunately, neither the Bezout identity nor the standard Euclidean division are valid anymore in *K[X,Y]*, even if the GCD can still be defined as in any factorial ring [10]. Consequently, the computation of a quasi-GCD raises serious problems for polynomials of several variables, and to our knowledge there exist no numerically stable algorithm carrying out this task.

2.3. Solution for p=2 sources

Assume $e_i(t)$ are uncorrelated to each other and that the transform F we are looking for is orthogonal. Then F may be written as a Givens rotation

$$F = c\begin{pmatrix} 1 & \theta \\ -\theta & 1 \end{pmatrix},\ c = 1/\sqrt{1+\theta^2},\ \text{and}\ B_0 = \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix}. \tag{8}$$

Since x(t) is zero-mean, so is e(t), and the cumulants of order three of $\{e_1(t), e_2(t)\}$ are identically zero if one component $e_i(t)$ has a symmetric probability density. If normal densities are rarely observed in the real world, actual densities are often near symmetric, and the cumulants of order 3 may be very badly conditioned. This motivates us to utilize cumulants of order 4.

Denote γ_{ij} the cumulants $\mathrm{cum}^{(i+j)}\{e_1(t)^i, e_2(t)^j\}$, and m_{ij} the moments $E\{e_1(t)^i, e_2(t)^j\}$. Similarly denote in upper case $\Gamma_{ij}=\mathrm{cum}^{(i+j)}\{s_1(t)^i, s_2(t)^j\}$, and $M_{ij}=E\{s_1(t)^i, s_2(t)^j\}$. Now equations (7) write:

$$\Gamma_{31}/c^4 = -\gamma_{13}\theta^4 + (\gamma_{04} - 3\gamma_{22})\theta^3 + 3(\gamma_{13} - \gamma_{31})\theta^2 - (\gamma_{40} - 3\gamma_{22})\theta + \gamma_{31}, \tag{9-a}$$

$$\Gamma_{13}/c^4 = -\gamma_{31}\theta^4 - (\gamma_{40} - 3\gamma_{22})\theta^3 + 3(\gamma_{31} - \gamma_{13})\theta^2 + (\gamma_{04} - 3\gamma_{22})\theta + \gamma_{13}, \tag{9-b}$$

$$\Gamma_{22}/c^4 = \gamma_{22}\theta^4 + 2(\gamma_{31} - \gamma_{13})\theta^3 + (\gamma_{40} - 4\gamma_{22} + \gamma_{04})\theta^2 + 2(\gamma_{13} - \gamma_{31})\theta + \gamma_{22}. \tag{9-c}$$

The problem consists of solving this system for θ. Since in this case we are dealing with polynomials of a single variable, we could compute the quasi GCD of the system and then solve it for θ. However, this would hide some interesting insights and there is a better way to proceed. Denote κ_{ij} the cumulants $\mathrm{cum}^{(i+j)}\{x_1(t)^i, x_2(t)^j\}$; then input cumulants γ_{ij} satisfy:

$$\begin{aligned}
\gamma_{40} &= \kappa_{40}\alpha^4 + \kappa_{04}\beta^4 \\
\gamma_{31} &= -\kappa_{40}\beta\alpha^3 + \kappa_{04}\beta^3\alpha. \\
\gamma_{22} &= (\kappa_{40} + \kappa_{04})\beta^2\alpha^2. \\
\gamma_{13} &= -\kappa_{40}\beta^3\alpha + \kappa_{04}\beta\alpha^3. \\
\gamma_{04} &= \kappa_{40}\beta^4 + \kappa_{04}\alpha^4.
\end{aligned} \tag{10}$$

By eliminating α, β and the κ_{ij}'s in (10), it can be shown that input cumulants satisfy the additional property:

$$(\gamma_{13}-\gamma_{31})^2 - (\gamma_{04}+\gamma_{40})\gamma_{22} + 2\gamma_{22}^2 = 0. \tag{11}$$

This relation imposes the vector $[\gamma_{40},\gamma_{31},\gamma_{22},\gamma_{13},\gamma_{04}]$ to lie on a 4-dimensional manifold in $\mathbb{R}^5$. Though they are difficult to produce, similar relations exist when p>2. In fact, if n_e and n_d denote the number of equations and the number of degree of freedom in B_0 respectively, there exist in general n_e-n_d free relations between the γ_{ij} due to linearity. With the help of this result, it can be seen that equation of (9-c) has two double roots, θ_o and $1/\theta_o$, where θ_o is in]-1,1]:

$$\theta_o = \rho/2 - \mathrm{sign}(\rho)\sqrt{\rho^2/4+1}, \text{ with } \rho = (\gamma_{13}-\gamma_{31})/\gamma_{22}. \tag{12}$$

It can be checked that the roots (12) are the unique common solution of system (9). Thus, we have a direct theoretical solution for the transform F in the noiseless case. Now when the cumulants are estimated from actual data, there will not exist an exact common root to (9), but only an approximate one. For instance, equation (9-c) may have either no real roots, or four distinct real roots. Solution (12) can still be computed in this case, and is actually a good approximation to the Least-Squares solution (LS). When p=2, solution (12) has always been accurate in all our simulations.

2.4. More than 2 sources

2.4.1. Weaker independence criterion

In the following, instead of cancelling all the possible cross-cumulants of order four, we shall only cancel those involving pairs of outputs. Pairwise independence is known to be a weaker condition in general, but our conjecture is that both notions are equivalent in our problem. In fact, e(t) is a *linear* transform of a process with *mutually independent* components, **x**(t).

2.4.2. Initialization

The problem becomes much more difficult since we have to deal with p(p-1)/2>1 unknowns (the entries of the orthogonal matrix, F). The first approach would be to search the minima of a single functional formed of the sum of the squares of all polynomials (there are 3p(p-1)/2 of them if we consider only pairwise independence at order 4). This procedure, already proposed in [11] in the p=2 case, has two inconveniences. First, one adds spurious roots by elevating the degree up to 8. And second, we are forced to carry out an exhaustive search in a p(p−1)/2-dimensional space, which is very costly. A search seems difficult to avoid, but a local search is preferred to an exhaustive one. In any *local* search, the most important issue is initialization, and it is performed as described in the following.

The orthogonal matrix that we are left to find can be expressed as the product of p(p-1)/2 Givens rotations in a canonical ordering, followed by a signs-matrix, Δ [3]. Yet, a Givens rotation acts in a plane and may be defined by a single parameter, its tangent θ for instance. This is one way of parametrizing any orthogonal matrix up to an undetermination of the type ΔP. One can perform one sweep by estimating each of these rotations in turn via the technique described in section 2.3 [3]. Our initialization procedure consists of a single sweep, and does not provide in general the best approximation to mutual independence of the outputs, but rather imposes independence of the last pair processed.

2.4.3. Least-Squares refinement

This is why a refinement processing may be necessary when p>2. Let us describe the case p=3, the general case being a straight extension. Denote F_0 the orthogonal matrix estimated in the initialization step. At iteration k, we look for a matrix $\delta F(k)$, close to identity, such that

$$F_k = \delta F(k)\, F_{k-1} \tag{13}$$

is improving the solution of system (7) in the LS sense. Since $\delta F(k)$ is not imposed to be orthogonal, it is necessary to add the equations insuring cancellation of second-order moments to the system, as well as those imposing unit variance outputs. For p=3, the system given by pairwise cumulants is:

$$[\, \Gamma_{310}, \Gamma_{301}, \Gamma_{031}, \Gamma_{130}, \Gamma_{103}, \Gamma_{013} \,]^T = \mathbf{0}, \tag{14-a}$$

$$[\, \Gamma_{110}, \Gamma_{101}, \Gamma_{011}, \Gamma_{200}, \Gamma_{020}, \Gamma_{002} \,]^T = [\, 0, 0, 0, 1, 1, 1 \,]^T. \tag{14-b}$$

Note that in (14) we have omitted symmetric pairwise cumulants, such as Γ_{220}, because of their poor conditioning (their zeros are indeed also stationary points). The linearized system is of the form

$$\Gamma(k)\, \mathbf{z}(k) = -\mathbf{y}(k), \tag{15}$$

and is solved in the LS sense for $\mathbf{z}(k)$. For the sake of clarity, here is the part of (15) defined by (14-a):

$$\delta F(k) = I + \begin{pmatrix} f_{11} f_{12} f_{13} \\ f_{21} f_{22} f_{23} \\ f_{31} f_{32} f_{33} \end{pmatrix},\ \Gamma(k)^T = \begin{pmatrix} 3\gamma_{220} & 3\gamma_{211} & \gamma_{400} & \gamma_{301} & 0 & 0 \\ 3\gamma_{211} & 3\gamma_{202} & 0 & 0 & \gamma_{400} & \gamma_{310} \\ 0 & 0 & 3\gamma_{121} & 3\gamma_{022} & \gamma_{130} & \gamma_{040} \\ \gamma_{040} & \gamma_{031} & 3\gamma_{220} & 3\gamma_{121} & 0 & 0 \\ \gamma_{013} & \gamma_{004} & 0 & 0 & 3\gamma_{202} & 3\gamma_{112} \\ 0 & 0 & \gamma_{103} & \gamma_{004} & 3\gamma_{112} & 3\gamma_{022} \end{pmatrix},\ \mathbf{z}(k) = \begin{pmatrix} f_{12} \\ f_{13} \\ f_{21} \\ f_{23} \\ f_{31} \\ f_{32} \end{pmatrix}, \tag{16}$$

$$\mathbf{y}(k) = [\, \gamma_{310}, \gamma_{301}, \gamma_{031}, \gamma_{130}, \gamma_{103}, \gamma_{013} \,]^T.$$

In the formula above, γ_{ijr} and Γ_{ijr} denote the input and output cumulants at iteration k. Of course, the complete form of system (15) also includes equations provided by (14-b) so that (15) is overdetermined. Notations could be simplified by resorting to Kronecker product; however, some entries would be repeated, hence according different weights to each cumulants, which is not suitable for a Total Least Squares solution. In a real-time processing context, $\Gamma(k)$ and $\mathbf{y}(k)$ are updated at each time step, and (15) solved only once at each time step. The solution of (15) serves as tracking algorithm. On the other hand, in a batch processing, one can run several steps of (15).

The issue of updating $\Gamma(k)$ and $\mathbf{y}(k)$ must be addressed. Of course, we do not want to compute the output cumulants from the outputs since: (i) in the real-time processing, the iterative solving of (15) makes outputs non-stationary (ii) in the batch processing it would be too costly to recompute from scratch all the cumulants from long data records. Instead, we can compute output cumulants as polynomial functions of input cumulants by just plugging back the value of the new transform F_k given by (13). The update of input cumulants, relevant for real-time processing, can be performed via exponential averaging [3]. In batch processing, they are computed once for all.

Let us turn to stopping criteria. We impose $\|z(k)\|\geq\varepsilon$, otherwise the transform is negligible and should not be applied; and also $\|z(k)\|\leq\eta$, because z(k) must not be too large in order for the linear approximation to keep valid. Denote σ_{max} and σ_{min} the extreme singular values of $\Gamma(k)$. The first condition is equivalent to $\sigma_{max} \leq \|y(k)\|/\varepsilon$, and can be checked easily before actually completing the solution of the system, once the singular values of Γ have been computed. The second is faced by retaining only sufficiently large singular components such that: $\sigma_{min} \geq \|y(k)\|/\eta$. If this condition cannot be satisfied, even by σ_{max}, then we stop. Otherwise, the LS solution is computed by setting the singular values smaller than $\|y(k)\|/\eta$ to zero. Of course a test on the maximum number of iterations must be included.

3. COLORED TRANSFORMS

The first possibility to deal with colored transfer functions, would be to use stationarity and take the Fourier transform. As a matter of fact, the problem could be solved for each bin independently in the spectral domain with help of our algorithm described by (12-16). However, some problems arise: first, one should impose continuity conditions to $P(\omega)$ and $\Delta(\omega)$, and second, the computation of the solution becomes extremely costly for p>2. This procedure would eventually come to use particular properties of the trispectrum. The third approach we have retained is parametric.

4. IDENTIFICATION OF SCALAR MA MODELS

MA modelization is useful since it allows for taking into account delays and absorption in propagation of waves. As in [7], the goal is to obviate the minimum phase condition. Our contribution here consists of a modification of the algorithm proposed by Giannakis et al in [6]. The improvement is especially sensible for short data lengths. The observed process satisfies:

$$e(t) = \sum_{k=0}^{q} b_k \, x(t-k), \tag{16}$$

Denote $C(i,j)=\mathrm{cum}\{e(t),e(t+i),e(t+j),e(t+q)\}$ and $C_x=\mathrm{cum}\{x(t)^4\}$. Assuming that the process x(t) is white up to fourth order, we have:

$$C(i,j) = b_0\, b_i\, b_j\, b_q\, C_x \tag{17}$$

so that we deduce the following relations

$$b_j\, C(i,k) - b_i\, C(j,k) = 0,\ \forall k,\ 0 \le k \le q \tag{18-a}$$

$$b_j \sqrt{C(i,i)} - b_i \sqrt{C(j,j)} = 0,\ \forall i,j,\ 0 \le i < j \le q \tag{18-b}$$

Equation (18-a) and (18-b) provide us with $p(p+1)^2/2$ and $p(p+1)/2$ linear relations, respectively. The vector of coefficients, $\mathbf{b}=[b_1,b_2,\dots b_q]^T$, may be obtained by solving in the Least-Squares sense an overdetermined linear system $\mathbf{C}\ \mathbf{b} = \mathbf{c}$, where $\mathbf{c}$ contains all the factors of b_0, which is assumed equal to one. In [7] [6], only those p equations where i=k=0 were utilized. We find it very useful to also include the set of equations (18-a). On the other hand, equations (18-b) have shown to be worse conditioned, and thus would decrease the precision of the solution obtained if incorporated in the LS solving. For instance, the system (18-a) obtained for q=2 would be:

$$[b_2\ b_1] \begin{pmatrix} 0 & 0 & 0 & C(0,0) & C(0,1) & C(0,2) & C(1,0) & C(1,1) & C(1,2) \\ C(0,0) & C(0,1) & C(0,2) & 0 & 0 & 0 & -C(2,0) & -C(2,1) & -C(2,2) \end{pmatrix} =$$

$$b_0 \begin{pmatrix} C(1,0) & C(1,1) & C(1,2) & C(2,0) & C(2,1) & C(2,2) & 0 & 0 & 0 \end{pmatrix}.$$

See last section for simulations.

5. MULTIVARIATE MA MIXTURE

For the multivariate case, we agree with Giannakis et al in that the Kronecker formalism is of better use than the tensor formalism. In a similar manner as in the previous section, denote $C(i,j) = \mathrm{cum}\{\mathbf{e}(t+i) \otimes \mathbf{e}(t+j) \otimes \mathbf{e}(t+q) \otimes \mathbf{e}(t)\}$ and $\mathbf{C}_x = \mathrm{cum}\{\mathbf{x}(t) \otimes \mathbf{x}(t) \otimes \mathbf{x}(t) \otimes \mathbf{x}(t)\}$. Then we have:

$$C(i,j) = [\mathbf{B}_i \otimes \mathbf{B}_j \otimes \mathbf{B}_q \otimes \mathbf{B}_0]\ \mathbf{C}_x \tag{19}$$

Now denote $\bar{C}(i,j)=\mathbf{unvec}_p\{C(i,j)\}$ and $\bar{C}_x=\mathbf{unvec}_p\{C_x\}$, where subscript ($_p$) indicates the number of rows of the matrix $\bar{C}$ obtained; i.e. $\bar{C}$ is p^3xp since C is p^4x1. Now using the property that for any matrices B, G and $\bar{C}$ with appropriate dimension, $[B \otimes G]\mathrm{vec}\{\bar{C}\} = \mathbf{vec}\{G\bar{C}B^T\}$ [6] :

$$\bar{C}(i,k) = [\ \mathbf{B}_k \otimes \mathbf{B}_q \otimes \mathbf{B}_0]\ \bar{C}_x\, \mathbf{B}_i^T,\ \forall i,k,\ 0 \le \{i,k\} \le q. \tag{20}$$

Now two cases can be considered. First, if matrices B_i are regular, one can form the $p(p+1)^2/2$ block-equations and solve them for the $\mathbf{B}_i^{-1}$'s:

$$\bar{C}(i,k)\ \mathbf{B}_i^{-T} = \bar{C}(j,k)\ \mathbf{B}_j^{-T},\ 0 \le i < j \le q,\ 0 \le k \le q. \tag{21}$$

This result is the multivariate version of (18-a). Now matrices $\mathbf{B}_i$ are not necessarily regular and we must find another system valid under more general assumptions. It involves $p(p-1)(p+1)/2$ block-equations (somewhat less, but still largely overdetermined), and is obtained only by using the fact that $\mathbf{B}_0$ is regular (true by hypothesis):

$$\bar{C}(j,k) = \bar{C}(0,k)\ \mathbf{B}_0^{-T}\ \mathbf{B}_j^T,\ 0 < j \le q,\ 0 \le k \le q. \tag{22}$$

System (22) provides more accurate solution (in the TLS sense) than Giannakis's which considers only equations for which k=0 [6].

6. ARMA MIXTURES

The AR part of the model allows to include reverberation in the modeling. It is well known that it can be estimated in a first step by utilizing the Yule-Walker equations, for delays larger than q in order to get rid of the MA contribution. The use of higher-order statistics does not seem to be crucial at this point. Then the residual $\bar{e}(t)$ may be computed and satisfies the MA model:

$$\bar{e}(t) = e(t) + \sum_{j=1}^{r} A_j\, e(t-j) = \sum_{i=0}^{q} B_i\, x(t-i); \tag{24}$$

Denote $\bar{x}(t) = B_0\, x(t)$ and $\bar{B}_k = B_k B_0^{-1}$, then $\bar{e}(t)$ satisfies the monic MA model

$$\bar{e}(t) = \bar{x}(t) + \sum_{k=1}^{q} \bar{B}_k\, x(t-k). \tag{25}$$

Hence, the coefficient matrices $\bar{B}_k$ can be identified by the method proposed in section 5 since $\bar{B}_0 = I$. At last, the original process $x(t)$ is estimated from the equality $\bar{x}(t) = B_0\, x(t)$ by using the technique described in section 2.

7. SIMULATION RESULTS

7.1. Instantaneous mixture[2]

One of the most demonstrative example is the recovery of three noises identically distributed. We show below the results obtained for three uniform white noises and for various observation lengths (250, 500, 1000, 1500 and 2000). The nine pairwise cross-cumulants are considered, and we report here the norm of the 9-dimensional corresponding vector, γ. After the orthogonal transformation B_0, we have inputs cumulants such that $\| \gamma \|_2 = 0.701$ and $\| \gamma \|_\infty = 0.49$. The application of the inverse transform estimated as in section 2.4.2 leads to output cumulants represented on figure 2 (the average results obtained over 100 trials are plotted; variances are represented with errors bars). For a 2000 sample, the norms of source cumulants were equal to $\| \kappa \|_2 = 0.091$, and $\| \kappa \|_\infty = 0.050$, as indicated on figure 2. As a performance criterion, one may look also at the product F^*B_0, which is indeed very close to a $P\Delta$ form:

$$F^*B_0 = \begin{pmatrix} -0.026 & -0.999 & -0.034 \\ 0.999 & -0.026 & -0.001 \\ 0.001 & -0.034 & 0.999 \end{pmatrix}, \quad B_0 = \begin{pmatrix} -0.279 & -0.666 & 0.692 \\ 0.717 & -0.623 & -0.310 \\ 0.638 & 0.410 & 0.652 \end{pmatrix}.$$

[2]: The author wishes to thank his student Gilles Favre for his help in running the simulations of this section.

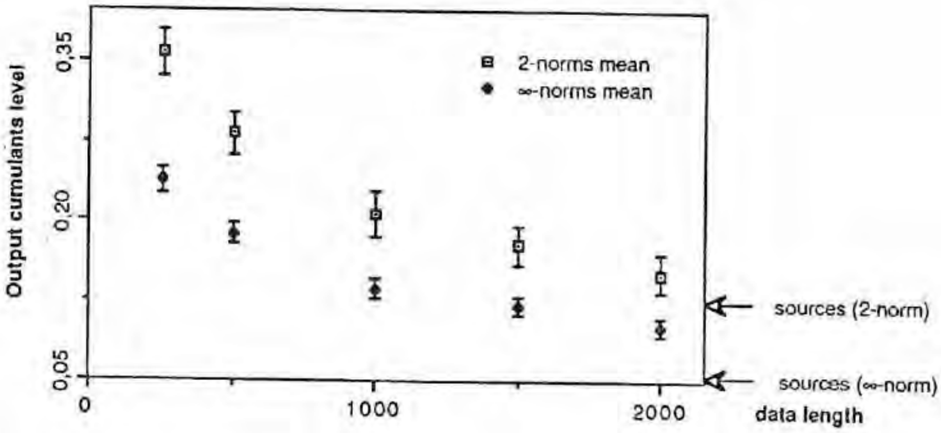

Figure 2

Recall the limitations of the methods proposed in [1] and [8] that work for different and symmetric probability density functions, respectively. Our algorithm performs successfully even for unsymmetric and equal pdf's. Moreover, in [8] the algorithm is iterative and fairly slow [4], and the elegant solution proposed in [2] requires the eigendecomposition of a fourth-order tensor, which might be very costly.

7.2 Scalar MA identification

In this section, we compare five different methods, obtained by solving systems of equations more and more overdetermined:

#1: this is the one of Giannakis et al [6], using p equations (square system of linear equations),

#2: the system is formed of the $p(p+1)$ first equations in (18-a), namely those for which $j=0$,

#3: the system contains all the $p(p+1)^2/2$ equations defined by (18-a),

#4: the system is (18-a) completed with the p equations of (18-b) for which $j=0$,

#5: the system is formed of all the $p(p+1)^2/2$ equations of (18-a) and the $p(p+1)/2$ square-root equations of (18-b). Thus, it contains $p(p+1)^2(p+2)/2$ equations.

Denote B and $\hat{B}$ the true and estimated (p+1)-vectors of coefficients, respectively. In our simulations, the criterion is the two-norm $\| B - \hat{B} \|$, and it has been computed for 50 independent realizations of a uniform zero-mean white noise, for a 2nd order MA filter defined by $B=[B_2,B_1,B_0]=[1.4792,-1.72,2.0]$. The results are summarized by the median and the standard deviation of the quantity $\| B - \hat{B} \|$, over the 50 realizations. Figure 3 shows the results obtained for data lengths 250, 500, 1000, 1500 and 2000 samples respectively. For short data lengths, the best estimator is #5, whereas for longer data records, all methods are roughly equivalent but one may prefer estimator #3 or #4. For an all-purpose use, estimator #3 seems the most attractive, and this probably holds true in the multivariate case (estimators #4 and #5 are not defined in the multivariate case in this paper).

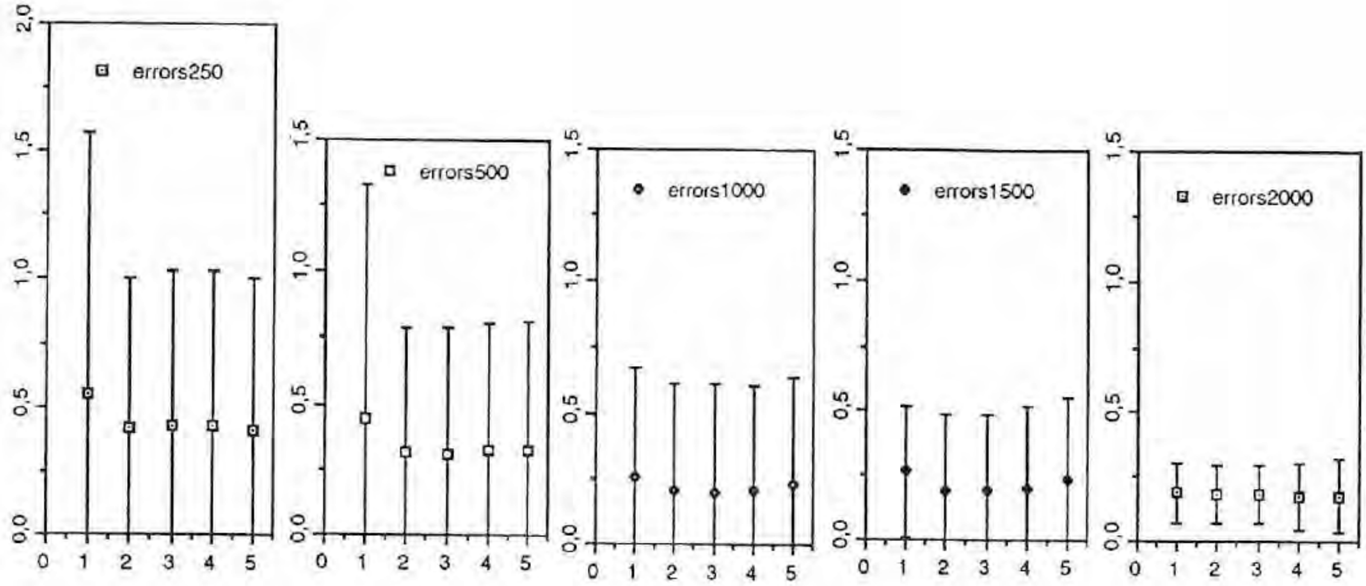

Figure 3

8. REFERENCES

[1] J.F.CARDOSO, "Source separation using higher-order moments", *Conference ICASSP*, Glasgow, may 1989.

[2] J.F.CARDOSO, "Blind identification of independent signals", *Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, june 1989.

[3] P.COMON, "Separation of stochastic processes", *Workshop on Higher-Order Spectral Analysis*, Vail, Colorado, june 1989.

[4] P.COMON, "Statistical approach to the Jutten-Herault algorithm", *NATO Workshop on Neuro-Computing*, les Arcs, France, february 1989.

[5] L.FETY, "Méthodes de traitement d'antenne adaptées aux radiotélécommunications", ENST PhD thesis, Paris, june 1988.

[6] G.GIANNAKIS, Y.INOUYE and J.MENDEL, "Cumulant-Based identification of multichannel moving-average models", to appear in *IEEE Trans on Auto. Control*,1989.

[7] G.GIANNAKIS, "Cumulants: A Powerful Tool in signal processing", *Proceedings of the IEEE*, vol.75, n°9, sept 1987, 1333-1334.

[8] C.JUTTEN and J.HERAULT, "Une solution Neuromimétique au problème de séparation de sources", *Traitement du Signal*, vol.5, n°6, 1989, 389-404.

[9] C.JUTTEN and J.HERAULT, "Independent components versus principal components analysis", *Conference EUSIPCO*, Grenoble, France, september 1988.

[10] D.E.KNUTH, *The Art of Computer Programming*,vol.2, Addison Wesley, 1981.

[11] J.L.LACOUME and P.RUIZ, "Sources identification: a solution based on cumulants", *IEEE ASSP workshop*, Mineapolis, aug 1988.

PROCEEDINGS
SPIE—The International Society for Optical Engineering

Advanced Algorithms and Architectures for Signal Processing IV

Franklin T. Luk
Chair/Editor

8–10 August 1989
San Diego, California

Sponsored by
SPIE—The International Society for Optical Engineering

Cooperating Organizations
Applied Optics Laboratory/New Mexico State University
Center for Applied Optics Studies/Rose-Hulman Institute of Technology
Center for Applied Optics/University of Alabama in Huntsville
Center for Electro-Optics/University of Dayton
Center for Excellence in Optical Data Processing/Carnegie Mellon University
Jet Propulsion Laboratory/California Institute of Technology
Optical Sciences Center/University of Arizona
Optoelectronic Computing Systems Center/University of Colorado,
Colorado State University

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 206/676-3290 (Pacific Time) • Telex 46-7053

SPIE PROCEEDINGS SERIES

Volume 1152