Blind Source Separation with Time Dependent Mixtures

Nestor Parga
Departamento de Física Teórica
Universidad Autónoma de Madrid
Cantoblanco, 28049 Madrid, Spain
nestor@ftnp.ft.uam.es

Jean-Pierre Nadal
[to whom correspondence should be addressed]
*Laboratoire de Physique Statistique de l'ENS**
Ecole Normale Supérieure
24, rue Lhomond, 75231 Paris Cedex 05, France
nadal@lps.ens.fr

Abstract

We address the problem of blind source separation in the case of a time dependent mixture matrix. For a slowly and smoothly varying mixture matrix, we propose a systematic expansion which leads to a practical algebraic solution when stationary and ergodic properties hold for the sources.

Resumé

Nous considérons le problème de la séparation aveugle de sources dans le cas d'une matrice de mélange variant lentement et continuement avec le temps. Nous proposons un développement systématique conduisant à une solution algébrique dans le cas où les sources satisfont à certaines conditions de stationarité et d'ergodicité.

To appear in *Signal Processing* [

*Laboratoire associé au C.N.R.S. (U.M.R. 8550), à l'ENS, et aux Universités Paris VI et Paris VII.

Title = Blind Source Separation with Time Dependent Mixtures

Running title = BSS with Time Dependent Mixtures

Keywords

Blind Source Separation

Symbols

N number of inputs and sources
$\mathbf{S}(t) = \{S_j(t), j = 1, ..., N\}$ input data
$\boldsymbol{\sigma}(t) = \{\sigma_a(t), a = 1, ..., N\}$ unknown sources
$\mathbf{M} = \{M_{j,a}, a = 1, ..., N\ j = 1, ..., N\}$ mixture matrix
$\mathbf{J}$ $N \times N$ matrix defining a linear filter applied to the data
$\delta_{a,b}$ Kronecker symbol
$< . >$ empirical expectation: average on a given time window.
$< . >_c$ second order cumulant (e.g. $< S_1 S_2 >_c = < S_1 S_2 > - < S_1 >< S_2 >$ with $< . >$ defined above)
T length of time window on which averaged are taken
τ time delay used for measuring time correlations
Δ typical time scale of the process generating the sources.
$\mathbf{K}^0 = \{K^0_{a,b}, a = 1, ..., N\ b = 1, ..., N\}$ second order cross-cumulant matrix of the sources
$\mathbf{K}(\tau) = \{K(\tau)_{a,b}, a = 1, ..., N\ b = 1, ..., N\}$ second order cross-cumulant matrix of the sources at a time delay τ
$\boldsymbol{C_0}$ second order cross-cumulant matrix of the input data
$\mathbf{C}(\tau)$ second order cross-cumulant matrix at a time delay τ of the input data
$\mathbf{C}^+(\tau)$ symmetric part of the matrix $\mathbf{C}(\tau)$
$\boldsymbol{\Omega}, \mathcal{O}$ $N \times N$ orthogonal matrices
$\boldsymbol{\Lambda}$ diagonal matrix (eigenvalues of $\boldsymbol{C_0}$)
C_A second order cumulant of an arbitrary quantity A.

Number of pages: 12

Figures and tables: none.

Introduction

The problem of source separation arises in many different field, both in signal processing (speech, radar,...) and in neural computation ("coktail-party" effect, separation of odors,...). In all these cases one has to separate different independent "sources" (voices, odors,...) that appear linearly superposed (mixed) when gathered by a set of sensors. Although the data have a linear structure, the difficulty of the task is that the "mixture matrix", that is the set of coefficients in the linear superposition, is unknown - hence the name of "blind source separation" (**BSS**). Since the early proposals of Herault and Jutten [10] and of Bar-ness [2], a lot of effort have been devoted to the search of efficient algorithms for performing BSS (see, e.g., [7, 5, 6, 8, 11, 12, 3, 13, 1]). In the standard case, the mixture matrix is a constant (it does not change with time). Then the linear structure of the data allows to perform BSS by applying some constant linear filter to the output of the array of sensors, which is computed from some analysis of the statistics of these outputs.

In the present paper we address the issue of BSS for time-dependent mixture matrices. This more complicated situation may arise if, e.g., the sources are moving with respect to the data collecting system [10]. Clearly, with no other prior knowledge, one cannot expect in that case to separate the sources at each instant of time. However, through some adaptive procedure, one may hope to obtain good performances *in average*, or a reasonable *prediction* of the mixture matrix based on previous observations. In this paper we propose such an approach to blind source separation with a time dependent mixture matrix for the particular situation where the following properties are expected to hold: (i) the time dependency of the mixture matrix is smooth and slow as compared to the typical time scale of the sources; (ii) the source dynamics satisfy some stationary and ergodic properties. More precisely we will propose a systematic expansion leading to an algebraic solution based on the measure of a limited number of correlations between a well chosen set of combinations of input data.

1 Blind Source Separation: A Reminder

1.1 Linear mixtures of independent sources

The standard paradigm of **BSS** is the following. The input data are assumed to be a *linear mixture of independent sources.* More precisely, at each time t the observed data $\mathbf{S}(t)$ is an N dimensional vector given by

$$S_j(t) = \sum_{a=1}^{N} M_{j,a}\sigma_a(t), \;\; j = 1, ..., N \tag{1}$$

(in vector form $\mathbf{S} = \mathbf{M}\boldsymbol{\sigma}$) where the σ_a are N statistically independent variables, of unknown probability distributions, and $\mathbf{M}$ is an unknown $N \times N$ matrix, called the *mixture matrix.* In the simplest case, $\mathbf{M}$ is time independent.

By hypothesis, all the source cumulants are diagonal, in particular the second order cumulant at equal time $\mathbf{K}^0$ is of the form:

$$K^0_{a,b} \equiv < \sigma_a(t)\sigma_b(t) >_c = \delta_{a,b} K^0_{a,a} \tag{2}$$

where $\delta_{a,b}$ is the Kronecker symbol. Without loss of generality, one can always assume that the sources have zero average:

$$< \sigma_a >= 0, \quad a = 1, ..., N \tag{3}$$

(otherwise one has to estimate the average of each input, and substract it from that input).

Performing **BSS** means finding the linear filter, that is a $N \times N$ matrix $\mathbf{J}$, such that the N-dimensional filter ouput $\mathbf{h}$

$$h_i(t) = \sum_{j=1}^{N} J_{i,j} S_j(t), \; i = 1, ..., N \tag{4}$$

gives a reconstruction of the sources: ideally, one would like to have $\mathbf{J} = \mathbf{M}^{-1}$. However, as it is well known and clear from the above equations, one can recover the sources only up to an arbitrary permutation, and up to a multiplicative factor of arbitrary sign for each source. In particular, this means that the cumulant $\mathbf{K}^0$ is arbitrary: one can always assume the sources to have unit variance, $K^0_{a,a} = 1, a = 1, ..., N$.

As it is usually done in the study of source separation, one assumes that the number of sources is known (there are N observations, e.g. N captors, for N independent sources), and one assumes $\mathbf{M}$ to be invertible. The difficulty comes from the fact that the statistics of the sources are not known, the mixture matrix is not known and is not necessarily (and in general it is not) an orthogonal matrix.

1.2 BSS From Time Correlations

A lot of work has been done in order to define efficient BSS algorithms (see, e.g., [10, 7, 5, 6, 8, 11, 12, 3, 13, 1]. We will not at all make a review of known algorithms. For our purpose, it will however be convenient to consider one particular technique, namely the algebraic approach based on time correlations [9, 14, 4, 11, 13].

We thus assume that the 2nd order cross cumulant matrix $\mathbf{K}(\tau)$ for some time delay $\tau > 0$,

$$K(\tau)_{a,b} \equiv < \sigma_a(t) \; \sigma_b(t-\tau) >_c \tag{5}$$

has non zero diagonal elements:

$$K(\tau)_{a,b} = \delta_{a,b} \; K_{a,a}(\tau) \tag{6}$$

Then $\mathbf{J}$ diagonalizes $\boldsymbol{C_0}$ and $\mathbf{C}(\tau)$ simultaneously, where $\boldsymbol{C_0}$ is the 2nd order cumulant at equal times,

$$\boldsymbol{C_0} \equiv < \mathbf{S}\mathbf{S}^T >_c = \mathbf{M}\mathbf{K}^0\mathbf{M}^T \tag{7}$$

and $\mathbf{C}(\tau)$ is the 2nd order cumulant of the inputs at time delay τ:

$$\mathbf{C}(\tau) \equiv < \mathbf{S}(t) \; \mathbf{S}^T(t-\tau) >_c = \; \mathbf{M}\,\mathbf{K}(\tau)\,\mathbf{M}^T. \tag{8}$$

Finding $\mathbf{J}$ is then an easy to solve algebraic problem [9, 14, 4, 11, 13]. One possible way for computing $\mathbf{J}$ is to first perform the principal component analysis of the data (diagonalization of $\boldsymbol{C_0}$), which determines $\mathbf{J}$ up to an orthogonal matrix [6]. This orthogonal matrix is then obtained as the one which diagonalizes the matrix $\mathbf{C}(\tau)$ projected onto the principal components (see e.g. [13] for details).

2 Time dependent mixtures

2.1 Formulation of the Problem

Let us consider now the case of a time-dependent mixture matrix, $\mathbf{M} = \mathbf{M}(t)$, which is a smoothly and slowly varying as compared to the typical time scale of the sources. More precisely, we assume that there exists some time scale T such that, on any time window of size T, one has a sufficient statistics of the sources (that is their cross-cumulants estimated by averages over this time window are null), and the mixture matrix is almost constant: the norm of the matrix $d\mathbf{M}/dt$ is small compared to $1/T$. Here and in the following, *average* (at time t) of a quantity A will thus mean an average over the time window $[t-T,\ t]$, and we will compute it as

$$< A(t) >_T \equiv \int_0^T \frac{dt'}{T} A(t-t'). \tag{9}$$

Hence, we have for example

$$\begin{aligned} \boldsymbol{C_0} &= < \mathbf{S}(t)\ \mathbf{S}^T(t) >_c = < \mathbf{M}(t)\boldsymbol{\sigma}(t)\boldsymbol{\sigma}^T(t)\mathbf{M}^T(t) >_c \\ < \mathbf{S}(t)\ \mathbf{S}^T(t) > &= \int_0^T \frac{dt'}{T} \mathbf{M}(t-t')\boldsymbol{\sigma}(t-t')\boldsymbol{\sigma}^T(t-t')\mathbf{M}^T(t-t'), \end{aligned} \tag{10}$$

and

$$\begin{aligned} \mathbf{C}(\tau) &= < \mathbf{S}(t)\ \mathbf{S}^T(t-\tau) >_c = < \mathbf{M}(t)\boldsymbol{\sigma}(t)\boldsymbol{\sigma}^T(t-\tau)\mathbf{M}^T(t-\tau) >_c \\ < \mathbf{S}(t)\ \mathbf{S}^T(t-\tau) > &= \int_0^T \frac{dt'}{T} \mathbf{M}(t-t')\boldsymbol{\sigma}(t-t')\boldsymbol{\sigma}^T(t-\tau-t')\mathbf{M}^T(t-\tau-t') \end{aligned} \tag{11}$$

2.2 Optimal constant filter

If the time dependency of $\mathbf{M}$ is very weak, one can try an adiabatic approximation: *for each time window* $[t-T,\ t]$ one can compute a time independent filter matrix $\mathbf{J}_t$. That is, on that particular time window, one analyses the data *as if* they were generated by some linear mixture with a time independent mixture matrix. The subscript t added to $\mathbf{J}$ is a reminder that $\mathbf{J}$, computed in this way, is associated to this particular time interval $[t-T,\ t]$ (since the true mixture matrix is evolving with time, the constant matrix $\mathbf{J}_{t'}$ that will be computed from data of a different time interval $[t'-T,\ t']$ will be different). One may ask $\mathbf{J}_t$ to perform source separation *in average* over that time window $[t-T,\ t]$ (e.g. one can ask for the outputs of the filter to have cross-cumulants as small as possible when these cumulants are computed from time averages over this time window).

One possibility would be to compute the filter matrix as the one which minimizes some convenient criterium measuring the quality of source separation. This is not what we will do, since here and in the following we want to make an explicite use of the hypothesis that the mixture matrix is slowly varying. A convenient choice is then to compute $\mathbf{J}_t$ as the common set of left eigenvectors of the two cumulants $\boldsymbol{C_0}$ and $\mathbf{C}(\tau)$, exactly as one would do it if $\mathbf{M}$ was time independent (see section (1.2)). But, in order to use the same technique as for a constant $\mathbf{M}$, one has to deal with two symmetric matrices. However $\mathbf{C}(\tau)$ may not be a symmetric matrix due to the time-dependency of $\mathbf{M}$ (and this asymmetry is in fact a signature of the non constancy of the mixture matrix). One can rather compute $\mathbf{J}_t$ from the diagonalization of $\boldsymbol{C_0}$ and $\mathbf{C}^+(\tau)$, where $\mathbf{C}^+(\tau)$ is the symmetric part of $\mathbf{C}(\tau)$:

$$\mathbf{C}^+(\tau) \;=\; \frac{1}{2}\,[\mathbf{C}(\tau) \;+\; \mathbf{C}(\tau)^T]. \tag{12}$$

The strategy for computing $\mathbf{J}_t$ is then as follows (as shortly explained in section (1.2) and detailed in [13]). One first perform the principal component analysis, which means computing the orthogonal matrix $\mathbf{\Omega}$ such that $\boldsymbol{C_0} = \mathbf{\Omega}^T\mathbf{\Lambda}\mathbf{\Omega}$ where $\mathbf{\Lambda}$ is the diagonal matrix whose diagonal elements are the eigenvalues of $\boldsymbol{C_0}$. Then $\mathbf{J}_t$ is searched for as $\mathbf{J}_t = \mathcal{O}\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Omega}$, where $\mathcal{O}$ is another orthogonal matrix. This matrix is chosen as the one which diagonalizes the matrix $\mathbf{C}^+(\tau)$ after projection onto the principal components, that is the matrix $\mathbf{\Lambda}^{-\frac{1}{2}}\mathbf{\Omega}\mathbf{C}^+(\tau)\mathbf{\Omega}^T\mathbf{\Lambda}^{-\frac{1}{2}}$. In such a way $\mathbf{J}_t\boldsymbol{C_0}\mathbf{J}_t^T$ and $\mathbf{J}_t\mathbf{C}^+(\tau)\mathbf{J}_t^T$ are diagonal matrices, which is the desired result.

In addition we will see that this way of computing a constant filter matrix is precisely what we need for the expansion we propose in the next section.

2.3 Towards a systematic expansion

In order to get a better estimate of the mixture matrix (or of its inverse) for a given time window $[t-T,\ t]$, one may try to estimate $\mathbf{M}(t-t')$ for t' between 0 and T in a linear expansion in t'. Instead of computing one matrix $\mathbf{J}_t$, one will then compute two matrices. More generally, performing an expansion up to some given order k in t'/T, one will have to compute $k+1$ matrices. In term of the assumptions presented in section (2.1), as we will show in the next section such an expansion will lead to performing BSS on $[t-T,\ t]$ up to a given order in $Td\mathbf{M}/dt$.

The above strategy requires to store the data during the time window $[t-T,\ t]$ in order to compute the optimal filter matrix for that time window $[t-T,\ t]$: this implies an *off line* processing. Let us however comment briefly on the use of such an approach for *online processing*. One may use as a prediction for $\mathbf{J}_{t+T}$ the result of the computation done on the previous window $[t-T,\ t]$. Then it is clear that, for instance, the optimal constant filter matrix $\mathbf{J}_{t+T}$ will be obtained from the linear expansion performed on $[t-T,\ t]$. More generally, one may compute $\mathbf{J}$ for the time window $[t,\ t+T]$ at order k in t'/T from an expansion at order $k+1$ on $[t-T,\ t]$.

We now come back to the (off line) processing of data for one given window, dealing with the linear expansion.

3 First Order Expansion

3.1 Linear approximation within a time window

The hypothesis on the slow evolution of the mixture matrix means that, at first order, $\mathbf{M}(t-t')$ for t' within 0 and T can be written as

$$\mathbf{M}(t-t') = \mathbf{M}_t^0 \; - \; \frac{t'}{T}\mathbf{M}_t^1 \tag{13}$$

where $\mathbf{M}_t^0$ and $\mathbf{M}_t^1 = T\ d\mathbf{M}/dt$ are two unknown matrices to be determined from statistics on the time window $[t-T,\ t]$. What we want to do is to see whether one can measure a limited number of correlations of the input data in order to determine $\mathbf{M}_t^0$ and $\mathbf{M}_t^1$. Let first consider the 2nd order cumulant matrix at some time delay τ, $\mathbf{C}(\tau)$ as defined in (11) (note that $\mathbf{C}(\tau=0) \equiv \boldsymbol{C_0}$). Replacing $\mathbf{M}$ given by (13), taking τ smaller than T, with the averages defined as in (9), we have

$$\begin{aligned} < \mathbf{S}(t)\ \mathbf{S}^T(t-\tau) > \; &= \; \int_0^T \frac{dt'}{T} \{ \ \mathbf{M}_t^0 \ \boldsymbol{\sigma}(t-t')\boldsymbol{\sigma}^T(t-\tau-t')\ \mathbf{M}_t^{0T} \\ &- \; \frac{t'}{T}\mathbf{M}_t^1 \ \boldsymbol{\sigma}(t-t')\boldsymbol{\sigma}^T(t-\tau-t')\ \mathbf{M}_t^{0T} \\ &- \; \frac{t'}{T}\mathbf{M}_t^0 \ \boldsymbol{\sigma}(t-t')\boldsymbol{\sigma}^T(t-\tau-t')\ \mathbf{M}_t^{1T} \\ &- \; \frac{\tau}{T}\mathbf{M}_t^0 \ \boldsymbol{\sigma}(t-t')\boldsymbol{\sigma}^T(t-\tau-t')\ \mathbf{M}_t^{1T} \ \} \end{aligned} \tag{14}$$

and one gets

$$\begin{aligned} \mathbf{C}(\tau) \; &= \; \mathbf{M}_t^0\mathbf{K}^{(0)}(\tau)\mathbf{M}_t^{0T} \; - \; \mathbf{M}_t^1\mathbf{K}^{(1)}(\tau)\mathbf{M}_t^{0T} \\ & \quad - \mathbf{M}_t^0\mathbf{K}^{(1)}(\tau)\mathbf{M}_t^{1T} \; - \; \frac{\tau}{T}\mathbf{M}_t^0\mathbf{K}^{(0)}(\tau)\mathbf{M}_t^{1T} \end{aligned} \tag{15}$$

where the $\mathbf{K}^{(k)}(\tau)$ are the generalized source cumulants:

$$\begin{aligned} \mathbf{K}^{(k)}(\tau) \equiv \int_0^T \frac{dt'}{T} \left(\frac{t'}{T}\right)^k \ \boldsymbol{\sigma}(t-t')\boldsymbol{\sigma}^T(t-\tau-t') \\ - \left[\int_0^T \frac{dt'}{T}\ \boldsymbol{\sigma}(t-t') \right] \left[\int_0^T \frac{dt'}{T} \left(\frac{t'}{T}\right)^k \ \boldsymbol{\sigma}^T(t-\tau-t') \right]. \end{aligned} \tag{16}$$

Note that $\mathbf{K}^{(0)}(\tau) \equiv \mathbf{K}(\tau)$. One can see from the last term in the r.h.s of (15) that $\mathbf{C}(\tau)$ is indeed not symmetric for non zero τ. From the above equations, it appears that measuring $\boldsymbol{C_0}$ and $\mathbf{C}(\tau)$ will not be sufficient in order to estimate $\mathbf{M}_t^0$ and $\mathbf{M}_t^1$, since we have $\mathbf{K}^{(1)}(\tau)$ as an additional unknown. However, measuring, say, $< \frac{t'}{T}\ \mathbf{S}(t-t')\ \mathbf{S}^T(t-\tau-t') >_c$, will not help, since this average will depend on a new source cumulant, namely $\mathbf{K}^{(2)}(\tau)$.

3.2 A Tractable Case

We now make use of the assumption that the sources are generated according to some *stationary and ergodic* process (at least up to the second order statistics), together with the hypothesis that the time window T is large compared to the typical time scale of this process.

Let us consider some quantity $A(t)$ of interest, such as the vector $\boldsymbol{\sigma}(t)$ or the matrix $\boldsymbol{\sigma}(t)\boldsymbol{\sigma}^T(t-\tau)$, and the integral

$$Z(t) \;=\; \int_0^T \frac{dt'}{T}\, f(\frac{t'}{T})\; A(t-t') \tag{17}$$

where $f(u)$ is any positive function defined on $[0,1]$, such as $f(u) = u^k$ for k integer. A being a random variable, we can consider its average value and the fluctuations around it. From ergodicity we have:

$$< Z > = < A > \int_0^1 du\; f(u) \tag{18}$$

where the average $< A >$ does not depend on the time t (stationarity).

Now let us consider the fluctuations around the mean as characterized by the second cumulant:

$$< Z^2 >_c \equiv < Z^2 > - < Z >^2 = \int_0^T dt_1\; f(t_1) \int_0^T dt_2\; f(t_2)\;\; < A(t-t_1)A(t-t_2) >_c \tag{19}$$

From the stationarity hypotheses the 2nd order cumulant matrix $C_A \equiv < A(t-t_1)A(t-t_2) >_c$ is a function of the time difference alone:

$$< A(t-t_1)A(t-t_2) >_c = \; C_A[v \equiv t_1 - t_2] \tag{20}$$

One then has

$$< Z^2 >_c = \int_0^1 du\; f(u) \int_0^1 du'\; f(u')\; C_A[T(u-u')] \tag{21}$$

In order to characterize the typical time scale of the process, we make the more explicit hypotheses that for some a and Δ, positive and finite numbers, the norm of the correlation matrix is bounded according to

$$\text{for any } v, \quad |\; C_A[v] \;| \;<\; a\; \Psi[\frac{|v|}{\Delta}] \tag{22}$$

where $\Psi(x) = \Psi(-x)$ is some function which goes quickly to zero as $|x|$ goes to infinity, e.g.,

$$\Psi(x) \;=\; \exp -|x|^{2r} \tag{23}$$

for some $r > 0$. We may also assume that Δ is taken as the smallest value for which (22) holds. Hence we have

$$< Z^2 >_c < \; a \int_0^1 du\; f(u) \int_0^1 du'\; f(u')\Psi(\frac{T}{\Delta}|u-u'|) \tag{24}$$

Our hypothesis for T means that T is large compared to Δ;

$$T \quad >> \quad \Delta \tag{25}$$

At lowest order, that is in the limit $\frac{T}{\Delta} \to \infty$, $< Z^2 >_c$ vanishes, so that Z is a non fluctuating quantity: it is almost surely equal to its mean value $< Z >$. In replacing Z by $< Z >$, one is in fact neglecting terms (at worst) of order $\sqrt{\frac{\Delta}{T}}$. More precisely, for $\frac{T}{\Delta}$ large one can write

$$\Psi(\frac{T}{\Delta}|v|) \;\; = \;\; \frac{\Delta}{T}\,\alpha\,\delta(v) + \left[\frac{\Delta}{T}\right]^3 \;\beta\,\delta''(v) \tag{26}$$

with δ the Dirac distribution and

$$\alpha = \int_{-\infty}^{\infty} dx\;\Psi(x), \quad \beta = \int_{-\infty}^{\infty} dx\;\Psi(x)x^2 \tag{27}$$

As a result,

$$| < Z^2 >_c | \sim \frac{\Delta}{T}\,a\,\alpha \int_0^1 du\; f(u)^2. \tag{28}$$

In the above derivation we have been working with A as a function fo t alone. In the case of a quantity such as $A = \boldsymbol{\sigma}(t)\boldsymbol{\sigma}^T(t-\tau)$, A is in fact a function of t and $t-\tau$. One can make exactly the same analysis as above, in which one will get expressions with, e.g., $\Psi(\frac{T}{\Delta}|u - u' - \frac{\tau}{T}|)$. Hence the derivation will apply as well if the limit of large $\frac{T}{\Delta}$ is taken at a given value of $\frac{\tau}{T}$. This is equivalent to state that the scaling regime of interest is

$$\Delta \;<<\; \tau \;<<\; T \tag{29}$$

so that one can neglect terms of order $\frac{\Delta}{T}$ even when taking into account terms of order $\frac{\tau}{T}$.

3.3 Solution at first order

According to the above discussion one can replace, under the integrals defining the cumulants $\mathbf{K}^{(k)}(\tau)$, the source terms by their average, that is:

$$\mathbf{K}^{(k)}(\tau) \;\; = \;\; \int_0^T \frac{dt'}{T}\left(\frac{t'}{T}\right)^k \mathbf{K}^{(0)}(\tau) \;\; = \;\; \frac{1}{k+1}\,\mathbf{K}^{(0)}(\tau) \tag{30}$$

As a result, correlations at a time delay τ involve only $\mathbf{K}^{(0)}(\tau)$. Moreover, one sees that the only products of matrices that appear for a given τ are $\mathbf{M}^0_t\mathbf{K}^{(0)}(\tau)\mathbf{M}^{0T}_t$, $\mathbf{M}^0_t\mathbf{K}^{(0)}(\tau)\mathbf{M}^{1T}_t$ and $\mathbf{M}^1_t\mathbf{K}^{(0)}(\tau)\mathbf{M}^{0T}_t$. Since $\mathbf{K}^0 \equiv \mathbf{K}^{(0)}(0)$ is arbitrary as explained in section (1.1), we need to measure only three combinations of correlations. Denoting by $\mathbf{C}^{(1)}(\tau)$ the cumulants

$$\begin{aligned}\mathbf{C}^{(1)}(\tau) \;\; &= \;\; \int_0^T \frac{dt'}{T}\frac{t'}{T}\,\mathbf{S}(t-t')\,\mathbf{S}^T(t-\tau-t') \\ &- \left[\int_0^T \frac{dt'}{T}\,\mathbf{S}(t-t')\right]\left[\int_0^T \frac{dt'}{T}\frac{t'}{T}\,\mathbf{S}^T(t-\tau-t')\right],\end{aligned} \tag{31}$$

and adding a subscript + (resp. −) to denote the symmetric (resp. antisymmetric) part of a matrix, one obtains easily the following relations:

$$4\,\boldsymbol{C_0} \;-\; 6\,\boldsymbol{C_0}^{(1)} \;=\; \mathbf{M}_t^0\mathbf{K}^0\mathbf{M}_t^{0T} \tag{32}$$

$$(4+\frac{3\tau}{T})\,\mathbf{C}^+(\tau) \;-\; 6(1+\frac{\tau}{T})\,\mathbf{C}^{(1)+}(\tau) \;=\; \mathbf{M}_t^0\mathbf{K}^{(0)}(\tau)\mathbf{M}_t^{0T} \tag{33}$$

and

$$\mathbf{C}^-(\tau) \;+\; \frac{3\tau}{T}\{\mathbf{C}^+(\tau) \;-\; 2\mathbf{C}^{(1)+}(\tau)\} \;=\; \frac{\tau}{T}\,\mathbf{M}_t^1\mathbf{K}^{(0)}(\tau)\mathbf{M}_t^{0T} \tag{34}$$

The matrices $\mathbf{M}_t^0$ and $\mathbf{K}^{(0)}(\tau)$ are obtained from the first two equations (32) and (33), where the data appear on the r.h.s. in the form of two symmetric matrices: to get $\mathbf{M}_t^0$ and $\mathbf{K}^{(0)}(\tau)$ one can then use exactly the same techniques as those used in order to obtain the matrices $\mathbf{M}$ and $\mathbf{K}(\tau)$ for a time independent mixture, see section (1.2). Eventually, $\mathbf{M}_t^1$ is easily computed from equation (34).

Two final remarks.

- One should note that the l.h.s. of equations (32,33,34) are cross-correlations of particular combinations of the input data: one can then measure these correlations directly, instead of computing separately each matrix appearing in these equations.

- as a definition of the average $< . >$, one may use instead of (9), any similar definition such as

$$< A(t') >_t \equiv \int_0^\infty dt'\ \Psi(t')\ A(t-t'). \tag{35}$$

  where $\Psi(t')$ is sufficiently zero for t' larger than T. The above derivation can be easily adapted, provided Ψ is such that the numerical factors $\int_0^\infty\ dt'\ \Psi(t')\ \left(\frac{t'}{T}\right)^k$ do depend on k - if not, all the correlations $\mathbf{C}(\tau)$, $\mathbf{C}^{(1)}(\tau)$ give the same equation, and thus one cannot combine them to get formulae similar to (32 - 34). Note that this excludes in particular the choice $< A(t') >_t \equiv \int_0^\infty\ \frac{dt'}{T}\ \exp(-t'/T)\ A(t-t')$.

Conclusion

In this paper we considered the problem of BSS with a time varying mixture. We showed that, under some assumptions on the statistics of the sources, when the mixture matrix is slowly varying it is possible to obtain a solution making use of techniques derived for time independent mixtures. We proposed a systematic expansion, and we presented in detail the evaluation of the first order expansion. The validity of the expansion can be checked *a posteriori*: one has to check that the first order term $\mathbf{M}_t^1$, which is in fact $T\ d\mathbf{M}/dt$, is indeed small, and whether the outputs do appear independent on any time window of size T. As we explained it, the expansion at a given order k, performed on a given time window $[t-T,\ t]$, can be used either for *off line* processing - that is in order to process the data of that time window -, or for *on line* processing, predicting the mixture matrix at order $k-1$ for the next time window.

It is cumbersome, but not difficult, to derive the equations at any order - one has just to follow the same strategy, expanding the mixture matrix up to the required order within a time window, as explained in section (2.3).

For simplicity, we have worked with second order cumulants at equal time and at some delay τ ($\tau < T$). It would be interesting to adapt our method to other approaches to BSS. For example, one may use criteria based on higher cumulants at equal time which have been proposed for performing BSS in the case of a constant mixture matrix [7, 6, 13]. With averages defined as time-window averages as above (section (2.1)), one has to perform the necessary expansions of the cumulants under consideration. Again, the zeroth order will be given by the solution of the system as if $\mathbf{M}$ was constant.

Numerical simulations remain to be done in order to test the efficiency of the proposed expansion.

Acknowledgements

This work has been partly supported by the French-Spanish program "Picasso", the E.U. grant CHRX-CT92-0063, the Universidad Autónoma de Madrid, the Université Paris VI and Ecole Normale Supérieure. NP and JPN thank, respectively, the Laboratoire de Physique Statistique (ENS) and the Departamento de Física Teórica (UAM) for their hospitality.

References

[1] S. I. Amari , A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation", in D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, ed., *Advances in Neural Information Processing Systems 8*, MIT Press, 1996.

[2] Y. Bar-Ness, "Bootstrapping adaptive interference cancelers: some practical limitations", *Proc. of The Globecom Conf.*, Miami November 1982, paper F3.7 pp. 1251–1255.

[3] A. Bell and T. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution", *Neural Comp.*, Vol. 7, 1995, pp. 1129–1159.

[4] A. Belouchrani A. and K. Abed Meraim, "Séparation aveugle au second ordre de sources temporellement corrélées", *Proc. GRETSI'93*, Juan-Les-Pins, September 1993, vol. 1, pp. 309-312.

[5] G. Burel, "Blind separation of sources: a nonlinear neural algorithm", *Neural Networks*, Vol. 5, 1992, pp. 937–947.

[6] J.-F. Cardoso, "Source separation using higher-order moments", *Proc. Internat. Conf. Acoust. Speech Signal Process.-89*, Glasgow, 1989, pp. 2109–2112.

[7] P. Comon, "Independent component analysis, a new concept ?" *Signal Processing*, Vol. 36, 1994, pp. 287–314.

[8] N. Delfosse and Ph. Loubaton, "Adaptive blind separation of independent sources: a deflation approach", *Signal Processing*, Vol. 45, 1995, pp. 59–83.

[9] L. Féty, "Méthodes de traitement d'antenne adaptée aux radio-communications", In *PhD Thesis*, ENST Paris, 1988.

[10] C. Jutten and J. Herault, "Blind separation of sources, part i: An adaptive algorithm based on neuromimetic architecture", *Signal Processing*, Vol. 24, 1991, pp. 1–10.

[11] L. Molgedey and H.G. Schuster, "Separation of a mixture of independent signals using time delayed correlations", *Phys. Rev. Lett.*, Vol. 72, 1994, pp. 3634–3637.

[12] J.-P. Nadal and N. Parga, "Nonlinear neurons in the low noise limit: a factorial code maximizes information transfer", *Network*, Vol. 4, 1994, pp. 565-582.

[13] J.-P. Nadal and N. Parga, "Redundancy reduction and independent component analysis: Conditions on cumulants and adaptive approaches", *Neural Computation*, Vol. 9(7), 1997, pp. 1421-1456.

[14] L. Tong , V. Soo, R. Liu, and T. Huang, "Amuse: a new blind identification algorithm", *Proc. ISCAS*, New Orleans, USA, 1990.