Authors' final version of a paper published in "Signal Processing"

Paper reference:

Y. Deville, M. Benali, F. Abrard, "Differential source separation for underdetermined instantaneous or convolutive mixtures: concept and algorithms", Signal Processing, Vol. 84, Issue 10, pp. 1759-1776, Oct. 2004, http://authors.elsevier.com/sd/article/S0165168404001215

Differential source separation for underdetermined instantaneous or convolutive mixtures: concept and algorithms

Yannick Deville[1], Mohammed Benali, Frédéric Abrard

Université Paul Sabatier
Laboratoire d'Acoustique, Métrologie, Instrumentation
Bât. 3R1B2
118, Route de Narbonne
31062 Toulouse Cedex
France
E-mail: ydeville@cict.fr

Abstract:
This paper concerns the underdetermined or noisy case of the blind source separation (BSS) problem, i.e. the situation when the number of observed mixed signals is lower than the number of sources, which is of high practical interest. We first propose a general differential BSS concept to handle this case. This approach applies to linear instantaneous and convolutive mixtures. It uses optimization criteria based on differential parameters so as to achieve "partial BSS", i.e. so as to make some sources invisible in these criteria and to perform the exact separation of the other sources only. In other words, each output signal is thus reduced to a mixture of: i) only one visible source and ii) all invisible sources. Various BSS methods may be derived from this concept. We illustrate it by applying this concept to a specific criterion and associated algorithms, which exploit the assumed non-stationarity of some sources. The resulting approach applies to convolutive mixtures and uses the second-order statistics of the signals. It adapts the filters of a direct BSS system so as to cancel the "differential cross-correlation function" (introduced in this paper) of signals derived by this system. We analyze the stability of this approach, by using the Ordinary Differential Equation method, and we show its performance by means of numerical tests.

Résumé:
Cet article concerne le cas sous-déterminé ou bruité du problème de séparation aveugle de sources (SAS), c-à-d la situation où le nombre de signaux mélangés observés est inférieur au nombre de sources, qui est d'un grand intérêt pratique. Nous proposons d'abord un concept général de SAS différentielle pour traiter ce cas. Cette approche s'applique aux mélanges linéaires instantanés et convolutifs. Elle utilise des critères d'optimisation fondés sur des paramètres différentiels afin de réaliser une "SAS partielle", c-à-d afin de rendre certaines sources invisibles dans ces critères et d'obtenir seulement la séparation exacte des autres sources. En d'autres termes, chaque signal de sortie est ainsi réduit à un mélange de : i) une seule source visible et ii) toutes les sources invisibles. Diverses méthodes de SAS peuvent être déduites de ce concept. Nous l'illustrons en appliquant ce concept à un critère spécifique et à des algorithmes associés, qui exploitent la non-stationnarité supposée de certaines sources. L'approche résultante s'applique aux mélanges convolutifs et utilise les statistiques d'ordre 2 des signaux. Elle adapte les filtres d'un système direct de SAS de manière à annuler la "fonction d'inter-corrélation

[1] Corresponding author.

différentielle" (introduite dans cet article) de signaux déterminés par ce système. Nous analysons la stabilité de cette approche, en utilisant la méthode des Equations Différentielles Ordinaires, et nous illustrons ses performances à l'aide de tests numériques.

Keywords:
blind signal separation,
differential criterion and algorithm,
differential correlation function,
instantaneous or convolutive mixture,
non-stationary source,
ordinary differential equation,
partial source separation,
stability analysis,
underdetermined or noisy mixture.

1 Problem Statement

Blind source separation (BSS) methods aim at restoring a set of N source signals $x_j(n)$ from a set of P observed signals $y_i(n)$, which are mixtures of these source signals [5],[7],[12],[17]. The mixed signals $y_i(n)$ are often provided by a set of sensors, and the mixing phenomenon then results from the simultaneous propagation of all signals from their emission locations to all sensors. This gives rise to two major classes of BSS problems, depending on the features of the considered propagation. In the so-called linear instantaneous mixture model, each propagation channel from source j to sensor i is represented by a scalar coefficient a_{ij}, which typically reflects attenuation during propagation. The overall relationship between the column vectors $x(n)$ and $y(n)$ of sources and observations is then expressed in the discrete time domain as:

$$y(n) = Ax(n), \tag{1}$$

where the scalar mixing matrix A consists of the coefficients a_{ij}. In the more general convolutive mixture model, each source-to-sensor channel is represented by a transfer function $A_{ij}(z)$, which may also account for propagation delays and multipath propagation. The overall source-to-observation relationship then reads in the $\mathcal{Z}$ domain:

$$Y(z) = A(z)X(z), \tag{2}$$

where the elements of the mixing matrix $A(z)$ are the transfer functions $A_{ij}(z)$.

Most BSS investigations have been performed in the case when: i) the number P of observed signals is equal to the number N of source signals, so that the considered mixing matrix A or $A(z)$ is square, and ii) this matrix is invertible. The BSS problem then consists in estimating the inverse of this mixing matrix (up to some indeterminacies [5],[7],[12]). This may first be shown for linear instantaneous mixtures by considering a vector $s(n)$ of output signals of a BSS system, obtained by multiplying the available mixed signal vector by a scalar separating matrix C, i.e:

$$s(n) = Cy(n). \tag{3}$$

Combining (1) and (3) yields

$$s(n) = CAx(n), \tag{4}$$

which shows that when C is made equal to A^{-1}, all the output signals $s_j(n)$ resp. become exactly equal to all the source signals $x_j(n)$ to be restored. The same analysis may also be performed in the $\mathcal{Z}$ domain for convolutive mixing and separating matrices.

Various methods have been proposed for estimating the inverse of the mixing matrix. They are esp. based on the assumed statistical independence or uncorrelation of the source signals [5],[7],[12],[17]. Many of these methods consist in maximizing, minimizing or cancelling statistical parameters of the output signals of a BSS system, such as their second-order or higher-order moments or cumulants, which are classical parameters in the higher-order statistics field [15],[16].

Most of these investigations were performed under the assumption $P = N$, as stated above. In many practical situations however, only a limited number of sensors is acceptable, due e.g. to cost constraints or physical configuration, whereas these sensors receive a larger number of sources (possibly including "noise sources"). In this paper, we consider this underdetermined situation corresponding to $P < N$, and we require that

$P \geq 2$. A few analyses and BSS methods have been reported for this case (see e.g. [1],[4],[9],[11],[13],[18]), mainly for linear instantaneous mixtures. However, they set major restrictions on the source properties (discrete sources are especially considered) and/or on the mixing conditions. We here introduce a general concept and resulting practical algorithms, which only set some (non-)stationary constraints on the sources, and which combine the following features:

- They apply to continuous sources with unknown distributions.
- They may be used for linear instantaneous and convolutive mixtures.
- They are not restricted to specific mixing conditions resulting in particular values in the mixing matrix.
- They only perform linear (instantaneous or convolutive) combinations of the available mixed signals, thus providing linear combinations of the sources. We set this condition because in various applications, such as speech enhancement, it has been observed that the artefacts created by non-linear processing are even more disturbing than the "noise" present in the initial mixed signals, so that linear processing should be preferred. The price to pay for this linear operation is that the outputs of the BSS system introduced hereafter still contain some "noise". More precisely, among the N mixed sources, only P of these sources are considered as the sources of interest, whereas the other $N-P$ sources are seen as "noise" sources. The proposed approach then separates each of the P sources of interest from the other $P-1$ such sources, i.e. it reduces each output signal to a mixture of: i) only one of the P sources of interest and ii) the $N-P$ "noise" sources.

The remainder of this paper is organized as follows. In Section 2, we introduce the proposed general concept, which makes it possible to derive various BSS methods intended for underdetermined or "noisy" mixtures. As an application of this concept, Section 3 and Appendix A describe a resulting criterion and associated algorithms for basic convolutive mixtures. The experimental performance of these algorithms is reported in Section 4 and conclusions are drawn from this investigation in Section 5.

2 BSS limitation and proposed concept

For the sake of simplicity, we present the proposed concept in the linear instantaneous case in this section. This approach may be adapted to the convolutive case by transposing the following discussion from the time domain to the $\mathcal{Z}$ domain. The application of our general concept to a specific BSS method will be detailed in Section 3 for convolutive mixtures.

We showed above that if $P = N$ an exact separation of all sources is theoretically possible (and is achieved when C is made exactly equal to A^{-1}). It is clearly also possible when $P > N$, but not when $P < N$ as will now be shown. In the latter case, A is not square. Each output signal $s_i(n)$ defined by (3) is a linear instantaneous combination of the P mixed signals. For an arbitrary matrix A, the highest number of source contributions that may be cancelled in such a (non-zero) combination is $P-1$. This optimum case, where each output is still a mixture of the remaining $(N-P+1)$ sources, is called "partial BSS" hereafter.

This only shows that partial BSS is theoretically possible, in the sense that it is achieved for adequate combination coefficients c_{ij} in the matrix C. To actually achieve it in practice, we then need algorithms which are able to estimate these adequate values of the coefficients c_{ij}. It may be shown that the classical methods, that have been developed for the case when $P = N$, do not meet this requirement: whereas their principles (such as the above-mentioned cumulant optimization) coincide with the separation of the source signals when $P = N$, these BSS methods yield outputs signals which are still mixtures of all sources when applied to arbitrary observed signals such that $P < N$.

This paper therefore aims at introducing a concept for achieving the above-defined partial BSS when $P < N$. By "concept", we mean that we do not propose a single criterion (and/or algorithm) but a general way to derive new partial BSS methods from various existing approaches developed for the case when $P = N$. We therefore start from one of the latter methods, based on a given signal parameter, such as those defined in Section 1: for example, we may start from a method based on the optimization of a given set of cumulants of output signals. We then set the following additional constraints: i) two occurences of the considered signal parameter should be available and ii) the considered sources should consist of two types with respect to corresponding source parameters, i.e:

1. first-type sources: for (at most) P sources, the source parameter should take different values in the two occurences,
2. second-type sources: for the other $N - P$ sources (at least), this parameter should take the same value in the two occurences.

The above-mentioned two "occurences" may be obtained in various ways, which allows one to derive various methods from the proposed concept. They may e.g. consist of the two values of a parameter, such as a (zero-lag) cumulant, resp. obtained for two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. These domains are then defined as follows. When considering these theoretical cumulant values themselves, which are combinations of signal expectations [15],[16], each of these time domains $\mathcal{D}_i$ is restricted to the time n_i when these expectations are considered. On the contrary, the corresponding cumulant estimates used in practice are resp. obtained by time averaging over two domains $\mathcal{D}_1$ and $\mathcal{D}_2$ which consist of non-overlapping bounded time intervals, assuming that each source is stationary over each such domain (this is referred to as "short-term stationarity" below). The second type of sources then consists of source signals whose cumulant values do not vary from one of the considered time domains to the other one, i.e. sources which are in addition "long-term stationary". On the contrary, first-type sources then consist of long-term non-stationary source signals.

The proposed concept then consists in deriving new signal parameters from the initial ones in such a way that the effect of the second-type sources disappears in these new parameters. This is achieved as follows. For the sake of simplicity, we consider an initial parameter which is a linear function of corresponding source parameters (for example, an output cumulant is indeed a linear function of cumulants of the independent sources, as illustrated hereafter). The new parameter that we then define exploits the difference of behavior between the two types of sources concerning the variations of their initial parameter from one occurence to the other one. More precisely, we here focus on the case when the new parameter that we define is the difference between the two values of the initial parameter resp. associated with the two available occurences (e.g. the difference between the two output cumulants corresponding to the two time domains). As each second-type source yields the same contribution in the two initial parameter values, its effect is cancelled in the corresponding parameter difference, i.e. the second-type sources are thus indeed made invisible in this differential parameter.

For the sake of clarity, let us describe in some more detail the application of this general principle to a specific linear instantaneous BSS method. Consider observed signals defined by (1) and processed by a BSS system which yields the output signals defined by (3). We select as the initial BSS method one of the classical approaches based on the optimization of specific output signal cumulants, say for example their $(3,1)$ zero-lag cross-cumulants, i.e:

$$CUM^{(3,1)}_{s_i s_j}(n) = CUM(s_i(n), s_i(n), s_i(n), s_j(n)). \tag{5}$$

Starting from this approach, which does not apply to underdetermined mixtures, the general principle that we introduced above for underdetermined mixtures consists in deriving a differential BSS method based on the differential cumulants that we associate to (5), that we define as:

$$DCUM^{(3,1)}_{s_i s_j}(n_1, n_2) = CUM^{(3,1)}_{s_i s_j}(n_2) - CUM^{(3,1)}_{s_i s_j}(n_1). \tag{6}$$

Let us show that, whereas the initial parameter $CUM^{(3,1)}_{s_i s_j}(n)$ depends on all sources, its differential version $DCUM^{(3,1)}_{s_i s_j}(n_1, n_2)$ only depends on the first-type sources. Eq. (4) may be expressed as:

$$s(n) = Bx(n), \tag{7}$$

where the matrix $B = CA$ includes the effects of the mixing and separating stages. Denoting b_{pq} the elements of B, (7) implies that two arbitrary output signals $s_i(n)$ and $s_j(n)$ may be expressed with respect to all sources as:

$$s_i(n) = \sum_{k=1}^{N} b_{ik} x_k(n) \tag{8}$$

$$s_j(n) = \sum_{l=1}^{N} b_{jl} x_l(n). \tag{9}$$

Using cumulant properties and the assumed independence of all sources, the $(3,1)$ cross-cumulant of the above output signals may be shown easily to be equal to:

$$CUM^{(3,1)}_{s_i s_j}(n) = \sum_{k=1}^{N} b_{ik}^3 b_{jk} CUM^{(4)}_{x_k}(n), \tag{10}$$

where the 4^{th}-order (zero-lag) auto-cumulant of source x_k is defined as:

$$CUM^{(4)}_{x_k}(n) = CUM(x_k(n), x_k(n), x_k(n), x_k(n)). \tag{11}$$

The standard output cumulant expressed in (10) therefore actually depends on the 4^{th}-order cumulants of all sources. The differential output cumulant associated to (10), as defined in (6), may then be expressed as:

$$DCUM^{(3,1)}_{s_i s_j}(n_1, n_2) = \sum_{k=1}^{N} b_{ik}^3 b_{jk} DCUM^{(4)}_{x_k}(n_1, n_2), \tag{12}$$

where we define the 4^{th}-order (zero-lag) differential auto-cumulant of source x_k as:

$$DCUM^{(4)}_{x_k}(n_1, n_2) = CUM^{(4)}_{x_k}(n_2) - CUM^{(4)}_{x_k}(n_1). \tag{13}$$

Let us now take into account the fact that we are considering the underdetermined situation, when P observed mixtures of $N > P$ source signals are available and when P sources, say x_1 to x_P, are of the firt stype (i.e. long-term non-stationary), while the other sources, i.e. x_{P+1} to x_N, are of the second type (i.e. long-term stationary). The standard 4^{th}-order auto-cumulant $CUM_{x_k}^{(4)}(n)$ of each of the sources x_{P+1} to x_N then takes the same values for $n = n_1$ and $n = n_2$, so that $DCUM_{x_k}^{(4)}(n_1, n_2) = 0$. The expression (12) then reduces to:

$$DCUM_{s_i s_j}^{(3,1)}(n_1, n_2) = \sum_{k=1}^{P} b_{ik}^3 b_{jk} DCUM_{x_k}^{(4)}(n_1, n_2). \tag{14}$$

This shows explicitly that this differential parameter only depends on the first-type sources.

Let us stress again that the description that we just provided should only be regarded as an example of the proposed general principle. Another example will be analyzed in more detail in the next section. Beyond these examples, the main result obtained at this stage is the above-defined general principle itself, which makes it possible to create new parameters which do not depend on the corresponding parameters of the second-type sources. This is of major interest because the initial BSS configuration is thus transformed in a configuration where (at most) P sources are visible (i.e. the first-type sources) from the point of view of the considered new parameter, and P mixtures of these sources are available. In other words, thanks to this approach we get back in the classical situation involving as many observed signals as source signals (concerning this new parameter). This suggests to propose, as the differential BSS method to be considered, a method whose definition is derived from that of the selected initial method by only replacing the initial parameter by its differential version in the optimization criterion of the initial method. Based on the above description, the resulting differential BSS method may be hoped to yield separating coefficients c_{ij} such that each output signal $s_i(n)$ contains a contribution from only one of the sources seen by this method (i.e. of the first-type sources), but we insist that this should then be checked: we thus defined a procedure for proposing potential differential BSS methods, but the actual behavior of any such method should then be analyzed, because the properties of the differential parameter on which it is based may be slightly different from those of the initial non-differential parameter. For example, a second-order zero-lag auto-cumulant, i.e. a signal variance, is always positive whereas its differential version may be negative. We will analyze the consequence of this modification on a specific method in Section 3. More precisely, in that section we will present in detail the application to a specific BSS method of the overall approach that we have just defined, including both the detailed derivation of the considered differential BSS method and the subsequent analysis of its properties. Before proceeding to this specific case, the following comments should be made about our general approach:

- As explained above, the differential methods succesfully derived from our approach yield output signals $s_i(n)$ which each contain a contribution from only one of the first-type sources. It should be clear that each such signal $s_i(n)$ also contains contributions from all second-type sources, corresponding to the combination of the mixed signals $y_i(n)$ by means of the obtained coefficients c_{ij}, as shown in (3). The proposed methods therefore achieve the above-defined partial BSS, and more specifically the sources that are thus completely separated are the first-type ones. These sources are the signals of interest in this partial BSS problem, as opposed to the "noise signals", i.e. the second-type sources, which still appear in all outputs.

- To summarize, the proposed concept consists of a differential BSS approach, which uses optimization criteria based on differential parameters, so as to make some sources invisible in these criteria and to perform the exact separation of the other sources only. This approach may therefore also be considered as follows. Various classical BSS methods identify the mixing matrix or its inverse by optimizing higher-order signal cumulants. They are insensitive to the presence of additional Gaussian noise in the observed signals, because such noise has null higher-order cumulants and is therefore inherently hidden in the considered optimization parameter. The approach proposed in this paper may be seen as a means to extend this behavior to any non-Gaussian stationary noise, since we modify optimization parameters so as to hide this initially visible type of noise.

3 Application to a second-order convolutive method

3.1 Redefining the classical approach

This section shows explicitly how the above general concept may be applied to an existing BSS method in order to derive its differential version. For the sake of simplicity, we consider a second-order approach, i.e. the classical decorrelation method that has been used by various authors in the basic configuration of the convolutive BSS problem [2],[6],[19],[20]. We therefore first redefine this classical method in a way which is suited to the approach that we will then use to extend it. The considered configuration involves two convolutive mixtures of two uncorrelated sources, defined in the $\mathcal{Z}$ domain as:

$$
\begin{aligned}
Y_1(z) &= X_1(z) + A_{12}(z)X_2(z) && (15)\\
Y_2(z) &= A_{21}(z)X_1(z) + X_2(z), && (16)
\end{aligned}
$$

where $A_{12}(z)$ and $A_{21}(z)$ are strictly causal MA filters and their orders are (at most) equal to M. These mixed signals are provided to a BSS system, which aims at restoring the source signals from them. The version of this system considered here is the direct structure shown in Fig. 1, where $C_{12}(z)$ and $C_{21}(z)$ are the transfer functions of strictly causal M^{th}-order MA filters. The coefficients of each of these filters evolve vs. time and the value of the k^{th} coefficient, with $k \in [1, M]$, at time n is denoted $c_{ij}(n,k)$. These coefficients are adapted so as to decorrelate the time-shifted intermediate signals $u_1(n)$ and $u_2(n)$ of the BSS system, i.e. so as to fulfill the following criterion:

$$
R_{u_i u_j}(n, n-k) = 0, \quad i \neq j \in \{1,2\}, k \in [1, M], \tag{17}
$$

where

$$
R_{vw}(m_1, m_2) = E\{v(m_1)w(m_2)\} \tag{18}
$$

denotes the cross-correlation (i.e. the second-order cross-cumulant) of any couple of signals and where all the signals considered in this paper are assumed to be zero-mean for simplicity. The classical stochastic algorithm used in practice to fulfill the above criterion consists in adapting each filter coefficient at each time n according to the rule:

$$
c_{ij}(n+1,k) = c_{ij}(n,k) + \mu u_i(n) u_j(n-k) \qquad i \neq j \in \{1,2\}, k \in [1, M], \tag{19}
$$

where μ is a positive adaptation gain. This algorithm indeed implements the above criterion, as its equilibrium points correspond to:

$$
E\{c_{ij}(n+1,k) - c_{ij}(n,k)\} = 0 \qquad i \neq j \in \{1,2\}, k \in [1, M], \tag{20}
$$

which is clearly equivalent to (17).

The motivation for selecting this criterion (and associated algorithm) may be explained as follows. The internal state of the BSS system of Fig. 1 is defined by the values of the transfer functions $C_{12}(z)$ and $C_{21}(z)$. The state of interest in the BSS problem is the so-called "separating state", which yields non-permuted separated source signals on the system outputs, i.e:

$$S_i(z) = X_i(z), \quad i \in \{1,2\} \tag{21}$$

and which is defined by [6]:

$$C_{ij}(z) = A_{ij}(z), \quad i \neq j \in \{1,2\}. \tag{22}$$

The criterion to be used for adapting the filters $C_{12}(z)$ and $C_{21}(z)$ of the separating system must be selected so that it is fulfilled at the separating state, otherwise the separating filters cannot converge to this state of interest. The criterion defined by (17) indeed meets this requirement: at the separating state, the intermediate signals $u_1(n)$ and $u_2(n)$ resp. depend only on the source signals $x_1(n)$ and $x_2(n)$, so that their cross-correlation function involved in (17) is equal to zero.

3.2 Limitations of the classical approach

Now consider the situation when the available two mixed signals $Y_1(z)$ and $Y_2(z)$ contain not only the above contributions from sources $X_1(z)$ and $X_2(z)$, but also contributions from an arbitrary number of additional uncorrelated zero-mean sources $X_3(z)$ to $X_N(z)$, i.e:

$$\begin{aligned} Y_1(z) &= X_1(z) + A_{12}(z)X_2(z) + \sum_{j=3}^{N} A_{1j}(z)X_j(z) && (23) \\ Y_2(z) &= A_{21}(z)X_1(z) + X_2(z) + \sum_{j=3}^{N} A_{2j}(z)X_j(z), && (24) \end{aligned}$$

where N is the overall number of sources and where the additional transfer functions $A_{ij}(z)$ introduced here are also strictly causal M^{th}-order (or less) MA filters. Before focusing on the structure of Fig. 1, we consider the complete class of BSS systems which process these mixed signals in a linear convolutive way, based on the motivations presented in Section 1. The desired "optimum operation" of these systems for the considered mixed signals may be defined as follows. By linearly combining the mixed signals, only a single source contribution may be cancelled in any (non-zero) output of such systems for arbitrary mixtures. If $X_1(z)$ and $X_2(z)$ are considered as the main signals of interest, i.e. the useful signals to be separated, one would like these signals to appear resp. only in the outputs $S_1(z)$ and $S_2(z)$ of the BSS system. In other words, $X_1(z)$ and $X_2(z)$ are then the signals that should be cancelled in system outputs. These outputs would then also contain contributions from sources $X_3(z)$ to $X_N(z)$, which would then be considered as additional undesirable sources, i.e. "noise". The considered BSS system would thus perform the partial BSS of $X_1(z)$ and $X_2(z)$.

The *structure* of the BSS system shown in Fig. 1 is potentially suited to this optimum operation, in the sense that it is able to achieve the partial BSS of $X_1(z)$ and $X_2(z)$ as will now be shown. By combining the mixing equations (23)-(24) and the internal equations of the BSS system of Fig. 1, the intermediate signals $u_i(n)$ of the latter system may be expressed in the $\mathcal{Z}$ domain as:

$$U_i(z) = \sum_{p=1}^{N} H_{ip}(z) X_p(z) \quad i \in \{1,2\}, \tag{25}$$

with:

$$H_{ip}(z) = A_{ip}(z) - C_{ij}(z) A_{jp}(z) \quad i \neq j \in \{1,2\}, \tag{26}$$

where

$$A_{11}(z) = 1 \qquad \text{and} \qquad A_{22}(z) = 1. \tag{27}$$

When the separating filters take the values defined by (22), (26)-(27) yield:

$$H_{12}(z) = 0 \qquad \text{and} \qquad H_{21}(z) = 0. \tag{28}$$

Eq. (25) then shows that, for these separating filter values, the intermediate signals $U_1(z)$ and $U_2(z)$, and therefore the resulting output signals, resp. do not depend on $X_2(z)$ and $X_1(z)$. The partial BSS of $X_1(z)$ and $X_2(z)$ is thus achieved and the "partial separating state" corresponds to (22). Although this shows the structure of this BSS system to be adequate, the overall classical approach defined in Subsection 3.1 is not able to achieve this partial BSS, due the *criterion* (and associated algorithm) that it uses, as will now be shown. Eq. (25) reads in the time domain:

$$u_i(n) = \sum_{p=1}^{N} \sum_{m=1}^{2M} h_{ip}(n,m) x_p(n-m), \tag{29}$$

where $h_{ip}(n,m)$ are the time-varying coefficients of the impulse responses of the overall filters $H_{ip}(z)$ defined in (26). For any state of the BSS system, the cross-correlation of the signals defined in (29) may then be expressed as:

$$\begin{aligned} R_{u_i u_j}(n, n-k) &= \sum_{p=1}^{N} \sum_{m=1}^{2M} \sum_{l=1}^{2M} h_{ip}(n,m) h_{jp}(n-k,l) R_{x_p}(n-m, n-k-l) \\ & \qquad\qquad i \neq j \in \{1,2\}, k \in [1, M]. \end{aligned} \tag{30}$$

At the partial separating state, (28) yields:

$$h_{12}(n,m) = 0 \qquad \forall\, m \in [1, 2M] \tag{31}$$

$$h_{21}(n,m) = 0 \qquad \forall\, m \in [1, 2M]. \tag{32}$$

The contributions of $X_1(z)$ and $X_2(z)$ in (30) then disappear, while the contributions of the noise sources remain. In other words, in the case of noisy mixed signals, the cross-correlation values $R_{u_i u_j}(n, n-k)$ are non-zero at the partial separating state. The classical approach of Subsection 3.1 cannot then converge to this state, as it adapts the separating filter coefficients so as to reach a state where the cross-correlation values $R_{u_i u_j}(n, n-k)$ are cancelled. This approach then fails to achieve partial BSS.

3.3 Proposed differential approach

We here still consider the noisy mixed signals introduced in Subsection 3.2. Based on the above results, we again use the structure of Fig. 1, but we here aim at introducing a new criterion (and associated algorithms) for adapting its filter coefficients. This criterion is designed so that the resulting method becomes able to achieve the partial BSS of $X_1(z)$ and $X_2(z)$. The approach proposed to this end is based on the general differential BSS concept that we introduced in Section 2. The associated stationarity requirements only concern the statistical signal parameters used in the considered approach, i.e. second-order statistics. In other words, the useful sources $X_1(z)$ and $X_2(z)$ (resp. the noise sources $X_3(z)$ to $X_N(z)$) should have different (resp. identical) second-order statistics at times n_1 and n_2 when these times are separated by a "long" period. This long period is defined by contrast with each short period associated to a single time n_1 or n_2, over which sample statistics of the signals are measured in practice. As shown above, the main idea of differential BSS then consists in considering the difference between the signal statistics resp. associated to n_1 and n_2. To apply this general idea to the specific approach introduced in this section, we first define the "differential correlation function", which reads:

$$DR_{vw}(n_1, n_2, l_1, l_2) \quad = \quad R_{vw}(n_2 - l_1, n_2 - l_2) - R_{vw}(n_1 - l_1, n_1 - l_2), \tag{33}$$

where n_1 and n_2 are two reference times and l_1 and l_2 are two lags. When considering the difference between the two values, resp. associated to $n = n_1$ and $n = n_2$, of the cross-correlation function involved in the classical criterion (17), we obtain:

$$\begin{aligned} R_{u_iu_j}(n_2, n_2 - k) \quad - \quad & R_{u_iu_j}(n_1, n_1 - k) = DR_{u_iu_j}(n_1, n_2, 0, k), \\ & i \neq j \in \{1, 2\}, k \in [1, M]. \end{aligned} \tag{34}$$

We then use the same separating filter values for $n = n_1$ and $n = n_2$, and the same principle is also applied to any time-shifted version of this couple of times. The first overall filter coefficient $h_{ip}(n, m)$ involved in (30) then has the same value for $n = n_1$ and $n = n_2$ and this common value is simply denoted $h_{ip}(n, m)$ hereafter. The coefficient $h_{jp}(n - k, l)$ in (30) leads to the same phenomenon, so that combining (30) with (34) yields:

$$\begin{aligned} DR_{u_iu_j}(n_1, n_2, 0, k) \quad &= \quad \sum_{p=1}^{N} \sum_{m=1}^{2M} \sum_{l=1}^{2M} h_{ip}(n, m) h_{jp}(n - k, l) DR_{x_p}(n_1, n_2, m, k + l), \\ & \qquad i \neq j \in \{1, 2\}, k \in [1, M]. \end{aligned} \tag{35}$$

where the differential auto-correlation functions $DR_{x_p}(.)$ are also defined according to (33). Eq. (35) holds for any state and any type of sources. When the noise sources $X_3(z)$ to $X_N(z)$ have the above-mentioned long-term stationarity property, their differential auto-correlations $DR_{x_p}(.)$ contained in (35) are equal to zero, so that only the useful sources $X_1(z)$ and $X_2(z)$ remain and (35) reduces to:

$$\begin{aligned} DR_{u_iu_j}(n_1, n_2, 0, k) \quad &= \quad \sum_{p=1}^{2} \sum_{m=1}^{2M} \sum_{l=1}^{2M} h_{ip}(n, m) h_{jp}(n - k, l) DR_{x_p}(n_1, n_2, m, k + l), \\ & \qquad i \neq j \in \{1, 2\}, k \in [1, M]. \end{aligned} \tag{36}$$

In other words, from the point of view of the new parameter $DR_{u_iu_j}(n_1, n_2, 0, k)$ that we introduced, we actually get back in the classical 2-source to 2-sensor configuration (but this new parameter depends on the differential auto-correlation functions of the sources, instead of the plain auto-correlations which appear in the classical approach). Combining (36) and (31)-(32) then shows that $DR_{u_iu_j}(n_1, n_2, 0, k) = 0$ at the partial separating state whereas, if the sources $X_1(z)$ and $X_2(z)$ are long-term non-stationary, $DR_{u_iu_j}(n_1, n_2, 0, k) \neq 0$ for all other states (except for some possible spurious states, which correspond to those that may exist for the classical approach). The criterion that we eventually propose for performing the partial BSS of $X_1(z)$ and $X_2(z)$ in the case of noisy mixtures therefore consists in adapting all separating filter coefficients $c_{ij}(n, k)$ so as to achieve:

$$DR_{u_iu_j}(n_1, n_2, 0, k) = 0, \quad i \neq j \in \{1, 2\}, k \in [1, M]. \tag{37}$$

To summarize, the proposed approach operates as follows. Only the useful supposedly long-term non-stationary sources $x_1(n)$ and $x_2(n)$ contribute to the parameter $DR_{u_iu_j}(n_1, n_2, 0, k)$. In this summary of our method, we therefore only have to consider the "visible part" of any signal $u_i(n)$ or $u_j(n)$, i.e. its overall component associated to the subset of sources $\{x_1(n), x_2(n)\}$. For arbitrary separating filter values, each of the signals $u_i(n)$ and $u_j(n)$ contains contributions from both useful sources. The visible parts of $u_i(n)$ and $u_j(n)$ are then correlated and (34) yields $DR_{u_iu_j}(n_1, n_2, 0, k) \neq 0$. On the contrary, the partial separating state as defined in Subsection 3.2 corresponds to separating filter values such that the visible part of $u_1(n)$ is restricted to a contribution from $x_1(n)$, whereas the visible part of $u_2(n)$ only consists of a contribution from $x_2(n)$. The visible parts of $u_1(n)$ and $u_2(n)$ are then uncorrelated and (34) yields $DR_{u_iu_j}(n_1, n_2, 0, k) = 0$. Therefore, by forcing the separating filters to evolve so that $DR_{u_iu_j}(n_1, n_2, 0, k) = 0$ as stated in (37), our approach forces the BSS system to reach the partial separating state. Each intermediate signal $u_i(n)$, and therefore each associated output signal $s_i(n)$, then only contains a contribution from the single useful source $x_i(n)$ (plus of course contributions from the noise sources).

The practical differential algorithms associated to the partial BSS criterion that we introduced in (37) are therefore zero-search algorithms, as they adapt the coefficients $c_{ij}(n, k)$ so as to cancel the differential correlation values which appear in (37). A simple adaptation algorithm for performing this zero search then consists in updating each coefficient $c_{ij}(n, k)$ according to[2]:

$$c_{ij}(n+1, k) \;=\; c_{ij}(n, k) + \mu_i DR_{u_iu_j}(n_1, n_2, 0, k) \qquad i \neq j \in \{1, 2\}, k \in [1, M], \tag{38}$$

where the adaptation gains μ_1 and μ_2 are selected by taking into account stability requirements that are derived in Appendix A. Using (18) and (33), this algorithm reads explicitly:

$$\begin{aligned} c_{ij}(n+1, k) \;=\; & c_{ij}(n, k) + \mu_i[E\{u_i(n_2)u_j(n_2 - k)\} - E\{u_i(n_1)u_j(n_1 - k)\}], \\ & i \neq j \in \{1, 2\}, k \in [1, M]. \end{aligned} \tag{39}$$

The two expectations $E\{.\}$ in (39) are resp. estimated over two time domains associated to n_1 or n_2, where the useful sources must have different statistics, as explained above.

[2]We here apply the classical method for associating practical global and stochastic zero-search algorithms to given criteria which are based on the cancellation of any output signal statistical parameters. This classical method was e.g. already used in one of the earliest BSS approaches, i.e. in the Hérault-Jutten BSS neural network [14], with a reference [8] to the general Robbins-Monro stochastic adaptation framework.

One may also use the stochastic version of the above global algorithm (39), which reads:

$$
\begin{aligned}
c_{ij}(n+1,k) &= c_{ij}(n,k) + \mu_i[u_i(n_2)u_j(n_2-k) - u_i(n_1)u_j(n_1-k)], \\
&\qquad\qquad\qquad\qquad i \neq j \in \{1,2\}, k \in [1,M]
\end{aligned} \tag{40}
$$

and deserves the following comments. The classical algorithm (19) performs a sweep over the data by using an increasing time index n. The algorithm (40) proposed here is also based on a single sweep, but each step of this sweep involves two points in the data time series, corresponding to the indices n_1 and n_2. The difference between these indices is typically kept constant (and "long", as defined above), so that the sweep is performed in parallel over two time-shifted versions of the data. The considered couple of times inside these data is then defined by a single index, denoted n in (40), which is e.g. equal to n_1 or n_2.

4 Numerical tests

4.1 Basic configuration

We first validated the proposed second-order stochastic and global differential algorithms by means of tests performed with two artificial convolutive mixtures of three synthetic random sources. In the first series of tests, the useful sources $x_1(n)$ and $x_2(n)$ consist of 100000 samples, split in two 50000-sample periods. In each of these periods, these sources are independent and each of them is stationary, binary-valued and equiprobable. They are equal to ± 1 in their first period and ± 2 in the second one. The noise source $x_3(n)$ is uniformly distributed over the range $[-1, +1]$ in both periods. These three sources are mixed according to (23)-(24), where the mixing filters associated to the useful sources are set to:

$$
A_{12}(z) \simeq -0.381z^{-1} + 0.136z^{-2} + 0.081z^{-3} \tag{41}
$$

$$
A_{21}(z) \simeq -0.327z^{-1} - 0.184z^{-2} + 0.027z^{-3}, \tag{42}
$$

whereas the mixing filters associated to the noise source are:

$$
A_{13}(z) = z^{-1} + \frac{2}{3}z^{-2} + \frac{1}{3}z^{-3} \tag{43}
$$

$$
A_{23}(z) = \frac{1}{3}z^{-1} + \frac{2}{9}z^{-2} + \frac{1}{9}z^{-3}. \tag{44}
$$

Fig. 2 represents the evolution of the coefficients $c_{ij}(n,k)$ of the separating filters, adapted by means of the proposed stochastic algorithm (40), when the above-defined parallel sweep is performed over both 50000-sample periods. These coefficients converge to values which are close to those of the mixing filters $A_{12}(z)$ and $A_{21}(z)$. This shows the ability of the proposed approach to achieve the partial BSS of $X_1(z)$ and $X_2(z)$ defined by (22), although the contributions from the noise source $X_3(z)$ contained in the observed signals cover significantly larger ranges than those of the useful sources.

The global version (39) of the proposed algorithm also succeeds in achieving partial BSS in these conditions. Moreover, the filter coefficients then converge in a more accurate and much smoother way, as shown in Fig. 3. More precisely, the separating filters thus obtained after convergence are provided in Table 1 (with $L = 50000$ samples here). The convergence error E may then be defined as the highest absolute difference between these experimental filter coefficients and their target values defined in (41)-(42). This error is here equal to 0.015, which is quite limited.

In our second series of tests, we restricted the number of source samples in each of the two considered periods to a value L lower than above. This aimed at determining the minimum "frame length" required for the approach to operate correctly, and therefore the mimimum duration over which the source signals are requested to be (short-term) stationary. Table 1 shows the resulting values of the separating filters after convergence for the proposed global algorithm. The above-defined convergence error E increases from 0.015 to 0.036 and eventually 0.128 when L is decreased from 50000 to 5000 and eventually 500 samples. The two stationarity periods should therefore preferably contain a few thousand samples in order to achieve good convergence accuracy in the considered conditions. The stochastic version of our algorithm leads to the same conclusion, as it yields almost the same convergence accuracy as the global version for $L < 50000$ samples (but still with less smooth trajectories).

4.2 Extensions

The experimental results reported above correspond to one of the simplest configurations that may be considered for the proposed approach. However, this method also applies to more complex situations in terms of the number of sources or of their stationarity properties, as already stated above. We here ilustrate these two aspects.

Let us first consider again the same configuration as in Subsection 4.1, with two 50000-sample stationary periods for the two useful sources $x_1(n)$ and $x_2(n)$. We again mix them with the above-defined uniform noise source $x_3(n)$, using the filters defined in (41)-(44). However, we here also mix all these signals with an additional "noise" (i.e. long-term stationary) source, in order to illustrate the applicability of our differential approach to an arbibrary number of noise sources. The available two mixed signals are then defined by (23)-(24), where N is here set to 4 and where the additional parameters to be specified are: i) the noise source $x_4(n)$, which is a realization of a gaussian centered unit-variance random process, and ii) its associated mixing filters, which are set to:

$$A_{14}(z) = \frac{1}{6}z^{-1} + \frac{1}{12}z^{-2} + \frac{1}{12}z^{-3} \tag{45}$$

$$A_{24}(z) = \frac{1}{2}z^{-1} + \frac{1}{4}z^{-2} + \frac{1}{4}z^{-3}. \tag{46}$$

Our method is also suited to this configuration, since: i) the number of long-term non-stationary (and therefore useful) sources does not exceed the number of observations and ii) this method sets no restrictions on the number of long-term stationary (i.e. noise) sources. The proposed algorithms are applied to this situation exactly in the same way as in Subsection 4.1 and the filters $C_{12}(z)$ and $C_{21}(z)$ are again intended to resp. converge towards $A_{12}(z)$ and $A_{21}(z)$ defined in (41)-(42), thus achieving the (partial) separation of the two useful sources $x_1(n)$ and $x_2(n)$. Fig. 4 shows the actual evolution of the coefficients of the filters $C_{12}(z)$ and $C_{21}(z)$ when using the above-defined global differential algorithm. This confirms that these coefficients converge towards the correct values, i.e. that the proposed approach also succeeds in achieving partial separation when two noise sources are present. Moreover, the addition of the second noise source $x_4(n)$ yields almost no performance degradation:

- Comparing Fig. 3 and 4 shows that the separating coefficients have almost the same evolutions in the cases considered in Subsection 4.1 and here.

- A more detailed analysis of these curves proves that the above-defined convergence error E is here equal to 0.016, which is only slightly higher than the value 0.015 which was obtained in Subsection 4.1 in the same conditions (and esp. for the same signal realizations) except that $x_4(n)$ was not present.

The second aspect of the operation of the proposed method to be considered here concerns the short-term stationary periods of the useful sources. For the sake of clarity, the configuration that we presented in Subsection 4.1 was based on the simplest situation required for the proposed approach, where two short-term stationary periods are considered for each useful source and these periods are the same for both sources. Let us stress that the proposed approach also applies to much more general situations. Indeed, this approach is esp. intended for piecewise stationary useful sources. We then set no conditions on the number of such "pieces", i.e. of short-term stationary periods (provided at least two long enough periods exist for each each source). We also set no restrictions on the positions in time of these periods, and especially the periods resp. associated to the two useful sources need not be synchronized: even when the useful sources consist of a series of non-synchronized stationary periods, the intersections of these periods form domains where *both* sources are stationary and where our differential method may therefore be applied (provided these periods are long enough).

We now illustrate these properties by presenting a test performed in the following configuration, which may be considered as an extension of the previous ones. Each useful source is again derived from the 100000-sample realization of a random ± 1 binary process. However, this realization is here split into a few periods, (i.e. 5 periods for $x_1(n)$ and 9 for $x_2(n)$) and a specific constant scaling factor is applied to each realization on each period, thus yielding the eventual sources considered hereafter. The detailed definitions of these sources are provided in Tables 2 and 3. It should be noted that no stationary period boundaries are shared by $x_1(n)$ and $x_2(n)$. These sources are shown in Fig. 5 a) and b). They are then mixed with the same two noise sources as above, using the same mixing filters. The resulting mixed signals $y_1(n)$ and $y_2(n)$ are provided in Fig. 5 c) and d).

Allowing the stationary periods of the useful sources to take arbitrary unknown positions yields a practical issue for the proposed approach, i.e. how to detect the time domains when both sources are stationary, in order to apply our differential approach to two of these domains. A simple practical solution to this problem consists in detecting these domains as the domains where both mixed signals have (almost) constant powers. For instance, we may operate as follows in order to find two such 5000-sample domains, where our differential BSS method will be applied. We first compute the mean powers of both observed signals $y_1(n)$ and $y_2(n)$ over successive adjacent 1000-sample domains. The resulting power profiles are shown in Fig. 5 e) and f). It should be noted that they are coherent with the envelopes of the mixed signals provided in Fig. 5 c) and d). These mean power profiles are also shown in more detail in Fig. 6 a) and 7 a). The latter figures confirm that these mean powers are "constant" (up to the statistical fluctuations of the signals over the considered 1000-sample windows) on the periods where both sources are stationary. An automated procedure must then be designed for deriving from these mean power profiles the two 5000-sample time domains where the mixed signals have the most constant powers. The procedure that we propose to this end operates as follows. For each mixed signal, we successively consider each time position associated to a 1000-sample window where one of the above mean power values was computed. For each such position, we compute the variance[3] associated to 5 values of the mean power of the considered mixed signal, where these 5 values resp. correspond to the considered 1000-sample window, to the two 1000-sample windows situated on its left and to the two 1000-sample windows situated on its right. We thus obtain two variance profiles, resp. associated to the two mixed signals, which are shown in Fig. 6 b) and 7 b). We then select the two 5000-sample periods which yield the lowest variance values for both mixed signals. For the signals considered in this test, these periods turned out to resp. correspond to the sample ranges [5001,10000] and [90000,95000].

When applying our global differential BSS approach to these periods, the convergence error E for the separating coefficients was 0.0056. This is coherent with the error that we reported in Subsection 4.1 for 5000-sample periods (i.e. $E = 0.036$), considering that we included an additional noise source here and that only a coarse comparison may be performed between these two tests, as the stationary periods used in these tests do not correspond to the same signal realizations.

5 Conclusions

Classical BSS methods only apply to the situation when the number of observed signals is at least equal to the number of sources to be separated. In this paper, we considered the opposite case, i.e. underdetermined or noisy mixtures. We introduced a general differential BSS concept which then performs partial BSS, i.e. which separates a subset of the considered sources, whereas the other sources are made invisible in BSS optimization criteria (so that no restriction is set on the number of such sources). This general concept may e.g. be used to derive extended versions of various classical BSS methods for instantaneous or convolutive mixtures. We illustrated this approach esp. by applying it to a convolutive criterion and associated algorithms based on the global or stochastic cancellation of the above-defined "differential cross-correlation function" of intermediate signals of the BSS system. This particular method, which is the differential version of the classical decorrelation approach, adapts the filters of a direct BSS system so as to perform the separation of two non-stationary sources from two convolutive mixtures which also contain an arbitrary number of stationary noise sources. We analyzed the stability of the proposed algorithms, which defines the signs of the adaptation gains to be used. We also demonstrated the effectiveness of these algorithms by means of numerical tests. Our future investigations will esp. concern the application of the proposed general concept to i) specific BSS algorithms which are suited to more general mixing conditions and to ii) differential parameters associated to other source properties than their non-stationarity.

Acknowledgments

The authors would like to thank Stéphane Savoldelli for his participation in this activity during his training period.

[3] A modified version of this approach would consist in normalizing this variance by the squared mean of the values over which this variance is computed, in order for this criterion not to depend on the average power of the signals over the considered domain.

A Stability analysis

We proved in Subsection 3.3 that (37) is met at the partial separating state. This state is therefore an equilibrium point of the proposed stochastic and global algorithms. This equilibrium point must be stable in addition. We here investigate in which conditions this requirement is met. To this end, we use the Ordinary Differential Equation (ODE) method [3], which makes it possible to analyze the local asymptotic behavior of adaptive systems whose overall updating algorithm may be expressed in vector form as:

$$\theta_{n+1} = \theta_n + H(\theta_n, \xi_{n+1}), \tag{47}$$

where θ_n, $H(\theta_n, \xi_{n+1})$ and ξ_{n+1} are the column vectors resp. composed of:

- the adaptive parameters of the system, which define its state,
- the updating terms for the parameters contained in θ_n,
- the signal values required to define the above updating terms.

The equilibrium points of (47) are all the constant state vectors θ^* for which

$$\lim_{n \to \infty} E_{\theta^*}[H(\theta^*, \xi_{n+1})] = 0, \tag{48}$$

where $E_{\theta^*}[.]$ denotes the mathematical expectation with respect to the probability law of the vector ξ_{n+1} for a given vector θ^*. Each equilibrium point θ^* of (47) may be stable or not, depending on the properties of the function H and on the statistics of the vectors $(\xi_n)_{n \geq 0}$. The ODE approach allows one to analyze stability for stationary sources by approximating the discrete-time recurrence (47), under some conditions[4] on H, by a continuous-time differential system that reads:

$$\frac{d\theta}{dt} = \lim_{n \to +\infty} E_\theta[H(\theta, \xi_{n+1})]. \tag{49}$$

This differential system is locally stable in the vicinity of an equilibrium point θ^* if and only if (iff) the associated tangent linear system:

$$\frac{d\theta}{dt} = J(\theta^*)(\theta - \theta^*) \tag{50}$$

is stable, i.e. iff all the eigenvalues of $J(\theta^*)$ have negative real parts. For any state θ, $J(\theta)$ denotes the corresponding Jacobian matrix of the system, i.e. the matrix composed of the partial derivatives

$$J_{ij}(\theta) = \lim_{n \to +\infty} \frac{\partial (E_\theta[H(\theta, \xi_{n+1})]^{(i)})}{\partial \theta^{(j)}}, \tag{51}$$

where $E_\theta[H(\theta, \xi_{n+1})]^{(i)}$ is the i^{th} component of $E_\theta[H(\theta, \xi_{n+1})]$ and $\theta^{(j)}$ is the j^{th} component of θ.

We first have to apply this ODE approach to a slightly extended version of the classical decorrelation approach (19), where we introduce two independent adaptation gains μ_1 and

[4] The adaptation gains should be sufficiently small. The other conditions on H concern its regularity [3].

μ_2 resp. for adapting the separating filters $C_{12}(z)$ and $C_{21}(z)$. The adaptation rule (19) then becomes:

$$c_{ij}(n+1,k) \quad = \quad c_{ij}(n,k) + \mu_i u_i(n) u_j(n-k) \qquad i \neq j \in \{1,2\}, k \in [1,M]. \tag{52}$$

This extended rule falls in the class of algorithms defined by (47) and corresponds to:

$$\begin{aligned}
\theta_n &= [c_{12}(n,1),\ldots,c_{12}(n,M),c_{21}(n,1),\ldots,c_{21}(n,M)]^T && (53)\\
H(\theta_n,\xi_{n+1}) &= [\mu_1 u_1(n)u_2(n-1),\ldots,\mu_1 u_1(n)u_2(n-M), \\
&\quad \mu_2 u_2(n)u_1(n-1),\ldots,\mu_2 u_2(n)u_1(n-M)]^T && (54)\\
\xi_{n+1} &= [y_1(n),y_2(n),u_1(n-1),\ldots,u_1(n-M), \\
&\quad u_2(n-1),\ldots,u_2(n-M)]^T. && (55)
\end{aligned}$$

The corresponding Jacobian matrix $J(\theta^s)$ at the separating state θ^s for two sources may then be derived from (53)-(55) by means of (51) (a more detailed description of a relatively similar analysis may be found in [6]). It has a complex expression, which does not allow one to easily derive its eigenvalues, as required by the ODE approach. In order to obtain more meaningful results, we focus on the case when the following two conditions are met:

1. The sources are temporally white (at order 2), i.e:

$$R_{x_i}(m) = 0 \quad \text{if } m \neq 0, \tag{56}$$

   where correlation functions $R_{x_i}(.)$ have a single argument in this part of the discussion, as the sources are supposedly stationary. The powers or variances $R_{x_i}(0)$ of these centered sources are denoted P_{x_i} hereafter.

2. The mixture ratio is low, i.e. the coefficients of the mixing filters $A_{12}(z)$ and $A_{21}(z)$ are very small.

The Jacobian matrix $J(\theta^s)$ at the separating state may then be shown to consist of four simple sub-matrices, i.e:

$$J(\theta^s) \simeq \begin{pmatrix} -\mu_1 P_{x_2} I_M & 0 \\ 0 & -\mu_2 P_{x_1} I_M \end{pmatrix}, \tag{57}$$

where I_M is the M^{th}-order identity matrix. These calculations only concern the (extended) classical algorithm. We presented them however, because they are a required first step of our analysis, which has not been reported in the literature to our knowledge, and from which we can then easily derive the results to be established for the differential approach proposed in this paper. As a by-product of this investigation, the stability condition for the (extended) classical algorithm may first be derived as follows from the above results. The matrix $J(\theta^s)$ obtained in (57) is diagonal and its eigenvalues are $-\mu_1 P_{x_2}$ and $-\mu_2 P_{x_1}$. As explained above, the separating state is a stable equilibrium point for this BSS algorithm iff these eigenvalues are negative. As the source signal powers P_{x_2} and P_{x_1} are always positive, this stability condition reads:

$$\begin{cases} \mu_1 > 0 \\ \mu_2 > 0 \end{cases} \tag{58}$$

This is the reason why the classical algorithm (19) uses $\mu_1 = \mu_2 = \mu > 0$.

The algorithm (40) proposed in this paper also falls in the class of adaptation rules defined by (47). Moreover, it is related in a simple linear way to the algorithm (52) that we just analyzed: the function value $H(\theta_n, \xi_{n+1})$ for our algorithm is the difference between the two values, resp. at times n_2 and n_1, of the function $H(\theta_n, \xi_{n+1})$ corresponding to the algorithm (52). Moreover, the ODE approach itself is also linear with respect to $H(\theta_n, \xi_{n+1})$. Therefore, when applying this ODE approach to the algorithm (40) that we proposed, the expressions obtained in the successive steps of this analysis are straighforwardly derived from those obtained above for the classical algorithm (52): the expressions obtained for the latter algorithm are replaced by the difference of their values between times n_2 and n_1. In other words, we here get the differential version of the previous analysis, which is natural as we consider the differential version of the previous algorithm. Especially, the eigenvalues of the Jacobian matrix here become $-\mu_1 DP_{x_2}(n_1, n_2)$ and $-\mu_2 DP_{x_1}(n_1, n_2)$, where we define the differential power of any signal $v(n)$ for times n_1 and n_2 as:

$$DP_v(n_1, n_2) = R_v(n_2, n_2) - R_v(n_1, n_1). \tag{59}$$

The proposed differential algorithm is then locally stable at the partial separating state iff:

$$\begin{cases} \mu_1 DP_{x_2}(n_1, n_2) > 0 \\ \mu_2 DP_{x_1}(n_1, n_2) > 0 \end{cases} \tag{60}$$

The differential powers cannot be removed from this condition, as they may be positive or negative, depending on the considered source signals. This should be contrasted with their classical, i.e. non-differential, counterparts which appeared in the classical approach and which are always positive. The signs of the adaptation gains μ_1 and μ_2 should therefore be selected according to the signs of the differential powers of the source signals (which may e.g. be estimated in practice by adapting the technique that we developed elsewhere for estimating the signs of source kurtosis [10]). For the sake of simplicity, both adaptation gains may have the same absolute value, i.e: $|\mu_1| = |\mu_2| = \mu > 0$.

References

[1] F. Abrard, Y. Deville, P. White, "From blind source separation to blind source cancellation in the underdetermined case: a new approach based on time-frequency analysis", Proc. 3rd Int. Conf. Independent Component Anal. and Signal Separation, San Diego, California, 9-13 December 2001, pp. 734-739

[2] M.J. Al-Kindi and J. Dunlop, "Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel", *Signal Process.*, Vol. 17, 1989, pp. 241-250

[3] A. Benveniste et al., "Adaptive algorithms and stochastic approximations", Applications of Mathematics, Vol. 22, Springer-Verlag, 1990

[4] X-R Cao and R-W Liu, "General approach to blind source separation", IEEE Trans. Signal Process., Vol. 44, No. 3, March 1996, pp. 562-571

[5] J.-F. Cardoso, "Blind signal separation: statistical principles", Proc. IEEE, Vol. 86, No. 10, Oct. 1998, pp. 2009-2025

[6] N. Charkani and Y. Deville, "Self-adaptive separation of convolutively mixed signals with a recursive structure. Part I: stability analysis and optimization of asymptotic behaviour", Signal Process., Vol. 73, No. 3, 1999, pp. 225-254

[7] A. Cichocki, S.-I. Amari, Adaptive blind signal and image processing. Learning algorithms and applications, Wiley, Chichester, England, 2002

[8] P. Comon et al., "Blind separation of sources, Part II: Problems statement", Signal Process., Vol. 24, No. 1, 1991, pp. 11-20

[9] P. Comon and O. Grellier, "Non-linear inversion of underdetermined mixtures", Proc. Int. Workshop Independent Component Anal. Blind Signal Separation, Aussois, France, 11-15 January 1999, pp. 461-465

[10] Y. Deville and M. Benali, "A criterion for deriving the sub/super Gaussianity of source signals from their mixtures", Proc. 10th European Signal Process. Conf., Tampere, Finland, 4-8 September 2000, session WedAmPO1

[11] F. Gamboa, "Separation of sources having unknown discrete supports", IEEE Workshop Higher-Order Stat., Begur, Spain, 12-14 June 1995, pp. 56-60

[12] A. Hyvarinen et al., Independent Component Analysis, Wiley, New York, 2001

[13] A. Jourjine, S. Rickard, O. Yilmaz, "Blind separation of disjoint orthogonal signals: demixing N sources from 2 mixtures", Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process., Istanbul, Turkey, 5-9 June 2000, pp. 2985-2988

[14] C. Jutten and J. Hérault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture", Signal Process., Vol. 24, No. 1, July 1991, pp. 1-10

[15] J.M. Mendel, "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications", Proc. IEEE, Vol. 79, No. 3, March 1991, pp. 278-305

[16] C.L. Nikias and J.M. Mendel, "Signal processing with higher-order spectra", IEEE Sig. Process. Mag., July 1993, pp. 10-37

[17] K. J. Pope and R.E. Bogner, "Blind signal separation II. Linear, convolutive combinations", Digital Signal Process., Vol. 6, 1996, pp. 17-28

[18] A. Taleb, C. Jutten, "On underdetermined source separation", Proc. IEEE Int. Conf. Acoust., Speech, and Signal Process., Phoenix, Arizona, 15-19 March 1999, pp. 1445-1448

[19] S. Van Gerven, D. Van Compernolle, "Feedforward and feedback in a symmetric adaptive noise canceler: stability analysis in a simplified case", Proc. 6th European Signal Process. Conf., Brussels, Belgium, 1992, pp. 1081-1084

[20] S. Van Gerven, D. Van Compernolle, "Signal separation by symmetric adaptive decorrelation: stability, convergence and uniqueness", IEEE Trans. Signal Process., Vol. 43, No. 7, July 1995, pp. 1602-1612

Table 1: Separating filters after convergence of proposed global algorithm vs. number L of source samples in each period.

L	$C_{12}(z)$ and $C_{21}(z)$
500	$C_{12}(z) \simeq -0.509z^{-1} + 0.045z^{-2} + 0.120z^{-3}$
	$C_{21}(z) \simeq -0.383z^{-1} - 0.087z^{-2} + 0.131z^{-3}$
5000	$C_{12}(z) \simeq -0.351z^{-1} + 0.172z^{-2} + 0.066z^{-3}$
	$C_{21}(z) \simeq -0.340z^{-1} - 0.190z^{-2} + 0.048z^{-3}$
50000	$C_{12}(z) \simeq -0.396z^{-1} + 0.142z^{-2} + 0.086z^{-3}$
	$C_{21}(z) \simeq -0.331z^{-1} - 0.184z^{-2} + 0.022z^{-3}$

Table 2: Definition of source $x_1(n)$: periods over which it is stationary and scaling factors applied to the original ± 1 binary random process to create $x_1(n)$.

period	scaling factor
[1,20000]	1
[20001,40000]	3
[40001,60000]	0.5
[60001,80000]	2
[80001,100000]	1.5

Table 3: Definition of source $x_2(n)$: same principle as previous table.

period	scaling factor
[1,5000]	2
[5001,15000]	1
[15001,30000]	2.5
[30001,45000]	0.5
[45001,55000]	3.5
[55001,65000]	1
[65001,75000]	2
[75001,90000]	3
[90001,100000]	1.5

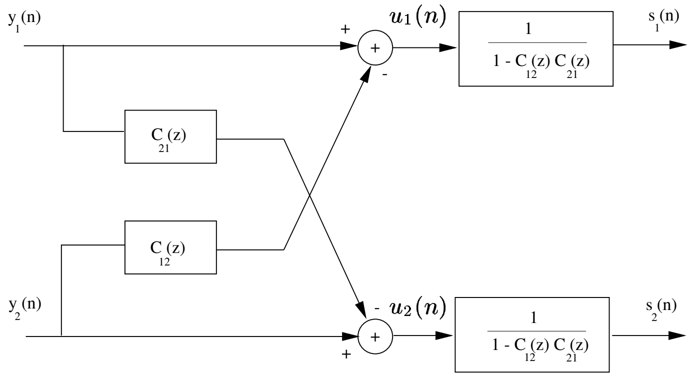

Figure 1: Separating system based on a direct structure.

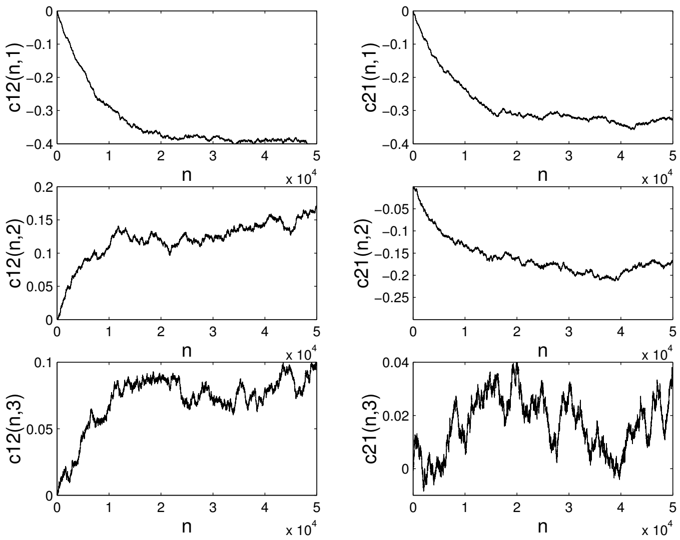

Figure 2: Evolution of the coefficients $c_{ij}(n, k)$ of the strictly causal MA separating filters adapted with the proposed stochastic algorithm, for one noise source (left column: $C_{12}(z)$, right column: $C_{21}(z)$).

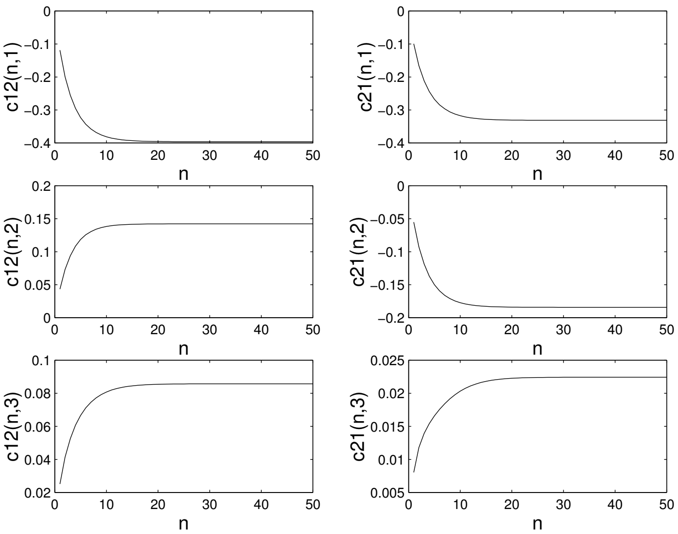

Figure 3: Evolution of the coefficients $c_{ij}(n,k)$ of the strictly causal MA separating filters adapted with the proposed global algorithm, for one noise source (left column: $C_{12}(z)$, right column: $C_{21}(z)$). n is the index of each update performed by taking into account all the available data.

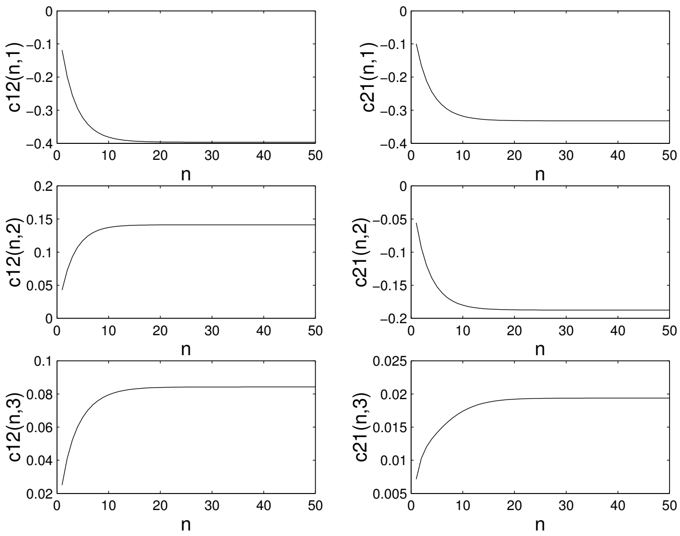

Figure 4: Evolution of the coefficients $c_{ij}(n,k)$ of the strictly causal MA separating filters adapted with the proposed global algorithm, for two noise sources (left column: $C_{12}(z)$, right column: $C_{21}(z)$).

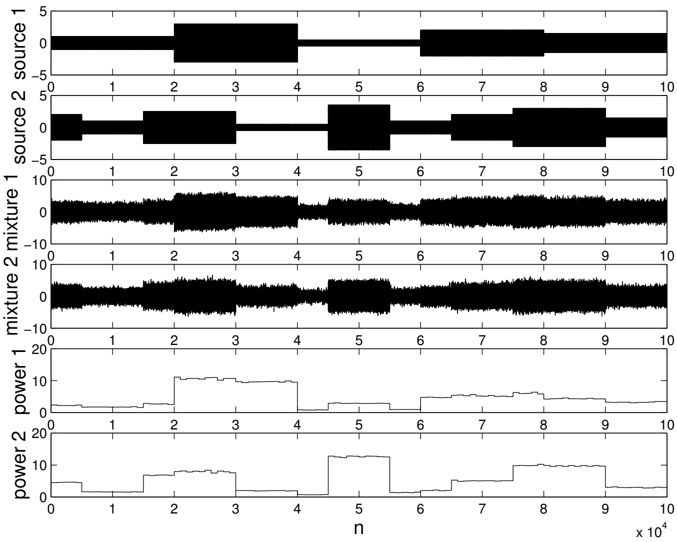

Figure 5: Top two plots = a) and b): two useful sources. Middle two plots = c) and d): two mixtures of these useful sources and of two noise sources. Bottom two plots = e) and f): mean powers of mixed signals on adjacent 1000-sample windows.

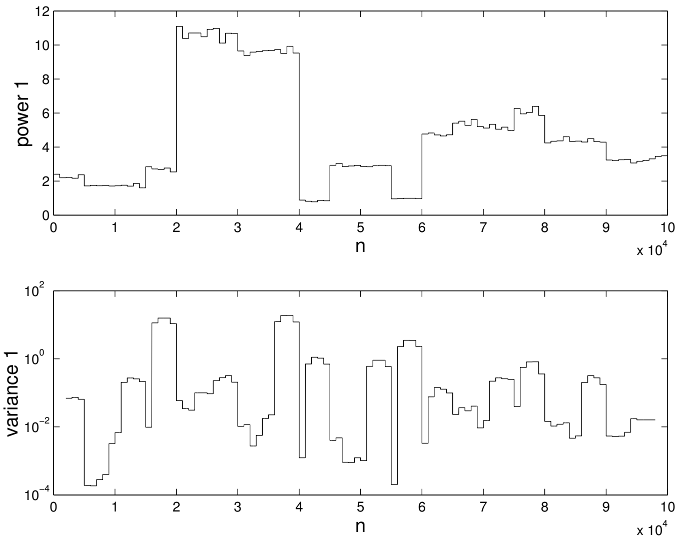

Figure 6: Top plot = a): mean powers of first mixed signal on adjacent 1000-sample windows. Bottom plot = b): variances of above mean powers on 5000-sample windows, with a 1000-sample stepsize.

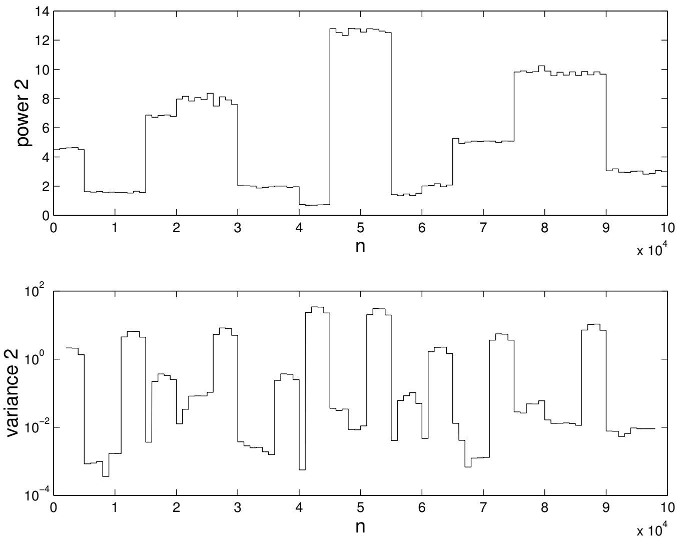

Figure 7: Same principle as previous figure, for second mixed signal.

Figure captions:

Fig. 1: Separating system based on a direct structure.

Fig. 2: Evolution of the coefficients $c_{ij}(n,k)$ of the strictly causal MA separating filters adapted with the proposed stochastic algorithm, for one noise source (left column: $C_{12}(z)$, right column: $C_{21}(z)$).

Fig. 3: Evolution of the coefficients $c_{ij}(n,k)$ of the strictly causal MA separating filters adapted with the proposed global algorithm, for one noise source (left column: $C_{12}(z)$, right column: $C_{21}(z)$). n is the index of each update performed by taking into account all the available data.

Fig. 4: Evolution of the coefficients $c_{ij}(n,k)$ of the strictly causal MA separating filters adapted with the proposed global algorithm, for two noise sources (left column: $C_{12}(z)$, right column: $C_{21}(z)$).

Fig. 5: Top two plots = a) and b): two useful sources. Middle two plots = c) and d): two mixtures of these useful sources and of two noise sources. Bottom two plots = e) and f): mean powers of mixed signals on adjacent 1000-sample windows.

Fig. 6: Top plot = a): mean powers of first mixed signal on adjacent 1000-sample windows. Bottom plot = b): variances of above mean powers on 5000-sample windows, with a 1000-sample stepsize.

Fig. 7: Same principle as previous figure, for second mixed signal.