Reprint of paper published in "Signal Processing"

Paper reference:
N. Charkani, Y. Deville, "Self-adaptive separation of convolutively mixed signals with a recursive structure. Part I: stability analysis and optimization of asymptotic behaviour", Signal Processing, vol. 73, no. 3, pp. 225-254, 1999.

Self-Adaptive Separation of Convolutively Mixed Signals with a Recursive Structure - Part I: Stability Analysis and Optimization of Asymptotic Behaviour

Nabil CHARKANI (1)(2) Yannick DEVILLE (3)†

(1) Laboratoires d'Electronique Philips S.A.S. (LEP)
22, Avenue Descartes, BP 15, 94453 Limeil-Brévannes Cedex, FRANCE
(2) Laboratoire de Traitement d'Images et Reconnaissance de Formes (LTIRF - INPG)
46, Avenue Félix Viallet, 38031 Grenoble Cedex, FRANCE
(3) Laboratoire d'Acoustique, de Métrologie et d'Instrumentation (LAMI),
Université Paul Sabatier, 38 Rue des 36 Ponts, 31400 Toulouse, FRANCE

e-mail: charkani@lep-philips.fr
e-mail: ydeville@cict.fr

† Author to whom correspondence about this paper should be sent.

Short title: convolutively mixed signal separation

Keywords: asymptotic behaviour optimization, blind source separation, convolutive mixtures, higher-order statistics, non-linear algorithms, ordinary differential equation, recurrent structure, self-adaptive algorithms, stability analysis.

Number of typewritten pages of the paper: 51

Number of tables of the paper: 0

Number of figures of the paper: 3

Figure captions:

Fig. 1: Basic mixture model for source separation

Fig. 2: Direct structure for the separation system

Fig. 3: Recurrent structure for the separation system

Abstract: In this paper, we investigate the self-adaptive source separation problem for convolutively mixed signals. The proposed approach uses a recurrent structure adapted by a generic rule involving arbitrary separating functions. A stability analysis of this algorithm is first performed. It especially applies to some classical rules for instantaneous and convolutive mixtures that were proposed in the literature but only partly analyzed. The expression of the asymptotic error variance is then determined for strictly causal mixtures. This enables to derive the optimum separating functions that minimize this error variance. They are shown to be only related to the probability density functions of the sources. To perform this error minimization, two normalization procedures that improve the algorithm properties are proposed. Their stability conditions and their asymptotic behaviour are analyzed.

Résumé: Dans cet article, nous traitons le problème de la séparation auto-adaptative de sources pour des mélanges convolutifs. L'approche proposée utilise une structure récurrente adaptée par une règle générique basée sur des fonctions séparatrices arbitraires. On effectue d'abord une analyse de la stabilité de cet algorithme. Elle s'applique notamment à plusieurs règles classiques pour des mélanges instantanés ou convolutifs qui n'ont été que partiellement analysées dans la littérature . L'expression de la variance de l'erreur asymptotique est ensuite déterminée dans le cas de mélanges strictement causaux. Ceci permet de calculer les fonctions séparatrices optimales au sens de la minimisation de la variance de l'erreur. On montre qu'elles ne dépendent que de la densité de probabilité des sources. Pour réaliser cette minimisation d'erreur, deux procédures de normalisation qui améliorent les propriétés de l'algorithme sont proposées. Leurs conditions de stabilité et leurs performances asymptotiques sont analysées.

1 Introduction

Multichannel blind (or self-adaptive) source separation is a basic topic in signal processing. It has recently received increased attention due to the importance of its potential applications. It arises in many fields of engineering and applied sciences including antenna array processing, geophysical data processing, noise reduction, speech processing, biological system analysis etc.... It consists in recovering signals emitted by unknown sources and mixed by an unkown medium, using only several observations of the mixtures. The only assumptions made are the linearity of the mixing system and the statistical independence of the original signals.

The methods proposed for achieving blind source separation may be classified in various ways. A possible classification can be made depending whether the mixtures are instantaneous or convolutive.

Instantaneous mixture: this corresponds to memoryless mixing systems, i.e. each observation is only a linear conbination of the source signals at the same time position. This problem was first formulated independently by Hérault, Jutten and Ans [14] in biological applications and by Bar-Ness, Carlin and Steinberger [2] in the satellite communication area. Then, many contributions from different authors emphasized all the potential applications of this technique and provided new solutions to this problem. Those studies introduced and developed the concept of Independent Component Analysis as a new method for data processing and time series analysis and as an extension of the well known Principal Component Analysis [7], [16]. The approaches presented in this domain explore the different possible formulations of statistical independence: cross-moment cancellation criteria [15], minimization of sums of cumulants [5],[8],[18],[20], minimization of constrast functions [7],[21], maximum likelihood approaches [12],[30] and more recently geometrical approaches [31] and criteria based on entropy minimization (see e.g. [3]).

Convolutive mixture: this corresponds to mixing systems with time memory. It represents a more general case than the mere instantaneous mixture assumption, and it especially concerns acoustic applications. The literature in this domain is relatively recent and is more reduced than in the instantaneous case. Nevertheless, some major contributions were proposed in both time and frequency domains. Especially, Al-Kindi et al. [1] and Van Gerven et al. [33] suggested to achieve source separation by decorrelating the outputs of the system. Nguyen et al. developed an extension of the Hérault-Jutten algorithms based on fourth-order cross-moment or fourth-order cross-cumulant cancellation [24]-[26]. Weinstein et al. [35] also developed a non-parametric approach using second-order cross-spectra, whereas Yellin et al. [37] applied the same idea to higher-order cross-spectra.

The difference of maturity between the fields of instantaneous and convolutive mixtures also concerns the results that have been reported about the properties of the proposed algorithms resulting from the criteria defined above. These properties concern the equilibrium points, their stability and the asymptotic behaviour at these points. They have been deeply investigated for the algorithms proposed for instantaneous mixtures, whereas the results that have been provided for convolutive mixtures are more restricted [1],[33],[34] because of the mathematical complexity of the required analyses.

In this paper, we present a detailed investigation of those yet unexplored aspects of convolutive source separation. In Section 2, we define the source separation structure and adaptation algorithm that we consider, and their connection with previously reported approaches. In Section 3, we consider the equilibrium points of this adaptation algorithm and we define conditions on its separating functions that guarantee that the separating point is an equilibrium point. The stability of this separating point is analyzed in Section 4. General conditions on the separating functions are first provided. As a by-product, they are then applied to specific algorithms which have been proposed in the literature for convolutive or instantaneous mixtures but only partly analyzed up to now. The asymptotic (i.e. steady-state) error of the proposed approach is described in Section 5. It is shown to depend on the selected separating functions. The optimum class of functions is then derived. Specific members of this class are obtained in Section 6 by using two alternative normalization procedures. Conlusions drawn from this investigation are presented in Section 7 and extensions for practical application are outlined.

2 Problem statement

2.1 Mixture model

Let us consider the two-dimensional convolutive mixture model illustrated in Fig. 1, where $(x_1(n), x_2(n))^T$ and $(y_1(n), y_2(n))^T$ are respectively the source signal vector and the observation vector. Despite its simplicity, this scheme is frequently used in source separation studies [24],[35],[37]. The source-observation relationship of this model can be expressed as follows in the Z-domain:

$$\begin{pmatrix} Y_1(z) \\ Y_2(z) \end{pmatrix} = \begin{pmatrix} 1 & A_{12}(z) \\ A_{21}(z) & 1 \end{pmatrix} \begin{pmatrix} X_1(z) \\ X_2(z) \end{pmatrix}, \tag{1}$$

where:

- $X_i(z)$ and $Y_i(z)$ are respectively the Z-transforms of $x_i(n)$ and $y_i(n)$.
- $A_{ij}(z), i \neq j \in \{1, 2\}$ is the transfer function of the channel that links source j to sensor i. The impulse response of this channel is denoted as $(a_{ij}(k))_{k \in \mathcal{Z}}$ hereafter.

The matrix in (1) that transforms $(X_1(z), X_2(z))^T$ into $(Y_1(z), Y_2(z))^T$ is called the *mixing matrix*. Furthermore, we assume that:

- (AS1) the filters $A_{ij}(z), i \neq j \in \{1, 2\}$ have a causal moving average (MA) structure with order M_i. The mixing matrix is thus causal and stable.
- (AS2) the sources $x_1(n)$ and $x_2(n)$ are stationary, zero-mean and statistically independent.

The time domain source-observation relationship can then be written as:

$$\begin{cases} y_1(n) = x_1(n) + \sum_{k=0}^{M_1} a_{12}(k) x_2(n-k) \\ y_2(n) = \sum_{k=0}^{M_2} a_{21}(k) x_1(n-k) + x_2(n) \end{cases} \tag{2}$$

2.2 Separation structures

The aim of the blind source separation technique is to recover the original sources $x_1(n)$ and $x_2(n)$ by using only the observations $y_1(n)$ and $y_2(n)$. This can be achieved by estimating the inverse of the mixing matrix [24],[35],[37]. This strategy uses a reconstruction structure called a *separation system*, which consists of an implementation of the inverse of the mixing matrix. The source separation problem is then reformulated as an inverse problem. It can be solved if the mixing model is "invertible" (i.e. if the mixing matrix is minimum phased, as explained at the end of the current subsection). Two basic structures may be used to implement the inverse of the mixing matrix, i.e. the direct structure [33],[35] (see Fig. 2) and the recurrent structure [24],[37] (see Fig. 3). Both schemes yield the same observation-output relationship, i.e. in the Z-domain:

$$\begin{pmatrix} S_1(z) \\ S_2(z) \end{pmatrix} = \frac{1}{1 - C_{12}(z)C_{21}(z)} \begin{pmatrix} 1 & -C_{12}(z) \\ -C_{21}(z) & 1 \end{pmatrix} \begin{pmatrix} Y_1(z) \\ Y_2(z) \end{pmatrix}. \tag{3}$$

Therefore, the source-output relationship for both schemes is:

$$\begin{pmatrix} S_1(z) \\ S_2(z) \end{pmatrix} = \frac{1}{1 - C_{12}(z)C_{21}(z)} \begin{pmatrix} 1 - C_{12}(z)A_{21}(z) & A_{12}(z) - C_{12}(z) \\ A_{21}(z) - C_{21}(z) & 1 - C_{21}(z)A_{12}(z) \end{pmatrix} \begin{pmatrix} X_1(z) \\ X_2(z) \end{pmatrix}. \tag{4}$$

To achieve source separation, the coefficients of the filters C_{ij} should be selected so that the output signals $s_i(n)$ are equal to the source signals $x_i(n)$, up to a permutation and a shaping filter. This yields two solutions:

$$C_{ij}(z) = A_{ij}(z), \quad i \neq j \in \{1, 2\} \implies S_i(z) = X_i(z), \quad i \in \{1, 2\} \tag{5}$$

$$C_{ij}(z) = \frac{1}{A_{ji}(z)}, \quad i \neq j \in \{1, 2\} \implies S_i(z) = \frac{X_j(z)}{A_{ij}(z)}, \quad i \neq j \in \{1, 2\}. \tag{6}$$

Hereafter, the filters $C_{ij}(z)$ are constrained to have a causal MA structure. In this case, the solution (6) cannot be reached and only the solution defined by (5) is valid. It is called *the separating solution* hereafter. However, (5) does not provide a practical means for choosing the filters of the separation structure, since the mixing filters $A_{ij}(z)$ are unknown. Therefore, criteria for assigning the filters C_{ij} must be defined. Such criteria are presented in the next subsection. Before this, we present some assumptions and notations that are used in the remainder of this paper.

Assumptions:

- (AS3) Order(C_{ij}) = Order$(A_{ij}) = M_i$, $i \neq j \in \{1, 2\}$.
- (AS4) $M_1 = M_2 = M$.

In fact, only the condition $Order(C_{ij}) \geq Order(A_{ij})$ is necessary if one wants the filters C_{ij} to be able to fit exactly the mixing filters A_{ij}. However, $(AS3)$ and $(AS4)$ are used to allow a simpler presentation of the results.

Notations:

- $C_{ij}(z) = \sum_{k=0}^{M} c_{ij}(k) z^{-k}, \quad i \neq j \in \{1, 2\}$
- $H(z) = 1 - C_{12}(z)C_{21}(z) = \sum_{k=0}^{2M} h(k) z^{-k}$
- $W(z) = \frac{1}{H(z)} = \sum_{k \geq 0} w(k) z^{-k}$

- $W_{eq}(z) = \dfrac{1}{1 - A_{12}(z)A_{21}(z)} = \sum_{k \geq 0} w_{eq}(k) z^{-k}$.

These notations deserve the following comments. The considered (i.e. direct or recurrent) separation system should be realizable. This first requires $W(z)$ to be causal (see Appendix A for a more detailed explanation). This is reflected in the above notations. This also requires the separation system to be stable. This system should especially meet these two requirements at the state of interest, i.e. at the separating solution. This has two consequences. On the one hand, the value of $W(z)$ at the separating solution, i.e. $W_{eq}(z)$, should be causal. This is also reflected in the above notations. On the other hand, the transfer function of the separation system at the separating solution is equal to the inverse of the mixing matrix. This inverse of the mixing matrix is thus required to be causal and stable. Moreover, the mixing matrix itself was already constrained above to be causal, stable and invertible. Combining all these requirements shows that the mixing matrix should be minimum phased [29] in the proposed approach.

2.3 Classical separation criteria

One way to assign the filters C_{ij} is to adapt their coefficients $(c_{ij}(k))_{k \geq 0}$ so that the outputs of the source separation structure become statistically independent. Theoretically, the statistical independence of two signals is achieved if and only if (iff) one of the following equivalent requirements is fulfilled:

- the joint probability density function of those signals is equal to the product of the marginal probability density functions,
- all the cross-cumulants of the signals are equal to zero,
- all the cross-spectra [28] of the signals are equal to zero,
- $E[x_1^p(n) x_2^q(n)] = E[x_1^p(n)] E[x_2^q(n)], \ \forall \ p, q \geq 0$,
- $E[f(x_1(n)) g(x_2(n))] = E[f(x_1(n))] E[g(x_2(n))]$, for any couple of functions f and g that ensures the existence of the above-mentioned mathematical expectations.

It is clear that the above requirements cannot be checked experimentally since it would require an infinite number of tests. Hence, all reported source separation algorithms use a limited number of necessary independence conditions with the hope that they are also sufficient. The intuitive motivation of this method is that only a finite set of equations is needed to determine a finite set of unknown parameters (the filter coefficients). The criteria based on this principle which have been reported may be summarized as follows.
Frequency-based criteria using the cross-spectra of the outputs of the separating structure were proposed by Weinstein et al. [35] and Yellin et al. [37]. Their approaches lead to recurrent algorithms that can be converted into stochastic adaptive algorithms based either on output decorrelation [35] or output cross-cumulant cancellation (some hints are provided in [37]). In the time domain, decorrelation-based algorithms have been presented by Al-Kindi et al. [1] and Van Gerven et al. [33]. Nguyen et al. [24]-[26] proposed two algorithms, based on *i)* the minimization of squared fourth-order cross-cumulants, and on *ii)* fourth-order cross-moment cancellation. Nguyen et al. also proposed an extension of the Hérault-Jutten rule to the convolutive domain that reads:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu_n f(s_i(n)) g(s_j(n-k)) \quad i \neq j \in \{1,2\}, k \in [0, M] \tag{7}$$

where $c_{ij}(n,k)$ is the k^{th} coefficient of filter C_{ij} at the n^{th} iteration, μ_n is a "small" positive adaptation gain, $f()$ and $g()$ are odd nonlinear functions. In addition s_i and s_j in (7) are the output signals of the considered recurrent structure (see Fig. 3). They are computed according to the following formula, which is derived in Appendix A:

$$s_i(n) = \frac{\left(y_i(n) - \sum_{k=1}^{M} c_{ij}(n,k)s_j(n-k)\right) - c_{ij}(n,0)\left(y_j(n) - \sum_{k=1}^{M} c_{ji}(n,k)s_i(n-k)\right)}{1 - c_{12}(n,0)c_{21}(n,0)}, \quad i \neq j \in \{1,2\}. \tag{8}$$

The Nguyen-Jutten fourth-order cross-moment cancellation algorithm is a specific case of (7), corresponding to $f(x) = x^3$ and $g(x) = x$, and this is in fact the only case that they experimentally studied.

2.4 Proposed separation criterion

As stated above, the rule (7) was proposed as an extension of the Hérault-Jutten algorithm to convolutive mixtures, but neither its convergence properties nor its asymptotic behaviour have been investigated up to now. This paper addresses this topic, and more generally speaking provides an analysis of the recurrent structure of Fig. 3 for a larger class of non-linear algorithms that reads:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu_n f_i(s_i(n)) g_i(s_j(n-k)) \qquad i \neq j \in \{1,2\}, k \in [0,M] \tag{9}$$

where $f_i()$ and $g_i()$ are arbitrary functions that are not necessarily odd.

The gain μ_n used for updating the filter coefficients may be set in different ways. Especially, two different classes of algorithms were defined in [4] depending on the properties of μ_n:

- *The asymptotically constant gain algorithm* corresponds to:

$$\left\{ \begin{array}{c} \mu_n \geq 0 \\ \mu = \lim_{n\to\infty} \mu_n > 0 \end{array} \right. \tag{10}$$

- *The decreasing gain algorithm* corresponds to:

$$\left\{ \begin{array}{l} \mu_n \geq 0 \\ \sum_n \mu_n^{\alpha} = \beta > 0 \quad \textit{for some } \alpha > 1 \\ \sum_n \mu_n = +\infty \end{array} \right. \tag{11}$$

Note that the classical choice $\mu_n = \mu, \forall n \geq 0$ belongs to (10), whereas $\mu_n = \frac{1}{n+1}, \forall n \geq 0$ belongs to (11). The results presented hereafter are valid for both algorithms. More detailed properties can be derived for the decreasing gain approach. However, we focus on the asymptotically constant gain algorithm because it allows to handle non-stationary systems and especially slowly varying ones [4].

3 Equilibrium states

The algorithm (9) can be formulated in vector form as:

$$\theta_{n+1} = \theta_n + \mu_n H(\theta_n, \xi_{n+1}), \tag{12}$$

where θ_n, ξ_{n+1} and $H(\theta_n, \xi_{n+1})$ are column vectors defined as:

$$\theta_n = [c_{12}(n,0), \ldots, c_{12}(n,M), c_{21}(n,0), \ldots, c_{21}(n,M)]^T, \tag{13}$$

$$\xi_{n+1} = [y_1(n), y_2(n), s_1(n-1), \ldots, s_1(n-M), s_2(n-1), \ldots, s_2(n-M)]^T, \tag{14}$$

$$\begin{aligned} H(\theta_n, \xi_{n+1}) \;=\; & [f_1(s_1(n))g_1(s_2(n)), \ldots, f_1(s_1(n))g_1(s_2(n-M)), \\ & f_2(s_2(n))g_2(s_1(n)), \ldots, f_2(s_2(n))g_2(s_1(n-M))]^T. \end{aligned} \tag{15}$$

The equilibrium points of (12) are defined as the vectors θ^* for which

$$E_{\theta^*}[H(\theta^*, \xi_{n+1})] = 0, \tag{16}$$

where $E_{\theta^*}[]$ denotes the mathematical expectation associated to the asymptotic probability law of the vector ξ_{n+1} for a given vector θ^*. Using (15), (16) becomes equivalent to[1]:

$$E[f_i(s_i(n))g_i(s_j(n-k))] = 0 \quad k \in [0, M],\; i \neq j \in \{1, 2\}. \tag{17}$$

The solutions of (17) depend on the separating functions f_i and g_i. Hence, looking for the exact location of all the equilibrium states requires to fix explicitly f_i and g_i. This is not within the scope of this paper, since our aim is to study the properties of algorithm (12) for the largest possible class of separating functions. Therefore, the following approach is used in this paper. We do not fix the separating functions f_i and g_i and therefore we do not investigate all equilibrium points. Instead, we only focus on the properties of the *separating point* θ^s corresponding to the separating solution, i.e. (see (5) and (13)):

$$\theta^s = [a_{12}(0), \ldots, a_{12}(M), a_{21}(0), \ldots, a_{21}(M)]^T, \tag{18}$$

and we especially determine at which condition this state is an equilibrium point of (12). When the separating state is reached, each output $s_i(n)$ for $i \in \{1, 2\}$ is equal to the source $x_i(n)$, due to (5). The equilibrium condition (17) then becomes:

$$E[f_i(x_i(n))g_i(x_j(n-k))] = 0, \quad k \in [0, M],\; i \neq j \in \{1, 2\}. \tag{19}$$

Using the mutual statistical independence of the sources $x_1(n)$ and $x_2(n)$, (19) is equivalent to:

$$E[f_i(x_i)]E[g_i(x_j)] = 0, \quad i \neq j \in \{1, 2\}. \tag{20}$$

In other words, by requiring the separating point to be an equilibrium state of the algorithm, we introduce restrictions on the functions f_i and g_i, i.e. they should be chosen so as to meet (20). Many such choices can then be made depending on the available information on the statistical properties of the sources. In order not to constrain both functions f_i and g_i, only assumptions on g_i are made hereafter, i.e:

$$E[g_i(x_j)] = 0, \quad i \neq j \in \{1, 2\}. \tag{21}$$

Additional conditions on the separating functions will result from the other constraints set on the considered algorithm in the subsequent sections.

[1]For readability, the subscript θ^* is omitted in the mathematical expectations $E[]$ below.

4 Stability Analysis

4.1 Stability condition for any equilibrium state

Each equilibrium state θ^* of (12) may be stable or not, depending on the properties of the functional H and on the statistics of the vectors $(\xi_n)_{n\geq 0}$. The approach used in this paper to analyze stability is the so-called Ordinary Differential Equation technique (ODE) [4], which approximates the discrete recurrence (12), under some conditions[2] on μ_n and H, by a continuous differential system that reads:

$$\frac{d\theta}{dt} = \lim_{n\to+\infty} E_\theta[H(\theta, \xi_{n+1})]. \tag{22}$$

The differential system (22) is locally stable in the vicinity of an equilibrium point θ^* iff the associated tangent linear system:

$$\frac{d\theta}{dt} = J(\theta^*)(\theta - \theta^*) \tag{23}$$

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

$$J_{ij}(\theta) = \lim_{n\to+\infty} \frac{\partial(E_\theta[H(\theta, \xi_{n+1})]^{(i)})}{\partial\theta^{(j)}}, \tag{24}$$

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 vector θ. An explicit formulation of $J_{ij}(\theta)$ at any fixed point is given in Appendix B.

4.2 Stability analysis at the separating point

In this sub-section, we apply the general results of Sub-section 4.1 to a specific state, i.e. the separating point (18) (with g_i chosen so that this point is an equilibrium state, as explained in Section 3). This analysis is first carried out for the general type of mixture defined above, i.e. causal convolutive mixtures. Then two specific types of mixtures are considered (namely strictly causal convolutive mixtures and instantaneous mixtures). The associated stability conditions are derived and analyzed.

4.2.1 Analysis for causal convolutive mixtures

The stability analysis requires the computation of the Jacobian matrix $J(\theta^s)$. Its expression can be simplified if we assume that:

$(AS5)$ *$x_1(n)$ and $x_2(n)$ are independent identically distributed (i.i.d) random sequences.*

This assumption is made in all this paper, while an extension to the case of coloured signals is provided in a companion paper [6]. Under $(AS5)$, the Jacobian matrix (24) can be written as a 2 by 2 block matrix (see Appendix B):

$$J(\theta^s) = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \tag{25}$$

where

[2] μ_n should be sufficiently small and it should meet e.g. (10) or (11). The conditions on H concern generally its regularity. They are globally met when using the assumptions and the models defined in this paper.

$$G_{ii} = \begin{pmatrix} -\alpha_i w_{eq}(0) + \varphi_i w_{eq}(0) & 0 & \dots & 0 \\ -\alpha_i w_{eq}(1) + \varphi_i w_{eq}(0) & -\alpha_i w_{eq}(0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha_i w_{eq}(M) + \varphi_i w_{eq}(0) & -\alpha_i w_{eq}(M-1) & \dots & -\alpha_i w_{eq}(0) \end{pmatrix}, \quad (26)$$

$$G_{ij} = \begin{pmatrix} -\beta_i w_{eq}(0) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \ i \neq j \in \{1,2\}, \quad (27)$$

and

$$\begin{cases} \alpha_i &= E[f_i'(x_i)]E[x_j g_i(x_j)], \quad \textit{where } j \textit{ is chosen so that } i \neq j \in \{1,2\} \\ \beta_i &= E[x_i f_i(x_i)]E[g_i'(x_j)], \quad \textit{where } j \textit{ is chosen so that } i \neq j \in \{1,2\} \\ \varphi_i &= a_{ji}(0)E[f_i(x_i)]E[x_j g_i'(x_j)], \quad \textit{where } j \textit{ is chosen so that } i \neq j \in \{1,2\} \\ w_{eq}(0) &= \dfrac{1}{1 - a_{12}(0)a_{21}(0)} \end{cases} \quad (28)$$

The eigenvalues of $J(\theta^s)$ are the roots of the associated characteristic polynomial, i.e. $P(\lambda) = det(J(\theta^s) - \lambda I)$. A compact expression of $P(\lambda)$ can be easily derived:

$$P(\lambda) = (\alpha_1 w_{eq}(0)+\lambda)^M(\alpha_2 w_{eq}(0)+\lambda)^M[\lambda^2 + w_{eq}(0)(\alpha_1+\alpha_2-\varphi_1-\varphi_2)\lambda + ((\alpha_1-\varphi_1)(\alpha_2-\varphi_2)-\beta_1\beta_2)w_{eq}^2(0)]. \quad (29)$$

Hence, the eigenvalues of $J(\theta^s)$ are[3]

$$\begin{array}{lll} 1) & -\alpha_1 w_{eq}(0) & (30) \\ 2) & -\alpha_2 w_{eq}(0) & (31) \\ 3) & \left[-(\alpha_1 - \varphi_1 + \alpha_2 - \varphi_2) \pm \sqrt{\Delta}\right] \frac{w_{eq}(0)}{2} \quad if\ \Delta > 0 & (32) \\ & \left[-(\alpha_1 - \varphi_1 + \alpha_2 - \varphi_2) \pm i\sqrt{-\Delta}\right] \frac{w_{eq}(0)}{2} \quad if\ \Delta < 0 & (33) \\ & -(\alpha_1 - \varphi_1 + \alpha_2 - \varphi_2) w_{eq}(0)/2 \quad if\ \Delta = 0 & (34) \end{array}$$

where

$$\Delta = [(\alpha_1 - \varphi_1) - (\alpha_2 - \varphi_2)]^2 + 4\beta_1\beta_2. \quad (35)$$

The stability condition (i.e. all eigenvalues having negative real parts) therefore reads as follows depending on the sign of Δ:

- if $\Delta > 0$

$$\begin{cases} \alpha_1 w_{eq}(0) > 0 \\ \alpha_2 w_{eq}(0) > 0 \\ (\alpha_1 - \varphi_1)(\alpha_2 - \varphi_2) > \beta_1\beta_2 \\ (\alpha_1 + \alpha_2) w_{eq}(0) > (\varphi_1 + \varphi_2) w_{eq}(0) \end{cases} \quad (36)$$

[3] In (33), i is such that $i^2 = -1$.

- if $\Delta \leq 0$

$$\left\{\begin{array}{c} \alpha_1 w_{eq}(0) > 0 \\ \alpha_2 w_{eq}(0) > 0 \\ (\alpha_1 + \alpha_2) w_{eq}(0) > (\varphi_1 + \varphi_2) w_{eq}(0) \end{array}\right. \tag{37}$$

4.2.2 Analysis for strictly causal convolutive mixtures

Here, we assume that the coupling mixture filters A_{ij} are strictly causal , i.e. that the first coefficients $a_{12}(0)$ and $a_{21}(0)$ are both null. The motivation for studying this case is that this condition is generally met in acoustic applications due to non-zero propagation time and it yields also a major simplification of the above results. More precisely, one takes advantage of this assumption in the separating structure by fixing $c_{12}(0) = c_{21}(0) = 0$. The first and $(M+2)^{th}$ components of θ_n and $H(\theta_n, \xi_{n+1})$ (see (13) and (15)) then disappear. The corresponding rows and columns of the Jacobian matrix $J(\theta^s)$ at the separating point (25) also disappear, and $J(\theta^s)$ becomes a $2M$ by $2M$ matrix defined as:

$$J(\theta^s) = \begin{pmatrix} -\alpha_1 G & 0_M \\ 0_M & -\alpha_2 G \end{pmatrix} \tag{38}$$

where 0_M is the M by M null matrix and G is defined as follows (taking into account that (28) here yields $w_{eq}(0) = 1$):

$$G = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ w_{eq}(1) & 1 & 0 & \dots & 0 \\ w_{eq}(2) & w_{eq}(1) & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w_{eq}(M-1) & w_{eq}(M-2) & w_{eq}(M-3) & \dots & 1 \end{pmatrix}. \tag{39}$$

It follows that the eigenvalues of $J(\theta^s)$ are $-\alpha_1$ and $-\alpha_2$, and so the stability condition is

$$\left\{\begin{array}{l} \alpha_1 > 0 \\ \alpha_2 > 0 \end{array}\right. \tag{40}$$

Using the notations of (28), the condition (40) can be rewritten as:

$$\left\{\begin{array}{l} E[f_1^{'}(x_1)]E[x_2 g_1(x_2)] > 0 \\ E[f_2^{'}(x_2)]E[x_1 g_2(x_1)] > 0 \end{array}\right. \tag{41}$$

Hence for strictly causal mixtures, the stability condition at the separating state does not depend on the properties of the mixing matrix (except that the mixing matrix is assumed to be invertible). It should be noted that (41) is met by any couple of odd functions (f_i, g_i) whatever the sources.

4.2.3 Analysis for instantaneous mixtures

The assumptions made above on the mixing and the separating structures remain valid in the case of instantaneous mixtures which only corresponds to setting $M = 0$ in the analysis presented in Sub-section 4.2.1 for causal convolutive mixtures. The Jacobian matrix (25) at the separating state then becomes:

$$J(\theta^s) = \begin{pmatrix} -(\alpha_1 - \varphi_1) w_{eq}(0) & -\beta_1 w_{eq}(0) \\ -\beta_2 w_{eq}(0) & -(\alpha_2 - \varphi_2) w_{eq}(0) \end{pmatrix}. \tag{42}$$

The characteristic polynomial (29) is then:

$$P(\lambda) = \lambda^2 + w_{eq}(0)(\alpha_1 + \alpha_2 - \varphi_1 - \varphi_2)\lambda + [(\alpha_1 - \varphi_1)(\alpha_2 - \varphi_2) - \beta_1\beta_2]\, w_{eq}^2(0). \quad (43)$$

Hence, the eigenvalues of $J(\theta^s)$ are those defined in (32)-(34). The stability condition becomes:

- if $\Delta > 0$

$$\begin{cases} (\alpha_1 - \varphi_1)(\alpha_2 - \varphi_2) > \beta_1\beta_2 \\ (\alpha_1 + \alpha_2)w_{eq}(0) > (\varphi_1 + \varphi_2)w_{eq}(0) \end{cases} \quad (44)$$

- if $\Delta \leq 0$

$$(\alpha_1 + \alpha_2)w_{eq}(0) > (\varphi_1 + \varphi_2)w_{eq}(0). \quad (45)$$

Note that the stability condition (36) or (37) in the convolutive case is a superset of the condition corresponding to instantaneous mixtures. Those conditions become equivalent for some particular source statistics and separating functions. This especially includes the case when the sources have even probability density functions and the following separating functions are used[4]:

$$\begin{cases} f_i(x) = x^{2m+1}, & i \in \{1,2\}, m \geq 0 \\ g_i(x) = x^{2n+1}, & i \in \{1,2\}, n \geq 0, n + m \neq 0 \end{cases} \quad (46)$$

This specific case was already described in the literature and is discussed more in detail in the following sub-section.

4.3 Analysis of classical adaptation rules

The generic approach presented for the separating point in Sub-section 4.2 especially applies to several specific adaptation rules which have been proposed in the literature. This allows to derive stability conditions for these algorithms, which have almost not been reported up to now. This method is applied to three classical algorithms hereafter.

4.3.1 Analysis of two classical algorithms for convolutive mixtures

Here, we analyze the stability of two classical algorithms, respectively based on output decorrelation and on fourth-order (3,1) cross-moment cancellation.

- **Analysis of the decorrelation algorithm**
  This algorithm can be seen as an extension of the classical Adaptive Noise Canceller adaptation strategy [36] to the Symmetric Adaptive Decorrelation scheme [1],[33],[34]. In the case of causal mixing filters, it corresponds to the following adaptation rule:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu_n s_i(n) s_j(n-k) \quad i \neq j \in \{1,2\}, k \in [0,M]. \quad (47)$$

  This adaptation rule is a particular case of the general framework defined in this paper (see (9)). It corresponds to the following choice for the separating functions:

$$\begin{cases} f_i(x) = x & i \in \{1,2\} \\ g_i(x) = x & i \in \{1,2\} \end{cases} \quad (48)$$

[4]In this case, (46) leads to $\alpha_i > 0, \beta_i > 0$ and $\varphi_i = 0$. The reader can easily check that the stability conditions in the intantaneous and convolutive cases are then the same.

Note that the assumption (21) made on the separating functions $g_i()$ is valid here, since it corresponds to the zero-mean hypothesis made on the sources, i.e:

$$E[x_i] = 0, \quad i \in \{1, 2\}. \tag{49}$$

Only some aspects of the convergence properties of (47) were investigated in the literature [33],[34]. Hereafter, we apply the generic results obtained in Sub-section 4.2 to this rule. The eigenvalues of the Jacobian matrix at the separating state are derived by applying (30)-(34) to the functions defined in (48), thus yielding:

$$\left\{ \begin{array}{c} -E[x_1^2]w_{eq}(0) \\ -E[x_2^2]w_{eq}(0) \\ -(E[x_1^2] + E[x_2^2])w_{eq}(0) \\ 0 \end{array} \right. \tag{50}$$

Since the eigenvalues are real, the stability condition corresponds to their negativeness which requires that:

$$w_{eq}(0) > 0 \tag{51}$$

or equivalently:

$$1 - a_{12}(0)a_{21}(0) > 0. \tag{52}$$

Nevertheless, the algorithm always yields a null eigenvalue. This implies that the algorithm is not asymptotically stable, but only globally stable with fluctuations. This means also that there exists a one-dimensional subspace, the Kernel of $J(\theta^s)$, associated to eigenvectors corresponding to the null eigenvalue, in which asymptotic convergence cannot be reached. In fact, from a computational point of view, the estimation associated to the null eigenvalue can take small but non null values that may be positive, leading then to instability.
In the case of strictly causal filters, the eigenvalues of the Jacobian matrix are $-E[x_1^2]$ and $-E[x_2^2]$ that are always negative. Hence, the decorrelation scheme yields an asymptotically stable separating state in this case and becomes then a potentially attractive separation procedure.

- **Analysis of the algorithm based on (3,1) cross-moment cancellation**
  We have shown above that the decorrelation criterion leads generally to a numerically unstable algorithm in the general case of causal filters. The cause of this problem is that the first-order lag coefficients $c_{12}(n, 0)$ and $c_{21}(n, 0)$ are updated by the same correcting term, i.e. $\mu_n s_1(n)s_2(n)$. There are different strategies to overcome this problem. A well-known one consists in using the algorithm (7) with at least one non-linear separating function $f()$ or $g()$. A classical algorithm of this family corresponds to the adaptation rule:

$$c_{ij}(n+1, k) = c_{ij}(n, k) + \mu_n s_i^3(n)s_j(n-k) \quad i \neq j \in \{1, 2\}, k \in [0, M]. \tag{53}$$

This algorithm is also a particular case of (9), obtained for:

$$\left\{ \begin{array}{ll} f_i(x) = x^3 & i \in \{1, 2\} \\ g_i(x) = x & i \in \{1, 2\} \end{array} \right. \tag{54}$$

(53) was proposed and experimentally studied by Nguyen et al. [24]-[26], but its stability analysis was not performed. We investigate this aspect in the following, by

applying (30)-(34) to the algorithm defined in (53). The eigenvalues of the Jacobian matrix at the separating state become:

$$\left\{ \begin{array}{c} -3E[x_1^2]E[x_2^2]w_{eq}(0) \\ -\left(3E[x_1^2]E[x_2^2] \pm \sqrt{E[x_1^4]E[x_2^4]}\right) w_{eq}(0) \end{array} \right. \tag{55}$$

and the stability condition is then:

$$\left\{ \begin{array}{c} w_{eq}(0) > 0 \\ E[x_1^4]E[x_2^4] < 9E^2[x_1^2]E^2[x_2^2] \end{array} \right. \tag{56}$$

The second condition in (56) implies that the sources must be globally sub-gaussian. It should be noted that (56) is the same stability condition as with the version of this algorithm for instantaneous mixtures [32].

4.3.2 Analysis of a classical algorithm for instantaneous mixtures

Here, we consider the Hérault-Jutten algorithm for intantaneous mixtures, i.e:

$$c_{ij}(n+1) = c_{ij}(n) + \mu_n f(s_i(n))g(s_j(n)) \quad i \neq j \in \{1,2\}. \tag{57}$$

The corresponding stability condition was first studied by E. Sorouchyari [32] for the limited class of separating functions defined by (46) and symmetrically distributed sources. Several later contributions [10],[11],[19],[22],[23] reached globally the same results. This paper extends these results by providing a stability condition at the separating point for possibly asymmetrically distributed sources and any separating functions f and g: this is a straightforward application of the results of Sub-section 4.2.3 to the case when $f_1 = f_2 = f$ and $g_1 = g_2 = g$.

5 Asymptotic behaviour analysis and optimization of the separating functions

This section is devoted to the study of the asymptotic behaviour of the algorithm (12), i.e. once the transient phase ended. The algorithm properties are shown to be closely related to the choice of the separating functions. Our goal, in this paper, is to determine the separating functions $f_i(x)$ and $g_i(x)$ that ensure the best matching of the mixing filters in the mean square sense, i.e. that minimize $E[|\theta_n - \theta^s|^2]$ for large n.
In this section, the mixing and the separating coupling filters are assumed to be strictly causal, based on the motivation presented in Sub-section 4.2.2, i.e. we assume that:

(AS6) $a_{12}(0) = a_{21}(0) = 0$ *and* $c_{12}(0) = c_{21}(0) = 0$.

Furthermore, we focus on the particular case of the *constant gain algorithm*, corresponding to the scalar gain condition $\mu_n = \mu,\ \forall n \geq 0$. (12) then becomes:

$$\theta_{n+1} = \theta_n + \mu H(\theta_n, \xi_{n+1}). \tag{58}$$

It is clear that all the results presented above (equilibrium conditions, stability conditions ...) especially apply to this specific version of the algorithm.

5.1 Some theoretical results

In Sub-section 4.1, we derived stability conditions for any equilibrium state θ^* which are assumed to be met hereafter. This was done by approximating the stochastic algorithm (12) by the continuous differential equation (22). θ^* was then a limit, when time tends to infinity, of the solution of the ODE (22). The asymptotic behaviour studied here refers to how far the stochastic algorithm varies from its ODE. This corresponds to the fluctuations of θ_n around the equilibrium state θ^* after a large number n of iterations, assuming that θ_n remains in the attraction domain of θ^*. To investigate this aspect, we use a theorem established by Benveniste et al. [4], but before this we recall some theoretical results.

Definition 1 *A given matrix K is said to be stable iff all its eigenvalues have negative real parts.*

Theorem 1 ([17]) *For a given real stable matrix A and a positive definite real matrix C, there exists a unique symmetric and positive definite solution B to the so-called Lyapunov equation*

$$A^T B + BA + C = 0. \tag{59}$$

Theorem 2 (adapted from Benveniste et al. [4]) *Let $\theta(t)$ be the solution of the ODE associated to (58). The vector θ_n, in (58), is then an approximation of $\theta(t_n)$ for $t_n = \sum_{k=1}^{n} \mu_k = n\mu$. Besides, let the normalized error variable be:*

$$Z_\mu(t_n) = \frac{\theta_n - \theta(t_n)}{\sqrt{\mu}} \tag{60}$$

and let $Z_\mu(t)$ be the continuous trajectory obtained by a linear interpolation of $Z_\mu(t_n)$. Then, for $\mu \longrightarrow 0$ and $t \longrightarrow \infty$, we have:

$$Z_\mu(t) \longrightarrow N(0, P), \tag{61}$$

where (61) corresponds to the convergence in law and N(0,P) denotes the Gaussian probability density function with zero mean and covariance matrix P. The matrix P is the unique symmetric and positive definite solution of the Lyapunov equation:

$$J(\theta^*)P + PJ^T(\theta^*) + R(\theta^*) = 0 \tag{62}$$

where $J(\theta^*)$ is the Jacobian matrix (24) at θ^* and $R(\theta^*)$ is defined by:

$$R(\theta^*) = \sum_{n \in \mathcal{Z}} Cov[H(\theta^*, \xi_{n+1}), H(\theta^*, \xi_0)] \tag{63}$$

where Cov[] denotes the covariance matrix.

This means that for a sufficiently small stepsize μ, θ_n is an asymptotically unbiased estimator of θ^* with a covariance matrix μP. Furthermore, we have:

$$\lim_{n \to +\infty} E[|\theta_n - \theta^*|^2] = \lim_{n \to +\infty} \sum_{i \in [1,M] \cup [M+2,2M]} E[(\theta_n^{(i)} - \theta^{*(i)})^2] = \mu\, Tr(P), \tag{64}$$

where $\theta_n^{(i)}$ and $\theta^{*(i)}$ are respectively the i^{th} order entries of θ_n and θ^*. Hence, the asymptotic error variance of the estimation of θ^* can be written as:

$$\sigma_\infty = \mu\, Tr(P). \tag{65}$$

5.2 Application to source separation

In the following, we consider the source separation algorithm (58) in its scalar form that will be called algorithm (N0) and denoted, from here on, by:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu f_i(s_i(n)) g_i(s_j(n-k)) \qquad i \neq j \in \{1,2\}, k \in [1,M]. \tag{66}$$

Moreover, we only consider the equilibrium state θ^s. As shown in Appendix C, the corresponding matrix $R(\theta^s)$ can be written as:

$$R(\theta^s) = \begin{pmatrix} \gamma_1 R_1 & 0 \\ 0 & \gamma_2 R_2 \end{pmatrix}, \tag{67}$$

where

$$R_i = \begin{pmatrix} 1 & a_i & \dots & a_i \\ a_i & 1 & \ddots & a_i \\ \vdots & \ddots & \ddots & a_i \\ a_i & \dots & a_i & 1 \end{pmatrix}, \tag{68}$$

$$a_i = \frac{E^2[f_i(x_i)]}{E[f_i^2(x_i)]}, \tag{69}$$

$$\gamma_i = E[f_i^2(x_i)] E[g_i^2(x_j)], \quad \textit{where } j \textit{ is chosen so that } i \neq j \in \{1,2\}. \tag{70}$$

Besides, the Jacobian matrix associated to (58) at the equilibrium state θ^s is defined in (38). The Lyapunov equation (62) associated to the algorithm (N0) becomes then:

$$\begin{pmatrix} -\alpha_1 G & 0 \\ 0 & -\alpha_2 G \end{pmatrix} P + P \begin{pmatrix} -\alpha_1 G^T & 0 \\ 0 & -\alpha_2 G^T \end{pmatrix} + \begin{pmatrix} \gamma_1 R_1 & 0 \\ 0 & \gamma_2 R_2 \end{pmatrix} = 0_{2M} \tag{71}$$

where 0_{2M} is the $2M$ by $2M$ null matrix.

The unique symmetric and positive definite solution of (71) can be determined as follows: let's consider the equation below, where matrix B is the unknown parameter:

$$-GB - BG^T + R_i = 0_M \tag{72}$$

where 0_M is the M by M null matrix. The matrix R_i is positive definite since its eigenvalues are $1 + (M-1)a_i$ and $1 - a_i$, and these values are strictly positive because[5] $0 \leq a_i < 1$. The real matrix $-G$ is stable since all its eigenvalues are equal to -1. Hence, (72) is a Lyapunov equation meeting the requirements of Theorem 1. Its unique symmetric and positive definite solution will be denoted by K_i.

The two equations (72) for $i \in \{1,2\}$ and their solutions $(K_i)_{i\in\{1,2\}}$ can be recombined in the following block matrix form:

$$\begin{pmatrix} -\alpha_1 G & 0 \\ 0 & -\alpha_2 G \end{pmatrix} \begin{pmatrix} \frac{\gamma_1}{\alpha_1} K_1 & 0 \\ 0 & \frac{\gamma_2}{\alpha_2} K_2 \end{pmatrix} + \begin{pmatrix} \frac{\gamma_1}{\alpha_1} K_1 & 0 \\ 0 & \frac{\gamma_2}{\alpha_2} K_2 \end{pmatrix} \begin{pmatrix} -\alpha_1 G^T & 0 \\ 0 & -\alpha_2 G^T \end{pmatrix} + \begin{pmatrix} \gamma_1 R_1 & 0 \\ 0 & \gamma_2 R_2 \end{pmatrix} = 0_{2M}. \tag{73}$$

[5] $0 \leq a_i \leq 1$ due to (69). Moreover, $a_i = 1$ corresponds to a constant function f_i; however such a function does not meet the stability requirements (40), so $a_i = 1$ cannot occur here.

Comparing (71) and (73) shows that:

$$K = \begin{pmatrix} \frac{\gamma_1}{\alpha_1}K_1 & 0 \\ 0 & \frac{\gamma_2}{\alpha_2}K_2 \end{pmatrix} \tag{74}$$

is a solution of (71). Moreover, K is symmetric and positive definite (because K_i is symmetric and positive definite and $\frac{\gamma_i}{\alpha_i} > 0$ for $i \in \{1,2\}$ due to (40) and (70)). Therefore K is the unique solution of (71) that was to be determined. The asymptotic error variance (65) is then equal to:

$$\sigma_\infty = \mu Tr(K) = \mu \sum_{i,j=1 \ i \neq j}^{i,j=2} \frac{E[f_i^2(x_i)]}{E[f_i'(x_i)]} \frac{E[g_i^2(x_j)]}{E[x_j g_i(x_j)]} Tr(K_i) \tag{75}$$

Besides, Appendix D shows that:

$$Tr(K_i) = q_{i2} + q_{i1}a_i, \ \ i \in \{1,2\}, \tag{76}$$

where q_{ij} for $i,j \in \{1,2\}$ are real values that depend only on the mixing matrix and a_i is defined in (69). Hence the asymptotic error variance (75) becomes:

$$\sigma_\infty = \mu \sum_{i,j=1 \ i \neq j}^{i,j=2} \frac{E[f_i^2(x_i)]}{E[f_i'(x_i)]} \frac{E[g_i^2(x_j)]}{E[x_j g_i(x_j)]} (q_{i2} + q_{i1}a_i). \tag{77}$$

The right term of (77) depends on four types of parameters:

- μ that can be set so as to achieve a tradeoff between the asymptotic error variance and the convergence speed,
- the separating functions f_i and g_i,
- the source statistics,
- q_{ij} that are related to the mixing matrix.

For a given adaptation gain μ and a given set of observations (source statistics and mixing channel properties), only the separating functions can be optimized in order to reduce the asymptotic error variance. The aim is then to choose the separating functions f_i and g_i that minimize σ_∞ as defined in (77). Appendix D shows that $q_{i2}+q_{i1}a_i > 0$ for $i \in \{1,2\}$. Therefore, when the stability condition (41) is met, the two terms $\frac{E[f_i^2(x_i)]}{E[f_i'(x_i)]}(q_{i2} + q_{i1}a_i)\frac{E[g_i^2(x_j)]}{E[x_j g_i(x_j)]}$ of σ_∞ respectively corresponding to $(i,j) = (1,2)$ and $(2,1)$ are positive and depend on different sets of separating functions, i.e. resp. (f_1, g_1) and (f_2, g_2). Therefore, the minimization of σ_∞ can be performed by minimizing these two terms separately. Similarly, each of these two positive terms can be minimized by minimizing separately its two independent contributions associated with f_i and g_i: $abs\left(\frac{E[f_i^2(x_i)]}{E[f_i'(x_i)]}(q_{i2} + q_{i1}a_i)\right)$ and $abs\left(\frac{E[g_i^2(x_j)]}{E[x_j g_i(x_j)]}\right)$. The minimization of each contribution is then equivalent to the minimization of its squared value, i.e. $\frac{E[f_i^2(x_i)]^2}{E[f_i'(x_i)]^2}(q_{i2} + q_{i1}a_i)^2$ and $\frac{E[g_i^2(x_j)]^2}{E[x_j g_i(x_j)]^2}$ that is a differentiable function of f_i or g_i. Note, however that this minimization problem suffers from a scaling ambiguity: by replacing the separating functions f_i and g_i by $\epsilon_{i1} f_i$ and $\epsilon_{i2} g_i$, where ϵ_{i1} and ϵ_{i2} are scaling factors, the quantities to be minimized become $\epsilon_{i1}^2 \frac{E[f_i^2(x_i)]^2}{E[f_i'(x_i)]^2}(q_{i2} + q_{i1}a_i)^2$ and $\epsilon_{i2}^2 \frac{E[g_i^2(x_j)]^2}{E[x_j g_i(x_j)]^2}$. They are then increasing functions of ϵ_{i1}^2 and ϵ_{i2}^2. Hence, decreasing $(\epsilon_{ij}^2)_{j \in \{1,2\}}$ enables to reduce the asymptotic error variance, but at the expense of lower convergence speed since ϵ_{i1} and

ϵ_{i2} appear as factors of μ in (66). Further constraints have then to be set on the separating functions to remove this scaling ambiguity. Such constraints are presented in the next section. Before this, only an extremum search is undertaken in the current section since it is insensitive to the scaling ambiguity mentioned above. The results obtained represent classes of functions that are determined up to constant scaling factors. This study of the separating functions that lead to extremum values of $\frac{E[f_i^2(x_i)]^2}{E[f_i'(x_i)]^2}(q_{i2}+q_{i1}a_i)^2$ and $\frac{E[g_i^2(x_j)]^2}{E[x_j g_i(x_j)]^2}$ is developed in Appendix E and yields:

$$f_{iext}(x) = -\nu_{i1}\frac{p'_{x_i}(x)}{p_{x_i}(x)}, \tag{78}$$

$$g_{iext}(x) = \nu_{i2}\, x, \tag{79}$$

where p_{x_i} is the probability density function of source x_i and (ν_{i1}, ν_{i2}) is a couple of arbitrary real constants. Moreover, it may be shown that (41) here yields $\nu_{i1}\nu_{i2} > 0$ if $p'_{x_i}(x) \longrightarrow 0$ when $x \longrightarrow \pm\infty$, which is most often the case. It should be noted that the same classes of functions were obtained for linear instantaneous mixtures of white sources based on a minimum likelihood approach [13],[30].

6 Two algorithms with normalized optimum separating functions

The analysis presented in Section 5 yields the extremum separating functions defined up to a scaling factor (ν_{i1} and ν_{i2}). One way of removing this ambiguity is to normalize these functions. Two alternative normalization methods are presented hereafter. Each of them yields a specific algorithm, which is analyzed in terms of stability and asymptotic behaviour.

6.1 Normalization scheme (N1)

One way of removing the scaling ambiguity mentioned above is to use scaling factors ϵ_{i1} and ϵ_{i2} such that:

$$E[\epsilon_{i1}^2 f_i^2(x)] = 1, \tag{80}$$

$$E[\epsilon_{i2}^2 g_i^2(x)] = 1, \tag{81}$$

which corresponds to[6]:

$$\epsilon_{i1} = \frac{1}{\sqrt{E[f_i^2(x)]}}, \tag{82}$$

$$\epsilon_{i2} = \frac{1}{\sqrt{E[g_i^2(x)]}}. \tag{83}$$

The associated algorithm (N1) is derived by replacing f_i and g_i respectively by $\epsilon_{i1} f_i$ and $\epsilon_{i2} g_i$ in (66), which yields for causal convolutive mixtures:

[6] The opposites of the values of ϵ_{i1} and ϵ_{i2} provided in (82) and (83) are also solutions of (80) and (81) but there is no use considering them: they would only allow to change the sign of the adaptation term in (84), but this degree of freedom is already available through the signs of f_i and g_i, which are free at this stage.

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu \frac{f_i(s_i(n))}{\sqrt{E[f_i^2(s_i)]}} \frac{g_i(s_j(n-k))}{\sqrt{E[g_i^2(s_j)]}} \quad i \neq j \in \{1,2\}, k \in [0,M] \tag{84}$$

or equivalently,

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu F_i(s_i(n)) G_i(s_j(n-k)) \quad i \neq j \in \{1,2\}, k \in [0,M] \tag{85}$$

where

$$\begin{aligned} F_i(x) &= \frac{f_i(x)}{\sqrt{E[f_i^2(x)]}}, \\ G_i(x) &= \frac{g_i(x)}{\sqrt{E[g_i^2(x)]}}. \end{aligned} \tag{86}$$

(84) can be seen as a zero-search procedure for the set of correlation coefficients between the random variables $f_i(s_i(n))$ and $g_i(s_j(n-k))$, instead of the classical zero-search procedure for the non-normalized correlation between these two variables used in (9). Moreover, the variance of the correcting term $\frac{f_i(s_i(n))}{\sqrt{E[f_i^2(s_i)]}} \frac{g_i(s_j(n-k))}{\sqrt{E[g_i^2(s_j)]}}$ is equal to one at the separating state. This value is independent from the scales and statistics of the sources and from the separating functions, so that the adaptation gain μ can be selected independently from these parameters, which is an attractive feature of this new algorithm (84).
However, this algorithm includes parameters which must be estimated in practical situations (i.e. $\sqrt{E[f_i^2(s_i)]}$ and $\sqrt{E[g_i^2(s_j)]}$), unlike the previous algorithm (66). Therefore, the results about stability and asymptotic behaviour presented in the previous sections cannot be applied directly to (84). However, (84) and the estimation of the associated energies ($\sqrt{E[f_i^2(s_i)]}, \sqrt{E[g_i^2(s_j)]}$) can be formulated as a relaxation scheme of the form (66) with a second-order perturbation term. This is a classical strategy, which is used for example in the estimation of the input covariance matrix in the Recursive Least Squares (RLS) algorithm [4]. It can also be shown that this normalization scheme does not modify fundamentally the results obtained in the previous sections. In fact, computations show that the stability condition and the asymptotic behaviour of this algorithm (84) can be deduced from the previous solutions by using the following transformation:

$$\begin{aligned} f_i(x) &\longrightarrow F_i(x), \\ g_i(x) &\longrightarrow G_i(x). \end{aligned} \tag{87}$$

The stability condition is then the same as (36) and (37) applied to F_i and G_i instead of f_i and g_i. Similarly, in the case of strictly causal filters, the stability condition (41) becomes explicitly:

$$E[F_1'(x_1)]E[x_2 G_1(x_2)] > 0, \tag{88}$$
$$E[F_2'(x_2)]E[x_1 G_2(x_1)] > 0, \tag{89}$$

or, by inserting (86) in the latter equations:

$$E[f_1'(x_1)]E[x_2 g_1(x_2)] > 0, \tag{90}$$
$$E[f_2'(x_2)]E[x_1 g_2(x_1)] > 0. \tag{91}$$

The asymptotic error variance is derived in the same way from (77), which yields:

$$\sigma_\infty = \mu \sum_{i,j=1 \;\; i \neq j}^{i,j=2} \frac{\sqrt{E[f_i^2(x_i)]}}{E[f_i'(x_i)]} \frac{\sqrt{E[g_i^2(x_j)]}}{E[x_j g_i(x_j)]}(q_{i2} + q_{i1} a_i). \tag{92}$$

Note that, in the transition from (77) to (92), a_i was unchanged since it is invariant up to a scaling factor applied to the function f_i. As in Sub-section 5.2, the optimum separating functions can be determined by minimizing independently $abs\left(\frac{\sqrt{E[f_i^2(x_i)]}}{E[f_i'(x_i)]}(q_{i2}+q_{i1}a_i)\right)$ and $abs\left(\frac{\sqrt{E[g_i^2(x_j)]}}{E[x_j g_i(x_j)]}\right)$ or their squared values for each couple (i, j). This leads to the optimum choice (see Appendix E):

$$f_{iopt}(x) = -\nu_{i1} \frac{p'_{x_i}(x)}{p_{x_i}(x)}, \tag{93}$$

$$g_{iopt}(x) = \nu_{i2}\, x. \tag{94}$$

where p_{x_i} is the probability density function of the source x_i and (ν_{i1}, ν_{i2}) is a couple of arbitrary real constants. Moreover, it may be shown that (90)-(91) here yield $\nu_{i1}\nu_{i2} > 0$ if $p'_{x_i}(x) \longrightarrow 0$ when $x \longrightarrow \pm\infty$.

The corresponding optimum functions F_i and G_i are then derived from (86), which yields[7]:

$$F_{iopt}(x) = -\frac{\frac{p'_{x_i}(x)}{p_{x_i}(x)}}{\sqrt{E[(\frac{p'_{x_i}(x)}{p_{x_i}(x)})^2]}}, \tag{95}$$

$$G_{iopt}(x) = \frac{x}{\sqrt{E[x^2]}}. \tag{96}$$

It should be noted that the scaling ambiguity indeed disappears in these functions.

6.2 Normalization scheme (N2)

A second strategy to remove the scaling ambiguity mentioned above is to use scaling factors ϵ_{i1} and ϵ_{i2} such that:

$$E[\epsilon_{i1} f_i'(x)] = \frac{1}{\sqrt{E[x^2]}}, \tag{97}$$

$$E[\epsilon_{i2}^2 g_i^2(x)] = 1, \tag{98}$$

which corresponds to[8]:

$$\epsilon_{i1} = \frac{1}{E[f_i'(x)]\sqrt{E[x^2]}}, \tag{99}$$

$$\epsilon_{i2} = \frac{1}{\sqrt{E[g_i^2(x)]}}. \tag{100}$$

[7]The solution obtained by changing the sign of both functions in (95) and (96) also exists (due to $\nu_{i1}\nu_{i2} > 0$), but it yields the same algorithm (85).

[8]Like with the algorithm (N1), there is no use considering the opposite value of ϵ_{i2}.

The associated algorithm (N2), for causal mixtures, can then be written as:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu \, \frac{f_i(s_i(n))}{E[f_i'(s_i)]\sqrt{E[s_i^2]}} \, \frac{g_i(s_j(n-k))}{\sqrt{E[g_i^2(s_j)]}} \quad i \neq j \in \{1,2\}, k \in [0,M], \tag{101}$$

which is equivalent to:

$$c_{ij}(n+1,k) = c_{ij}(n,k) + \mu F_i(s_i(n)) G_i(s_j(n-k)) \quad i \neq j \in \{1,2\}, k \in [0,M], \tag{102}$$

where

$$F_i(x) = \frac{f_i(x)}{E[f_i'(x)]\sqrt{E[x^2]}} \tag{103}$$

$$G_i(x) = \frac{g_i(x)}{\sqrt{E[g_i^2(x)]}}. \tag{104}$$

This normalization scheme is not a classical one and is based on the following principle. For strictly causal mixtures, when using the initial algorithm (N0) with extremum separating functions (78)-(79), the Jacobian matrix and therefore the convergence speed of this algorithm depend on f_{iext} and therefore on the source statistics. This is due to the expression of f_{iext} (while g_{iext} does not yield any restrictions). This is a drawback, because these source statistics are generally unknown. A solution to this problem consists in modifying the algorithm (N0) so that the Jacobian matrix does not depend any more on f_i. This is the target reached by the algorithm (N2) as shown below.
For causal convolutive mixtures, the stability condition and the asymptotic error variance can be deduced from (36) and (37) using the following tranformation:

$$\begin{array}{ccc} f_i(x) & \longrightarrow & F_i(x) \\ g_i(x) & \longrightarrow & G_i(x). \end{array} \tag{105}$$

Under the strict causality assumption, the stability condition derived from (41) reads:

$$E[x_i G_j(x_i)] > 0, \quad i \neq j \in \{1,2\} \tag{106}$$

or equivalently:

$$E[x_i g_j(x_i)] > 0, \quad i \neq j \in \{1,2\}. \tag{107}$$

It should be noted that this condition is independent from $f_i(x)$, as stated above. The asymptotic error variance associated to (N2) is:

$$\sigma_\infty = \mu \sum_{i,j=1 \; i \neq j=1}^{i,j=2} \frac{E[f_i^2(x_i)]}{E^2[f_i'(x_i)]} \frac{\sqrt{E[g_i^2(x_j)]}}{E[x_j g_i(x_j)]} \frac{1}{\sqrt{E[x_i^2]}} (q_{i2} + q_{i1} a_i). \tag{108}$$

The optimum separating functions correspond to the independent minimization of $\frac{E[f_i^2(x_i)]}{E^2[f_i'(x_i]}(q_{i2} + q_{i1} a_i), i \in \{1,2\}$ and $\frac{\sqrt{E[g_i^2(x_j)]}}{E[x_j g_i(x_j)]}, i \neq j \in \{1,2\}$. This leads also to the following optimum choice (see Appendix E):

$$f_{iopt}(x) = -\nu_{i1} \frac{p'_{x_i}(x)}{p_{x_i}(x)}, \tag{109}$$

$$g_{iopt}(x) = \nu_{i2} \, x, \tag{110}$$

where p_{x_i} is the probability density function of the source x_i and (ν_{i1}, ν_{i2}) is a couple of arbitrary real constants such that $\nu_{i2} > 0$ due to (107) and (110). The corresponding optimum functions F_i and G_i are then derived from (103), which yields:

$$F_{iopt}(x) = \frac{\frac{p'_{x_i}(x)}{p_{x_i}(x)}}{E[(\frac{p'_{x_i}(x)}{p_{x_i}(x)})']\sqrt{E[x^2]}}, \qquad (111)$$

$$G_{iopt}(x) = \frac{x}{\sqrt{E[x^2]}}, \qquad (112)$$

It should be noted that, like with the algorithm (N1), the scaling ambiguity disappears in these normalized functions.

7 Conclusions and prospects

This paper deals with the separation of two convolutively mixed white signals. The proposed approach is based on a recurrent separation structure adapted by a generic rule involving arbitrary separating functions. The following aspects of the convergence properties of this rule are analyzed.

A vector-based formulation of the adaptation scheme is first derived. We then determine conditions on the separating functions for the separating state to be a stable equilibrium state of the adaptation rule. This analysis especially applies to specific source separation algorithms for convolutive mixtures which have been proposed in the literature and experimentally studied, but for which almost no stability analyses have been reported up to now. It also allows to extend the theoretical results reported for the Hérault-Jutten algorithm for instantaneous mixtures.

The expression of the asymptotic error variance of the estimation of the mixture filters is then determined in the case of strictly causal mixtures. The separating functions which minimize the error variance are derived. They are shown to be only related to the probability density functions of the sources. This minimization of the error variance leads us to introduce two alternative normalization procedures. It should be noted that these procedures also allow to derive normalized versions of various adaptation rules with arbitrary separating functions (i.e. not necessarily optimum) which yield attractive features[9].

As stated above, the proposed approach is developed for white source signals and yields optimum functions which depend on the source statistics. For this approach to be applicable in real situations, it should be extended to the case of possibly-coloured sources with unknown statistics. This practical extension is presented in the second part of this paper [6]. The latter paper also describes the performance obtained with the basic and extended approaches in various situations ranging from synthetic data to audio signals measured in real conditions. This shows that the optimization of the separating functions yields a major improvement of the quality of the separation that can be achieved.

A Causality of $W(z)$ and $W_{eq}(z)$

In Subsection 2.2, we defined the filters $W(z)$ and $W_{eq}(z)$ and briefly showed that they should be causal in the proposed approach. In this appendix, we provide more detailed

[9]These normalization procedures are detailed for the case of strictly causal convolutive mixtures, which is the main topic of this paper, but they may be used in the same way for causal convolutive mixtures and instantaneous mixtures.

explanations and expressions about this property (some related information may also be found in [34]).

Let us first consider the direct structure shown in Fig. 2. This structure explicitly contains two blocks whose transfer functions are:

$$W(z) = \frac{1}{1 - C_{12}(z)C_{21}(z)}. \tag{113}$$

For this separation system to be realizable, these two blocks should themselves be realizable, and therefore causal. This condition may be expressed as follows in terms of the coefficients of the MA filters $C_{ij}(z)$. As these causal filters are both assumed to be of order M, $C_{12}(z)C_{21}(z)$ is a polynomial in z^{-1} of order $2M$. Moreover, its constant term is equal to[10]: $c_{12}(n,0)c_{21}(n,0)$. Therefore:

$$W(z) = \frac{1}{1 - c_{12}(n,0)c_{21}(n,0) + D(z)} \tag{114}$$

with:

$$D(z) = \sum_{k=1}^{2M} d(k) z^{-k}. \tag{115}$$

This yields two cases. If $1 - c_{12}(n,0)c_{21}(n,0) \neq 0$, a Taylor series expansion of (114) shows that $W(z)$ may be expressed as:

$$W(z) = \sum_{k \geq 0} w(k) z^{-k}. \tag{116}$$

$W(z)$ is then causal. If $1 - c_{12}(n,0)c_{21}(n,0) = 0$, let k_0, with $k_0 \geq 1$, be the lowest index of the non-zero terms $d(k)$. Then, (114) yields:

$$W(z) = z^{k_0} . \frac{1}{d(k_0)} . \frac{1}{1 + D'(z)} \tag{117}$$

with:

$$D'(z) = \sum_{k=1}^{2M-k_0} d'(k) z^{-k}. \tag{118}$$

Expanding the third term of (117) as a Taylor series again yields a causal term in (117), but then multiplying it by z^{k_0} results in a non-causal overall filter $W(z)$. As a conclusion, $W(z)$ is causal if and only if $1 - c_{12}(n,0)c_{21}(n,0) \neq 0$. It should be noted that this condition only concerns the instantaneous part (i.e. lag zero) of the filters $C_{ij}(z)$.

We now move to the recurrent structure shown in Fig. 3. The transfer function (113) and the associated causality condition do not appear explicitly in blocks of this structure, but they are implicitly present in its connections, and more precisely in its cross-couplings. They become explicit when considering the "realizability" of this structure, i.e. when investigating how to implement it in practice. This yields several approaches, as will now be shown. Let us first consider the Z-domain representation of this structure. This separation system then corresponds to the following equations:

[10] Here again, $c_{ij}(n,k)$ is the k^{th} coefficient of filter C_{ij} at the n^{th} iteration. Both indices are provided in the notations used for these filter coefficients in order to avoid any ambiguity. On the contrary, the index n is omitted in the notations for the time-dependent transfer functions $C_{ij}(z)$ of these filters, as they do not yield such an ambiguity.

$$\begin{cases} S_1(z) = Y_1(z) - C_{12}(z)S_2(z) \\ S_2(z) = Y_2(z) - C_{21}(z)S_1(z) \end{cases} \tag{119}$$

This may be considered as the theoretical definition of the recurrent structure, in the sense that it does not provide directly a practical means for computing the system outputs or their Z transforms $S_1(z)$ and $S_2(z)$. Such a means may be obtained by solving these equations in the Z domain, which yields (3). The separation system is then implemented as the cascade connection of the matrix which appears in (3) and of the filters defined by (113). The latter filters should then be realizable and therefore causal. The condition thus obtained is therefore the same as for the direct structure. This is normal, as the cascade connection that we thus introduced as a practical means for implementing the theoretical recurrent structure, based on the observation-output relationship (3), is in fact identical to the direct structure shown in Fig. 2.

Now consider the time-domain representation of the recurrent structure. It is defined by the time-domain counterpart of (119), i.e:

$$\begin{cases} s_1(n) = y_1(n) - \sum_{k=0}^{M} c_{12}(n,k)s_2(n-k) \\ s_2(n) = y_2(n) - \sum_{k=0}^{M} c_{21}(n,k)s_1(n-k) \end{cases} \tag{120}$$

These time-domain equations may be combined in the same way as in the above approach in the Z domain, which does not deserve any additional comments. However, they may also be handled in another way, i.e. by splitting the instantaneous terms (i.e. lag zero) of the convolution products in their right-hand terms, and grouping them with the left-hand terms of these equations, which yields:

$$\begin{cases} s_1(n) + c_{12}(n,0)s_2(n) = y_1(n) - \sum_{k=1}^{M} c_{12}(n,k)s_2(n-k) \\ c_{21}(n,0)s_1(n) + s_2(n) = y_2(n) - \sum_{k=1}^{M} c_{21}(n,k)s_1(n-k) \end{cases} \tag{121}$$

Solving this set of equations with $s_1(n)$ and $s_2(n)$ as unknowns provides a practical means for computing the latter quantities with respect to their previous values and current observations $y_1(n)$ and $y_2(n)$. The expressions thus obtained are provided in (8). This practical approach only holds if the set of equations (121) can be solved, i.e. if its determinant is non-zero, i.e. if $1 - c_{12}(n,0)c_{21}(n,0) \neq 0$. This approach therefore leads to the same condition as the previous ones. It should be noted that this condition is here directly obtained in terms of the instantaneous part of the filters $C_{ij}(z)$, because this part was handled in a specific way in this approach. This is to be contrasted with the previous approaches, where this condition was initially expressed in terms of the causality of the filter $W(z)$.

Up to this point, the analysis presented in this appendix was performed for any set of filters $C_{12}(z)$ and $C_{21}(z)$. It especially applies to the state of interest in this investigation, i.e. to the separating solution. At this state, $C_{12}(z) = A_{12}(z)$ and $C_{21}(z) = A_{21}(z)$. This analysis thus shows that in the proposed approach the mixing matrix is required to be such that

$$W_{eq}(z) = \frac{1}{1 - A_{12}(z)A_{21}(z)} \tag{122}$$

is causal, and that this condition is equivalent to: $1 - a_{12}(0)a_{21}(0) \neq 0$.

B Derivation of the Jacobian matrix at the separating state

B.1 Jacobian matrix at any fixed state

This appendix aims at providing details on the computation of the Jacobian matrix (see Section 4). The first part of this investigation is performed for an arbitrary state (i.e. not necessarily an equilibrium state). This state is assumed to be fixed, i.e. not to depend on the iteration index n. Therefore, this index is omitted hereafter, so that the definition (13) of this state becomes:

$$\theta = [c_{12}(0), \ldots, c_{12}(M), c_{21}(0), \ldots, c_{21}(M)]^T. \tag{123}$$

The elements $J_{ij}(\theta)$ of the Jacobian matrix are defined in (24). Their terms $E_\theta[H(\theta, \xi_{n+1})]^{(i)}$ and $\theta^{(j)}$ may be derived respectively from (15) and (123). This yields the following expressions, where the subscript θ of the mathematical expectation E_θ and the limit "$\lim_{n\to+\infty}$" are omitted for simplicity:

$$\begin{cases} J_{ij}(\theta) = E\left[\dfrac{\partial\left(f_1(s_1(n))g_1(s_2(n-i+1))\right)}{\partial c_{12}(j-1)}\right] & (i,j) \in [1, M+1] \times [1, M+1] \\ J_{ij}(\theta) = E\left[\dfrac{\partial\left(f_1(s_1(n))g_1(s_2(n-i+1))\right)}{\partial c_{21}(j-M-2)}\right] & (i,j) \in [1, M+1] \times [M+2, 2M+2] \\ J_{ij}(\theta) = E\left[\dfrac{\partial\left(f_2(s_2(n))g_2(s_1(n-i+M+2))\right)}{\partial c_{12}(j-1)}\right] & (i,j) \in [M+2, 2M+2] \times [1, M+1] \\ J_{ij}(\theta) = E\left[\dfrac{\partial\left(f_2(s_2(n))g_2(s_1(n-i+M+2))\right)}{\partial c_{21}(j-M-2)}\right] & (i,j) \in [M+2, 2M+2] \times [M+2, 2M+2] \end{cases} \tag{124}$$

It clearly appears from equation (124) that the computation of the Jacobian matrix requires the calculus of the partial derivatives of the outputs $(s_1(n), s_2(n))$ and their delayed versions versus the components of θ. For the fixed state θ, the input-ouput relationship of the recurrent structure i.e :

$$s_i(n) = y_i(n) - \sum_{k=0}^{M} c_{ij}(k)s_j(n-k), \quad i \neq j \in \{1, 2\} \tag{125}$$

leads to:

$$\begin{cases} \dfrac{\partial s_1(n)}{\partial c_{12}(k)} = -c_{12} * \dfrac{\partial s_2(n)}{\partial c_{12}(k)} - s_2(n-k) \\ \dfrac{\partial s_2(n)}{\partial c_{12}(k)} = -c_{21} * \dfrac{\partial s_1(n)}{\partial c_{12}(k)} \\ k \in [0, M] \end{cases} \tag{126}$$

where c_{12} and c_{21} denote respectively the impulse responses of the filters $(c_{12}(k))_{k\geq 0}$ and $(c_{21}(k))_{k\geq 0}$ and $*$ denotes the convolution product. Solving this set of two equations yields:

$$\begin{cases} \dfrac{\partial s_1(n)}{\partial c_{12}(k)} = -(w * s_2)(n-k) \\ \dfrac{\partial s_2(n)}{\partial c_{12}(k)} = (w * c_{21} * s_2)(n-k) \\ k \in [0, M] \end{cases} \tag{127}$$

where w denote the impulse response of the filter $(w(k))_{k\geq 0}$ defined in Sub-section 2.2. One derives in the same way:

$$\left\{ \begin{array}{rcl} \dfrac{\partial s_2(n)}{\partial c_{21}(k)} & = & -(w * s_1)(n-k) \\ \dfrac{\partial s_1(n)}{\partial c_{21}(k)} & = & (w * c_{12} * s_1)(n-k) \\ k \in [0, M] & & \end{array} \right. \tag{128}$$

Combining (124), (127) and (128) leads to the explicit formulation of the Jacobian matrix entries, i.e:

- for $(i,j) \in [1, M+1] \times [1, M+1]$

$$\begin{array}{rcl} J_{ij}(\theta) & = & -E\left[f_1'(s_1(n))(w * s_2(n-j+1))g_1(s_2(n-i+1)\right] \\ & & +E\left[f_1(s_1(n))g_1'(s_2(n-i+1))(w * c_{21} * s_2(n-i-j+2))\right], \end{array} \tag{129}$$

- for $(i,j) \in [1, M+1] \times [M+2, 2M+2]$

$$\begin{array}{rcl} J_{ij}(\theta) & = & E\left[f_1'(s_1(n))(w * c_{12} * s_1(n-j+M+2))g_1(s_2(n-i+1)\right] \\ & & -E\left[f_1(s_1(n))g_1'(s_2(n-i+1))(w * s_1(n-i-j+M+3))\right], \end{array} \tag{130}$$

- for $(i,j) \in [M+2, 2M+2] \times [1, M+1]$

$$\begin{array}{rcl} J_{ij}(\theta) & = & E\left[f_2'(s_2(n))(w * c_{21} * s_2(n-j+1))g_2(s_1(n-i+M+2)\right] \\ & & -E\left[f_2(s_2(n))g_2'(s_1(n-i+M+2))(w * s_2(n-i-j+M+3))\right], \end{array} \tag{131}$$

- for $(i,j) \in [M+2, 2M+2] \times [M+2, 2M+2]$

$$\begin{array}{rcl} J_{ij}(\theta) & = & -E\left[f_2'(s_2(n))(w * s_1(n-j+M+2))g_2(s_1(n-i+M+2)\right] \\ & & +E\left[f_2(s_2(n))g_2'(s_1(n-i+M+2))(w * c_{12} * s_1(n-i-j+2M+4))\right]. \end{array} \tag{132}$$

B.2 Jacobian matrix at the separating state for white sources

At the separating state, each output $s_i(n)$ is equal to the source $x_i(n)$ and $c_{ij}(k) = a_{ij}(k)$ for $k \in [0, M]$. Also taking into account the causality assumption ($w(k) = 0$ for $k < 0$), the statistical independence of the sources and the whiteness of each source (see (AS5)), (129) to (132) yield:

- for $(i,j) \in [1, M+1] \times [1, M+1]$

$$J_{ij}(\theta^s) = -E[f_1'(x_1)]E[x_2 g_1(x_2)]w_{eq}(i-j) + E[f_1(x_1)]E[x_2 g_1'(x_2)]w_{eq}(0)a_{21}(0)\delta_{j,1}, \tag{133}$$

- for $(i,j) \in [1, M+1] \times [M+2, 2M+2]$

$$J_{ij}(\theta^s) = w_{eq}(0)a_{12}(0)E[x_1 f_1'(x_1)]E[g_1(x_2)]\delta_{j,M+2} - E[x_1 f_1(x_1)]E[g_1'(x_2)]w_{eq}(0)\delta_{i,1}\delta_{j,M+2}, \tag{134}$$

- for $(i,j) \in [M+2, 2M+2] \times [1, M+1]$

$$J_{ij}(\theta^s) = w_{eq}(0)a_{21}(0)E[x_2 f_2^{'}(x_2)]E[g_2(x_1)]\delta_{j,1} - E[x_2 f_2(x_2)]E[g_2^{'}(x_1)]w_{eq}(0)\delta_{j,1}\delta_{i,M+2}, \tag{135}$$

- for $(i,j) \in [M+2, 2M+2] \times [M+2, 2M+2]$

$$J_{ij}(\theta^s) = -E[f_2^{'}(x_2)]E[x_1 g_2(x_1)]w_{eq}(i-j) + E[f_2(x_2)]E[x_1 g_2^{'}(x_1)]w_{eq}(0)a_{12}(0)\delta_{j,M+2}, \tag{136}$$

where $\delta_{p,k}$ is the Kronecker symbol defined as:

$$\begin{cases} \delta_{p,k} = 1 & if\ p = k \\ \delta_{p,k} = 0 & otherwise \end{cases} \tag{137}$$

The jacobian matrix can then be written as a block matrix,

$$J(\theta^s) = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \tag{138}$$

where

$$G_{ii} = \begin{pmatrix} -\alpha_i w_{eq}(0) + \varphi_i w_{eq}(0) & 0 & \dots & 0 \\ -\alpha_i w_{eq}(1) + \varphi_i w_{eq}(0) & -\alpha_i w_{eq}(0) & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ -\alpha_i w_{eq}(M) + \varphi_i w_{eq}(0) & -\alpha_i w_{eq}(M-1) & \dots & -\alpha_i w_{eq}(0) \end{pmatrix}, \tag{139}$$

and

$$G_{ij} = \begin{pmatrix} \eta_i w_{eq}(0) - \beta_i w_{eq}(0) & 0 & \dots & 0 \\ \eta_i w_{eq}(0) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \eta_i w_{eq}(0) & 0 & \dots & 0 \end{pmatrix}, \qquad i \neq j \in \{1,2\} \tag{140}$$

with the following notations:

$$\begin{cases} \alpha_i = E[f_i^{'}(x_i)]E[x_j g_i(x_j))] & where\ j\ is\ chosen\ so\ that\ j \neq i \in \{1,2\} \\ \beta_i = E[x_i f_i(x_i)]E[g_i^{'}(x_j))] & where\ j\ is\ chosen\ so\ that\ j \neq i \in \{1,2\} \\ \varphi_i = a_{ji}(0)E[f_i(x_i)]E[x_j g_i^{'}(x_j))] & where\ j\ is\ chosen\ so\ that\ j \neq i \in \{1,2\} \\ \eta_i = a_{ij}(0)E[x_i f_i^{'}(x_i)]E[g_i(x_j))] & where\ j\ is\ chosen\ so\ that\ j \neq i \in \{1,2\} \end{cases} \tag{141}$$

and where $w_{eq}(0)$ is the zero-lag coefficient of the filter $W_{eq}(z)$ defined in Sub-section 2.2 i.e: $w_{eq}(0) = \dfrac{1}{1 - a_{12}(0)a_{21}(0)}$.

In the case when the separating state θ^s is also an equilibrium point and under the condition (21), $\eta_i = 0$ so that the matrix G_{ij} for $i \neq j \in \{1,2\}$ becomes:

$$G_{ij} = \begin{pmatrix} -\beta_i w_{eq}(0) & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 \end{pmatrix}, \qquad i \neq j \in \{1,2\}. \tag{142}$$

C Derivation of $R(\theta^s)$

As shown in Sub-section 5.2, the estimation of the asymptotic error variance needs the computation of the matrix $R(\theta^s)$ defined by

$$R(\theta^s) = \sum_{n \in \mathcal{Z}} Cov[H(\theta^s, \xi_{n+1}), H(\theta^s, \xi_0)]. \tag{143}$$

Besides, taking into account (5) and (15) reduced to the case of strictly causal mixtures, $H(\theta^s, \xi_{n+1})$ can be written as:

$$H(\theta^s, \xi_{n+1}) = [\Sigma_{1,n}^s, \Sigma_{2,n}^s]^T, \tag{144}$$

where

$$\begin{cases} \Sigma_{1,n}^s = [f_1(x_1(n))g_1(x_2(n-1)), \ldots, f_1(x_1(n))g_1(x_2(n-M))] \\ \Sigma_{2,n}^s = [f_2(x_2(n))g_2(x_1(n-1)), \ldots, f_2(x_2(n))g_2(x_1(n-M))] \end{cases} \tag{145}$$

Using the above-mentioned notations, a new formulation of (143) can be derived:

$$\begin{aligned} R(\theta^s) &= \sum_{n \in \mathcal{Z}} \begin{pmatrix} Cov[\Sigma_{1,n}^s, \Sigma_{1,0}^s] & Cov[\Sigma_{1,n}^s, \Sigma_{2,0}^s] \\ Cov[\Sigma_{2,n}^s, \Sigma_{1,0}^s] & Cov[\Sigma_{2,n}^s, \Sigma_{2,0}^s] \end{pmatrix} && (146) \\ &= \begin{pmatrix} \gamma_1 R_1 & 0_M \\ 0_M & \gamma_2 R_2 \end{pmatrix}, && (147) \end{aligned}$$

where

$$R_i = \begin{pmatrix} 1 & a_i & \ldots & a_i \\ a_i & 1 & \ddots & a_i \\ \vdots & \ddots & \ddots & a_i \\ a_i & \ldots & a_i & 1 \end{pmatrix}, \tag{148}$$

$$a_i = \frac{E^2[f_i(x_i)]}{E[f_i^2(x_i)]}, \tag{149}$$

$$\gamma_i = E[f_i^2(x_i)]E[g_i^2(x_j)], \quad \textit{where } j \textit{ is chosen so that } i \neq j \in \{1, 2\}, \tag{150}$$

and 0_M is the M by M null matrix. Moreover, the transition from (146) to (147) uses the fact that θ^s was assumed to be the separating state and that $E[g_i(x_j)] = 0$, as reported in (21).

D Some properties of the matrices K_i

D.1 General formulation of K_i

In Section 5, we consider the symmetric and positive solution K_i of the Lyapunov equation (72) that reads:

$$-GK_i - K_iG^T + R_i = 0_M, \tag{151}$$

where G is a lower triangular matrix, defined in (39), that only depends on the mixing medium and R_i is defined in (68). In fact, R_i and therefore K_i are functions of a_i defined in (69) and in seek for clarity they are denoted respectively $R_i(a_i)$ and $K_i(a_i)$ hereafter. Taking the partial derivative of (151) with respect to a_i leads to:

$$-G\frac{\partial K_i}{\partial a_i}(a_i) - \frac{\partial K_i}{\partial a_i}(a_i)G^T + \frac{\partial R_i}{\partial a_i}(a_i) = 0_M \tag{152}$$

where:

$$\frac{\partial R_i}{\partial a_i}(a_i) = \begin{pmatrix} 0 & 1 & \dots & 1 \\ 1 & 0 & \ddots & 1 \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix}. \tag{153}$$

Besides, G and $\frac{\partial R_i}{\partial a_i}(a_i)$ are independent from a_i, i.e. from the sources statistics and the separating functions. Therefore, as a solution of (152), $\frac{\partial K_i}{\partial a_i}(a_i)$ has the same property, i.e. $\frac{\partial K_i}{\partial a_i}(a_i) = Q_{i1}$ where Q_{i1} is a real matrix that depends only on the mixture parameters. Furthermore, combining (151), (152) and (153), one gets:

$$-G(K_i(a_i) - a_i\,\frac{\partial K_i}{\partial a_i}(a_i)) - (K_i(a_i) - a_i\,\frac{\partial K_i}{\partial a_i}(a_i))G^T + I_M = 0_M. \tag{154}$$

(154) implies also that $K_i(a_i) - a_i\,\frac{\partial K_i}{\partial a_i}(a_i)$ depends only on the mixing medium, i.e. $K_i(a_i) - a_i\,\frac{\partial K_i}{\partial a_i}(a_i) = Q_{i2}$ where Q_{i2} is a real matrix that only depends on the mixing system. As a result, the matrix $K_i(a_i)$ can be written as:

$$K_i(a_i) = Q_{i2} + Q_{i1}a_i \tag{155}$$

It should be noted that $Q_{i2} = K_i(0)$, so that Q_{i2} is a symmetric matrix. Similarly, $Q_{i1} = \frac{\partial K_i}{\partial a_i}(a_i)$ is also symmetric.

D.2 Some properties of $Tr(K_i(a_i))$

For $a_i \in [0, 1[$, the matrix $K_i(a_i)$ is symmetric and positive definite by construction (see Sub-section 5.2). It has then the following properties:

- $(P1)$: for all non null vectors X, we have $X^T K_i(a_i)X > 0$.
- $(P2)$: all the eigenvalues of $K_i(a_i)$ are strictly positive.
- $(P3)$: the trace of $K_i(a_i)$ is strictly positive (since it is the sum of the eigenvalues of $K_i(a_i)$).

Besides, due to (155), the trace of the matrix $K_i(a_i)$ can be expressed as:

$$Tr(K_i(a_i)) = q_{i2} + q_{i1}a_i \tag{156}$$

where $q_{ij} = Tr(Q_{ij})$ for $i, j \in \{1, 2\}$. It should be noted that $q_{i2} = Tr(Q_{i2}) = Tr(K_i(0)) > 0$ due to $(P3)$.

In the following, we show that $q_{i2} + q_{i1}a_i > 0$ for $a_i \in [0, 1]$:

- $q_{i2} + q_{i1} a_i > 0$ for $a_i \in [0, 1[$ is due to $(P3)$.

- $q_{i2} + q_{i1} > 0$. This result is not an immediate consequence of the previous one. Let's define $K_i(1)$ by:

$$K_i(1) = \lim_{a_i \to 1} K_i(a_i) = Q_{i2} + Q_{i1} \tag{157}$$

$K_i(1)$ is therefore symmetric. Moreover, due to the continuity of $K_i(a_i)$ and $R_i(a_i)$ in the vicinity of $a_i = 1$, the limit of (151) when a_i tends to 1 shows that $K_i(1)$ is such that:

$$-GK_i(1) - K_i(1)G^T + R_i(1) = 0_M. \tag{158}$$

Besides, (155) shows that the matrix $K_i(a_i)$ is a linear function of a_i, so that it can be written as:

$$K_i(a_i) = (1 - a_i)K_i(0) + a_i K_i(1) \tag{159}$$

Therefore, whatever the vector X, we have:

$$X^T K_i(a_i)X = (1 - a_i)X^T K_i(0)X + a_i X^T K_i(1)X. \tag{160}$$

Let us consider a nonzero vector X. Then $X^T K_i(0)X > 0$ due to $(P1)$. Now assume that $X^T K_i(1)X < 0$. Then, when a_i is varied from 0 to 1, (160) shows that $X^T K_i(a_i)X$ varies continuously from $X^T K_i(0)X > 0$ to $X^T K_i(1)X < 0$. Therefore, there exists a value $a_i \in]0, 1[$ such that $X^T K_i(a_i)X = 0$. But this is uncompatible with $(P1)$, which means that the above assumption is false, i.e. whatever the nonzero vector X $X^T K_i(1)X \geq 0$. $K_i(1)$ is thus a semi-positive definite matrix and all its eigenvalues are positive or null. Moreover, since $K_i(1)$ is symmetric, it can be expressed as $K_i(1) = UDU^T$ where D is a diagonal matrix containing the eigenvalues of $K_i(1)$ and U is a unitary matrix. Therefore, if all these eigenvalues were null, D would be the null matrix, and so would $K_i(1)$. But this is not compatible with (158). Hence, at least one of the eigenvalues of $K_i(1)$ is strictly positive and so is their sum i.e. $Tr(K_i(1)) = q_{i2} + q_{i1} > 0$.

E Optimum separating functions

In Sections 5 and 6, the extremum or optimum separating functions were shown to be related to the minimization of the following expressions:

- $\left(\frac{E[f_i^2(x_i)]}{E[f_i'(x_i)]}(q_{i2} + \frac{E^2[f_i(x_i)]}{E[f_i^2(x_i)]}q_{i1})\right)^2$, $\left(\frac{\sqrt{E[f_i^2(x_i)]}}{E[f_i'(x_i)]}(q_{i2} + \frac{E^2[f_i(x_i)]}{E[f_i^2(x_i)]}q_{i1})\right)^2$ and $\frac{E[f_i^2(x_i)]}{E^2[f_i'(x_i)]}(q_{i2} + \frac{E^2[f_i(x_i)]}{E[f_i^2(x_i)]}q_{i1})$ for the choice of the function f_i respectively for algorithms (N0), (N1) and (N2).

- $\left(\frac{E[g_i^2(x_j)]}{E[x_j g_i(x_j)]}\right)^2$, $\left(\frac{\sqrt{E[g_i^2(x_j)]}}{E[x_j g_i(x_j)]}\right)^2$ and $\frac{E[g_i^2(x_j)]}{E^2[x_j g_i(x_j)]}$ for optimizing the function g_i respectively for algorithms (N0), (N1) and (N2).

In the following, suffixes i and j will be omitted to avoid a cumbersome presentation. Nevertheless, we must keep in mind that f_i depends only on the source x_i and g_i is only related to x_j.

Using the above-mentioned notations, all the expressions to be optimized can be expressed as: $L_1(f) = \frac{E^k[f^2(x)]}{E^2[f'(x)]}\left(q_2 + \frac{E^2[f(x)]}{E[f^2(x)]}q_1\right)^m$ for $(k, m) \in \{(1,1), (1,2), (2,2)\}$ or $L_2(g) = \frac{E^k[g^2(x)]}{E^2[xg(x)]}$ for $k \in \{1, 2\}$.

Hereafter, the following assumptions are assumed to be met:

- $(AS7)$ the probability density functions of the sources are defined for all real values, twice continuously derivable and such that $\lim_{|x|\to+\infty} p(x) = 0$ and $\lim_{|x|\to+\infty} p'(x) = 0$.
- $(AS8)$ $f(x)$ is a continuously derivable function.
- $(AS9)$ $g(x)$ is a continuous function.

Note that some less restrictive assumptions could be used at the expense of a more cumbersome presentation.

E.1 Optimum choice for f(x)

For given sources and mixture parameters, let us consider the functional L_1 defined by:

$$L_1(f) = \frac{E^k[f^2(x)]}{E^2[f'(x)]}\left(q_2 + \frac{E^2[f(x)]}{E[f^2(x)]}q_1\right)^m, \tag{161}$$

where $(k, m) \in \{(2,2), (1,2), (1,1)\}$ correspond respectively to (N0), (N1) and (N2).

(161) can be written in a more explicit form as:

$$L_1(f) = \frac{\left(\int_{-\infty}^{+\infty} f^2(x)p(x)dx\right)^k}{\left(\int_{-\infty}^{+\infty} f'(x)p(x)dx\right)^2}\left(q_2 + \frac{\left(\int_{-\infty}^{+\infty} f(x)p(x)dx\right)^2}{\int_{-\infty}^{+\infty} f^2(x)p(x)dx}q_1\right)^m. \tag{162}$$

E.1.1 Extremum condition

L_1 is a differentiable functional of f and so its extrema are reached under the necessary condition:

$$dL_1(f_{ext})(\psi) = 0, \tag{163}$$

for all the continuously derivable functions ψ. $dL_1(f)(\psi)$ denotes the differential of L_1 at f along the trajectory defined by ψ. Differentiating (162) versus f leads to:

$$dL_1(f)(\psi) = \eta_1 \int_{-\infty}^{+\infty} f(x)p(x)\psi(x)dx - \eta_2 \int_{-\infty}^{+\infty} \psi'(x)p(x)dx + \eta_3 \int_{-\infty}^{+\infty} p(x)\psi(x)dx \tag{164}$$

where $\psi(x)$ is a continuously derivable function and

$$\eta_1 = 2\frac{E^{k-1}[f^2(x)]}{E^2[f'(x)]}\left(q_1(k-m)a + kq_2\right), \tag{165}$$

$$\eta_2 = 2\frac{E^k[f^2(x)]}{E^3[f'(x)]}\left(q_2 + q_1 a\right), \tag{166}$$

$$\eta_3 = 2\frac{E^{k-1}[f^2(x)]E[f(x)]}{E^2[f'(x)]}\, q_1 m, \tag{167}$$

$$a = \frac{E^2[f(x)]}{E[f^2(x)]}. \tag{168}$$

Besides, for all the functions $\psi(x)$ that increase more smoothly than $p(x)$ decreases, i.e. which are such that:

$$\lim_{|x|\to+\infty} \psi(x)p(x) = 0, \tag{169}$$

we can derive the following result using a by-part integration:

$$\int_{-\infty}^{+\infty} \psi^{'}(x)p(x)dx = -\int_{-\infty}^{+\infty} \psi(x)p^{'}(x)dx. \tag{170}$$

The extremum condition (163) applied to (164) then becomes:

$$\eta_1 \int_{-\infty}^{+\infty} f_{ext}(x)p(x)\psi(x)dx + \eta_2 \int_{-\infty}^{+\infty} \psi(x)p^{'}(x)dx + \eta_3 \int_{-\infty}^{+\infty} p(x)\psi(x)dx = 0, \tag{171}$$

which is equivalent to:

$$\int_{-\infty}^{+\infty} (\eta_1 f_{ext}(x)p(x) + \eta_2 p^{'}(x) + \eta_3 p(x))\psi(x)dx = 0. \tag{172}$$

(172) should be valid for all the functions $\psi(x)$ that meet (169), which implies that:

$$\eta_1 f_{ext}(x) = -\eta_2 \frac{p^{'}(x)}{p(x)} - \eta_3. \tag{173}$$

Besides, the equality $\eta_1 = 0$ cannot hold since it would require a probability density function of the form $p(x) = v_1 exp(-\frac{\eta_3}{\eta_2}x)$ where v_1 is an integration constant. This is incompatible with (169) and so $\eta_1 \neq 0$. Hence, (173) can be rewritten as

$$f_{ext}(x) = -\frac{\eta_2}{\eta_1}\frac{p^{'}(x)}{p(x)} - \frac{\eta_3}{\eta_1}. \tag{174}$$

Taking the mathematical expectation of (174) and including the expressions of η_i for $i \in \{1, 2, 3\}$ provided in (165) to (167) leads to:

$$E[f_{ext}(x)](kq_2 + mq_1 + (k-m)q_1 a) = 0. \tag{175}$$

Two cases are to be considered:

- $m = k$ which corresponds to (N0) and (N2). (175) is then equivalent to:

$$E[f_{ext}(x)](q_1 + q_2) = 0, \tag{176}$$

which leads to $E[f_{ext}(x)] = 0$ since $q_1 + q_2 > 0$ according to Appendix D. This yields $\eta_3 = 0$ and therefore

$$f_{ext}(x) = -\frac{\eta_2}{\eta_1}\frac{p^{'}(x)}{p(x)}. \tag{177}$$

For algorithm (N2), this results in a unique normalized function as defined in (103), i.e.

$$F_{ext}(x) = \frac{\frac{p^{'}(x)}{p(x)}}{E[(\frac{p^{'}(x)}{p(x)})^{'}]\sqrt{E[x^2]}}. \tag{178}$$

- $m \neq k$, which corresponds to the case $(k,m)=(1,2)$ and the algorithm (N1). combining (161) and (174), the extremum value of L_1 can be rewritten as

$$L_{1ext} = \frac{E[(\frac{p'(x)}{p(x)}+\frac{\eta_3}{\eta_2})^2]}{E^2[(\frac{p'(x)}{p(x)})']}(q_2+q_1a)^2 \tag{179}$$

$$= \frac{E[(\frac{p'(x)}{p(x)})^2]+(\frac{\eta_3}{\eta_2})^2}{E^2[(\frac{p'(x)}{p(x)})']}(q_2+q_1a)^2. \tag{180}$$

Therefore:

1. if $E[f_{ext}(x)]=0$, (167) yields $\eta_3=0$ and (168) yields $a=0$. The corresponding extremum separating function f_{ext} is derived from (174) which yields:

$$f_{ext}^{(1)} = -\frac{\eta_2}{\eta_1}\frac{p'(x)}{p(x)}. \tag{181}$$

   The associated normalized function $F_{ext}^{(1)}$ defined in (86) is given by:

$$F_{ext}^{(1)}(x) = \pm\frac{\frac{p'(x)}{p(x)}}{\sqrt{E[(\frac{p'(x)}{p(x)})^2]}}. \tag{182}$$

   The corresponding value of L_{1ext} is then denoted $L_{1ext}^{(1)}$ and reads:

$$L_{1ext}^{(1)} = q_2^2\frac{E[(\frac{p'(x)}{p(x)})^2]}{E^2[(\frac{p'(x)}{p(x)})']}. \tag{183}$$

2. if $E[f_{ext}(x)]\neq 0$, (167) yields $\eta_3\neq 0$. This solution of (175) holds under the condition $a=2+\frac{q_2}{q_1}$ which implies that $0<-q_1<q_2\leq -2q_1$ (because $a\in[0,1[$ and $q_2>0$ as shown in Appendix D.2). This corresponds to the normalized separating function denoted $F_{ext}^{(2)}$ and defined by:

$$F_{ext}^{(2)}(x) = \pm\sqrt{1-a}\left[\frac{\frac{p'(x)}{p(x)}}{\sqrt{E[(\frac{p'(x)}{p(x)})^2]}}+\epsilon\sqrt{\frac{a}{1-a}}\right] \tag{184}$$

   where $\epsilon\in\{-1,1\}$. The associated value of L_{1ext} is denoted $L_{1ext}^{(2)}$ and is given by:

$$L_{1ext}^{(2)} = -4q_1(q_1+q_2)\frac{E[(\frac{p'(x)}{p(x)})^2]}{E^2[(\frac{p'(x)}{p(x)})']}. \tag{185}$$

   One can then verify easily that $L_{1ext}^{(1)} \geq L_{1ext}^{(2)}$ since $0<-q_1<q_2<-2q_1$. However, this is achieved at the expense of lower convergence speed, since the eigenvalues of the Jacobian matrix $J(\theta^s)$ associated to $F_{ext}^{(2)}$ are equal to those associated to $F^{(1)}_{ext}$ attenuated by the factor $\sqrt{1-a}$. The algorithm behaves as if the adaptation gain μ was replaced by $\mu\sqrt{1-a}$. In order to use a relevant comparison strategy, we consider the function $F^{(2)}_{ext}(x)$ associated to the adaptation gain $\frac{\mu}{\sqrt{1-a}}$ which leads to the same eigenvalues as $F^{(1)}_{ext}$ associated to the adaptation gain μ. The new extremum value of L_1 will then be $L^{(3)}_{1ext} = \frac{L^{(2)}_{1ext}}{1-a}$. This leads to:

$$L^{(3)}_{1ext} = 4q_1^2 \frac{E[(\frac{p'(x)}{p(x)})^2]}{E^2[(\frac{p'(x)}{p(x)})']} \implies L^{(3)}_{1ext} = 4(\frac{q_1}{q_2})^2 L^{(1)}_{1ext} \tag{186}$$

$$\implies L^{(3)}_{1ext} \geq L^{(1)}_{1ext}. \tag{187}$$

Hence for a given convergence speed, the separating function $F^{(1)}_{ext}$ behaves better than $F^{(2)}_{ext}$ from the point of view of the asymptotic error variance. This is therefore the function kept hereafter.

As a result, for all three algorithms (N0), (N1) and (N2) the separating functions to be retained are of the form:

$$f_i(x) \quad \propto \quad -\frac{p'_{x_i}(x)}{p_{x_i}(x)} \tag{188}$$

$$\tag{189}$$

where p_{x_i} is the probability density function of the source x_i.

E.1.2 Optimum functions

The separating functions retained in Sub-section E.1.1 were found using only an extremum condition which is necessary and not sufficient to guarantee the minimization of $L_1(f)$.
To prove that the extremum functions thus obtained are the optimum ones, we study the values of $L_1(f)$ for some typical functions hereafter. We start with algorithm (N2) because it is easier to handle due to the uniqueness of the normalized extremum function F_{ext} (see (178)). In this case, we just have to exhibit a function f such that $L_1(f) > L_{1ext}$.

- **Algorithm (N2)**
  Let us consider the class of separating functions $f^{(n)}(x) = x^n$ for $n \geq 1$. The associated values of L_1 are:

$$L_1^{(n)} = \frac{E[x^{2n}]}{n^2 E^2[x^{n-1}]}(q_2 + q_1 \frac{E^2[x^n]}{E[x^{2n}]}) \tag{190}$$

  Moreover, f_{ext} is associated to L_{1ext} given by:

$$L_{1ext} = q_2 \frac{E[(\frac{p'(x)}{p(x)})^2]}{E^2[(\frac{p'(x)}{p(x)})']} = q_2 \frac{1}{E[(\frac{p'(x)}{p(x)})^2]}. \tag{191}$$

Therefore, considering the function $f^{(1)}$ of the above-defined class yields:

$$\frac{L_1^{(1)}}{L_{1ext}} = E[x^2]E[(\frac{p'(x)}{p(x)})^2] \implies \frac{L_1^{(1)}}{L_{1ext}} \geq E^2[x\frac{p'(x)}{p(x)}] \qquad (192)$$

$$\implies \frac{L_1^{(1)}}{L_{1ext}} \geq 1, \qquad (193)$$

where the transition in (192) uses the Cauchy-Schwarz inequality.
The equality in (193) holds only for Gaussian probability density functions (p.d.f). Therefore, for non-Gaussian p.d.f, $f^{(1)}$ is such that $L_1(f) > L_{1ext}$. For Gaussian p.d.f, we consider another function, i.e. $f^{(3)}$, which yields:

$$\frac{L_1^{(3)}}{L_{1ext}} = \frac{E[x^6]}{9E^2[x^2]}E[(\frac{p'(x)}{p(x)})^2] \implies \frac{L_1^{(3)}}{L_{1ext}} = \frac{E[x^6]}{9E^3[x^2]} \qquad (194)$$

$$\implies \frac{L_1^{(3)}}{L_{1ext}} = \frac{5}{3}. \qquad (195)$$

Hence, under the regularity assumptions made on the p.d.f of the sources, we can always exhibit functions f such that $L_1(f) > L_{1ext}$. f_{ext} corresponds then to the minimum value of L_1 and so it is optimum.

- **Algorithm (N1)**
Let's consider the extremum function retained in Sub-section E.1.1, i.e. $f_{ext}(x) = -\nu\frac{p'(x)}{p(x)}$. The associated value of L_1 is then:

$$L_{1ext} = \frac{E[f_{ext}^2(x)]}{E^2[f_{ext}'(x)]}\left(q_2 + \frac{E^2[f_{ext}(x)]}{E[f_{ext}^2(x)]}q_1\right)^2. \qquad (196)$$

Let's consider now a separating function f. Like in sub-section E.1.1, to have a relevant comparison criterion for comparing f_{ext} and f, we consider the performance of these two functions when the algorithms have the same convergence speed at the vicinity of the equilibrium state[11]. Under this assumption, the algorithm (N1) associated to f behaves as if the adaptation gain μ was multiplied by $\frac{E[f_{ext}'(x)]}{E[f'(x)]}\sqrt{\frac{E[f_{ext}(x)^2]}{E[f^2(x)]}}$, which corresponds to the value of L_1:

$$L_1 = \frac{E[f_{ext}'(x)]}{E^2[f_{ext}^2(x)]}\frac{E^2[f^2(x)]}{E^4[f'(x)]}\left(q_2 + \frac{E^2[f(x)]}{E[f^2(x)]}q_1\right)^2. \qquad (197)$$

The minimization of (197) is then equivalent to the minimization of $\frac{E[f^2(x)]}{E^2[f'(x)]}\left(q_2 + \frac{E^2[f(x)]}{E[f^2(x)]}q_1\right)$ which corresponds to the value of L_1 associated to the normalization scheme (N2). Hence, the results are the same as those obtained for (N2), i.e. the optimum separating function is associated to $f_{ext}(x) = -\nu\frac{p'(x)}{p(x)}$.
Note also that, for zero-mean separating functions f, we have:

$$L_1(f) = \frac{E[f^2(x)]}{E^2[f'(x)]}q_2^2 = \frac{1}{\lambda_f^2}q_2^2 \qquad (198)$$

[11] for a given function g.

where $\lambda_f = \frac{E[f'(x)]}{\sqrt{E[f^2(x)]}}$ is the contribution related to f in the eigenvalues of the Jacobian matrix at the separating state $\theta^{(s)}$. Hence, for a given separating function g, the convergence speed in the vicinity of the equilibrium state is an increasing function to the absolute value of λ_f. Furthermore, the minimization of L_1 becomes coupled with the maximization of λ_f. The reader can verify easily that separating functions proportional to $-\frac{p'(x)}{p(x)}$ enable simultaneous maximization of the local convergence speed and the minimization of L_1 which is a very important property.

E.1.3 Conclusion for $f(x)$

Using the suffixes i, the optimum normalized separating functions F_i are eventually expressed as follows:

- for the algorithm (N1):

$$F_{iopt}(x) = \pm \frac{\frac{p'_{x_i}(x)}{p_{x_i}(x)}}{\sqrt{E[(\frac{p'_{x_i}(x)}{p_{x_i}(x)})^2]}} \tag{199}$$

- for the algorithm (N2):

$$F_{iopt}(x) = \frac{\frac{p'_{x_i}(x)}{p_{x_i}(x)}}{E[(\frac{p'_{x_i}(x)}{p_{x_i}(x)})']\sqrt{E[x^2]}} \tag{200}$$

where p_{x_i} is the probability density function of the source x_i.

E.2 Optimun choice for g(x)

For the algorithm (N0), the extrema of the functionnal

$$L_2(g) = \frac{E[g^2(x)]^2}{E[xg(x)]^2} \tag{201}$$

have to be determined under the constraint (21). An approach based on Lagrange multipliers then yields:

$$g_{ext}(x) = \eta x \tag{202}$$

where η is a constant.
For both algorithms (N1) and (N2) the optimum functions g are the ones which minimize:

$$L_2(g) = \frac{E[g^2(x)]}{E[xg(x)]^2}. \tag{203}$$

The investigation of L_2 can be done in two steps as for L_1. A more compact method can be based on the Cauchy-Schwarz inequality that enables to handle directly the minimization of L_2. In fact, the Cauchy-Schwarz inequality implies that:

$$E[xg(x)]^2 \leq E[g^2(x)]E[x^2] \Longrightarrow L_2(g) \geq \frac{1}{E[x^2]}. \tag{204}$$

The minimum value of $L_2(g)$ is reached when the equality holds in (204), which corresponds to:

$$g_{ext}(x) = \eta \ x. \tag{205}$$

where η is a constant that must meet the algorithm stability condition.
Using the notations with suffixes i and j, the optimum solution is:

$$g_{iopt}(x) = \eta_{gi} \ x. \tag{206}$$

where η_{gi} is a real constant.
The contribution of the function g in algorithms (N1) and (N2) is then:

$$G_{iopt}(x) = \frac{g_{iopt}(x)}{\sqrt{E[g_{iopt}^2(x)]}} = d\frac{x}{\sqrt{E[x^2]}} \tag{207}$$

where $d \in \{-1, 1\}$ depending on the stability condition.

Figures

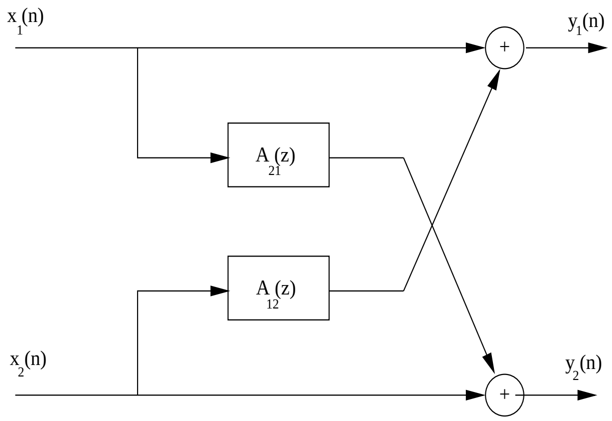

Figure 1: Basic mixture model for source separation.

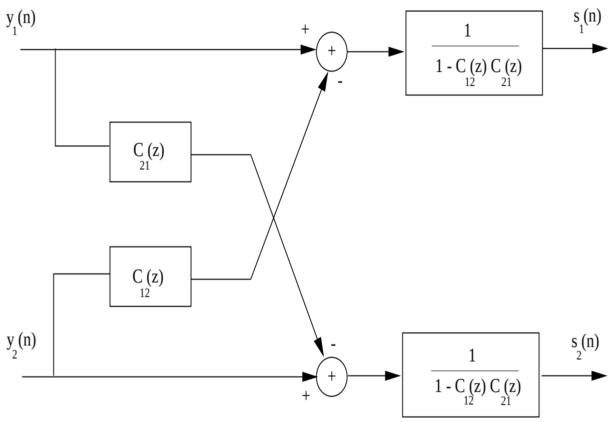

Figure 2: Direct structure for the separation system.

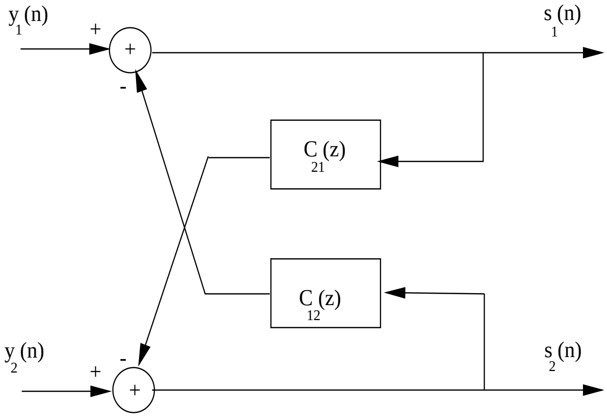

Figure 3: Recurrent structure for the separation system.

References

[1] M.J. Al-Kindi and J. Dunlop "Improved adaptive noise cancellation in the presence of signal leakage on the noise reference channel", Signal Processing, vol. 17, no. 3, pp. 241-250, July 1989.

[2] Y. Bar-Ness, J.W. Carlin and M.L. Steinberger, "Bootstrapping adaptive crosspol canceler for sattelite communication", Proc. ICC'82, International Conference on Communication, Miami, USA, paper 4F.2, June 1982.

[3] A.J. Bell and T. Sejnowski, "Fast blind separation based on information theory", in Proc. International Symposium on Nonlinear Theory and its Applications, pp. 43-47, Las Vegas, USA, 1995.

[4] A. Benveniste, M. Metivier and P. Priouret, *Adaptive algorithms and stochastic approximations,* Applications of Mathematics, vol. 22, Springer-Verlag, 1990.

[5] J.F. Cardoso , "Blind identification of independent signals", In Workshop on Higher-Order Spectral Analysis, Vail (CO), USA, June 1994.

[6] N. Charkani and Y. Deville, "Self-Adaptive Separation of Convolutively Mixed Signals with a Recursive Structure. Part II: Application of optimization analysis to synthetic and real signals", submitted to Signal Processing.

[7] P. Comon , "Independent Component Analysis, a new concept", Signal Processing, vol. 36, no. 3, pp. 287-314, April 1994.

[8] P. Comon , "separation of sources using higher-order cumulants", Advanced algorithms and architectures for signal processing, vol. 1152, August 1989.

[9] P. Comon , "MA identification using fourth order cumulants", Signal Processing, vol. 26, no. 3, pp. 1-10, March 1992.

[10] Y. Deville, "A unified stability analysis of the Hérault-Jutten source separation neural network", Signal Processing, vol. 51, no. 3, pp. 229-233, June 1996.

[11] J.C Fort, "Analyse de l'algorithme de séparation de sources de Jutten et Hérault", Traitement du signal, vol. 8, no. 1, pp. 35-42, 1991.

[12] M. Gaeta and J.L. Lacoume, "Estimateurs du maximum de vraisemblance étendus à la séparation de sources non gaussiennes", Traitement du signal, vol. 7, no. 5, pp. 419-434, 1990.

[13] Ph. Garat, "Approche statistique pour la séparation aveugle de sources", PhD thesis, UJF Grenoble, 1994.

[14] J. Hérault, C. Jutten and B. Ans, "détection des grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé", In Proc. Gretsi'85, Nice, France, pp. 1017-1022, Mai 1985.

[15] C. Jutten and J. Hérault, "Blind separation of sources, part I: an adaptive algorithm based on neuromimetic architecture", Signal Processing, vol. 24, no. 1, pp. 1-10, July 1991.

[16] C. Jutten, "Calcul neuromimétique et traitement du signal: analyse en composantes indépendantes", PhD thesis, UJF-INP Grenoble, 1987.

[17] T. Kailath, *Linear Systems,* NJ: Prentice-Hall, 1980.

[18] J.-L. Lacoume and P. Ruiz, "Source identification: A solution based on cumulants", IEEE ASSP Workshop V, Mineapolis, USA, August 1988.

[19] O. Macchi and E. Moreau , "Self-adaptive source separation, Part I: convergence analysis of a direct linear network controled by the Hérault-Jutten adaptation law", IEEE Transactions on Signal Processing, vol. 45, no. 4, pp. 918-926, April 1997.

[20] A. Mansour and C. Jutten, "Fourth-order criteria for blind sources separation", IEEE Trans. Signal Processing , vol. 43, no. 8, pp. 2022-2025, August 1995.

[21] E. Moreau and O. Macchi "New self-adaptive algorithms for source separation based on contrast functions", in Proc. HOS'93, Signal Processing Workshop on Higher-Order Statistics, pp. 215-219, (Lake Tahoe, USA), 1993.

[22] E. Moreau, "Apprentissage et adaptativité. Séparation auto-adaptative de sources indépendantes", PhD thesis, UP XI Orsay, 1995.

[23] E. Moreau and O. Macchi, "Self-adaptive source separation, Part II: comparison of the direct, feedback and mixed linear network", IEEE Transactions on Signal Processing, vol. 46, no. 1, pp. 39-50, Jan. 1998.

[24] H.L. Nguyen Thi, "Séparation aveugle de sources à large bande dans un mélange convolutif", PhD thesis, INP Grenoble, 1993.

[25] H.L. Nguyen Thi, C. Jutten and J. Caelen, "Speech enhancement: Analysis and comparison of methods in various real situations", In Eusipco'92, European Signal Processing Conference, Brussels, Belgium, pp. 303-306, August 1992.

[26] H.L. Nguyen Thi, C. Jutten, "Blind source separation for convolutive mixtures", Signal Processing, vol. 45, no. 2, pp. 209-229, March 1995.

[27] H.L. Nguyen Thi and C. Jutten, "Whitening for blind separation of sources", In Proc. 3rd Worshop ATHOS on Higher Order Statistics, Edinbourg, Scotland, September, 1994.

[28] C.L Nikias and J.M Mendel, "Signal processing with higher-order spectra", IEEE Signal Processing Magazine, pp. 10-37, July 1993.

[29] A. Papoulis, *Probability, random variables, and stochastic processes.* Singapore: McGraw-Hill, 1984.

[30] D.T. Pham, Ph. Garat and C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach", In Eusipco'92, European Signal Processing Conference, Brussels, Belgium, pp. 771-774, August 1992.

[31] C. Puntonet, A. Mansour and C. Jutten "A geometrical algorithm for blind separation of sources" in Actes du colloque GRETSI'95, (Juan-Les-Pins, France), 1995.

[32] E. Sorouchyari, "Blind separation of sources, part III: Stability analysis", Signal Processing, vol. 24, no. 1, pp. 21-29, July 1991.

[33] S. Van Gerven, D. Van Compernolle, "Feedforward and feedback in symmetric adaptive noise canceller: Stability analysis in a simplified case", In Eusipco'92, European Signal Processing Conference, Brussels, Belgium, pp. 1081-1084, August 1992.

[34] S. Van Gerven, D. Van Compernolle, "Signal separation by symmetric adaptive decorellation: stability, convergence and uniqueness", IEEE Transactions on Signal Processing, vol. 43, no. 7, pp. 1602-1612, July 1995.

[35] E. Weinstein, M. Feder and A.V. Oppenheim, "Multi-channel signal separation by decorrelation", IEEE Trans. Speech and Audio Processing , vol. 1, no. 4, pp. 405-413, October 1993.

[36] B. Widrow et al., "Adaptive noise cancelling: principles and applications", Proceedings of the IEEE, vol. 63, no. 12, pp. 1692-1716, 1975.

[37] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation", IEEE Trans. Signal Processing , vol. 42, no. 8, pp. 2158-2168, August 1994.