Authors' final version of a paper published in "Neurocomputing"

Paper reference:
F. Abrard, Y. Deville, J. Thomas, "Blind partial separation of underdetermined convolutive mixtures of complex sources based on differential normalized kurtosis", Neurocomputing, no. 71, pp. 2071-2086, 2008.

Elsevier on-line version:
http://dx.doi.org/10.1016/j.neucom.2007.07.033

Blind partial separation of underdetermined convolutive mixtures of complex sources based on differential normalized kurtosis

Frédéric Abrard (1), Yannick Deville[1] (2), Johan Thomas (2)

(1) i2e,
1030 av. Guillibert de la Lauziere
BP 140.000
13794 Les Milles - Aix en Provence
France

(2) Laboratoire d'Astrophysique de Toulouse-Tarbes
Observatoire Midi-Pyrénées - Université Paul Sabatier Toulouse 3 - CNRS
14 Av. Edouard Belin
31400 Toulouse
France
E-mail: yannick.deville@ast.obs-mip.fr

[1]Corresponding author. Phone: +33 5 61 33 28 24. Fax: +33 5 61 33 28 40.

Abstract: This paper concerns the blind separation of P complex convolutive mixtures of N statistically independent complex sources, with underdetermined or noisy mixtures i.e. $P < N$. Our approach exploits the assumed distinct statistical properties of the sources: P sources are non-stationary, while the others are stationary. Our method achieves the "partial separation" of the P non-stationary sources. It uses a deflation procedure including extraction and coloration stages. The original criteria introduced in these stages use our differential source separation concept. They consist in optimizing the differential normalized kurtosis and differential power that we introduce. To optimize these criteria, we propose Netwon-like algorithms. Experimental results prove the efficiency of our method.

Keywords:
convolutive mixture,
differential blind source separation,
differential normalized kurtosis,
non-stationary signal,
underdetermined or noisy mixture.

1 Introduction

Blind source separation (BSS) methods aim at restoring a set of N unknown source signals $s_j(n)$ from a set of P observed signals $x_p(n)$, which are mixtures of these source signals [12]. The mixed signals $x_p(n)$ are often provided by a set of sensors, and the mixing phenomenon then results from the simultaneous propagation of all source signals from their emission locations to all sensors. Provided mixing remains linear, it may be represented by a convolutive model, where each propagation channel from source j to sensor p is defined by a transfer function $A_{pj}(z)$. These transfer functions may account for propagation attenuations and delays, and multipath propagation. The overall source-observation relationship then reads in the $\mathcal{Z}$ domain

$$X(z) = A(z)S(z), \tag{1}$$

where $S(z) = [S_1(z) \dots S_N(z)]^T$ and $X(z) = [X_1(z) \dots X_P(z)]^T$ are the $\mathcal{Z}$ transforms of the source and observation vectors, and where the mixing matrix $A(z)$ consists of the above-defined transfer functions $A_{pj}(z)$. This general convolutive model especially includes linear instantaneous mixtures, which have been mainly considered in the literature and which correspond to the situation when all transfer functions $A_{pj}(z)$ are restricted to scalar coefficients.

Most BSS investigations have been performed in the case when: (i) the mixture is determined, i.e. the number P of observed signals is equal to the number N of source signals, so that the considered mixing matrix $A(z)$ is square, and (ii) this matrix is invertible. The BSS problem then consists in estimating the inverse of this mixing matrix, up to some indeterminacies [12]. Various methods have been proposed to this end. They are especially based on the assumed statistical independence or uncorrelation of the source signals [12]. Many of these methods consist in optimizing statistical parameters of the output signals of a BSS system, such as their second-order or higher-order moments or cumulants, which are classical parameters in the higher-order statistics field [14],[16].

In many practical situations however, only a limited number of sensors is acceptable, due e.g. to cost constraints or physical configuration, whereas these sensors receive a larger number of sources (possibly including "noise sources"). In this paper, we consider this underdetermined situation corresponding to $P < N$, and we require that $P \geq 2$. Some analyses and statistical BSS methods have been reported for this case (see e.g. [4],[5],[6],[10],[18]), mainly for linear instantaneous mixtures. However, they set major restrictions on the source properties (discrete sources are especially considered) and/or on the mixing conditions. Other reported approaches set various sparsity requirements on the source signals, especially in the time-frequency domain (see e.g. [1],[2],[9],[17],[20],[21]). In this paper, we aim at avoiding all above constraints, at the expense of only achieving "partial BSS" as explained hereafter.

In [8], we introduced a general differential BSS concept for processing underdetermined mixtures. In its standard version, this approach uses a statistical framework and takes advantage of the distinct properties that the sources are assumed to have from the point of view of stationarity. More precisely, we investigate the situation when (at most) P of the N mixed sources are non-stationary, while the other $N - P$ sources (at least) are stationary. The P non-stationary sources are the signals of interest in this approach, while the $N - P$ stationary sources are considered as disturbance, i.e. "noise sources". This may e.g. correspond to the situation when P microphones provide mixtures of P speech signals superimposed with $N - P$ stationary noise signals. The differential BSS concept that we proposed in [8] then makes it possible to derive output signals which each contain only one of the P sources of interest, still superimposed with some residual components from the $N-P$ noise sources. This optimum case [8], when each output is still a mixture of $(N-P+1)$ sources, is called "partial BSS" (for the P sources of interest). The features thus obtained were described in [8].

Although we first defined this differential BSS concept in a quite general framework in [8], we then only applied it to a specific BSS system and to associated separation criterion and algorithms, based on second-order statistics. We focused on that specific version of our approach because of its simplicity. However, the resulting BSS method is thus limited to the situation when: (i) only $P=2$ mixtures are considered, (ii) the mixing filters include no instantaneous nor non-causal parts and (iii) the mixing matrix is minimum-phase. In the current paper, we therefore aim at deriving differential BSS criteria and associated algorithms which apply to much more general conditions, i.e. to an arbitrary number of observed signals and to arbitrary mixing filters. To this end, we especially resort to higher-order statistical signal parameters, namely to the differential normalized kurtosis that we introduce to this end hereafter.

The remainder of this paper is organized as follows. In Section 2, we derive the two criteria used in the extraction and coloration stages of the proposed differential BSS method and we define its overall structure, based on a deflation procedure. Practical algorithms for optimizing the above two criteria are introduced in Section 3. The experimental performance achieved by the proposed method is presented in Section 4 and conclusions are drawn from this investigation in Section 5.

2 Proposed differential BSS criteria and overall method

2.1 A new extraction method based on differential normalized kurtosis

In this paper, we consider the general case of underdetermined convolutive mixtures of complex-valued signals. We aim at introducing a new partial BSS method for this configuration. To this end, we use our general differential BSS concept, which is described in detail in Section 2 of [8], and which may be applied to various statistical parameters. We here propose a new application of this concept, where the considered parameter is the normalized kurtosis. This classical higher-order statistical parameter has been used by J.K. Tugnait [19] as a criterion for achieving blind source extraction in the case when $P=N$, assuming statistically independent sources. We here aim at developing a differential extension of that BSS method.

2.1.1 Mixing and extraction stages

We consider the configuration involving N source signals $s_j(n)$ and P observed signals $x_p(n)$, with $P<N$. The mixing stage is assumed to be complex and convolutive: each entry of the mixing matrix $A(z)$ is a complex-valued, possibly non-causal, Moving Average (MA) filter. Each propagation channel from source j to observation p is therefore represented by the transfer function

$$A_{pj}(z)=\sum_{l=-L_1}^{L_2} a_{pj}(l)z^{-l}, \tag{2}$$

where the coefficients $a_{pj}(l)$ are complex-valued. The parameters $L_1 \geq 0$ and $L_2 \geq 0$ define the highest allowed order for all these filters.

Each output signal $y(n)$ of the extraction stage of our BSS method is computed by a feedforward structure, with P inputs and thus P extraction filters, whose transfer functions are denoted $B_p(z)$, with $p = 1, \ldots, P$. These filters are possibly non-causal and MA[2], i.e.

$$B_p(z) = \sum_{l=-L'_1}^{L'_2} b_p(l) z^{-l} \tag{3}$$

with complex-valued coefficients $b_p(l)$, and with $L'_1 \geq 0$ and $L'_2 \geq 0$. This structure allows us to extract one signal at once. The ideal values of the filters $B_p(z)$ are such that this signal then consists of a contribution of one and only one of the P sources of interest, added with contributions of the stationary sources.

For any filters $B_p(z)$, the considered output $y(n)$ of the extraction stage reads

$$\begin{aligned} y(n) &= \sum_{p=1}^{P} [B_p(z)] x_p(n) && (4) \\ &= \sum_{p=1}^{P} \sum_{j=1}^{N} [B_p(z) A_{pj}(z)] s_j(n) && (5) \\ &= \sum_{j=1}^{N} [C_j(z)] s_j(n) && (6) \end{aligned}$$

where[3]

$$\begin{aligned} C_j(z) &= \sum_{p=1}^{P} B_p(z) A_{pj}(z) && (7) \\ &= \sum_{l=-L''_1}^{L''_2} c_j(l) z^{-l} && (8) \end{aligned}$$

with $L''_1 = L_1 + L'_1$ and $L''_2 = L_2 + L'_2$. Eq. (6) and (8) yield

$$y(n) = \sum_{j=1}^{N} \sum_{l=-L''_1}^{L''_2} c_j(l) s_j(n-l). \tag{9}$$

The set of filters $C_j(z)$, with $j = 1 \ldots N$, represents the combined effects of propagation and extraction for $y(n)$. These filters may be defined either in terms of their transfer functions $C_j(z)$ or impulse response coefficients $c_j(l)$, as shown by (8).

[2]For the sake of simplicity, we only consider Finite Impulse Response (FIR) mixing and separating filters in this paper. Infinite Impulse Response (IIR) filters e.g. correspond to setting $L'_1 \to +\infty$ and $L'_2 \to +\infty$. Conversely, FIR filters may be considered as approximations of IIR filters. They therefore only make it possible to approximately reach separating points, except in specific cases [19]. The approximation thus introduced can be made negligible by e.g. choosing L'_1 and L'_2 to be high enough. This phenomenon is therefore ignored hereafter. For more details on this topic, the reader may refer to [19].

[3]The brackets "[ ]" around each filter transfer function $B_p(z)$ in (4) mean that we here consider the operator defined by this filter and $y(n)$ is the result obtained when applying this operator to the corresponding signal $x_p(n)$ and summing over p. This compact notation stands for the actual time-domain convolution of $x_p(n)$ by the impulse response of the considered filter. The same notation is used in the next two equations.

We here assume that the source signals are zero-mean, temporally and mutually independent. Moreover, based on our above-defined differential BSS concept, the overall set of N sources is assumed to consist of[4]:

1. P sources of interest $s_j(n)$, with $j = 1, \ldots, P$. Their assumed properties with respect to stationarity may be defined as follows. In the theoretical analysis provided below, we consider two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. Each of these domains $\mathcal{D}_i$ consists of a bounded time interval around a time n_i: we e.g. take into account the set of adjacent times $(n_i - l)$, with $l = -L_1'' \ldots L_2''$, when considering the values of s_j associated to $y(n_i)$ in (9). Each source signal is assumed to have the same statistical properties over all these times around n_i. This corresponds to short-term stationarity. We select the gap between the times n_i to be significantly larger than $L_1'' + L_2''$. Each source signal of interest is then assumed to have different statistical properties in the resulting time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. This corresponds to long-term non-stationarity.

2. $N-P$ noise sources $s_j(n)$, with $j = P+1, \ldots, N$. Each of these sources is supposed to have the same statistical properties for any time n, thus exhibiting both short-term and long-term stationarity.

2.1.2 Differential cumulants

Since each source is zero-mean, (9) yields $E[y(n)] = 0$, where $E[.]$ stands for expectation. We now derive the expressions of the 2^{nd} and 4^{th}-order zero-lag output cumulants with respect to the source cumulants, for any given time n. Denoting complex conjugates with the superscript *, the 2^{nd}-order zero-lag cumulant (or power) of the zero-mean signal $y(n)$ reads

$$\begin{aligned}
CUM_2(y,n) &= E[y(n)y^*(n)] && (10)\\
&= E\left[\sum_{j=1}^{N}\sum_{l=-L_1''}^{L_2''} c_j(l)s_j(n-l)\sum_{k=1}^{N}\sum_{m=-L_1''}^{L_2''} c_k^*(m)s_k^*(n-m)\right] && (11)\\
&= \sum_{j=1}^{N}\sum_{l=-L_1''}^{L_2''}\sum_{k=1}^{N}\sum_{m=-L_1''}^{L_2''} c_j(l)c_k^*(m)E\left[s_j(n-l)s_k^*(n-m)\right]. && (12)
\end{aligned}$$

Since the sources s_j are zero-mean, temporally and mutually independent, (12) becomes

$$\begin{aligned}
CUM_2(y,n) &= \sum_{j=1}^{N}\sum_{l=-L_1''}^{L_2''} c_j(l)c_j^*(l)E\left[s_j(n-l)s_j^*(n-l)\right] && (13)\\
&= \sum_{j=1}^{N}\sum_{l=-L_1''}^{L_2''} c_j(l)c_j^*(l)CUM_2(s_j, n-l). && (14)
\end{aligned}$$

Let n be equal to the time n_i associated to one of the considered time domains $\mathcal{D}_i$. Then, due to the above-mentioned stationarity properties of the sources s_j, the term $CUM_2(s_j, n-l)$ in (14) takes the same value for all time indices $(n-l)$ inside $\mathcal{D}_i$. It

[4] The indices here resp. assigned to the stationary and non-stationary sources are only selected for the sake of simplicity: the relevance of our method does not depend on these indices.

is therefore denoted as $CUM_2(s_j, \mathcal{D}_i)$ hereafter. Consequently, $CUM_2(y, n_i)$ can also be denoted as $CUM_2(y, \mathcal{D}_i)$ and reads

$$CUM_2(y, \mathcal{D}_i) = \sum_{j=1}^{N} CUM_2(s_j, \mathcal{D}_i) \sum_{l=-L_1''}^{L_2''} c_j(l) c_j^*(l). \tag{15}$$

Now consider the 4^{th}-order zero-lag cumulant (or non-normalized kurtosis) of $y(n)$, with two conjugate terms, i.e. [13]

$$CUM_4(y, n) = CUM(y(n), y(n), y^*(n), y^*(n)). \tag{16}$$

Using the multilinearity properties of cumulants and their nullity for independent random variables, one derives in the same way as above

$$CUM_4(y, \mathcal{D}_i) = \sum_{j=1}^{N} CUM_4(s_j, \mathcal{D}_i) \sum_{l=-L_1''}^{L_2''} c_j^2(l) [c_j^*(l)]^2. \tag{17}$$

We now introduce the 2^{nd}-order zero-lag differential cumulant (or differential power) of $y(n)$ associated to the two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$, that we define as[5]

$$\Delta CUM_2(y) \quad = \quad CUM_2(y, \mathcal{D}_2) - CUM_2(y, \mathcal{D}_1). \tag{18}$$

Eq. (15) then yields

$$\Delta CUM_2(y) \quad = \quad \sum_{j=1}^{N} \Delta CUM_2(s_j) \sum_{l=-L_1''}^{L_2''} c_j(l) c_j^*(l), \tag{19}$$

where $\Delta CUM_2(s_j)$ is defined in the same way as in (18). The 4^{th}-order zero-lag differential cumulant (or differential non-normalized kurtosis) of $y(n)$ associated to $\mathcal{D}_1$ and $\mathcal{D}_2$ is defined by using the same approach and (17) yields

$$\begin{aligned} \Delta CUM_4(y) &= CUM_4(y, \mathcal{D}_2) - CUM_4(y, \mathcal{D}_1) & (20) \\ &= \sum_{j=1}^{N} \Delta CUM_4(s_j) \sum_{l=-L_1''}^{L_2''} c_j^2(l) [c_j^*(l)]^2. & (21) \end{aligned}$$

Let us now take into account that $s_{P+1}(n)$ to $s_N(n)$ are long-term stationary. The standard 2^{nd}-order cumulant $CUM_2(s_j, \mathcal{D}_i)$ of each of the sources $s_j(n)$ with $j = P+1, \ldots, N$ then takes the same values for $\mathcal{D}_i = \mathcal{D}_1$ and $\mathcal{D}_i = \mathcal{D}_2$, so that $\Delta CUM_2(s_j) = 0$. The same phenomenon occurs for 4^{th}-order cumulants.

On the contrary, $s_1(n)$ to $s_P(n)$ are long-term non-stationary. More precisely, we assume that they have non-zero 2^{nd} and 4^{th}-order differential cumulants for the considered time domains $\mathcal{D}_1$ and $\mathcal{D}_2$, i.e. that the following conditions are met: [6]

[5] $\Delta CUM_2(y)$ depends on the selected time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. We omit these domains in the notation $\Delta CUM_2(y)$ and in subsequent expressions, for the sake of readability and because a single application of our differential method only uses a single couple of domains $\mathcal{D}_1$ and $\mathcal{D}_2$.

[6] Note that, similarly, the stationarity constraints on sources $s_j(n)$ with $j = P+1, \ldots, N$ in fact only concern their 2^{nd} and 4^{th}-order differential cumulants.

Assumption 1 $\Delta CUM_2(s_j) \neq 0$, $\forall\, j = 1, \ldots, P$.

Assumption 2 $\Delta CUM_4(s_j) \neq 0$, $\forall\, j = 1, \ldots, P$.

Eq. (19) and (21) then reduce to

$$\Delta CUM_2(y) = \sum_{j=1}^{P} \Delta CUM_2(s_j) \sum_{l=-L_1''}^{L_2''} c_j(l) c_j^*(l) \tag{22}$$

$$\Delta CUM_4(y) = \sum_{j=1}^{P} \Delta CUM_4(s_j) \sum_{l=-L_1''}^{L_2''} c_j^2(l) [c_j^*(l)]^2. \tag{23}$$

This shows explicitly that the 2^{nd} and 4^{th}-order differential cumulants of the output signal $y(n)$ only depend on the non-stationary sources.

2.1.3 Differential normalized kurtosis: definition and global maximum

Given two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$, we define the corresponding differential normalized kurtosis of any signal y as

$$k_D(y) = \frac{\Delta CUM_4(y)}{[\Delta CUM_2(y)]^2}. \tag{24}$$

We here consider the above-defined extracted signal $y(n)$ and we investigate the global maximum of the absolute value of its differential normalized kurtosis $k_D(y)$, with respect to the filters $C_j(z)$. To this end, we derive from (22) and (23)

$$|\Delta CUM_2(y)| = \left| \sum_{j=1}^{P} \Delta CUM_2(s_j) \sum_{l=-L_1''}^{L_2''} c_j(l) c_j^*(l) \right| \tag{25}$$

and

$$\begin{aligned} |\Delta CUM_4(y)| &= \left| \sum_{j=1}^{P} \Delta CUM_4(s_j) \sum_{l=-L_1''}^{L_2''} [c_j(l)]^2 [c_j^*(l)]^2 \right| && (26) \\ &= \left| \sum_{j=1}^{P} \Delta CUM_4(s_j) \sum_{l=-L_1''}^{L_2''} |c_j(l)|^4 \frac{|\Delta CUM_2(s_j)|^2}{|\Delta CUM_2(s_j)|^2} \right| && (27) \\ &\leq \sum_{j=1}^{P} |k_D(s_j)| \sum_{l=-L_1''}^{L_2''} |c_j(l)|^4 \, |\Delta CUM_2(s_j)|^2. && (28) \end{aligned}$$

We also have

$$\sum_{j=1}^{P} |k_D(s_j)| \sum_{l=-L_1''}^{L_2''} |c_j(l)|^4 \, |\Delta CUM_2(s_j)|^2 \tag{29}$$

$$= |k_D(s_j)|_{max} \sum_{j=1}^{P} \frac{|k_D(s_j)|}{|k_D(s_j)|_{max}} \sum_{l=-L_1''}^{L_2''} |c_j(l)|^4 \, |\Delta CUM_2(s_j)|^2 \tag{30}$$

with $|k_D(s_j)|_{max} = max_{1\leq j\leq P} |k_D(s_j)|$. We thus have

$$\frac{|k_D(s_j)|}{|k_D(s_j)|_{max}} \leq 1 \quad \forall\, j = 1, \ldots, P. \tag{31}$$

Combining this with (28) and (30) then yields

$$|\Delta CUM_4(y)| \quad \leq \quad |k_D(s_j)|_{max} \sum_{j=1}^{P} \sum_{l=-L_1''}^{L_2''} |c_j(l)|^4 \, |\Delta CUM_2(s_j)|^2. \tag{32}$$

Using (25), we get

$$\frac{|\Delta CUM_4(y)|}{|\Delta CUM_2(y)|^2} \leq |k_D(s_j)|_{max} \frac{\sum_{j=1}^{P} \sum_{l=-L_1''}^{L_2''} |c_j(l)|^4 \, |\Delta CUM_2(s_j)|^2}{\left|\sum_{j=1}^{P} \sum_{l=-L_1''}^{L_2''} |c_j(l)|^2 \, \Delta CUM_2(s_j)\right|^2}. \tag{33}$$

The 2^{nd}-order differential cumulants of the sources of interest are supposed to meet Assumption 1. In addition, we consider the following case hereafter (at the end of this section, we provide a method for selecting time domains $\mathcal{D}_1$ and $\mathcal{D}_2$ such that this assumption is met):

Assumption 3 *The 2^{nd}-order differential source cumulants $\Delta CUM_2(s_j)$, with $j = 1, \ldots, P$, have the same sign.*

Under the above assumptions, we have

$$\sum_{j=1}^{P} \sum_{l=-L_1''}^{L_2''} |c_j(l)|^4 \, |\Delta CUM_2(s_j)|^2 \leq \left|\sum_{j=1}^{P} \sum_{l=-L_1''}^{L_2''} |c_j(l)|^2 \, \Delta CUM_2(s_j)\right|^2 \tag{34}$$

with equality if and only if

$$\text{there exist } j_0 \in \{1, \ldots, P\} \text{ and } l_0 \in \{-L_1'', \ldots, L_2''\} \text{ such that } c_j(l) = d\delta(j-j_0)\delta(l-l_0), \tag{35}$$

where: (i) d is an arbitrary complex constant, (ii) $\delta(j - j_0) = 1$ if $j = j_0$, and $\delta(j - j_0) = 0$ otherwise and (iii) $\delta(l - l_0)$ is defined in the same way as $\delta(j - j_0)$.

Using Eq. (24), (33) and (34), we get

$$|k_D(y)| = \frac{|\Delta CUM_4(y)|}{|\Delta CUM_2(y)|^2} \leq |k_D(s_j)|_{max} \tag{36}$$

with equality if and only if (35) is met and under the condition that the value j_0 be such that $|k_D(s_{j_0})| = |k_D(s_j)|_{max}$. When these conditions are met, (9) shows that the output of the extraction stage reads

$$y(n) = ds_{j_0}(n - l_0) + \sum_{j=P+1}^{N} c_j(n) * s_j(n). \tag{37}$$

This output signal then contains a contribution from only one of the P sources of interest, added with a convolutive mixture of the $N-P$ stationary sources. We thus exactly achieve the above-defined partial source separation for one of the sources of interest.

We thus showed, under the condition that the 2^{nd}-order differential cumulants of the sources of interest have the same sign, that when the global maximum of the function

$$|k_D(y)| = k_D(y).sign(k_D(y)) \tag{38}$$

is obtained, the output signal $y(n)$ only contains one of the sources of interest (plus noise sources), and that this source is one with the highest value of $|k_D(y)|$. That suggests that $|k_D(y)|$ might then used as a cost function to achieve partial BSS from the considered underdetermined convolutive mixtures. The suitability of this cost function also depends on its other extrema, however, as explained below in Section 2.1.4. Before proceeding to this topic, it should be noted that if the 2^{nd}-order differential cumulants of the sources of interest do not have the same sign, Eq. (34) and thus (36) are no longer true. In that case, the global maximum of $|k_D(y)|$ cannot be used as a BSS criterion, because it does not guarantee anymore partial source separation. The signs of the 2^{nd}-order differential cumulants of the sources of interest are therefore essential for the validity of the approach that we propose in this paper and they should be checked when using this approach. This may be done by applying the method that we already used in [8]: that method makes it possible to derive only from the observations the signs of the 2^{nd}-order differential cumulants of the sources of interest for any given time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. This makes it possible to test various domains and to select two domains, if any, such that the 2^{nd}-order differential cumulants of the sources of interest have the same sign.

2.1.4 Local extrema

In practice, the extraction filters $B_p(z)$ are adapted by an optimization algorithm which aims at maximizing $|k_D(y)|$. We just showed that the global maximum of this cost function corresponds to a valid point, i.e. to an output signal $y(n)$ containing only one of the sources of interest. However, optimization algorithms may converge towards local extrema of this cost function, if any. We must therefore also determine whether the other extrema of this cost function $|k_D(y)|$ also correspond to partial source separation. This important topic requires detailed calculations and is therefore presented in Appendix A. Still under Assumption 3, this yields the following result concerning the gradient of $|k_D(y)|$ with respect to the (scaled) coefficients $\bar{c}_j(l)$ of the global filters associated to the sources of interest: the only points where this gradient is equal to zero and which are extrema of $|k_D(y)|$ correspond to situations when only one coefficient $\bar{c}_j(l)$ with $j \in \{1, \ldots, P\}$ and $l \in \{-L_1'', \ldots, L_2''\}$ is non-zero. These points are therefore all defined by Eq. (35), so that they also yield an output signal defined by (37), i.e. a signal which contains a contribution from only one of the P sources of interest, added with a mixture of the $N-P$ stationary sources. At these points, we therefore indeed exactly achieve the above-defined partial source separation for one of the sources of interest.

2.1.5 Interpretation with respect to differences in source statistical properties

We showed above that the local maximization of $|k_D(y)|$ leads to global filters $C_j(z)$ which satisfy (35). The extracted signal (37) then contains only one of the P sources which have different statistical properties between the considered two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$ and a combination of the $N-P$ stationary sources, which have no influence on the identification of the filters $C_j(z)$ associated to the P non-stationary sources.

This can be compared to the classical results concerning Gaussian sources in higher-order-statistic BSS methods for determined mixtures: Gaussian signals have zero 4^{th}-order cumulants and they have no influence on the 4^{th}-order cumulant of an output signal of a linear BSS system, so that (i) Gaussian sources cannot be extracted using the 4^{th}-order cumulant as a BSS criterion and (ii) additive Gaussian noise contributions in the observations do not have any (theoretical) impact on the identification of the mixing system with that BSS criterion, but they yield corresponding noise contributions in the outputs of the BSS system. Such classical higher-order BSS methods thus yield distinct behaviors for sources which have different statistical properties from the point of view of Gaussianity (more precisely, from the point of view of the nullity/non-nullity of their 4^{th}-order cumulants). Gaussian sources are thus inherently invisible in the identification stage of such classical BSS methods. Our approach based on differential normalized kurtosis has similar properties, but with respect to differential cumulants instead of classical cumulants. Stationary sources are invisible in its identification stage, thus allowing extraction filters to converge towards values which achieve the (partial) separation of all non-stationary sources in the underdetermined configuration, under the above-mentioned assumptions. Our differential BSS method thus yields distinct behaviors for sources which have different statistical properties from the point of view of stationarity (more precisely, from the point of view of the nullity/non-nullity of their 2^{nd} and 4^{th}-order differential cumulants).

2.2 New coloration stage and overall BSS method

The method that we introduced above covers the first stage of our overall BSS approach, i.e. it yields a single output signal which extracts one of the sources of interest, with added noise source contributions. We then aim at extracting the other sources of interest. To this end, we use a deflation scheme. The resulting overall BSS method consists of the following steps:

1. Extract one source of interest, as explained above i.e. by adapting the filters $B_p(z)$ so as to maximize $|k_D(y)|$ (a corresponding practical optimization algorithm is described below in Section 3). This yields an output signal $y(n)$.

2. Estimate the contributions of this extracted source in the observed signals. This is achieved by a coloration stage, which consists in deriving a set of coloration filters $H_p(z)$. Each such filter corresponds to an observation $x_p(n)$ and to the considered extracted signal $y(n)$. It is applied to $y(n)$, and the resulting signal $h_p(n) * y(n)$ contains a filtered version of the extracted source. We aim at making this contribution equal to the contribution in $x_p(n)$ of the extracted source, by adequately selecting $h_p(n)$. The coloration method used to this end is presented further in this section.

3. Cancel the contributions of the extracted source in the observed signals, by computing the modified observations

$$x'_p(n) = x_p(n) - h_p(n) * y(n). \tag{39}$$

   We thus obtain a new BSS configuration, where the number of sources of interest has been decreased by one.

4. If the configuration obtained at this stage still involves more than one source of interest, go to Step 1, i.e. apply this procedure again to the modified observations (the number of observations may be decreased by one). Otherwise, end this deflation procedure.

Such deflation approaches have been used in the literature for determined mixtures, e.g. in [7],[11],[19]. They require modifications in our underdetermined context, not only in the extraction stage that we presented above, but also in the coloration stage, which is the last building block that we have to develop in order to obtain our overall deflation-based BSS method. As explained above, this stage aims at identifying the filters $H_p(z)$. The methods used to this end in the literature are based on the second-order statistics of the extracted signal $y(n)$ and/or of each observation $x_p(n)$, such as their cross-correlation (see e.g. [19]). They cannot be used directly here, because these signals contain contributions from all noise sources, which yield unacceptable second-order terms. We therefore introduce a modified, differential, coloration method. Our differential extraction stage yields the extracted signal $y(n)$ defined by (37), where d and l_0 are arbitrary constants. In the following calculations, we set

$$d = 1 \text{ and } l_0 = 0. \tag{40}$$

This case is considered without loss of generality, since it only consists in redefining $H_p(z)$ so that it includes the scale factor d and time shift l_0 of s_{j_0} in (37) when taking into account the signal $h_p(n) * y(n)$ hereafter. Under the assumption that the sources are independent and zero-mean, the 2^{nd}-order zero-lag cumulants (or powers) of the signals $x'_p(n)$ in the time domains $\mathcal{D}_i$ read

$$E\left[\left|x'_p(n)\right|^2\right]_{\mathcal{D}_i} = E\left[\left|x_p(n) - h_p(n) * y(n)\right|^2\right]_{\mathcal{D}_i} \tag{41}$$

$$\begin{aligned} &= E[|(a_{pj_0}(n) - h_p(n)) * s_{j_0}(n)|^2|]_{\mathcal{D}_i} \\ &+ \sum_{\substack{j \leq P \\ j \neq j_0}} E[|a_{pj}(n) * s_j(n)|^2]_{\mathcal{D}_i} \\ &+ \sum_{j=P+1}^{N} E[|(a_{pj}(n) - h_p(n) * c_j(n)) * s_j(n)|^2]_{\mathcal{D}_i}. \end{aligned} \tag{42}$$

We then introduce the 2^{nd}-order zero-lag differential cumulants (or differential powers) of these signals between the domains $\mathcal{D}_1$ and $\mathcal{D}_2$, which are denoted ΔW_p below, i.e.

$$\Delta W_p = \Delta CUM_2(x'_p) \tag{43}$$

$$= E\left[\left|x'_p(n)\right|^2\right]_{\mathcal{D}_2} - E\left[\left|x'_p(n)\right|^2\right]_{\mathcal{D}_1}. \tag{44}$$

The differential cumulants of sources s_{P+1} to s_N are equal to zero. Therefore, we have

$$\begin{aligned} \Delta W_p = \; & E[|(a_{pj_0}(n) - h_p(n)) * s_{j_0}(n)|^2|]_{\mathcal{D}_2} - E[|(a_{pj_0}(n) - h_p(n)) * s_{j_0}(n)|^2|]_{\mathcal{D}_1} \\ & + \sum_{\substack{j \leq P \\ j \neq j_0}} E[|a_{pj}(n) * s_j(n)|^2]_{\mathcal{D}_2} - \sum_{\substack{j \leq P \\ j \neq j_0}} E[|a_{pj}(n) * s_j(n)|^2]_{\mathcal{D}_1}. \end{aligned} \tag{45}$$

ΔW_p is therefore independent from stationary sources. We now analyze the extrema of this differential cumulant. Let us consider possibly non-causal MA filters $H_p(z)$ defined by[7]

$$H_p(z) = \sum_{l=-L_1'''}^{L_2'''} h_p(l) z^{-l} \tag{46}$$

[7]Remember that the lags l in $h_p(l)$ here include the time shift l_0. The orders L_1''' and L_2''' should be set to high enough values in order to ensure (47) when the highest accepted time shift l_0 is taken into account.

with complex coefficients $h_p(l)$ and

$$L_1''' \geq L_1 \text{ and } L_2''' \geq L_2. \tag{47}$$

Using the same approach as in Section 2.1.2, Eq. (45) becomes

$$\begin{aligned}\Delta W_p = \ & \Delta CUM_2(s_{j_0}) \textstyle\sum_{l=-L_1'''}^{L_2'''} |(a_{pj_0}(l) - h_p(l))|^2 \\ & + \sum_{\substack{j \leq P \\ j \neq j_0}} E[|a_{pj}(n) * s_j(n)|^2]_{\mathcal{D}_2} - \sum_{\substack{j \leq P \\ j \neq j_0}} E[|a_{pj}(n) * s_j(n)|^2]_{\mathcal{D}_1}.\end{aligned} \tag{48}$$

Considering that the only variable here is the filter $H_p(z)$ that we are adapting, we can rewrite this expression as

$$\Delta W_p \quad = \Delta CUM_2(s_{j_0}) \textstyle\sum_{l=-L_1'''}^{L_2'''} |(a_{pj_0}(l) - h_p(l))|^2 + Ct, \tag{49}$$

where Ct is a constant value. The cost function ΔW_p is therefore quadratic with respect to the coefficients $h_p(l)$ and then only has one extremum, which is a maximum or a minimum depending on the sign of $\Delta CUM_2(s_{j_0})$. This extremum is reached when

$$h_p(l) = a_{pj_0}(l), \ \forall \ l \ \in \{-L_1''', \ldots, L_2'''\}. \tag{50}$$

Combining this condition with (37) and (40) shows that the contribution of the extracted source $s_{j_0}(n)$ in the colored output signal $h_p(n) * y(n)$ is then equal to the contribution of this source in observation $x_p(n)$. This proves that a method for performing the coloration operation required in our underdetermined configuration consists in looking for the only extremum of the cost function which consists of the above-defined 2^{nd}-order differential cumulant, or differential power, ΔW_p.

Eq. (41) shows that the power of $x'_p(n)$ may also be interpreted as the mean square error between the signals $x_p(n)$ and $h_p(n)*y(n)$. Due to Eq. (44), the differential power of $x'_p(n)$ is then the differential mean square error between these signals $x_p(n)$ and $h_p(n) * y(n)$. The criterion used in our coloration stage, i.e. the optimization of ΔW_p, may therefore also be interpreted as the optimization of the differential mean square error between the above signals.

2.3 Extension to colored sources

For the sake of simplicity, we only considered temporally independent sources in the above description of all stages of our BSS method. More generally speaking, the proposed approach is directly applicable to sources $S_j(z)$ which are filtered versions of temporally independent processes $U_j(z)$, i.e.

$$S_j(z) = F_j(z)U_j(z) \tag{51}$$

where $F_j(z)$ are MA filters with complex, possibly non-causal, impulse responses. The same extraction stage as above is then used and, in its analysis, the source filters $F_j(z)$ are combined with the mixing filters, so that we get back to the same configuration as above but with respect to the temporally independent processes $U_j(z)$ instead of the source signals. Thus, when partial separation is reached, the extracted signal $y(n)$ restores the temporally independent process $U_j(z)$ associated to one of the sources of interest (again with added noise source contributions). The above-defined coloration stage is then used in this general case too. It still yields the contributions in all observations of the source whose process $U_j(z)$ was extracted. These contributions may then be subtracted from all observations in the framework of the same deflation approach as above.

3 Optimization algorithms

The differential BSS method proposed in this paper is based on the above-defined deflation procedure. We here describe the algorithms that we use to optimize the two cost functions resp. involved in the extraction and coloration stages of this procedure. The corresponding tools are defined in Appendix B.

3.1 Extraction stage

3.1.1 Cost function

The cost function to be maximized in the extraction stage of our BSS method is the absolute value of the differential kurtosis of the signal $y(n)$, which may be expressed as follows, based on (24),

$$
\begin{aligned}
|k_D(y)| &= \left| \frac{\Delta CUM_4(y)}{[\Delta CUM_2(y)]^2} \right| \\
&= sign(\Delta CUM_4(y)).\frac{\Delta CUM_4(y)}{[\Delta CUM_2(y)]^2}. \qquad (52)
\end{aligned}
$$

We then derive, from (18), (20) and the expression of the 4^{th}-order cumulant of a complex zero-mean random variable [12],[19],

$$
\Delta CUM_2(y) = E\left[y(k)y^*(k)\right]_{\mathcal{D}_2} - E\left[y(k)y^*(k)\right]_{\mathcal{D}_1} \qquad (53)
$$

$$
\begin{aligned}
\Delta CUM_4(y) &= E\left[y^2(k).[y^*(k)]^2\right]_{\mathcal{D}_2} - 2E\left[y(k).y^*(k)\right]^2_{\mathcal{D}_2} \\
&\quad -E\left[y^2(k)\right]_{\mathcal{D}_2} E\left[y^2(k)\right]^*_{\mathcal{D}_2} \\
&\quad -E\left[y^2(k).[y^*(k)]^2\right]_{\mathcal{D}_1} + 2E\left[y(k).y^*(k)\right]^2_{\mathcal{D}_1} \\
&\quad +E\left[y^2(k)\right]_{\mathcal{D}_1} E\left[y^2(k)\right]^*_{\mathcal{D}_2}. \qquad (54)
\end{aligned}
$$

3.1.2 Derivatives of $k_D(y)$

In order to simplify the equations, we use the following notations

$$
\begin{aligned}
A &= \nabla_{b_p(l)} CUM_2(y) \\
&= E\left[y^*(k)x_p(k-l)\right] \qquad (55)
\end{aligned}
$$

$$
\begin{aligned}
B &= \nabla_{b_p(l)} CUM_4(y) \\
&= 2E\left[x_p(k-l)y^*(k).|y(k)|^2\right] - 4E\left[|y(k)|^2\right] E\left[y^*(k).x_p(k-l)\right] \\
&\quad -2E\left[y^{*2}(k)\right] E\left[x_p(k-l)y(k)\right] \qquad (56)
\end{aligned}
$$

$$
C = \nabla_{b_p(l)b_p(l)} CUM_2(y) = 0 \qquad (57)
$$

$$
\begin{aligned}
D &= \nabla_{b_p(l)b_p^*(l)} CUM_2(y) \\
&= E[|x_p(k-l)|^2] \qquad (58)
\end{aligned}
$$

$$
\begin{aligned}
&= D^* \\
E &= \nabla_{b_p(l)b_p(l)} CUM_4(y) \\
&= 2E[x_p^2(k-l)y^{*2}(k)] - 4E[y^*(k)x_p(k-l)]^2 \\
&\quad -2E[y^*(k)]E[x_p^2(k-l)]
\end{aligned} \tag{59}
$$

$$
\begin{aligned}
F &= \nabla_{b_p(l)b_p^*(l)} CUM_4(y) \\
&= 4E[|x_p^*(k-l)y^*(k)|^2] - 4|E[x_p(k-l)y^*(k)]|^2 \\
&\quad -4E[|y(k)|^2]E[|x_p(k-l)|^2] - 4|E[x_p(k-l)y(k)]|^2.
\end{aligned} \tag{60}
$$

The differential versions of the derivatives (55) to (60) may then be expressed as follows, where the subscripts $\mathcal{D}_1$ and $\mathcal{D}_2$ refer to the time domains where the expectations are considered,

$$
\begin{aligned}
\Delta A &= \nabla_{b_p(l)} \Delta CUM_2(y) \\
&= A_{\mathcal{D}_2} - A_{\mathcal{D}_1}
\end{aligned} \tag{61}
$$

$$
\begin{aligned}
\Delta B &= \nabla_{b_p(l)} \Delta CUM_4(y) \\
&= B_{\mathcal{D}_2} - B_{\mathcal{D}_1}
\end{aligned} \tag{62}
$$

$$
\Delta C = \nabla_{b_p(l)b_p(l)} \Delta CUM_2(y) = 0 \tag{63}
$$

$$
\begin{aligned}
\Delta D &= \nabla_{b_p(l)b_p^*(l)} \Delta CUM_2(y) \\
&= D_{\mathcal{D}_2} - D_{\mathcal{D}_1} \\
\Delta E &= \nabla_{b_p(l)b_p(l)} \Delta CUM_4(y) \\
&= E_{\mathcal{D}_2} - E_{\mathcal{D}_1}
\end{aligned} \tag{64}
$$

$$
\begin{aligned}
\Delta F &= \nabla_{b_p(l)b_p^*(l)} \Delta CUM_4(y) \\
&= F_{\mathcal{D}_2} - F_{\mathcal{D}_1}.
\end{aligned} \tag{65}
$$

First-order derivative $k'_D(y)$

Considering the first-order derivative of $k_D(y)$ with respect to a single complex extraction coefficient $b_p(l)$, Eq. (130) here yields

$$
k'_D(y) = 2\nabla_{b_p^*(l)} k_D(y). \tag{66}
$$

Detailed calculations then show that, using the above-defined notations, we have

$$
\nabla_{b_p^*(l)} k_D(y) = \frac{\Delta CUM_2^2(y)\Delta B^* - 2\Delta CUM_4(y)\Delta CUM_2(y)\Delta A^*}{[\Delta CUM_2(y)]^4}. \tag{67}
$$

Second-order derivative $k''_D(y)$

In order to use the Newton-like algorithm defined in Appendix B, we also need the second-order derivative of $k_D(y)$ with respect to a single complex extraction coefficient $b_p(l)$. Eq. (135) here yields

$$
\begin{aligned}
k''_D(y) &= 2\mathcal{R}e\{\nabla^2_{b_p(l)b_p^*(l)} k_D(y) + \nabla^2_{b_p(l)b_p(l)} k_D(y)\} \\
&\quad +2i\mathcal{R}e\{\nabla^2_{b_p(l)b_p^*(l)} k_D(y) - \nabla^2_{b_p(l)b_p(l)} k_D(y)\}.
\end{aligned} \tag{68}
$$

Additional calculations then yield

$$\begin{aligned}
\nabla^2_{b_p(l)b_p^*(l)} k_D(y) &= \{[-2\Delta A \Delta B^* + \Delta CUM_2(y)\Delta F^* - 2\Delta B \Delta A^* \\
&\quad -2\Delta CUM_4(y)\Delta D^*]\,\Delta CUM_2(y) \\
&\quad +6\Delta CUM_4(y)|\Delta A|^2\Big\}\,[\Delta CUM_2(y)]^{-4} \\
&= \{[-2\mathcal{R}e\{\Delta A \Delta B^*\} + \Delta CUM_2(y)\Delta F^* \\
&\quad -2\Delta CUM_4(y)\Delta D]\,\Delta CUM_2(y) \\
&\quad +6\Delta CUM_4(y)|\Delta A|^2\Big\}\,[\Delta CUM_2(y)]^{-4} \\
\nabla^2_{b_p^*(l)b_p^*(l)} k_D(y) &= \{[-4\Delta A^* \Delta B^* + \Delta CUM_2(y)\Delta E^*]\,\Delta CUM_2(y) \\
&\quad +6\Delta CUM_4(y)\Delta A^{*2}\Big\}\,[\Delta CUM_2(y)]^{-4}\,. \qquad (69)
\end{aligned}$$

The extraction stage therefore operates as follows:

1. Initialize all extraction coefficients $b_p(l)$ according to

$$b_p(l) = \delta(l) \qquad (70)$$

   as in [19]. This is a reasonable choice, since it corresponds to

$$y(n) = \sum_{p=1}^{P} x_p(n). \qquad (71)$$

2. Independently adapt each of these complex coefficients $b_p(l)$ according to the modified Newton algorithm (141), applied to $|k_D(y)|$, using the above expressions. [8]

3.2 Coloration stage

As shown above, the coloration filters $H_p(z)$ are updated so as to reach the only extremum of the cost function ΔW_p defined in (43). This extremum is searched using the complex version of Newton's algorithm once again. Since this extremum may be a minimum or maximum, we use the original form (137) of this algorithm, for each coloration coefficient $h_p(k)$. The derivatives to be used in (137) are expressed in the same way as above with respect to the corresponding complex gradient terms, which may here be shown to read as follows

$$\begin{aligned}
\nabla_{h_p^*(k)} \Delta W_p &= -E[y^*(n-k)(x_p(n) - h_p(n) * y(n))]_{\mathcal{D}_2} \\
&\quad +E[y^*(n-k)(x_p(n) - h_p(n) * y(n))]_{\mathcal{D}_1} \qquad (72) \\
\nabla^2_{h_p(l)h_p^*(l)} \Delta W_p &= E[|y(n-k)|]_{\mathcal{D}_2} - E[|y(n-k)|]_{\mathcal{D}_1} \qquad (73) \\
\nabla^2_{h_p(l)h_p(l)} \Delta W_p &= 0. \qquad (74)
\end{aligned}$$

These filters are also initialized according to $h_p(k) = \delta(k)$, which corresponds to initially applying no coloration to $y(n)$.

[8] Note that we thus consider each variable $b_p(l)$ separately. Therefore, instead of using the exact Hessian matrix of the cost function, we approximate it by its diagonal. We thus in fact obtain an approximate Newton-like algorithm. The same comment applies to the coloration stage presented below.

4 Experimental results

We ran various tests using different source signals and mixing matrices in a 3-source to 2-observation configuration. In the first set of tests, sources 1 and 2 have different cumulant values in the selected two time domains, whereas source 3 is stationary over these domains. In domain D_1, Source 1 consist of an artificial binary-valued signal, while Source 2 is the sum of such an artificial binary-valued signal and of a Gaussian signal. In domain D_2, Sources 1 and 2 are artificial MSK communication signals. Source 3 is an artificial binary-valued signal in both domains. Each domain contains 30000 samples. Each test is performed with randomly generated mixing filters, and their orders are set to $L_1 = 0$ and $L_2 = 2$. Our differential BSS method is here applied as follows:

1. Starting from observations $x_1(n)$ and $x_2(n)$, the extraction stage yields a signal $y(n)$, which only contains one of the sources of interest[9] (plus the noise source, and some residuals from the other source of interest in practice, due to estimation errors).

2. The coloration stage is then applied to $y(n)$ and to observation $x_1(n)$. This yields an output signal denoted $z_1(n)$ hereafter, which contains the contribution of the extracted source in $x_1(n)$ (plus the noise source ...).

3. The signal $z_1(n)$ is subtracted from $x_1(n)$. This directly yields an output signal denoted $z_2(n)$ hereafter, which contains the contribution of the other source of interest in $x_1(n)$ (plus the noise source ...) and completes our BSS method in this simple setup.

We apply this method with extraction filters $B_p(z)$ and coloration filters $H_p(z)$ containing 4 coefficients, i.e. 3 for their causal part and 1 for their non-causal part. We use 50 iterations in Newton's algorithms for both the extraction and coloration stages.

Performance is here measured as follows. The Signal-to-Interference Ratios (SIR) available from the observations are first used to measure the quality of the inputs of our BSS system. More precisely, we define two types of input SIRs:

1. The "global input SIRs" are defined for each source $s_j(n)$ and each observation $x_i(n)$ as the ratios of the "Signal power" and "Interference power", when the contribution of $s_j(n)$ in $x_i(n)$ is considered as the "Signal" and the contributions of the other two sources in $x_i(n)$ are considered as "Interferences", i.e.

$$SIR_{in}^{glob}(x_i, s_j) = \frac{E\{|a_{ij}(n) * s_j(n)|^2\}}{E\{|x_i(n) - a_{ij}(n) * s_j(n)|^2\}}, \qquad i = 1,\ 2,\ j = 1, \ldots, 3 \tag{75}$$

$$= \frac{E\{|a_{ij}(n) * s(j)|^2\}}{E\{|\sum_{k \in \{1,\ldots,3\},\ k \neq j} a_{ik}(n) * s_k(n)|^2\}}. \tag{76}$$

2. The "partial input SIRs" are defined in the same way as the above global input SIRs, except that all contributions of the stationary source $s_3(n)$ are removed, in order to only consider the two non-stationary sources that we aim at separating one from the

[9]As explained above, we obtain a time-shifted and scaled version of that source.

other. i.e.

$$SIR_{in}^{part}(x_i, s_j) = \frac{E\{|a_{ij}(n) * s(j)|^2\}}{E\{|a_{ik}(n) * s_k(n)|^2\}}, \qquad i = 1,\ 2,\ j = 1,\ 2,$$
$$k = 1,\ 2,\ k \neq j. \tag{77}$$

The performance improvement achieved by our BSS method is then measured by comparing the above input SIRs to the output SIRs measured from the outputs of our BSS system. We here consider two types of output SIRs. Again, both of them are computed with signals where all contributions of the stationary source $s_3(n)$ are removed, in order to only take into account the two non-stationary sources that we aim at separating one from the other. It should be clear that $s_3(n)$ is only removed when computing these performance figures, but is indeed present when we first apply our differential BSS method to identify the mixing filters. The considered output SIRs are:

1. The "estimation output SIRs" associated to each source $s_j(n)$ and each colored output $z'_i(n)$, where $z'_i(n)$ consists of the above-defined signal $z_i(n)$ without its contribution associated to $s_3(n)$. These SIRs are again defined as the ratios of the "Signal power" and "Interference power", where (i) the "Signal" is the ideal value of the output signal $z'_i(n)$ if this signal extracts source $s_j(n)$, which is equal to $a_{1j}(n) * s_j(n)$ as explained above, and (ii) the "Interference" is the difference between this ideal value and the actual signal $z'_i(n)$. These SIRs therefore read

$$SIR_{out}^{est}(z'_i, s_j) \;=\; \frac{E\{|a_{1j}(n) * s_j(n)|^2\}}{E\{|z'_i(n) - a_{1j}(n) * s_j(n)|^2\}}, \qquad i = 1,\ 2,\ j = 1,\ 2. \tag{78}$$

   These SIRs take into account two types of deviations in $z'_i(n)$, i.e. (i) the undesired components in $z'_i(n)$ associated to the *other* source of interest, which really concern the *separation* of one of these sources from the other, (ii) the differences between the filters applied to source $s_j(n)$ in $z'_i(n)$ and in its ideal value $a_{1j}(n) * s_j(n)$. This type of SIR therefore completely measures with which accuracy $z'_i(n)$ estimates the specific filtered version $a_{1j}(n) * s_j(n)$ of $s_j(n)$. This criterion is relevant in applications where any additional filter applied to an extracted source should be considered as performance degradation (e.g. in speech enhancement applications). On the contrary, in other fields only the contribution from the other source of interest should be regarded as degradation, i.e. only the *separation* aspect of the above criterion should be taken into account in performance criteria (this e.g. concerns communication applications where filtered versions of the extracted source are acceptable, because the output of a BSS system is then applied to an equalizer anyway). To handle the latter case, we therefore introduce an alternative output SIR hereafter.

2. Denoting $[z'_i]_{s_j}(n)$ the component of $s_j(n)$ in $z'_i(n)$, we define the "separation output SIRs" associated to each source $s_j(n)$ and each colored output $z'_i(n)$ as

$$SIR_{out}^{sep}(z'_i, s_j) = \frac{E\{|[z'_i]_{s_j}(n)|^2\}}{E\{|[z'_i]_{s_k}(n)|^2\}}, \qquad i = 1,\ 2,\ j = 1,\ 2,$$
$$k = 1,\ 2,\ k \neq j. \tag{79}$$

In addition to the differential kurtotic method suited to underdetermined mixtures that we proposed in this paper, we also apply the classical, i.e. non-differential, kurtotic method only intended for determined mixtures which was introduced in [19]. The results obtained with both methods for three mixing matrices are provided as Tests no. 1 to 3 of Tables 1 to 5. The following conclusions may especially be drawn from these tests. The classical method fails to achieve partial source separation: it yields estimation output SIRs which are most often lower, sometimes slightly higher, than the partial input SIRs (here and below, we take into account the best SIR among the two values associated to both sources, for each observation or output signal). On the contrary, our differential method yields significant performance improvement:

1. It yields estimation output SIRs which range from about 10 to 17 dB, while the partial SIRs of the processed mixed signals range from 2 to 7 dB. This therefore typically corresponds to a 10 dB SIR improvement.

2. The output performance is of course even better if only considering the separation aspect measured by separation output SIRs, which range from 11 to 22 dB.

In order to further analyze the performance of our method, we ran two additional tests. In Test no. 4, we used the same mixing configuration as in Test no. 1, except that the length parameter L_2 of the mixing filters was set to 4 (more precisely, the same mixing filter coefficient values as in Test no. 1 were used for lags 0 to 2, and additional non-zero values were added for lags 3 and 4). Tables 1 to 5 show that our differential method still yields good output SIRs for that higher filter length, while the non-differential method does not. Eventually, in Test no. 5, we used the same configuration as in Test no. 1, except that the stationary Source 3 had a Laplacian distribution. Again, Tables 1 to 5 show that only our differential method yields high output SIRs in this configuration.

5 Conclusion

In this paper, we introduced new criteria and associated algorithms for the Blind Source Separation problem. The proposed approach is designed to handle the difficult case when the source signals are complex-valued and their mixtures are complex-valued, convolutive, and underdetermined (i.e. N sources and $P < N$ observations). It makes it possible to achieve the partial separation of P sources, by taking advantage of the assumed distinct statistical properties of the sources: the P sources of interest should be non-stationary, while the other $(N - P)$ "noise sources" should be stationary. This method is based on the differential BSS concept that we introduced in [8]. It may be seen as a differential extension of the approach proposed by J.K. Tugnait in [19] and it is especially based on the differential normalized kurtosis that we introduce in this paper. We therefore call this BSS method "DNKurt".

More precisely, we considered P convolutive mixtures of a set of N sources, composed (i) of P sources whose 2^{nd} and 4^{th}-order cumulants take different values in two time domains, and which are therefore non-stationary and (ii) of $(N - P)$ sources which are stationary from the point of view of these cumulants. We studied the absolute value $|k_D(y)|$ of the differential normalized kurtosis, considered between these two time domains, of a signal extracted by our BSS system. We demonstrated that all maxima of $|k_D(y)|$ are obtained for filter values which achieve the partial separation of the P non-stationary sources, i.e. which yield an output signal containing contributions from only one of these P sources of interest plus a mixture of the $(N-P)$ noise sources[10]. We therefore introduced a new extraction stage, based on the maximization of $|k_D(y)|$. We then proposed a new coloration stage, which relies on differential power optimization (or, equivalently, differential mean square error optimization). This stage yields the contributions of the extracted source in all observations. The other sources are then derived using a traditional deflation procedure.

In order to optimize the cost functions that we proposed in the above extraction and coloration stages, we introduced Newton-like algorithms, for real functions of complex variables.

Experimental tests in a 3-source to 2-observation configuration clearly show the efficiency of the proposed method: while the classical, i.e. non-differential, kurtotic method fails to achieve partial source separation, our approach typically yields a 10 dB improvement in terms of estimation SIR (en even more in terms of separation SIR).

Our future investigations will especially concern (i) the application of the proposed method to real-world signals exhibiting the required short-term stationarity and long-term non-stationarity, and (ii) the development of modified versions of this approach which reduce its computational load.

A Local extrema

In this appendix, we show that all the extrema of the cost function $|k_D(y)|$ are obtained for filters $C_p(z)$ satisfying (35) [11]. To this end, we analyze the global filters $C_j(z)$ associated to the sources of interest, i.e. with $j = 1, \ldots, P$, in the situation when the gradient of the cost function $|k_D(y)|$ is equal to zero. The approach that we develop in this appendix is similar to the method which was previously used by J. K. Tugnait [19] to analyze the variations of the classical normalized kurtosis with respect to the filters associated to all sources in the case of determined mixtures. Here, we precisely aim at proving that a differential extension of this type of approach may be developed for underdetermined mixtures.

To simplify the notations, we make use of the complex gradient operator [3] that we denote ∇. Let ν_r and ν_i be the real and imaginary parts of a complex scalar variable ν and $J(\nu)$ a real function of that variable ν. The complex gradient of $J(\nu)$ with respect to ν has the following properties:

$$\nabla_\nu J(\nu) = \frac{1}{2}\left(\frac{\partial J(\nu)}{\partial \nu_r} - i\frac{\partial J(\nu)}{\partial \nu_i}\right) \tag{80}$$

$$\nabla_{\nu^*} J(\nu) = \frac{1}{2}\left(\frac{\partial J(\nu)}{\partial \nu_r} + i\frac{\partial J(\nu)}{\partial \nu_i}\right). \tag{81}$$

This operator follows the same rules as the classical real gradient operator except that: (i) the derivative must be computed with respect to the complex variable ν^* and (ii) the variable ν^* must be considered as independent from ν in this derivation.

We now consider the cost function (38). We study the complex gradient of this cost function with respect to each coefficient of the global filters $C_j(z)$. For given indices j and l, we have

$$\nabla_{c_j^*(l)}|k_D(y)| = \nabla_{c_j^*(l)} k_D(y).sign(k_D(y)). \tag{82}$$

[10] This result holds only if the differential 2^{nd}-order cumulants of the sources of interest have the same sign. This is a not a problem however, since the considered time domains may be selected accordingly, as explained above in this paper.

[11] Note that the source which is partially separated for such a set of filters is not necessarily the one with the highest differential kurtosis value.

We first consider the term $\nabla_{c_j^*(l)} k_D(y)$ in (82). Eq. (24) yields

$$\nabla_{c_j^*(l)} k_D(y) = \frac{[\Delta CUM_2(y)]^2 \nabla_{c_j^*(l)} \Delta CUM_4(y) - 2\Delta CUM_2(y)\Delta CUM_4(y)\nabla_{c_j^*(l)}\Delta CUM_2(y)}{[\Delta CUM_2(y)]^4}. \tag{83}$$

Due to (22), (23) and (24), the differential normalized kurtosis $k_D(y)$ does not depend on the coefficients $c_j(l)$ with $j > P$. Therefore, the corresponding first and second-order derivatives of $k_D(y)$ are equal to zero. Thus, the stationary sources do not have any effect on the extrema of $k_D(y)$ and are therefore no longer considered hereafter.

For the remaining sources, i.e. those with $1 \leq j \leq P$, the first derivative of the 4^{th}-order cumulant expressed in (23) is

$$\begin{aligned} \nabla_{c_j^*(l)} \Delta CUM_4(y) &= \nabla_{c_j^*(l)} \left[\sum_{m=1}^{P} \Delta CUM_4(s_m) \sum_{k=-L_1''}^{L_2''} [c_m(k)]^2 [c_m^*(k)]^2 \right] \\ &= 2\Delta CUM_4(s_j)[c_j(l)]^2 c_j^*(l). \end{aligned} \tag{84}$$

Similarly, (22) leads to

$$\nabla_{c_j^*(l)} \Delta CUM_2(y) = \Delta CUM_2(s_j) c_j(l). \tag{85}$$

Using (83), (84) and (85), we can rewrite Eq. (82) as

$$\nabla_{c_j^*(l)} |k_D(y)| = \frac{2c_j(l)[\Delta CUM_4(s_j)|c_j(l)|^2 \Delta CUM_2(y) - \Delta CUM_4(y)\Delta CUM_2(s_j)] sign(k_D(y))}{[\Delta CUM_2(y)]^3}. \tag{86}$$

We now analyze the local extrema of this cost function. Thanks to Assumption 3, the 2^{nd}-order differential cumulants of all sources of interest have the same sign. As explained above, this sign may be derived from the observations $X_i(z)$. Moreover, if these differential cumulants are negative, they can be changed into positive values just by permuting the selected time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. Therefore, without any loss of generality, we consider from now the 2^{nd}-order differential cumulants of all sources of interest to be positive. We then define

$$\bar{c}_j(k) = \sqrt{\Delta CUM_2(s_j)}\, c_j(k), \quad j = 1, \ldots, N. \tag{87}$$

We therefore have

$$\bar{c}_j(k) = 0 \quad \forall\, j > P, \quad \forall\, k. \tag{88}$$

Moreover, (22) yields

$$\Delta CUM_2(y) = \sum_{j=1}^{P} \sum_{k=-L_1''}^{L_2''} \bar{c}_j(k) \bar{c}_j^*(k) \tag{89}$$

and (23) and (24) result in

$$\Delta CUM_4(y) = \sum_{j=1}^{P} k_D(s_j) \sum_{k=-L_1''}^{L_2''} [\bar{c}_j(k)]^2 [\bar{c}_j^*(k)]^2. \tag{90}$$

Dividing Eq. (86) by $\sqrt{\Delta CUM_2(s_j)}$, we obtain

$$\frac{\nabla_{\bar{c}_j^*(l)}|k_D(y)|}{\sqrt{\Delta CUM_2(s_j)}} = \frac{2\bar{c}_j(l)[k_D(s_j)|\bar{c}_j(l)|^2\Delta CUM_2(y) - \Delta CUM_4(y)]sign(k_D(y))}{[\Delta CUM_2(y)]^3}. \tag{91}$$

All the points where the gradient of the cost function $|k_D(y)|$ with respect to the global filters $\bar{c}_j(l)$ becomes equal to zero are the solutions of

$$\nabla_{\bar{c}_j^*(l)}|k_D(y)| = 0 \qquad \forall\, j = 1,\ \ldots,\ P \quad \text{and} \quad \forall\, l = -L_1'',\ \ldots,\ L_2''. \tag{92}$$

For given j and l, the latter equation combined with (91) leads to

$$\text{either} \qquad \bar{c}_j(l) = 0 \tag{93}$$

$$\text{or} \qquad k_D(s_j)|\bar{c}_j(l)|^2 = \frac{\Delta CUM_4(y)}{\Delta CUM_2(y)}. \tag{94}$$

We define

$$\underline{c}^{(M)} = [\ldots, \underbrace{\bar{c}_1(-1)\ldots\bar{c}_P(-1)}_{l=-1}, \underbrace{\bar{c}_1(0)\ldots\bar{c}_P(0)}_{l=0}, \underbrace{\bar{c}_1(1)\ldots\bar{c}_P(1)}_{l=1}, \ldots]^T$$

and we set its entries so that (92) holds. M is the number of non-zero entries of $\underline{c}^{(M)}$. Therefore, M is also the number of possibly time-shifted contributions of different or identical sources of interest in the extracted signal, as shown by (9). Since the extraction process provides a result up to an arbitrary scale factor, we can set $\underline{c}^{(M)}$ so that

$$\|\underline{c}^{(M)}\|^2 = \sum_{j=1}^{P}\sum_{k=-L_1''}^{L_2''} |\bar{c}_j(k)|^2 = 1. \tag{95}$$

Combined with (89), the latter equation implies

$$\Delta CUM_2(y)|_{\underline{c}=\underline{c}^{(M)}} = 1. \tag{96}$$

Eq. (94) then becomes

$$|\bar{c}_j^{(M)}(l)|^2 = \frac{\beta_M}{k_D(s_j)} \tag{97}$$

with

$$\beta_M = \Delta CUM_4(y)|_{\underline{c}=\underline{c}^{(M)}}. \tag{98}$$

Thus for a given M, the solutions of (92) are the vectors $\underline{c}^{(M)}$ which are such that

$$|\bar{c}_j^{(M)}(l)|^2 = \begin{cases} \beta_M/k_D(s_j) & \text{if } (j,l) \in I_M \\ 0 & \text{otherwise} \end{cases} \tag{99}$$

where I_M is a M-element subset of couples (j,l), with $j \in \{1,\ldots,P\}$, $l \in \{-L_1'',\ldots,L_2''\}$ and (j,l) such that $\bar{c}_j(l) \neq 0$.

We now prove that all the vectors $\underline{c}^{(M)}$ with $M \geq 2$ and $\|\underline{c}^{(M)}\|^2 = 1$ are saddle points of the cost function $|k_D(y)|$. We first show these points are not local maxima. For any $\underline{c}^{(M)}$, with an associated set I_M and $M \geq 2$, we define

$$\begin{aligned} B(\underline{c}^{(M)}) = \ & \{\underline{c} \,/\, \|\underline{c}\|^2 = \|\underline{c}^{(M)}\|^2 = 1,\ c_j(l) = 0 \text{ for } (j,l) \notin I_M \\ & \text{and } c_j(l) \text{ unspecified for } (j,l) \in I_M\}. \end{aligned} \tag{100}$$

Let (j_1, l_1) and (j_2, l_2) be two distinct couples from I_M and consider $\underline{c} \in B(\underline{c}^{(M)})$ such that, for $\epsilon > 0$ (ϵ small) and x real-valued,

$$\begin{aligned} \bar{c}_{j_1}(l_1) &= (1-\epsilon)\bar{c}_{j_1}^{(M)}(l_1) && (101) \\ \bar{c}_{j_2}(l_2) &= x\bar{c}_{j_2}^{(M)}(l_2) && (102) \\ \bar{c}_j(l) &= \bar{c}_j^{(M)}(l) \quad \text{if } (j,l) \neq (j_1, l_1) \text{ and } (j,l) \neq (j_2, l_2) && (103) \end{aligned}$$

Since $\underline{c} \in B(\underline{c}^{(M)})$ and $\underline{c}^{(M)} \in B(\underline{c}^{(M)})$, we have

$$||\underline{c}||^2 = ||\underline{c}^M||^2. \tag{104}$$

Therefore

$$|(1-\epsilon)\bar{c}_{j_1}^{(M)}(l_1)|^2 + |x\bar{c}_{j_2}^{(M)}(l_2)|^2 = |\bar{c}_{j_1}^{(M)}(l_1)|^2 + |\bar{c}_{j_2}^{(M)}(l_2)|^2. \tag{105}$$

The latter equation can be rewritten as

$$x^2 = 1 + \nu(2\epsilon - \epsilon^2) \tag{106}$$

with

$$\nu = \left| \frac{\bar{c}_{j_1}^{(M)}(l_1)}{\bar{c}_{j_2}^{(M)}(l_2)} \right|^2. \tag{107}$$

Eq. (24), (89) and (90) yield

$$\begin{aligned} k_D(y)_{\underline{c}^{(M)}} &= \frac{\Delta CUM_4(y)_{\underline{c}=\underline{c}^{(M)}}}{\left[\Delta CUM_2(y)_{\underline{c}=\underline{c}^{(M)}}\right]^2} && (108) \\ &= \frac{\sum_{j=1}^{P} k_D(s_j) \sum_{l=-L_1''}^{L_2''} |\bar{c}_j^{(M)}(l)|^4}{\left[\sum_{j=1}^{P} \sum_{l=-L_1''}^{L_2''} |\bar{c}_j^{(M)}(l)|^2\right]^2} && (109) \end{aligned}$$

and

$$k_D(y)_{\underline{c}} = \frac{\sum_{j=1}^{P} k_D(s_j) \sum_{l=-L_1''}^{L_2''} |\bar{c}_j(l)|^4}{\left[\sum_{j=1}^{P} \sum_{l=-L_1''}^{L_2''} |\bar{c}_j(l)|^2\right]^2}. \tag{110}$$

Moreover $\underline{c} \in B(\underline{c}^{(M)})$. Therefore $||\underline{c}||^2 = ||\underline{c}^{(M)}||^2 = 1$. The denominators of (109) and (110) are thus both equal to 1. These equations then become

$$\begin{aligned} k_D(y)_{\underline{c}^{(M)}} &= \sum_{j=1}^{P} k_D(s_j) \sum_{l=-L_1''}^{L_2''} |\bar{c}_j^{(M)}(l)|^4 && (111) \\ k_D(y)_{\underline{c}} &= \sum_{j=1}^{P} k_D(s_j) \sum_{l=-L_1''}^{L_2''} |\bar{c}_j(l)|^4. && (112) \end{aligned}$$

Let us introduce

$$\Gamma_M = k_D(y)_{\underline{c}} - k_D(y)_{\underline{c}^{(M)}}. \tag{113}$$

Eq. (101), (102), (103), (111) and (112), lead to

$$\Gamma_M = k_D(s_{j_1})|\bar{c}_{j_1}^{(M)}(l_1)|^4[(1-\epsilon)^4 - 1] + k_D(s_{j_2})|\bar{c}_{j_2}^{(M)}(l_2)|^4[x^4 - 1]. \tag{114}$$

In order to remove x from this expression, we first derive from (106)

$$\begin{aligned} x^4 - 1 &= [1 + \nu(2\epsilon - \epsilon^2)]^2 - 1 \\ &= 4\nu\epsilon + 2\epsilon^2[-\nu + 2\nu^2] + O(\epsilon^3). \end{aligned} \tag{115}$$

In addition, we have

$$(1-\epsilon)^4 - 1 = -4\epsilon + 6\epsilon^2 + O(\epsilon^3). \tag{116}$$

We then simplify (114), using (99),(107),(115) and (116). This yields

$$\Gamma_M = \beta_M |\bar{c}_{j_1}^{(M)}(l_1)|^2 4\epsilon^2(1+\nu) + O(\epsilon^3). \tag{117}$$

Eq. (24) and (98) result in

$$k_D(y)_{\underline{c}^{(M)}} = \frac{\beta_M}{\left[\Delta CUM_2(y)|_{\underline{c}=\underline{c}^{(M)}}\right]^2}. \tag{118}$$

Also taking into account (96), we obtain $k_D(y)_{\underline{c}^{(M)}} = \beta_M$. Moreover, for $\underline{c}$ in a neighborhood of $\underline{c}^{(M)}$, we have

$$sign(k_D(y)_{\underline{c}}) = sign(k_D(y)_{\underline{c}^{(M)}}) = sign(\beta_M). \tag{119}$$

Therefore, using (113) and (117) and for ϵ sufficiently small, we deduce that in a neighborhood of $\underline{c}^{(M)}$, the following conditions hold

- if $\beta_M > 0$ then $\underbrace{k_D(y)_{\underline{c}}}_{>0} = \underbrace{k_D(y)_{\underline{c}^{(M)}}}_{>0} + \underbrace{\Gamma_M}_{>0}$
- if $\beta_M < 0$ then $\underbrace{k_D(y)_{\underline{c}}}_{<0} = \underbrace{k_D(y)_{\underline{c}^{(M)}}}_{<0} + \underbrace{\Gamma_M}_{<0}$.

Therefore, whatever β_M

$$|k_D(y)_{\underline{c}}| > |k_D(y)_{\underline{c}^{(M)}}|. \tag{120}$$

Thus, any point $\underline{c}^{(M)}$ with $M \geq 2$ cannot be a local maximum of $|k_D(y)|$.

The same analysis must be carried out to show that a point $\underline{c}^{(M)}$ with $M \geq 2$ cannot be a local minimum of $|k_D(y)|$. To this end, we choose the vector $\underline{c}$ so that

$$\bar{c}_j(l) = \begin{cases} \sqrt{1-\epsilon}\,\bar{c}_j^{(M)}(l), & \forall (j,l) \in I_M \\ \sqrt{\epsilon} & \text{for a single couple } (j,l) = (j_1, l_1) \text{ with } (j_1, l_1) \notin I_M \\ 0 & \text{otherwise} \end{cases} \tag{121}$$

with a small positive ϵ. It may be shown easily that, since $||\underline{c}^{(M)}|| = 1$, we have $\|\underline{c}\| = 1$. Therefore, (111) and (112) also apply here. The latter equation yields

$$k_D(y)_{\underline{c}} = (1-\epsilon)^2 \sum_{j=1}^{P} k_D(s_j) \sum_{l=-L_1''}^{L_2''} |\bar{c}_j^{(M)}(l)|^4 + k_D(s_{j_1})\epsilon^2. \tag{122}$$

Combining the latter equation with (98),(108),(111) and $||\underline{c}^{(M)}|| = 1$ yields

$$k_D(y)_{\underline{c}} = (1-\epsilon)^2\beta_M + k_D(s_{j_1})\epsilon^2 \tag{123}$$
$$= \beta_M(1-2\epsilon) + O(\epsilon^2). \tag{124}$$

Thus, for ϵ sufficiently small

$$\begin{aligned} |k_D(y)_{\underline{c}}| &\simeq |\beta_M|(1-2\epsilon) \\ &= |k_D(y)_{\underline{c}^{(M)}}|(1-2\epsilon) \\ &< |k_D(y)_{\underline{c}^{(M)}}|. \end{aligned} \tag{125}$$

Therefore, a point $\underline{c}^{(M)}$ with $M \geq 2$ cannot be a local minimum of $|k_D(y)|$. As an overall result, it can only be a saddle point.

Therefore, the only points where the gradient of $|k_D(y)|$, with respect to the coefficients $\bar{c}_j(l)$ of the global filters associated to the sources of interest, is equal to zero and which are extrema of $|k_D(y)|$ are obtained for $M = 1$. In other words, they are such that only one coefficient $\bar{c}_j(l)$ with $j \in \{1, \ldots, P\}$ and $l \in \{-L_1'', \ldots, L_2''\}$ is non-zero. They are therefore all defined by Eq. (35).

B Tools for optimization algorithms

We here define the tools required in Section 3.

B.1 Complex derivatives

B.1.1 First-order complex derivative

Let $f(\theta)$ be a real scalar function of a complex scalar variable θ. We here define its first-order complex derivative with respect to θ, denoted $f'(\theta)$. This derivative describes the evolution of $f(\theta)$ along the real and imaginary axes. This operation is achieved by considering separately the real and imaginary parts of θ and then defining

$$f'(\theta) = \frac{\partial}{\partial \mathcal{R}e\{\theta\}} f + i \frac{\partial}{\partial \mathcal{I}m\{\theta\}} f. \tag{126}$$

To simplify the notations, we rewrite this expression as

$$f'(\theta) = \nabla_{\mathcal{R}e\{\theta\}} f + i \nabla_{\mathcal{I}m\{\theta\}} f. \tag{127}$$

We here use the complex derivative operator defined in [3]. This operator, denoted ∇, yields

$$\nabla_{\theta^*} f = \tfrac{1}{2}\left(\tfrac{\partial f}{\partial \mathcal{R}e\{\theta\}} + i \tfrac{\partial f}{\partial \mathcal{I}m\{\theta\}}\right) \tag{128}$$

i.e., using the above-defined notations,

$$\nabla_{\theta^*} f = \frac{1}{2}\left(\nabla_{\mathcal{R}e\{\theta\}} f + i \nabla_{\mathcal{I}m\{\theta\}} f\right). \tag{129}$$

Comparing this expression to (127), we obtain

$$f'(\theta) = 2\nabla_{\theta^*} f. \tag{130}$$

Note that deriving $f(\theta)$ with respect to θ instead of θ^* leads to

$$\begin{aligned} \nabla_\theta f &= \tfrac{1}{2}\left(\nabla_{\mathcal{R}e\{\theta\}} f - i \nabla_{\mathcal{I}m\{\theta\}} f\right) \\ &= \left(\nabla_{\theta^*} f\right)^*. \end{aligned} \tag{131}$$

B.1.2 Second-order complex derivative

We now define the second-order complex derivative of $f(\theta)$ with respect to θ as

$$f''(\theta) = \nabla^2_{\mathcal{R}e\{\theta\}\mathcal{R}e\{\theta\}} f + i\nabla^2_{\mathcal{I}m\{\theta\}\mathcal{I}m\{\theta\}} f. \tag{132}$$

The calculations presented in [3] then yield

$$\begin{aligned} \nabla^2_{\mathcal{R}e\{\theta\}\mathcal{R}e\{\theta\}} f &= \nabla^2_{\theta\theta^*} f + \nabla^2_{\theta^*\theta^*} f + \nabla^2_{\theta\theta} f + \nabla^2_{\theta^*\theta} f \\ &= 2\mathcal{R}e\{\nabla^2_{\theta\theta^*} f + \nabla^2_{\theta\theta} f\} \end{aligned} \tag{133}$$

$$\begin{aligned} \nabla^2_{\mathcal{I}m\{\theta\}\mathcal{I}m\{\theta\}} f &= \nabla^2_{\theta\theta^*} f - \nabla^2_{\theta^*\theta^*} f - \nabla^2_{\theta\theta} f + \nabla^2_{\theta^*\theta} f \\ &= 2\mathcal{R}e\{\nabla^2_{\theta\theta^*} f - \nabla^2_{\theta\theta} f\}. \end{aligned} \tag{134}$$

We thus eventually obtain

$$f''(\theta) = 2\mathcal{R}e\{\nabla^2_{\theta\theta^*} f + \nabla^2_{\theta\theta} f\} + 2i\mathcal{R}e\{\nabla^2_{\theta\theta^*} f - \nabla^2_{\theta\theta} f\}. \tag{135}$$

B.2 Newton-like algorithms

B.2.1 Newton's classical algorithm

The classical version of Newton's algorithm applies to a real scalar function $f(\theta)$ of a real vector θ. For a scalar variable θ, a slightly extended form [15] of this algorithm reads

$$\theta_{n+1} = \theta_n + \mu \frac{\left.\frac{\partial f}{\partial \theta}\right|_{\theta_n}}{\left.\frac{\partial^2 f}{\partial \theta^2}\right|_{\theta_n}}, \tag{136}$$

where θ_n is the value of θ for the n^{th} iteration and μ is the adaptation gain. This gain is equal to -1 in Newton's most classical algorithm, and this algorithm may then converge towards a maximum or a minimum of the function $f(\theta)$. The gain μ is assigned as explained below in modified versions of this method considered in this paper.

B.2.2 Complex version of Newton's algorithm

Now consider a real scalar function $f(\theta)$ of a complex scalar variable θ. Using the complex derivatives defined in Section B.1, the above algorithm becomes

$$\theta_{n+1} = \theta_n + \mu \left.\left(\frac{\mathcal{R}e\{f'(\theta)\}}{\mathcal{R}e\{f''(\theta)\}} + i.\frac{\mathcal{I}m\{f'(\theta)\}}{\mathcal{I}m\{f''(\theta)\}} \right)\right|_{\theta_n}. \tag{137}$$

B.2.3 Modified Newton algorithms for maximum search

In the extraction stage of our BSS method, we need an algorithm which is guaranteed to only converge towards a maximum, since this stage aims at maximizing a cost function. On the contrary, Newton's classical algorithm may converge towards a maximum or a minimum of $f(\theta)$, as mentioned above. It may be checked easily that this behavior is related to the dependence of its update term with respect to the second-order derivative of $f(\theta)$, and that this problem may therefore be avoided by using the modified form of algorithm (136) that we propose, i.e.

$$\theta_{n+1} = \theta_n + \mu \frac{\left.\frac{\partial f}{\partial \theta}\right|_{\theta_n}}{\left|\left.\frac{\partial^2 f}{\partial \theta^2}\right|_{\theta_n}\right|}, \tag{138}$$

where μ is a positive adaptation gain: the modified algorithm (138) may only converge towards local maxima of $f(\theta)$, not towards its minima. Similarly, for a complex scalar variable θ, the modified form of (137) for finding a maximum of $f(\theta)$ reads

$$\theta_{n+1} = \quad \theta_n + \mu. \left.\left(\frac{\mathcal{R}e\{f'(\theta)\}}{|\mathcal{R}e\{f''(\theta)\}|} + i.\frac{\mathcal{I}m\{f'(\theta)\}}{|\mathcal{I}m\{f''(\theta)\}|}\right)\right|_{\theta_n}. \tag{139}$$

Now consider the situation encountered in this paper, i.e. the maximization of a function g defined with respect to another function f as

$$g = |f| = sign(f)f. \tag{140}$$

The corresponding maximization algorithm reads

$$\theta_{n+1} = \quad \theta_n + \mu.sign(f(\theta)). \left.\left(\frac{\mathcal{R}e\{f'(\theta)\}}{|\mathcal{R}e\{f''(\theta)\}|} + i.\frac{\mathcal{I}m\{f'(\theta)\}}{|\mathcal{I}m\{f''(\theta)\}|}\right)\right|_{\theta_n}. \tag{141}$$

References

[1] F. Abrard, Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources", Signal Processing, Vol. 85, Issue 7, pp. 1389-1403, July 2005.

[2] P. Bofill, M. Zibulevsky, "Underdetermined blind source separation using sparse representations", Signal Processing, vol. 81, pp. 2353-2362, 2001.

[3] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory", Proc. Inst. Elec. Eng., Vol. 130, No. 1, pp. 11-16, February 1983.

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

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

[6] P. Comon, "Blind identification and source separation in 2x3 under-determined mixtures", IEEE Transactions on Signal Processing, vol. 52, no. 1, pp. 11-22, Jan. 2004.

[7] N. Delfosse, P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach", Signal Processing, vol. 45, no. 1, pp. 59-84, July 1995.

[8] Y. Deville, M. Benali, F. Abrard, "Differential source separation for underdetermined instantaneous or convolutive mixtures: concept and algorithms", Signal Processing, Vol. 84, Issue 10, pp. 1759-1776, Oct. 2004.

[9] Y. Deville, M. Puigt, "Temporal and time-frequency correlation-based blind source separation methods. Part I: determined and underdetermined linear instantaneous mixtures", to appear in Signal Processing.

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

[11] A. Hyvarinen, E. Oja, "A fast fixed-point algorithm for Independent Component Analysis", Neural Computation, vol. 9, pp. 1483-1492, 1997.

[12] A. Hyvarinen, J. Karhunen, E. Oja, "Independent Component Analysis", Wiley, New York, 2001.

[13] J.-L. Lacoume, P.-O. Amblard, P. Comon, Statistiques d'ordre supérieur pour le traitement du signal, Masson, Paris, 1997.

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

[15] F. Michaud, Méthodes adaptatives pour le signal, Hermès, Paris, 1992.

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

[17] M. Puigt, Y. Deville, "Time-frequency ratio-based blind separation methods for attenuated and time-delayed sources", Mechanical Systems and Signal Processing, vol. 19, pp. 1348-1379, 2005, http://authors.elsevier.com/sd/article/S0888327005001123

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

[19] J. K. Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria", IEEE Trans. on Signal Processing, Vol. 45, No. 3, pp. 658-672, March 1997.

[20] L. Vielva, Y. Pereiro, D. Erdogmus, J.C. Principe, "Inversion techniques for underdetermined BSS in an arbirary number of dimensions", Proceedings of the Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA2003), pp. 131-136, Nara, Japan, April 1-4, 2003.

[21] O. Yilmaz, S. Rickard, "Blind separation of speech mixtures via time-frequency masking", IEEE Transactions on Signal Processing, vol. 52, no. 7, pp. 1830-1847, July 2004.

Table 1: Global input SIR (dB).

		Source index		
Test no.	Observation index	1	2	3
1	1	-9.3	-12.92	7.42
	2	-13.40	-9.46	7.70
2	1	-8.32	-16.00	7.44
	2	-16.00	-9.35	8.32
3	1	-8.43	-13.02	6.81
	2	-11.83	-9.54	7.17
4	1	-6.86	-11.35	5.03
	2	-11.11	-6.76	4.86
5	1	-3.95	-8.13	1.43
	2	-8.60	-4.02	1.70

Table 2: Partial input SIR (dB).

		Source index	
Test no.	Observation index	1	2
1	1	3.33	-3.33
	2	-3.70	3.70
2	1	7.17	-7.17
	2	-6.29	6.29
3	1	4.20	-4.20
	2	-2.13	2.13
4	1	3.96	-3.96
	2	-3.88	3.88
5	1	3.33	-3.33
	2	-3.70	3.70

Table 3: Estimation output SIR (dB) of non-differential BSS method.

		Source index	
Test no.	Output index	1	2
1	1	1.41	-0.76
	2	2.57	-1.92
2	1	1.40	-2.45
	2	4.73	-5.77
3	1	5.23	-4.01
	2	0.18	1.03
4	1	1.35	0.02
	2	3.98	-2.61
5	1	1.57	1.81
	2	5.13	-1.76

Table 4: Estimation output SIR (dB) of differential BSS method.

		Source index	
Test no.	Output index	1	2
1	1	13.50	-5.30
	2	-1.98	10.17
2	1	16.73	-8.10
	2	-0.91	9.55
3	1	14.61	-5.50
	2	-1.30	10.42
4	2	12.83	-5.64
	1	-1.68	8.88
5	1	13.94	-5.21
	2	-1.89	10.61

Table 5: Separation output SIR (dB) of differential BSS method.

		Source index	
Test no.	Output index	1	2
1	1	21.84	-21.84
	2	-11.25	11.25
2	1	18.87	-18.87
	2	-14.07	14.07
3	1	16.13	-16.13
	2	-15.25	15.25
4	1	13.41	-13.41
	2	-15.07	15.07
5	1	19.68	-19.68
	2	-12.20	12.20