DIFFERENTIAL SOURCE SEPARATION: CONCEPT AND APPLICATION TO A CRITERION BASED ON DIFFERENTIAL NORMALIZED KURTOSIS

Yannick Deville, Mohammed Benali
Laboratoire d'Acoustique, de Métrologie, d'Instrumentation (LAMI),
Université Paul Sabatier, 38 Rue des 36 Ponts, 31400 Toulouse, FRANCE
Tel: +33 5 61 55 63 33; fax: +33 5 61 25 94 78
e-mail: ydeville@cict.fr

ABSTRACT

This paper concerns the underdetermined case of the blind source separation problem, i.e. the situation when the number of observed mixed signals is lower than the number of sources. The general concept that we propose in this case consists of a differential source separation approach, which uses optimization criteria based on differential parameters, so as to make some sources invisible in these criteria and to perform an exact separation of the other sources only. We illustrate this partial source separation concept on a new criterion based on the "differential normalized kurtosis" that we introduce to this end. We then validate the performance of this criterion by means of experimental tests.

1 INTRODUCTION

In the blind source separation problem, one aims at estimating a set of n_s independent source signals $X_i(t)$ from a set of n_o observed signals $Y_i(t)$, which are mixtures of these source signals [1]. The mixed signals $Y_i(t)$ are typically provided by sensors, and in the so-called linear instantaneous mixture model, each source-to-sensor propagation path is represented by a scalar coefficient a_{ij} applied to the considered source signal. The overall relationship between the column vectors $X(t)$ and $Y(t)$ of sources and observations then reads:

$$Y(t) = AX(t), \tag{1}$$

where the mixing matrix A consists of the coefficients a_{ij}. Most investigations have been performed in the case when: i) $n_o = n_s$, so that the matrix A is square, and ii) this matrix is invertible. The source separation problem then basically consists in determining an estimate of the inverse of A: consider a vector $S(t)$ of outputs signals of a source separation system, obtained by multiplying the available mixed signals by a separating matrix C, i.e:

$$S(t) = CY(t). \tag{2}$$

Combining (1) and (2) shows that, when C is made equal to A^{-1}, all the output signals $S_i(t)$ resp. become exactly equal to all the source signals $X_i(t)$ to be restored. Various methods have been proposed to estimate A^{-1}, based on the assumed statistical independence of the source signals (see e.g. the survey in [1]). Most of them consist of maximizing or cancelling parameters of the outputs signals such as their moments, cumulants or normalized kurtosis, which are classical parameters in the higher-order statistics field [2].

As stated above, most of these investigations were performed under the assumption $n_o = n_s$. 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. A few authors have considered this case when $n_o < n_s$ (see e.g. [3] and references therein, [4]). This paper also concern this case but, unlike in previously reported investigations, we here introduce a practical approach which combines the following features:

- It applies to continuous sources with unknown distributions.
- It is not restricted to specific mixing conditions.
- It only performs linear (instantaneous) combinations of the available mixed signals, thus providing linear combinations of the sources. We initially set this condition because of the eventual applications that we have in mind (although the proposed approach is not yet developed up to that point in this paper): in various applications, such as speech enhancement, it has been observed that the artefacts created by non-linear processing are (e.g. physiologically) even more disturbing than the "noise" present in the initial mixed signals, so that linear processing should be preferred. The price to pay for this linear operation is that some "noise" is still present in the outputs of the separating system introduced hereafter. More precisely, this approach only performs a "partial source separation", as defined in Section 2.

2 SOURCE SEPARATION LIMITATION AND PROPOSED CONCEPT

For the sake of simplicity, we focus on the case of linear instantaneous mixtures in this paper. We showed above that if $n_o = n_s$ an exact separation of all sources is theoretically possible (and is achieved when C is made exactly equal to A^{-1}). It is clearly also possible when $n_o > n_s$, but not when $n_o < n_s$ as will now be shown. In the latter case, A is not square. Each output signal $S_i(t)$ defined by (2) is a linear instantaneous combination of the n_o mixed signals. For an arbitrary matrix A, the highest number of source contributions that may be cancelled in such a (non-zero) combination is $n_o - 1$. This optimum case, where each output is still a mixture of the remaining $(n_s - n_o + 1)$ sources, is called partial source separation hereafter.

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

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

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

The above-mentioned two "occurences" may be obtained in various ways, allowing one to derive various methods from the proposed concept. For example, they may correspond to the two values of a parameter, such as a cumulant, resp. obtained for two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$. These domains may be defined as follows. When considering these theoretical (zero-lag) cumulants themselves, which consist of mathematical expectations of signals [2], each of these time domains $\mathcal{D}_i$ is restricted to the time position t_i when the corresponding mathematical expectations are considered. On the contrary, the corresponding cumulant estimates are resp. obtained by time averaging over two domains $\mathcal{D}_1$ and $\mathcal{D}_2$ which then consist of non-overlapping bounded time intervals, assuming that each source exhibits stationarity over each such domain (this is referred to as "short-term stationarity" below). When the two parameter occurences are defined in such a way, the second type of sources consists of sources whose cumulant values do not vary from one of the considered time domains to the other, i.e. sources which in addition exhibit "long-term stationarity". On the contrary, first-type sources then consist of long-term non-stationary sources.

Whatever the origin of these occurences, the proposed concept consists in deriving new signal parameters from the initial ones in such a way that the effect of the second type of sources disappears in them, using a method that is detailed further in this section. The initial problem is thus transformed in a problem where (at most) n_o sources are visible (i.e. the first-type sources) from the point of view of the new parameter, and n_o mixtures of these sources are available. In other words, thanks to this approach we get back in the classical configuration involving as many observed signals as source signals. Adapting the initial source separation method to the optimization of the new parameter then yields separating coefficients c_{ij} such that each output signal $S_i(t)$ contains a contribution from only one of the sources seen by the approach, i.e. of the first-type sources. Of course, each such signal $S_i(t)$ also contains contributions from all second-type sources, corresponding to the combination of the mixed signals $Y_i(t)$ by means of the above-mentioned coefficients c_{ij}, as shown in (2). This approach therefore achieves the above-defined partial source separation, and more specifically the sources that are thus completely separated are the first-type one. These sources are the signals of interest in this partial source separation problem, as opposed to the "noise signals", i.e. the second-type sources, which are still present in all outputs.

We now explain how to derive new parameters in which the second-type sources are invisible. For the sake of simplicity, we assume that the initial parameter is a linear function of the corresponding source parameters (for example, an output cumulant is indeed a linear function of the cumulants of the independent sources). The new parameter that we then define exploits the difference of behavior between the two types of sources concerning the variations of their initial parameter from one occurence to the other. More precisely, we here focus on the case when the new parameter that we define is the difference between the two values of the initial parameter resp. associated with the two available occurences (e.g. the difference between the two output cumulants corresponding to the two time domains). As each second-type source yields the same contribution in the two initial parameters, its effect is cancelled in the corresponding parameter difference (see the example in the next section), i.e. the above-defined goal is thus reached.

To summarize, the proposed concept consists of a differential source separation approach, which uses optimization criteria based on differential parameters, so as to make some sources invisible in these criteria and to perform an exact separation of the other sources only.

3 A CRITERION BASED ON DIFFERENTIAL NORMALIZED KURTOSIS

We now apply the above concept to the source separation criterion that we defined in [5] for the case $n_o = 2$ mixtures of $n_s = 2$ sources. As this criterion is based on the notions of cumulants and normalized kurtosis, we first have to define the new parameters used in the differential source separation method derived below, i.e. the differential versions of the above parameters. Given an arbitrary zero-mean signal u and two time domains $\mathcal{D}_1$ and $\mathcal{D}_2$ introduced in Section 2, we define the corresponding (zero-lag) m^{th}-order differential cumulant as:

$$\begin{aligned} DCUM_m(u, \mathcal{D}_1, \mathcal{D}_2) &= CUM_m(u, \mathcal{D}_2) \\ &\quad - CUM_m(u, \mathcal{D}_1), \end{aligned} \quad (3)$$

where $CUM_m(u, \mathcal{D}_i)$ are the corresponding classical (i.e. non-differential) cumulants, which may be found e.g. in [2]. Then, based on the expression of the classical (i.e. non-differential) normalized kurtosis on a domain $\mathcal{D}_i$, i.e [2]

$$K(u, \mathcal{D}_i) = \frac{CUM_4(u, \mathcal{D}_i)}{[CUM_2(u, \mathcal{D}_i)]^2}, \quad (4)$$

we define the corresponding differential normalized kurtosis as:

$$DK(u, \mathcal{D}_1, \mathcal{D}_2) = \frac{DCUM_4(u, \mathcal{D}_1, \mathcal{D}_2)}{[DCUM_2(u, \mathcal{D}_1, \mathcal{D}_2)]^2}. \quad (5)$$

The source separation configuration that we then consider corresponds to $n_o = 2$ observed signals $Y_i(t)$ which are linear instantaneous mixtures of an arbitrary number n_s of sources signals. Two of these source signals, denoted $X_1(t)$ and $X_2(t)$, have long-term non-stationarity: these are the signals of interest to be separated. The other source signals, denoted $V_3(t)$ to $V_{n_s}(t)$, have long-term stationarity: these are the noise signals. In the remainder of this section, we consider the centered versions of all signals, denoted with lower-case letters. The (centered) mixed signals then read:

$$y_i(t) = a_{i1}x_1(t) + a_{i2}x_2(t) + \sum_{j=3}^{n_s} a_{ij}v_j(t), \quad i \in 1, 2. \quad (6)$$

Now consider a linear instantaneous combination of the source signals:

$$z(t) = \alpha_1 x_1(t) + \alpha_2 x_2(t) + \sum_{j=3}^{n_s} \alpha_j v_j(t). \quad (7)$$

This signal may e.g. be an output of the considered source separation system, created as a linear instantaneous combination of the available two mixed signals and here expressed as:

$$z(t) = y_1(t) - cy_2(t). \quad (8)$$

The α coefficients in (7) then combine the mixing and separating coefficients a_{ij} and c. The m^{th}-order cumulant of $z(t)$ at each time position t_i is then derived from (7) by using the properties of cumulants [2] and the assumed statistical independence of the source signals. This yields:

$$\begin{aligned} CUM_m(z, t_i) &= \alpha_1^m CUM_m(x_1, t_i) \\ &\quad + \alpha_2^m CUM_m(x_2, t_i) \\ &\quad + \sum_{j=3}^{n_s} \alpha_j^m CUM_m(v_j, t_i). \end{aligned} \quad (9)$$

The corresponding differential cumulant $DCUM_m(z, t_1, t_2)$ then follows from the difference defined in (3). The contribution of each noise signal $v_j(t)$ disappears in this difference, as its cumulants $CUM_m(v_j, t_i)$ corresponding to the two time positions t_1 and t_2 have the same value. All noise signals therefore also disappear in the differential normalized kurtosis (5), which may also be expressed as:

$$DK(z, t_1, t_2) = \frac{DK(x_1, t_1, t_2) + q^2 DK(x_2, t_1, t_2)}{[1+q]^2} \quad (10)$$

where we introduce the variable q defined as:

$$q = \frac{\alpha_2^2 DCUM_2(x_2, t_1, t_2)}{\alpha_1^2 DCUM_2(x_1, t_1, t_2)}. \quad (11)$$

Comparing (10)-(11) with the expressions what we derived in [5] for our initial (i.e. non-differential) source separation criterion shows that they are identical, except that: i) the initial cumulants and normalized kurtosis are replaced by their differential counterparts, and therefore ii) the variable p that we introduced in our initial criterion is replaced by the variable q. Note that, whereas p always covers the range $[0, +\infty]$ when the α coefficients are varied for given sources, q covers either the same range or $[-\infty, 0]$, depending on the signs of the differential 2^{nd}-order cumulants of the considered sources. Due to space limitations, we here restrict ourselves to the case when it covers the range $[0, +\infty]$. The solution that we propose for the partial source separation problem considered here is then the differential counterpart of the initial one, i.e: the separation of the signals of interest is guaranteed to be achieved by adjusting each coefficient c of the separation system so as to reach one of the extrema of $DK(z, t_1, t_2)$, where the types of extrema to be considered depend on the signs of the here-differential normalized kurtosis of the sources [5]. More precisely:

1. If $DK(x_i, t_1, t_2) < 0$ for both sources of interest, then the two values of c for which partial source separation is achieved coincide with the two values of c which minimize $DK(z, t_1, t_2)$.
2. Symmetrically, if $DK(x_i, t_1, t_2) > 0$ for both sources, partial source separation coincides with the two maxima of $DK(z, t_1, t_2)$.
3. Eventually, if $DK(x_1, t_1, t_2)$ and $DK(x_2, t_1, t_2)$ have oposite signs, partial source separation coincides with the two extrema (i.e. its only minimum and maximum) of $DK(z, t_1, t_2)$.

4 EXPERIMENTAL RESULTS

We have validated the above approach by means of mixed signals obtained by combining three sources, i.e. two binary-valued (-1/+1) signals of interest $X_1(t)$ and $X_2(t)$ and a stationary noise signal $V_3(t)$ uniformly distributed in the range $[-1, +1]$. Each signal $X_i(t)$ is stationary over each considered time interval, but its statistics are modified from one interval to the other by changing the probability p_i that $X_i(t) = +1$. These probabilities were selected so that the centered versions $x_1(t)$ and $x_2(t)$ of both useful signals are sub-Gaussian in the first interval $\mathcal{D}_1$ and super-Gaussian in the second interval $\mathcal{D}_2$. More precisely, p_1 and p_2 were both set to $\frac{1}{2}$ in $\mathcal{D}_1$, which results in the following theoretical[1] parameter values for each centered source $x_i(t)$: $CUM_2(x_i, \mathcal{D}_1) = 1$, $CUM_4(x_i, \mathcal{D}_1) = -2$ and therefore $K(x_i, \mathcal{D}_1) = -2$ as shown by (4). Similarly, we used $p_1 = p_2 = \frac{1}{10}$ in $\mathcal{D}_2$, so that $CUM_2(x_i, \mathcal{D}_2) = 0.36$, $CUM_4(x_i, \mathcal{D}_2) = 0.6624$ and therefore $K(x_i, \mathcal{D}_2) \simeq 5.11$. The differential normalized kurtosis of both centered sources are then $DK(x_i, \mathcal{D}_1, \mathcal{D}_2) = 6.5$, i.e. these sources are "differentially super-Gaussian". These signals were mixed and centered according to:

$$y_1(t) = a_{11}x_1(t) + a_{12}x_2(t) + a_{13}v_3(t) \quad (12)$$
$$y_2(t) = a_{21}x_1(t) + a_{22}x_2(t) + a_{23}v_3(t). \quad (13)$$

In all experiments, we used $a_{11} = 1$, $a_{12} = 0.9$, $a_{21} = 0.8$, and $a_{22} = 1$, so that the two values of c corresponding to partial source separation are $a_{12}/a_{22} = 0.9$ and $a_{11}/a_{21} = 1.25$ [5]. On the contrary, the mixture coeficients a_{13} and a_{23} were varied in the experiments, so as to investigate the influence of the contributions of the noise signals on the performance of the proposed approach.

[1] Noticable deviations from the theoretical distributions defined in this paragraph were observed in the empirical signals and appear in the figures provided hereafter.

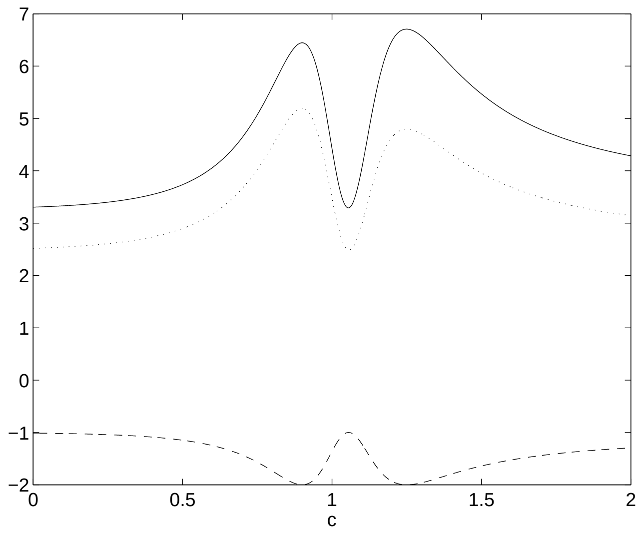

Figure 1: Variations vs c of non-differential output kurtosis in each domain $\mathcal{D}_i$ and of differential kurtosis between these domains. Dashed line: $K(z, \mathcal{D}_1)$, dotted line: $K(z, \mathcal{D}_2)$, solid line: $DK(z, \mathcal{D}_1, \mathcal{D}_2)$. Mixture coefficients: $a_{13} = a_{23} = 0$.

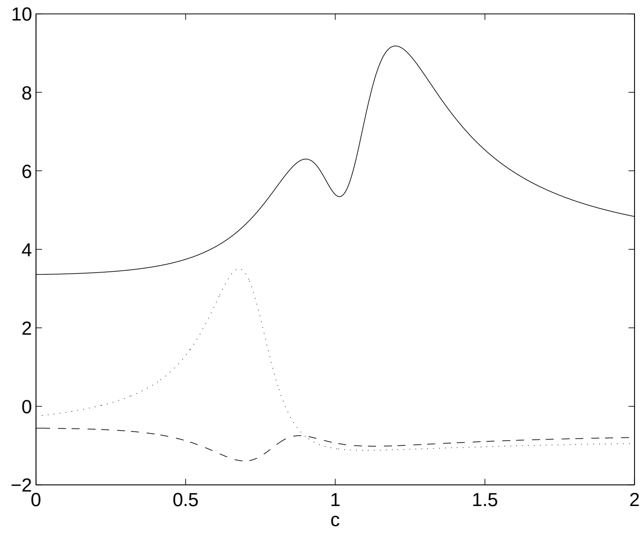

Figure 2: Same legend as in Fig. 1, except: $a_{13} = 2$, $a_{23} = 3$.

The very first experiment was performed with no noise contributions, i.e. $a_{13} = a_{23} = 0$. The theoretical discussion provided above shows that, in this preliminary noiseless case, the two sources of interest may then be separated by three approaches, i.e. either by using our non-differential approach in any of the considered two time domains, or by applying our new differential method between these domains. This was confirmed experimentally by considering the centered source separation system output and by determining the variations vs c of its experimental normalized kurtosis (differential or not) associated with the considered three approaches. When varying c with a step of 10^{-3}, $K(z, \mathcal{D}_1)$ has two minima situated at $c = 0.900$ and $c = 1.250$ (see Fig. 1), whereas $K(z, \mathcal{D}_2)$ and $DK(z, \mathcal{D}_1, \mathcal{D}_2)$ have two maxima situated at $c = 0.900$ and $c = 1.250$, which is in full agreement with the discussion presented at the end of Section 3 (including its coutnerpart for the non-differential approach).

The second experiment was then carried out with moderate noise contributions, i.e. $a_{13} = 0.2$ and $a_{23} = 0.3$. The non-differential approaches are affected by this noise: the two minima of $K(z, \mathcal{D}_1)$ now occur for $c = 0.889$ and $c = 1.268$ and the maxima of $K(z, \mathcal{D}_2)$ correspond to $c = 0.872$ and $c = 1.311$. On the contrary, the differential approach proposed in this paper is completely insensitive to this noise, as the maxima of $DK(z, \mathcal{D}_1, \mathcal{D}_2)$ remain exactly at $c = 0.900$ and $c = 1.250$.

The last experiment was performed with large noise contributions, i.e. $a_{13} = 2$ and $a_{23} = 3$. The non-differential approaches then do not apply at all: the two minima of $K(z, \mathcal{D}_1)$ here occur for $c = 0.697$ and $c = 1.129$, which is quite different from the suitable values. Moreover, $K(z, \mathcal{D}_2)$ then only has a single maximum, situated at $c = 0.679$ (see Fig. 2). On the contrary, the differential approach is almost insensitive to this noise, as the maxima of $DK(z, \mathcal{D}_1, \mathcal{D}_2)$ are then only shifted to $c = 0.902$ and $c = 1.201$.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have introduced a concept which allows to perform partial source separation with a reduced number of observed signals. We applied this differential approach to a criterion based on the "differential normalized kurtosis" that we introduced to this end, and we validated its performance by means of experimental tests. Optimization algorithms associated to this criterion are presented in another paper [6]. Our future investigations will esp. concern the application of the proposed concept to convolutive source separation methods.

References

[1] J.-F. Cardoso, "Blind signal separation: statistical principles", Proceedings of the IEEE, vol. 86, no. 10, pp. 2009-2025, Oct. 1998.

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

[3] P. Comon, O. Grellier, "Non-linear inversion of underdetermined mixtures", Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99), Aussois, France, Jan. 11-15, 1999, pp. 461-465.

[4] X-R Cao, R-W Liu, "General approach to blind source separation", IEEE Transactions on Signal Processing, vol. 44, no. 3, pp. 562-571, March 1996.

[5] Y. Deville, "A source separation criterion based on signed normalized kurtosis", Proceedings of the 4th International Worshop on Electronics, Control, Measurement and Signals (ECMS'99), Liberec, Czech Republic, May 31 - June 1, 1999, pp. 143-146.

[6] F. Abrard, Y. Deville, M. Benali, "Numerical and analytical solutions to the differential source separation problem", to appear in the Proceedings of EUSIPCO'2000, Tampere, Finland, Sept. 5-8, 2000.