NUMERICAL AND ANALYTICAL SOLUTIONS TO THE DIFFERENTIAL SOURCE SEPARATION PROBLEM

Frédéric ABRARD, Yannick DEVILLE and Mohammed BENALI
Laboratoire d'Acoustique, de Métrologie et d'Instrumentation (LAMI)
Université Paul Sabatier, 38 Rue des 36 Ponts, 31400 Toulouse, FRANCE.
Phone: (+33) 5 61 55 65 33 Fax: (+33) 5 61 25 94 78
e-mail: abrard@cict.fr

ABSTRACT

The aim of this paper is to present several algorithms which solve the blind source separation (BSS) problem when stationary noises are added to the source signals in the linear instantaneous mixture context. We use a new criterion based on the *differential normalized kurtosis* that we developed in a previous paper. The different algorithms are then applied to mixtures of two sources signals and an additive noise.

1 INTRODUCTION

Blind source separation is now a classical subject in signal processing. It consists in 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. We focus here on the case of linear instantaneous mixtures. The overall relationship between the source and the observation vectors then reads:

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

where A represents the mixing matrix.

Among the different proposed methods, higher-order statistics are the most frequently used. A presentation of the principal methods can be found in [1]. It appears that most investigations have been performed considering that the number of source and observed signals are equal (i.e. $n_s = n_o$).

We recently introduced a new criterion based on the optimization of the signed normalized kurtosis of each separation system output [2]-[3]. Like classical approaches, this method works properly only when the number n_s of source signals and the number n_o of observed signals are equal. This restriction implies that the observed signals must be free from noise.

To go further, we modified this criterion to adapt it to the case when stationary noise is added to the source signals. This new problem can be expressed as: $n_s > n_o$ with $(n_s - n_o)$ stationary noise signals. The theoretical concepts of the modified criterion are developed in [4]. This new criterion is based on a parameter that we called the *differential normalized kurtosis.*

The aim of this paper is to present different algorithms associated to the above modified criterion, which solve the differential blind source separation problem in noisy configurations. We will present two different ways:

- iterative algorithms based on gradient and Newton methods,
- an analytical resolution.

To illustrate these methods, we will apply them to noisy mixtures of binary samples. It should be clear however that the proposed approach also applies to continuous sources.

2 A CRITERION BASED ON THE DIFFERENTIAL NORMALIZED KURTOSIS

2.1 Problem Statement

In most practical situations, the number of sensors is limited and each one receives signals from a large set of sources. More comfortable configurations are when we exactly know the number of sources and we then put the same number of sensors. But we all know that we will always have additive noise which can be considered as another source signal.

For the sake of simplicity, we focus here on the case of 3 independent sources and 2 sensors, more especially 2 desired source signals X_1 and X_2 added with a stationary noise X_3:

$$\begin{aligned} Y_1(t) &= a_{11}X_1(t) + a_{12}X_2(t) + a_{13}X_3(t) \\ Y_2(t) &= a_{21}X_1(t) + a_{22}X_2(t) + a_{23}X_3(t) \end{aligned} \tag{2}$$

where a_{ij} are the coefficients of the mixture.
In the remainder of this paper, we consider the centered versions of all signals, denoted with lower-case letters.
The observed signals $y_i(t)$ are connected to the same separation system as in [2]-[3] which is showed in Fig. 1.

We can express its output $s(t)$ with respect to the three source signals:

$$s(t) = \alpha_1 x_1(t) + \alpha_2 x_2(t) + \alpha_3 x_3(t) \tag{3}$$

with

$$\begin{aligned} \alpha_1 &= a_{11} - ca_{21} \\ \alpha_2 &= a_{12} - ca_{22} \\ \alpha_3 &= a_{13} - ca_{23} \end{aligned} \tag{4}$$

If $x_3(t) = 0$, the criterion that we introduced in [2]-[3] tells us that two extrema, (maxima and/or minima depending on the types of sources) of the signed normalized kurtosis $k_s(c)$ appear for the two values of c which realize the extraction of each source:

$$\begin{aligned} c = c_{s1} = a_{11}/a_{21} \\ c = c_{s2} = a_{12}/a_{22} \end{aligned} \tag{5}$$

If we apply this kurtosis criterion when $x_3(t) \neq 0$ then the corresponding output signals are still mixtures of all the different source signals. Because we thus do not solve the problem in this configuration, we defined a new criterion based on the *differential normalized kurtosis* [4].

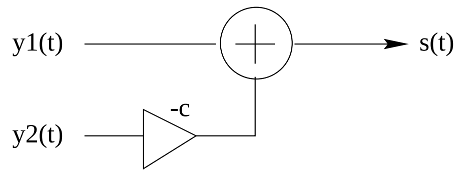

Figure 1: Separation system.

2.2 Differential Normalized Kurtosis

We consider the case when both x_1 and x_2 have long term non-stationarity but $x_3(t)$, the noise signal, has long term stationarity. We take two successive time domain $\mathcal{D}_1$ and $\mathcal{D}_2$ such that x_1 and x_2 are stationary onto $\mathcal{D}_1$ and then onto $\mathcal{D}_2$. In this case the 2^{nd} and 4^{th}-order cumulants are constant for each time domain and then only considered at time $t_1 \in \mathcal{D}_1$ and $t_2 \in \mathcal{D}_2$.
We pointed out in [4] that it is possible to cancel the noise influence during the search of the separation system coefficient c by simple differences between 2^{nd} and 4^{th}-order cumulants at time t_1 and t_2 .
The *differential normalized kurtosis* Δk_s is expressed as:

$$\Delta k_s = \frac{\Delta \text{cum}_4(s)}{[\Delta \text{cum}_2(s)]^2} = \frac{\text{cum}_4(s^2) - \text{cum}_4(s^1)}{[\text{cum}_2(s^2) - \text{cum}_2(s^1)]^2} \quad (6)$$

where Δ means the differential value between two instants t_1 and t_2. s^1 and s^2 represent the output of the separation system for the two successive instants t_1 and t_2.
It appears that because $x_3(t)$ has long term stationarity, its n^{th} order cumulant is a constant value for every time domain. Then the influence of this source, i.e. the noise, disappears in Δk_s.
We may then express Δk_s as:

$$\Delta k_s = \frac{\Delta k_{x_1} + \Delta k_{x_2} p^2}{(1+p)^2} \quad (7)$$

where

$$p = \frac{\alpha_2^2 \Delta \text{cum}_2(x_2)}{\alpha_1^2 \Delta \text{cum}_2(x_1)} \quad (8)$$

This expression is similar to the one for the case when $n_o = n_s$ obtained in our previous papers [2], [3]. But there are two main differences:

- First we consider here the cumulant differences between two samples taken at two successives instants t_1 and t_2.
- Then the ratio p can now be negative or positive, depending on the source signals properties. The concequence is that if $p = -1$ then the ratio (8) is infinite which is a problem. A future paper will show how this problem may be avoided.

We focus here on the case when $p > 0$. The expression and variations of Δk_s are then directly derived from those of k_s presented in [2]-[3] by replacing all cumulants by their differential versions. This entails that all extrema of given types of Δk_s vs c appear for the two values of c which solve the BSS problem between x_1 and x_2. Due to the limited number of sensors ($n_o < n_s$), the output $s(t)$ still includes noise. But it is important to notice that we extract only the desired source added with noise and not a mixture of the three source signals. We thus achieve a partial source separation limited to x_1 and x_2.
We now introduce different algorithms for searching these extrema.

3 DIFFERENTIAL SEPARATION ALGORITHMS

3.1 Iterative Methods

Iterative approaches can be used to solve our problem. Typically, with an initial value x_0, such algorithms compute other values x_1, x_2, ..., x_i until required convergence is reached. The main features of these algorithms are convergence speed, stability and accuracy.

We used two types of iterative algorithms: *Gradient* and *Newton*. We modified them in order to improve their performances, as we will see below.

3.1.1 Gradient Method

This method is used in the case of extremum search. We consider a function of only one variable called f.

Usual Version: The above-defined series $x_1, \ldots, x_j$ is built with the following expression:

$$x_{n+1} = x_n - \mu f'(x_n) \quad (9)$$

with $\mu > 0$ (resp. $\mu < 0$) when searching a minimum (resp. maximum) of the function f. Stability, convergence speed and accuracy depend on the choice of the initial value x_0 and adaptation gain μ. Typically, convergence towards the j^{th} extremum is garanteed only if the adaptation gain μ is bounded [5], i.e:

$$0 < |\mu| < \mu_j \quad (10)$$

where μ_j depends on the shape of f around the considered extremum.

Modified Version: We consider for example the case of a maximum search. If the function f contains several maxima, we can find all of them only if the adaptation gain μ satisfies (10) for each one. We must therefore choose μ so that:

$$\mu_{\text{chosen}} \leq min\{\mu_1, \mu_2, \ldots\} \quad (11)$$

But using the same value of μ for finding all extrema penalizes overall convergence speed as will now be shown. The standard equation (9) implies that the parameter μ should be set to a low value to avoid divergence in the areas where the first derivative f' is high. But this leads to slow convergence in low-derivative areas. These conflicting conditions lead one to use a trade-off value.

In order to avoid this kind of problem, we modified the standard algorithm. Two choices were possible:

- make μ as an adaptive gain,
- reduce the dynamics of the function f.

The first solution requires us to start from a low value of μ and may cause some stability problems if the range of the values of f varies with time before convergence is reached. Typically a value of μ which is adequate for some values of f might become too high when f varies and then might make the algorithm diverge. This explain the need to start from a low value of μ and the need to converge before the value of f changes. So, we choose the 2^{nd} solution. But this one has a main drawback: because μ is fixed, after a convergence period, the algorithm will alternately go from a value x_L situated on the left of the considered extremum to another value x_R situated on the right of the extremum. These two values define what we will call a "convergence tube". Its range, and therefore the convergence accuracy, depend on the value of the parameter μ. It can be shown that with better convergence speed (higher μ) we will have a lower accuracy.

The problem is now to find a way to reduce the high variations of f' without changing the monotony and the extrema positions of f. We propose two modified versions of the initial algorithm:

$$x_{n+1} = x_n - \mu \tanh(\eta f'(x_n)) \quad \text{with } \eta \geq 1 \quad (12)$$

$$x_{n+1} = x_n - \mu \text{sign}(f'(x_n)) \sqrt[k]{|f'(x)|} \quad (13)$$

The hyperbolic tangent is a regular function, defined onto $\Re$ and bounded between -1 and +1. That means that for every kind of function f', $\tanh(f')$ will have its variations between -1 and +1.
The k^{th} root gives also a high reduction of the dynamics of f but does not have any higher bound.

3.1.2 Newton Method

Classically, the Newton algorithm can be used for zero search or extremum search. Depending on which version is considered, by using the fact that an extremum of the function f coincide with a zero of its first derivative f', its expression is slightly different. We here apply it to the search of the extrema of the kurtosis Δk_s.
Considering a function f, the resulting algorithm can be expressed as:

$$x_{n+1} = x_n - \mu \frac{f'(x_n)}{f''(x_n)} \quad (14)$$

To avoid the same problems as with the gradient algorithm when the ratio $\frac{f'(x_n)}{f''(x_n)}$ takes high and low values, we modified it as:

$$x_{n+1} = x_n - \mu \tanh\left(\eta \frac{f'(x_n)}{f''(x_n)}\right) \quad (15)$$

3.2 Analytical Resolution

The values of c which realize partial source separation coincide with some extrema of the differential normalized kurtosis and therefore with zeros of its first derivative. In our configuration (two sources of interest for two sensors), the differential normalized kurtosis $\Delta k_s(c)$ may be expressed as a 4^{th}-order rational function in c, depending on observed signals' differential cumulants:

$$\Delta k_s(c) = \frac{\Delta\text{cum}_4(y_2)c^4 - 4\Delta\text{cum}_{13}(y_1,y_2)c^3}{(\Delta\text{cum}_2(y_2)c^2 - 2\Delta\text{cum}_{11}(y_1,y_2)c + \Delta\text{cum}_2(y_1))^2}$$

$$+ \frac{6\Delta\text{cum}_{22}(y_1,y_2)c^2 - 4\Delta\text{cum}_{31}(y_1,y_2)c + \Delta\text{cum}_4(y_1)}{(\Delta\text{cum}_2(y_2)c^2 - 2\Delta\text{cum}_{11}(y_1,y_2)c + \Delta\text{cum}_2(y_1))^2} \quad (16)$$

Its first derivative, can be expressed as:

$$\frac{d\Delta k_s}{dc} = \frac{(2a_1b_2 - a_2b_1)c^4 + (4a_1b_3 + a_2b_2 - 2a_3b_1)c^3}{(b_1c^2 + b_2c + b_3)^3}$$

$$+ \frac{(3a_2b_3 - 3a_4b_1)c^2 + (2a_3b_3 - a_4b_2 - 4a_5b_1)c}{(b_1c^2 + b_2c + b_3)^3}$$

$$+ \frac{(a_4b_3 - 2a_5b_2)}{(b_1c^2 + b_2c + b_3)^3} \quad (17)$$

with

$$\begin{array}{ll} a_1 = \Delta\text{cum}_4(y_2) & a_5 = \Delta\text{cum}_4(y_1) \\ a_2 = \Delta\text{cum}_{13}(y_1,y_2) & b_1 = \Delta\text{cum}_2(y_2) \\ a_3 = \Delta\text{cum}_{22}(y_1,y_2) & b_2 = -2\Delta\text{cum}_{11}(y_1,y_2) \\ a_4 = \Delta\text{cum}_{31}(y_1,y_2) & b_3 = \Delta\text{cum}_2(y_1) \end{array}$$

The zeros of the numerator can be found with Ferarri's method [6]. This method returns the exact roots of the equation, i.e. the values of c corresponding to the extrema of the function. It is important to notice that these exact solutions are found without any iterative loop as opposed to the algorithms that we defined above. But exact analytical solutions exist only if the order of the polynomial is equal to or lower than 4. This is true in our configuration with only two sources and two observations but in the other cases we must use iterative solutions.

4 EXPERIMENTAL RESULTS

In order to apply these algorithms, we use three signals. The desired source signals x_1 and x_2 are binary and their statistics change between each time domain with $\frac{\Delta\text{cum}_2(x_2)}{\Delta\text{cum}_2(x_1)}$ positive. The other one, x_3, considered as noise is stationary and uniform. The mixture used is:

$$\begin{array}{l} y_1(t) = x_1(t) + 0.9x_2(t) + 0.2x_3(t) \\ y_2(t) = 0.8x_1(t) + x_2(t) + 0.3x_3(t) \end{array} \quad (18)$$

We use the above separation system, giving us a signal $s(t)$.
With the considered signals, the two values of c which realize the separation between x_1 and x_2 correspond to the two maxima of $\Delta k_s(c)$. By combining (5) and (18), we see that the theoretical values of c achieving partial source separation are $c_1 = 1.25$ and $c_2 = 0.9$, as can be shown in Fig. 2. In our simulations, we search the value $c_1 = 1.25$ starting from $c = 5$.

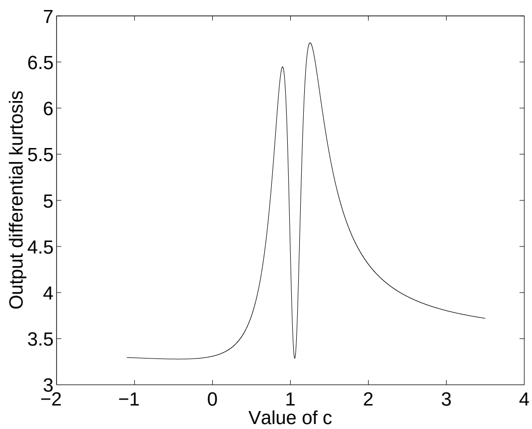

Figure 2: Variations of the differential normalized kurtosis of the separation system vs c.

4.1 Gradient Method

We here apply the gradient algorithms to search the maxima of $\Delta k_s(c)$. In order to compare the speed, stability and accuracy of the different algorithms, we considered a set of

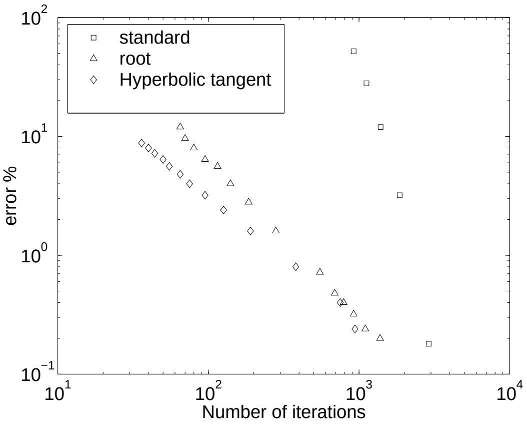

Figure 3: Number of iterations before convergence and obtained accuracy.

values for μ. Fig. 3 shows the resulting number of iterations before c remains in the "convergence tube" and the error percentage obtained. The latter parameter is defined as: $100 \times |c_R - c_L|/c_{\text{real}}$ where c_{real} is the theoretical searched value.

We can see on Fig. 3 that in every case the accuracy decreases exponentially vs the required number of iterations when we modify μ in order to improve the convergence speed. With the two modified versions we always need less iterations for the same precision than with the original one, with an advantage for the hyperbolic tangent.

In our tests, we started from a far initial value $c = 5$ situated in a low-derivating area as can be shown in Fig. 2. With the standard version of the algorithm we were then forced to take a low μ to avoid divergence. Convergence is then first very slow and becomes faster when $\frac{dk_s}{dc}$ increases. The compression introduced by the hyperbolic tangent or the 4^{th}-order root avoids this problem. This means that for a fixed value of μ we have fast convergence in both low and high-derivative areas.

Moreover, when we are simultaneously searching for the two values of c with the same μ, starting from two different initial values, we verify that the standard version of the algorithm easily diverges for one of the two extrema if we want to increase convergence speed. In this case, the parameter μ must be unreasonably low to hope any real-time convergence computation. The modified versions avoid this problem by being stable for a whole range of μ.

4.2 The Newton Algorithm

This algorithm has better convergence properties than gradient ones. The main difference, is that we do not have any "convergence tube". Due to the variations of the 2^{nd} derivative of $\Delta k_s(c)$ vs c, the adaptation may reach a higher speed than with gradient algorithms. For example, with the standard version we can reach a 0.001% accuracy in 100 iterations, which is much better than with gradient algorithms. With the modified version, the same accuracy can be reached in only 70 iterations. If we want "only" a 0.5% accuracy, the two algorithms give the result in 79 iterations for the standard version and 33 iterations for the modified one. Other experiments show that we approximately have a 1/2 ratio between the convergence speed of the modified and standard Newton algorithms.

But the most important feature is that convergence towards the two extrema with the same μ is very difficult in the standard version whereas the modified one is very much stable. This is the same result as we found with gradient algorithms, demonstrating once more time the robustness of the hyperbolic tangent version.

4.3 Analytical Approach

We obtained the two solutions $c_1 = 1.2521$ and $c_2 = 0.8987$. These solutions are very close to the theoretical values. These slight deviations result from the difference between the experimental and theoretical statistics of the considered source signals.

5 CONCLUSION AND FUTURE WORK

In this paper we have introduced and compared different methods for solving the source separation problem in a linear instantaneous mixture context with additive noise. We use the criterion introduced in [4], based on the *differential normalized kurtosis*. This criterion permits to achieve partial source separation of the desired source from noisy mixture. The obtained result consists of the desired source with noise, which can then be processed with classical noise cancellers.

Two numerical algorithms have been first studied: *Global Gradient* and *Newton*. In the two cases we modified their expression in order to improve their robustness and convergence speed. We showed that the Newton approach has better properties for solving our problem. We also presented an analytical solution, which can be used when we have a *two useful sources for two sensors* configuration. We investigate now the case of noisy convolutive mixtures.

References

[1] J.F. Cardoso, "Blind source separation: statistical principles", Proceedings of the IEEE. vol. 86, no. 10, October 1998.

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

[3] Y. Deville, M. Benali, "Un critère de séparation de sources fondé sur les kurtosis normalisés signés", GRETSI, Vannes, France, Sept. 13-17, 1999, vol. 4, pp. 1145-1148.

[4] Y. Deville, M. Benali, "Differential source separation: concept and application to a criterion based on differential normalized kurtosis", accepted to EUSIPCO, Tampere, Finland, 5-8 Sept. 2000.

[5] S. Haykin, Adaptive filter theory, 3^d edition. Upper Saddle River, New Jersey: Prentice Hall international editions, , 1996.

[6] E. Durand, "Solutions numériques des équations algébriques", Tome 1, Masson et C^{IE}, 1960.