Differential Fast Fixed-Point BSS for Underdetermined Linear Instantaneous Mixtures

Yannick Deville, Johanna Chappuis, Shahram Hosseini, and Johan Thomas

Laboratoire d'Astrophysique de Toulouse-Tarbes,
Observatoire Midi-Pyrénées - Université Paul Sabatier,
14 Av. Edouard Belin, 31400 Toulouse, France
{ydeville, shosseini, jthomas}@ast.obs-mip.fr,
johanna.chappuis@airbus.com

Abstract. This paper concerns underdetermined linear instantaneous blind source separation (BSS), i.e. the case when the number P of observed mixed signals is lower than the number N of sources. We propose a partial BSS method, which separates P supposedly non-stationary sources of interest one from the others (while keeping residual components for the other $N-P$, supposedly stationary, "noise" sources). This method is based on the general differential BSS concept that we introduced before. Unlike our previous basic application of that concept, this improved method consists of a differential extension of the FastICA method (which does not apply to underdetermined mixtures), thus keeping the attractive features of the latter algorithm. Our approach is therefore based on a differential sphering, followed by the optimization of the differential kurtosis that we introduce in this paper. Experimental tests show that this differential method is much more robust to noise than standard FastICA.

1 Introduction

Blind source separation (BSS) methods [9] aim at restoring a set of N unknown source signals $s_j(n)$ from a set of P observed signals $x_i(n)$. The latter signals are linear instantaneous mixtures of the source signals in the basic case, i.e.

$$x(n) = As(n) \tag{1}$$

where $s(n) = [s_1(n) \dots s_N(n)]^T$ and $x(n) = [x_1(n) \dots x_P(n)]^T$ are the source and observation vectors, and A is a constant mixing matrix. We here assume that the signals and mixing matrix are real-valued and that the sources are centered and statistically independent. Moreover, we consider the underdetermined case, i.e. $P < N$, and we require that $P \geq 2$. Some analyses and statistical BSS methods have been reported for this difficult case (see e.g. [2],[3],[4],[7],[10]). However, they set major restrictions on the source properties (discrete sources are especially considered) and/or on the mixing conditions. Other reported approaches use in several ways the assumed sparsity of the sources (see e.g. [1] and references

J. Rosca et al. (Eds.): ICA 2006, LNCS 3889, pp. 48–56, 2006.

therein). In [6], we introduced a general differential BSS concept for processing underdetermined mixtures. In its standard version, we consider 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 "noise sources". Our differential BSS concept then achieves the "partial BSS" of the P sources of interest, i.e. it yields output signals which each contain contributions from only one of these P sources, still superimposed with some residual components from the noise sources (this is described in [6]).

Although we first defined this differential BSS concept in a quite general framework in [6], we then only applied it to a simple but restrictive BSS method, which is especially limited to $P = 2$ mixtures and based on slow-convergence algorithms. We here introduce a much more powerful BSS criterion and associated algorithms, based on differential BSS. This method is obtained by extending to underdetermined mixtures the kurtotic separation criterion [5] and the associated, fast converging, fixed-point, FastICA algorithm [8], thus keeping the attractive features of the latter algorithm.

2 Proposed Differential BSS Method

2.1 A New BSS Criterion Based on Differential Kurtosis

The standard FastICA method [8], which is only applicable to the case when $P = N$ (or $P > N$), extracts a source by means of a two-stage procedure. The first stage consists in transferring the observation vector $x(n)$ through a real PxP matrix M, which yields the vector

$$z(n) = Mx(n). \tag{2}$$

In the standard FastICA method, M is selected so as to sphere the observations, i.e. so as to spatially whiten and normalize them. The second stage of that standard method then consists in deriving an output signal $y_i(n)$ as a linear instantaneous combination of the signals contained by $z(n)$, i.e

$$y_i(n) = w^T z(n) \tag{3}$$

where w is a vector, which is constrained so that $\| w \|= 1$. This vector w is selected so as to optimize the (non-normalized) kurtosis of $y_i(n)$, defined as its zero-lag 4th-order cumulant

$$K_{y_i}(n) = cum(y_i(n), y_i(n), y_i(n), y_i(n)). \tag{4}$$

Now consider the underdetermined case, i.e. $P < N$. We again derive an output signal $y_i(n)$ according to (2) and (3). We aim at defining how to select M and w, in order to achieve the above-defined partial BSS of the P sources of interest. To this end, we apply the general differential BSS concept that we described in [6] to the specific kurtotic criterion used in the standard FastICA method.

We therefore consider two times n_1 and n_2. We then introduce the differential (non-normalized) kurtosis that we associate to (4) for these times. We define this parameter as

$$DK_{y_i}(n_1, n_2) = K_{y_i}(n_2) - K_{y_i}(n_1). \tag{5}$$

Let us show that, whereas the standard parameter $K_{y_i}(n)$ depends on all sources, its differential version $DK_{y_i}(n_1, n_2)$ only depends on the non-stationary sources. Eq. (1), (2) and (3) yield

$$y_i(n) = v^T s(n) \tag{6}$$

where the vector

$$v = (MA)^T w \tag{7}$$

includes the effects of the mixing and separating stages. Denoting v_q, with $q = 1 \dots N$, the entries of v, (6) implies that the output signal $y_i(n)$ may be expressed with respect to all sources as

$$y_i(n) = \sum_{q=1}^{N} v_q s_q(n). \tag{8}$$

Using cumulant properties and the assumed independence of all sources, one derives easily

$$K_{y_i}(n) = \sum_{q=1}^{N} v_q^4 K_{s_q}(n) \tag{9}$$

where $K_{s_q}(n)$ is the kurtosis of source $s_q(n)$, again defined according to (4). The standard output kurtosis (9) therefore actually depends on the kurtoses of all sources. The corresponding differential output kurtosis, defined in (5), may then be expressed as

$$DK_{y_i}(n_1, n_2) = \sum_{q=1}^{N} v_q^4 DK_{s_q}(n_1, n_2) \tag{10}$$

where we define the differential kurtosis $DK_{s_q}(n_1, n_2)$ of source $s_q(n)$ in the same way as in (5). Let us now take into account the assumption that P sources are non-stationary, while the other sources are stationary. We denote by $\mathcal{I}$ the set containing the P unknown indices of the non-stationary sources. The standard kurtosis $K_{s_q}(n)$ of any source $s_q(n)$ with $q \notin \mathcal{I}$ then takes the same values for $n = n_1$ and $n = n_2$, so that $DK_{s_q}(n_1, n_2) = 0$ [1]. Eq. (10) then reduces to

$$DK_{y_i}(n_1, n_2) = \sum_{q \in \mathcal{I}} v_q^4 DK_{s_q}(n_1, n_2). \tag{11}$$

[1] Note that the "complete" stationarity of the sources $s_q(n)$ with $q \notin \mathcal{I}$ is sufficient for, but not required by, our method: we only need their differential kurtoses (and their differential powers below) to be zero for the considered times.

This shows explicitly that this differential parameter only depends on the non-stationary sources. Moreover, for given sources and times n_1 and n_2, it may be seen as a function $f(.)$ of the set of variables $\{v_q,\ q \in \mathcal{I}\}$, i.e $DK_{y_i}(n_1, n_2)$ is equal to

$$f(v_q,\ q \in \mathcal{I}) = \sum_{q \in \mathcal{I}} v_q^4 \alpha_q \tag{12}$$

where the parameters α_q are here equal to the differential kurtoses $DK_{s_q}(n_1, n_2)$ of the non-stationary sources. The type of function defined in (12) has been widely studied in the framework of standard kurtotic BSS methods, i.e. methods for the case when $P = N$, because the standard kurtosis used as a BSS criterion in that case may also be expressed according to (12) [2]. The following result has been established (see [9] p. 173 for the basic 2-source configuration and [5] for a general proof). Assume that all parameters α_q with $q \in \mathcal{I}$ are non-zero, i.e. that all non-stationary sources have non-zero differential kurtoses for the considered times n_1 and n_2. Consider the variations of the function in (12) on the P- dimensional unit sphere, i.e. for $\{v_q,\ q \in \mathcal{I}\}$ such that

$$\sum_{q \in \mathcal{I}} v_q^2 = 1. \tag{13}$$

The results obtained in [5],[9] imply in our case that the maxima of the absolute value of $f(v_q,\ q \in \mathcal{I})$ on the unit sphere are all the points such that only one of the variables v_q, with $q \in \mathcal{I}$, is non zero. Eq. (8) shows that the output signal $y_i(n)$ then contains a contribution from only one non-stationary source (and contributions from all stationary sources). We thus reach the target partial BSS for one of the non-stationary sources.

The last aspect of our method that must be defined is how to select the matrix M and to constrain the vector w (which is the parameter controlled in practice, unlike v) so that the variables $\{v_q,\ q \in \mathcal{I}\}$ meet condition (13). To this end, we define the differential correlation matrix of $z(n)$ as

$$DR_z(n_1, n_2) = R_z(n_2) - R_z(n_1) \tag{14}$$

where $R_z(n) = E\{z(n)z(n)^T\}$ is its standard correlation matrix. The differential correlation matrix $DR_s(n_1, n_2)$ of the sources is defined in the same way. It is diagonal, since the sources are assumed to be uncorrelated and centered, and its non-zero entries are the differential powers of the non-stationary sources, i.e.

$$DP_{s_q}(n_1, n_2) = E\{s_q^2(n_2)\} - E\{s_q^2(n_1)\}. \tag{15}$$

The BSS scale indeterminacy makes it possible to rescale these differential powers up to *positive* factors. Therefore, provided the diagonal elements of $DR_s(n_1, n_2)$

[2] In standard approaches, the summation for $q \in \mathcal{I}$ in (12) is performed over all $P = N$ sources and the parameters α_q are equal to the standard kurtoses $K_{s_q}(n)$ of all these sources. However, this has no influence on the discussion below, which is based on the general properties of the type of functions defined by (12).

corresponding to the P sources of interest are strictly positive for the considered times n_1 and n_2, they may be assumed to be equal to 1 without loss of generality. We then select the matrix M so that

$$DR_z(n_1, n_2) = I \tag{16}$$

and we control w so as to meet $\| w \| = 1$. This method is the differential extension of the sphering stage of the FastICA approach. As shown in Appendix A, these conditions on M and w guarantee that the constraint (13) is satisfied.

2.2 Summary of Proposed Method

The practical method which results from the above analysis operates as follows:

Step 1 Select two non-overlapping bounded time intervals (which correspond to n_1 and n_2 in the above theoretical analysis) such that all non-stationary[3] sources have non-zero differential kurtoses and positive differential powers (15). These intervals may be derived by only resorting to the observed signals, $x_i(n)$, e.g. as explained in [6].

Step 2 Compute an estimate $\widehat{DR}_x(n_1, n_2)$ of the differential correlation matrix of the observations, defined in the same way as in (14). Then perform the eigendecomposition of that matrix. This yields a matrix Ω whose columns are the unit-norm eigenvectors of $\widehat{DR}_x(n_1, n_2)$ and a diagonal matrix Λ which contains the eigenvalues of $\widehat{DR}_x(n_1, n_2)$. Then derive the matrix $M = \Lambda^{-1/2}\Omega^T$. This matrix performs a "differential sphering" of the observations, i.e. it yields a vector $z(n)$ defined by (2) which meets (16).

Step 3 Create an output signal $y_i(n)$ defined by (3), where w is a vector which satisfies $\| w \| = 1$ and which is adapted so as to maximize the absolute value of the differential kurtosis of $y_i(n)$, defined by (5). Various algorithms may be used to achieve this optimization, especially by developing differential versions of algorithms which were previously proposed for the case when $P = N$. The most classical approach is based on gradient ascent [9]. We here preferably derive an improved method from the standard fixed-point FastICA algorithm [8], which yields several advantages with respect to the gradient-based approach, i.e. fast convergence and no tunable parameters. Our differential fast fixed-point algorithm then consists in iteratively performing the following couple of operations:

[3] "non-stationary" here means "long-term non-stationary". More precisely, all sources should be stationary inside each of the two time intervals considered here, so that their statistics may be estimated for each of these intervals, by time averaging. This corresponds to "short-term stationarity". The above-mentioned "sources of interest" (resp. "noise sources") then consist of source signals whose statistics are requested to vary (resp. not to vary) from one of the considered time intervals to the other one, i.e. sources which are "long-term non-stationary" (resp. "long-term stationary").

1) Differential update of w

$$w = \left[E\{z(w^Tz)^3\} - 3w\right]_{n_2} - \left[E\{z(w^Tz)^3\} - 3w\right]_{n_1} \tag{17}$$

$$= \left[E\{z(w^Tz)^3\}\right]_{n_2} - \left[E\{z(w^Tz)^3\}\right]_{n_1} \tag{18}$$

where the expressions $\left[E\{z(w^Tz)^3\}\right]_{n_i}$ are resp. estimated over the two considered time intervals.

2) Normalization of w, to meet condition $\| w \| = 1$, i.e

$$w = w/||w||. \tag{19}$$

Step 4 The non-stationary source signal extracted as $y_i(n)$ in Step 3 is then used to subtract its contributions from all observed signals. The resulting signals are then processed by using again the above complete procedure, thus extracting another source, and so on until all non-stationary sources have been extracted. This corresponds to a deflation procedure, as in the standard FastICA method [8], except that a *differential* version of this procedure is required here again. This differential deflation operates in the same way as the standard deflation, except that the statistical parameters are replaced by their differential versions, as in (17). Here again, a parallel (differential) approach [8] could be considered instead of deflation.

3 Experimental Results

We now illustrate the performance of the proposed method for a configuration involving 2 linear instantaneous mixtures of 3 artificial sources. Each of the 2 non-stationary sources $s_1(n)$ and $s_2(n)$ consists of two 5000-sample time windows. Both sources have a Laplacian distribution $p(x) = 1/2\exp(-|x|)$ in the first window and a uniform distribution over $[-0.5, 0.5]$ in the second window. The "noise" source $s_3(n)$ has the same distribution over all 10000 samples.

The overall relationship between the original sources and the outputs of our BSS system reads $y(n) = Cs(n)$, where $C = [c_{ij}]$ is here a 2x3 matrix. If $s_1(n)$ and $s_2(n)$ appear without permutation in $y(n)$, c_{12} and c_{21} correspond to the undesired residual components of $s_2(n)$ and $s_1(n)$ resp. in $y_1(n)$ and $y_2(n)$ and should ideally be equal to zero. The "error" associated to the partial BSS of $s_1(n)$ and $s_2(n)$ may then be measured by the parameter $(E\{c_{12}^2\} + E\{c_{21}^2\})$, where the expectations $E\{.\}$ are estimated over a set of 100 tests hereafter. Equivalently, the quality of this partial BSS may be measured by the inverse of the above error criterion, i.e

$$Q = \frac{1}{E\{c_{12}^2\} + E\{c_{21}^2\}}. \tag{20}$$

We investigated the evolution of this criterion with respect to the input Signal to Noise Ratio (SNR) associated to the observed mixed signals, defined as

$$SNR_{in} = \sqrt{SNR_{in}^1 \ . \ SNR_{in}^2} \tag{21}$$

where the input SNR associated to each mixed signal $x_i(n)$ reads

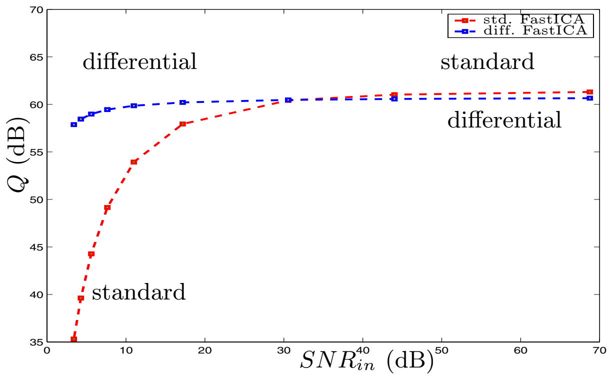

Fig. 1. Separation quality criterion Q of standard and differential FastICA methods vs input SNR

$$SNR_{in}^{i} = \frac{a_{i1}E\{s_1^2\} + a_{i2}E\{s_2^2\}}{a_{i3}E\{s_3^2\}} \quad i \in \{1, 2\}. \tag{22}$$

The input SNR was varied in our tests by changing the magnitude of the noise source $s_3(n)$. Fig. 1 shows the performance of the proposed differential BSS method and of the standard FastICA algorithm. This proves the effectiveness of our differential approach, since its quality criterion Q remains almost unchanged down to quite low input SNRs, i.e less than 5 dB, whereas the performance of standard FastICA already starts to degrade around 30 dB input SNR[4].

4 Conclusion

In this paper, we considered underdermined BSS. By using our differential BSS concept, we proposed a partial BSS method which has the same general structure as the kurtotic methods (especially FastICA) which have been developed for the case when $P = N$: it consists of a first stage which uses the second-order statistics of the signals, followed by a second stage which takes advantage of their fourth-order statistics. However, these stages are here based on new statistical parameters, that we introduce as the differential versions of the standard parameters. The proposed BSS method thus basically consists of a differential sphering, followed by the optimization of the differential kurtosis of an output signal. This optimization may especially be performed by using our differential version of the fast fixed-point algorithm which has been introduced in the standard FastICA approach, thus keeping the advantages of the latter algorithm. This has been

[4] For very high input SNRs (which is not the target situation for our approach !) standard FastICA performs slightly better than its differential counterpart. This probably occurs because the differential statistical parameters involved in the latter approach are estimated with a slightly lower accuracy than their standard version, partly because each expectation in the differential parameters is only estimated over one half of the available signal realization.

confirmed by our experimental tests, which show that our method is much more robust to noise than standard FastICA. Our future investigations will especially aim at extending our differential BSS method to convolutive mixtures.

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. 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.
3. P. Comon and O. Grellier, "Non-linear inversion of underdetermined mixtures", Proc. ICA'99, pp. 461-465, Aussois, France, 11-15 January 1999.
4. P. Comon, "Blind identification and source separation in 2x3 under-determined mixtures", IEEE Trans. on Sig. Proc, vol. 52, no. 1, pp. 11-22, Jan. 2004.
5. N. Delfosse, P. Loubaton, "Adaptive blind separation of independent sources: a deflation approach", Signal Processing, vol. 45, no. 1, pp. 59-84, July 1995.
6. 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.
7. F. Gamboa, "Separation of sources having unknown discrete supports", IEEE Workshop Higher-Order Stat., pp. 56-60, Begur, Spain, 12-14 June 1995.
8. A. Hyvarinen, E. Oja, "A fast fixed-point algorithm for Independent Component Analysis", Neural Computation, vol. 9, pp. 1483-1492, 1997.
9. A. Hyvarinen, J. Karhunen, E. Oja, "Independent Component Analysis", Wiley, New York, 2001.
10. A. Taleb, C. Jutten, "On underdetermined source separation", Proc. ICASSP-99, pp. 1445-1448, Phoenix, Arizona, 15-19 March 1999.

A Proof of Condition (13)

We first introduce the matrix H, defined as the diagonal matrix with entries equal to 1 for indices $q \in \mathcal{I}$ and 0 otherwise. We then define the vector

$$\tilde{v} = Hv \tag{23}$$

which is such that

$$||\tilde{v}||^2 = \sum_{q \in \mathcal{I}} v_q^2. \tag{24}$$

Besides, Eq. (23) and (7) yield

$$||\tilde{v}||^2 = w^T (MA) H (MA)^T w. \tag{25}$$

Moreover, Eq. (1) and (2) yield

$$DR_z(n_1, n_2) = (MA) DR_s(n_1, n_2) (MA)^T. \tag{26}$$

In addition, the properties of $DR_s(n_1, n_2)$ provided in Section 2.1 mean that $DR_s(n_1, n_2) = H$. Eq. (26) and (25) then yield

$$||\tilde{v}||^2 = w^T DR_z(n_1, n_2)w. \tag{27}$$

Therefore, (27) and (24) show that, if M is selected so that (16) is met and w is controlled so as to meet $\| w \| = 1$, then the requested condition (13) is met.