Fixed-point algorithms for convolutive blind source separation based on non-gaussianity maximization

Algorithmes à point fixe pour séparation aveugle de sources convolutive fondée sur la maximisation de non-gaussianité

Johan THOMAS, Yannick DEVILLE, Shahram HOSSEINI

Laboratoire d'Astrophysique de Toulouse-Tarbes
Université Paul Sabatier Toulouse 3 - CNRS
14 av. Edouard Belin, 31400 Toulouse, France

jthomas@ast.obs-mip.fr, ydeville@ast.obs-mip.fr, shosseini@ast.obs-mip.fr

Abstract This paper presents a new approach to the problem of blind separation of independent components in the case of MA convolutive mixtures of MA processes. It consists of an extension of the well-known Fast-ICA algorithm developed by Hyvärinen and Oja for instantaneous mixtures. We introduce a new type of sphering (convolutive sphering) that allows the use of non-gaussianity criteria and associated parameter-free fast fixed-point algorithms for the estimation of the source innovation processes. We prove the relevance of these criteria by reformulating the mixtures as linear instantaneous ones. We then describe associated kurtotic and negentropic time-domain algorithms. Test results are presented for artificial coloured signals and for speech signals.

Résumé Cet article présente une nouvelle approche pour la séparation aveugle de composantes indépendantes dans le cas des mélanges convolutifs MA de processus MA. Cette méthode peut être considérée comme une extension de l'algorithme Fast-ICA développé par Hyvärinen et Oja pour les mélanges instantanés. Nous introduisons un nouveau type de blanchiment ("sphering" convolutif) qui permet l'utilisation de critères de non gaussianité associés à des algorithmes rapides de type point fixe sans paramètre à ajuster, afin d'estimer les processus d'innovation des sources. Nous prouvons la pertinence de ces critères en reformulant le mélange sous forme instantanée. Nous dérivons ensuite les algorithmes à base de kurtosis et de négentropie qui en découlent. Des résultats de test sont présentés pour des signaux colorés artificiels et pour des signaux de parole.

1 Introduction

Blind source separation (BSS) consists in estimating a set of N unobserved source signals from P observed mixtures of these sources where the mixture parameters are unknown. Let us denote by $\mathbf{s}(n) = [s_1(n), \cdots, s_N(n)]^T$ the vector of sources and by $\mathbf{x}(n) = [x_1(n), \cdots, x_P(n)]^T$ the observations. In this paper, we suppose that the sources are zero mean and mutually statistically independent. We consider convolutive mixtures defined by a set of $P \times N$ unknown causal FIR filters which form a supposedly non-singular matrix $\mathbf{H}(z) = [H_{ij}(z)]$. The overall relationship between the sources and the observations then reads in the $\mathcal{Z}$ domain

$$\mathbf{X}(z) = \mathbf{H}(z).\mathbf{S}(z) \tag{1}$$

In this paper, each source $s_j(n)$ is assumed to be expressed in the $\mathcal{Z}$ domain as

$$S_j(z) = F_j(z).U_j(z) \tag{2}$$

where $F_j(z)$ corresponds to a causal FIR filter and $U_j(z)$ is the $\mathcal{Z}$ transform of an i.i.d. process $u_j(n)$, which is the innovation process of $s_j(n)$.

We can then express the mixing equation (1) in another way

$$\mathbf{X}(z) = \mathbf{G}(z).\mathbf{U}(z) \tag{3}$$

where $\mathbf{G}(z) = \mathbf{H}(z).\begin{pmatrix} F_1(z) & & 0 \\ & \ddots & \\ 0 & & F_N(z) \end{pmatrix}$.

The goal of convolutive BSS is typically to estimate the contributions of all sources in each observation, i.e. $H_{ij}(z).S_j(z)$. In deflation-based methods such as [1], this is achieved by using the following procedure

1. estimate the innovation process $u_j(n)$ of a source $s_j(n)$ from the observations.
2. identify and apply P coloring filters to $u_j(n)$ to recover the contributions of $s_j(n)$ in each observation.
3. subtract these contributions from all the observations.
4. set $N \leftarrow N-1$ and go back to step(1) if $N \neq 1$, in order to extract another source.

We here consider BSS methods which use non-gaussianity as a criterion to realize the first step of the above procedure. In the next section, we analyze the principles and limitations of the existing methods and we propose an approach to extend them so as to obtain the currently missing fast-converging kurtotic and negentropic methods for convolutive mixtures. The experimental performance of the proposed methods is presented in Section 3 and conclusions are drawn from this investigation in Section 4.

2 Analysis and extension of BSS methods based on non-gaussianity

2.1 Previously reported approaches

Delfosse and Loubaton [2] proposed the first deflation-based kurtotic BSS method for *linear instantaneous* mixtures, where the filters $H_{ij}(z)$ are replaced by simple scalar coefficients. This method first consists in deriving a sphered version of the observations at time n, i.e. a set of linear combinations of these observations composed of signals which are mutually uncorrelated at time n and which have unit variances. A first output signal is then derived as a linear combination y of the sphered observations, with a normalized coefficient vector $\mathbf{w}$ selected so as to maximize the square (or the absolute value) of the *non-normalized* kurtosis of y defined by $kurt(y) = E\left\{y^4\right\} - 3\left(E\left\{y^2\right\}\right)^2$ for a zero-mean signal. Delfosse and Loubaton proved in [2] that the local maxima of this criterion correspond to the separation points. They used a gradient-like method to maximize the above-mentioned kurtotic criterion. This requires one to select an adequate adaptation gain and anyway yields slow convergence. Hyvärinen and Oja solved this problem by introducing a fixed-point algorithm for optimizing the above criterion [3]. To put it briefly, this algorithm takes advantage of the fact that a constrained optimization of the considered criterion is to be performed and iteratively sets the adaptive vector $\mathbf{w}$ of combination coefficients to a normalized version of the gradient of the kurtotic criterion on the constraint surface. This algorithm does not require one to tune any parameter (such as the above adaptation gain in gradient-like methods) and was shown to converge very rapidly.

A different approach was proposed by Tugnait for *convolutive* mixtures [1]. It directly operates on the observations, i.e. without first sphering them, but then uses the absolute value of the *normalized* kurtosis of the output signal y, i.e. $kurt_N(y) = \frac{kurt(y)}{(E\{y^2\})^2}$, as the separation criterion. Tugnait proved that the separation points correspond to local maxima of this criterion. He proposed to optimize this criterion by using a gradient-based approach, which again yields slow convergence. The tests performed in our team [4] showed that, even when using Newton's optimization scheme, convergence remains slow, especially for high-order mixing filters.

This paper therefore aims at filling the gap which results from the above approaches, i.e. at introducing fast-converging kurtotic or negentropic methods for convolutive mixtures. To this end, we investigate how to extend to convolutive mixtures the approach based on sphering and fixed-point optimization of non-normalized kurtosis which has been proposed for instantaneous mixtures. Note that one could think of using the alternative approach, i.e. trying to introduce a fast fixed-point algorithm in the method without sphering and with normalized kurtosis which was proposed by Tugnait. However, this approach is not applicable because, unlike Hyvärinen's solution, Tugnait's method is not based on constrained optimization.[1]

2.2 A new method for estimating an innovation process

All above methods require a normalization, as the non-normalized kurtosis of y tends to infinity when the power of y tends to infinity. In Tugnait's approach, the criterion itself is normalized. Instead, the other two approaches use non-normalized kurtosis and are based on a normalization of the power of y. This results from the sphering stage of these linear instantaneous approaches, which yields $E\left\{y^2\right\} = \|\mathbf{w}\|^2$, so that by selecting $\mathbf{w}$ with $\|\mathbf{w}\|^2 = 1$, we

[1]As explained in [1], the filters in the BSS system are only normalized for practical reasons, but the criterion $kurt_N(y)$ does not depend on the overall scale.

guarantee that $E\left\{y^2\right\} = 1$. We here aim at extending this method to convolutive mixtures. As the first step of our approach, we therefore introduce a "convolutive sphering" of the observations, defined as follows. At any time n, we consider the column vector

$$\mathbf{v}(n) = [x_1(n+R), \cdots, x_1(n-R), \cdots, x_P(n+R), \cdots, x_P(n-R)]^T \tag{4}$$

which contains $M = (2R+1)P$ entries. We derive the M-entry column vector $\mathbf{x}'(n) = [x'_1(n), \cdots, x'_M(n)]^T$ defined as

$$\mathbf{x}'(n) = \mathbf{B}\mathbf{v}(n) \tag{5}$$

where the $M \times M$ matrix $\mathbf{B}$ is chosen so that

$$\begin{cases} E\left\{(x'_i(n))^2\right\} = 1\,, & \forall i \in \{1, \cdots, M\} \\ E\left\{x'_i(n)x'_j(n)\right\} = 0\,, & \forall i \neq j \in \{1, \cdots, M\} \end{cases} \tag{6}$$

With respect to $\mathbf{v}(n)$, the operation (5) may therefore be considered as conventional sphering, e.g. performed by a classical Principal Component Analysis and normalization. Now, with respect to the original observations $x_i(n)$, this may be interpreted differently. Indeed, Eq. (4) and (5) show that the signals $x'_i(n)$ are convolutive mixtures of the $x_i(n)$. Eq. (6) and (7) then mean that the signals $x'_i(n)$ are created so as to have unit variances and to be mutually uncorrelated, which may be seen as a spatio-temporal whitening and normalization of the observations $x_i(n)$.

With this pre-processing stage, Appendix A proves the equivalence

$$E(y^2) = 1 \Longleftrightarrow \|\mathbf{w}\| = 1 \tag{7}$$

where

$$y(n) = \sum_{m=1}^{M} w_m . x'_m(n) \tag{8}$$

and $\mathbf{w}$ is an extended vector of extraction coefficients which, together with (5), yields a convolutive combination y of the observations.

Our method then consists in maximizing the absolute value of the kurtosis of y defined by (9) under the constraint

$$\|\mathbf{w}\|^2 = \sum_{m=1}^{M} w_m^2 = 1. \tag{9}$$

Appendix B shows that this criterion lets us extract an estimate $e_j(n)$ of a delayed and scaled source innovation process $\alpha_j u_j(n-r)$, under some conditions.

Moreover, powerful algorithms for optimizing that criterion may then be straightforwardly derived from those previously reported for linear instantaneous mixtures, because Appendix B shows that the convolutive mixtures $v(n)$ studied in this paper may be reformulated as instantaneous mixtures in the considered conditions. Especially, the convolutive kurtotic fixed-point Fast-ICA algorithm that we propose as an extension of [3] based on our modified vector $\mathbf{w}$ then uses the iterations

1) $\mathbf{w} = E\left\{\mathbf{x}'(\mathbf{w}^T\mathbf{x}')^3\right\} - 3\mathbf{w}$
2) $\mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$

Similarly, for instantaneous mixtures, instead of using the kurtosis, another contrast function based on negentropy was proposed by Hyvärinen to estimate non-gaussianity [6]. It has been shown to yield better robustness and lower variance than the kurtotic approach. In particular, it is more robust to extreme values than the kurtosis criterion which involves a fourth-order moment, whose estimation is very sensitive to outliers. Furthermore, a fast and reliable algorithm was also developed by Hyvärinen for this type of function. It is based on Newton's method and is closely connected to the above fixed-point algorithm. We extend this negentropic algorithm to convolutive mixtures in the same way as the above kurtotic approach

1) $\mathbf{w} = E\left\{\mathbf{x}'g(\mathbf{w}^T\mathbf{x}')\right\} - E\left\{g'(\mathbf{w}^T\mathbf{x}')\right\}\mathbf{w}$
2) $\mathbf{w} = \frac{\mathbf{w}}{\|\mathbf{w}\|}$

where g and g' are the first and second derivatives of the contrast function G.

2.3 Overall proposed BSS methods

The above extraction stage provides an estimate $e_j(n)$ of a source innovation process up to a delay and a scale factor, that we can colour to obtain the contributions of the j^{th} source in each observation. This can be done by seeking the non-causal coloration filters $C_k(z) = \sum_{i=-R'}^{R'} \gamma_i z^{-i}$ which make the signals $c_k(n) * e_j(n)$ be the closest to $x_k(n)$ in the mean square sense [4]. One might use Newton's algorithm, which converges in one iteration for the criterion $E\left\{(c_k(n) * e_j(n) - x_k(n))^2\right\}$ since this criterion can be expressed as $\mathbf{c_k^T H c_k} + \mathbf{b^T c_k} + d$.

After subtracting these contributions from all observations, we obtain another mixture configuration with $N-1$ sources. The first step must then be iterated as explained in Section 1.

3 Experimental results

In this section, we illustrate the performance of our methods on several examples. These algorithms are first tested for $P = 2$ convolutive mixtures of $N = 2$ artificial coloured signals containing 100000 samples. The innovation processes $u_j(n)$ have uniform distributions. For this type of signals, the kurtosis appeared to be the best optimization criterion, as compared to negentropy.

Figure 1 shows the resulting output Signal to Interference Ratio (SIR) depending on the model order K that we define as the sum of the mixture and innovation coloration filter orders. For each source, this SIR is averaged over the two estimated source contributions. For each value of K, 100 experiments were made by varying the mixture and coloration filter coefficients with a uniform distribution. The order value R used to estimate the source innovation processes was set to $R = K$, since our tests showed that this yields a good trade-off between performance (SIR) and computational cost. Similarly, the order of the non-causal filters used to color the estimated innovation processes was set to $R' = 2K$. Fig. 1 shows that the means of SIRs are about 13 dB for the first source and a little lower for the second source. Comparisons showed that this performance is slightly better than with Tugnait's algorithm [4] and the processing time is much smaller (about 20 times smaller in some cases).

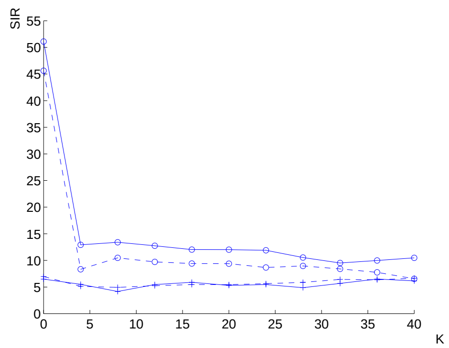

Fig 1. SIR of the kurtotic algorithm as a function of the model order K.

Circles : mean SIR
Plus signs : standard deviation of SIR

Solid lines : contributions of the first source
Dashed lines : contributions of the second source

In the next series of experiments, we fixed the model order to $K = 20$ to analyze the influence of the value of R on performance. Fig. 2 shows that a good compromize is reached between the mean and the standard deviation of SIR for the value $R = 20$. This confirms that $R = K$ is a good choice.

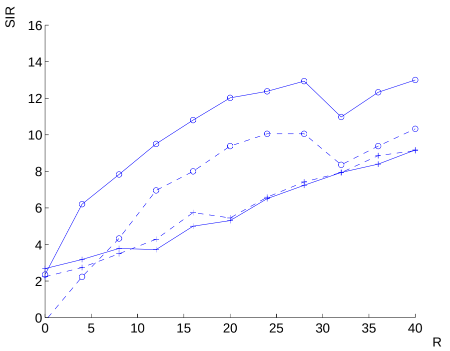

Fig 2. SIR of the kurtotic algorithm as a function of R.

Circles : mean SIR
Plus signs : standard deviation of SIR

Solid lines : contributions of the first source
Dashed lines : contributions of the second source

Another set of experiments was then made, still with artificial sources, by testing the case when the relationship (14) is satisfied. To this end, we used $P = 3$ observed mixtures of the above 2 sources and we set $R_1 = K - 1$ non causal lags and $R_2 = K$ causal lags in the vector $\mathbf{v}$ defined in (4). Fig. 3 shows that the mean SIR is about 5 dB higher than with $P = 2$ but so is the standard deviation. This high standard deviation may result from the higher number of parameters to be estimated here.

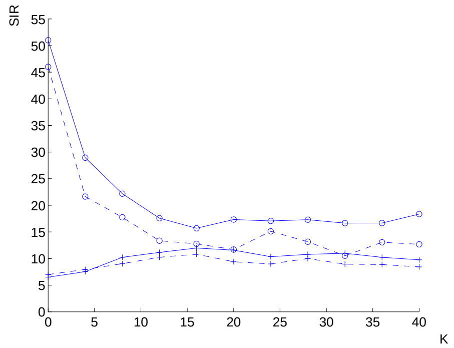

Fig 3. SIR of the kurtotic algorithm as a function of the model order K.

Circles : mean SIR
Plus signs : standard deviation of SIR

Solid lines : contributions of the first source
Dashed lines : contributions of the second source

The last series of experiments was carried out with 2 English speech sources sampled at 20 kHz during 5 seconds. As in the first set of experiments, we varied the order of the mixing filters $H_{ij}(z)$ and we again performed 100 experiments for each filter order. For these audio signals, the negentropic optimization criterion with $G(x) = e^{-\frac{x^2}{2}}$ turned out to yield better performance. In this case, the performance (Fig. 4) is more dependent on the mixture order and lower than in Fig. 1. The performance is also rather limited for the second extracted source. This could result from the MA process model (2) which is only approximately relevant for speech sources and involves higher filter orders.

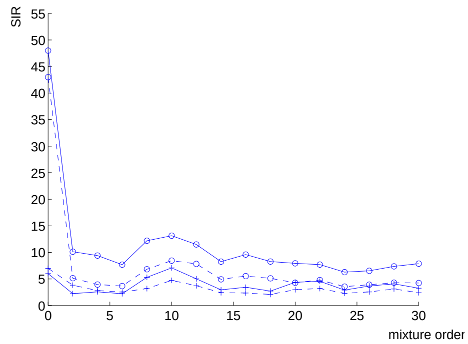

Fig 4. SIR of the negentropic algorithm as a function of the mixture order.

Circles : mean SIR
Plus signs : standard deviation of SIR

Solid lines : contributions of the first source
Dashed lines : contributions of the second source

4 Conclusion

In this paper, we have introduced new methods for Blind Source Separation in the convolutive case. These methods are based on a new type of sphering (convolutive sphering) which allows the use of fixed-point algorithms to maximize non-gaussianity criteria. Several test series have been performed for artificial coloured signals and for speech signals. The performance in terms of Signal to Interference Ratio appears slightly better than with Tugnait's method, whereas the processing time is much smaller.

Appendix A. Power normalization

The power of the output signal $y(n)$ defined by (9), reads

$$E\left\{y^2\right\} = E\left\{\left(\sum_{m=1}^{M} w_m.x'_m(n)\right)^2\right\} \tag{10}$$

$$= \sum_{m=1}^{M} w_m^2 E\left\{\left(x'_m(n)\right)^2\right\} + \sum_{m_1 \neq m_2} w_{m_1}.w_{m_2} E\left\{x'_{m_1}.(n).x'_{m_2}(n)\right\} \tag{11}$$

By constraining $x'(n)$ so as to meet (6) and (7), we get $E\left\{y^2\right\} = \sum_{m=1}^{M} w_m^2 = \|\mathbf{w}\|^2$. Therefore

$$E\left\{y^2\right\} = 1 \Longleftrightarrow \|\mathbf{w}\| = 1 \tag{12}$$

Appendix B. Relevance of considered criterion

The P considered observations $x_i(n)$ are expressed with respect to the N innovation processes $u_j(n)$ according to (3). They are therefore causal K^{th}-order FIR mixtures of these processes. Now consider the vector $\mathbf{v}(n)$ defined in (4) and composed of delayed observations. The analysis e.g. provided in [5] implies that if

$$PL \geq N(K+L) \tag{13}$$

where $L = (2R+1)$ is the number of lags, then $\mathbf{v}(n)$ may also be interpreted as a set of linear instantaneous mixtures of corresponding sources, which are here delayed and scaled versions of the innovation processes $u_j(n)$.

Therefore, if (14) is met, the investigation for instantaneous mixtures provided in [2] proves rigorously that, by maximizing the absolute value of the non-normalized kurtosis of the signal $y(n)$ defined in (9), we extract a delayed and scaled innovation process $\alpha_j u_j(n-r)$, whose practical estimate is denoted $e_j(n)$ hereafter.

If (14) is not met, the reformulated instantaneous BSS problem is underdetermined, i.e. it involves less observations than sources (note that this is especially the case when $P = N$). This underdetermination is related to the finite order of the extraction filters applied to the observations by the processing stages defined by (5) and (9). Some approximations are necessary in this underdeterminated case. However, when the ratio $\frac{PL}{N(K+L)}$ tends to 1 (which is the case when $P = N$ and L is high), a delayed and scaled innovation process may still be approximately estimated as a linear combination of the available observations whose absolute non-normalized kurtosis is maximum.

References

[1] J.K. Tugnait, "Identification and Deconvolution of Multichannel Linear Non-Gaussian Processes Using Higher Order Statistics and Inverse Filter Criteria", IEEE Transactions on Signal Processing, vol. 45, no. 3, pp. 658-672, March 1997.

[2] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach", Signal Processing, vol. 45, pp. 59-83, 1995.

[3] A. Hyvärinen and E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis", Neural Computation, vol. 9, pp. 1483-1492, 1997.

[4] F. Abrard, "Méthodes de séparation aveugle de sources et applications", Ph. D. thesis, pp. 29-41, University of Toulouse, France, 2003.

[5] C. Févotte, "Approche temps-fréquence pour la séparation aveugle de sources non-stationnaires", Ph. D. thesis, pp. 80-82, University of Nantes, France, 2003, http://www-sigproc.eng.cam.ac.uk/~cf269.

[6] A. Hyvärinen, "Fast and Robust Fixed-Point Algorithms for Independent Component Analysis", IEEE Transactions on Neural Networks, vol. 10, no. 3, pp. 626-634, May 1999.