Reprint of paper published in the International Journal of Adaptive Control and Signal Processing

Paper reference:
Y. Deville, O. Albu, N. Charkani, "New self-normalized blind source separation networks for instantaneous and convolutive mixtures", International Journal of Adaptive Control and Signal Processing, vol. 16, pp. 753-761, 2002.

New self-normalized blind source separation networks for instantaneous and convolutive mixtures

Yannick DEVILLE (1), Ovidiu ALBU (1), Nabil CHARKANI (2)

(1) Université Paul Sabatier
Laboratoire d'Acoustique, de Métrologie, d'Instrumentation
Bât. 3R1B2, 118 Route de Narbonne
31062 Toulouse Cedex
France.
Phone: +33 5 61 55 65 34.
Fax : +33 5 61 55 81 54.
E-mail: ydeville@cict.fr

(2) Philips Consumer Communications
Advanced Development
Route d'Angers
72081 Le Mans Cedex 9
France.

Abstract: The main blind source separation networks proposed in this paper apply to convolutive mixtures (including instantaneous ones). They have a recurrent or direct structure and they may use channel-specific separating functions. They are based on a self-normalized weight adaptation rule, which adaptively estimates the average powers of nonlinear functions of the network outputs. This allows us to control several aspects of the operation of these networks, esp. their convergence speed/accuracy trade-off. It also makes them more robust with respect to non-stationary situations. We analyze their convergence properties. We validate all these results by means of experimental tests performed with these networks, classical ones, and additional proposed linear instantaneous direct networks based on a normalization of their outputs. These tests esp. show that the proposed networks improve the convergence trade-off and that only these networks apply to highly mixed non-stationary sources.

Keywords: blind source separation, convergence analysis, convolutive mixtures, self-normalized algorithms.

1 Introduction

Blind source separation (BSS) methods aim at estimating a set of source signals from a set of mixtures of these signals[1]. We here especially consider neural BSS approaches. The first such neural network, intended for linear instantaneous mixtures and based on a recurrent structure, was proposed by Hérault and Jutten[2] (HJ). Moreau and Macchi[3] (MM) then adapted this approach to a direct structure. Cichocki, Kasprzak and Amari[4] (CKA) also defined related single- or multi-layer neural networks, which contain additional self-adaptive weights updated so as to normalize the "scales" of the network outputs.

We here introduce new BSS networks which share some features with the above ones and we extend them to convolutive mixtures. Their principles are defined in Section 2. Their convergence properties are then analyzed in Section 3. Section 4 describes the experimental performance of the classical networks, our above-mentioned extensions, and the networks with self-normalized outputs that we also introduce. Eventually, Section 5 presents the major conclusions drawn from this investigation.

2 Proposed networks with self-normalized weight updating

2.1 Networks for linear instantaneous mixtures

We here propose BSS neural networks for processing N linear instantaneous mixtures of N sources. They have the same direct or recurrent N-input N-output structures as the HJ and MM approaches and they may also be extended to several layers. Their original feature is the self-normalized algorithm used to update their weights c_{ij}. The "theoretical" discrete-time version of this algorithm reads:

$$c_{ij}(n+1) = c_{ij}(n) - a\frac{f[s_i(n)]}{\sqrt{E[f^2(s_i)]}}\,\frac{g[s_j(n)]}{\sqrt{E[g^2(s_j)]}} \quad i \neq j \in \{1 \dots N\} \tag{1}$$

and is to be contrasted with the classical rule used in the HJ and MM networks, i.e:

$$c_{ij}(n+1) = c_{ij}(n) - af[s_i(n)]g[s_j(n)] \quad i \neq j \in \{1 \dots N\}. \tag{2}$$

In both rules, a is a positive adaptation gain, $s_i(t)$ and $s_j(t)$ are the (estimated) centered versions of the network outputs $S_i(t)$ and $S_j(t)$, $E[.]$ denotes mathematical expectation, f and g are the so-called "separating functions". Two couples of such functions have been shown to be attractive in the classical networks, i.e:

$$f = (.)^3 \qquad \text{and} \qquad g = (.), \tag{3}$$

$$f = (.) \qquad \text{and} \qquad g = (.)^3, \tag{4}$$

because they are simple and resp. allow one to separate globally sub-Gaussian and super-Gaussian couples of sources. Comparing (1) and (2) shows that the proposed approach consists in normalizing the formely used signals $f[s_i(n)]$ and $g[s_j(n)]$ by their average powers. This yields several attractive features, mainly because the resulting algorithm is self-normalized from various points of view, as explained in Subsection 2.3.

In practice, the values of the above average powers used as normalizing terms are most often unknown, as the sources are unknown. They are therefore estimated, e.g. by first-order low-pass filtering. In this case, at each time step n, the terms $E[f^2(s_i)]$ and $E[g^2(s_j)]$

in (1) are resp. replaced by their adaptive estimates $N_{f,i}(n+1)$ and $N_{g,j}(n+1)$, which are updated according to the rules:

$$N_{f,i}(n+1) = N_{f,i}(n) + \eta(f^2[s_i(n)] - N_{f,i}(n)) \quad i \in \{1 \dots N\} \tag{5}$$
$$N_{g,j}(n+1) = N_{g,j}(n) + \eta(g^2[s_j(n)] - N_{g,j}(n)) \quad j \in \{1 \dots N\} \tag{6}$$

where η is a positive adaptation gain. The two single-layer practical networks thus obtained, resp. based on a Direct and a Recurrent structures, and both operating with Normalized Weight Updating terms, are called D-NWU and R-NWU hereafter.

2.2 Extension to convolutive mixtures

We now consider more general, i.e. convolutive, mixtures. We focus on a classical configuration (see Reference 5 and references therein), expresssed in the $\mathcal{Z}$ domain as:

$$Y_1(z) = X_1(z) + A_{12}(z)X_2(z) \tag{7}$$
$$Y_2(z) = A_{21}(z)X_1(z) + X_2(z), \tag{8}$$

where $Y_1(z)$ and $Y_2(z)$ are two observed mixed signals, $X_1(z)$ and $X_2(z)$ are two unknown statistically independent source signals and $A_{12}(z)$ and $A_{21}(z)$ are the transfer functions of two unknown causal Mth-order moving average (MA) filters. We consider two structures for processing these mixed signals, i.e. the recurrent structure in Fig. 1 and the direct structure with post-processing (shaping) filters in Fig. 2, where $C_{12}(z)$ and $C_{21}(z)$ are causal Mth-order MA separating filters. Several non-normalized rules have been reported for updating the weights $c_{ij}(n,k)_{k\in[0,M]}$ of these filters (see Reference 5 and references therein). We here introduce a self-normalized rule, defined as the convolutive extension of the normalization scheme that we proposed above for linear instantaneous mixtures. We also extend this rule by introducing separating functions f_i and g_i which may be specific to each channel i of the separating system[1]. This aims at keeping the required flexibility for subsequently optimizing these functions with respect to the sources extracted on each channel, as explained in Reference 5. The resulting theoretical rule reads [2]:

$$c_{ij}(n+1,k) = c_{ij}(n,k) - a\frac{f_i[s_i(n)]}{\sqrt{E[f_i^2(s_i)]}}\,\frac{g_i[s_j(n-k)]}{\sqrt{E[g_i^2(s_j)]}},\; i \neq j \in \{1,2\}, k \in [0,M]. \tag{9}$$

In practical situations, the normalizing terms $E[f_i^2(s_i)]$ and $E[g_i^2(s_j)]$ of (9) are again resp. replaced by $N_{f,i}(n+1)$ and $N_{g,j}(n+1)$, which are updated by means of (5)-(6).

2.3 Properties of the proposed networks

We here briefly review the major properties which result from the proposed self-normalization of the adaptation rule (more details are available in Reference 6), referring to the proposed rule (1) for simplicity. The expectation of each correcting term $\frac{f[s_i(n)]}{\sqrt{E[f^2(s_i)]}}\,\frac{g[s_j(n)]}{\sqrt{E[g^2(s_j)]}}$ of this rule is the correlation coefficient of the random variables $f[s_i(n)]$ and $g[s_j(n)]$ (assuming they are centered). The expectation of the weight update

[1] This extension also applies to the linear instantaneous networks that we introduced above.

[2] More precisely, when the direct structure is used, the signals $s_i(.)$ and $s_j(.)$ in (9) are typically replaced by the signals available before the post-processing stage of Fig. 2.

$\Delta c_{ij}(n) = c_{ij}(n+1) - c_{ij}(n)$ thus ranges between $-a$ and a. The maximum magnitude of the average weight update is therefore freely chosen by the user of the proposed approach, by selecting the desired gain value a. It does not depend on the source scales and statistics nor on the mixture coefficients, which is an attractive feature as these parameters are unknown (note that it is also independent from f and g). The magnitude of the fluctuations of the weights around an equilibrium point achieving BSS has similar properties, because the variance of the above correcting term is equal to one at such a point.

The convergence speed and accuracy of the proposed approach also have some independence properties with respect to the separating functions and source and mixture parameters, whereas they are controlled by the adaptation gain a. Especially, a common positive scale factor applied to both source signals has no influence on convergence speed and accuracy when using

$$f(x) = \lambda x^m \quad \text{and} \quad g(x) = \mu x^n, \tag{10}$$

where λ and μ are arbitrary positive scale factors (note that this includes the classical couples of functions defined in (3)-(4)).

The proposed networks are also robust with respect to non-stationary situations, which may result from time-varying mixtures or non-stationary sources, such as speech signals, in which large-magnitude periods alternate with low-magnitude ones. More precisely, by using short-term estimates of the mean powers $E[f^2(s_i)]$ and $E[g^2(s_j)]$ in the adaptation rule (1), these networks automatically keep their above-defined features despite (slow enough) modifications of the mixture coefficients and/or source parameters. On the contrary, the HJ and MM networks may only be used with care and yield degraded performance in such situations: their adaptation gain a should be set to a low value, to avoid low convergence accurary and even divergence during the periods when the signals have large (unknown) magnitudes, but for such a fixed adaptation gain convergence is very slow during the periods corresponding to low signal magnitudes. The output scale self-normalization used in the CKA network then yields even more fundamental limitations: the low- and large-magnitude periods of the sources result in the same output level, i.e. the restored sources are "compressed", which is not acceptable for the considered signals, such as speech.

3 Convergence of the practical networks

The locations and stability of the equilibrium points of the proposed networks define the types of sources that these networks can separate. These properties are derived by analyzing the local asymptotic behavior of the adaptation rules of these networks, by means of the Ordinary Differential Equation (ODE) method[7]. This analysis is here carried out for stationary centered sources, mixed in a convolutive way. Some steps are skipped due to space limitations, but they may be adapted e.g. from References 5 and 8. Especially, by using the same type of first-order approximation as in Reference 8, we here express the overall convolutive practical algorithm to be analyzed in vector form as:

$$\theta_{n+1} = \theta_n + H(\theta_n, \xi_{n+1}), \tag{11}$$

where θ_n, ξ_{n+1} and $H(\theta_n, \xi_{n+1})$ are column vectors defined as:

$$\begin{aligned} \theta_n = \; & [c_{12}(n,0), \ldots, c_{12}(n,M), c_{21}(n,0), \ldots, c_{21}(n,M), \\ & N_{f,1}(n), N_{f,2}(n), N_{g,1}(n), N_{g,2}(n)]^T, \end{aligned} \tag{12}$$

$$\xi_{n+1} = [y_1(n), y_2(n), s_1(n-1), \ldots, s_1(n-M), s_2(n-1), \ldots, s_2(n-M),]^T, \quad (13)$$

$$\begin{aligned} H(\theta_n, \xi_{n+1}) = & \Bigg[-a \frac{f_1[s_1(n)]}{\sqrt{N_{f,1}(n)}} \frac{g_1[s_2(n)]}{\sqrt{N_{g,2}(n)}}, \ldots, -a \frac{f_1[s_1(n)]}{\sqrt{N_{f,1}(n)}} \frac{g_1[s_2(n-M),]}{\sqrt{N_{g,2}(n)}}, \\ & -a \frac{f_2[s_2(n)]}{\sqrt{N_{f,2}(n)}} \frac{g_2[s_1(n)]}{\sqrt{N_{g,1}(n)}}, \ldots, -a \frac{f_2[s_2(n)]}{\sqrt{N_{f,2}(n)}} \frac{g_2[s_1(n-M),]}{\sqrt{N_{g,1}(n)}}, \\ & \eta(f_1^2[s_1(n)] - N_{f,1}(n)), \eta(f_2^2[s_2(n)] - N_{f,2}(n)), \\ & \eta(g_2^2[s_1(n)] - N_{g,1}(n)), \eta(g_1^2[s_2(n)] - N_{g,2}(n))]^T. \end{aligned} \quad (14)$$

The equilibrium points of (11) are all the constant state vectors θ^* for which

$$\lim_{n \to \infty} E_{\theta^*}[H(\theta^*, \xi_{n+1})] = 0, \quad (15)$$

where $E_{\theta^*}[.]$ denotes the mathematical expectation with respect to the probability law of the vector ξ_{n+1} for a given vector θ^*. When applying the condition (15) to the specific function $H(\theta_n, \xi_{n+1})$ defined in (14), the first $2(M+1)$ components of this function yield equations which implicitly define the separating filter weights corresponding to equilibrium points. These equations read as follows[3]:

$$E_{\theta^*}\left[\frac{f_i[s_i(n)]}{\sqrt{N_{f,i}(n)}} \frac{g_i[s_j(n-k)]}{\sqrt{N_{g,j}(n)}} \right] = 0, \quad i \neq j \in \{1,2\}, k \in [0, M]. \quad (16)$$

The expectation in (16) is taken for a fixed vector $\theta_n = \theta^*$, and therefore for fixed $N_{f,i}(n)$ and $N_{g,j}(n)$. The equilibrium condition (16) is therefore equivalent to:

$$E_{\theta^*}[f_i[s_i(n)]\ g_i[s_j(n-k)]] = 0, \quad i \neq j \in \{1,2\}, k \in [0, M]. \quad (17)$$

(17) is exactly the same as the equilibrium condition previously derived for the classical non-normalized networks. As the proposed networks also have the same input/output relationship as the classical ones, it may be shown easily that their filter weights at equilibrium points have the same expressions as in the classical networks. These expressions may therefore be found in the above-cited papers which describe the classical networks.

The analysis of the stability of these equilibrium points θ^* is based on the Jacobian matrix $J(\theta^*)$ of the system, as detailed e.g. in the above-cited papers. The corresponding calculations are outlined hereafter. $J(\theta^*)$ may first be split into sub-matrices, i.e:

$$J(\theta^*) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \quad (18)$$

where the sub-matrices A, B, C, D resp. have the following dimensions: $2(M+1)$x$2(M+1)$, $2(M+1)$x4, 4x$2(M+1)$, 4x4. It may be shown that only sub-matrix A has an influence on stability. This matrix A itself may be split into four $(M+1)$x$(M+1)$ sub-matrices A_{ij}. The top left sub-matrix of A, i.e. A_{11}, may be expressed as:

$$A_{11} = \alpha_{12}\beta_{11}, \quad (19)$$

[3]For readability, the limit $\lim_{n \to \infty}$ is often omitted in the mathematical expectations $E_{\theta^*}[.]$ below.

where :

$$\alpha_{ij} = \frac{1}{\sqrt{n_{f,i} n_{g,j}}} \tag{20}$$

$$n_{f,i} = E_{\theta^*}[f_i^2(s_i(n))] \quad i \in \{1,2\} \tag{21}$$

$$n_{g,j} = E_{\theta^*}[g_i^2(s_j(n))] \quad i \neq j \in \{1,2\} \tag{22}$$

and β_{11} is the matrix whose element (k,l), with $0 \leq k \leq M$ and $0 \leq l \leq M$, reads:

$$\lim_{n \to +\infty} E_{\theta^*}\left[-a \frac{\partial (f_1[s_1(n)] g_1[s_2(n-k)])}{\partial c_{12}(n,l)}\right]. \tag{23}$$

The same approach applies to the other matrices A_{ij} and leads to:

$$A = \begin{pmatrix} \alpha_{12}\beta_{11} & \alpha_{12}\beta_{12} \\ \alpha_{21}\beta_{21} & \alpha_{21}\beta_{22} \end{pmatrix}. \tag{24}$$

The matrices β_{ij} are the components of the Jacobian matrix of the non-normalized counterpart of the approach considered in this paper. Their explicit expressions are therefore available from our previous investigation[5]. This relationship between the non-normalized and normalized approaches may also be interpreted as follows. Combining (19)-(23) shows that each element (k,l) of A_{11} has the same expression (23) as the corresponding element of β_{11}, except that the original separating functions f_i and g_i are resp. replaced by the corresponding normalized functions F_i and G_i, according to the transform:

$$f_i(x) \rightarrow F_i(x) = \frac{f_i(x)}{\sqrt{E_{\theta^*}[f_i^2(x)]}} \quad \text{and} \quad g_i(x) \rightarrow G_i(x) = \frac{g_i(x)}{\sqrt{E_{\theta^*}[g_i^2(x)]}}. \tag{25}$$

The same principle applies to all other sub-matrices A_{ij}. Therefore, the overall matrix A for the proposed normalized algorithm operating with f_i and g_i is identical to the already known matrix for the non-normalized algorithm, but here operating with the functions F_i and G_i defined in (25). The resulting stability condition for the normalized algorithm is therefore directly derived from the one that we established in Reference 5 for the non-normalized algorithm, by using the transform (25).

More explicit results may then be derived by considering specific separating functions. For linear instantaneous mixtures, one thus e.g. shows that the R-NWU network operating with the separating functions (3) has exactly the same equilibrium points (in terms of weight values) as the corresponding HJ network, and exactly the same stability condition at each such point. It is therefore able to separate the same sources as the latter network, i.e. the globally sub-Gaussian couples of signals.

4 Experimental results

The experiments described below were performed with the following seven networks:

- The classical HJ, MM and CKA networks.
- The R-NWU and D-NWU instantaneous networks that we proposed above.
- The instantaneous version of the structure shown in Fig. 2, adapted with the classical rule (2). This modified version of the MM network yields self-Normalized Outputs, as explained below, and is therefore denoted MM-NO.

- An original network, based on the same structure as the MM-NO one, but adapted with our Normalized Weight Updating rule (1) and therefore denoted D-NWU-NO.

We here introduce the MM-NO and D-NWU-NO networks because, unlike the MM and D-NWU networks, they extract the sources with the same scale factor as the HJ and R-NWU networks. They therefore provide a more relevant way to compare the performance of direct and recurrent structures. Their outputs have "self-normalized scales", i.e. these signals are equal to the components of the sources contained by the observed signals.

All these networks are here operated with the separating functions defined by (4) because the considered sources are real speech source signals, and are therefore super-Gaussian. Performance is defined by the trade-off achieved by each considered network between convergence time and accuracy. The convergence time T_c is the number of samples required for all network weights to have converged to their equilibrium values. The convergence accuracy is assessed by means of the Signal-to-Noise Ratio Improvement $SNRI$ (defined in Part II of Reference 5), measured as follows. The source signals are sampled at 8 kHz over a period of 25 s, yielding 200000 samples. The network weights are permanently adapted, but the $SNRI$ is only measured over the last 10 seconds, i.e. 80000 samples, thus allowing the networks to previously converge in up to about 100000 samples.

We here only detail a single set of experiments, aiming at "easy conditions" so that all considered approaches apply and can be compared. More precisely, the two sources are here first rescaled to the range $[-1, 1]$, and only moderately mixed, i.e. the coefficients of the artificial mixing matrix are set to: $a_{11} = a_{22} = 1$, $a_{12} = 0.3$ and $a_{21} = 0.4$. The convergence speed/accuracy trade-off of the considered networks is illustrated in Fig. 3[4]. We here focus on its part of higher practical interest, i.e. on its left part, which corresponds to lower T_c. This figure then shows that, for convergence times lower than 20000 samples, our normalized networks (i.e. R-NWU, D-NWU and D-NWU-NO) yield a better convergence trade-off than their non-normalized counterparts (i.e. resp. HJ, MM, MM-NO). Moreover, the convergence times of the HJ, MM and MM-NO networks cannot be made lower than about 7000 samples, as these networks diverge when further increasing their adaptation gain. On the contrary, the R-NWU, D-NWU and D-NWU-NO networks can combine a T_c of only a few hundred samples with an acceptable $SNRI$. These results confirm the discussion of the properties of these networks provided in Subsection 2.3: whereas the non-normalized networks yield a dilemma between slow convergence and reduced accuracy for non-stationary signals, their normalized counterparts avoid this problem. Therefore, only the latter networks make it possible to achieve the low convergence time which is required in practical applications (not only to provide separated signals quickly enough to the end-user in the case of constant mixtures, but also to be able to track fast-time-varying mixtures which occur in practice). As expected, the other self-normalized approach, i.e. the CKA network, yields very poor performance for non-stationary sources: its $SNRI$ is lower than 0 dB, i.e. applying the mixed signals to this network in fact further decreases their quality.

Many more test results are detailed in Reference 6. They esp. show that for higher mixture ratios than above only the R-NWU, D-NWU and D-NWU-NO networks operate correctly, and that only these networks are insensitive to a common source scale factor.

[4]The MM and D-NWU networks are omitted in this figure, as they resp. yield almost the same performance as the MM-NO and D-NWU-NO networks.

5 Conclusions

We introduced several BSS networks in this paper. While some of them mainly aim at comparing relevant approaches in our tests and are restricted to direct instantaneous structures, our main contribution concerns direct and recurrent networks for instantaneous and convolutive mixtures. These networks are mainly based on a self-normalized weight adaptation rule, which yields various advantages over classical approaches. Especially, their convergence speed/accuracy trade-off is controlled by means of their adaptation gain. Moreover, these networks are thus robust with respect to non-stationary situations, whereas the classical networks yield degraded performance in this case. Our experimental tests proved that the proposed networks yield a better convergence trade-off than the classical ones and even showed that, for highly mixed non-stationary sources, these networks are not only attractive but required, as only they apply while the classical ones fail.

We also analyzed the convergence properties of the proposed networks. This investigation is based on the ODE approach, which here takes a more complex form than in the case of the classical networks, due to the additional adaptive parameters that we introduced. This analysis first proved that these networks resp. have the same equilibrium points as their classical counterparts. It also showed the effect of the proposed normalization on their stability conditions: the original separating functions are replaced by their normalized versions in these conditions. Specifically, we showed that these conditions, and therefore the separable sources, are the same as for classical networks in standard situations. This investigation opens the way to the subsequent optimization of the channel-specific separating functions of the proposed self-normalized networks, using the approach that we developed in Reference 5 for non-normalized networks.

References

1. Cardoso JF. Blind signal separation: statistical principles. *Proceedings of the IEEE* 1998; **86**(10):2009-2025.

2. Jutten C, Hérault J. Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing* 1991; **24**(1):1-10.

3. Macchi O, Moreau E. Self-adaptive source separation, Part I: convergence analysis of a direct linear network controled by the Hérault-Jutten algorithm. *IEEE Transactions on Signal Processing* 1997; **45**(4):918-926.

4. Cichocki A, Kasprzak W, Amari SI. Multi-layer neural networks with a local adaptive learning rule for blind separation of source signals. *Proceedings of the 1995 International Symposium on Nonlinear Theory and Its Applications (NOLTA '95).* Las Vegas: U.S.A, 1995; 61-65.

5. Charkani N, Deville Y. Self-adaptive separation of convolutively mixed signals with a recursive structure. Part I: stability analysis and optimization of asymptotic behaviour. *Signal Processing* 1999; **73** (3):225-254.

6. Deville Y, Albu O, Charkani N. New self-normalized source separation networks. *Lab. technical report, LAMI, University of Toulouse.* France, May 2002.

7. Benveniste A, Metivier M, Priouret P. *Adaptive algorithms and stochastic approximations. Applications of Mathematics vol. 22.* Springer-Verlag, 1990.

8. Deville Y, Charkani N. Convergence of source separation neural networks operating with self-normalized weight updating terms. *Proceedings of the International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99).* Aussois, France, 1999;227-232.

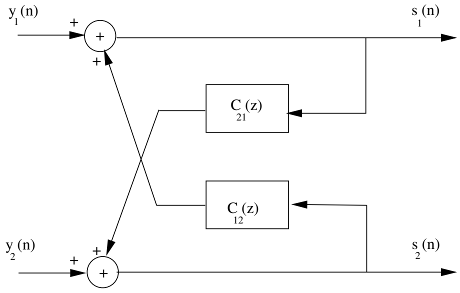

Figure 1: Recurrent BSS network for convolutive mixtures.

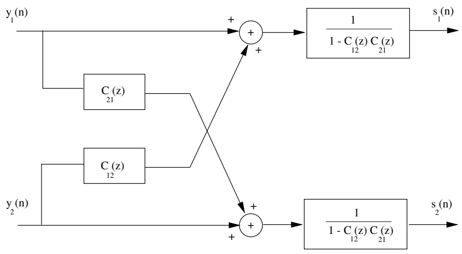

Figure 2: Direct BSS network for convolutive mixtures.

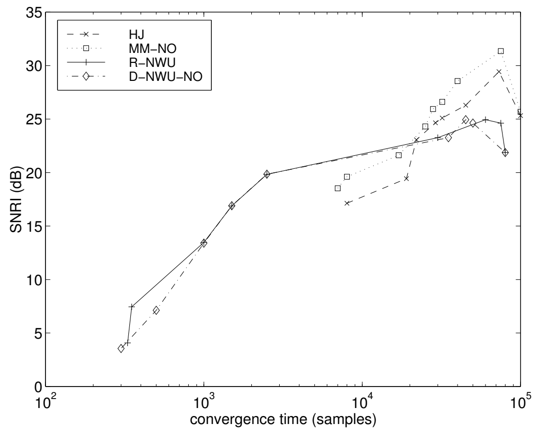

Figure 3: $SNRI$ vs convergence time T_c. The gain a is increased until: i) $SNRI < 5$ dB (R-NWU and D-NWU-NO networks; the CKA network yields $SNRI < 0$ dB whatever a and is therefore not shown) or ii) divergence occurs (HJ and MM-NO networks).