CONVERGENCE OF SOURCE SEPARATION NEURAL NETWORKS OPERATING WITH SELF-NORMALIZED WEIGHT UPDATING TERMS

Yannick DEVILLE

Laboratoire d'Acoustique, de Métrologie, d'Instrumentation (LAMI)
Université Paul Sabatier
38 Rue des 36 Ponts, 31400 Toulouse
France. E-mail: ydeville@cict.fr

Nabil CHARKANI

Philips Consumer Communications
Advanced Development
Route d'Angers
72081 Le Mans Cedex 9, France
E-mail: nabil.charkani@dev-lme.pcc.philips.com

ABSTRACT

In this paper, we define extended versions of two classical source separation neural networks, which provide self-normalized operation. We then analyze their convergence properties. We thus show how the conditions defining the positions and stability of their equilibrium points are modified by the proposed normalization of the weight updating terms. Especially, we prove that the standard version of a proposed network applies to globally sub-Gaussian i.i.d source signals.

1. INTRODUCTION

Blind source separation is a generic signal processing technique, which applies e.g. to antenna or microphone array processing [1]. In its "basic configuration", two signals $Y_1(t)$ and $Y_2(t)$ are available (e.g. from sensors), and these signals are unknown linear instantaneous mixtures of two unknown independent source signals $X_1(t)$ and $X_2(t)$, i.e:

$$Y_1(t) = a_{11}X_1(t) + a_{12}X_2(t) \quad (1)$$
$$Y_2(t) = a_{21}X_1(t) + a_{22}X_2(t), \quad (2)$$

where the terms a_{ij} are unknown real-valued constant mixture coefficients. The source separation problem then consists in estimating the source signals $X_j(t)$ from the mixed signals $Y_i(t)$, up to an arbitrary permutation and an arbitrary scaling factor.

Many solutions to this problem have been proposed since the beginning of the eighties. For a survey of these approaches, the reader may e.g. refer to [2]. In this paper, we restrict ourselves to a class of methods inspired from the field of artificial neural networks. The first of these networks was by the way one of the very first solutions to the source separation problem. This recurrent network was proposed by Hérault and Jutten [1] more than a decade ago and its convergence properties were analytically studied a few years later. Several papers were thus published by independent authors about its convergence in the "basic configuration" defined above. Sorouchyari [3] and Fort [4] used the same type of method, based on the Ordinary Differential Equation (ODE) technique [5]. This approach was then revisited and somewhat extended by Moreau and Macchi [6],[7]. Comon et al. [8] presented another method which yields different results. An approach bridging the gap between these two methods was then proposed by Deville [9], so that the convergence properties of the simplest versions of this network are now well defined.

Two extensions of the Hérault-Jutten network were also proposed in the literature for performing linear instantaneous source separation. On the one hand, Moreau and Macchi [6],[7],[10] introduced a direct (i.e. non-recurrent) version of the Hérault-Jutten network, based on the same adaptation rule. This direct network is attractive because it avoids the matrix inversion which must be performed with the recurrent version in order to derive the network outputs from its inputs and weights. Moreau and Macchi also studied the convergence properties of this network in the "basic configuration", esp. using the ODE approach[1].

On the other hand, Cichocki, Kasprzak and Amari [11] defined neural networks which may be considered as extensions of the above-mentioned ones. These extended networks contain additional self-adaptive weights, which are updated so as to normalize the "scales" of the network outputs. These networks were claimed to be thus able to process ill-conditioned mixtures and badly-scaled source signals to which the Hérault-Jutten network would not apply. Both the direct and recurrent versions of this type of neural networks were described, and it was also proposed to cascade them in a multilayer neural network in order to improve performance. The convergence properties of such networks in the "basic configuration" were analyzed using the ODE method in [12].

In this paper, we propose another type of self-normalized networks, i.e. the normalization here concerns the magnitudes of the weight updating terms (more precisely, their variances). As these networks are direct extensions of the Hérault-Jutten and Moreau-Macchi ones, the features of the latter networks which are of importance in the frame of this paper are first summarized in Section 2. The principles of the proposed networks are then depicted in Section 3 and their convergence properties are analyzed in Section 4. The

This work was partly performed when the authors were with the Laboratoires d'Electronique Philips S.A.S (LEP), at Limeil-Brévannes, France.

[1] In addition, Moreau and Macchi proposed and analyzed a mixed version of this network [6],[7].

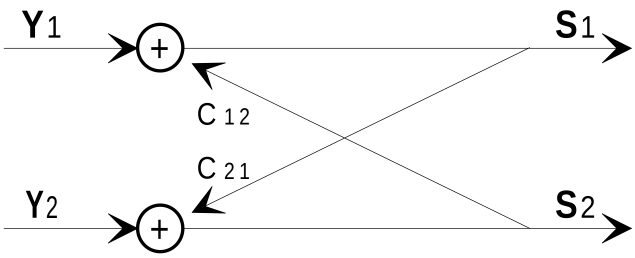

Figure 1: Hérault-Jutten recurrent neural network.

conclusions drawn from this investigation are presented in Section 5.

2. MAIN FEATURES OF THE CLASSICAL NETWORKS

As stated in Section 1, we here summarize the features of the two classical source separation artificial neural networks which are of interest in the frame of their extension to the proposed networks.

The Hérault-Jutten (HJ) approach [1] is based on the recurrent network shown in Fig. 1, where c_{12} and c_{21} are the adaptive weights of this network[2]. These weights are updated according to the following nonlinear unsupervised learning rule, based on the higher-order statistics of the output signals:

$$\frac{dc_{ij}(t)}{dt} = -af[s_i(t)]g[s_j(t)], \tag{3}$$

or in a discrete-time version:

$$c_{ij}(n+1) = c_{ij}(n) - af[s_i(n)]g[s_j(n)], \tag{4}$$

where 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 $S_j(t)$, and f and g are typically odd functions. Briefly, the motivation for this learning rule is to force the network outputs $S_1(t)$ and $S_2(t)$ to become (almost) statistically independent, thus making them become respectively proportional to the sources $X_1(t)$ and $X_2(t)$, or *vice-versa.*

When arbitrary odd nonlinear functions f and g are used, the network is only able to separate (some types of) symmetric sources [13]. As shown in [13], this restriction may be avoided by using either $f = (.)$ or $g = (.)$ (and not both because this would result in using only the second-order statistics of the signals and it would not guarantee that this algorithm reaches separation [1]). Especially, two sets of functions are attractive, due to their simplicity and to the type of sources to which they apply, i.e:

$$f = (.)^3 \quad \text{and } g = (.), \tag{5}$$

and

$$f = (.) \quad \text{and } g = (.)^3. \tag{6}$$

[2] As compared to the original papers by Hérault and Jutten, the signs of the weights c_{12} and c_{21} have been changed in Fig. 1, in order to be homogeneous with the other figure of the current paper. The rules (3) or (4) used to update these weights have been modified accordingly.

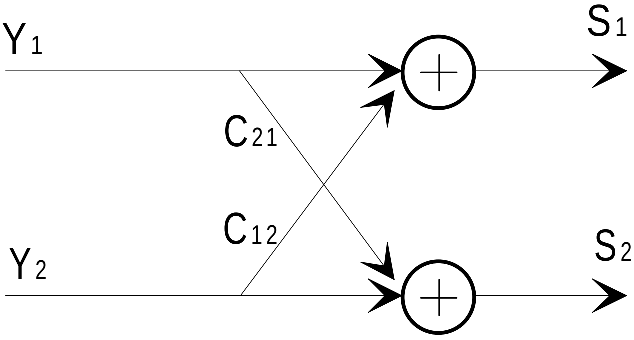

Figure 2: Moreau-Macchi direct neural network.

The choice between these two sets of functions is to be made depending on the considered type of sources (to ensure that the network weights converge to values which yield separated signals at the network outputs). More precisely, (5) applies to globally sub-Gaussian sources [3]-[7],[9], whereas (6) applies to globally super-Gaussian sources[3].

As stated in Section 1, the approach proposed by Moreau and Macchi (MM) [6],[7],[10], is based on a direct (i.e. non-recurrent) version of the HJ network (see Fig. 2), adapted with the same rule (3) as the latter network. Moreau and Macchi showed that for this network too, the functions defined in (5) allow to separate sub-Gaussian signals.

3. PRINCIPLES OF THE PROPOSED NETWORKS

As stated in Section 1, we here propose source separation neural networks which operate with self-normalized weight updating and which are derived from the HJ and MM networks. The need to introduce such a normalization in the latter networks may be justified as follows. The weights of these networks are updated according to the rule (3). The selection of the adaptation gain a of this rule leads to the classical convergence speed/accuracy trade-off of adaptive systems, which may be summarized as follows. If this gain is high, the increments of the network weights, which correspond to the right term of (3), are large. This has two consequences. On the one hand, the weights may thus converge quickly towards the values for which these networks provide separated signals, which is a desirable feature. But on the other hand, the fluctuations of these weights around their equilibrium values once convergence has been reached remain large. The networks thus provide "noisy" partly mixed signals on their outputs instead of cleanly separated sources. On the contrary, a low adaptation gain yields good convergence accuracy at the expense of low convergence speed.

Moreover, the achieved convergence trade-off also depends on the functions f and g of the rule (3). In particular, multiplying any of these functions by a given scaling factor is equivalent to multiplying the adaptation gain a by this factor. One could therefore think of jointly selecting the values of the gain a and functions f and g when designing the

[3] This may be shown e.g. by adapting the approach of [3] to the functions defined in (6).

considered system, so as to trade-off the convergence speed and accuracy of the network, depending on the requirements set on these two parameters in the considered application. However, (3) shows that the convergence compromise also depends on the magnitudes of the output signals (through the resulting values of $f[s_i(t)]$ and $g[s_j(t)]$) and therefore on the magnitudes of the mixed signals. For instance, if the functions f and g are set to (5), applying a scaling factor λ to both mixed signals (and therefore to $s_i(t)$ and $s_j(t)$) is equivalent to applying a factor λ^4 to the adaptation gain a. This means that these signal magnitudes have a huge influence on the convergence trade-off. Unlike the previous parameters, these magnitudes are not fixed when designing the considered system, as they depend on the intrinsic emission "levels" (or energies) of the sources and on their locations with respect to the sensors which measure their mixtures. Consequently, the convergence speed and accuracy are not controlled, which is a major drawback of the HJ and MM networks. A practical system based on these networks should therefore include some additional means for ensuring that its operation does not depend on the magnitudes of the measured mixed signals. In the solution to this problem proposed below, this feature (and others [14]) is inherently provided by the considered source separation neural networks.

More precisely, the type of networks that we propose is based on the same structures as above: they may contain one or several layers, and each layer may have a recurrent or a direct forms. These forms are resp. shown in Fig. 1 and 2 for the case of two source signals, and are extended to the case of a higher number of sources in the same way as for the HJ and MM networks. The proposed networks differ from the previous ones in the algorithm used to update their weights, which here reads:

$$\frac{dc_{ij}(t)}{dt} = -a\frac{f[s_i(t)]}{\sqrt{E[f^2(s_i)]}}\,\frac{g[s_j(t)]}{\sqrt{E[g^2(s_j)]}} \qquad (7)$$

or in a discrete-time version:

$$c_{ij}(n+1) = c_{ij}(n) - a\frac{f[s_i(n)]}{\sqrt{E[f^2(s_i)]}}\,\frac{g[s_j(n)]}{\sqrt{E[g^2(s_j)]}}, \qquad (8)$$

where $E[.]$ stands for mathematical expectation. When an equilibrium point achieving source separation is reached, the variance of the correcting term:

$$\frac{f[s_i(n)]}{\sqrt{E[f^2(s_i)]}}\,\frac{g[s_j(n)]}{\sqrt{E[g^2(s_j)]}} \qquad (9)$$

of the rule (8) is equal to one, thanks to the normalizing terms. This value is independent from the scales (and statistics) of the source and mixed signals, which is the main motivation for introducing this rule here, as explained above. Moreover, this variance is independent from the separating functions. The achieved convergence trade-off therefore only depends on the selected adaptation gain a, so that it can be fixed when designing the system. The proposed networks thus solve the problem of the classical approaches reported above.

In practice, the normalizing terms $\sqrt{E[f^2(s_i)]}$ and $\sqrt{E[g^2(s_j)]}$ are most often unknown and are therefore estimated using first-order low-pass filtering. In other words, at each time step, the terms $E[f^2(s_i)]$ and $E[g^2(s_j)]$ in (8) are resp. replaced by $N_{f,i}(n+1)$ and $N_{g,j}(n+1)$ which are updated according to the rules:

$$N_{f,i}(n+1) = N_{f,i}(n) + \eta(f^2[s_i(n)] - N_{f,i}(n)) \qquad (10)$$

$$N_{g,j}(n+1) = N_{g,j}(n) + \eta(g^2[s_j(n)] - N_{g,j}(n)) \qquad (11)$$

where η is a positive adaptation gain. The weight adaptation rule then reads:

$$c_{ij}(n+1) = c_{ij}(n) - a\frac{f[s_i(n)]}{\sqrt{N_{f,i}(n+1)}}\,\frac{g[s_j(n)]}{\sqrt{N_{g,j}(n+1)}}. \qquad (12)$$

For the sake of brevity, the single-layers versions of these two types of new networks, resp. based on a Direct and a Recurrent structures, and both operating with Normalized Weight Updating, are called D-NWU and R-NWU in the remainder of this paper.

4. CONVERGENCE PROPERTIES

The main properties of the proposed networks concern their convergence, i.e. the positions and stability of their equilibrium points, as this defines the types of sources that these networks can separate. These properties are derived by analyzing the asymptotic behavior of the adaptation rules of these networks. Unlike for the classical networks, the overall set of rules to be analyzed here not only includes the adaptation (12) of their weights, but also the adaptation (10)-(11) of the normalizing terms that we introduced. This extended case still remains in the scope of the general ODE convergence analysis method mentioned in Section 1, although it leads to more complex calculations, as will now be shown. This analysis is carried out for stationary independent identically distributed (i.i.d) centered[4] sources.

The weight adaptation rule (12) is first slightly modified as follows. For a small adaptation gain η, (10) yields the following first-order approximation:

$$\frac{1}{\sqrt{N_{f,i}(n+1)}} \simeq \frac{1}{\sqrt{N_{f,i}(n)}}\left[1 - \frac{\eta}{2}\left(\frac{f^2[s_i(n)]}{N_{f,i}(n)} - 1\right)\right]. \qquad (13)$$

A similar expression may be derived for $\frac{1}{\sqrt{N_{g,j}(n+1)}}$. Inserting both expressions in (12), one gets:

[4] In practical situations, the source, mixed and network output signals are not necessarily centered. As explained above, estimates of the mean values of the network outputs are then adaptively determined, so as to derive estimated centered versions of these signals, to be used in the adaptation rules. The overall adaptation algorithm of the system then includes additional updated parameters, i.e. the estimated mean outputs. The latter parameters should also be taken into account in the convergence analysis. This is achieved by using the same method as for the adaptive normalizing terms taken into account hereafter. Therefore, for the sake of clarity, we here restrict ourselves to a situation where only the latter terms need be introduced. More precisely, the signals are here assumed to be centered, so that the output signals $S_1(t)$ and $S_2(t)$ are used directly in the adaptation rules (10)-(12) as the centered signals $s_1(t)$ and $s_2(t)$ and the algorithm does not use estimated mean values for network outputs.

$$c_{ij}(n+1) = c_{ij}(n) - a\frac{f[s_i(n)]}{\sqrt{N_{f,i}(n)}}\frac{g[s_j(n)]}{\sqrt{N_{g,j}(n)}} - a\eta\epsilon(n), \quad (14)$$

where $\epsilon(n)$ corresponds to a small perturbation and depends on the fonctions f and g and on the network outputs. The ODE method also applies to this generalized version of the weight updating rule. More precisely, it can be shown [5] that its perturbation term $-a\eta\epsilon(n)$ may be neglected in the analysis as $\epsilon(n)$ remains uniformly bounded in a fixed compact (this results from the fact that the perturbation term then becomes negligible when a and η are very small). Combining this approximated weight adaptation rule with (10)-(11) yields the overall algorithm to be analyzed, which reads explicitly:

$$\left\{\begin{array}{lcl} c_{12}(n+1) & = & c_{12}(n) - a\frac{f[s_1(n)]}{\sqrt{N_{f,1}(n)}}\frac{g[s_2(n)]}{\sqrt{N_{g,2}(n)}} \\ c_{21}(n+1) & = & c_{21}(n) - a\frac{f[s_2(n)]}{\sqrt{N_{f,2}(n)}}\frac{g[s_1(n)]}{\sqrt{N_{g,1}(n)}} \\ N_{f,1}(n+1) & = & N_{f,1}(n) + \eta(f^2[s_1(n)] - N_{f,1}(n)) \\ N_{f,2}(n+1) & = & N_{f,2}(n) + \eta(f^2[s_2(n)] - N_{f,2}(n)) \\ N_{g,1}(n+1) & = & N_{g,1}(n) + \eta(g^2[s_1(n)] - N_{g,1}(n)) \\ N_{g,2}(n+1) & = & N_{g,2}(n) + \eta(g^2[s_2(n)] - N_{g,2}(n)) \end{array}\right. \quad (15)$$

This overall algorithm may be expressed in vector form as:

$$\theta_{n+1} = \theta_n + H(\theta_n, \xi_{n+1}), \quad (16)$$

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

$$\theta_n = [c_{12}(n), c_{21}(n), N_{f,1}(n), N_{f,2}(n), N_{g,1}(n), N_{g,2}(n)]^T, \quad (17)$$

$$\xi_{n+1} = [y_1(n), y_2(n)]^T, \quad (18)$$

$$\begin{array}{lcl} H(\theta_n, \xi_{n+1}) & = & \Bigg[-a\frac{f[s_1(n)]}{\sqrt{N_{f,1}(n)}}\frac{g[s_2(n)]}{\sqrt{N_{g,2}(n)}}, \\ & & -a\frac{f[s_2(n)]}{\sqrt{N_{f,2}(n)}}\frac{g[s_1(n)]}{\sqrt{N_{g,1}(n)}}, \\ & & \eta(f^2[s_1(n)] - N_{f,1}(n)), \\ & & \eta(f^2[s_2(n)] - N_{f,2}(n)), \\ & & \eta(g^2[s_1(n)] - N_{g,1}(n)), \\ & & \eta(g^2[s_2(n)] - N_{g,2}(n))\Bigg]^T. \end{array} \quad (19)$$

4.1. Equilibrium points

The equilibrium points of (16) are defined as the vectors θ^* for which

$$\lim_{n\to\infty} E_{\theta^*}[H(\theta^*, \xi_{n+1})] = 0, \quad (20)$$

where $E_{\theta^*}[.]$ denotes the mathematical expectation associated to the asymptotic probability law of the vector ξ_{n+1} for a given vector θ^*. When applying the condition (20) to the specific function $H(\theta_n, \xi_{n+1})$ defined in (19), the first two components of this function yield two equations which implicitly define the network outputs corresponding to equilibrium points and which read as follows[5]:

$$E\left[\frac{f[s_i(n)]}{\sqrt{N_{f,i}(n)}}\frac{g[s_j(n)]}{\sqrt{N_{g,j}(n)}}\right] = 0, \quad i \neq j \in \{1,2\}. \quad (21)$$

As the sources are supposedly i.i.d, the network outputs at any given equilibrium point are also i.i.d. Moreover, (15) shows that $N_{f,i}(n)$ and $N_{g,j}(n)$ are only derived from the previous values of the network outputs and are therefore here statistically independent from $s_i(n)$ and $s_j(n)$. The equilibrium condition (21) is therefore equivalent to:

$$E\left[\frac{1}{\sqrt{N_{f,i}(n)N_{g,j}(n)}}\right] E[f[s_i(n)]g[s_j(n)]] = 0, \quad i \neq j \in \{1,2\}, \quad (22)$$

which may be simplified into:

$$E[f[s_i(n)]g[s_j(n)]] = 0, \quad i \neq j \in \{1,2\}. \quad (23)$$

The explicit expressions of the network weights at each equilibrium point, with respect to the source statistics and mixture coefficients, may then be derived by combining (23) with: i) the mixture equations (1)-(2), which do not depend on the considered network, and ii) the specific input/output relationship of the considered network. However, these calculations may be avoided here by noting that (23) is exactly the same as the equilibrium condition derived for the classical non-normalized (i.e. HJ and MM) networks. As the HJ and R-NWU networks also share the same input/output relationship, defined by their common recurrent structure, one directly concludes that they have exactly the same equilibrium points. Similarly, the D-NWU network has the same equilibrium points as the MM one, due to their common direct structure.

4.2. Stability analysis

Each equilibrium point θ^* of (16) may be stable or not, depending on the properties of the function H and on the statistics of the vectors $(\xi_n)_{n\geq 0}$. The ODE approach used in this paper to analyze stability approximates the discrete recurrence (16), under some conditions[6] on H, by a continuous differential system that reads:

$$\frac{d\theta}{dt} = \lim_{n\to+\infty} E_\theta[H(\theta, \xi_{n+1})]. \quad (24)$$

The differential system (24) is locally stable in the vicinity of an equilibrium point θ^* if and only if (iff) the associated tangent linear system:

$$\frac{d\theta}{dt} = J(\theta^*)(\theta - \theta^*) \quad (25)$$

is stable, i.e. iff all the eigenvalues of $J(\theta^*)$ have negative real parts. For any state θ, $J(\theta)$ denotes the Jacobian matrix

[5]For readability, the limit $\lim_{n\to\infty}$ and the subscript θ^* are omitted in the mathematical expectations $E[.]$ below.

[6]The adaptation gains a and η should be sufficiently small. The other conditions on H concern its regularity [5].

of the system, i.e. the matrix of partial derivatives with entries:

$$J_{ij}(\theta) = \lim_{n\to+\infty} \frac{\partial(E_\theta[H(\theta,\xi_{n+1})]^{(i)})}{\partial\theta^{(j)}}, \qquad (26)$$

where $E_\theta[H(\theta,\xi_{n+1})]^{(i)}$ is the i^{th} component of $E_\theta[H(\theta,\xi_{n+1})]$ and $\theta^{(j)}$ is the j^{th} component of vector θ. The expression of $J(\theta^*)$ corresponding to the function H defined in (19) and to any fixed equilibrium point θ^* may be derived as follows. Considering the different natures of the components of H (and θ), we split $J(\theta^*)$ into sub-matrices, i.e:

$$J(\theta^*) = \begin{pmatrix} A & B \\ C & D \end{pmatrix}, \qquad (27)$$

where the sub-matrices A, B, C, D resp. have the following dimensions: 2x2, 2x4, 4x2, 4x4. Each of these sub-matrices is successively considered hereafter. As shown is (26), A is derived by taking into account only the first two components of H and θ, which are related to the weights and their adaptation rule (as opposed to the other adaptive parameters, i.e. the normalizing terms of the adaptation rules). Moreover, assuming that the order of the operations can be changed in (26), the calculation of A consists of two major steps, i.e: i) the computation of the partial derivatives $\frac{\partial(H(\theta,\xi_{n+1})^{(i)})}{\partial\theta^{(j)}}$, and ii) the derivation of the asymptotic expectations of the latter expressions. For the considered function H, the first step yields the intermediate matrix M composed of the following elements:

$$m_{11} = -a\frac{1}{\sqrt{N_{f,1}(n)N_{g,2}(n)}}\frac{\partial(f[s_1(n)]g[s_2(n)])}{c_{12}(n)} \qquad (28)$$

$$m_{12} = -a\frac{1}{\sqrt{N_{f,1}(n)N_{g,2}(n)}}\frac{\partial(f[s_1(n)]g[s_2(n)])}{c_{21}(n)} \qquad (29)$$

$$m_{21} = -a\frac{1}{\sqrt{N_{f,2}(n)N_{g,1}(n)}}\frac{\partial(f[s_2(n)]g[s_1(n)])}{c_{12}(n)} \qquad (30)$$

$$m_{22} = -a\frac{1}{\sqrt{N_{f,2}(n)N_{g,1}(n)}}\frac{\partial(f[s_2(n)]g[s_1(n)])}{c_{21}(n)}. \qquad (31)$$

As a second step, the matrix A is derived by taking the asymptotic expectation of M. To this end, it should be noted that the factors $\frac{1}{\sqrt{N_{f,i}(n)N_{g,j}(n)}}$ and $\frac{\partial(f[s_i(n)]g[s_j(n)])}{c_{kl}(n)}$ of each element of M are independent, due to the fact that the sources are i.i.d (this is based on the same principle as in the above analysis of equilibrium points). Taking the asymptotic expectation of M therefore yields:

$$A = \begin{pmatrix} \alpha_{12}\beta_{11} & \alpha_{12}\beta_{12} \\ \alpha_{21}\beta_{21} & \alpha_{21}\beta_{22} \end{pmatrix} \qquad (32)$$

with:

$$\alpha_{ij} = \lim_{n\to+\infty} E_{\theta^*}\left[\frac{1}{\sqrt{N_{f,i}(n)N_{g,j}(n)}}\right] \qquad (33)$$

$$\beta_{ij} = \lim_{n\to+\infty} E_{\theta^*}\left[-a\frac{\partial(f[s_i(n)]g[s_k(n)])}{c_{jl}(n)}\right] \quad \text{with } k \neq i,\ l \neq j. \qquad (34)$$

Moreover, the factors β_{ij} may be interpreted as follows. One may easily derive how the above presentation would be modified if the normalizing terms $N_{f,i}$ and $N_{g,j}$ were not included in H and θ: the Jacobian matrix would then only consist of A, which would still be expressed as in (32), except that it would not include the factors α_{ij} (which is equivalent to setting $\alpha_{ij} = 1$ in (32)). In other words, the factors β_{ij} are the elements of the Jacobian matrix of the non-normalized counterpart of the network considered here (i.e. of the HJ or MM network). Their explicit expressions are therefore available from the literature.

Using the same approach, it may be shown that all the elements of sub-matrix B are proportional to:

$$\lim_{n\to+\infty} E_{\theta^*}[f[s_i(n)]g[s_j(n)]]. \qquad (35)$$

Since θ^* is an equilibrium point, all these elements, and therefore the complete sub-matrix B, are zero. $J(\theta^*)$ is thus block-triangular, so that its Eigenvalues consist of all the Eigenvalues of its sub-matrices A and D. Therefore, its sub-matrix C need not be determined. Eventually, straightforward calculations show that:

$$D = -\eta I_4, \qquad (36)$$

where I_4 is the fourth-order identity matrix.

All the Eigenvalues of D are equal to $-\eta$. As η is set to a positive value, their real part is always negative, i.e. they always meet the above-defined stability condition. The stability of the considered networks therefore only depends on the Eigenvalues of A. Deriving them from (32) yields the eventual stability condition, which reads:

$$\begin{cases} \beta_{11}\beta_{22} - \beta_{12}\beta_{21} > 0 \\ \alpha_{12}\beta_{11} + \alpha_{21}\beta_{22} < 0 \end{cases} \qquad (37)$$

if $\Delta \geq 0$, where Δ is defined as:

$$\Delta = (\alpha_{12}\beta_{11} - \alpha_{21}\beta_{22})^2 + 4\alpha_{12}\alpha_{21}\beta_{12}\beta_{21}. \qquad (38)$$

If $\Delta < 0$, the stability condition is restricted to:

$$\alpha_{12}\beta_{11} + \alpha_{21}\beta_{22} < 0. \qquad (39)$$

These conditions are direct extensions of the ones for the non-normalized networks. More precisely, the expression of the Jacobian matrix of the latter networks is obtained by setting $\alpha_{ij} = 1$ in the generalized matrix (32), as explained above. As the above stability conditions were derived from (32), this entails that the stability conditions for the non-normalized networks are also obtained by setting $\alpha_{ij} = 1$ in (37)-(39), thus yielding:

$$\begin{cases} \beta_{11}\beta_{22} - \beta_{12}\beta_{21} > 0 \\ \beta_{11} + \beta_{22} < 0 \end{cases} \qquad (40)$$

if $\Delta \geq 0$, with $\Delta = (\beta_{11} - \beta_{22})^2 + 4\beta_{12}\beta_{21}$, and

$$\beta_{11} + \beta_{22} < 0 \tag{41}$$

if $\Delta < 0$. In other words, the stability conditions for the R-NWU and D-NWU networks are extensions of the conditions for the the HJ and MM networks, obtained by inserting the factors α_{ij} in the latter equations.

To make these results more explicit, we now apply them to the most clasical case, i.e. to the R-NWU network operating with the separating functions (5) (the other cases are treated in the same way). To this end, we use the expression of the Jacobian matrix of the corresponding HJ network at any equilibrium state provided in [3]. This expression especially entails that, for the R-NWU network:

$$\begin{aligned} \Delta &> 0 && (42) \\ \beta_{11} &= \beta_{22}. && (43) \end{aligned}$$

Due to (42), the stability condition for the R-NWU network is here (37). Besides, it should be noted that (33) implies $\alpha_{ij} > 0$. Combining this property with (43), (37) may be simplified in:

$$\begin{cases} \beta_{11}\beta_{22} - \beta_{12}\beta_{21} > 0 \\ \beta_{11} < 0 \end{cases} \tag{44}$$

This should be compared to the stability condition for the corresponding HJ network. For this network too, $\Delta > 0$. Its stability condition is therefore (40), which may here be simplified by using (43). This turns out to yield again (44). This means that, in the specific case considered here, the factors α_{ij} have no influence on the stability condition. Moreover, the overall result thus obtained (for i.i.d sources), is that the R-NWU network operating with the separating functions (5) has exactly the same equilibrium points as the corresponding HJ network, and exactly the same stability condition at each such point. It is therefore able to separate the same (supposedly i.i.d) sources as the latter network, i.e. the globally sub-Gaussian signals.

5. CONCLUSION

In this paper, we have defined extended versions of the classical HJ and MM source separation neural networks, which provide self-normalized operation (especially with respect to the source scales). We have analyzed their convergence properties by means of the ODE technique. We have thus shown (for i.i.d sources) that the proposed normalization of the weight updating rule does not modify the positions of the equilibrium points of the networks but yields a generalized form for their stability condition. In specific cases however, this condition turns out to be identical to the original one. Especially, we have shown that the R-NWU network operating with the classical functions (5) applies to globally sub-Gaussian signals. An industrial application of these results will be reported in another paper [15].

6. REFERENCES

[1] C. Jutten, J. Hérault, "Blind separation of sources, Part I: An adaptive algorithm based on neuromimetic architecture", Signal Processing, vol. 24, no. 1, pp. 1-10, July 1991.

[2] K. J. Pope, R.E. Bogner, "Blind signal separation I. Linear, instantaneous combinations", Digital Signal Processing, vol. 6, pp. 5-16, 1996.

[3] E. Sorouchyari, "Blind separation of sources, Part III: Stability analysis", Signal Processing, vol. 24, no.1, pp. 21-29, July 1991.

[4] J.-C. Fort, "Stabilité de l'algorithme de séparation de sources de Jutten et Hérault", Traitement du Signal, vol. 8, no. 1, pp. 35-42, Jan. 1991.

[5] A. Benveniste, M. Metivier and P. Priouret, "Adaptive algorithms and stochastic approximations", Applications of Mathematics, vol. 22, Springer-Verlag, 1990.

[6] E. Moreau, O. Macchi, "Self-adaptive source separation, Part II: comparison of the direct, feedback and mixed neural network", submitted to IEEE Transactions on Signal Processing.

[7] E. Moreau, "Apprentissage et adaptativité. Séparation auto-adaptative de sources indépendantes", PhD Thesis, Paris XI (Orsay), Feb 1, 1995.

[8] P. Comon, C. Jutten, J. Hérault, "Blind separation of sources, Part II: Problems statement", Signal Processing, vol. 24, no. 1, pp. 11-20, 1991.

[9] Y. Deville, "A unified stability analysis of the Hérault-Jutten source separation neural network", Signal Processing, vol. 51, no. 3, pp. 229-233, June 1996.

[10] O. Macchi, E. Moreau, "Self-adaptive source separation, Part I: convergence analysis of a direct linear network controled by the Herault-Jutten algorithm", submitted to IEEE Transactions on Signal Processing.

[11] A. Cichocki, W. Kasprzak, S.I. Amari, "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, Dec. 10-14, pp. 61-65.

[12] Y. Deville, "Analysis of the convergence properties of a self-normalized source separation neural network", Proceedings of the First IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC '97), pp. 69-72, Paris, France, April 16-18, 1997.

[13] Y. Deville, "Application of the Hérault-Jutten source separation neural network to multi-tag radio-frequency identification systems" (invited paper), Proceedings of Ecole des Techniques Avancées en Signal Image Parole, pp. 265-272, Grenoble, France, Sept. 2-6, 1996.

[14] N. Charkani, " Séparation auto-adaptative de sources pour les mélanges convolutifs. Application à la téléphonie mains-libres dans les voitures", PhD thesis, INP Grenoble, France, 1996.

[15] Y. Deville, J. Damour, N. Charkani, "Improved multi-tag radio-frequency identification systems based on new source separation neural networks", submitted to the International Workshop on Independent Component Analysis and Blind Signal Separation (ICA'99), Aussois, France, Jan. 11-15, 1999.