Recurrent Source Separation Structures as Iterative Methods for Solving Nonlinear Equation Systems

R. A. Ando, L. T. Duarte, D. C. Soriano, R. Attux, R. Suyama, Y. Deville, C. Jutten

Abstract— **In this work, we present a new interpretation of recurrent separation strategies in terms of local inversion based on iterative methods for solving nonlinear equation systems. From this interpretation, we firstly obtain a fresh perspective on the important methods proposed by Hérault and Jutten and by Hosseini and Deville for solving, respectively, linear and nonlinear blind source separation problems. Afterwards, a new separation structure for dealing with linear-quadratic models is proposed, leading to improvements from the convergence standpoint with respect to the basic network proposed by Hosseini and Deville. Finally, elements of a dynamical analysis of the proposed network are provided, indicating, aside from the well-known risk of divergence towards infinity, the possibility of convergence to limit cycles and strange attractors.**

Keywords— Source Separation, Nonlinear Mixtures, Iterative Methods for Equation Solving, Recurrent Networks.

I – INTRODUCTION

Recurrent structures have been used for source separation since the very origin of this field of research, although, presently, their use is generally related to applications involving nonlinear mixing models. In the context of linear-quadratic (LQ) mixtures, structures of this kind have been studied in great detail in a number of works (see for instance [5,8]), which allowed the establishment of a theoretical foundation for effectively dealing with the ensuing information retrieval task.

In this work, we present an interpretation of recurrent separating structures in terms of a local mixture inversion based on the iterative solution of nonlinear algebraic equation systems. This interpretation, which can be applied to any kind of memoryless mixing model, is then used to show that the Hérault-Jutten and Hosseini-Deville canonical structures are, in their respective domains, equivalent to numerical methods based on first-order (linear) approximations. Later, a line search strategy is used to derive a new recurrent structure for LQ mixtures, which is shown, with the aid of simulations performed for two different experimental setups, to provide performance improvements insofar as stability is concerned. Finally, elements of a dynamical analysis of the new networks are exposed, revealing *inter alia* the possibility of convergence to limit cycles and strange attractors. This analysis confirms the relevance of intrinsic stabilizing mechanisms for the operation of the Hosseini-Deville architecture [8]. The perspectives for future work are vast, as the proposed interpretation, in practice, puts the whole repertoire of numerical equation-solving methods at the disposal of those interested in designing separating systems for nonlinear scenarios.

The paper is organized as follows. In Section II, we discuss the basic aspects of the problem of source separation and also introduce the linear and linear-quadratic mixing models. In Section III, we present the relationship between recurrent separating structures and iterative numerical methods, revisiting the classical Hérault-Jutten and Hosseini-Deville approaches. Section IV contains the exposition of the obtained simulation results and of their analysis and Section V brings some conclusions and topics for future research efforts.

II - SOURCE SEPARATION

A recurrent problem in signal processing is to recover a set of source signals based on mixed versions of them. When only a very limited amount of information about the sources and the mixing process is available, this problem is referred to as blind source separation (BSS) [1,2]. Since its formulation, which dates back to the 1980's [3], the problem of BSS has attracted considerable attention, which is justifiable both in view of its theoretical and practical relevance. For the sake of illustration, it can be mentioned that BSS is part of applications in fields like biomedical signal processing, audio analysis and telecommunications, just to cite some emblematic instances [1].

Essentially, the derivation of a BSS method encompasses three steps. The first one is related to the definition of a device that will act as the separating system. Then, in a second step, one must define a separation criterion. For example, in BSS methods based on independent component analysis (ICA) [4], which are founded on the assumption that sources can be modeled as independent random variables, the separation task is fulfilled by building a device that somehow retrieves the statistical independence property lost after the mixing process. Finally, once the separating structure and a sound criterion have been chosen, one ends up with an optimization problem in hand. Therefore, one must choose a suitable numerical method in order to establish an effective framework.

The focus of this work is on the first step mentioned in the last paragraph, namely the definition of the separating structure. While this problem is simple in the classical scenario considered in BSS – linear instantaneous mixtures and equal number of sources and mixtures –, it becomes a major issue when the mixing process is nonlinear. Indeed, for some types

R. A. Ando, R. Attux and D. C. Soriano are with DCA/FEEC/University of Campinas (UNICAMP). L. T. Duarte is with FCA/UNICAMP. R. Suyama is with CECS/Federal University of ABC (UFABC). Y. Deville is with IRAP/Université de Toulouse. C. Jutten is with the Gipsa-lab/Université Joseph Fourier and also with the Institut Universitaire de France. E-mails: assato@dca.fee.unicamp.br, attux@dca.fee.unicamp.br, leonardo.duarte@fca.unicamp.br, soriano@dca.fee.unicamp.br, ricardo.suyama@ufabc.edu.br, Yannick.Deville@irap.omp.eu, christian.jutten@gipsa-lab.grenoble-inp.fr.

of nonlinear mixing systems, it may be difficult to obtain a separating device that is able to perfectly compensate for the mixing effects without violating the so-called separability requirements [1]. For instance, in order to deal with general nonlinear BSS problems, one could think of an ICA method applied to separating devices such as multilayer perceptrons (MLPs) or other systems endowed with universal approximation capability. However, such a strategy is risky, since a too flexible separating system may recover independent components that still correspond to mixed versions of the original sources [5]. In view of this problem, the research on nonlinear BSS has been mostly conducted on a case-by-case basis, in which the separating structure is defined according to the nature of the nonlinear mixing system.

Among the most studied nonlinear mixing systems are those based on the linear-quadratic (LQ) model [5], for which one can find applications in the context of several problems, like separation of scanned images [9], design of smart chemical sensor arrays [10] and hyperspectral image processing [11]. Later in this section, we shall briefly review the LQ model, paying special attention to the recurrent separating system proposed in [5]. This solution was inspired by the classical linear BSS method proposed in [3], which is discussed in the sequel.

II.1 – Linear Mixtures

Let the vectors $\mathbf{s}(n)$ and $\mathbf{x}(n)$ represent the sources and the mixtures, respectively. Classically, for an instantaneous linear mixing process, we have:

$$\boldsymbol{x}(n) = \boldsymbol{A}\boldsymbol{s}(n) \tag{1}$$

where $\boldsymbol{A}$ denotes the so-called mixing matrix. Most BSS methods are developed to deal with the determined case, in which the number of sources is equal to the number of mixtures and, as a consequence, $\boldsymbol{A}$ is a square (and supposedly full-rank) matrix.

The first solution to the linear BSS problem expressed in Equation (1) was obtained by Hérault, Jutten and Ans [3]. Essentially, their proposal was to cancel, by adjusting the coefficients m_{12} and m_{21} (see Equation (2) below), the nonlinear correlation between the fixed points, obtained for each $\boldsymbol{x}(n)$, of a recurrent network, which, for the case of two sources, is given by

$$\begin{aligned} y_1(m+1) &= x_1(n) - m_{12}y_2(m) \\ y_2(m+1) &= x_2(n) - m_{21}y_1(m) \end{aligned} \tag{2}$$

Note that source separation in this case can be carried out by simply considering the separating system given by

$$\boldsymbol{y}(n) = \boldsymbol{W}\boldsymbol{x}(n) \tag{3}$$

where $\boldsymbol{W}$ is the separating matrix. However, there are some interesting aspects regarding the recurrent network presented in (2). For instance, this recursive system can be easily implemented on analog systems. Moreover, this network has inspired solutions in more complex cases, such as those arising from LQ models, which will be discussed in the next section.

II.2 – Linear-Quadratic Mixtures

A natural extension with respect to the linear model defined by Equation (1) can be introduced by considering quadratic terms. For instance, in the so-called linear-quadratic (LQ) model, the mixing process, considering the case of two sources and two mixtures, is given by:

$$\begin{aligned} x_1(n) &= s_1(n) - L_{12}s_2(n) - Q_1 s_1(n)s_2(n) \\ x_2(n) &= s_2(n) - L_{21}s_1(n) - Q_2 s_1(n)s_2(n) \end{aligned} \tag{4}$$

Contrarily to the linear case, the definition of a separating device for the mixing process (4) is not direct [5], especially when the number of sources is greater than two. In view of this problem, Hosseini and Deville [5] proposed a separating strategy based on the following recurrent system:

$$\begin{aligned} y_1(m+1) &= x_1(n) + L_{12}y_2(m) + Q_1 y_1(m)y_2(m) \\ y_2(m+1) &= x_2(n) + L_{21}y_1(m) + Q_2 y_1(m)y_2(m) \end{aligned} \tag{5}$$

The rationale behind this network is that, when the mixing parameters are known in advance (for example, for the non-blind case), then, for a given sample of mixtures, $\boldsymbol{x}(n)$, the sources $s_1(n)$ and $s_2(n)$ will correspond to an equilibrium solution of (5). An important question, therefore, is to analyze whether this equilibrium point is stable or not. This issue is discussed in [5]. In addition to the structure (5), Hosseini and Deville proposed an extended version of their basic network (5), which includes a self-feedback that stabilizes the network [8]. In this work, however, we shall use only the basic version of the network, with the intention of analyzing the extended version in a future work.

Of course, in the context of BSS, the coefficients of the recurrent network expressed in (5) must be adjusted. This can be done by setting up, for instance, an ICA-based method that tries to maximize the independence between the signals retrieved by the network. In the present work, though, given that the focus is on the network itself, we will study only the non-blind case - for which the network coefficients are known - treated, for instance, in [8],

III – Source Separation Methods and Equation Solving

The main idea behind source separation methods is, given a certain mixing model and a number of mixtures, to be able to find the original sources up to information-preserving ambiguities (typically related to permutation and scaling [1]). Each signal can be considered to be a vector containing a set of samples acquired during a certain period of time.

As previously mentioned, the source separation problem can either be said to be non-blind, for which the mixing coefficients are either known a priori or potentially estimated with the aid of a set of source and mixture values (as explained in [8]); or it can be said to be blind, in which case the coefficients are unknown and cannot be estimated in accordance with the previously described (non-blind) strategy.

In the case of linear mixtures, as shown in (1), the mixing process can be understood in terms of a matrix $\boldsymbol{A}$, and, in the context defined by (3), the canonical separating solution would be a matrix $\boldsymbol{W}$ equal to the inverse of $\boldsymbol{A}$, without forgetting the aforementioned potential ambiguities. For general, memoryless non-linear mixtures, however, even in a case for which the mixing model is known, an analytical solution can be difficult or even impossible to find, which establishes a clear demand for iterative methods. Due to the memoryless nature of the model, we can consider the mixture samples one at a time, without having to keep track of the previous values. For each instant n, a dynamical process regulated by an alternative time index m can be used to, ideally, reconstruct the original sources from the successive reached equilibria.

Interestingly, as will be shown in the following, this dynamical approach can be understood not only as a means to reach certain equilibrium points, but as a general strategy to solve nonlinear algebraic equation systems and locally perform an inversion of the mixing process. Firstly, we will show that this interpretation brings a new perspective regarding the equivalence between the Hérault-Jutten and Hosseini-Deville approaches in their respective validity domains, and, moreover, indicates the possibility of using a number of different numerical methods for dealing with virtually any nonlinear mixing model.

III.1 – Analysis of the Hérault-Jutten Approach

Let us consider the linear mixing model shown in (1). As already stated, it can be globally inverted by using a system of the form of (3), if the inverse of $\boldsymbol{A}$ can be estimated. However, let us consider the problem from another point of view. If we rewrite (1) as:

$$F(\boldsymbol{s}(n)) = \boldsymbol{x}(n) - \boldsymbol{A}\boldsymbol{s}(n) = \boldsymbol{0} \tag{6}$$

we may consider the problem of estimating the sources as being equivalent to the task of numerically finding, for each available value of $\mathbf{x}$(n), an estimate $\mathbf{y}$(n) of $\mathbf{s}$(n). In other words, the problem can be understood in terms of finding the roots of $F(.)$ or of iteratively solving a system of linear equations. A sort of "first-order" method to accomplish this task assumes the form of:

$$\boldsymbol{y}(m+1) = \boldsymbol{y}(m) - \aleph F[\boldsymbol{y}(m)] \tag{7}$$

considering $\aleph$ as a diagonal matrix whose non-null elements account for the effective use of different, positive step-sizes for each direction. Using (6) and (7), we have

$$\begin{aligned}\boldsymbol{y}(m+1) &= \boldsymbol{y}(m) - \aleph\boldsymbol{x}(n) + \aleph\boldsymbol{A}\boldsymbol{y}(m) \\ &= \boldsymbol{G}\boldsymbol{x}(n) + (\boldsymbol{I} - \boldsymbol{G}\boldsymbol{A})\boldsymbol{y}(m)\end{aligned}$$

where $\boldsymbol{G} = -\aleph$. This structure is equivalent, for a specific choice of $\boldsymbol{G}$, to a discrete-time version of the classical Hérault-Jutten approach to source separation [3], as described, for instance, in [5]. Therefore, in a certain sense, aside from the canonical interpretation of the validity of this approach in terms of equilibrium points, it is possible to interpret it as an iterative process to solve a linear system of algebraic equations.

III.2 – Analysis of Recurrent Networks for LQ Model

For the LQ model, we will now show that the separating system proposed by Hosseini and Deville [5], described in section II.2, is also equivalent to a "first-order" method for iteratively inverting the mixing process. For the sake of clarity of exposition, considering a two-source case, the non-linear system $F(.)$ for which we want to find the roots can be written as:

$$F[\boldsymbol{y}(m)] = \begin{bmatrix} y_1(m) - L_{12}y_2(m) - Q_1y_1(m)y_2(m) - x_1(n) \\ y_2(m) - L_{21}y_1(m) - Q_2y_1(m)y_2(m) - x_2(n)\end{bmatrix} \tag{8}$$

Applying (8) to the "first-order" method described in (7), and considering that $\aleph$ is the identity matrix, we obtain:

$$\begin{aligned} y_1(m+1) = x_1(n) + L_{12}y_2(m) + Q_1y_1(m)y_2(m) \\ y_2(m+1) = x_2(n) + L_{21}y_1(m) + Q_2y_1(m)y_2(m)\end{aligned} \tag{9}$$

which is essentially equivalent to the network structure proposed by Hosseini and Deville in (5). As a consequence, this network can be understood as operating as a "first-order" iterative method to locally invert the system of nonlinear algebraic equations engendered by the mixing process. This appears to us as a valid theoretical perspective *per se*, but, in the next section, we will show that it can be useful to building new separating schemes in view of the possibility of resorting to the repertoire of available numerical methods.

III.3 – Proposal of a Recurrent Network for LQ Mixtures Based on the Use of a Variable Step Size

The use of recurrent networks for inverting LQ models is, to a certain extent, limited by the menace that the underlying dynamical process does not reach the desired equilibrium point. This problem, which is related to the stability of the employed network, was discussed and carefully addressed in [8]. In the following, we will show that the use of certain equation-solving numerical methods can lead to interesting results insofar as convergence is concerned.

Let us return, once more, to the "first-order" method shown in (7), assuming, this time, a scalar step-size α:

$$\boldsymbol{y}(m+1) = \boldsymbol{y}(m) - \alpha\, \boldsymbol{F}[\boldsymbol{y}(m)] \tag{10}$$

where α is a positive value lower than or equal to 1. The reason why we decided to use the scalar step-size α is that by decreasing the variation performed on each iteration, we can increase the method's stability.

Using the two-source case again for the sake of clarity, we have:

$$\begin{aligned} y_1(m+1) &= (1-\alpha)y_1(m) \\ &\quad + \alpha\big(x_1(n) + L_{12}y_2(m) + Q_1y_1(m)y_2(m)\big) \\ y_2(m+1) &= (1-\alpha)y_2(m) \\ &\quad + \alpha\big(x_2(n) + L_{21}y_1(m) + Q_2y_1(m)y_2(m)\big)\end{aligned} \tag{11}$$

Notice that the above expression is basically a weighted average between the current value of y_1 and y_2, and the value that would be admitted as the next on the Hosseini-Deville recurrent network given by (9). If $\alpha = 1$, then (11) is equivalent to (9). However, for $\alpha \neq 1$, the obtained network resembles the extended network introduced in [8].

To choose effective values for α on an iteration-by-iteration basis, we can use a simple one-dimensional search for a solution capable of generating an update leading to a decrease in $\|F(\boldsymbol{y}(m+1))\|^2$. This approach, which is an adaptation of well-known ideas related to the subject of line search, can be based on elaborate methods, e.g. the Fibonacci and Golden Section algorithms [7].

For our implementation, however, we used a simple method of iteratively decreasing the value of α (by multiplying it by a constant factor lower than 1) and recalculating the cost function $\|F(\boldsymbol{y}(m+1))\|^2$ until it stops decreasing, since at this point, we can assume α to be the desired minimizing step. This produces the same results as the elaborate method mentioned, but slower. Since the method is already fast enough for the two-source cases, the elaborate versions of the algorithm were not required.

IV – SIMULATION RESULTS

IV.1 – Performance for different scenarios

The proposed network, which includes a method for iteratively choosing the step size α, has interesting "stabilizing properties" that, in some cases, are responsible for performance improvements with respect to the canonical Hosseini-Deville networks. This will be illustrated with the aid of some concrete examples.

In the first simulation, the sources are uniformly distributed over the range [-1,1], and the mixing parameters are $L_{12} = Q_1 = -0.8$ and $L_{21} = Q_2 = -0.9$. These values for the source amplitude and coefficients are known to cause the original Hosseini-Deville network [5] to be potentially unstable, as can be seen in the upper two panels of Fig. 1. In contrast, it can be seen, in the lower panel, that the network based on the variable step size strategy was able to circumvent the stability problems, thus providing good estimates of the original sources.

In the simulation, both recurrent networks had an upper limit of 4000 iterations. We have obtained mean-squared error (MSE) values for the network with variable step size of 2.58×10^{-4} and 1.21×10^{-5} for sources 1 and 2 respectively, which are very small compared to the source amplitudes. For the original network, however, the MSE could not be properly estimated due to the existence of several points that diverged to infinity.

If we consider only the points that converged, however, we obtain MSEs of 2.67×10^{-2} and 2.65×10^{-2}, which are much higher than those obtained for the network with the variable step size. We have also verified that only 67.4% of the points converged. It should also be noticed that, here and in all cases treated in this work, both networks are initialized with the null vector, which ensures a fair convergence comparison.

In the second simulation, we have used sources uniformly distributed in the [-2,2] interval, and mixing coefficients with values $L_{12} = 0.5$, $L_{21} = 0.8$, $Q_1 = -0.5$ and $Q_2 = 0.8$. None of the structures was able to have a satisfactory convergence for all points, but, as can be seen in Fig. 2, the proposed network had a better overall performance, as confirmed by the MSE values: 0.24 and 0.64 for the Hosseini-Deville network (for sources 1 and 2 respectively), and 0.11 and 0.50 for the proposed network. Moreover, the proposed network converged for 87% of the points, whereas the Hosseini-Deville network converged for 68.1% of the points.

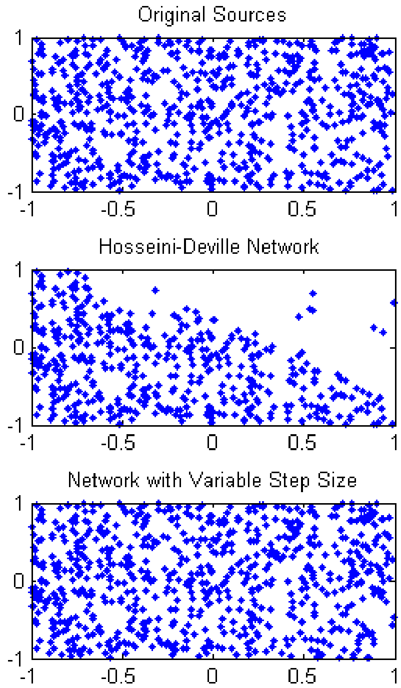

Figure 1 – Distribution of the original sources and of the source estimates generated using the canonical Hosseini-Deville network and the variable step size approach described in Section III.3.

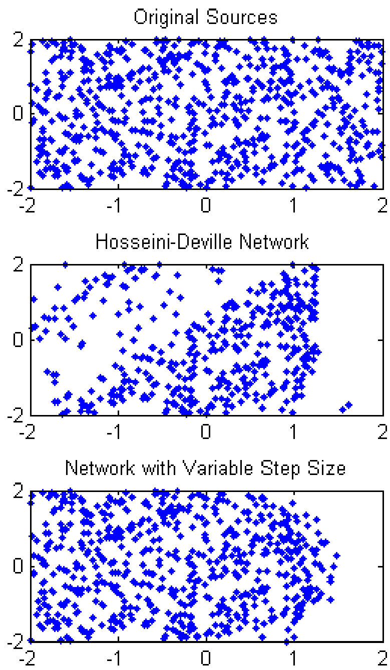

Figure 2 - Distribution of the original sources and of the source estimates generated using the canonical Hosseini-Deville network and the variable step size approach described in Section III.3.

IV.2 – Elements of a More Detailed Dynamical Analysis

In order to analyze in more detail the stability of the proposed network, we will consider a setup for which the canonical Hosseini-Deville network does not reach a fixed point [8]. The following parameters are used: $L_{12} = L_{21} = 0$, $Q_1 = Q_2 = 0.5$. A single time instant, in which the source samples are $s_1 = -1$, $s_2 = -2$, is considered.

In the upper panel of Fig. 3, we present the bifurcation diagram obtained from one of the state variables of the proposed network by varying the step size α. Qualitatively, one notices that, for values of α smaller than a threshold approximately equal to 0.8, there is convergence to the desired fixed point (-1, the value of the sample associated with the source s_1). Interestingly, when α exceeds this limit, there is a cascade of period-doubling bifurcations that eventually lead to chaos (notice the "dense regions" of the diagram), which characterizes a typical Feigenbaum scenario [6]. This analysis is rigorously confirmed by the lower panel of Fig. 3, which brings an estimate of the value of the largest Lyapunov exponent associated with the network [6]. The existence of positive values for this exponent confirms the existence of chaotic behavior, whereas the negative values correspond to convergence to fixed points or limit cycles. Finally, the null values indicate the values for which there occur bifurcations. It must be said that, after the region of chaotic behavior, we have a conventional pattern of divergence towards infinity.

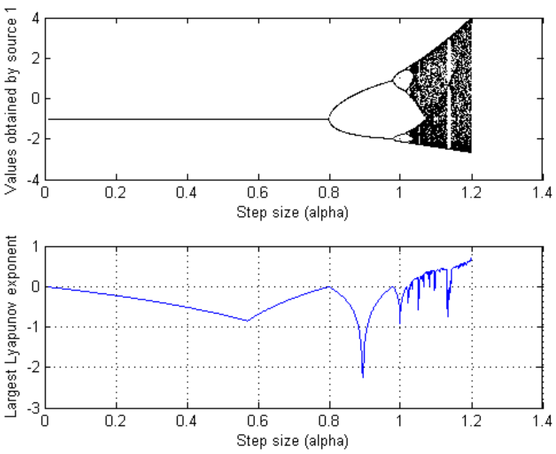

Figure 3 – Bifurcation Diagram and Largest Lyapunov Exponent

This analysis, albeit preliminary, clearly reveals the complexity of possible behaviors arising from recurrent networks following the Hosseini-Deville structure and indicates the importance of using control strategies like the variable step size proposed in this work, or alternatively, the self-feedback existing in the extended version of Hosseini-Deville recurrent network [8].

V – Conclusion

In this work, we presented an interpretation of recurrent methods for source separation in terms of local inversion based on iterative methods for solving nonlinear systems of algebraic equations. Firstly, it was shown that both the Hérault-Jutten and the basic version of Deville-Hosseini approaches are, in their respective application domains, equivalent to "first-order" numerical strategies.

In the following, a new network, based on a line search strategy devised to seek step sizes capable of minimizing the error, was proposed. Based on simulations performed for two different scenarios, it was shown that the new network was capable of having a stable behavior for a wider range of values in comparison with the canonical network structure used for linear-quadratic models. Finally, a preliminary dynamical analysis of the proposed network revealed the existence of limit cycles and chaos for the basic Deville-Hosseini architecture and gave further support to the conclusion that stabilizing mechanisms are highly desirable whenever iterative methods like those discussed here are employed.

As perspectives for future work, we indicate the extension of the idea presented here to "second-order" methods, like the Newton-Raphson method, as well as the analysis of other mixing system orders and models. We also intend to compare these networks with the extended Deville-Hosseini network. A detailed study concerning the blind case and an extension of the presented dynamical analysis are also in our plans.

Acknowledgements

The authors thank FAPESP and CNPq for the financial support.

References

1. Comon, P., Jutten, C., eds.: Handbook of blind source separation, independent component analysis and applications. Academic Press (Elsevier) (2010)

2. Romano, J.M.T., Attux, R.R.F., Cavalcante, C.C., Suyama, R.: Unsupervised signal processing: channel equalization and source separation. CRC Press (2011)

3. J. Hérault, C. Jutten, and B. Ans, Détection de grandeurs primitives dans un message composite par une architecture de calcul neuromimétique en apprentissage non supervisé," in Proceedings of the GRETSI, 1985, pp. 1017-1022.

4. Comon, P.: Independent component analysis, a new concept? Signal Processing 36 (1994) 287–314

5. Hosseini, S., Deville, Y.: Blind separation of linear-quadratic mixtures of real sources using a recurrent structure. In: Proc of 6th International Work-conference in Artificial And Natural Neural Networks (IWANN). (2003) 289–296

6. Fiedler-Ferrara, N., Prado, C.P.C.: Caos Uma Introdução. Edgard Blücher (1994)

7. Luenberger, D., G.: Linear and Nonlinear Programming. Addison-Wesley (1984).

8. Deville, Y., Hosseini, S.: Recurrent networks for separating extractable-target nonlinear mixtures. part i: Non-blind configurations. Signal Processing 89 (2009) 378–393

9. Merrikh-Bayat, F., Babaie-Zadeh, M., Jutten, C.: Linear-quadratic blind source separating structure for removing show-through in scanned documents. International Journal on Document Analysis and Recognition (2010) 1–15

10. Bedoya, G.: Nonlinear blind signal separation for chemical solid-state sensor arrays. PhD thesis, Universitat Politecnica de Catalunya (2006)

11. Meganem, I., Deville, Y., Hosseini, S., Déliot, P., Briottet, X., Duarte, L.T.: Linear-quadratic and polynomial non-negative matrix factorization ; application to spectral unmixing. In: Proc. of the 19th European Signal Processing Conference (EUSIPCO 2011). (2011)