Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE @ DIRECT®

Signal Processing 84 (2004) 217–229

SIGNAL PROCESSING

www.elsevier.com/locate/sigpro

Three easy ways for separating nonlinear mixtures?

Christian Jutten[a,*], Massoud Babaie-Zadeh[a,b], Shahram Hosseini[a,c]

[a] *Laboratoire des Images et des Signaux (CNRS UMR No. 5083), INPG-46, Avenue Félix Viallet, 38031 Grenoble Cedex, France*
[b] *Electrical Engineering Department, Sharif University of Technology, Tehran, Iran*
[c] *Laboratoire d'Acoustique, Métrologie, Instrumentation, Université Paul Sabatier, Toulouse, France*

Received 21 February 2003; received in revised form 10 June 2003

Abstract

In this paper, we consider the nonlinear Blind Source Separation BSS and independent component analysis (ICA) problems, and especially uniqueness issues, presenting some new results. A fundamental difficulty in the nonlinear BSS problem and even more so in the nonlinear ICA problem is that they are nonunique without a suitable regularization. In this paper, we mainly discuss three different ways for regularizing the solutions, that have been recently explored.

Keywords: Source separation; Independence; Nonlinear; Mutual information

1. The nonlinear ICA and BSS problems

1.1. Model and problem

Consider N samples of the m-dimension observed random vector $\mathbf{x}$, modeled by

$$\mathbf{x} = \mathscr{F}(\mathbf{s}) + \mathbf{n}, \tag{1}$$

where $\mathscr{F}$ is an unknown mixing mapping assumed invertible, $\mathbf{s}$ is an unknown n-dimensional source vector containing the source signals $s_1, s_2, \ldots, s_n$, which are assumed to be statistically independent, and $\mathbf{n}$ is an additive noise, independent of the sources.

Such a model is usual in multidimensional signal processing, where each sensor receives an unknown superimposition of unknown source signals at time instants $t = 1, \ldots, N$. Then, the goal is to recover the n unknown actual source signals $s_j(t)$ which have given rise to the observed mixtures. This is referred to as the blind source separation (BSS) problem, *blind* since no or very little prior information about the sources is required. Since the only assumption is the independence of sources, the basic idea in blind source separation consists in estimating a mapping $\mathscr{G}$, only from the observed data $\mathbf{x}$, such that $\mathbf{y} = \mathscr{G}(\mathbf{x})$ are statistically independent. The method, based on statistical independence, constitutes a generic approach called independent component analysis (ICA). The key question addressed in this paper is the following: under which conditions is the vector $\mathbf{y}$ provided by ICA equal to the unknown sources $\mathbf{s}$?

In the following, we assume that there are as many mixtures as sources ($m = n$) and that noise is zero.

* Corresponding author. Tel.: +33-4-76-57-43-51; fax: +33-4-76-57-47-90.
E-mail addresses: christian.jutten@inpg.fr (C. Jutten), mbzadeh@yahoo.com (M. Babaie-Zadeh), hosseini@cict.fr (S. Hosseini).

0165-1684/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/j.sigpro.2003.10.011

1.2. Nonlinear mixtures

The general nonlinear ICA problem then consists in estimating a mapping $\mathscr{G} : (\mathbb{R})^n \to (\mathbb{R})^n$ that yields components

$$\mathbf{y} = \mathscr{G}(\mathbf{x}) \tag{2}$$

which are statistically independent, only using the observations $\mathbf{x}$.

A fundamental characteristic of the nonlinear ICA problem is that, in the general case, solutions always exist and they are highly nonunique. In other words, ICA and BSS are not equivalent: one can easily design a nonlinear mapping which mixes the sources and provides statistically independent variables y_i

$$y_i = h_i(\mathbf{s}) \neq h_i(s_{\sigma(i)}), \tag{3}$$

where $\sigma(i)$ is a permutation over $\{1, 2, \ldots, n\}$ (see Section 2).

Moreover, the separation is achieved, if each estimated output y_i only depends on a unique source $y_i = h_i(s_{\sigma(i)})$. Then, strong distortions can still occur, due to the mapping $h_i(.)$. One reason for this is that if u and v are two independent random variables, any of their functions $f(u)$ and $g(v)$ (where f and g are invertible functions) are independent too.

In the nonlinear BSS problem, one would like to find the original source signals $\mathbf{s}$ that have generated the observed data $\mathbf{x}$ with weaker indeterminacies, e.g. like the usual scale and permutation indeterminacies in linear mixtures. This would be a clearly more meaningful and unique problem than the nonlinear ICA problem defined above. But is it possible? In other words, can prior information on the sources and/or the mixing mapping be sufficient for this?

In other words, if some arbitrary independent components are found for the data generated by (1), they may be quite different from the true source signals. Generally, using ICA for solving the nonlinear BSS problem requires additional prior informations or suitable regularizing constraints.

1.3. Regularization examples

1.3.1. Linear mixtures

In linear mixtures, the unknown invertible mapping $\mathscr{F}$ is modeled by a square regular matrix $\mathbf{A}$

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \tag{4}$$

and the separating structure $\mathscr{G}$ is parameterized by a matrix $\mathbf{B}$. The global mapping is then a matrix $\mathbf{BA}$: it constitutes a very efficient regularization constraint which insures equivalence between ICA and BSS [15]. The basic linear case is now understood quite well [10,13,22]. Since 1985 (see [26] for a historical review and early references), several well-performing BSS and ICA algorithms [4,8,11,12,14,23,25] have been developed and applied to an increasing number of applications in biomedical [30,47], instrumentation [17,33], acoustics [32,44], sparse coding and feature extraction [21], analysis of color [46] or hyperspectral images [34], etc. Basically, these algorithms are based on the optimization of an independence criterion, and require statistics of order higher than 2. Many more references on linear ICA and BSS can be found in the recent books [13,22].

1.3.2. Using extra information

If the source signals are random variables having a temporal structure, linear blind source separation can be achieved by using either temporal correlations [9,43] or nonstationarity [31,35]. With these weak (since many signals satisfy them) assumptions, the independence criterion involves more equations, and it acts like a regularization. It also allows to simplify it: it is well known that, under the above assumptions, second order statistics are sufficient for insuring separation.

1.4. Outline

This paper is organized as follows. In Section 2, we explain with more details and examples why ICA and BSS are different in nonlinear mixtures. Then, we consider three ways for regularizing the problem of source separation in nonlinear mixtures. In Section 3, we show that smoothing mappings is not a sufficient constraint. In Section 4, we study constrained structures of nonlinear mixtures, which are separable, i.e. for which BSS and ICA are equivalent. In Section 5, we show that priors on sources can also regularize the solutions. Concluding remarks, in Section 6, summarize the key points of the paper.

2. Existence and uniqueness of nonlinear ICA and BSS

Several authors [18,24,26,39,40] have recently addressed the important issues on the existence and uniqueness of solutions for the nonlinear ICA and BSS problems. Their main results, which are direct consequences of Darmois's results on factorial analysis [16], are reported in this section.

2.1. Indeterminacies

Recall first the definition of a random vector with independent components.

Definition 2.1.1. A random vector $\mathbf{x}$ has statistically independent components if its joint probability density function (pdf) $p_{\mathbf{x}}(\mathbf{u})$ satisfies $p_{\mathbf{x}}(\mathbf{u})=\prod_i p_{x_i}(u_i)$, where $p_{x_i}(u_i)$ are the marginal pdfs of the random variables x_i.

Clearly, the product of a permutation matrix $\mathbf{P}$ by any diagonal mapping both preserves independence and insures separability.

Definition 2.1.2. A one-to-one mapping $\mathscr{H}$ is called *trivial*, if it transforms *any* independent random vector $\mathbf{s}$ into an independent random vector.

The set of trivial transformations will be denoted by $\mathfrak{T}$. Trivial mappings preserve the independence property of *any* random vector, i.e. for any distributions. One can easily show the following result [39].

Theorem 2.1.3. *A one-to-one mapping $\mathscr{H}$ is trivial if and only if it satisfies*

$$\mathscr{H}_i(u_1,u_2,\ldots,u_n)=h_i(u_{\sigma(i)}), \quad i=1,2,\ldots,n, \tag{5}$$

where the $h_i(.)$ are arbitrary scalar functions and σ is any permutation over $\{1,2,\ldots,n\}$.

From this result, we can deduce immediately the corollary.

Corollary 2.1.4. *A one-to-one mapping $\mathscr{H}$ is trivial if and only if its Jacobian matrix is diagonal up to a permutation.*

This result establishes a link between the independence assumption and the objective of source separation. It becomes soon clear that the objective of source separation is to make the global mapping $\mathscr{H}=\mathscr{G}\circ\mathscr{F}$ trivial using the independence assumption.

However, from (5), it is clear that sources can only be separated up to a permutation and a nonlinear function. For any invertible mapping $\mathscr{F}(\mathbf{x}) = [f_1(\mathbf{x}),\ldots,f_n(\mathbf{x})]^T$ such that each of its components is a scalar nonlinear mapping $f_i(\mathbf{x}) = f_i(x_i)$, $i = 1,\ldots,n$, it is evident that if $p_{\mathbf{x}}(\mathbf{u})=\prod_i p_{x_i}(u_i)$, then $p_{\mathscr{F}(\mathbf{x})}(\mathbf{v})=\prod_i p_{f_i(x_i)}(v_i)$.

Unfortunately, as we shall see in the next subsection, independence can be preserved with non trivial mappings.

2.2. Results from factor analysis

Considering the factorial representation of the random vector $\mathbf{x}$ in a random vector $\boldsymbol{\zeta}$ with independent components ζ_i:

$$\mathbf{x}=\mathscr{F}_1(\boldsymbol{\zeta}), \tag{6}$$

the uniqueness study of the representation consists in addressing the following question. Is there another factorial representation of $\mathbf{x}$ in random vector $\boldsymbol{\omega}$ with independent components ω_i such that

$$\mathbf{x}=\mathscr{F}_1(\boldsymbol{\zeta})=\mathscr{F}_2(\boldsymbol{\omega}), \tag{7}$$

where $\boldsymbol{\zeta}$ and $\boldsymbol{\omega}$ are different random vectors with independent components? If yes, uniqueness is wrong.

In the general case, when the mapping $\mathscr{H}$ has no particular form, a well-known statistical result shows that preserving independence is not a strong enough constraint for ensuring separability in the sense of Eq. (5). This result, based on a simple constructive method (detailed below) similar to a Gram–Schmidt orthogonalization procedure, has been established already early in the 1950s by Darmois [16]. It has also been used in [24] for designing parameterized families of nonlinear ICA solutions.

Let $\mathbf{x}$ be any random vector, and $\mathbf{y} = \mathscr{G}(\mathbf{x})$ the independent random vector provided by the invertible mapping $\mathscr{G}$. Since $\mathbf{y}$ is independent, one can write:

$$p_{\mathbf{Y}}(\mathbf{y}) = p_{\mathbf{X}}(\mathbf{x})/|J_{\mathscr{G}}(\mathbf{x})|. \tag{8}$$

Without loss of generality, one can assume that y_i, $i=1,\ldots,n$, is a uniform random variable. Then, Eq. (8) reduces to the condition

$$p_{\mathbf{X}}(\mathbf{x}) = |J_{\mathscr{G}}(\mathbf{x})|. \tag{9}$$

Looking for solutions of the form:

$$g_1(\mathbf{x}) = g_1(x_1),$$

$$g_2(\mathbf{x}) = g_2(x_1, x_2),$$

$$\vdots$$

$$g_n(\mathbf{x}) = g_n(x_1, x_2, \ldots, x_n), \tag{10}$$

Eq. (8) becomes

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} \frac{\partial g_i(\mathbf{x})}{\partial x_i} \tag{11}$$

or, using the Bayes theorem

$$p_{X_1}(x_1) p_{X_2/X_1}(x_1, x_2) \ldots p_{X_n/X_1,\ldots,X_{n-1}}(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} \frac{\partial g_i(\mathbf{x})}{\partial x_i}. \tag{12}$$

By simple integration of (12), one obtains the following solution

$$g_1(x_1) = F_{X_1}(x_1),$$

$$g_2(x_1, x_2) = F_{X_2/X_1}(x_1, x_2),$$

$$\vdots$$

$$g_n(x_1, x_2, \ldots, x_n) = F_{X_n/X_1,\ldots,X_{n-1}}(x_1, x_2, \ldots, x_n), \tag{13}$$

where F_{X_1} denotes the marginal cumulative density function of the random variable X_1, and $F_{X_{k+1}/X_1,\ldots,X_k}$ denotes the conditional cumulative density function of the random variable X_{k+1}, given $X_1, \ldots, X_k$. Mapping (13) is then a nontrivial mapping, (since its Jacobian is not diagonal) which maps the random vector $\mathbf{x}$ to an independent (uniform) random vector $\mathbf{y}$.

Darmois's result is negative in the sense that it shows that, for any random vector $\mathbf{x}$, there exists at least one nontrivial transformation $\mathscr{H}$ [1] which "*mixes*" the variables while still preserving their statistical independence. Hence blind source separation is simply *impossible* for general nonlinear transformations by resorting to statistical independence only without constraints on the transformation model.

We can then conclude like Darmois in [16]: "*These properties* [...] *clarify the general problem of factor analysis by showing the great indeterminacies it presents as soon as one leaves the already very wide field of linear diagrams.*"

[1] $\mathscr{H}$ depends on $\mathbf{x}$, and generally it is not a mapping preserving independence for another random vector $\mathbf{u} \neq \mathbf{x}$.

2.3. A simple example

We give a simple example of mixing mappings preserving independence, derived from [40]. Suppose $s_1 \in \mathbb{R}^+$ is a Rayleigh distributed variable with pdf $p_{s_1}(s_1) = s_1 \exp(-s_1^2/2)$, and s_2 is independent of s_1, with a uniform pdf $s_2 \in [0, 2\pi)$. Consider the nonlinear mapping

$$[y_1, y_2] = \mathscr{H}(s_1, s_2) = [s_1 \cos(s_2), s_1 \sin(s_2)] \tag{14}$$

which has a nondiagonal Jacobian matrix

$$\mathbf{J} = \begin{pmatrix} \cos(s_2) & -s_1 \sin(s_2) \\ \sin(s_2) & s_1 \cos(s_2) \end{pmatrix}. \tag{15}$$

The joint pdf of y_1 and y_2 is

$$p_{y_1,y_2}(y_1, y_2) = \frac{p_{s_1,s_2}(s_1, s_2)}{|\mathbf{J}|} = \frac{1}{2\pi} \exp\left(\frac{-y_1^2 - y_2^2}{2}\right) = \left(\frac{1}{\sqrt{2\pi}} \exp\frac{-y_1^2}{2}\right)\left(\frac{1}{\sqrt{2\pi}} \exp\frac{-y_2^2}{2}\right)$$

This shows that the random variables y_1 and y_2 are independent, although they are still nonlinear mixtures of the sources. $\mathscr{H}$ preserves the independence but only for the random variables s_1 and s_2 (Rayleigh and uniform).

Other examples can be found in the literature (see for example [29]), or can be easily constructed.

2.4. Conclusion

In nonlinear mixtures, ICA does not insure separation, and emphasizes very large indeterminacies so that the nonlinear BSS problem is ill-posed. A natural idea, for reducing the indeterminacies, is to add regularization. This can be done according to a few ways, e.g. constraining the transformation $\mathscr{H}$ in a certain set of transformations $\mathfrak{Q}$ or using priors on sources.

3. Smooth mappings

Recently, multi-layer perceptron (MLP) networks (see [19]) have been used in [2,48] for estimating the generic nonlinear mappings $\mathscr{H}$. Especially, Almeida conjectured that smooth mappings providing by MLP networks leads to a regularization sufficient for ensuring that nonlinear ICA leads to nonlinear BSS, too. However, the following example [5] shows that smoothness alone is not sufficient for separation.

Without a loss of generality, consider two independent random variables $\mathbf{s} = (s_1, s_2)^{\mathrm{T}}$ which are both uniformly distributed in the interval $[-1, 1]$, and the nonlinear smooth mapping represented by the matrix

$$\mathbf{R} = \begin{pmatrix} \cos(\theta(r)) & -\sin(\theta(r)) \\ \sin(\theta(r)) & \cos(\theta(r)) \end{pmatrix}, \tag{16}$$

where $r \triangleq \sqrt{s_1^2 + s_2^2}$. This is a rotation for which the rotation angle $\theta(r)$ depends on the radius r as follows:

$$\theta(r) = \begin{cases} \theta_0(1-r)^q, & 0 \leqslant r \leqslant 1, \\ 0, & r > 1, \end{cases} \tag{17}$$

where $q \geqslant 2$. Fig. 1 shows the transformation of the region $\{-1 \leqslant s_1 \leqslant 1, -1 \leqslant s_2 \leqslant 1\}$ under this mapping for $q = 2$ and $\theta_0 = \pi/2$.

It can be seen [5] that the Jacobian matrix of this smooth mapping is

$$\mathbf{J_R} = \begin{pmatrix} \cos(\theta(r)) & -\sin(\theta(r)) \\ \sin(\theta(r)) & \cos(\theta(r)) \end{pmatrix} \times \begin{pmatrix} 1 - s_2 \dfrac{\partial\theta}{\partial s_1} & -s_2 \dfrac{\partial\theta}{\partial s_2} \\ s_1 \dfrac{\partial\theta}{\partial s_1} & 1 + s_1 \dfrac{\partial\theta}{\partial s_2} \end{pmatrix}. \tag{18}$$

Computing the determinant

$$\det(\mathbf{J_R}) = 1 + s_1 \frac{\partial\theta}{\partial s_2} - s_2 \frac{\partial\theta}{\partial s_1} \tag{19}$$

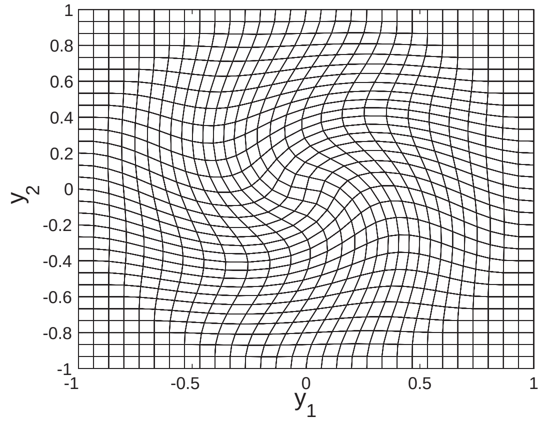

Fig. 1. The smooth mapping **R** is a mixing mapping which preserves independence of any uniform random vector. The curves are the transforms of lines $s_1 = \mathrm{cst}$ and $s_2 = \mathrm{cst}$ by the mapping **R** points out these properties.

and since

$$s_1 \frac{\partial\theta}{\partial s_2} = s_2 \frac{\partial\theta}{\partial s_1} = \frac{s_1 s_2}{r} \theta'(r) \tag{20}$$

one finally gets $\det(\mathbf{J_R} = 1)$, and hence:

$$p_{y_1 y_2}(y_1, y_2) = p_{s_1 s_2}(s_1, s_2). \tag{21}$$

From (18) the Jacobian matrix of this smooth mapping is not diagonal (the mapping is then mixing). However, from (21) the mapping preserves the independence of the two uniform random variables on $[-1, 1]$. This counterexample proves that restricting the mapping to be smooth is not sufficient.

In fact, it mainly means that smoothness is a too vague property, and one has to explore further for defining sufficient conditions, and discovering a (third) way for separating nonlinear mixtures. Hyvärinen and Pajunen gave a partial answer to this question in [24], proving that a unique solution (up to a rotation) can be obtained in the two-dimensional special case if the mixing mapping $\mathscr{F}$ is constrained to be a conformal mapping. In fact, such a mapping is characterized by much stronger constraints that simple smoothness, since it is analytic and nonzero: a well-known property is that angles remain unaltered by conformal mappings.

4. Structural constraints

A natural way of regularizing the solution consists in looking for separating mappings belonging to a specific subspace $\mathfrak{Q}$. To characterize the indeterminacies for this specific model $\mathfrak{Q}$, one must solve the tricky independence preservation equation which can be written

$$\forall E \in \mathfrak{M}_n,$$
$$\int_E \mathrm{d}F_{s_1}\,\mathrm{d}F_{s_2}\cdots\mathrm{d}F_{s_n} = \int_{\mathscr{H}(E)} \mathrm{d}F_{y_1}\,\mathrm{d}F_{y_2}\cdots\mathrm{d}F_{y_n}, \tag{22}$$

where $\mathfrak{M}_n$ is the set of all the measurable compacts in $\mathbb{R}^n$ (in other words, $\mathfrak{M}_n$ is a σ-algebra on $\mathbb{R}^n$), and F_{s_i} denotes the distribution function of the random variable s_i .

Let $\mathfrak{P}$ denote the set [2]

$$\mathfrak{P} = \{(F_{s_1}, F_{s_2}, \ldots, F_{s_n}), / \exists \mathscr{H} \in \mathfrak{Q} \setminus (\mathfrak{T} \cap \mathfrak{Q}) : \mathscr{H}(s) \text{ has independent components}\} \tag{23}$$

of all source distributions for which there exists a non-trivial (i.e. not belonging to the set of trivial mappings $\mathfrak{T}$) mapping $\mathscr{H}$ belonging to the model $\mathfrak{Q}$ and preserving the independence of the components of the source vector $\mathbf{s}$.

Ideally, $\mathfrak{P}$ should be empty and $\mathfrak{T} \cap \mathfrak{Q}$ should contain the identity as a unique element. However, in general this is not fulfilled. We then say that source separation is possible when the distributions of the sources belong to the set $\bar{\mathfrak{P}}$, which is the complement of $\mathfrak{P}$. The sources are then restored up to a trivial transformation belonging to the set $\mathfrak{T} \cap \mathfrak{Q}$. Solving (22), i.e. determining $\mathfrak{P}$, is generally a very difficult problem, except for simple models $\mathfrak{Q}$, like linear invertible mappings.

4.1. Example: linear models

In the case of regular linear models, the transformation $\mathscr{F}$ is linear and can be represented by (4), where $\mathbf{A}$ is a square invertible matrix. In this case it suffices to constrain the separating model $\mathscr{G}$ to lie in the subspace of invertible square matrices, and one has to estimate a matrix $\boldsymbol{B}$ such that $\boldsymbol{y} = \boldsymbol{Bx} = \boldsymbol{Hs}$ has independent components. The global transform $\mathscr{H}$ is then restricted to the subspace $\mathfrak{Q}$ of invertible square matrices.

The set of *linear* trivial transformations $\mathfrak{T} \cap \mathfrak{Q}$ is the set of matrices equal to the product of a permutation and a diagonal matrices. From the Darmois–Skitovich theorem [16], it is clear that the set $\mathfrak{P}$ contains the distributions having at least two Gaussian components. Thus we end up with Comon's well-known theorem [15]: blind source separation is possible whenever we have at most one Gaussian source, and the sources can then be restored up to a permutation and a diagonal matrix.

[2] In Eq. (23), $\setminus$ denotes the difference between two sets.

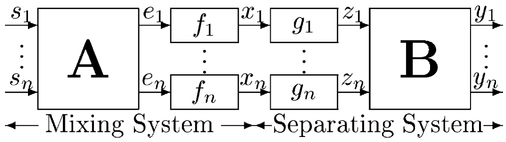

Fig. 2. The mixing-separating system for PNL mixtures.

4.2. Separability of PNL mixtures

In the post-nonlinear (PNL) model, the nonlinear observations have the following specific form (Fig. 2):

$$x_i(t) = f_i\left(\sum_{j=1}^{n} a_{ij} s_j(t)\right), \quad i = 1, \ldots, n. \tag{24}$$

One can see that the PNL model consists of a linear mixture followed by a componentwise nonlinearity f_i acting on each output independently from the others. The nonlinear functions (distortions) f_i are assumed to be invertible.

Besides its theoretical interest, this model belonging to the L-ZMNL [3] family suits perfectly for a lot of real-world applications. For instance, such models appear in sensors array processing [33], satellite and microwave communications [37], and in many biological systems [28].

[3] L stands for Linear and ZMNL stands for Zero-Memory Non-Linearity: it is a separable model with a linear stage followed by a nonlinear (static) distortion.

As discussed before, the most important thing when dealing with nonlinear mixtures is the separability issue. First, the separation structure $\mathscr{G}$ must be constrained so that:

1. It can invert the mixing system in the sense of Eq. (5).
2. It should be as simple as possible for reducing the residual distortions h_i, which result from using the independence assumption only.

Under these two constraints, we have no other choice that selecting for the separating system $\mathscr{G}$ the mirror structure of the mixing system $\mathscr{F}$ (see Fig. 2). The global transform $\mathscr{H}$ is then restricted to the subspace $\mathfrak{Q}$ of transforms, which consists of a cascade of an invertible linear mixture (regular matrix $\mathbf{A}$) followed by componentwise invertible distortions and again an invertible linear mixture (regular matrix $\mathbf{B}$). In [40], it has been shown that these mixtures are separable for distributions having at most one Gaussian source (the set $\mathfrak{P}$ contains the distributions having at least two Gaussian components), with the same indeterminacies as linear mixtures (the set of *linear* trivial transformations $\mathfrak{T} \cap \mathfrak{Q}$ is the set of matrices equal to the product of a permutation and a diagonal matrices) if $\mathbf{A}$ has at least 2 nonzero entries on each row or column.

Separability of PNL mixtures can be generalized to convolutive PNL (CPNL) mixtures, in which the instantaneous mixture (matrix $\mathbf{A}$) is replaced by linear filters (matrix of filters $\mathbf{A}(z)$), where each source is independent and identically distributed (iid) [6]. In fact, denoting $\mathbf{A}(z) = \sum_k \mathbf{A}_k z^{-k}$, and defining:

$$\mathbf{s} \triangleq (\ldots, \mathbf{s}^{\mathrm{T}}(k-1), \mathbf{s}^{\mathrm{T}}(k), \mathbf{s}^{\mathrm{T}}(k+1), \ldots)^{\mathrm{T}}, \qquad (25)$$

$$\mathbf{x} \triangleq (\ldots, \mathbf{x}^{\mathrm{T}}(k-1), \mathbf{x}^{\mathrm{T}}(k), \mathbf{x}^{\mathrm{T}}(k+1), \ldots)^{\mathrm{T}} \qquad (26)$$

we have

$$\mathbf{x} = \mathbf{f}(\bar{\mathbf{A}}\mathbf{s}), \qquad (27)$$

where $\mathbf{f}$ acts componentwise, and

$$\bar{\mathbf{A}} = \begin{bmatrix} \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & \mathbf{A}_{k+1} & \mathbf{A}_k & \mathbf{A}_{k-1} & \cdots \\ \cdots & \mathbf{A}_{k+2} & \mathbf{A}_{k+1} & \mathbf{A}_k & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots \end{bmatrix}. \qquad (28)$$

The iid nature of the source samples, i.e. the temporal independence of $s_i(k)$, $i = 1, \ldots, n$, insures the spatial independence of $\mathbf{s}$. Then, the CPNL mixtures can be viewed as a particular PNL mixtures. For FIR mixing matrix $\mathbf{A}(z)$, (27) corresponds to a finite dimension PNL mixture and the separability holds. For more general filter (IIR) matrix, (27) is an infinite dimension PNL mixture, and the separability can be conjectured.

Moreover, using a suitable parameterization, Wiener systems [4] can be viewed as particular PNL mixtures. Consequently, separability of PNL mixtures ensures blind invertibility of Wiener systems [41]. Theis et al. have studied in [42] separability of a cascade of PNL stages, constituting a structure similar to multi-layer perceptron networks.

4.3. Other separable nonlinear mixtures

Due to the interesting Darmois's result for linear mixtures, it is clear that nonlinear mixtures which can be reduced to linear mixtures with a simple mapping should be separable.

4.3.1. A simple example

As an example, consider multiplicative mixtures:

$$x_j(t) = \prod_{i=1}^{n} s_i^{\alpha i}(t), \quad j = 1, \ldots, n, \qquad (29)$$

where the $s_i(t)$ are positive independent sources. Taking the logarithm yields to

$$\ln x_j(t) = \sum_{i=1}^{n} \alpha_i \ln s_i(t), \quad j = 1, \ldots, n \qquad (30)$$

which is a linear model for the new independent random variables $\ln s_i(t)$. For instance, this type of mixtures can be used for modeling the dependency between the temperature and magnetic field in Hall silicon sensor [3], or gray-level images as a product of incident light and reflected light [18]. Considering in more details the former example, the Hall voltage [36] is equal to

$$V_H = kBT^{\alpha} \qquad (31)$$

[4] Restricted to the cascade of a linear time invariant filter and a memoryless invertible nonlinear mapping.

where α depends on the semiconductor type, since the temperature effect is related to the mobility of the majority carriers. Then, using two types (N and P) of sensors, we have

$$V_{H_N}(t) = k_N B(t) T^{\alpha_N}(t),$$
$$V_{H_P}(t) = k_P B(t) T^{\alpha_P}(t). \tag{32}$$

For simplifying the equations, we now drop the variable t out. Because the temperature T is positive but the sign of the magnetic field B can vary, taking the logarithm leads then to the equations

$$\ln|V_{H_N}| = \ln k_N + \ln|B| + \alpha_N \ln T,$$
$$\ln|V_{H_P}| = \ln k_P + \ln|B| + \alpha_P \ln T. \tag{33}$$

These equations describe a linear mixture of the two sources $\ln|B|$ and $\ln T$. They can be easily solved even with a simple decorrelation approach since B appears with the same power in the two equations. It is even simpler to directly compute the ratio of the above two equations:

$$R = \frac{V_{H_N}}{V_{H_P}} = \frac{k_N}{k_P} T^{\alpha_N - \alpha_P} \tag{34}$$

which depends only on the temperature T. For separating the magnetic field, it is sufficient to estimate the parameter k so that $V_{H_N} R^k$ becomes uncorrelated with R. From this, one can deduce $B(t)$ up to a multiplicative constant. Final estimation of the values of B and T requires sign reconstruction and calibration steps.

Note that the idea of this subsection is usual in homomorphic filtering [38] of images, for separating illumination and reflectance components of the images, with a simple low-pass filtering after a log-transform of the images.

4.4. Generalization to a class of mappings

Extension of the Darmois–Skitovic theorem to nonlinear functions has been addressed by Kagan et al. in [27]. Their results have recently been revisited within the framework of BSS of nonlinear mixtures by Eriksson and Koivunen [18]. The main idea is to consider particular mappings $\mathscr{F}$ satisfying an *addition theorem* in the sense of the theory of functional equations. As a simple example of such a mapping, consider the nonlinear mixture of the two independent random variables s_1 and s_2:

$$x_1 = (s_1 + s_2)(1 + s_1 s_2)^{-1},$$
$$x_2 = (s_1 - s_2)(1 - s_1 s_2)^{-1}. \tag{35}$$

Now, using the variable transforms $u_1 = \tan^{-1}(s_1)$ and $u_2 = \tan^{-1}(s_2)$, the above nonlinear model becomes

$$x_1 = \tan(u_1 + u_2),$$
$$x_2 = \tan(u_1 - u_2). \tag{36}$$

Applying again the transformation $\tan^{-1}$ to x_1 and x_2 yields

$$v_1 = \tan^{-1}(x_1) = u_1 + u_2,$$
$$v_2 = \tan^{-1}(x_2) = u_1 - u_2 \tag{37}$$

which is now a linear mixture of the two independent variables u_1 and u_2. This nice result is due to the fact that $\tan(a + b)$ is a mapping of $\tan a$ and $\tan b$.

More generally, this property will hold provided that there exists a mapping $\mathscr{F}$ and an invertible function f satisfying an addition theorem:

$$f(s_1 + s_2) = \mathscr{F}[f(s_1), f(s_2)]. \tag{38}$$

Let $u \in \mathfrak{S}$ be in the range $[a, b]$. The basic properties required for the mapping $\mathscr{F}$ (in the case of two variables, but extension is straightforward) are the following:

- $\mathscr{F}$ is continuous at least separately for the two variables;
- $\mathscr{F}$ is commutative, i.e. $\forall (u, v) \in \mathfrak{S}^2$, $\mathscr{F}(u, v) = \mathscr{F}(v, u)$;
- $\mathscr{F}$ is associative, i.e. $\forall (u, v, w) \in \mathfrak{S}^3$, $\mathscr{F}(\mathscr{F}(u, v), w) = \mathscr{F}(u, \mathscr{F}(v, w))$;
- There exists an identity element $e \in \mathfrak{S}$ such that $\forall u \in \mathfrak{S}$, $\mathscr{F}(u, e) = \mathscr{F}(e, u) = u$;
- $\forall u \in \mathfrak{S}$, there exists an inverse element $u^{-1} \in \mathfrak{S}$ such that $\mathscr{F}(u, u^{-1}) = \mathscr{F}(u^{-1}, u) = e$.

In other words, denoting $u \circ v = \mathscr{F}(u, v)$, these conditions imply that the set $(\mathfrak{S}, \circ)$ is an Abelian group. Under this condition, Aczel [1] proved that there exists a monotonic and continuous function $f : \mathbb{R} \to [a, b]$ such that

$$f(x + y) = \mathscr{F}(f(x), f(y)) = f(x) \circ f(y). \tag{39}$$

Clearly, applying f^{-1} (which exists since f is monotonic) to the above equation leads to

$$x+y=f^{-1}(\mathscr{F}(f(x),f(y))) \\ =f^{-1}(f(x)\circ f(y)). \quad (40)$$

Using associativity and the above property (39), setting $y=x$, one can directly define a product $\star$ with integer which can be extended to real variables:

$$f(cx)=c\star f(x). \quad (41)$$

Taking the inverse f^{-1} and denoting $f(x)=u$, this yields

$$cf^{-1}(u)=f^{-1}(c\star u). \quad (42)$$

Then for any constants $c_1,\dots,c_n$ and random variables $u_1,\dots,u_n$, the following relation holds:

$$c_1f^{-1}(u_1)+\cdots+c_nf^{-1}(u_n) \\ =f^{-1}(c_1\star u_1\circ\cdots\circ c_n\star u_n). \quad (43)$$

Finally, Kagan et al. [27] stated the following theorem:

Theorem 4.4.1. *Let $u_1,\dots,u_n$ be independent random variables such that*

$$x_1=a_1\star u_1\circ\cdots\circ a_n\star u_n, \\ x_2=b_1\star u_1\circ\cdots\circ b_n\star u_n \quad (44)$$

are independent, and the operators $\star$ and $\circ$ satisfy the above conditions. Denoting by f the function defined by the operator $\circ$, $f^{-1}(u_i)$ is Gaussian if $a_ib_i\neq 0$.

This theorem can be easily extended to source separation, and with such mixtures the separation algorithm consists of 3 practical steps [18]:

1. Apply f^{-1} to the nonlinear observations for providing linear mixtures in $s_i=f^{-1}(u_i)$.
2. Solve the linear mixtures in s_i by any BSS method.
3. Restore the actual independent sources by applying $u_i=f(s_i)$.

Unfortunately, this algorithm is not blind since the function f must be known. If f is not known, a suitable separation structure is a cascade of identical nonlinear componentwise blocks (able to approximate f^{-1}) followed by a linear matrix $\mathbf{B}$ able to separate the sources in linear mixtures. This stage is further followed by identical nonlinear componentwise blocks (which approximate f) for restoring the actual sources. We remark that the two first blocks of this structure are identical to the separation structure of PNL mixtures (in fact slightly simpler, since all the nonlinear blocks are similar). We can then estimate the independent distorted sources s_i with a PNL mixture separation algorithm. After computing f from the nonlinear block estimates (which approximate f^{-1}), one can then restore the actual sources.

The PNL mixtures are close to these mappings. They are in fact more general since the nonlinear functions f_i can be different and unknown. Consequently, algorithms developed for separating sources in PNL mixtures (e.g. [40]) can be used for *blindly* separating these nonlinear mappings, avoiding the above step 1. Other examples of mappings satisfying the addition theorem are given in [18,27]. However, realistic mixtures belonging to this class seem unusual, except for the PNL mixtures (24) and the multiplicative mixtures (29).

5. Prior information on the sources

In this section we show that prior information on the sources can simplify or relax the indeterminacies. The first example takes into account that sources are bounded. The second example exploits the temporal correlation of the sources.

5.1. Bounded sources in PNL mixtures

Let us consider sources whose pdf has a bounded support, with nonzero values on the edges of the support. For example the uniform distribution or the distribution of a randomly sampled sine wave satisfy this condition. For simplicity, we discuss only PNL mixtures (Fig. 2) of two sources, but the results can be easily extended to more sources. From $p_{s_1s_2}(s_1,s_2)=p_{s_1}(s_1)p_{s_2}(s_2)$ we deduct that the joint distribution of the two sources $\mathbf{s}$ is contained in a rectangle. After the linear mixing $\mathbf{A}$, the joint distribution of $\mathbf{e}=\mathbf{As}$ lies inside a parallelogram. After the componentwise invertible nonlinear distortions f_i, the joint distribution of $\mathbf{x}$ (the PNL mixtures) is contained in a “distorted” parallelogram (see Fig. 3).

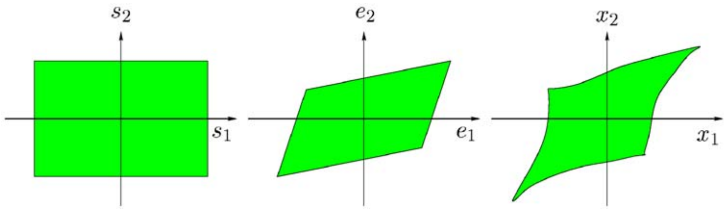

Fig. 3. Joint distributions of the signals at different locations in a PNL model.

Babaie-Zadeh et al. [7] proved the following theorem:

Theorem 5.1.1. *Consider the transformation*:

$$x_1 = h_1(e_1),$$

$$x_2 = h_2(e_2), \tag{45}$$

where h_1 and h_2 are analytic functions.[5] *If the borders of a parallelogram in the (e_1, e_2) plane are transformed to the borders of a parallelogram in the (x_1, x_2) plane, and the borders of these parallelograms are not parallel to the coordinate axes, then, there exist real constants a_1, a_2, b_1 and b_2 such that*

$$h_1(u) = a_1 u + b_1,$$

$$h_2(u) = a_2 u + b_2. \tag{46}$$

Remark 1. The requirement that the borders of the parallelograms must not be parallel to the coordinate axes emphasizes that the sources must be really mixed before the nonlinear distortion. This means that there must exist at least two nonzero elements in each row, or each column of the mixing matrix **A**, as stated in Section 4.2.

Remark 2. The existence of the constants b_1 and b_2, emphasizes on a "DC" indeterminacy (on the sources) in separating PNL mixtures. This indeterminacy exists also in linear mixtures but is generally skipped since one assumes zero-mean sources. In other words, in linear as well as PNL mixtures, one only can recover the "AC" part of the sources with the classical scale and permutation indeterminacies.

[5] A function is called analytic on an interval, if it can be expressed with a Taylor series on that interval.

Remark 3. The theorem provides another separability proof for PNL mixtures of bounded sources.

This theorem suggests a 2-step geometric approach for separating PNL mixtures:

- Find invertible functions g_1 and g_2 which transform the scatter plot of the observations to a parallelogram. From the above theorem, this step insures compensation of the nonlinear distortions.
- Separate the resulting linear mixture, by means of any linear ICA algorithm.

Details of the algorithm and experimental results are given in [7]. An important point about this method is that it proves by using simple prior information, the nonlinear distortions can be estimated without using the independence assumption. In other words, bounded sources provide useful extra information for simplifying separation algorithms in PNL mixtures: the nonlinear and linear parts can be optimized independently with two different criteria.

5.2. *Time correlated sources in nonlinear mixtures*

As we discussed in the previous sections, in general, applying the independence hypothesis for separating the nonlinear mixtures is not sufficient: it may lead to *good solutions* where the estimated independent components are the trivial mappings of the original sources, or to *bad solutions* where the estimated independent components are still the mixtures of the original sources. The interesting question is: how to distinguish the good solutions from the bad ones? Hosseini and Jutten suggest [20] that the temporal correlation between the successive samples of each source may be used for this purpose.

5.2.1. A simple example revisited

Let us return to simple example presented in Section 2.3, where we suppose now that the sources s_1, s_2, y_1 and y_2 are signals. If the signals $s_1(t)$ and $s_2(t)$ are temporally correlated and independent, one can write that

$$E[s_1(t_1)s_2(t_2)] = E[s_1(t_1)]E[s_2(t_2)], \quad \forall t_1, t_2. \quad (47)$$

Consequently, in addition to $E[y_1(t)y_2(t)] = E[y_1(t)]E[y_2(t)]$ many equations can be used and allow to reject the *bad solutions* $y_1(t)$ and $y_2(t)$:

$$E[y_1(t_1)y_2(t_2)] = E[y_1(t_1)]E[y_2(t_2)] \quad \forall t_1 \neq t_2. \quad (48)$$

It is evident that if $y_1(t)$ and $y_2(t)$ are the actual sources (or a trivial mapping of them), the above equality is true $\forall t_1, t_2$. Moreover, if the independent components are obtained from mappings (14) (this mapping is a mixing nonlinear mapping, preserving independence for random variables), the right side of (48) is equal to zero because y_1 and y_2 are zero mean Gaussian variables. The left side of (48) is equal to

$$E[s_1(t_1)\cos(s_2(t_1))s_1(t_2)\sin(s_2(t_2))]$$
$$= E[s_1(t_1)s_1(t_2)]E[\cos(s_2(t_1))\sin(s_2(t_2))]. \quad (49)$$

If $s_1(t)$ and $s_2(t)$ are temporally correlated, it is highly probable it exists t_1, t_2 such that (49) is not equal to zero (it depends evidently on the nature of the temporal correlation between two successive samples of the two sources) so that the equality (48) is false, and the solution can be rejected. In fact, the two stochastic *processes* $y_1(t)$ and $y_2(t)$ obtained from (14), are not statistically independent although their samples at each time instant (which are two random *variables*) are independent. This simple example shows how, using the temporal correlation, we can distinguish the trivial and nontrivial mappings preserving the independence, or at least cancel a few nontrivial mappings. Note that here we used only the cross-correlation (second order) of the signals, which is a first (but coarse) step toward independence. We also could add more equations for improving the independence test, and consider cross-correlations of order higher than two of $y_1(t_1)$ and $y_2(t_2)$ which must satisfy:

$$E[y_1^p(t_1)y_2^q(t_2)] = E[y_1^p(t_1)]E[y_2^q(t_2)], \quad \forall t_1, t_2,$$
$$\forall_{p,q} \neq 0. \quad (50)$$

5.2.2. Darmois decomposition with colored sources

As we mentioned in Section 2.2, another example used for illustrating the non-separability of the nonlinear mixtures is the Darmois decomposition procedure. Consider two independent and identically distributed random signals, $s_1(t)$ and $s_2(t)$ and suppose $x_1(t)$ and $x_2(t)$ are the nonlinear mixtures of them. Using the Darmois decomposition procedure [16,24], one can construct new signals $y_1(t)$ and $y_2(t)$ which are statistically independent although the underlying mapping is still a mixing (nontrivial) mapping:

$$y_1(t) = F_{X_1}(x_1(t)),$$
$$y_2(t) = F_{X_2|X_1}(x_1(t), x_2(t)). \quad (51)$$

Here F_{X_1} and $F_{X_2|X_1}$ are respectively the marginal and conditional cumulative distribution functions of the observations. If the sources are temporally correlated, Hosseini and Jutten show [20] that the independent components y_1 and y_2 obtained from the above procedure do not generally satisfy the following equality, for $t_1 \neq t_2$, where p_{Y_1} and p_{Y_2}, are the marginal pdfs and p_{Y_1,Y_2} is the joint pdf:

$$p_{Y_1,Y_2}(y_1(t_1)y_2(t_2)) = p_{Y_1}(y_1(t_1))p_{Y_2}(y_2(t_2)) \quad (52)$$

while the trivial transformations of the real sources, in the forms of $y_1 = f_1(s_1)$ and $y_2 = f_2(s_2)$, satisfy obviously the above equality because of the independence of the two sources. Thus, the above equations can be used to reject (or at least to restrict) the nontrivial ICA solutions obtained from the Darmois decomposition.

Of course, this theoretical result does not give any proof for the separability of nonlinear mixtures of temporally correlated sources, but it shows that even fairly weak prior informations on the sources can reduce the typical indeterminacies of ICA encountered in nonlinear mixtures. In fact, with this prior, ICA provides many equations (constraints) which can be used for regularizing the solutions and allow blind source separation.

6. Concluding remarks

In this paper, we have considered ICA and BSS problems for nonlinear mixture models. It appears clearly BSS and ICA are difficult and ill-posed problems, and regularization is necessary for actually achieving ICA solutions which coincide to BSS.

In this purpose, two main ways can be used. First, solving the nonlinear BSS problem appropriately using only the independence assumption is possible only if mixtures as well as separation structure are structurally constrained: for example post-nonlinear mixtures, or mappings satisfying addition theorem (4.4). Second, prior information on sources, for example bounded or temporally correlated sources, can simplify the algorithms or reduce the indeterminacies in the solutions. A third way, based on smooth mappings, is probably suitable, but accurate conditions on the mappings have to be defined.

A lot of work remains to be done in studying the nonlinear ICA and BSS problems. First, regularization methods based on constraints can be studied further, but other approaches, especially incorporation of temporal statistics [49] (only sketched in this paper) and variational Bayesian ensemble learning [45]. Secondly, remember a better modeling of the relationship between the independent components or sources and the observations is essential for choosing a suitable separation structure and subsequently for studying separability. Finally, up to now, the research has addressed mainly theoretical problems. The results will become more widely interesting only if they can be validated on realistic problems using real-world data.

Acknowledgements

This work has been supported by the European Commission project BLISS (IST-1999-14190). The authors thank Dr. A. Taleb for contributing many of the results presented in this paper, and the three anonymous reviewers for their very relevant and detailed comments which contributed to improve this paper.

References

[1] J. Aczel, Lectures on Functional Equations and Their Applications, Academic Press, New York, 1966.
[2] L. Almeida, Linear and nonlinear ICA based on mutual information, in: Proceedings of IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (AS-SPCC), Lake Louise, Canada, October 2000, pp. 117–122.
[3] L. Almeida, C. Jutten, H. Valpola, Realistic models of nonlinear mixtures, BLISS (IST1999-14190) Project Report D5, HUT, INESC, INPG, May 2001.
[4] S. Amari, A. Cichocki, H. Yang, A new learning algorithm for blind signal separation, in: Advances in Neural Information Processing Systems, Denver, Colorado, December 1996, pp. 757–763.
[5] M. Babaie-Zadeh, On blind source separation in convolutive and nonlinear mixtures, Ph.D. Thesis, INPG, Grenoble, France, September 2002.
[6] M. Babaie-Zadeh, C. Jutten, K. Nayebi, Separating convolutive post non-linear mixtures, in: Proceedings of the Third Workshop on Independent Component Analysis and Signal Separation (ICA2001), San Diego, CA, USA, 2001, pp. 138–143.
[7] M. Babaie-Zadeh, C. Jutten, K. Nayebi, A geometric approach for separating post nonlinear mixtures, in: Proceedings of the XI European Signal Processing Conference (EUSIPCO 2002), Vol. II, Toulouse, France, 2002, pp. 11–14.
[8] A. Bell, T. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, Neural Computation 7 (6) (1995) 1004–1034.
[9] A. Belouchrani, K. Abed Meraim, J.-F. Cardoso, E. Moulines, A blind source separation technique based on second order statistics, IEEE Trans. Signal Processing 45 (2) (1997) 434–444.
[10] J.-F. Cardoso, Blind signal separation: statistical principles, Proc. IEEE 9 (10) (1998) 2009–2025.
[11] J.-F. Cardoso, B. Laheld, Equivariant adaptive source separation, IEEE Trans. Signal Processing 44 (12) (1996) 3017–3030.
[12] J.-F. Cardoso, A. Souloumiac, Blind beamforming for non gaussian signals, IEE Proc.-F 140 (6) (1993) 362–370.
[13] A. Cichocki, S.-I. Amari, Adaptive Blind Signal and Image Processing—Learning Algorithms and Applications, Wiley, New York, 2002.
[14] A. Cichocki, R. Unbehauen, E. Rummert, Robust learning algorithm for blind separation of signals, Electron. Lett. 30 (17) (1994) 1386–1387.
[15] P. Comon, Independent component analysis, a new concept? Signal Processing 36 (3) (1994) 287–314.
[16] G. Darmois, Analyse des liaisons de probabilité, in: Proceedings of International Statistics Conferences 1947, Vol. III A, Washington, DC, 1951, p. 231.
[17] Y. Deville, J. Damour, N. Charkani, Multi-tag radio-frequency identification systems based on new blind source separation neural networks, Neurocomputing 49 (2002) 369–388.
[18] J. Eriksson, V. Koivunen, Blind identifiability of class of nonlinear instantaneous ICA models, in: Proceedings of the XI European Signal Processing Conference (EUSIPCO 2002), Vol. 2, Toulouse, France, September 2002, pp. 7–10.
[19] S. Haykin, Neural Networks—A Comprehensive Foundation, 2nd Edition, Prentice-Hall, Englewood Cliff, NJ, 1998.
[20] S. Hosseini, C. Jutten, On the separability of nonlinear mixtures of temporally correlated sources, IEEE Signal Processing Lett. 10 (2) (2003) 43–46.

[21] A. Hyvärinen, P.O. Hoyer, Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces, Neural Computation 12 (7) (2000) 1705–1720.
[22] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.
[23] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Computation 9 (7) (1997) 1483–1492.
[24] A. Hyvärinen, P. Pajunen, Nonlinear independent component analysis: existence and uniqueness results, Neural Networks 12 (3) (1999) 429–439.
[25] C. Jutten, J. Hérault, Blind separation of sources, Part I: an adaptive algorithm based on a neuromimetic architecture, Signal Processing 24 (1) (1991) 1–10.
[26] C. Jutten, A. Taleb, Source separation: from dusk till dawn, in: Proceedings of the Second International Workshop on Independent Component Analysis and Blind Source Separation (ICA2000), Helsinki, Finland, 2000, pp. 15–26.
[27] A. Kagan, Y. Linnik, C. Rao, Extension of Darmois–Skitovic theorem to functions of random variables satisfying an addition theorem, Comm. Statist. 1 (5) (1973) 471–474.
[28] M. Korenberg, I. Hunter, The identification of nonlinear biological systems: LNL cascade models, Biol. Cybern. 43 (12) (December 1995) 125–134.
[29] E. Lukacs, A characterization of the Gamma distribution, Ann. Math. Statist. (26) (1955) 319–324.
[30] S. Makeig, T.-P. Jung, A.J. Bell, D. Ghahramani, T. Sejnowski, Blind separation of auditory event-related brain responses into independent components, Proc. Natl. Acad. Sci. (USA) 94 (1997) 10979–10984.
[31] K. Matsuoka, M. Ohya, M. Kawamoto, A neural net for blind separation of nonstationary signals, Neural Networks 8 (3) (1995) 411–419.
[32] H.L. Nguyen Thi, C. Jutten, M. Kabre, J. Caelen, Separation of sources: a method for speech enhancement, Appl. Signal Processing 3 (1996) 177–190.
[33] A. Parashiv-Ionescu, C. Jutten, G. Bouvier, Source separation based processing for integrated hall sensor arrays, IEEE Sensors J. 2 (6) (December 2002) 663–673.
[34] L. Parra, C.D. Spence, S. Sajda, A. Ziehe, K.-R. Müller, Unmixing hyperspectral data, in: S.A. Solla, T.K. Lee, K.-R. Müller (Eds.), Advances in Neural Information Processing Systems, Vol. 12, MIT Press, Cambridge, 2000, pp. 942–948.
[35] D.T. Pham, J.-F. Cardoso, Blind separation of instantaneous mixtures of nonstationary sources, IEEE Trans. Signal Processing 49 (9) (2001) 1837–1848.
[36] R. Popovic, Hall-Effect Devices, Adam Hilger, Bristol, 1991.
[37] S. Prakriya, D. Hatzinakos, Blind identification of LTI-ZMNL-LTI nonlinear channel models, IEEE Trans. Signal Processing 43 (12) (December 1995) 3007–3013.
[38] T.G. Stockham, T.M. Cannon, R.B. Ingerbretsen, Blind deconvolution through digital signal processing, Proc. IEEE 63 (1975) 678–692.
[39] A. Taleb, A generic framework for blind source separation in structured nonlinear models, IEEE Trans. Signal Processing 50 (8) (2002) 1819–1830.
[40] A. Taleb, C. Jutten, Source separation in post-nonlinear mixtures, IEEE Trans. Signal Processing 47 (10) (1999) 2807–2820.
[41] A. Taleb, J. Sole, C. Jutten, Quasi-nonparametric blind inversion of Wiener systems, IEEE Trans. Signal Processing 49 (5) (2001) 917–924.
[42] F. Theis, E. Lang, Maximum entropy and minimal mutual information in a nonlinear model, in: Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA2001), San Diego, CA, USA, 2001, pp. 669–674.
[43] L. Tong, V. Soon, Y. Huang, R. Liu, AMUSE: a new blind identification algorithm, in: Proceedings of IEEE International Symposium on Circuits and Systems (ISCAS'90), New Orleans, LA, USA, 1990.
[44] K. Torkkola, Blind separation for audio signals—are we there yet? in: Proceedings of the International Workshop on Independent Component Analysis and Signal Separation (ICA1999), Aussois, France, 1999, pp. 239–244.
[45] H. Valpola, J. Karhunen, An unsupervised ensemble learning method for nonlinear dynamic state-space models, Neural Computation 14 (11) (2002) 2647–2692.
[46] J.H. van Hateren, D.L. Ruderman, Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex, Proc. R. Soc. Ser. B 265 (1998) 2315–2320.
[47] R. Vigário, J. Särelä, V. Jousmäki, M. Hämäläinen, E. Oja, Independent component approach to the analysis of eeg and meg recordings, IEEE Trans. Biomed. Eng. 47 (5) (2000) 589–593.
[48] H. Yang, S.-I. Amari, A. Cichocki, Information-theoretic approach to blind separation of sources in non-linear mixture, Signal Processing 64 (3) (1998) 291–300.
[49] A. Ziehe, M. Kawanabe, S. Harmeling, K.-R. Müller, Separation of post-nonlinear mixtures using ACE and temporal decorrelation, in: Proceedings of the International Conference on Independent Component Analysis and Signal Separation (ICA2001), San Diego, USA, 2001, pp. 433–438.